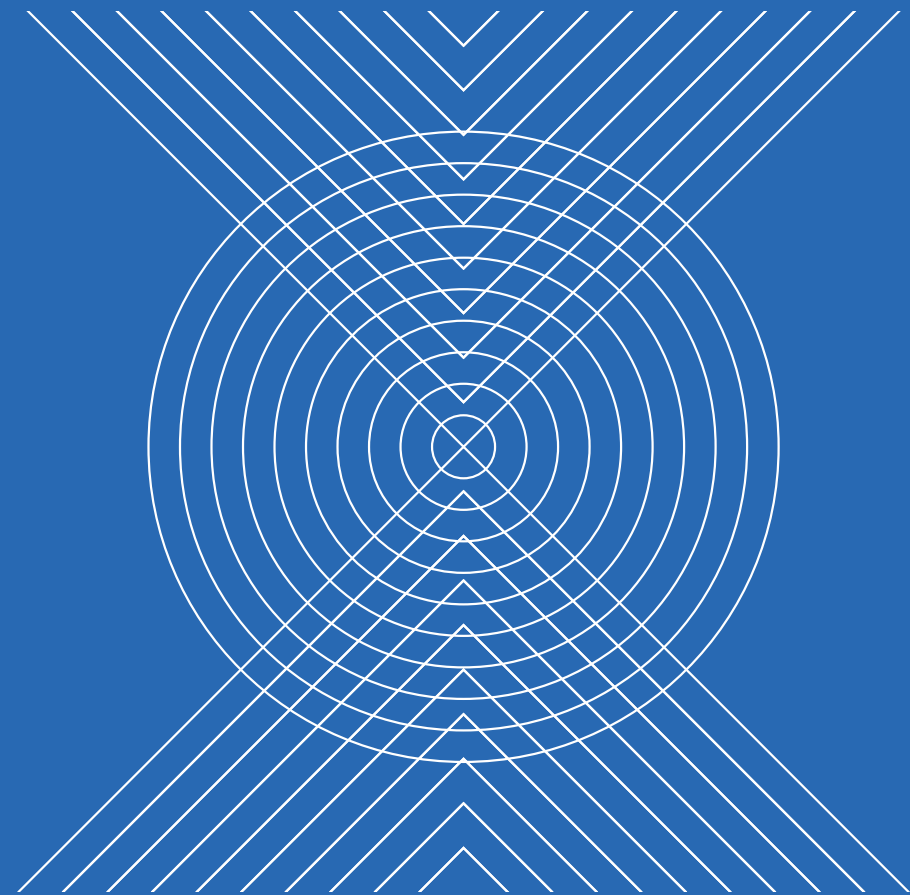


Environmental Engineering and Computer Application

Editor: Kennis Chan



Editor
Chan

Environmental Engineering
and Computer Application

Environmental Engineering and Computer Application includes up-to-date information on environmental and computer related issues. The book is divided on 5 sections:

- Environmental engineering and resources;
- Remote sensing, GIS and applications;
- Power electronics and energy engineering;
- Chemical, Biological and Agricultural Engineering, and
- Computer Algorithm, Simulation and Information Technologies

The book will be of interest to academics and professionals involved in the above mentioned subject areas.



CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
Schipholweg 107C
2316 XC Leiden, NL
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK



ISBN 978-1-138-02807-4



9 781138 028074

an informa business



CRC Press
Taylor & Francis Group
A BALKEMA BOOK

ENVIRONMENTAL ENGINEERING AND COMPUTER APPLICATION

PROCEEDINGS OF THE 2014 INTERNATIONAL CONFERENCE ON ENVIRONMENTAL
ENGINEERING AND COMPUTER APPLICATION (ICEECA2014), HONG KONG,
25–26 DECEMBER 2014

Environmental Engineering and Computer Application

Editor

Kennis Chan

Advanced Science and Industry Research Center, Hong Kong



CRC Press

Taylor & Francis Group

Boca Raton London New York Leiden

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A BALKEMA BOOK

CRC Press/Balkema is an imprint of the Taylor & Francis Group, an informa business

© 2015 Taylor & Francis Group, London, UK

Typeset by diacriTech, Chennai, India

All rights reserved. No part of this publication or the information contained herein may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording or otherwise, without written prior permission from the publishers.

Although all care is taken to ensure integrity and the quality of this publication and the information herein, no responsibility is assumed by the publishers nor the author for any damage to the property or persons as a result of operation or use of this publication and/or the information contained herein.

Published by: CRC Press/Balkema
P.O. Box 11320, 2301 EH Leiden, The Netherlands
e-mail: Pub.NL@taylorandfrancis.com
www.crcpress.com – www.taylorandfrancis.com

ISBN: 978-1-138-02807-4 (Hardback)
ISBN: 978-1-315-68538-0 (eBook PDF)

Table of contents

| | |
|-----------------------|------|
| Preface | xi |
| Organizing committees | xiii |

1. Environmental engineering and resources

| | |
|---|----|
| Environmental benefits of China's conversion of cropland to forest program and its viability: A case study of Meixian County of Shaanxi Province <i>H.Y. Han & S. Xia</i> | 3 |
| Traditional forest knowledge, rubber plantation and climate change <i>Y.H. Du, L. Guo, D.Y. Xue, F. Liu & X.Z. Yu</i> | 9 |
| Economic and thermal technical comparison of classical foundation and foundation on air gap–crawl space <i>Š. Korbelová, M. Vodičková & V. Šípková</i> | 15 |
| Gold refining process and its impact on the environment <i>S. Kaspin & N. Mohamad</i> | 19 |
| Semiempirical modeling of iron oxide pollution in the mine soil using geochemical analyses and spectroradiometric measurements <i>A. Khalil & A. Bannari</i> | 23 |
| Research on the strategy of controlling Ordos industrialization and eco-environment system based on SD model <i>L.M. Li, J.H. Zhen, S.S. Luo & S.P. He</i> | 29 |
| Environmental aspects of raw materials SMART processing <i>J. Spišák & M. Zelko</i> | 35 |
| Variation of extreme Total column ozone events over Seoul, South Korea <i>D. Amrita, J.K. Park & J.H. Park</i> | 41 |
| Application of artificial neural network ensembles for city ecology forecasting using air chemical composition information <i>D.I. Khritonenko & E.S. Semekin</i> | 45 |
| An improved method for calculating atmospheric refraction coefficient for eliminating influence of meteorological factors <i>L. Liu, C.L. Mi & J.S. Han</i> | 51 |
| Cerium-induced genotoxicity in Arabidopsis at environmentally relevant doses <i>L.Z. Sun, P.D. Tai, W. Liu, X.J. Li, C.Y. Jia, Z.H. Liu & X. Wang</i> | 55 |
| Determination of the cycle duration for 155 years in the Nikolaevsk-on-Amur meteorological station <i>E.E. Kholoden, O.M. Morina & S.A. Lobanov</i> | 59 |
| Soil erosion spatial distribution simulation of the Jinghe river basin in the Loess Plateau of China <i>G.S. Wu, D.L. Wang, Y.C. Wu, X.Y. Chen & W.S. Wei</i> | 63 |

| | |
|---|-----|
| Predicting sinking resistance of jacked piles with CPT in a multilayer soil at the pearl river delta alluvial plain <i>Y.Q. Hu, L.S. Tang, H.T. Sang & Z.Z. Li</i> | 71 |
| Numerical simulation of oil spill caused by ship collision in Dongjiakou Port <i>R.J. Hu, N. Wang, S.H. Jiang, L.H. Zhu, J.Z. Wu & F. Ma</i> | 77 |
| Statistics and analysis on causes for well blowout accident <i>W.F. Ge, L.F. Qian, F.B. Song, F. Zhang, H.G. Liu, X. Zhang & G.Y. Zhao</i> | 81 |
| Water quality and growth trend of aquatic plant duckweed: Management and benefits <i>M.W. Shammout & H. Zakaria</i> | 85 |
| Estimation of influence of mine waters of cassiterite-sulfide deposits on hydrosphere in Komsomol'sk and Kavalerovo districts of the far east <i>V.P. Zvereva, E.A. Yastremskaya & A.D. Pyatakov</i> | 89 |
| Effect of sodium sulfite and humic acid on CR(VI) content in artificial wastewater with or without CR-hyperaccumulator <i>Leersia hexandra</i> Swartz <i>Z.G. Cheng, X.F. Zhang & C. Tian</i> | 95 |
| Low cost adsorbent to reduce disinfection by-products from drinking water in small communities <i>T. Husain & M. Ahmad</i> | 99 |
| A submersible automatic monitoring system of water quality employing a spectral scanning sensor <i>Y.Q. Zhao, X. Li, J. Gu, F. He & Y.J. Fang</i> | 105 |
| Characteristics of dissolved trace elements in the Jinshui River of the South Qinling Mts., China <i>H.M. Bu & Q.F. Zhang</i> | 111 |
| Progress and prospect of migration character about insoluble sedimentary pollutants <i>L.X. Han, F.F. Yan & B. Chen</i> | 115 |
| Toward a more sustainable future: A critique on the mainstream of coastal tourism development since the 1960s <i>H. Zhang, Y.B. Wei & C.H. Fu</i> | 121 |
| Design of earthquake-data processing platform <i>S.Y. Zhang, Y. Li & Y. Zhao</i> | 127 |
| Activated sludge model parameters calibration based on multi-objective PSO <i>R.R. Wang, L.L. Cao & M.D. Liu</i> | 133 |
| Optimization of hexavalent chromium (Cr(VI)) reduction by granulated nano zero-valent iron (nZVI) <i>Y.J. Shih, C.W. Chen, K.F. Hsia & C.D. Dong</i> | 139 |
| Effect of calcination temperature and time of thermo-hydrolytic TiO ₂ on photocatalytic degradation of methylene blue <i>C.W. Chen, C.C. Su, Y.C. Chen & C.D. Dong</i> | 145 |

2. Remote sensing, GIS and application

| | |
|--|-----|
| Simulation of soil moisture fluctuation based on the signal-to-noise ratio of GPS <i>R.H. Hu, Y.J. Hu & X.Y. Yu</i> | 153 |
| Sorted run length coding-application to Meteosat image compression <i>M. Cherifi, M. Lahdir & S. Ameer</i> | 159 |
| Application of GIS to Tibetan Buddhist monastery data platform <i>Z.M. Caiji, L. Guo, D.Y. Xue & Y.J. Wang</i> | 165 |

| | |
|---|-----|
| The application of GIS in electric vehicle charging/battery swap infrastructure and operation <i>D.S. Bai, G. Chen, X.Q. Li, J. Liu, W.F. Li & W.N. Yuan</i> | 169 |
| Validation of LMSA and SAM for geological mapping: Comparative study among BGIS-2000, TM and ASTER <i>A. Bannari, G. Girouard & A. Saquaque</i> | 173 |
| Monitoring of disaster and restoration of eastern Japan tsunami by satellite images <i>H. Hashiba</i> | 179 |
| Application of PCA and KLE in high-rate GPS positioning <i>X.Y. Yu, Y.J. Hu, K.H. Ding & M.S. Ao</i> | 187 |
| Spatio-temporal processing of geoinformation for decision making in critical situations <i>K. Janecka, L. Berardi & A. Cimbelli</i> | 193 |
| Spatial distribution of ethnic villages in Qiandongnan mountainous region <i>D. Wang, L. Guo, B.S. Wu & L. Lv</i> | 205 |
| Study of AIS beacon control system based on ARM <i>Q. Hu, X.J. Li & S.F. Zhang</i> | 211 |

3. Power electronics and energy engineering

| | |
|--|-----|
| Stochastic optimal dispatch model considering wind turbine shut down events under extreme wind condition <i>Y.X. Zhang, Z.Y. Dong & H.M. Yang</i> | 219 |
| A machine learning approach to increase energy efficiency in district heating systems <i>F. Dalipi, S.Y. Yayilgan & A. Gebremedhin</i> | 223 |
| Modeling of hypergene processes proceeding in the tailing dumps of Krasnorechenskaya concentrating mill of Dalnegorsky district in the temperature range of 0 to 45°C using new thermodynamic data <i>V.P. Zvereva, A.M. Kostina & A.I. Lysenko</i> | 227 |
| Research on steel industry, energy-saving efficiency evaluation of Tangshan <i>X.H. Wan, H.L. Wang & L.H. Zhou</i> | 233 |
| Selected aspects of wind- power plants building influence on the operation of radio communications system <i>M. Džunda & N. Kotianová</i> | 239 |
| Development of a solar cell robot optimization program GUI <i>M.J. Kim, T.W. Park, J.H. Lee, K.S. Sim, I.K. Hwang & D.I. Park</i> | 243 |
| Research and design of a portable phase detector with power source based on the phase difference <i>T.K. Sun, D. Zhang, W. Liu & W. Liu</i> | 247 |
| Three-level full-bridge bi-directional DC-DC converter <i>C. Yin, Y. Rao & K. Jin</i> | 251 |
| Fabrication of <i>n</i> -DLC/ <i>i</i> -SiC/ <i>p</i> -Si photovoltaic cells by pulsed laser deposition at the electrical field <i>S. Kurumi, K. Suzuki & S. Yonekawa</i> | 255 |
| Applying the technique of data analysis to estimate the electrical consumption for household in Taiwan <i>F.F. Chen & T.K. Lu</i> | 259 |
| Design of electric vehicle charging battery swap station power quality data transmission system based on lab view <i>D.S. Bai, G. Chen, X.Q. Li, J. Liu, W.F. Li & H. Xing</i> | 263 |
| Research on intelligent integrated charging equipment for electric vehicle <i>X.Q. Li, H.L. Wang, S.Y. Li & K.X. Li</i> | 267 |

| | |
|---|-----|
| Design and development of driver protection circuit of self turn off devices of electric vehicle DC charger | 271 |
| <i>D.S. Bai, X.Q. Li, G. Chen, J. Liu, W.F. Li & H.L. Wang</i> | |

4. Chemical, biological and agricultural engineering

| | |
|---|-----|
| Drying of agricultural products using thin layer principle furnace | 279 |
| <i>J. Spišák, A. Olijár, M. Reisová & E. Oravcová</i> | |
| Effect of gamma irradiation on saccharification of <i>Undaria</i> sp. biomass | 285 |
| <i>J. Choi</i> | |
| <i>Pythium oligandrum</i> colonization enhances wheat growth and resists disease caused by <i>Pythium arrhenomanes</i> | 289 |
| <i>Y. Zhu, Z.Y. Wang, H.S. Qi, H. Yang, J.P. Zhou & Z.B. Wang</i> | |
| Hyperspectrum detection for nitrogen content of apple leaves | 295 |
| <i>J. An, G.Q. Yao & H.L. Deng</i> | |
| Effect of heavy metals on growth and metal accumulation of an energy crop, <i>Pennisetum purpureum</i> schum | 299 |
| <i>C. Tian, X.F. Zhang, Y.H. Su, P.H. Tan, H.F. Zhang & G.Z. Zhang</i> | |
| Research in disproportionation reaction mechanism of methyltrichlorosilane and trimethylchlorosilane catalyzed by modified catalyst AlCl ₃ /ZnCl ₂ -ZSM-5 | 303 |
| <i>W.Y. Xu, J.X. Zhou, Y.P. Liu, H.H. Wan & S.G. Hong</i> | |
| The correlation studies about histone modification and H2A. Z nucleosome positioning | 307 |
| <i>Z.H. Shan, J.H. Feng, P.F. Chen, H.H. Wei & H. Hu</i> | |
| Molecular cloning, overexpression, purification, sequence analysis of matrix Gla protein from the Giant Panda (<i>Ailuropoda melanoleuca</i>) | 311 |
| <i>P.P. Wang, X. Ding, Y.L. Hou, Y.X. Zhu, H.Q. Zhu, L. Fu & N. Zhang</i> | |
| The kinetic and thermodynamic study on the OH-initiated degradation mechanism of polychlorinated dibenzo- <i>p</i> -dioxins | 315 |
| <i>X. Zhang, C.X. Zhang, L.Y. Kang, C. Gon, X.M. Sun & J. Yu</i> | |
| Thermal, electrical and mechanical properties of Lead-free Tin solder alloys containing Bi and In | 319 |
| <i>S. Attia</i> | |

5. Computer algorithm, simulation and information technologies

| | |
|---|-----|
| A deletion cost based compression algorithm for linear vector data | 329 |
| <i>Q.X. Chen, Y. Hou & N. Wu</i> | |
| A design of highly-available and cost-effective data center infrastructure | 335 |
| <i>Y.J. Chen & W.J. Hsu</i> | |
| A method to detect memory leak of C++ class field member based on static analysis | 343 |
| <i>S.W. Liu & D.H. Jin</i> | |
| SSD lifespan-oriented cache optimization in SSD/HDD hybrid storage system | 347 |
| <i>X. Wang, Q.X. Zhang, Y.A. Tan & J.Q. Qu</i> | |
| A collaborative filtering algorithm based on the improved similarity measure method and rating matrix pre-filling | 353 |
| <i>D.W. Cai, Y. Liu & T. Cao</i> | |
| Testing the radio proximity fuse with hardware-in-loop simulation | 359 |
| <i>Y.J. Teng & H.Y. Liu</i> | |

| | |
|---|-----|
| An improved moving object tracking algorithm based on Camshift for dynamic background <i>G.W. Yuan, H. Zhou, D. Xu, Y. Zhao & Q.S. Tao</i> | 363 |
| Study on sustainability and life cycle assessment of RFID antenna <i>R.K. Kanth, P. Liljeberg, H. Tenhunen, L.R. Zheng, Y. Amin & H. Kumar</i> | 369 |
| Probabilistics data model based on privacy tipping points <i>M.V. Rallapalli</i> | 373 |
| Research on hybrid parallel computing model and key technology <i>J.P. Dai, L.P. Zhu, B. Hu & A.H. Zhou</i> | 377 |
| Scheme study of construction of TSMDB <i>H.K. Tang, M.J. Liu, Y.L. Fan & C.W. Li</i> | 381 |
| Damage diagnosis of derrick steel structures based on vibration parameters and improved genetic algorithm <i>D.Y. Han, S.M. Wei, N. Zhang & P.M. Shi</i> | 385 |
| A resource allocation economic model for Diffserv network <i>B. Zeng, L. Yao & W. Hu</i> | 391 |
| A study of active steering system for safety of a railway vehicle <i>K.S. Sim, T.W. Park, I.K. Hwang & M.J. Kim</i> | 397 |
| A development of the wheel-rail dynamic analysis program model with two-point contact algorithm <i>I.K. Hwang, T.W. Park, K.S. Sim & M.J. Kim</i> | 401 |
| Tight frames and iterative shrinkage based measurement matrix optimization for compressive sensing <i>E. Oey</i> | 405 |
| Amplitude control of the flow tube based on PID control <i>J. Yu & Z.B. Gui</i> | 411 |
| The noise reduction by the application of the vibration isolators <i>A.A. Igolkin, L.V. Rodionov & F.S. Pomatilov</i> | 417 |
| Intelligent management platform of forestry based on LBS cloud services <i>F. Li, Y. Ma, X. Zhang, M.B. Zhang, X.W. Yu & P.F. Feng</i> | 423 |
| Focus+context visualization with multi-resolution meshes <i>Y.P. Zha, Z.J. Li, Y.M. Liu, F.X. Li & K. Yang</i> | 427 |
| Modeling channel variations at 433MHz for wireless sensor networks in indoor scenarios <i>M.W. Khan, T. Zahid, Y.M. Zhou & G.Q. Xu</i> | 433 |
| A key management scheme for wireless sensor networks based on bridge and ECC <i>J. Zheng, F.F. Zhao, D.X. Wang & S.C. Xu</i> | 439 |
| Proposal of a universal model for managing research and development center <i>J. Spišák, M. Reisová, E. Oravcová & S. Poláková</i> | 447 |
| Research on human resources complex standardization system of large state-owned enterprises <i>L. Zhao</i> | 453 |
| How to cope with the challenges of internet-based finance: New finance mode of commercial bank in China <i>P.H. Wang</i> | 457 |
| The empirical study of construction of public rental based tenants' attention <i>X.M. Yang & L.L. Zhao</i> | 463 |
| Rhyme of ink: To composite effects of ink painting in after effects <i>Z.Y. Li, J. Kang & R.P. Cao</i> | 469 |
| Author index | 475 |

Preface

The awareness of environment protection is for sure another great achievement of human's; it is an expression of our self-awareness. Even though the idea of living while protecting the environment is not new, it has never been so widely and deeply practiced by any nations in history like it is today. From the late 90s in the last century, the surprisingly fast development of computer science has brought us so many new technologies which have already been applied in countless fields like manufacturing, military, aerospace and pharmacy. Now, another important field-the environment engineering is welcoming the introduction of these technologies.

The 2014 International Conference on Environment Engineering and Computer Application (ICEECA2014), by bringing together researchers in environment fields and in computer fields, is an important attempt to find potential uses for computer technologies in environment protection fields.

ICEECA2014 has been successfully held on December 25–26, 2014 in Hong Kong; it is a grand gathering attended by experts and researchers from all around the world. The participants at ICEECA2014 have generously shared their inspiring work with each other and exchanged their innovative ideas of how to help the environment science benefit from the development of computer technology and vice versa. The presentations at the convention also gave people much to think about.

This book, as one of the many accomplishments for ICEECA2014, is the cooperative work of many people's and organizations'. It comprises all the papers accepted by ICEECA2014. These papers are accepted after our reviewers' careful reading. I am sure that they will prove to be valuable in other's researches in the future.

Papers in this book are put under several different chapters. Under each chapter, one specific aspect was discussed in depth. By discussing different aspects separately, we are able to get a more thorough understanding of them all. These chapters include subjects like environmental engineering, remote sensing, GIS and application, graphic and image processing and computational science and application.

Lastly, I would like to acknowledge the contribution and help for ICEECA2014 from many people. First, I deeply thank all the contributors in this book; they have shown us the typical spirit of passion and devotion to science which all researchers and scholars should possess. Secondly, I appreciate the support from the CRC Press for publishing this book and the organizers and the committee for organizing ICEECA2014. Finally, those who have helped us in our preparation for ICEECA2014 but are not mentioned above also deserve my sincerest gratitude.

Organizing committee

GENERAL CHAIR

J. Yeh, *Tallinn University of Technology, Estonia*

PROGRAM CHAIR

C.J. Lee, *Science and Engineering Research Center, Hong Kong*

PROGRAM CHAIR

Kennis Chan, *Advanced Science and Industry Research Center, Hong Kong*

INTERNATIONAL SCIENTIFIC COMMITTEE

Michela Longo, *Politecnico di Milano, Italy*
H. Yaghoubi, *Director of Iran Maglev Technology (IMT), Iran*
P. Velayutham, *Sathyabama University, India*
Wen-Tsai Sung, *National Chin-Yi University of Technology, Taiwan*
Wei-Chen Wu, *Hsin Sheng College of Medical Care and Management, Taiwan*
Arabinda Sharma, *BRCM College of Engineering & Technology, India*
Veera Jyothi. B, *Chaitanya Bharati Institute of Technology, India*
M. Subramanyam, *National Institute of Technology, India*
M.S. Chen, *Da-Yeh University, Taiwan*
M.V. Raghavendra, *Adama Science & Technology University, Ethiopia*
J. Ye, *Hunan University of Technology, China*
Q. Yang, *University of Science and Technology Beijing, China*
X. Lee, *Hong Kong Polytechnic University, Hong Kong*
K. Weller, *Science and Engineering Research Center, Hong Kong*
J. Xu, *Northeast Dianli University, China*
Z.Y. Jiang, *University of Wollongong, Australia*
V.K. Jain, *Indian Institute of Technology, India*
Q. Zhang, *Shenzhen University, China*
O.P. Rishi, *University of Kota, India*
K.S. Rajesh, *Defence University College, India*
M.M. Kim, *Chonbuk National University, Korea*

1. Environmental engineering and resources

Environmental benefits of China's conversion of cropland to forest program and its viability: A case study of Meixian County of Shaanxi Province

H.Y. Han & S. Xia

COLLEGE OF MANAGEMENT/Center for Agricultural and Rural Development, Hangzhou, Zhejiang, China

ABSTRACT: The viability of China's Conversion of Cropland to Forest Program (CCFP) depends on farmers' subjective perceptions due to its "quasi-voluntary" characteristics. Farmers' willingness to pay for CCFP conservation is estimated based on a contingent valuation method. Although farmers have benefited from environmental improvements in terms of air purification and water conservation, there is a gap between farmers' perceived environmental services and government subsidy. It is critical for policy design to integrate the environmental goals with agricultural restructuring and poverty reduction.

1 INTRODUCTION

The Conversion of Cropland to Forest Program (CCFP), or Grain-for-Green, is an example of an ambitious environmental services program in China. Like many ongoing agri-environmental payment programs in Asia that aim to increase forest cover and promote community development (Toma et al., 2004), the CCFP is facing issues such as government budgetary limitations and inefficiencies in its implementation and monitoring mechanisms (Bennett and Wang, 2008). Under budget constraints, the critical challenge for the CCFP is to maintain the program's integrity in addressing the multiple objectives of alleviating poverty, increasing agricultural incomes, and reducing sediment inflows to the Yellow River (State Council, 2007).

Yet there is little evidence to indicate that Asian policy-makers have taken into account the benefits of environmental improvement (Dudgeon, 2000). The overall goal of this paper is to examine the viability of the CCFP by examining the participating farmers' subjective evaluations of the CCFP's environmental value by using data from 237 rural households. It is shown that the environmental benefits perceived by local farmers are less than the expected subsidies from the central government, and participating farmers' subjective perceptions of the environmental gains from the CCFP do not indicate the long-term viability of the CCFP.

2 BACKGROUND OF THE PAPER

2.1 *Study site*

Due to a shortage of grain and lack of other economic opportunities, farmers in western China have converted vast and steeply sloped tracts of forest and grassland into farmland over the past decades (Feng et al., 2005). Seventy percent of the 6.07 million

hectares of agricultural land that is situated on slopes >25 degrees is located in the Yangtze River and Yellow River regions (PTFDSSD, 2003). Two-thirds of the two billion tons of mud and sand that flow into the Yangtze and Yellow Rivers annually is attributed to steep land along the river valleys.

The reasons for selecting Meixian County as the study area are as follows. First, the CCFP initiated a pilot site in Meixian County of Shaanxi Province in 1999. As an agricultural products base, the annual household income per capita in Meixian County was 2,811 CNY (US\$399, US\$1 = 8.3 CNY; Meixian Bulletin of Economics and Social Development 2007, Meixian Statistical Bureau). By 2007, 4,639 hectares were enrolled in the CCFP across Meixian County. A further goal of the CCFP, in addition to environmental conservation, was to restructure the rural economy so that participating farmers could gradually shift into more environmentally and economically sustainable activities such as livestock breeding and off-farm work (SFA, 2003).

Second, the CCFP planned to pay farmers cash and grain subsidies for five years for commercial trees, and eight years for ecological trees on sloping farmland. The interviewed households in Meixian County began taking part in the CCFP at the end of 2002 by planting ecological trees. The enrolled land reached a peak of 1,333 hectares at the end of 2002. The subsidization of enrolled lands is due to end soon. According to the State Council notice on improving the CCFP issued on August 9 August 9, 2007, half of the previous cash subsidy will be provided for the renewal of converted lands. Whether or not farmers will reconvert the forest land back into crop farming will be determined by both the economic situation of participating farmers and by their subjective perceptions of the environmental benefits from the CCFP.

2.2 Questionnaire design

To examine whether the CCFP is an effective means of conserving land and meeting environmental targets, a three-part questionnaire is designed. The first section we collect background information about households. The second section we are concerned with the relationship between the CCFP and environmental services, and illustrate the current state of the ecosystem and management methods. The third section of the survey we ask farmers about their perceptions of the CCFP and their WTP for conserving forests. A hypothetical situation is put forward in which the CCFP is assumed to be partly supported by residents' donations due to government budget limitations. The survey respondents are then asked to choose among different options. Payment cards present a range of monetary values, beginning with 5 CNY and increasing in fixed intervals up to the maximum of 175 CNY.

2.3 Data collection

The contingent valuation method (CVM) involves surveying a sample of people who are likely to be affected by proposed changes in the environment to inquire if they are willing to pay a specified amount to secure a specific change or to avoid it. Prior to the formal field survey, focus group discussions were held and a pretest was conducted in November 2008 at Huaiya Township, Meixian County.

The respondents' WTP was ascertained by asking which the suggested amounts best described the

maximum they would be willing to pay in the form of a donation over the next three years.

3 THE VIABILITY OF THE CCFP

3.1 How much are farmers willing to pay for environmental services arising from the CCFP?

Farmers' subjective perceptions of the environmental contributions provided by the CCFP (Sharp and Bromley, 1979) could play a role in their decisions regarding converted land. The most widely used empirical method for measuring the demand for non-market public goods is the contingent valuation method (Morrison, 1997). To estimate farmers' WTP for the conservation of forests, a cumulative logistic probability function is employed as follows:

$$p_{yes} = \frac{e^{\alpha + \beta y}}{1 + e^{\alpha + \beta y}}, \ln \frac{p_{yes}}{1 - p_{yes}} = \alpha + \beta y + r \cdot BID$$

where p_{yes} is the probability of payment, α and β are coefficients, and $y = 1$ denotes yes, otherwise 0.

$$WTP_{mean} = \int_0^{BID_{Max}} \frac{e^{\alpha + \beta s + r \cdot BID}}{1 + e^{\alpha + \beta s + r \cdot BID}} = \frac{1}{r} \ln \frac{1 + e^{\alpha + \beta s + r \cdot BID_{MAX}}}{1 + e^{\alpha + \beta s}}$$

SPSS 16.0 was employed in our regression analysis. The regression results are presented in Table 2.

Of the 237 respondents, 186 indicated a positive WTP amount; 61 of these were willing to make a financial donation, and 125 were willing to contribute through obligatory working days, reflecting the lack of work opportunities for farmers outside the agricultural sector. The opportunity cost for local workers is 30 CNY per day, making the two forms of WTP interchangeable.

The average WTP per household is calculated as 146 CNY. At the end of 2008, there were 77,194 households in Meixian County; thus, with 78.5 per cent of householders indicating a positive WTP, the total WTP amounts to 8,847,204 CNY per year. By comparison, the total area of the CCFP was 4639 ha in Meixian County, making the total expected subsidy from the central government $4639 \times 2400 = 11,133,600$ CNY per year.

Because farmers perceive the amount of environmental value to be less than the opportunity costs in terms of forgone net revenue, it could lead to the discontinuation of the CCFP in Meixian County. Even without taking into account possible yea-saying, there is a gap between farmers' WTP for environmental services and the subsidy provided by the Chinese Government. Without sufficient subsidy, the farmers' perceived environmental

Table 1. Questionnaire.

| | | | | | | | |
|---|-----------------------------|-----|----|-----|----|-----|-----|
| Without maintenance and operation, there will be a loss of environmental services provided by local planted forests. Would you be willing to make a contribution environmental improvement in the form of donation? | | | | | | | |
| Yes <input type="checkbox"/> | No <input type="checkbox"/> | | | | | | |
| If no, the reason is (please tick): | | | | | | | |
| <input type="checkbox"/> My household cannot afford to pay | | | | | | | |
| <input type="checkbox"/> I have little interest in these environmental services | | | | | | | |
| <input type="checkbox"/> The government should pay for it rather than me | | | | | | | |
| <input type="checkbox"/> I am not confident of the purpose of my payment | | | | | | | |
| <input type="checkbox"/> Others | | | | | | | |
| If yes, which of the amounts listed below best describes your maximum willingness to pay in the coming three years? | | | | | | | |
| 15 | 30 | 45 | 60 | 75 | 90 | 150 | 210 |
| If your donation is given in the form of obligatory labor, how days do you want to in the coming three years? | | | | | | | |
| 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 5 | 7 |

Table 2. Results of the CVM analysis.

| Explanatory variables | Unstandardized coefficient | S.E | B index |
|--|----------------------------|-------|---------|
| Bid (Initial payment) | -0.019*** | 0.004 | 0.981 |
| Leader (0 = Ordinary household, 1= Household with a leader or member of the communist party) | 2.320** | 1.128 | 10.172 |
| Participation (1 = CCFP participants, 0 = non-participants) | 1.224** | 0.481 | 3.402 |
| Sex (1 = male, 0 = female) | -0.751 | 0.532 | 0.472 |
| Importance (1 = belief in the positive effect of the on the local environment, 0 = others) | 1.459*** | 0.521 | 4.303 |
| Education1 (4–6 years, 0 = others) | 0.352 | 0.713 | 1.421 |
| Education2 (7–9 years, 0 = others) | 0.286 | 0.641 | 1.331 |
| Education3 (10–12 years, 0 = others) | 2.0722** | 0.902 | 7.942 |
| Age1 (1 = 31-years old, 0 = others) | -0.424 | 1.159 | 0.654 |
| Age (1 = above 60 years old, 0 = others) | -0.869 | 1.611 | 0.419 |
| Income1 (1 = 1000–2000 CNY, 0 = others) | -1.102 | 0.859 | 0.332 |
| Income2 (1 = 2000–3000 CNY, 0 = others) | -1.117 | 0.825 | 0.327 |
| Income3 (1 = 3000–6000 CNY, 0 = others) | 0.888 | 0.867 | 2.431 |
| Constant | 1.772 | 1.646 | 5.884 |

Note: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

benefits could not be a valuable indicator of the long-term viability of the CCFP.

3.2 The viability of the CCFP

The field survey shows that non-participants in the CCFP had higher incomes than those of participants in both 2003 and 2008. The income analysis illustrates that the average household real net income decreased after participating in the CCFP (see Table 3). This conflicting information demonstrates the urgent need to examine the reasons underlying the changes in farmers' incomes.

One source is the increase in crop price. According to the field survey, in 2003 the prices for maize and wheat at the farm gate were 1.12 and 1.2 CNY/kg, respectively, compared to 1.2 and 1.8 CNY/kg in 2008. Farmers plant wheat as much as they can in response to market signals to offset the effect of the reduced crop area on their family incomes. The increase in price, rather than the CCFP, seems to be responsible for the increase in household incomes in different time periods.

Afforestation in Meixian County does not generate direct income for local farmers and the absence of cash income sources at the end of the program

Table 3. Per capita net income of households participating and not participating in the CCFP, 2003 and 2008 (unit: CNY).

| Income Component | (1) Non participants | | (2) Participants | | (3) Income difference T-value | |
|--------------------------|----------------------|----------------------|---------------------|----------------------|-------------------------------|-------------------|
| | 2003 | 2008 | 2003 | 2008 | 2003 | 2008 |
| Cropping without subsidy | 691.15 (572.54) | 794.63 (697.18) | 578.71 (417.28) | 753.68 (709.32) | 112.44 (1.74*) | 40.95 (0.43) |
| Cropping with subsidy | N.A | 864.18 (733.78) | N.A | 820.81 (723.89) | N.A | 43.38 (0.44) |
| Livestock farming | 67.37 (133.06) | 189.82 (571.46) | 111.91 (252.19) | 286.10 (642.67) | -44.53 (-1.51) | -96.28 (-1.15) |
| Off-farm | 1059.80 (1004.0) | 2206.00 (1805.63) | 718.11 (-838.25) | 1842.00 (1676.58) | 341.7 (2.80)*** | 364.00 (1.56) |

Note: 1.(1) and (2)present the means and (S.E.) 2.(3)values in brackets are T-values, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

may induce the reconversion of the set-aside plots. Economic viability may be possible if households find more profitable uses for their additional labor (SFA, 2003). According to the government, the CCFP not only aims to conserve soil and water in China's ecologically fragile areas, but also aims to restructure the rural economy so that participating farmers can gradually shift into more environmentally and economically sustainable activities, such as livestock breeding and off-farm work. Table 4 indicates that the livestock production income of participants increased from 111.91 CNY in 2003 to 286.1 CNY in 2008. By comparison, the livestock farming income of non-participants increased from 67.37 CNY in 2003 to 189.82 CNY in 2008.

There is little evidence that the program has induced participants to shift agricultural production from cropping towards animal husbandry and other agricultural activities. Compared with non-participants in the CCFP, the planted area of participants' cash crops increased from 0.08 ha in 2003 to 0.15 ha in 2008 and their cropping land-holdings declined from 0.33 ha to 0.28 ha per household. The completion of the CCFP has resulted in a decrease in the current cultivated area (Bennett, 2008).

Table 3 indicates that in both 2003 and 2008 the off-farm income of participants in the CCFP—the portion of household income obtained off the farm in terms of non-farm wages—was less than that of non-participants. In particular, in 2008 the non-agricultural income of participants was 19.76% less than that of non-participants in the CCFP because participating farmers had to carry out daily maintenance of the planted trees to ensure full compensation, which was conditional on the trees' survival rates, making it impossible for farmers to work outside of the agricultural sector. This is why, even with full compensation, the total income of participating farmers in 2008 was less than that of non-participating farmers.

Participants in the CCFP have suffered financial losses because of limited structural changes and the reduction of land available for planting crops. Our concern is that due to these constraints, participants may decide to exit the program and reconvert the set-aside land back to crop farming after the program ends, which will result in a net loss of environmental services. This outcome was clearly illustrated in Ningxia, where 79% of 316 participating households indicated that they would not have enough grain after the subsidized period, and only 8% said they would not reconvert their land back to cropland once the compensation ended (Yeh, 2009). If the program does not fairly compensate farmers, the enrolled land may be returned to crop farming and the conservation benefits could not be sustained (Chen et al., 2009).

4 CONCLUSION

Many environmental conservation programs fail to achieve their goals of ecosystem conservation because they lack a clear understanding of the interdependency between resources and the people that live among them. This paper clearly shows that by aligning political commitments with farmers' willingness to participate, farmers' resistance to participation in the CCFP has been partially overcome by perceiving private environmental benefits. Although participating farmers might continue the CCFP if they perceive themselves as benefiting from the associated environmental improvements, our results show that there is a gap between farmers' WTP and the subsidy provided by the central government. This means that enrolled land is likely to be returned to crop farming and the conservation benefits will not be sustained. Furthermore, the CCFP is a project in which the poor west provides ecological services to the relatively wealthy east. This redistributive effect will marginalize the already politically and economically marginalized groups in less-developed regions (Yeh, 2009).

Despite bringing about slight improvements, the economic situation of participating farmers was worsened due to limited structural changes, insufficient compensation, and a lack of arable land. Economic incentives in the form of subsidies are crucial for the viability of the CCFP because afforestation projects that enhance environmental benefits will not be undertaken voluntarily by participating farmers. Our analysis indicates that almost half of the farmers did not receive full compensation, which the central government estimated according to the opportunity costs of local arable land without taking into consideration the environmental services provided. The successful design and adoption of agri-environmental policies must ensure the simultaneous objectives of environmental conservation and economic development (Adhikari, 2009).

The long-term viability of the CCFP needs a complete reconsideration of human-environmental interactions. The redesign of the compensation system for environmental projects is a challenging issue for the central government to fulfill its environmental goals. Green payments should be made to agricultural producers in compensation for the environmental benefits (Hanrahan and Zinn, 2006). Beyond reducing sedimentation from farming, not only the CCFP should reward producers for adopting practices that actually conserve the ecosystem, but also non-market values should also be a larger part of the environmental contribution from the converted lands (Bennett, 2008). The optimal subsidy rates should reflect tradeoffs among regulators' administrative abilities, transfer of income, farmers' participation,

abatement costs, and environmental benefits (Horan and Claassen, 2007).

Knowledge of the potential environmental benefits, conservation costs, and producers' willingness to comply are all necessary to specify appropriate payment levels (Claassen et al., 2005). Restoring ecosystem services from agricultural lands requires not only site-specific information on the relationship between alternative production practices and environmental quality, but also an understanding of which incentives could encourage producers' participation (Batie, 2009). As this study highlights, the CCFP was largely based on the government's environmental motivations, and thus neglected the daily necessities of participants. This situation poses a threat to the viability of the rehabilitation efforts. Further research is needed to provide a solid basis for future environmental policy improvements in terms of ecological modernization in China.

ACKNOWLEDGMENT

This paper is supported by The major project of 2014 National Social Science Fund(14ZDA070) .

REFERENCES

- [1] Adhikari, B., 2009. Market-Based Approaches to Environmental Management: A Review of Lessons from Payments for Environmental Services in Asia. Asian Development Bank Institute, Working Paper Series.
- [2] Bennett, M.T., 2008. China's Sloping Land Conversion Program: Institutional Innovation or Business as Usual? *Ecological Economics* 65, pp. 699–711.
- [3] Bennett, J., Wang, X.H., 2008. The Way Ahead, In: Bennett et al. (eds), *Environmental Protection in China-land-use Management*, Edward Elgar Publishing Limited, USA and UK.
- [4] Chen, X.D., Lupi, F., He, G.M., Ouyang, Zh. Y., Liu, J.G., 2009. Factors Affecting Land Reconversion Plans Following a Payment for Ecosystem Service Program. *Biological Conservation* 142, pp. 1740–1747.
- [5] Claassen, R., Cattaneo, A., Johansson, R., 2005. Cost-Effective Design of Agri-environmental Payment Programs: U.S. Experience in Theory and Practice, Paper to be presented at ZEF-CIFOR workshop on payments for environmental services in developed and developing countries. Titisee, Germany, June 16–18, 2005.
- [6] Hanrahan, C.E., Zinn, J., 2006. Green Payments in U.S. and European Union Agricultural Policy, CRS Report for Congress, Order Code RL32624.
- [7] Horan, R.D., Claassen, R., 2007. Targeting Green Payments Under a Budget Constraint. *Land Economics* 83(3), pp. 319–330.
- [8] Morrison, G., 1997. Willingness to Pay and Willingness To Accept: Some Evidence of An Endowment Effect. *Appl. Econom.* 29, pp. 411–417.
- [9] Project Team for the Forestry Development Strategy for Sustainable Development in China (PTFDSSD), 2003. *Forestry Development Strategy for Sustainable Development in China*, China Forestry Publishing House, Beijing.
- [10] Sharp, B.M.H., and D.W. Bromley. 1979. Agricultural Pollution: The Economics of Coordination. *American Journal of Agricultural Economics* 61, pp. 591–600.
- [11] State council, 2007. A Notice on the Improvement of Sloping Land Conversion Program, http://www.china.com.cn/news/txt/2007-09/11/content_8856020.htm.
- [12] State Forestry Administration, 2003. Master Plan for the Sloping Land Conversion Program.
- [13] Toma, T., Nawir, A., Sabogal, C., Chokkalingham, U., De Jong, W., Gumartini, T., 2004. Review of Forest Rehabilitation Initiatives-Lessons from the Past. Executive Summary for CIFOR REHAB Country Syntheses.
- [14] Yeh, E.T., 2009. Greening Western China: A Critical View. *Geoforum*, 40, pp. 884–894.

Traditional forest knowledge, rubber plantation and climate change

Y.H. Du & L. Guo

College of Life and Environmental Sciences, Minzu University of China, Beijing, China

D.Y. Xue

College of Life and Environmental Sciences, Minzu University of China, Beijing, China
Nanjing Institute of Environmental Sciences, MEP, Nanjing, China

F. Liu & X.Z. Yu

Nabanhe National Nature Reserve, Jinghong, Yunnan, China

ABSTRACT: As a result of the development of modern industrial civilization, the rubber industry of Xishuangbanna has expanded greatly, leading to the replacement of traditional natural forests with rubber “forests”. The introduction of rubber plantations to Xishuangbanna has resulted in a change in traditional ethnic and agricultural patterns as well as climate conditions. Traditional knowledge related to the forest has gradually disappeared and the rainy season has become shorter and shorter. This article summarizes the locally adopted strategies to counter these changes. In particular, it focuses on the management of forest resources. This article also explores traditional knowledge on the sustainable use of the forest. Lastly the article makes a conclusion on the effects of rubber plantations based on the above mentioned details.

KEYWORDS: Xishuangbanna, traditional forest knowledge, rubber, adaptation satieties.

1 INTRODUCTION

Forests are vital economic and cultural resources for many indigenous people and tribes. Forests are frequently zoned by the community, with access and use regulated by local traditional norms. Traditional forest knowledge (TFK) is the name given to knowledge and beliefs transmitted either orally or through observation. Traditional forest knowledge (TFK) is not just about how people use forests, but also how they perceive and regulate forests in their livelihoods and communities. It includes a system of classification, a set of empirical observations about the local forest ecosystem and a system of self-management that governs forest resource use. Indigenous peoples are culturally invested in specific values, meanings, and identities that are linked with the natural landscape (Daigle and Putnam 2009).

After the industrial revolution, climate change is making extreme weather episodes more common, such as drought and flood. As a result of these climate changes, extreme changes to the ecosystem have also occurred. Forest conversion and habitat degradation, all a result of climate change, now greatly threaten tribal and indigenous access to valued resources (Ryan et al. 2008). The tribes and indigenous people depend on these resources, including water, plants,

animals, forest products (fungi, medicinal materials) to maintain their cultures and livelihoods (Garrit et al. 2013). It has been proven that these resources and habitats are invaluable to the cultural, economic, medicinal, and communal health of these peoples.

Indigenous/local communities have lived in areas of changing climate for hundreds of years. They have made their livelihoods in this volatile environment and thus are the first to discern and suffer from its changes. However, these people are *not just* victims of climate change. They have adapted a variety of methods to forecast weather patterns and have developed countermeasures to these changes. The culmination of these adaptations to weather patterns has actually had an effect on tribal culture; the requirement to develop new and unique strategies to access resources limited by weather changes has become a key part in tribal tradition.

Overall however, climate change has been detrimental to the culture and traditions of indigenous cultures. With travel being easier, more and more indigenous people opt to migrate to other areas rather than adapt to their situation. This, in turn, results in less and less people practicing and knowing indigenous culture. To understand potential climate change adaptation strategies, the article explores traditional forest knowledge

and historical local adaptive approaches to forest resource management. This article examines the impact of rubber plantations on forests and the resulting effects on traditional forest knowledge and local climate change.

2 MATERIALS AND METHODS

2.1 Location and method

Xishuangbanna-Dai-Autonomous-Prefecture is one of eight autonomous prefectures in Yunnan Province allocated for the protection of national minorities. The region (21°36'42"-58"N, 101°34'26"-47"E) is located in the south of Yunnan in southwest China. It is bordered on the south and southeast by Laos and on the southwest by Burma. Xishuangbanna is located at the northern limit of the SE Asian tropics and as such has a seasonal climate (Cao et al., 2006). The tropical forests in this region are floristically similar to the rain forests from the lower latitude seasonal tropics of SE Asia (Zhu et al., 2006; Zheng et al., 2006). Xishuangbanna harbors 16% of the plant diversity of China (Liu et al., 2002), and is part of the Indo-Burma diversity hotspots (Myers et al., 2000; Cao et al., 2006). It is rich in natural, historical and cultural resources, noted for its folklore, rain forests, rare plants and wildlife.



Figure 1. Location of Xishuangbanna in Yunnan.

Much of southwestern China, including Yunnan Province, is inhabited by non Han Chinese ethnic groups. There are 23 different ethnic groups in Yunnan Province. In Xishuangbanna, there are 13 minorities (include Han people), including Dai, Han, Aini, Yi, Lahu, Bulang, Jinuo, Yao, Miao, Hui, Wa, Zhuang, Jinpo etc. They have all lived in this

area for over 100 years. Dai is the largest nationality in the region which comprises around 35% of the total population. They have a rich trove of traditional knowledge on dealing with the changing weather and environment.

2.2 Study site selection and methods

This study began with interviews of natives from 20 local ethnic villages in Jinghong and Menghai in Xishuangbanna. The natives were questioned on their traditional forest knowledge and on how they responded to environmental change. These villages were selected for being fairly representative of the average village in the area.

This case study used data obtained from interviews and surveys conducted on the above mentioned villages. Some price estimations from local markets were made based on this data.

Semi-structured interviews were carried out with selected key members of each village's community. These include: local Religious Master, village heads and village elders. Those chosen to be interviewed were often respected leaders within the community who were able to offer rich explanations of natural and historic events. Also, being as they were village leaders, they were also able to inform us of village regulations, forest access regulations and the ever changing forest management styles.

Group discussions were organized in each of the three villages; the groups were a mix of people from all facets of village life including the old, young, male and female. With our guidance, the groups created village land-use maps, forest management maps, and historical timelines which recorded many important events affecting forest management.

3 RESULTS

Local peoples of the tropical rainforest belt are very dependent on the forest. Many of them have already been hurt by the impacts of environmental and climate change and are struggling to adapt to these changes.

3.1 Adaptation strategies to climate change

As environmental change has already affected their livelihoods in the past, in many cases indigenous peoples have developed specific strategies to cope with these extreme weather variations. Some examples of such adaptation strategies include:

- a Crop diversification: In order to minimize the risk of harvest failure, a variety of crops are planted in different environmental/field locations.

- b Changes of living area and a variety of movement patterns are used to deal with climatic variability.
- c Changes of hunting and gathering periods to adapt to changing animal migration and fruiting periods.
- d Changes of varieties and species. Livestock varieties may be changed to take account of new disease challenges.
- e Changes in food habits. People who are close to a town might trade or barter with neighboring villages or traders/markets.

The forest is an extremely important resource to the local people. The locals realize this and have thus developed an extensive knowledge of the forest.

3.2 Local traditional forest knowledge

3.2.1 Cultural and religious beliefs

Traditional knowledge of forest management and protection has been passed from generation to generation through legends, folklore, and festivals. The ethnic groups in Xishuangbanna view forests as integral components of their lives. Trees in private farm plots are respected by some as ancestors. Violation of the forest use regulations can result in serious punishment. These management regulations are formulated by the local people and are well enforced.

3.2.2 Classification of forests and trees

Forest categorization reflects knowledge about how to manage and use forests. In local communities, the forests are categorized as holy trees, temple forests, grave forests, watershed forests, family contracted forests and trees on farm land.

3.2.3 Customary law

Each village has its customary, which oversees the protection of collective forests and sacred trees. Unauthorized destruction of trees (whether intentionally felled or damaged by livestock) results in swift punishment for the offenders.

3.2.4 The roles of forests in the lives and livelihoods

Forest ecosystem provides food, medical resource, construction material (timber), fuel, shelter and so on for local ethnic groups. It is very important in their daily life.

3.3 Rubber plantation

In the past, Rubber grew exclusively in the Amazon rainforest. However, the advance of economic

globalization and the expansion of industrial production have led to the introduction of rubber plantations in Xishuangbanna. This in turn has had an effect on local traditional agricultural patterns. Many locals have given up their rural livelihood in search of work on these new rubber plantations.

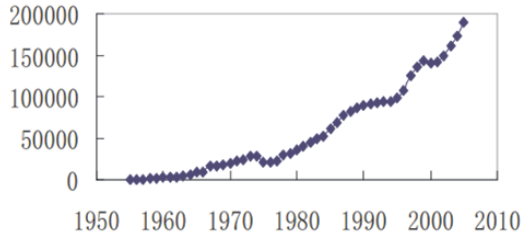


Figure 1. Changes in rubber cover of Xishuangbanna (1950–2010).

The booming rubber economy has also led to the disappearance of local forest and rainforest in Xishuangbanna. The local people are also gradually giving up their traditional ways of life because of economic benefits. This creates a chain reaction style effect. As more and more people forgo traditional knowledge in favor of the rubber plantations, the forests become less and less protected. As the forests become less protected, it allows for easier encroachment by the rubber plantations.

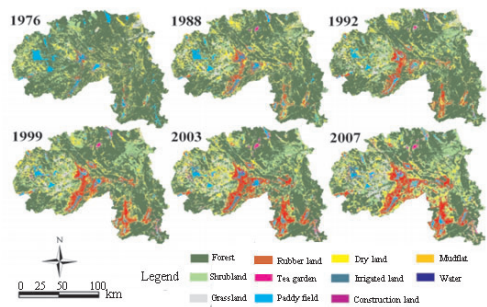


Figure 2. The changing of land use type of Xishuangbanna.

3.3.1 The climate change of Xishuangbanna

The last 50 years has seen a consistent temperature increase and precipitation decrease in Xishuangbanna. At the same time, the rubber planting area increases significantly in Yunnan province, and the rubber area in Xishuangbanna makes up about 80% of Yunnan. The expansion of rubber coincides with an increase in dry weather and drought as well as a decrease in the rainy season.

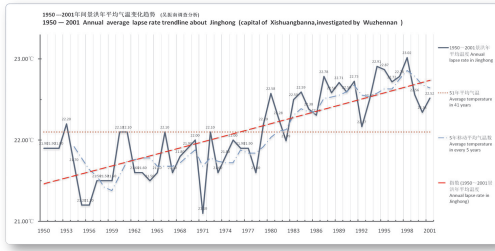


Figure 3. Annual average lapse rate trend line in Jinghong of Xishuangbanna (By Wu zhennan).

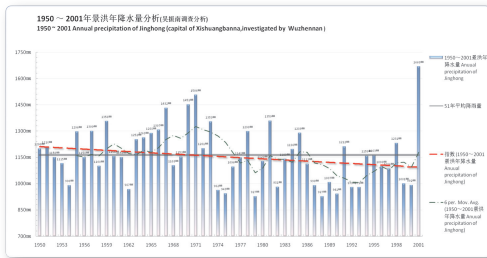


Figure 4. Annual precipitations in Jinghong of Xishuangbanna (By Wu zhennan).

3.3.2 Biodiversity loss and ecosystem service decrease

Excellent climate environment and rich biological resources in Xishuangbanna have allowed it with high tropical forest coverage. However, with the increase of population and the drive of economic interests, the natural forest has been replaced by rubber forest. Now, the swidden fields and farm land to forest have been replaced by rubber plants. The reduction in holy forest area is also serious; even many villages no longer have their original holy forest.

In order to achieve a high and stable yield of rubber, specific efforts were made so that only rubber plants grew in the rubber forests. The lack of secondary plants in the rubber forests resulted in its single community structure, and this in turn causes poor water conservation. The biological population in rubber forests is much lower than that of natural forests. Animals who are dependent on underlying vegetation cannot find them in rubber forests.

As an added effect, the lack of extra land of other crops has also resulted in a loss of unique rice varieties.

4 DISCUSSION AND CONCLUSIONS

The case study presented here demonstrates that local ethnic groups have their own traditional knowledge

related with forest utilization, protection, and management regimes. And traditional forest knowledge has demonstrated its significance in the protection of various forest types and tree species, contributing to the conservation of biodiversity.

Rubber planting however, has adversely changed local traditional beliefs and customs. Rubber plantations have caused the following changes: 1) transition of large areas of land for industrial use and development; 2) simplification of original forest plant community; 3) possible cause for the overall change in local regional climate. From the perspective of cultural change and social development, the rubber plantation has led to a transition from traditional living to modern living. This has caused traditional culture closely related to social life and agriculture to lose its original meaningful existence.

Despite that the economic pressure that rubber presents, many ethnic groups in Xishuangbanna are still engaged in their traditional productions and practices. Disruption of forest areas and changes in species composition as a result of climate change could lead to the loss of culturally important resources. This would negatively impact tribal culture and economy. In essence, the loss of cultural diversity and the loss of biodiversity go hand in hand (Figure 5).

Despite the importance and traditional knowledge developed in Xishuangbanna and the threat it faces from human activities, there is still little effort to study the changing climate and culture of this area.

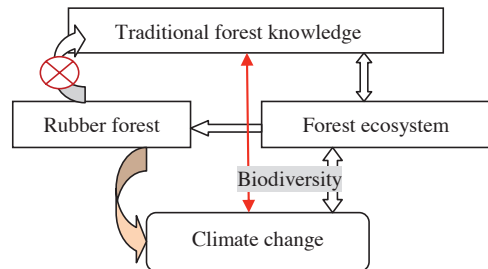


Figure 5. Relation among TFK, forest, rubber and climate change.

ACKNOWLEDGMENT

The work presented in this paper was supported by the Fundamental Research Funds for the Central Universities, the National Natural Science Foundation of China (No.31370480) and 111 Project (B08044). We thank the people provide us with data and help, the ethnic people we interviewed, and the staffs of NNR together with us in the field research.

We also thank the editors and anonymous reviewers for their helpful comments.

REFERENCES

- [1] Cao, M., Zou, X., Warren, M., Zhu, H., 2006. Tropical forests of Xishuangbanna, *China.Biotropica* 38, 306–309.
- [2] Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858.
- [3] Ryan MG, Archer SR, Birdsey R, Dahm C, Heath L, Hicke J, Hollinger D, Huxman T, Okin G, Oren R, Randerson J, and Schlesinger W (2008) Land Resources. In: U.S. Climate Change Science Program. The effects of climate change on agriculture, land resources, water resources, and biodiversity in the United States.
- [4] <http://www.climatechange.gov/Library/sap/sap4-3/final-report/>. Accessed 20 January 2013.
- [5] Garrit Voggesser, Kathy Lynn, John Daigle, Frank K. Lake and Darren Ranco (2013) Cultural impacts to tribes from climate change influences on forests. *Climatic Change* 120:615–626.
- [6] Zheng, Z., Feng, Z.L., Cao, M., Li, Z.F., Zhang, J.H., 2006. Forest structure and biomass of a tropical seasonal rainforest in Xishuangbanna, Southwest China. *Biotropica* 38, 318–327.
- [7] Zhu, H., Cao, M., Hu, H.B., 2006. Geological history, flora, and vegetation of Xishuangbanna, southern Yunnan, China. *Biotropica* 38, 310–317.

Economic and thermal technical comparison of classical foundation and foundation on air gap–crawl space

Š. Korbelová, M. Vodičková & V. Šípková

VSB—Technical University of Ostrava, Faculty of Civil Engineering, Ostrava - Poruba, Czech Republic

ABSTRACT: Foundation of the house is one of the most important things. There are two significant aspects. First is static one and the second is our aim, that is, thermal technical aspect. The main topic of the article is setting up wooden buildings over an air gap–crawl space. This type of foundation is commonly used in United States, and we would like to verify it in climatic condition of the Czech Republic. Crawl space has advantages such as ventilation of water vapor or readily available technical networks. Furthermore, this type of foundation is compared with usual method foundation on footings (strips). There are two modeled details of foundation that differ in the way of the foundation and composition of the floor. Construction details are analyzed in the software and then evaluated in terms of thermal technical and economic point of view.

KEYWORDS: Crawl space, foundation engineering, air gap.

1 INTRODUCTION

The construction can be found in several ways. The article deals with two ways. The most common is foundation through footings or slabs from reinforced concrete. For wooden structures, it is possible to use an alternative foundation pillars with crawl space.

Demands on the energy performance of buildings (Council 2010/31/EU) are still increased. Foundation of wooden houses using crawl space level from a technical point of view brings certain advantages. These include some possibility of control and easier repairs for technical distribution, which lead to the house; there is a possibility of ventilation water vapor or radon. Due to foundation, using wooden frame saved concrete and limited wet processes.

Height of the air gap should be at least 600 mm, rather than 900 mm (Šála J. & Smola J., 2013). Air temperature in the air gap is corresponding approximately to the outside air. For this reason, it is necessary to ensure the technical solution of the floor. It is also necessary to deal with a solution that would avoid increasing air humidity in the crawl space or water vapor in the floor structure.

Therefore, it is important to ensure proper behavior of the air gap. Principle of ventilated air gap is mainly dependent on different temperatures below the house and the outside environment and also at different temperatures on different sides of the house. If a plan of the house is larger, it is necessary for proper function of the air gap fitting exhaust stacks that use the “chimney” effect due to the different air pressures. The vent chimneys can add fans (Růžička M., 2014).

In case of foundation through footings, there are designed sufficiently large venting holes. These openings are sealed in winter (Šála J. & Smola J., 2013).

Wooden family house in the village Kramolna (region Náchod) has builtup area of 48.99 m². It is based above the ground using the foundation block with dimensions 500 × 500 mm. The air gap between the ground floor and the ground has a range of 350–750 mm (from north to south). The supporting structure of the house is built by the system STEICO with supporting wooden pillars in size 140 × 140 mm. On the load-bearing wall were used beams STEICO wall SW45 and the supporting joist floor construction STEICO joist SJ60. Thermal insulation of the whole house was done using blown insulation CLIMATIZER Plus, which is made from natural cellulose fibers. Peripheral walls of the house have two kinds of finishes—wooden siding and a clay plaster. To assess the details of the walls, wooden siding was selected.

For the house that is mentioned earlier is designed on variant foundation on a field. The thermal transmittance of the floor is the same as both variants. The value of thermal transmittance of the floor is 0.13 W/m²K. Both methods of establishment are analyzed and compared in terms of thermal technical and economic terms.

Types of peripheral wall and floors are given in [Tables 1–3](#).

Regarding the details of the lower corner, two alternative solutions are given, which vary the composition and methods based on floor construction. Detail no. 1 is the establishment of an air gap with the

Table 1. The composition external wall from the interior.

| Material | Thickness (mm) |
|---|----------------|
| Clay plaster | 10 |
| OSB board 3 SUPERFINISH ECO | 18 |
| Beams STEICO wall SW45 + thermal insulation CLIMATIZER Plus | 300 |
| Diffusion foil JUTADACH 115 | - |
| Wooden siding | 19 |

Table 2. The composition floor 1 from the interior.

| Material | Thickness (mm) |
|--|----------------|
| Linoleum | 4 |
| Mirelon | 2 |
| OSB board 3 SUPERFINISH ECO | 18 |
| OSB board 3 SUPERFINISH ECO | 25 |
| Beams STEICO joist SJ60 + Thermal insulation CLIMATIZER Plus | 300 |
| OSB board 3 | 12 |

Table 3. The composition floor 2 from the interior.

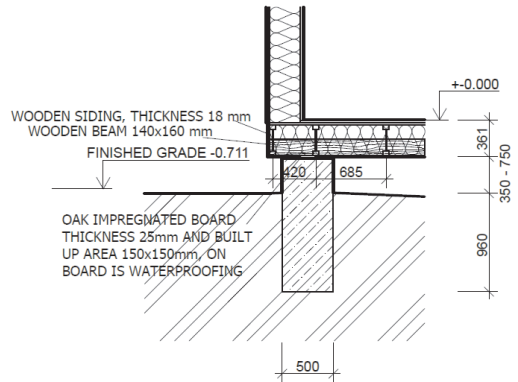
| Material | thickness mm |
|---|--------------|
| Linoleum | 4 |
| Cement screed | 50 |
| Vapor barrier | - |
| Thermal insulation Styrodur 3035 CS, thickness 2 × 140 mm | 280 |
| Waterproofing—elastek 50 special mineral | 5 |
| The base reinforced concrete | 150 |

appropriate composition floor (see Table 2). Detail no. 2 is based on footings from reinforced concrete with appropriate composition floor (see Table 3). Scheme of details are shown in Figure 1.

2 THERMAL TECHNICAL ASSESSMENT OF THE STRUCTURES IN THE SOFTWARE

Each house construction must comply with the requirements specified ČSN 73 0540-2. These requirements include the heat transfer coefficient and the condensation of water vapor, on which walls and

Detail no.1



Detail no.2

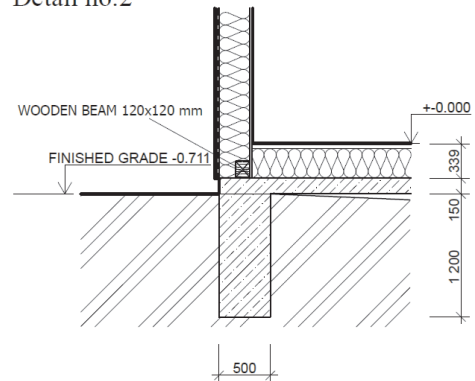


Figure 1. Scheme of solved details.

floor of the house were also assessed using a software Teplo 2011.

Boundary conditions for the assessment of structures in the program according to ČSN 730540-3 are as follows: design internal temperature $\theta_i = 20^\circ\text{C}$, internal air temperature $\theta_{ai} = 20.6^\circ\text{C}$ is design internal temperature increased by a surcharge according to Table ČSN 73 0540-3, the design external temperature in winter $\theta_e = -17^\circ\text{C}$ (region Náchod), internal resistance of heat transfer of the building wall construction $R_{si} = 0.13 \text{ m}^2\text{K/W}$, external resistance of heat transfer of the building floor construction $R_{se} = 0.17 \text{ m}^2\text{K/W}$, resisting transfer to the outside of the building structure $R_{sc} = 0.04 \text{ m}^2\text{K/W}$.

The heat thermal transmittance must satisfy the condition:

$$U \leq U_N \text{ (W/m}^2\text{K)}, \quad (1)$$

where $U_N \text{ (W/m}^2\text{K)}$ is the required heat thermal transmittance and $U \text{ (W/m}^2\text{K)}$ is the heat thermal transmittance.

Assessment of condensation of water vapor, where for building structure comprising wood element is determined by the condition according to ČSN 73 0540-3:

$$M_c = 0 \text{ (kg/m}^2\text{a)} \quad (2)$$

where M_c (kg/m²) is the annual amount of condensed water vapor inside the structure.

The boundary conditions correspond to the climatic conditions in the Czech Republic. Detail was been modeled in the Area 2011 software. According to ČSN 73 0540-3, boundary conditions are as follows: design internal temperature $\theta_i = 20^\circ\text{C}$ and internal air temperature $\theta_{ai} = 20.6^\circ\text{C}$, the design external temperature in winter $\theta_e = -17^\circ\text{C}$ (region Náchod), external temperature in winter in the soil $\theta_c = +5^\circ\text{C}$ is located at a depth of 3 m below finished grade, resistance to transfer to the inside of the building structure $R_{si} = 0.25 \text{ m}^2\text{K/W}$, resistance to transfer to the outdoor of the building structure $R_{se} = 0.04 \text{ m}^2\text{K/W}$, resistance in soil $R_{sc} = 0 \text{ m}^2\text{K/W}$ is located at a depth of 3 m below finished grade.

The assessment and evaluation of detail were performed according to requirements ČSN 73 0540-2.

$$f_{Rsi} > f_{Rsi,N} [-] \quad (3)$$

where $f_{Rsi,N} [-]$ is the lowest required temperature factor of the inner surface and $f_{Rsi} [-]$ is determined temperature factor of the inner surface.

3 THE ECONOMIC ASPECT

The price is an important factor for investor during the decision-making on appearance of the building and the way of foundation, both methods of founding were quoted according to RTS list price for the construction mentioned earlier. Price includes only the foundation of the construction with floor without finishing.

4 SUMMARY

The results of the evaluation according to requirements ČSN 730540-2 are shown in Tables 4 and 5 and Figure 2.

According to thermal technical evaluation, it has little difference.

Cost of classical foundation is 166,420 CZK. Cost of foundation on air gap is 149,100 CZK. The prices are without VAT. Difference in price between

Table 4. Evaluation of the construction of the software Teplu 2011.

| Name of construction | Determined value U (W/m ² K) | Required value U_N (W/m ² K) | Evaluation $U \leq U_N$ |
|----------------------|---|---|-------------------------|
| Wall | 0.13 | 0.30 | Fulfilled |
| Floor 1 | 0.13 | 0.24 | Fulfilled |
| Floor 2 | 0.13 | 0.45 | Fulfilled |

Table 5. Evaluation of details of the software Area 2011.

| Name | The temperature factor of the inner surface $f_{Rsi} [-]$ | Required lowest temperature factor $f_{Rsi,N} [-]$ | Evaluation $f_{Rsi} > f_{Rsi,N}$ |
|--------------|---|--|----------------------------------|
| Detail no. 1 | 0.915 | 0.760 | Fulfilled |
| Detail no. 2 | 0.931 | 0.760 | Fulfilled |

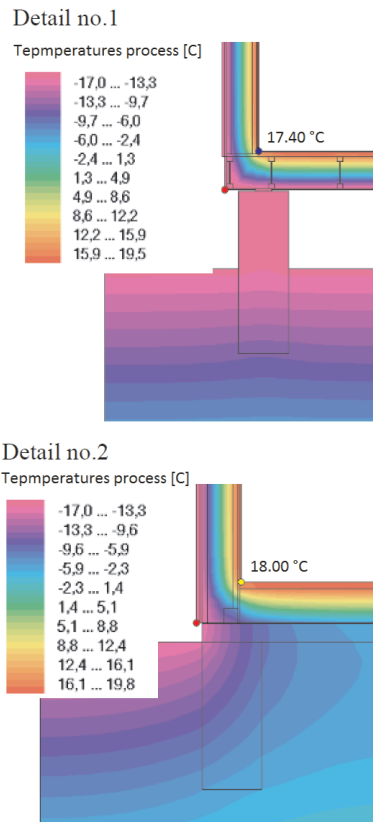


Figure 2. Illustration of temperatures process, software Area 2011.

the foundation on air gap and classic baseplate is minimal.

Choice of the foundation is only for investor.

ACKNOWLEDGMENT

The work was supported by the Student Grant Competition VSB-TUO. Project registration number is SP2014/167.

REFERENCES

- [1] Růžička, M. 2014. Moderní dřevostavba. Prague: Grada Publishing.
- [2] Šála, J. & Smola, J. 2013. Foundation of wooden structures above aired crawl space. Thermal protecting of buildings, 3: 49–52.

- [3] ČSN 73 0540 – 3, 2005. Thermal protecting of buildings – part 3: Design value quantities. Prague: The Office for Standards, Metrology and Testing.
- [4] ČSN 73 0540 – 2, 2012. Thermal protecting of buildings – part 2: Requirements. Prague: The Office for Standards, Metrology and Testing, including changes Z1: 2012.
- [5] Directive of the European Parliament and of the Council 2010/31/EU on the energy performance of buildings
- [6] Information on <http://www.prirodnibydeni.cz/crawl-space/>

Gold refining process and its impact on the environment

S. Kaspin

Metal Design Contemporary Department, Faculty of Art & Design, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia

N. Mohamad

Fine Art Department, Faculty of Art & Design, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia

ABSTRACT: The nitrogen dioxide that is formed during the traditional gold recovery process is hazardous to health, causing air pollution and being able to pose a serious risk on health and safety. Thus, the research objective is to identify and develop a suitable method that is safe and “eco-friendly” with a “Green” philosophy and most importantly, feasible to be used for both small-scale and in-house gold recovery process. Throughout the research process, the key problem that is derived from the handling of the quality control aspect of respective methods is used by the refiners. The research also concludes that the key area of concern would be on whether the proposed recovery method is feasible and viable to implement and be used by the respective traditional or small-scale jewelry workshop at their premises. To comprehend its real issues, a full cycle laboratory based physical and chemical testing on both traditional and sophisticated recovery methods have been carried rigorously to validate the research findings and conclusions.

1 GOLD REFINING PROCESS

1.1 Introduction

Refining is defined as the process of removing impurities from the precious metal. Gold metal going through refineries may either be a recycled scrap, which is purified and upgraded or in the final stage of its transformation from ore in the mine of bullion (Kenneth Blackmore, 1982).

The first recovery technique has been found and introduced in the Ancient Egyptian era; about 500 B.C. It is based on the result of the chemical testing analysis that determines the ages and level of purity of the sample gold items founded during archeological activities. It is evidenced that the sample has higher gold purity element compared with normal gold alloy, that proved the usage of the gold refining techniques (J.H.F Notton, 1974).

1.2 Gold jewelry scrap

The sources of gold scrap can be originated from jewelry fabrication process whether by start-of-art machine or handmade process such as cutting, grinding, filing and manual polishing on buffing wheels, sawing, and others. The waste particles can be found at the workbench and the floor, and airborne particles that settle as dust on work-surfaces, pipes, and operative clothing or may even be vented to the outside and lost (Dr. Christopher W. Corti, 2002).

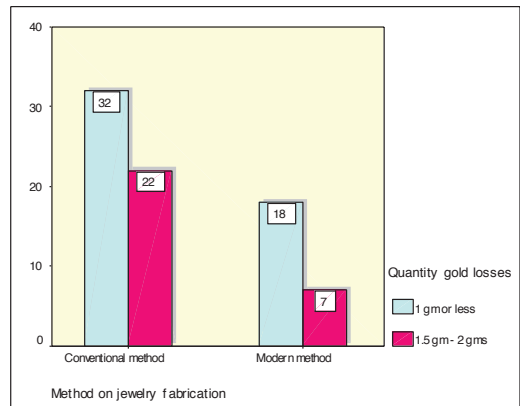


Figure 1. Cross tab of method on jewelry fabrication and quantity of gold losses.

Figure 1 shows that the most popular method used in jewelry fabrication is the conventional method at 68.4%. The conventional method records the highest production of gold scrap with 1 g or less at 40.5% followed by 1.5–2 g at 27.8%.

The gold scrap is categorized as higher-grade and lower-grade values (Figure 2) with reference to its quality such as composition, cleanliness from contamination, and others (Dr. Christopher W. Corti, 2002).

Usually, the jewelry process has indirect impact to the jewelry weight. The process can cause the quantity and the weight of the raw material (gold)



Figure 2. Low grade of gold scrap.
Source: Courtesy of Jewelers House in Machang, Arked Uda and Kota Bharu area. Taken by researcher on 2004.

to reduce slightly compared with the original weight measurement. The metal scrap or waste is very difficult to locate due to its changing physical existence in dusty form. Thus, the manner, the wastages, and the scrap have been handled during the process, which is something to be pondered by the producer as inefficient process that could result in the lost of revenue. Therefore, it is interesting to note that the waste made of gold scrap is able to be refined back to its metal condition through recycling and recovery process (Dr. Christopher W. Corti, 2002a, Saadiah Kaspin, 2008a, 2010 b).

Hence, a reliable and viable refining and recovery method is key to claim and to extract the missing gold scrap from the waste, through a safe and eco-friendly process in alloy manufacturing.

2 SMALL-SCALE GOLD REFINING PROCESS

Based on researcher's observation during the manufacturing site visits, we conclude that current refinery practice and methodology are seemed to be risky, unsafe, and dangerous due to improper handling of acidic chemical substances and unhealthy refining procedures during the process and non-hygienic workplace.

The process typically produces heavy smokes, poisonous fumes and bad odors that could possess a threat to human's health, safety as well as air pollution, largely without proper control measures (Dr. Christopher W. Corti, 2002).

Through the case study on refining methods category used by refiners in Kelantan district in Malaysia, it is found that there are two types of methods that usually used in traditional gold recovery process (methods *A* and *B*). The empirical lab testing has been conducted to identify the effectiveness of the two methods of producing higher purity of gold alloy that can go as high as 99.99%.

2.1 Gold refining process (method A)

Based on the findings, the main weakness for the method A is seen on the impact of the chemical reaction between gold metals with nitric acid that causes acrid brown fumes, which is heavily polluted

the refining compound. The pollution is so thick and heavy and poses a serious health hazard to the workers and the public surrounding the areas, and the dangerous fumes should be inhaled during the recovery process period.

The traditional recovery process is the higher risk of failure due to inaccurate calculation of gold caratage. The process to recover pure gold extract is found not that successful due to higher proportions of silver compared with pure gold extract at the end of the process.

The root cause for the issues is analyzed, and the result findings show that there is a clear indicator on the unsystematic procedures used during the gold refining and recovery process. Good hygienic procedures that are environmental friendly with greater emphasis on health and safety may be able to mitigate the dangerous elements and possibly increase the rate of recovery success (Dr. Christopher W. Corti, 2002).

2.2 Gold refining process (method B)

The rigorous laboratory testing on sample method B has reportedly incurring a declining weight of gold during precipitation and filtering processes. If the gold is not fully precipitated, it will cause other elements or impurities to mix with gold alloy and affect the gold purity, while the pure gold may risk in being flushed out during the filtering process. This method will take several days to complete as the refiners need to be sure that the gold has been fully precipitated.

Recovery process by method B finds that it has some positive impact on air pollution with minimal effect on environment (lesser smokes) and public at large, due to no heating process. The amount of nitric oxide is reported lower and lesser compared with method A.



Similar root cause is found in method B as the final analysis recorded a substantial proof and evidence that linked the pollution, potential health hazard, and economic losses (lower gold extract) to unsystematic procedures and suitable equipment/facilities to carry out the recovery process effectively and safely.

From [Table 1](#), it is clear that the chemical composition with the highest percentage of gold element is



Figure 3. An acrid fumes (NO_2) produced during refining process.

Table 1. Chemical composition of the solid gold samples from gold refining methods A and B using XRF.

| Traditional method | Sample | Weight (g) | Element (weight %) | | |
|--------------------|---|------------|---------------------|--------------|--------------|
| | | | Gold as Au | Silver as Ag | Copper as Cu |
| A |  | 4.57 | 99.0 | 0.89 | 0.11 |
| B |  | 9.66 | 95.3 | 1.78 | 2.95 |

the sample from method A with 99% purity, whereas the sample from method B shows 95.3%.

3 NEW IMPROVEMENT ON GOLD REFINING PROCESS

A high-tech laboratory experiment is conducted to repeat the traditional process copied from the local refiners to be tested in a more structured and systematic approach. The intent of the laboratory experiment is to revalidate the traditional process and to give evidence of its capability to extract gold from the scrap/waste. The purpose is to improvise the gold recovery process to be more safe and systematic with the highest result of gold purity.

The idea to design the lab equipment to be more safe and systematic process to recover gold is referred by the Ostwald process. Nitric acid is produced industrially by the Ostwald process, and through this method, the nitric acid can be reused in other industry applications with lower percentage (Tables 2 and 3).

Table 2. Laboratory experiment showed the reaction of gold grain with nitric acid.









| Gold recovery process | | | | |
|---|---|---|---|---|
| (STEP B) | | | | |
| Nitric acid 50 ml | | | | |
| 5.45 gm | 5.50 gm | 6.05 gm | 6.15 gm | 6.20 gm |
|  |  |  |  |  |
| Before heating | Starting heating | 60°C | 100°C | 107°C |

Table 3. The concentration of nitrogen dioxide (NO₂) is reduced when reacts with water (H₂O) and can be recycled for other applications in industry.

| Chemical reaction: nitrogen dioxide (NO ₂) with water (H ₂ O) | | |
|--|---|---|
| Water 100 ml | | |
|  |  |  |
| 6.05 pm | 6.15 pm | 6.20 pm |

The finding shows that the experiment has successfully eliminated nitrogen dioxides (NO₂) through “Ostwald process” and allowed the nitric acid recycling even in a lower percent of concentration. The other findings from the laboratory experiment show an improvement on the quality of gold purity that can reach as high as 100% in purity with a more controlled and a safer process.

4 RESULTS AND FINDINGS

The analysis of each samples from traditional method of gold recovery process and laboratory experiment results are validated through a rigorous laboratory testing process such as XRF, thermogravimetric, hardness vickers, and microstructure to measure whether the traditional method is effective to recover and extract high-quality solid gold. This research has successfully profiled the strength and weaknesses of both traditional and laboratory testing methods of gold recovery and refining method.

The shortcoming and challenges faced via traditional gold recovery process is the key reason why local jewelers are unable to refine their own gold scrap; the process is dangerous and involves of acidic material that can produce the nitrogen dioxides (NO₂), the acrid, and toxic fumes (Figure 3).

The nitrogen dioxide produced from the traditional gold recovery process has caused the air pollution and leads to the high risk in safety issues. This problem relates to the quality control of each method used by the refiners. The findings also conclude that this particular method is unable to filter out all silver and copper elements and in the same time affects the quality of gold in terms of percentage.

To verify whether the traditional gold recovery process can be reviewed to avoid the massive risk on air pollution as well as to increase the percentage of gold purity, further enhancement on laboratory experiments is tested and implemented. In fact, the finding shows that the experiment has eliminated

nitrogen dioxides (NO₂) through “Ostwald process” and allows the nitric acid recycling even in a lower percent of concentration. The findings from the laboratory experiment also reimprove the quality of gold recovered as high as 100% in purity with a more controlled and a safer process.

5 DISCUSSION AND CONCLUSION

We can conclude that this research finding is able to assist the local refiners to improve on the quality in their recovery process and at the same time enhance the possibility of in-house refining and recovery process by local jewelers.

Other than that, the traditional recovery process through the usage of nitric acid and precipitation process must be improved to avoid longer processing period, which is taken to complete the process and most importantly, to avoid the gold losses during filtering process.

To deploy the possibility of implementing a recycling strategy to recover precious metals such as gold, it is important to consider the cost of recovery as well as its commercial viability. There is no commercial sense in spending more in processing costs than the value of metal recovered. To conclude, the processing of wastes to reduce wastages and concentrate on the precious metal is essential for economic recovery, particularly to the small-scale jewelers.

ACKNOWLEDGMENT

The authors are grateful for support from local refiners, Associate Prof. Dr. Zaini Hamzah from Faculty of Applied Science, University Technology MARA (UiTM) and Dr. Christopher W. Corti from World Gold Council (WGC). Last but not least, a full appreciation to the University Technology MARA (UiTM) for their support in making this research successfully.

REFERENCES

- [1] J.H.F Notton, “Ancient Egyptian Gold Refining”, Gold Bulletin, (Mac 1974).
- [2] Kenneth Blackmore in ‘Buying Jewelry: A Practical Guide’(1982).
- [3] Dr. Christopher W. Corti, Recovery and Refining of Gold Jewellery Scraps and Waste, The Santa Fe Symposium on Jewelry Manufacturing Technology (May), Wold Gold Council. (2002).
- [4] Saadiah Kaspin, An Improvement of Traditional Gold Recovery Process for Jewellery Scrap, Faculty Of Art and Design, University Technology MARA (UiTM) Shah Alam, Selangor, Malaysia. (2008).
- [5] Saadiah Kaspin, An Investigation on The Effectiveness of Traditional Gold Recovery process for Jewellery Scrap, Faculty Of Art and Design, University Technology MARA (UiTM) Shah Alam, Selangor, Malaysia. (2010).

Semiempirical modeling of iron oxide pollution in the mine soil using geochemical analyses and spectroradiometric measurements

A. Khalil

FST-Gueliz, Cadi Ayyad University, Marrakech, Morocco

A. Bannari

Geoinformatics Department, Arabian Gulf University, Manama, Kingdom of Bahrain

ABSTRACT: This study focuses on the semiempirical modeling of soil iron oxide pollution (IOP) generated by acid mine drainage (AMD). To achieve our goal, 62 soil samples representing different IOP classes have been selected. Spectroradiometric measurements were acquired over each sample using an ASD spectrometer (350–2500 nm). These measurements were resampled and convolved according to the spectral bands of BGIS-2000, TM, and ASTER sensors. Then, IOP content was evaluated using geochemical analyses. Five spectral indices for iron oxide prediction were correlated to geochemical analyses using second-order regression analysis. The results show that only the model based on Madani index (IM) offers the most significant correlation coefficient (90%). This model could be derived from TM and ASTER data but not from BGIS-2000 data because the SWIR bands are absent for this sensor.

1 INTRODUCTION

Acid mine drainage (AMD) from mine wastes and contaminations of soils and water with metals are considered as major environmental problems in mining areas [1, 2]. This AMD is produced by interactions of water, air, and sulfidic mine wastes. AMD results from a series of chemical and biochemical oxidation reactions of sulfide minerals, for example, pyrite and pyrrhotite [3, 4]. These reactions lead to acidity as well as the dissolution of toxic and heavy metals (Fe, Mn, Cu, etc.) from tailings, waste rock piles, and open pits. Soil and aquatic ecosystems could be contaminated, and, consequently, human health and wildlife will be affected [5]. Furthermore, secondary minerals, typically formed during weathering of mine waste storage areas when the concentration of soluble constituents exceeds the corresponding solubility product, are also important. The most common secondary mineral compositions are hydrous iron oxide (goethite, etc.) and hydrated iron sulfate (jarosite, etc.) [6, 7].

Since 1960, the AMD impacts on the environment were recognized as a major pollution source [8]. The First International Tailings Symposium held in 1972 in Tucson (Arizona) marked the beginning of international awareness of mine tailing treatment. Consequently, considerable efforts have been expended in seeking solutions to minimize the AMD impact on environment [9, 2]. AMD contamination

and its spatial migration in time need to be understood for a better rehabilitation and perfect treatment actions. The geochemical method is generally the most used and the most accurate one to study soil pollution [10]. This technique requires systematic sampling followed by a laboratory analyses, and the interpolation of the obtained results in GIS and then spatial distribution maps are compiled. However, this approach is time consuming and expensive for covering a limited area. Remote sensing and spectroradiometric measurements offer a less expensive alternative to collect, to process, and to evaluate soil contamination [2]. Indeed, several studies have highlighted the application of field and imaging for direct identification of mine soils containing pollutants (e.g., heavy metals) [11,12]. Spectral measurement analyses could be an excellent tool to study soils properties and to predict AMD pollution impact [13]. According to some authors [14,15, 16], the absorption features associated with electronic transitions of Fe^{3+} and Fe^{2+} ions in Fe-bearing minerals can be found in the visible or NIR regions, and the absorption features associated with molecular vibrations of organic and inorganic functional groups (e.g., metal-OH, H_2O , etc.) can be found in NIR and/or SWIR regions. Other authors [17, 18, 19] have investigated the relationship between spectroradiometric measurements and geochemical analyses to identify soil pollution by heavy metals. Therefore, different spectral indices were developed to discriminate iron

oxides in soils, such as hematite index (IHm; [18]), ferrous index (IFe; [18]), coloration index (IC; [20]), Madani index (IM; [19]), and Segal and Merin index (ISM; 20)).

In this study, the potential of semiempirical modeling for IOP in the soil using geochemical analyses, spectroradiometric measurements, and different iron oxide spectral indices was investigated using three different sensors: BGIS-2000 of Quickbird, TM of Landsat-5, and ASTER of TERRA. To achieve our objectives, 62 soil samples representing different IOP classes have been selected. Spectroradiometric measurements were acquired over each sample using an analytical spectral device (ASD) covering the spectral domain from 350 to 2500 nm [22]. These measurements were resampled and convolved according to the spectral bands of the three used sensors. The IOP content in each sample was evaluated by geochemical analyses in the laboratory. The five spectral indices of iron oxides presented in the literature and summarized in Section 2.6 were correlated with geochemical analysis results using second-order regression. The abandoned Kettara mine, Morocco, was selected as a case study because the spatial extent of the IOP in the soil was known [3]. Figure 1 illustrates the used research methodology.

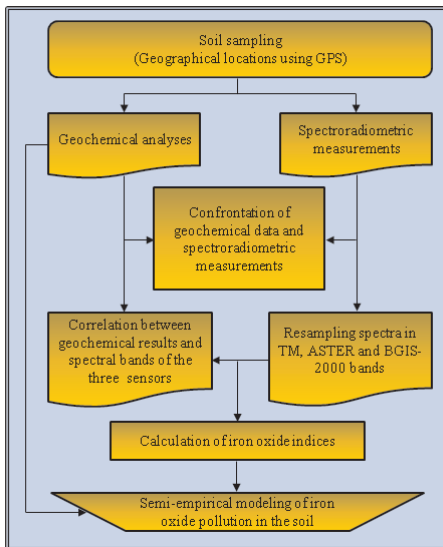


Figure 1. Research methodology.

2 MATERIAL AND METHODS

2.1 Study area

The abandoned Kettara mine, located approximately 35 km northwest of Marrakech city (Morocco), was

chosen as study area. The population of Kettara village is around 2000 people. During 44 years, from 1938 to 1981, the mine was exploited for iron oxide and iron sulfide minerals. The mining exploitation and the ore enrichment process had generated more than 3 Mt of mine wastes stockpiled over an area of 30 ha. These mine wastes contain 1.6%–14.5% of sulfur, mainly sulfide minerals [3]. They have generated significant AMD. Previous studies have shown that groundwater and surface drainage water at the Kettara mine site are contaminated [3]. Furthermore, Kettara surrounding soils are contaminated by heavy metals (Fe, Cu, etc.) as well as by secondary minerals [3, 10].

2.2 Soil sampling

Special consideration was given to the used criteria to select sampling points' locations. After a review of topographic and geologic maps, and according to previous studies [3, 10] on the abandoned Kettara mine, soil sampling was established in three different areas based on the location of the mine wastes. Three groups of soil samples have been collected: (i) samples located upstream of the mine wastes (13 samples); (ii) samples located nearby mine wastes (27 samples); and (iii) samples located downstream of the mine site (22 samples). Finally, 62 soil samples were collected within the study site, approximately 6 km². Geographic coordinates of sampling points were measured using a GPS within ± 5 m accuracy, with the "Lambert North Morocco" map projection.

2.3 Spectroradiometric measurements

Spectroradiometric measurements were acquired above 62 soils samples using the spectroradiometer ASD at the University of Ottawa. This instrument is equipped with three detectors operating in the VNIR and SWIR (350–2500 nm). It allows a continuous spectra with 1.4 nm sampling step in the spectral areas from 350 to 1000 nm and 2 nm in the areas of 1000–2500 nm [23]. The system resamples the measurements with 1 nm step, which allows the acquisition of 2151 contiguous bands per spectrum. The sensor is characterized by the programming capacity of the integration time, which allows an excellent performance of the signal-to-noise ratio, as well as a great stability [22]. Measurements were taken at the laboratory using two halogen lamps of 500 W each equipped with an electrical current regulator. The data were acquired at the nadir with a FOV of 25° and an illumination angle of approximately 5° from the vertical. The spectroradiometer was installed on a tripod with a height of approximately 40 cm of the target, which makes possible to observe a surface of 79 cm² approximately. The reflectance factor of each

soil sample was calculated according to Jackson et al.'s method [24].

2.4 Geochemical analysis

The collected samples were dried in an oven at 100°C, sieved through a 2-mm mesh, and subsequently pulverized to completion of 63-µm grain sizes. Geochemical analyses of the IOP were carried out at Acme Analytical Laboratories Ltd. (Vancouver, Canada), accredited under ISO 9002. Soil samples were digested using a strong multiacid method that dissolves most minerals. Then, 0.25 g split was heated in HNO₃-HClO₄-HF to fuming and taken to dryness. The residue was further dissolved in HCl. Then, solutions were analyzed using inductively coupled plasma mass spectrometry (ICP-MS).

2.5 Considered sensors

Three sensors were selected: BGIS-2000 of Quickbird, TM of Landsat-5 satellite, and ASTER of TERRA. This choice will highlight their spatial and spectral potentials to discriminate between different levels of IOP in the soil.

Table 1. Iron oxide indices proposed in the literature.

| index | Sensors | | | Authors |
|----------|-------------------------------|-------------------------------|-------------------------------|-----------------------|
| | BGIS2000 | TM | ASTER | |
| I_{SM} | $\frac{B3}{B2}$ | $\frac{B3}{B2}$ | $\frac{B2}{B1}$ | Segal & Merin (1989) |
| I_C | $\frac{(B3 - B2)}{B3}$ | $\frac{(B3 - B2)}{B3}$ | $\frac{(B2 - B1)}{B2}$ | Escadafal (1993) |
| I_{Fe} | $\frac{(B3 - B2)}{(B3 + B2)}$ | $\frac{(B3 - B2)}{(B3 + B2)}$ | $\frac{(B2 - B1)}{(B2 + B1)}$ | Madeira et al. (1997) |
| I_{Hm} | $\frac{B3^2}{(B1 * B2^2)}$ | $\frac{B3^2}{(B1 * B2^2)}$ | $\frac{B2^2}{B1^3}$ | Madeira et al. (1997) |
| IM | NA | $\frac{B4}{B5}$ | $\frac{B3}{B4}$ | Madani (2009) |

2.6 Spectral indices

Chemical composition and color are the most important criteria for the soil characterization, because they arise from their mineralogical composition and their organic matter content [25]. Indeed, the latter is responsible for soil dark color, while the iron oxide and hydroxides give them a red color, and the carbonates make them bright [26]. In the literature, several studies have shown a strong correlation between the soil colors and their reflectances in the visible bands [27,28,29]. Exploiting this correlation, Escadafal et al. [20] developed a simple normalized ratio of red and blue bands, which is called the coloration index (I_c). The reflectance in this part of the spectra is highly marked by the absorption phenomena caused by iron oxides (hematite and goethite). Moreover, to quantify and to map the iron oxides in lateritic soils in Brazil, Madeira et al. [18] developed two spectral indices: ferrous index (IFe), called also Hue index, and hematite index (IHm). This allows the estimation of the hematite content in soils, and the IFe allows the prediction of hematite/(hematite + goethite) ratio. Furthermore, Madani [19] showed that the ETM+ bands 4 and 5 (NIR and MIR) ratio discriminate favorably the rich iron oxide areas in Saudi Arabia. He reported that the success of this ratio index (IM) is attributing to absorption of pyrite, hematite, goethite, and magnetite in the band 4. Finally, according to Secal and Merin [21], the ratio index (ISM) of the red and green TM bands has been successfully used to map iron-bearing rocks in Lisbon Valley, Utah. The equations of these five indices are listed in Table 1. The considered indices will be tested in this study to detect and map IOP generated by the AMD in the Kettara soil.

3 RESULTS AND DISCUSSION

The exploratory data analysis statistics has revealed three pollution classes: the nonpolluted soil class, which is characterized by low iron oxide content, less than 5.69%; the moderate polluted soils class, where the iron oxide content is varied between 5.69 and 10%; and finally, the highly contaminated soil class, which is characterized by iron oxide content superior to 10% and reaching 37.49%.

The spectral signatures of the soil samples show many absorption features of iron oxides. Indeed, the hematite absorption features occur at the 550, 630, and 860 nm, and the goethite ones occur at the 480, 650, and 920 nm. Obviously, the depth of these absorption features is a function of the IOP content in the studied samples. The moderate polluted soils spectra exhibit many absorption features of iron oxide with relatively moderate depths, especially in the VNIR.

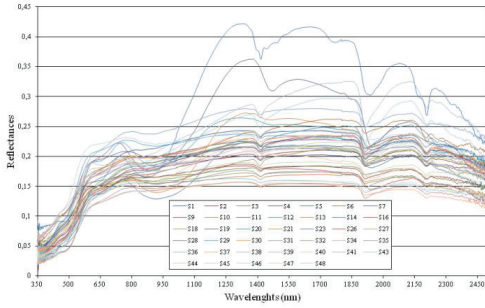


Figure 2. Spectral signatures of soil samples.

Moreover, in the SWIR regions (1000–2500 nm), the contaminated soils differ significantly from moderate or not polluted soils. Except the water and hydroxyl absorption bands existing in all soil spectra, some polluted soils show the Kaolinite and Illite absorption peaks. Although the shape of spectra varies depending on several soil intrinsic variables (texture, color, brightness, etc.), the spectral signatures of highly contaminated soils are higher than moderate and non-contaminated soils.

Table 2. Semiempirical models established between spectral indices and the geochemical analyses.

| Index | Sensor | Semiempirical models | R^2 |
|-------|--------|----------------------------------|-------|
| I_M | TM | $y = 170.80x^2 - 289.80x + 128$ | 0.90 |
| | ASTER | $y = 171.40x^2 - 292.13x + 130$ | 0.93 |
| IOPI | TM | $y = 2228.90x^2 + 29.74x + 5.09$ | 0.73 |
| | ASTER | $y = 666.24x^2 + 155.30x + 2.44$ | 0.65 |

The spectroradiometric measurements were resampled and convolved in the three sensors bands. Generally, independently to the used sensor, in all bands, the reflectance values vary significantly depending on the IOP content in the soil, especially in the SWIR bands. The correlation between the spectral bands and the geochemical analyses using a second-order regression analysis shows that the VNIR bands of the three sensors do not correlate significantly. Therefore, it is unlikely that these bands can detect IOP in the soil. However, the MIR bands (TM-5 and ASTER-4) are more sensitive to the IOP. They show a very good correlation, $R^2 = 87\%$.

Moreover, the SWIR band TM-7 and its homologous ASTER-7 show a significant correlation, respectively, $R^2 = 61\%$ and $R^2 = 63\%$. Finally, the bands 5, 6, and 8 of ASTER exhibit acceptable R^2 , respectively, 50%, 51%, and 57%.

The spectral indices described in Section 2.6 for the iron oxide detection were correlated with geochemical analyses; this was done by considering all soil samples and only the contaminated soils. Regardless of the used sensor and the soil samples, the R^2 of ISM, IC, and IFe never exceed 22%, which is insignificant. In addition, IHm shows no sensitivity to the IOP in the soil. Indeed, its R^2 does not exceed 5% for BGIS-2000 and TM sensors and 12% for ASTER. According to these results, these indices can never be sensitive to IOP in the soil. Yet, Madani index (IM) showed excellent R^2 , 90% for TM and 93% for ASTER. Although this index has been developed for iron mineral exploration, it seems to be a good potential for modeling IOP generated by AMD. The IM exploits the MIR band, which is significantly correlated (87%) with geochemical analyses. This index could not be calculated for BGIS-2000 sensor due to the absence of MIR band.

In addition to the considered indices, multiple spectral band combinations in the VNIR and SWIR and geochemical analyses were correlated. A simple difference between TM-5 and TM-7 (and their homologous ASTER-4 and ASTER-7), named iron oxide pollution index (IOPI), was significantly correlated, 73% for TM and 65% for ASTER. The semiempirical model equations established between IM and IOPI and geochemical analyses for IOP prediction in the soil using TM and ASTER sensors are listed in Table 2.

4 CONCLUSIONS

The obtained results show that whatever the used sensor, the MIR bands (TM-5 and ASTER-4) are the most sensitive to the IOP, $R^2 = 87\%$. In the SWIR region, TM-7 and ASTER-7 bands correlate significantly with 61% and 63%, respectively. Furthermore, independently to the used sensor and the IOP level in the soil, the R^2 do not exceed 22% for the indices based on the VNIR bands, such as: ISM, IC, IHm, and IFe. Only the indices that include the MIR bands in their equations correlate significantly. Indeed, the IM index provided the best prediction model of IOP in the soil with $R^2 = 90\%$ for TM and $R^2 = 93\%$ for ASTER. Moreover, the IOPI proposed in this study, based on a simple difference between TM-5 and TM-7 or ASTER-4 and ASTER-7, shows a significant prediction model with $R^2 = 73\%$. Unfortunately, no model was established for BGIS-2000 sensor because the SWIR bands are absent.

REFERENCES

- [1] U.S. Bureau of Mines. Proceedings of the International Land Reclamation and Mine Drainage Conference and 3rd International Conference on the Abatement of Acidic Drainage; U.S. Bureau of Mines Special Publication SP 06D-94; pp. 1–4, 1994.
- [2] J. Shang, B. Morris, P. Howarth, J. Lévesque, K. Staenz, and B. Neville, 'Mapping mine tailing surface mineralogy using hyperspectral remote sensing'. *Canadian Journal of Remote Sensing*, vol. 35, no. 1, pp. 126–141, 2009.
- [3] R. Hakkou, M. Benzaazoua, and B. Bussière, 'Acid Mine Drainage at the Abandoned Kettara Mine (Morocco): Environmental Characterization'. *Mine Water Env.*, vol. 27, pp. 145–159, 2008.
- [4] B. Bussière, 'Acid mine drainage from abandoned mine sites: problematic and reclamation approaches'. *Proceeding of International Symposium on Geo-environmental Engineering, ISGE2009, Septembre 8–10, Hangzhou, China*, 16 pages, 2009.
- [5] F.G. Bell, S.E.T. Bullock, T.F.J. Halbach, and P. Lindsay, 'Environmental impacts associated with an abandoned mine in the Witbank Coalfield, South Africa'. *International Journal of Coal Geology*, vol. 45, pp. 195–216, 2001.
- [6] C.N. Alpers, D.W. Blowes, D.K. Nordstrom, and J.L. Jambor, 'Secondary minerals and acid-mine water chemistry'. *Short Course Handbook on Environmental Geochemistry of Sulfide Mine Wastes*, Mineralogical Assoc of Canada, Waterloo, Ontario, Canada, pp. 247–270, 1994.
- [7] K.D. Nordstrom, and C.N. Alpers, 'Geochemistry of acid mine waters'. In: Plumlee GS, Logsdon MJ (eds), *The Environmental Geochemistry of Mineral Deposits, Part A: Processes, Techniques, and Health Issues*, *Rev Econ Geol.*, vol. 6A, pp. 133–160, 1999.
- [8] C.L. Aplin, and G.O. Argall (Editors), 'Tailing Disposal Today. Proceedings of the 1st International Tailing Symposium. Miller Freeman Publications, Inc., San Francisco, USA, 861 pages, 1973.
- [9] M.J. Puro, W.B. Kipkie, R.A. Knapp, T.J. McDonald, and R.A. Stuparyk, 'Inco's Copper Cliff tailings area'. In *Proceedings of the Sudbury'95 Conference on Mining and the Environment, 28 May – 1 June 1995, Sudbury, Ontario*. Edited by T.P. Hynes and M.C. Blanchette. Laurentian University, Sudbury, Ontario, Canada, vol. 1, pp. 181–191, 1995.
- [10] A. Khalil, L. Hanich, A. Bannari, L. Zouhri, O. Pourret, and R. Hakkou, 'Assessment of soil contamination around an abandoned mine in a semi-arid environment using geochemistry and geostatistics: Pre-work of geochemical process modeling with numerical models'. *Journal of geochemical exploration*, vol. 125, pp. 117–129.
- [11] G.A. Swayze, K.S. Smith, R.N. Clark, S.J. Sutley, R.M. Pearson, J.S. Vance, P.L. Hageman, P.H. Briggs, A.L. Meier, M.J. Singleton, and S. Roth, 'Using imaging spectroscopy to map acid mine waste'. *Environmental Science and Technology*, vol. 34, pp. 47–54, 2000.
- [12] D.J. Williams, J.M. Bigham, C.A. Cravotta, S.J. Traina, J.E. Anderson, and J.G. Lyon, 'Assessing mine drainage pH from the color and spectral reflectance of chemical precipitates'. *Applied Geochemistry*, vol. 17, pp. 1273–1286, 2002.
- [13] X.-L. Xie, X.-Z. Pan, and B. Sun, 'Visible and Near-Infrared Diffuse Reflectance Spectroscopy for Prediction of Soil Properties near a Copper Smelter'. *Pedosphere*, vol. 22, no. 3, pp. 351–366, 2012.
- [14] G.R. Hunt, and J.W. Salisbury, 'Visible and near-infrared spectra of minerals and rocks: I. Silicate minerals'. *Modern Geol.*, vol. 1, pp. 283–300, 1970.
- [15] G.R. Hunt, J.W. Salisbury and C.J. Lenhoff, 'Visible and near-infrared spectra of minerals and rocks: III. Oxides and hydroxides'. *Modern Geol.*, vol. 2, pp. 195–205, 1971.
- [16] Ben-Dor, 'Quantitative remote sensing of soil properties'. *Adv. Agron.*, vol. 75, pp. 173–243, 2002.
- [17] M. Pouget, J. Madeira, E. Le-Floch, and S. Kamal, 'Caractéristiques spectrales des surfaces sableuses de la région côtière nord-ouest de l'Égypte: application aux données satellitaires SPOT'. *Colloques et Séminaires ORSTOM Ilèmes Journées de Télédétection*, Paris, France, du 4 au 6 Décembre (Paris: Éditions ORSTOM), pp. 27–38, 1991.
- [18] J. Madeira, A. Bedidi, B. Cerveille, M. Pouget, and N. Flay, 'Visible spectrometric indices of hematite (Hm) and goethite (Gt) content in lateritic soils: the application of a Thematic Mapper (TM) image for soil-mapping in Brasilia, Brazil'. *International Journal of Remote Sensing*, vol. 18, no. 13, pp. 2853–2852, 1997.
- [19] A. Madani, 'Utilization of Landsat ETM+ Data for Mapping Gossans and Iron Rich Zones Exposed at Bahrah Area, Western Arabian Shield, Saudi Arabia'. *JKAU: Earth Science*, vol. 20, no. 1, pp. 35–49, 2009.
- [20] R. Escadafal, A. Belghit, et A. Ben-Moussa, 'Indices spectraux pour la télédétection de la dégradation des milieux naturels en Tunisie aride'. In G. Guyot (réd.) *Actes du 6^{ème} Symposium international sur les mesures physiques et signatures en télédétection*, Val d'Isère (France), 17–24 janvier, ISPRS-CNES, pp. 253–259, 1994.
- [21] D.B. Segal, and I.S. Merin, 'Successful Use of Landsat Thematic Mapper Data for Mapping Hydrocarbon Microseepage-Induced Minearologic Alteration, Lisbon Valley, Utah'. *Photogrammetric Engineering and R. S.*, vol. 55, no. 8, pp. 1137–1145, 1989.
- [22] ASD Inc., 'Precision Solution To Your Spectroradiometer and Near-IR Analyser Needs: Technical Guide'. <http://www.asdi.com>, 1999.
- [23] A. Bannari, K.S. Khurshid, K. Staenz, and J. Schwarz, 'A Comparison of Hyperspectral Chlorophyll Indices for Wheat Crop Chlorophyll Content Estimation Using Laboratory Reflectance Measurements'. *IEEE Transaction On Geosciences and Remote Sensing*, vol. 45, no. 10, pp. 3063–3073, 2007.
- [24] R.D. Jackson, P.J. Ninter, J. Paul, R.J. Reganato, J. Robert, and S.B. Adso, 'Haud-held radiometry'. Department of Agriculture, Science and Education Administration, Agricultural reviews and Manuals ARM-W-19, Oakland (California), 84 pages, 1980.
- [25] J.R. Irons, R.A. Weismiller, and G.W. Petersen, 'Soil Reflectance'. In: Asrar, G. Editor (1989). *Theory and Applications of Optical Remote Sensing*. John Wiley & Sons, Inc., N.Y., pp. 66–106, 1989.

- [26] J.P. Legros, 'Cartographies des sols : De l'analyse spatiale à la gestion des territoires'. Collection Gérer l'environnement. Presses Polytechniques et Universitaires Romandes, Lausanne, Suisse, 321 pages, 1996.
- [27] R. Escadafal, M.C. Girard, and D. Courault, 'Munsell Soil Color and Reflectance in the Visible Spectral Bands of Landsat MSS and MT Data'. *R. S. of Environment*, vol. 27, no. 1, pp. 37–46, 1989.
- [28] D.F. Post, E.H. Horvath, W.M. Lucas, S.A. White, M.J. Ehasz, and A.K. Batchily, 'Relation Between Soil Color and Landsat Reflectance on Semiarid Rangelands'. *Soil Science Society of America Journal*, vol. 58, no. 6, pp. 1809–1816, 1994.
- [29] A. Bannari, A.R. Huete, D. Morin, et F. Zagolski, 'Effets de la couleur et de la brillance du sol sur les indices de végétation'. *International Journal of Remote Sensing*, vol. 17, no. 10, pp. 1885–1906, 1996.

Research on the strategy of controlling Ordos industrialization and eco-environment system based on SD model

L.M. Li, J.H. Zhen, S.S. Luo & S.P. He

College of Geography Science, Inner Mongolia Normal University, Hohhot, China

ABSTRACT: Based on the connotative meaning of industrialization and eco-environment, the current situation of Ordos, and the simulated status of the industrialization and eco-environment system, the author chose five kinds of typical development patterns to make scenario analysis using SD model. The results showed that: SD model solved the complexity, feedback, and time-varying problems of industrialization and eco-environment system. In the future, Ordos should take the road of industrial economy and proceed resources development and environmental protection coordinately. However, in terms of the discussion of the conclusion and the theory of how to set the various indexes, its effectiveness and correctness remain to be tested by the practice.

1 INTRODUCTION

The development of regional industrialization promotes the economic level, but inevitably it also brings eco-environment problems. In 1995, according to the statistic study of the industrialization level, economic growth, and environmental quality, Grossman and Krueger put forward the hypothesis of environmental Cruz Nez curve (EKC curve) [1]. Then, foreign scholars have done a lot of empirical researches to examine EKC curve's existence. At the same time, many domestic scholars [2-3] also studied the relationship between industrialization and eco-environment and made a lot of achievements.

In view of the above-mentioned researches, scholars mostly focused on the analysis of the relationship between industrialization and eco-environment. However, their recommendations about coordinated development of the two excessively focused on the description of the policy based on the qualitative research, and they lacked the researches of the quantitative control strategy. Based on this, the research views Ordos, which has fragile eco-environment and rich natural resources as the research object. Using system dynamics (SD) model, the industrialization and eco-environment are brought into the unified system. Based on the current situation of simulation, we used the quantitative control strategy for exploring their coordinative development.

2 THE CONSTRUCTION OF INDUSTRIALIZATION AND ECO-ENVIRONMENT SYSTEM BASED ON SD MODEL

System dynamics (SD) was created by the scholar Jay. W. Forrester of Massachusetts Institute of Technology in America in 1956. It is a kind of simulation experimental approach based on the feedback control theory and mathematical computer stimulation technology to research and analyze the feedback system of complex dynamics [4]. Because the industrialization and eco-environment system is a complex giant system, this system is divided into five subsystems including population subsystem, resource subsystem, economic and social subsystem, environment subsystem, and the industrial subsystem. An in-depth analysis of each subsystem contributes to the overall grasp of the macro system.

2.1 *The determination of the system boundary*

System dynamics thinks of systemic behavior stems from the structure of the system, and the external environment's influence on the system behavior is put into effect by the internal structure [5]. Therefore, the key of system modeling is to reasonably determine the system boundary and the analysis of its structure. According to the purpose of model established, this study determines the boundary space simulation is

Ordos. The total land area is 87,000 square kilometers. Model of time boundary is from 2010 to 2025. The benchmark for the year is 2010. The interval of the simulation time is 1 year.

2.2 System structure analysis

Because the industrialization and eco-environment system are divided into population—resources—environment—social and economic—industrialization, this study follows the systematic, dynamic, scientific, and feasibility principles, considering the modeling objective, combined with the actual of the Ordos, and the five subsystems can be further subdivided into population, GDP, social development, investment in fixed assets, per capita resources, regional water consumption, total industrial energy consumption, environmental pollution, environmental governance, and the total industrial output value of 10 modules. Because the flow chart of the system is too large, this will not give.

2.3 Establish and model equation of the system parameters

The determination method of system parameters in this study are the use of statistical data system history, using the arithmetic average table function to determine the parameters, using the regression method to determine the parameters, and using the GM (1,1) model updating parameters and so on. Use of statistical data system history parameter arithmetic average determined are population mortality, cultivated area, and so on; using table parameter function determined mainly has the natural population growth rate, the rate of GDP growth, coal consumption coefficient, and so on; under the support of SPSS software, using regression method for the determination of the parameters of the main fiscal expenditure, industrial labor productivity, liquefied petroleum consumption, gross value of industrial output growth rate, and so on; using GM (1,1) parameters of model updating are mainly single bit output value of industrial “three wastes” emissions and the “three wastes” treatment rate. The model equation is based on determining the feedback loop flow position and flow rate, using auxiliary variables, table function, and parameters, to determine the number of each variable in the relationship between flow chart of mutual response to quantitative analysis and the dynamic behavior of a system. According to the model, flow chart, the model contains 85 variables and 23 parameters, a total of 108 equations. Due to limited space, this is no longer the one that one lists.

2.4 Model test

After completing the reality check test using Vensim PLE software, the statistical data from 2000 to 2010

Ordos is based on historical test model to verify the reliability of the model. Five indexes are selected in this study including population, GDP, the total water consumption, emissions life trash quantity, and gross value of industrial output, taking 2000 as the basic data and taking 1 year to validate the model. The simulation results show that the relative error in the -9.23%–9.60%, visible model simulation value, relative error within reasonable limits and the actual value, the simulation results is more reliable and has better behavior replication capacity; simulation research scenario can be carried out by industrialization and eco-environment system in Ordos.

3 MODEL SIMULATION

According to the actual situation and the principle of Ordos industrialization and eco-environment system to determine the control parameters, the choice of the natural population growth rate, the growth rate of GDP, fixed assets investment, education accounts for the proportion of financial expenditure, the proportion of GDP accounted for the proportion of the funds for science and technology, the growth rate of industrial water consumption, coal consumption of 10 indicators, industrial power consumption coefficient, coefficient of environmental protection and control investment accounted for the proportion of GDP, the proportion of industrial investment as the control parameter, simulation of various schemes, the simulation scheme of the control variables in Table 1. Run a simulation and automatically simulate on change module using the Vensim PLE software, simulated by five adjustment schemes of 10 control variable values to the model, so as to get the result control strategy of each simulation program of industrialization and the eco-environment system Ordos. Because the index number is more, so here only eight indexes were chosen to express the core simulation results in the simulation scheme (Table 2).

4 THE RESULTS OF SIMULATION ANALYSIS

4.1 Natural development pattern (I)

In the assumed model, the initial values of control parameters are unchanged, and it simulates the future to the current development model. The results show (Table 2) that this mode of development has achieved rapid growth in total population and economy. The regional total water consumption is in the acceptable range considering the total amount of water, but the pressure on water resources is more obvious. Pollution's

Table 1. Industrialization and eco-environment system regulation strategy of each simulation scheme of control variable value table in Ordos.

| The parameter name, | unit, | Natural | Rapid population | The rapid growth of | Resources and | Industrial resources |
|--|-------|-----------------|----------------------|---------------------|-----------------|-------------------------|
| | | development | and economic | industrial economic | environment | environment and economy |
| | | pattern (I) , | growth mode (II) , | model (III) , | protection mode | coordinated development |
| | | (IV) , | | | model (V) , | |
| Natural population growth rate, | %, | 7.59, | 9.59, | 7.59, | 6.59, | 7.59, |
| The growth rate of GDP, | %, | 7.50, | 9.00, | 8.50, | 5.50, | 6.50, |
| Education accounts for the proportion of | | | | | | |
| fiscal expenditure, | %, | 10.54, | 11.54, | 11.54, | 12.54, | 11.54, |
| Science and technology funds spending | | | | | | |
| accounts for the proportion of GDP, | %, | 0.05, | 0.06, | 0.065, | 0.07, | 0.06, |
| Investment in fixed assets ratio, | %, | 68.10, | 70.10, | 69.10, | 67.10, | 68.10, |
| Water consumption growth rate, | %, | 0.93, | 1.13, | 1.33, | 0.75, | 0.88, |
| Industrial coal consumption coefficient, | —, | 0.1858, | 0.2058, | 0.2258, | 0.1558, | 0.1808, |
| Industrial electricity consumption | | | | | | |
| coefficient, | —, | 0.3817, | 0.4017, | 0.4217, | 0.3517, | 0.3717, |
| Investment of environmental pollution | | | | | | |
| treatment in the proportion of GDP | | | | | | |
| accounted for, | %, | 0.056, | 0.053, | 0.050, | 0.086, | 0.076, |
| Industrial investment proportion, | %, | 53.73, | 54.73, | 55.73, | 51.73, | 54.73, |

Table 2. Control strategy of each simulation scheme running result of industrialization and eco-environment system of Ordos.

| | 2012, | | | | | 2025, | | | | |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| | I, | II, | III, | IV, | V, | I, | II, | III, | IV, | V, |
| The number of population (Ten thousand | | | | | | | | | | |
| people), | 1.4791, | 1.482, | 1.479, | 1.4776, | 1.4791, | 1.7807, | 1.843, | 1.7901, | 1.7405, | 1.772, |
| GDP (Billion yuan) , | 312.586, | 316.948, | 315.494, | 306.77, | 309.678, | 800.349, | 971.7, | 911.124, | 615.312, | 702.191, |
| The regional total water consumption (10000 | | | | | | | | | | |
| tons), | 122223, | 122465, | 122708, | 122005, | 122163, | 137853, | 141727, | 145702, | 134451, | 136900, |
| Industrial wastewater emissions (10000 tons), | 1503.93, | 1520.53, | 1523.21, | 1476.34, | 1488.56, | 3792.26, | 3884.26, | 4151.03, | 2519.81, | 3551.93, |
| Industrial emissions (million m ³), | 235.364, | 241.963, | 243.97, | 237.385, | 236.637, | 642.437, | 666.735, | 713.02, | 428.533, | 603.155, |
| Industrial solid wastes discharged (10000 tons), | 6.50836, | 6.74472, | 6.82351, | 6.65702, | 6.59568, | 17.849, | 18.6416, | 19.8951, | 11.993, | 16.7657, |
| Pollution influence on economy and Society | | | | | | | | | | |
| (Million yuan), | 440.346, | 451.924, | 455.369, | 442.685, | 441.888, | 1199.2, | 1243.19, | 1329.9, | 799.218, | 1125.76, |
| Total industrial output value (Billion yuan), | 125.228, | 135.376, | 138.546, | 131.011, | 130.039, | 350.179, | 366.665, | 398.575, | 228.583, | 331.28, |

effects on the economy and society have increased year by year indicating that the development model of pressure on the ecological environment is larger. Rapid growth of total industrial output value of the emissions of industrial waste and pollution impact

on society and economy is increasing. Thus, natural development pattern is the general development mode. Although this model has achieved rapid growth of the total industrial output value, the pressure on the eco-environment is obvious.

4.2 *Rapid population and economic growth mode (II)*

To realize the rapid growth of population and economy, we make the corresponding adjustment of control parameters (Table 1). The simulation results show that this model has realized the high speed growth of economy and population: the population growth from 1.482 million in 2012 to 1.843 million, GDP increased from 316.948 billion yuan to 971.7 billion yuan, total industrial output value increased from 125.228 billion yuan in 2012 to 366.665 billion yuan in 2025. However, this pattern of regional consumption, emissions of industrial waste, and pollution effect on the economy and society are five kinds of development patterns of second, coupled with a large number of population growth, it is certainly not conducive to the economic and social and ecological environment for the sustained and healthy development.

4.3 *The rapid growth of industrial economic model (III)*

We use the rapid growth of the industrial economy to promote economic and social development. Value of each control parameter adjustment is shown in Table 1. Through the simulation, the mode of development makes the total industrial output value increase, from 138.546 billion yuan in 2012, an increase of 398.575 billion yuan in 2025, an average annual growth rate of 13.41%. However, in this mode, the influence of regional consumption, industrial waste emissions and pollution on the economic society in the five modes is the highest (Table 2), bringing great pressure to the ecological environment. From the characteristics of the mode of development, it is suitable for the early stages of industrialization. However, in Ordos, now in the middle stage of industrialization, so the development mode is not suitable for the present stage.

4.4 *Resources and environment protection mode (IV)*

The model takes resource intensively and highly efficient utilization and environmental improvement as the goal. Table 1 shows the values of each control parameter adjusted. Through the simulation, the resources and environmental protection model indeed largely reduce water and energy consumption, reduce industrial waste emissions, substantially weakening the pollution effect on the economy and society, to control the growth of population, and the blind GDP (Table 2). However, the drawbacks of this model are the most excessive

pursuit of resource conservation and environmental protection, coordinated development of economy and society seriously enough, and easy to make some parts of poverty. This mode is suitable for the economic and material highly developed in the later stages of industrialization.

4.5 *Industrial resources environment and economy coordinated development model (V)*

This model is aimed at the coordinated development of industrial economy and resource environment. Value of each control parameter adjustment is shown in Table 1. This mode of control parameters, the simulation, and this model is not only a very good control of the unchecked growth of the population but greatly reduced the water resources and energy consumption and reduced the impact of industrial waste emissions and pollution on the economic society to a great extent (Table 2). Although the GDP and gross value of industrial output will be affected, it effectively controls the resource consumption and environmental damage, and it is of benefit to the more sustainable and coordinated development in Ordos and the future's economic benefits of society and environment. Therefore, it is the most suitable mode in the industrialization of Ordos.

5 DISCUSSION AND CONCLUSION

In this study, the following conclusions are obtained and discussed.

- 1 SD model achieves a better situation and forecast for the industrialization and eco-environment system of Ordos to solve quantitative simulation, the coordinated development of industrialization, and the eco-environment system control strategy of fewer problems. However, because of the limitations of the level of theory, the analysis of the system structure is not fully comprehensible, and parameters determined the method of regression analysis and GM (1,1) model and other restrictions. Therefore, the inevitable emergence of certain deviation exists in the final results of every simulation scheme.
- 2 After this research in the analysis of the simulation results of each scheme, it obtains the optimal control strategy of industrial resource environment and economy coordinated development model (V) for future industrial and eco-environment system coordination of Ordos sustainable development. However, this conclusion is only a theoretical study; the validity and correctness remain to be tested by the practice.

ACKNOWLEDGMENTS

This paper is funded by the project of National Natural Science Fund (41261109) and Inner Mongolia Natural Science Foundation Project (2012MS0617).

REFERENCES

- [1] Grossman G M, Krueger A B 1995. Economic growth and the environment. *Quarterly Journal of Economics* 110(2):353–377.
- [2] Zhao, Yue-long, Liu, Yan-hua 1996. A stubby on The relationship of fragile eco-environment and industrialization. *Economic Geography* 16(2):86–90.
- [3] Zhang, Huan 2014. The plight of resource and Environment and breakthrough ideas in the construction of ecological civilization in the industrialization progress of Hubei province. *Territory & Natural Resources Study* 27(1):27–29.
- [4] Wang, Zhen-jiang 1988. An introduction to system dynamics. Shanghai: Shanghai science and technology literature press.
- [5] Wang, Qi-fan 1995. The high-level system dynamics. Beijing: Tsinghua University Press.

Environmental aspects of raw materials SMART processing

J. Spišák & M. Zelko

Development and realization workplace of raw materials extracting and treatment, Faculty of Mining, Ecology, Process Control and Geotechnology, Technical University of Kosice, Kosice, Slovak Republic

ABSTRACT: Comprehensive environmental observation, eco-innovation, and informatization seeking to smartization are essential to ensure the delivery of the long-term data and information required to address the shift toward a green economy. Digitalization, informatization, and smartization could be used to assess and predict the condition, status, and trends of natural resources including raw materials, climate, ecosystems, and ecosystem services, as well as to evaluate low-carbon, climate mitigation, adaptation policies, and options across all sectors of the economy. Smartization and knowledge gained from these systems would be used to stimulate the smart use of strategic resources, to support the development of evidence-based directions, to foster new environmental services, and to develop new opportunities in global markets.

1 INTRODUCTION

EU countries, societies, companies, and individuals face a major challenge to establish a sustainable balance between human needs and the environment. Environmental resources, including water, air, biomass, fertile soils, biodiversity, ecosystems, and the services they provide, underpin the functioning of the European and global economy and quality of life. Research in this area must contribute to reversing the trends that damage the environment and to ensuring that ecosystems continue to provide the resources, goods, and services that are essential for well-being, and economic prosperity. The aim is to therefore provide knowledge for the management of natural resources, which achieves a sustainable balance between limited resources and the needs of society and the economy.

Research and innovation within this field should focus on the further understanding of the functioning of ecosystems, their interactions with social systems, and their role in sustaining the economy and human well-being. Society's actions risk triggering changes in the environment that are irreversible and alter the character of ecosystems (Zelko, Petruš and Spišák, 2010).

1.1 *Ensuring the sustainable supply of raw materials*

Recent trends indicate that demand for raw materials will be driven by the development of emerging economies and by the rapid diffusion of key enabling technologies.

Europe has to ensure a sustainable management and secure a sustainable supply of raw materials from inside and outside its borders for all sectors that depend on access to raw materials.

Policy targets for critical raw materials are outlined in the Commission's Raw Materials Initiative 26 (Horizon 2020, 2011).

The aim of this activity is to therefore improve the knowledge based on raw materials and develop innovative solutions for the cost-effective and environmentally friendly exploration, extraction, processing, recycling, and recovery of raw materials and for their substitution by economically attractive alternatives with a lower environmental impact.



Figure 1. Introduced list of critical raw materials (CRMs) in 2013.

Research and innovation within this raw material area should focus on the following (Lavrin and Zelko, 2006):

- 1 improving the knowledge based on the availability of raw materials;

- 2 promoting the sustainable supply and use of raw materials, covering exploration, extraction, processing, recycling, and recovery;
- 3 finding alternatives for critical raw materials; and
- 4 improving societal awareness and skills on raw materials.

1.2 *Shift toward a green economy through eco-innovation*

The EU cannot prosper in a world of ever increasing resource consumption, environmental degradation, and biodiversity loss. Decoupling growth from the use of natural resources requires structural changes in how such resources are used, reused, and managed, while safeguarding our environment. Eco-innovations will enable us to reduce pressure on the environment, increase resource efficiency, and put EU on the path to a resource and energy efficient economy. Eco-innovation also creates major opportunities for growth and jobs and increases European competitiveness within the global market.

The aim of this activity is to therefore foster all forms of eco-innovation that enable the transition to a green economy by strengthening eco-innovative technologies, processes, services, and products and by boosting their market uptake. All forms of innovation (incremental and radical, combining technological, organizational, societal, behavioral, business and policy innovation, and participation of civil society) will be supported. This will underpin a more circular economy, while reducing environmental impacts and taking account of rebound effects on the environment. This will include business models, industrial symbiosis, product service systems, product design, full life cycle, and cradle-to-cradle approaches. The aim is to improve resource efficiency by reducing, in absolute terms, inputs, waste, and the release of harmful substances along the value chain and foster reuse, recycling, and resource substitution. Emphasis will be given to facilitate the transition from research to market, involving industry and notably SMEs, from the development of prototypes to their introduction in the market and replication. Networking among eco-innovators will also seek to enhance the dissemination of knowledge and better to link supply with demand.

Structural and institutional changes to shift toward a green economy need to address the main barriers to societal and market change and should aim to empower consumers, business leaders, and policy makers to adopt innovative and sustainable behavior. We also need to develop robust and

transparent tools, methods, and models to assess and enable economic, societal, and institutional changes needed to achieve a paradigm of green economy. Research and development would explore sustainable consumption patterns, socio-economic research, behavioral science, user engagement, public acceptance of innovation, and activities to improve communication and public awareness. Full use needs demonstration actions!

It is also necessary to develop robust indicators at all appropriate scales, methods, and systems to support and assess shift toward a green economy. Driven by a life-cycle approach, research and innovation will improve the quality and availability of data, measurement methods, and systems relevant to resource efficiency and eco-innovation and facilitate the development of innovative schemes. Technology assessment methodologies and integrated modeling are essential to support resource efficiency and eco-innovation policies at all levels. The results enable monitoring, assessment, and reduction of material and energy flows through the production and consumption and enable managers to integrate environmental costs and externalities into their decisions.

Innovations in digitalization and informatization can constitute a key tool to support process effectiveness. Innovative ICT tools could contribute to significant efficiency gains in productivity, notably through automated processes, real-time monitoring, and decision support systems. Through the digitalization and informatization, we can accelerate a progressive dematerialization of the economy, by increasing the shift toward digital services, and to facilitate business models of the future.

We should develop comprehensive and sustained global environmental observation by means of informatization.

Comprehensive environmental observation and informatization is essential to ensure the delivery of the long-term data and information required to address this challenge. Digitalization, informatization, and smartization could be used to assess and predict the condition, status, and trends of natural resources including raw materials, climate, ecosystems, and ecosystem services, as well as to evaluate low-carbon and climate mitigation and adaptation policies and options across all sectors of the economy. Smartization and knowledge gained from these systems would be used to stimulate the smart use of strategic resources, to support the development of evidence-based directions, to foster new environmental services and to develop new opportunities in global markets (Dorcak and Spisak, 2004).

2 CASE STUDY—THE EUROPE’S BIGGEST RESEARCH AND DEVELOPMENT PROJECT I2MINE

The European industry faces a number of challenges that require technical solutions that encompass the entire network and include the economic, social, and environmental issues. The industry needs a new image based on pioneering solutions and a modern structure that can exploit minerals at greater depths and promote both high productivity and safe working conditions. The concept of an intelligent underground mine of tomorrow would need to move from a static control model to a dynamic control system that is capable of processing large quantities of a wide range of data. This mass flow and logistics model would be based on plant design, economic parameters, and an intelligent technological logistics system that maximizes performance (Spisak and Zelko, 2010). The concept of such a future mine addresses the large international Program/Project I²Mine: Innovative technologies and concepts for intelligent deep mine of the future. Members of this project are the most prominent companies, associations, and university departments from mining sector, respectively, from sector of raw materials extraction and treatment. This project offers a solution of a comprehensive mine wide network dealing with integrated information management and control systems and all with a strong accent to address environmental impacts. Due to the large amount of various kinds of data, for online monitoring, short- and long-term planning, safe, economic, and environmental production, a new decision support instruments are incorporated there. To get significant decisions during the strategic planning process, a set of techniques have been proposed, including the expected “green” technologies. Therefore, a reference models and engineering tools for performance, environmental issues, risks, and cost analysis in the long-term raw material planning have been developed.

Our team VRP (Development and Realization Workplace of Raw Materials Extraction and Treatment), which is an independent and self-financing department of mining/BERG Faculty of Technical University of Kosice, is a member of this project and participates in three main activities. The first activity is called technological logistics system and is based on material flow monitoring, the development of mine-wide information and management systems, logistics management, and information flow. What is also included is the determination of the concept of technological logistics (TL) and mass flow control models based on economic management parameters. The final goal of this activity is to

propose a new methodology for the application of logistics principles into the technological process. In the second activity, a project titled “New methods of underground mining and processing” is solved by two subsurface method tasks of processing copper, coal, and magnesite. Magnesite processing activity is fully the responsibility of our VRP in our department. The first task is to focus on the development of new technology for heat treatment of fine-grained shares of mining and the second to develop new methods of mining processing based on advanced technologies. The third and most important activity is the demonstration and validation of solutions in concrete terms the practice—in our case, in the mining company SMZ, a.s. Jelsava. In principle, this activity represents the identification of suitable plant to demonstrate the various activities of the aforementioned working groups. Demonstration of the first activity includes and verifies a new transport and logistics system for the whole production process, based on the process model development, a new logistics system for the whole production process based on the balance model development, a new intelligent physical model based on typical elements of logistics technology, and the gravitational principle. All demonstrates on ITA furnace in Jelsava (Figure 2). Part of demonstration is a new economic cost model based on the principles of measurable systems for various specific conditions and requirements.



Figure 2. Integrated thermal apparatus.

Demonstrations are expected on the physical model built in our (VRP) detached center in SMZ, a.s. Jelsava mine and economic cost model development takes place at the Technical University in Kosice.

The core of the project is to develop breakthrough technologies for autonomous, highly selective, continuous mineral extraction processes, and machinery based on new sensor technologies, facing front

separation as well as innovative concepts for mass flow management and transportation integrating state-of-the-art technologies. The main idea—the concept of an invisible, zero impact mines requires a refined process underground that selectively extracts the mineral, thereby reducing waste. For this reason, improved extraction machines and near to face processing methods, including backfill procedures, are developed. These developments include rock mechanics and ground control solutions, incorporating health, safety, and environmental issues. The I2Mine project is developing new autonomous machinery and equipment enabling for automated extraction and road heading. Due to boundary layer detection, the amount of waste rock cut during mining operations could be reduced. Automated operation in the longer run would reduce costs. The integrated near-to-face processing approach provides significantly higher concentration of valuable mineral, thus reducing transportation costs. Integrated monitoring and on-line maintenance activities as well as advanced decision-making support based on reliable information of the entire mine would also increase productivity through reduction of downtime and could be seen as a new ecosystem services. Altogether, an increase of productivity of eco-friendly activities of at least 20% is the aim and in some operations toward 30% is envisaged (I2Mine, 2011). The work of I2Mine is dedicated to developing highly innovative tools, methods, and machinery for underground mining operations. The work creates significant advances in the knowledge-based sector, which can contribute to reaching or sometimes maintaining worldwide technology leadership in the field. The results of project would, for the mining companies, provide for the ability for selling mineral products at lower prices and increase the available mineable reserves. The equipment manufacturers can maintain and most likely improve their global leadership in providing mining technology. Universities and research facilities could build upon their knowhow, experience, research, and development as well as to attract world-class researches and students to the industry, as well as to build a health respect of what industry requires. Considering all the above facts, I2Mine can definitely transform the global competitiveness of the European raw materials and equipment producing industries with respect to the environment.

3 CONCLUSIONS

Smartization and knowledge gained from this smart, green, and integrated resource efficiency and raw materials supervising systems would be used to stimulate the smart use of strategic resources, to support the development of evidence-based directions, to foster new environmental services, and to develop

new opportunities in global markets. Digitalization, informatization, and smartization could be used to assess and predict the condition, status and trends of natural resources including raw materials, climate, ecosystems, and ecosystem services, as well as to evaluate low-carbon and climate mitigation and adaptation policies and options across all sectors of the economy. It seems that it is only a promising way of how to meet these goals. Some of our experiences from working within this field and thus solution-based projects promise to achieve such a smart supervision system attributes where smartization supports a multidimensional system management. Let us mention at least an intelligent advanced solutions and concepts, which we deal with in developing new solutions for the need of industry practice. (Digital Factory, SMART Factory, Intelligent Mine of the Future) and which play a crucial role in meeting these demands. It enables real-time distillation of raw data into actionable information and makes those insights available at any decision point, at any time. This new approach is well supported by the above-mentioned FP7 project I2Mine (Innovative Technologies and Concepts for the Intelligent Deep Mine of the Future) [3]. The project marks the start of a series of activities designed to realize the concept of an invisible, zero-impact mine (the Technical University of Kosice participates in the project). We are the member of another project, which is a part of the cross-border co-operation programmer HU-SK: Virtual reality laboratory for factory of the future. Our workplace is well established in the European structures (ETP SMR, EIP on RM, KIC on RM) and also prepared of new commitments on RM dealing with environmental issues (Magnesium Chloride and I²Mine-2 and I²Mine-pilot).

ACKNOWLEDGMENT

This contribution/publication is the result of the project implementation Innovative Technologies and Concepts for the Intelligent Deep Mine of the Future (280855).

REFERENCES

- [1] Horizon 2020 (2011). The Framework Programme for Research and Innovation, COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS, Brussels, 2011, COM(2011) 808.
- [2] Spisak J. & Zelko M., The Advanced Technologies Development Trends for the Raw Material Extraction and Treatment Area, Products and Services; from

- R&D to Final Solutions, SCIYO, Croatia, 2010, pp. 257–278;
- [3] Zelko M. & Lavrin A., New trends improving the management and technological base of production on Earth resources - 2011. In: Transactions of the Universities of Kosice, Slovakia, 2011, No. 1, pp. 18–22.
- [4] Zelko M. & Petruš M., Spišák J., Environment and risk factors as the part of the European technology platform for sustainable mineral resources, Proceedings from conference, Slovakia, 2010, pp. 131–136.
- [5] Lavrin A. & Zelko M., Knowledge sharing in regional digital ecosystems, 2006, Slovenia, Organizacija. Vol. 39, no. 3, p. 191–199.

Variation of extreme Total column ozone events over Seoul, South Korea

D. Amrita, J.K. Park & J.H. Park

Department of Rural and Agricultural Engineering, Chungbuk National University, South Korea

ABSTRACT: This article focuses on the spatial-temporal structure of total column ozone (TCO) over Korea. The ozone is capable of absorbing UV radiation wavelengths of biologically damaging ultraviolet light. Therefore, TOC plays a very important role in climate and weather on regional and global spatial scales. The phenomenon “depletion of ozone layer” is not only prominent over Antarctic region but also an established fact for mid- and high-latitude areas of the earth. Reduction in ozone amount leads to the increase in elevated solar UV level, which has adverse effects on human health and our ecosystem. A quite number of studies have been conducted for different parts of the world noticing a degrading trend of stratospheric ozone all over the world including South Korea. However, in this present work rather than average ozone change, attention is given to the variation of extreme TCO events over Seoul as extreme weather conditions, which are now recognized as key aspects of climate change study. The present study deals with daily total column ozone data from TOMS Nimbus-7 (1979–1992), Earth Probe (1996–2004), and OMI (2005–2013) developed by National Aeronautics and Space Administration, Goddard Space Flight Center. Days with TCO higher and lower than or equal to $\pm 15\%$ of average ozone (1979–2013) are considered as higher and lower ozone days, respectively. Stratospheric ozone is maximum in March and minimum in October, whereas UV index is maximum in June and July and minimum in December. If one can see a variation of high and low ozone days, it is clear that not only average TCO but also high ozone events are decreasing with time. Although other previous studies suggest that this kind of extreme events frequency is high in winter months than in summer months, in Korea, it is observed that frequency is more in spring (February to May) than that in winter (November to January). It is also observed that there is a jet stream in lower stratosphere south to Korea during April, 2010, and in contrast, its location during March, 2012 events is far north-west from Korean peninsula. On the contrary, during ozone high episode, it shifts toward Europe and may have small or no influence on TCO over Korea.

1 INTRODUCTION

The phenomenon “depletion of ozone layer” is not only prominent over Antarctic region but also an established fact for mid- and high-latitude areas of the earth. Reduction in ozone amount leads to increase in elevated solar UV level, which has adverse effects on human health and our ecosystem (Bojokov, et al. 1995; Salby and Callaghan, 1993). A quite number of studies have been conducted for different parts of the world noticing a degrading trend of stratospheric ozone all over the world including South Korea (Chen and Nunez, 1998, Den et al 2005; Krzyscin, 2002). However, in this present work rather than average ozone change, attention is given to the variation of extreme total column ozone (TCO) events over Seoul as extreme weather conditions, which are now recognized as key aspects of climate change study.

2 DATA

The present study deals with daily total column ozone data from TOMS Nimbus-7 (1979–1992), Earth Probe (1996–2004), and OMI (2005–2013)

developed by National Aeronautics and Space Administration (NASA), Goddard Space Flight Center. Days with TCO higher and lower than or equal to $\pm 15\%$ of average ozone (1979–2013) are considered as higher and lower ozone days, respectively. To remove the seasonal bias, mean TCO for each month is calculated and used in this study. Other meteorological parameters Geopotential height, tropopause temperature, and U-wind anomalies data are taken from NCEP-NCAR reanalysis from 1948 to 2013.

3 RESULTS AND DISCUSSIONS

The recent variation of TCO and UV erythemal index (monthly mean, cloud free condition) are shown in [Figure 1](#). There is a clear seasonal change in both stratospheric ozone and UV index.

[Figure 1](#) shows the monthly evolution of TCO over Korea for 10 years. It can be seen that seasonal TCO variation is characterized by a buildup through the winter and early summer seasons and a decline through middle summer. April represents the end of the buildup period (>370 DU). After reaching this

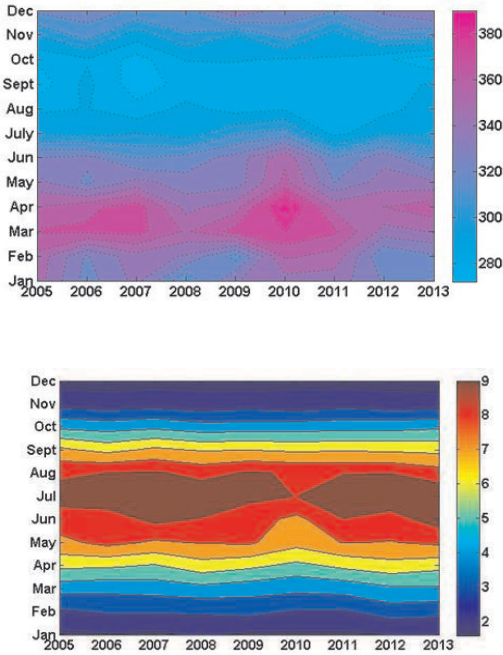


Figure 1. Annual and interannual variation of TCO and UV index.

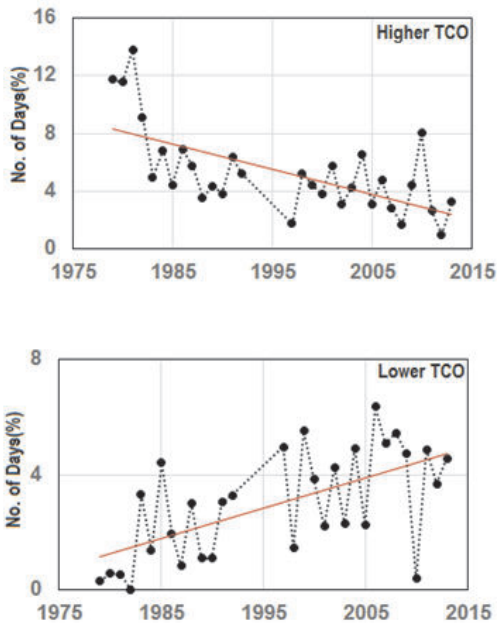


Figure 2. Variation of high and low ozone days.

maximum, TOC steadily decreases toward a minimum (<280 DU) that is achieved during September.

Stratospheric ozone is maximum in March and minimum in October, whereas UV index is maximum in June and July and minimum in December. Considering the light skin color of maximum population of Korea, UV index shows that the index that is higher than 7 can put them in high risk. Sudden low ozone events can expose the population to a high UV level, which they will not aware of. Number of days with high ozone and low ozone from 1979 to 2013 are shown in Figure 2.

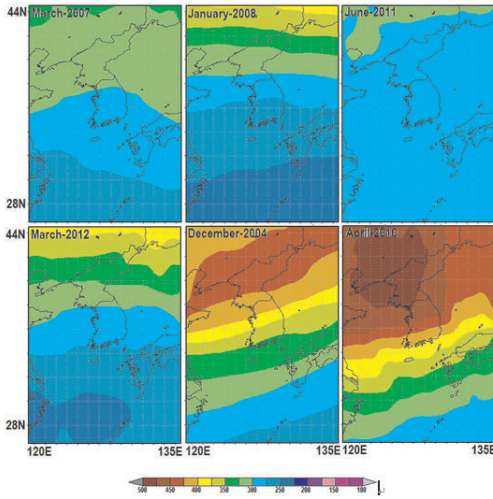
It is clear that not only average TCO but also high ozone events are decreasing with time. However, what is alarming is the sharp increase in the number of low ozone events. Except 2010, in recent years, the number of low ozone events is much higher than that observed in last two decades. Some recent observations of TCO rich and poor days for consecutively three days or more are summarized in Table 1 and also shown in Figure 3.

Table 1. Recent episode of ozone-rich and poor events.

| Ozone-rich episode | |
|----------------------|-----------------|
| Date | Average anomaly |
| December 29–31, 2004 | 23.02% |
| April 27–30, 2010 | 29.81% |
| Ozone poor episode | |
| Date | Average anomaly |
| March 1–4, 2007 | -17.26% |
| January 4–7, 2008 | -17.48% |
| June 23–29, 2011 | -16.87% |
| March 1–4, 2012 | -18.60% |

What is interesting is that both the number of ozone poor and rich days are maximum in February (Figure 4). Although other previous studies suggest that this kind of extreme events frequency is higher in winter months than in summer months, in Korea, it is observed that frequency is more in spring (February to May) than that in winter (November to January). As the UV index during April and May is moderately high, further decrease in stratospheric ozone can be very detrimental for human population. As the fact that this occasional low and high index in TCO is mainly guided by atmospheric circulation and turbulence, stratospheric dynamics during these days may be unstable, which leads to the decrease in TCO amount. Therefore, one ozone with high events in April, 2010 and low events in March, 2012 is chosen for further study.

Geographical distribution of TCO on April, 2010 events clearly show an ozone maximum just over Korean peninsula, whereas in case of March, 2012,



TCO in DU

Figure 3. TCO variations on recent ozone poor and rich events.

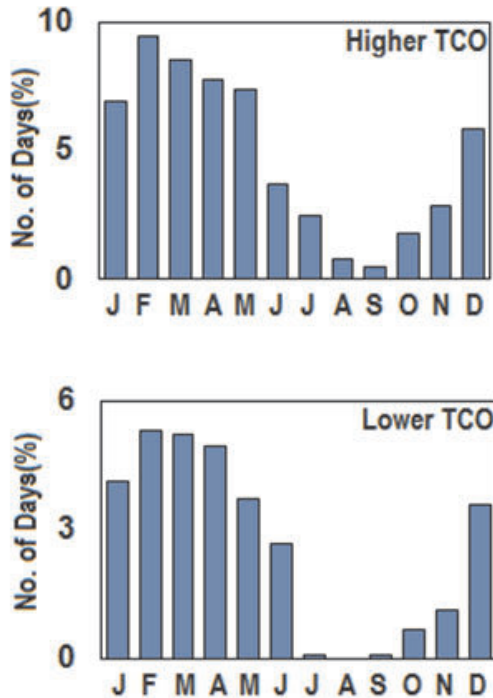


Figure 4. Monthly frequencies of high and low TCO days.

there is an ozone minimum over China south east to the peninsula. The variation meteorological condition in the two episodes is depicted in Figures 5 and 6.

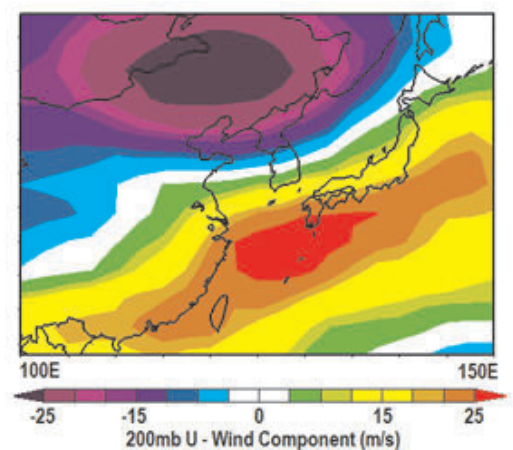
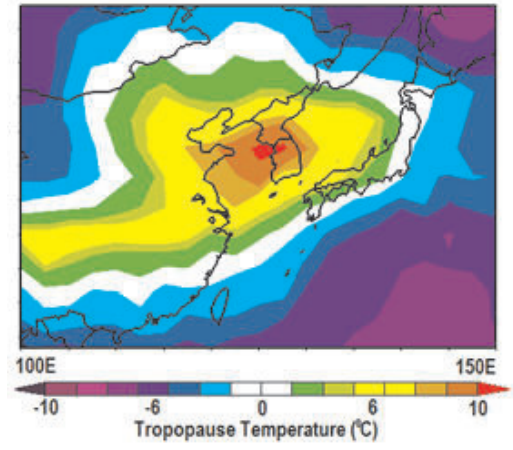
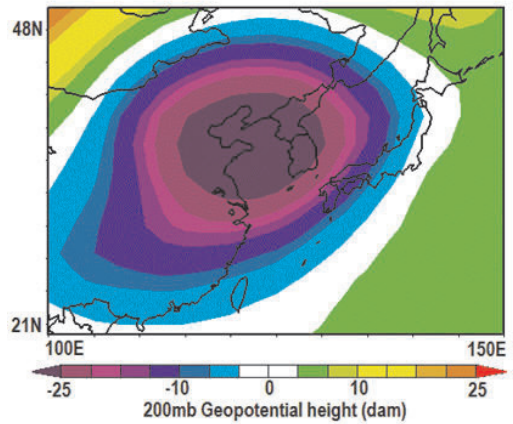


Figure 5. Climatological conditions during 27–30 April, 2010.

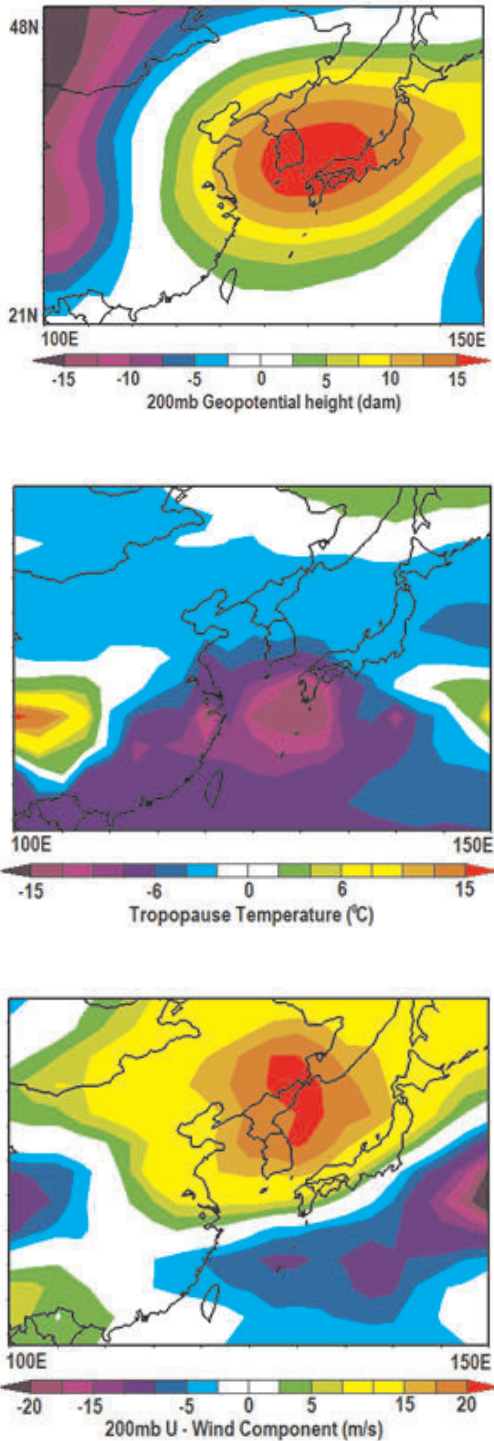


Figure 6. Climatological conditions during 1–4 March, 2012.

Concerning the spatial and temporal TCO variability, the analysis of the long-term TCO change over Korea between 2004 and 2013 indicates a statistically significant TCO decline for the months between May and October, with a maximum linear trend value in April.

4 CONCLUSIONS

The analysis of the spatial and temporal TCO variability over Korea using TOMS data provides several relevant conclusions. Reason behind the sudden low and high events of TCO is attributed mainly to the dynamic characteristics in UTLS and stratosphere. The meteorological conditions during the chosen ozone low and high days are totally contradictory. The low geopotential height during April 27–30, 2010 indicates the presence of stratospheric trough, which can build up a higher amount of ozone in nearby areas. The situation is completely reverse during low ozone events. The high geopotential height indicates a stratospheric ridge, which pushes the ozone poor air from southern region and creates low TCO observed over Korean peninsula. The scenario of tropopause temperature is also reverse in the two different episodes. It is higher during high ozone days and low during low ozone days. It is also observed that there is a jet stream in lower stratosphere south to Korea during April, 2010, and in contrast, its location during March, 2012 events is far north-west from Korean peninsula. Geopotential height at 50 mb clearly shows the position of polar vortex during these two events. It seems to be shifted toward Asian and American continent during March 2012, which can influence the ozone minimum over South Korea. In contrast, during ozone-high episode, it shifts toward Europe, which may have small or no influence on TCO over Korea.

REFERENCES

- [1] Bojkov, R.D., Bishop, L and Floletov, V.E., 1995, Total ozone trends from quality controlled ground-based data (1964-1994), *J. of Geophys. Res.*, 100, 25867–25876.
- [2] Chen, D., and Nunez, M. 1998, Temporal and spatial variability of total ozone in southwest Sweden revealed by two ground-based instruments, *Int. J. of Climatology*, 18, 1237–1246.
- [3] Den Outer, P. N., Slaper, H., and Tax, R. B., 2005, UV radiation in the Netherlands Assessing long-term variability and trends in relation to ozone and clouds. *J. of Geophys. Res.*, 110, D02203.
- [4] Krzyscin, J.W., 2002, long-term changes in ozone minimum event frequency over the northern hemisphere derived from ground based measurements, *Int. J. of Climatology*, 22, 1425–1439.
- [5] Salby, M.L. and Callaghan, P.F., 1993, Fluctuations of total ozone and their relationship to stratospheric air motions, *J. of Geophys. Res.*, 98, 2715–2727.

Application of artificial neural network ensembles for city ecology forecasting using air chemical composition information

D.I. Khritonenko & E.S. Semenkin

Siberian State Aerospace University, Krasnoyarsk, Russia

ABSTRACT: The constant growth of the population, the growth of the number of cars, the huge amount of harmful products, technogenic factors—these are some of the reasons for the bad ecological situation in modern megalopolis. The high level of harmful substances in the air makes the average lifetime shorter in cities and as a consequence damages the economy and city development. In this article, solving the problem of predicting the ecological condition of a city based on the air chemical composition is considered. The procedure of solving the problem with artificial neural networks, designed automatically with an evolutionary algorithm, is described. A set of modifications, which allow an increase in the accuracy of the solution of the data analysis problem, are presented and tested. A comparison with analogues is fulfilled. The analysis of the presented results shows the efficiency of the proposed approach when solving test problems. The application of this algorithm for predicting the ecological state of the city shows competitive efficiency compared with other algorithms.

1 INTRODUCTION

The modern ecological condition of cities leads to the necessity of finding the risk factors and developing a monitoring and control system for solving the technogenic problems. However, today the methods for estimating the risk level for urban territories, including geo-ecological risks, are still not developed properly. The lack of generalizing representations, concepts, and general methodological principles is one of the main disadvantages in the field of ecological risk estimation.

As most industrial and urban territories are located in basins, there may be increasing local circulation and aggregation of impurities in lower parts of the relief (Taseiko et al. 2011). Adverse orographic and meteorological features in connection with the atmosphere air quality lead to a higher air pollution level in the following cases: industrial sites are close to residential areas; the number of cars is increasing; houses are continuously built and green areas are destroyed. The high prevalence of diseases related to ecology among the populations of large industrial centres and a trend of an increasing pathology level determine the special importance of studying the health of city populations.

Concerning this, there appears a real necessity of including software for analyzing information about the environmental conditions in the ecological monitoring system. This necessity is due to the multidimensionality and multicoupling of ecological data. Many ecological processes are characterized

by nonlinearity and uncertainty that makes estimation and forecasting the ecological situation more complex. To increase the efficiency and accuracy of the decision-making process, artificial intelligence technologies capable of working under conditions of uncertain input information are used nowadays. One of the perspective directions of the use of computational intelligence technologies in ecological monitoring is the application of artificial neural networks in problems of predicting and forecasting ecological situations (Potilitsina et al. 2013).

2 THE PROBLEM DESCRIPTION

There is a sample (x_i, y_i) , $i = 1, 2, \dots, n$, where n is the number of instances in the sample, x_i is a vector of size m , where m is the number of sample attributes, and y_i is an observed value. It is also known that y_i may not only depend on its previous k values but also on values of x_i . The dependence between $x_i, x_{i-1}, \dots, x_{i-k}, y_i, y_{i-1}, \dots, y_{i-k}$ and y_{i+1} should be defined, which means predicting the values of y_{i+1} or solving a time series forecasting problem (Semenkin & Shabalov 2013).

The sample is a matrix with 39 columns, 28 of which are problem attributes, and 10 are the values to be predicted and 32 rows, where 32 is the number of instances. One of the columns is the year when the attributes are measured. There are also missing values in the database. The attributes show the constituents of different harmful materials in the air: copper, zinc,

iron, chlorine, and so on. All other measures show different diseases and the mortality because of these.

The data available in the sample is obtained in the environment pollution monitoring centre of GU “Krasnoyarsk CGSM-R,” which monitors the atmosphere air quality at eight stationary posts in the city of Krasnoyarsk, the Russian Federation. The yearbooks of the “Condition of the air pollution of the cities on the territory of the Krasnoyarsk region, and the republics of Khakassia and Tuva” from 1999 to 2012 are processed. Sanitary-demographic statistics is obtained in the territorial authority of the Federal Service of State Statistics in the Krasnoyarsk region (Krasnoyarskstat).

3 ARTIFICIAL NEURAL NETWORK ENSEMBLES

3.1 *Automatic design of artificial neural networks*

An artificial neural network (ANN) is a combination of mathematical models of neurons of the human brain, which allows the most complex data analysis problems to be efficiently solved (Kriesel 2005). However, when using this approach, the researcher always faces with the problem of choosing the optimal structure of the neural network and the optimal learning procedure according to the specified efficiency criterion. Usually a “trial and error” method is not a successful choice in this case, as it takes a huge amount of computational resources. The deep knowledge and experience of the expert in the field of neural networks are needed; nevertheless, the “trial and error approach” might result in insufficient modeling accuracy. Today, for solving the given problem, methods imitating the evolutionary process of the real world—evolutionary algorithms—are used. They allow the researcher to design the artificial neural networks automatically and with high efficiency (Shabalov et al. 2012). Besides, today there are a set of modifications of evolutionary algorithms that allow an increase in their efficiency. These algorithms do not require any parameter tuning, so they are easy to use for many researchers even if they do not have specific knowledge in the field of evolutionary modeling (Semenkin & Semenkina 2012-1).

We will now consider the main features of generating the neural networks with evolutionary algorithms. Usually the artificial neural network consists of the input layer, the output layer, and several hidden layers. The number of neurons of the input layer is usually equal to the number of inputs of the problem, and the number of neurons of the output layer is equal to the number of outputs of the problem to be solved. There can be several hidden layers, and they may contain different numbers of neurons. The following modifications are useful for such a network: feature selection, adjusting the links between the neurons in

hidden layers, choosing the number of neurons in the network, and choosing the types of activation functions for each of the neurons. We will now describe the genetic algorithm, which implements these modifications. The genetic algorithm operates with binary strings of a finite length. These strings will contain three parts: the first one is the vector $\overline{F} = (f_1, f_2, \dots, f_m)$, where m is the overall number of attributes of the problem. The “one” at the i th position of this vector means that the feature is used in the current ANN. In the other case, there is “zero” at this position. This part is responsible for choosing informative features. Connections between neurons of all layers are encoded using adjacency matrixes. As the recurrent ANN will not be used in this work, the matrix can be decreased to a triangular shape. This allows a decrease in the search complexity for the genetic algorithm. This kind of matrix will be added to the genotype string by string. The last block encodes the activation function types for every neuron. For every neuron, three bits are allocated, which allows the encoding of eight different types of activation functions. As a result, the ANNs are encoded as a vector of length $m + 0.5 \cdot n(n - 1) + 3 \cdot n$, where n is the maximal possible number of neurons in the network. The implementation of these modifications is also made for a genetic programming algorithm (Semenkin & Semenkina 2012-2). The main difference between these algorithms is the manner of data representation.

3.2 *Designing ANN ensembles*

The standard scheme for solving data analysis problems (Popov et al. 2012) with intelligence information technologies (IIT), in particular, neural networks, involves two stages. On the first stage, the structure and parameters of the neural networks are adjusted, and in the second, the best of the neural networks is used for solving the problem. However, using several neural networks simultaneously by combining them into an ensemble can be useful (Semenkin & Semenkina 2013). Increasing the quality of the obtained solutions with such an approach was observed long ago. The combining of neural networks into ensembles was first reported in the works of Hansen and Salamon (Hansen & Salamon 1990), where the efficiency of such approaches was also shown.

Today using ensembles is one of the promising directions in automated learning methods (Zhou & Jiang 2004). To design an ensemble, three tasks should be performed:

- 1 Choosing the method of designing individual ANNs;
- 2 Selection of ANNs for forming the ensemble from the set of ANNs obtained;
- 3 Determining the method for an utilisation of the solution given by all members of the ensemble.

In 1988, the problem of weak learning capabilities was studied. The problem was solved by strengthening the weak algorithms with boosting (Schapire 1990). Its essence is that if this algorithm is in a specific method for forming the learning sample for every next neural network, the learning sample includes incorrectly solved examples from the previous one. Therefore, the examples are processed by a consecutive set of ANNs. In 1993, an efficient learning algorithm, based on this principle, was developed (Freund & Schapire 1995). In 1996, Breiman proposed a bagging method, which learned every IIT with a sample, randomly (with returns) formed from the original one. This method is considered to be the most effective for samples of small size, as the lack of the part of learning examples in these samples leads to significant diversity in the received IIT. With a large number of features and a small number of learning examples, the method of random subspaces should be used (Skurichina & Duin 2002). In such a method, separate IITs are trained on different sets of attributes, which are selected randomly. The listed methods allow the design of different IITs (in particular, ANNs) for forming ensembles.

The question of how to consider results of different ensemble members is still to be considered. The basic work in this field describes the voting method (Hansen & Salamon 1990) and also the averaging method (Scheme 1). There is a modification of this method that allows errors of every ensemble member to be taken into account. We will now describe two such methods (Scheme 2):

- 1 The outputs of the ensemble members are summed and divided by their number.
- 2 The error of each ensemble member is taken into account. Each of them is associated with a weight, inversely proportional to the error (Perrone & Cooper 1993). The output values are summed and divided by their number.

There is a scheme (Scheme 3), allowing an error to be obtained, which is not higher than the best of the ensemble members (Popov et al. 2012):

- 1 The point in the learning sample is found, which is the closest to the point where we need to get the value;
- 2 The IIT, which has the lowest error, is determined at this point;
- 3 The selected IIT makes the decision about the value at the point of interest.

The forming of the ANN ensemble is also possible by means of the self-configured genetic programming algorithm (Semenkin & Semenkina 2013). For this purpose, the functional set is filled with different mathematical operators and functions (sine, cosine, addition, subtraction, and so on), and the terminal set is filled with different adjustable coefficients

and inputs of the neural networks selected for the ensemble. Using such an instrument does not require additional costs for its adjustment and allows the ensemble to be formed in automatic mode.

Thus, applying evolutionary algorithms may allow the design of efficient ensembles of artificial neural networks in automated mode without using special knowledge of neuro-evolutionary modeling. The efficiency of the program realization of the presented algorithms for ecological condition forecasting using the chemical air composition will be presented in the next section of the article.

4 EXPEREMENTAL RESULTS

The experiments were held using a special library with open source code «Rminer» (Cortez 2014), as well as the approach, which describes that it was implemented using C++. For testing of the algorithm, a set of experimental data were taken from the UCI machine learning repository (Bache & Lichman 2013).

For testing and comparing the algorithms, the following efficiency criteria were used:

$$\text{MAD} = \frac{1}{n} \cdot \sum_{i=1}^n |y_i - y_i^*| \quad (1)$$

$$\text{MRD} = \frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{y_i - y_i^*}{y_i^*} \right| \cdot 100\% \quad (2)$$

where y_i is the obtained value at the i th point; y_i^* is the true value; n is the sample size, $y_i^* \neq 0 \forall i$. In the case of solving classification problems, the accuracy measure is used, which shows the average percentage of correctly classified objects:

$$\text{Accuracy} = \left(1 - \frac{\text{ERROR}}{n} \right) 100\% \quad (3)$$

In this equation, ERROR is the number of incorrectly classified objects and n is the overall number of objects in the sample.

The efficiency of the developed artificial neural network ensemble forming methods was checked on the following problems: iris classification, wine classification, bank scoring (Australian and German variants), and forest fires area prediction (Cortez & Morais 2007). Each criteria value was calculated using the cross-validation method.

Choosing the efficient scheme of organizing the ensemble is one of the main problems of the approach.

From one point of view, any of the schemes existing nowadays can be chosen, and in most cases, it will be better than any single neural network. In contrast, the genetic programming algorithm allows it to be done in automatic mode, by sacrificing the computational resources. In the case when the accuracy of the resulting model is important, this approach is preferable. We will now show its (ENS_GP) efficiency (proportion of correctly classified items) compared with one of the existing methods of ensemble organizing on some test problems.

Table 1. Testing results.

| Algorithm | Irises | Wines | Australian Credit | German Credit |
|-----------|--------|-------|-------------------|---------------|
| Scheme 1 | 94.9 | 95.0 | 89.3 | 78.5 |
| Scheme 2 | 95.6 | 95.7 | 89.3 | 78.9 |
| Scheme 3 | 97.7 | 97.0 | 90.0 | 80.0 |
| ENS_GP | 97.7 | 97.3 | 90.1 | 80.0 |

The efficiency criterion (3), presented in Table 1, shows the advantages of generating ensembles with genetic programming. This means that we can hope to receive a more accurate model for the problem of predicting the ecological state of the city.

The testing of the proposed approach was conducted on a problem of predicting the area of forest fires. The comparison with some data analysis algorithms implemented in «Rminer» is presented in Table 2. As the efficiency criterion, the mean absolute deviation (MAD) is used.

Table 2. Comparison with analogues.

| Algorithm | MAD | Algorithm | MAD |
|-----------|-------|-----------|-------|
| MR | 13.00 | NN | 13.08 |
| DT | 13.81 | SVM | 12.71 |
| RF | 12.93 | ENS_GP | 12.53 |

In this table, MR is a multiple regression model; DT is a method of decision trees; RF is a method of random forest, NN is a neural network; and SVM is a support vector machine. As we can see from the presented data, the proposed approach (ENS_GP) has better prediction accuracy than existing analogues.

Now, we return to solving the problem of ecological forecasting condition. In this problem, the criterion of interest for the researcher is the mean relative deviation (MRD).

The database for the problem has missing values, as it has already been mentioned. Given the small size of the sample, and also the fact that the missing values are presented in a significant part of the sample,

the recovery of missing values is necessary. One of the promising approaches for the described conditions is the use of evolutionary algorithms (Snituk 2006), and in particular, the genetic programming algorithm for building up a nonlinear regression. For this problem, the terminal and the functional sets of the genetic programming algorithm are filled with different mathematical operators and functions and also some numerical values and feature variables of the problem. Here, we provide the comparison of the prediction results obtained with the perceptron automatically designed with a genetic programming algorithm for the cases with and without the recovery of missing values. In the case without the use of the recovery procedure, the incomplete data were excluded, as well as the problem attributes. All the predicted values were numbered (Out1–Out10). The testing results are presented in Table 3.

Table 3. Missing values recovery algorithm efficiency.

| | Out1 | Out2 | Out3 | Out4 | Out5 |
|------------------------|------|------|------|------|-------|
| With missing values | 9.8 | 12.7 | 6.1 | 4.2 | 6.3 |
| Without missing values | 9.2 | 12.5 | 5.4 | 2.4 | 5.4 |
| | Out6 | Out7 | Out8 | Out9 | Out10 |
| With missing values | 1.5 | 0.5 | 5.3 | 4.7 | 10.5 |
| Without missing values | 1.0 | 0.2 | 4.8 | 3.8 | 8.5 |

As we can see from these results, the missing values recovery procedure allows information about predicted values to be saved as well as the increased efficiency of the obtained predictors.

The results of solving the problem with the ensemble approach and with the basic model are presented in Table 4.

As we can see, the proposed approach has the minimum error for most of the presented predicted values, and so, this modification can be used for the similar problems. A comparison with each of the analogues, presented in Table 2, also shows the advantages of the suggested approach. Using ensembles requires significantly computational resources. Applying this approach cannot be justified if the problem can be solved with required accuracy by alternative techniques. In the case of error minimization, this approach has shown itself to be an efficient method for designing an ensemble of different data analysis methods (Popov et al. 2012).

Table 4. Comparison with basic model.

| | Out1 | Out2 | Out3 | Out4 | Out5 |
|-------------|------|------|------|------|-------|
| Basic model | 9.2 | 12.5 | 5.4 | 2.4 | 5.4 |
| ENS_GP | 5.1 | 9.3 | 5.4 | 1.9 | 3.9 |
| | Out6 | Out7 | Out8 | Out9 | Out10 |
| Basic model | 1.0 | 0.2 | 4.8 | 3.8 | 8.5 |
| ENS_GP | 1.0 | 0.2 | 2.8 | 3.6 | 7.6 |

Further steps to solve this problem include using recurrent neural networks, as they have higher flexibility for problems of predicting time series. Also, the need for further extension of the data sample should be mentioned. This procedure will allow an increase in the accuracy of existing models and the generation of new ones for predicting the ecological state of the city.

5 CONCLUSIONS

In this article, an evolutionary approach for the automated forming of artificial neural networks and their ensembles was described. The standard methods of forming ensembles were described. The best in the means of classification and prediction error was obtained by genetic programming. The universality of this method is one of the reasons why it is widely used in data analysis problems. The problems of restoring missing values, generating neural networks, and forming ensembles are efficiently solved by this method in this article. The comparison of the proposed approach with other data analysis methods has shown its efficiency for the problem of predicting the area of forest fires. The practical problem of predicting the ecological state of the city using information on the air chemical composition was successfully solved. As a result of testing, the proposed modification has shown better performance than the basic algorithm for most of the predicted values. For a set of predicted values, the efficiency of the approach did not change. This may result from the small volume of the sample and also the absence of key variables of the problem. The further expansion of the sample will allow this problem to be solved more precisely. Recovering the missed values with the genetic programming algorithm also allows an

increase in the efficiency of the approach. Generating recurrent neural networks and their ensembles is one of the possible ways for further developing the approach. However, it should be remembered that using a universal procedure for the evolutionary design of neural network predictors significantly increases the computational complexity, which can be unacceptable in some cases.

ACKNOWLEDGMENT

This work is supported by the Russian Foundation for Basic Research (Project No: 14-06-00256).

REFERENCES

- [1] Bache, K., Lichman, M. 2013. UCI Machine Learning Repository [URL]: <http://archive.ics.uci.edu/ml>
- [2] Breiman, L. 1996. Bagging predictors. *Machine Learning*. 24(2): 123–140.
- [3] Cortez P. 2014. RMiner: Data Mining with Neural Networks and Support Vector Machines using R. In R. Rajesh (Ed.), *Introduction to Advanced Scientific Software and Toolboxes*, In Press.
- [4] Cortez P., Morais A. 2007. A Data Mining Approach to Predict Forest Fires using Meteorological Data. *Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimarães, Portugal*, pp. 512–523.
- [5] Freund Y., Schapire R. E. 1995. A decision-theoretic generalization of on-line learning and an application to boosting. *European Conference on Computational Learning Theory*, pp. 23–37.
- [6] Hansen L.K., Salamon P. 1990. Neural network ensembles. *IEEE Trans. Pattern Analysis and Machine Intelligence*. 12(10): 993–1001.
- [7] Kriesel D. A. 2005. Brief Introduction to Neural Networks. [URL]: http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf
- [8] Perrone M.P., Cooper L.N. 1993. When networks disagree: ensemble method for neural networks. *Artificial Neural Networks for Speech and Vision, Chapman & Hall*. New York, pp.126–142.
- [9] Popov E. A., Semenkina M. E., Lipinskiy L. V. 2012. Decision making by an ensemble of intelligent information technologies. *VestnikSibGAU*. 5(45): 95–99. (in Russ.)
- [10] Potilitsina E. N., Lipinskiy L. V., Sugak E.V. 2013. Using artificial neural networks for solving applied ecological problems. *Modern problems of science and education* 4. (in Russ)
- [11] Schapire R. E. 1990. The strength of weak learnability. *Machine Learning*. 5: 197–227.
- [12] Semenkin E.S., Semenkina M.E. 2012. Self-configuring Genetic Algorithm with Modified Uniform Crossover Operator // *Advances in Swarm Intelligence*. Lecture Notes in Computer Science 7331. – Springer-Verlag, Berlin Heidelberg, 2012. – P. 414–421.

- [13] Semenkin E., Semenkina M. 2012. Self-Configuring Genetic Programming Algorithm with Modified Uniform Crossover. In: Proceedings of the Congress on Evolutionary Computations of the IEEE World Congress on Computational Intelligence (CEC WCCI 2012), Brisbane, Australia, 2012. – P. 1918–1923.
- [14] Semenkin E., Semenkina M. 2013. Classifier Ensembles Integration with Self-Configuring Genetic Programming Algorithm // Adaptive and Natural Computing Algorithms. Lecture Notes in Computer Science, Volume 7824. – Springer-Verlag, Berlin, Heidelberg, 2013. - Pp. 60–69.
- [15] Semenkin E.S., Shabalov A.A. 2013. Intelligent information technologies in time series forecasting. *VestnikSibGAU*4(50): 128-134. (in Russ.)
- [16] Shabalov, A., Semenkin, E., Galushin, P. 2012. Integration of Intelligent Information Technologies Ensembles for Modeling and Classification // Hybrid Artificial Intelligence Systems. Lecture Notes in Computer Science, Volume 7208/2012, pp. 365–374.
- [17] Skurichina M., Duin R. P. W. 2002. Limited bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis & Applications*. 5: 121–135.
- [18] Snituk V.E. 2006. Evolutionary method for missing data recovering. *Proceedings of the VI-th International conference "Intelligent data analysis"*. Kiev, pp. 262–271. (in Russ.)
- [19] Taseiko O. V., Mikhailuta S. V., Zakharov U. V. 2011. *Modeling local conditions of pollution dissipation among city buildings*. Saarbrücken: LAP LAMBERT Academic Publishing GmbH & Co. KG. (in Russ.)
- [20] Zhou Z.-H., Jiang Y. 2004. NeC4.5: Neural Ensemble Based C4.5. *IEEE Trans. Knowl. Data Eng.* 16(6): 770–773.

An improved method for calculating atmospheric refraction coefficient for eliminating influence of meteorological factors

L. Liu, C.L. Mi & J.S. Han

Shandong Linyi Bureau of Land Resources, China

ABSTRACT: Atmospheric refraction coefficient is an important factor to affect the accuracy of trigonometric leveling. This article summarizes major methods for calculating atmospheric refraction coefficient and their advantages and disadvantages. Based on the actual requirements of data processing, this article presents an improved method for calculating atmospheric refraction coefficient that can eliminate the influence of meteorological factors including temperature, atmospheric pressure, and so on. With this method, the accuracy of trigonometric leveling under different meteorological environments can be effectively improved.

1 INFLUENCE OF ATMOSPHERIC REFRACTION ON TRIGONOMETRIC LEVELING

On the surface or under the shaft, influence brought by atmospheric refraction cannot be avoided in electronic angle and distance measurement, so atmospheric refraction is an important factor to affect the accuracy of trigonometric leveling. When the angle between pitch angle and horizontal direction is 1°, 2°, 3°, 5°, 7°, 10°, 13°, 15°, 17°, 20°, 23°, 25°, 27°, and 30°, substituting data in the effect formula of atmospheric refraction coefficient yields the following error influence quality [1]. Figure 1 shows that when the observed slope distance exceeds 1 m, the effect on quality of atmospheric refraction is more than 1 mm. Therefore, during the precise leveling with the use of trigonometric leveling, the influence of atmospheric refraction must be taken into consideration.

2 COMPARISON OF ATMOSPHERE-REFRACTION CORRECTION METHODS

2.1 Practical measurement method

The practical measurement method^[2] is to measure a certain amount of atmospheric elements according to the knowledge of atmospheric physics and then to figure out the temperature gradient of pertinent measuring point and next to integrate in terms of equation and finally to get the atmospheric refraction coefficient in the direction of survey line. For example, electing an instrument at point “a” to observe the target at point b, the equation for calculating the atmospheric refraction coefficient (K_{ab}) of survey line (ab) is that:

$$K_{ab} = \frac{17.21P}{(T+273.16)^2} + \frac{1006.6P}{S^2(T+273.16)^2} \int_0^S \frac{dT}{dh}(S-x_i) dx \quad (1)$$

In equation (1), P (hPa) is the average atmospheric pressure between “a” and “b,” and T (°C) is the temperature of horizontal distance, and $\frac{dT}{dh}$ (°C/m) represents the temperature gradient at one point.

S represents the horizontal distance between “a” and “b,” and x_i is the horizontal distance from point “i” to observation station a in the direction of survey line.

2.2 Inversion method

Inversion method means to calculate in reverse the atmospheric refraction coefficient during the process of observation through the height difference observed

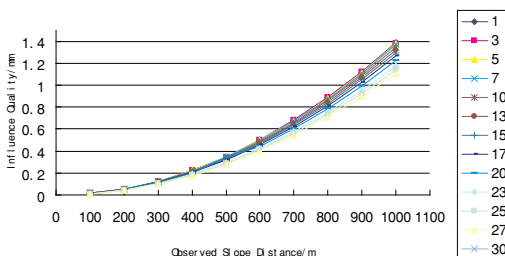


Figure 1. The influence quality of atmospheric refraction with different angles and distance on trigonometric leveling.

by electromagnetic distance measurement and trigonometric leveling, including height-difference inversion method with known precision, reciprocal observation height-difference inversion method, and so on. An inversion process of atmospheric refraction will be explained with examples of height-difference inversion methods of direct and reversed observation.

The instrument is erected at point “a” and the target of observation is at point “b.” h_{ab} and h_{ba} represent the values of the height differences of direct and reversed observation, respectively. S is the horizontal distance between point “a” and “b.” K_{ab} and K_{ba} represent separately the atmospheric refraction coefficients in directions of survey lines of direct and reversed observation. R is earth average curvature radius between point “a” and “b.” Ignoring vertical deflection and other observation errors produces the equation that

$$h_{ab} + h_{ba} + \frac{S^2}{R} - \frac{K_{ab} + K_{ba}}{2R} S^2 = 0 \quad (2)$$

Set $(K_{ab} + K_{ba}) / 2 = K$ and substituting it in equation (2) resulting in the equation as below:

$$K = 1 + \frac{h_{ab} + h_{ba}}{S^2} R \quad (3)$$

2.3 Simulation method

Simulation method^[5] is to measure atmospheric factors of some special points including temperature, atmospheric pressure, and so on in the direction of survey line and then calculate the refraction coefficient of these special points in terms of a equation. The atmospheric refraction coefficient of the whole survey line can be simulated by data of these special points with the use of the weight theory. The atmospheric refraction coefficient can be calculated by the following equation that

$$K = 2S_0^{-2} \int_0^{S_0} P_i k_i (S_0 - x) dx \quad (4)$$

In equation (4), S_0 is the total length of observing line; P_i is the effect weight on the whole survey line of refraction coefficient k_i at point “i” and it can be set that $P_i = \frac{S_0 - x_i}{S_0}$ (x_i is the horizontal distance between point “i” and observation station). The refraction coefficient k_i of atmospheric factors at special point “i” can be calculated by the following equation that $k_i = \frac{81.96RP}{T^2 \times 10^6} \left(0.0342 + \frac{dT_i}{dh_i} \right) \cos \alpha$. S_0 is the total length of optical path; P_i is the effect weight of k_i in K ,

and $P_i = \frac{S_0 - x_i}{S_0}$ (x_i is the horizontal distance between point “i” and observation station).

Practical measurement method is to figure out refraction coefficient through atmospheric factors on actual survey line. This method is rigorous in theory, small in errors, and credible in determination results. However, this method is seldom used, because in trigonometric leveling, it is necessary to measure temperature difference above and below the survey line with a result that the amount of work is huge and the measurement efficiency is reduced. Inversion method is to figure out the atmospheric refraction coefficient during the observation process with the height difference of two points known, so this method is suitable to be applied in experimental research instead of actual production^[3]. The accuracy of simulation method is determined mainly by the choice of weight. The refraction coefficient of points may be affected by many factors including terrain, landmark, temperature, wind, moisture, and so on, so it is hard to choose feature points.

As is known from the discussion above, current atmosphere-refraction correction method is accomplished by measuring atmospheric factors to figure out atmospheric refraction coefficient on survey line. However, atmospheric factors during observation process are hard to measure in real time, low in efficiency, and unknown in accuracy. This article discusses a new atmosphere-refraction correction approach with no need for measuring atmospheric factors.

3 A NEW ATMOSPHERE-REFRACTION CORRECTION APPROACH

Affected by constantly changing atmospheric factors, atmospheric refraction coefficients at different time and places are different. To reduce the influence of atmospheric factors on refraction coefficients, this article presents a new approach to improve the method for calculating atmospheric refraction coefficient.

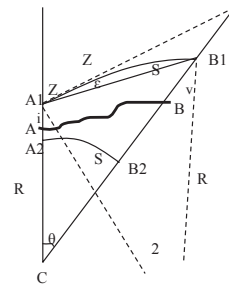


Figure 2. Influence of vertical refraction.

As is shown in **Figure 2**, erect the total station instrument at point A on ground, set the height of instrument as “ p ” and put a prism at point B with a height “ v ”. The visual arc caused by atmospheric refraction is A_1B_1 , corresponding zenith distance of A_1B_1 is Z in terms of the definition of zenith distance, and corresponding zenith distance of chord $\overline{A_1B_1}$ is Z' . For A_1B_1 , its curvature radius is set equal to R_0 and the arc length to S_0 , so the atmospheric refraction angle can be represented by the equation that

$$\varepsilon = \frac{S_0}{2R_0} \quad (5)$$

In **Figure 2**, if projections of points A and B on geoid surface are A_2 and B_2 , respectively, the length S of arc A_2B_2 on geoid surface is the distance from points A and B to geoid surface and the corresponding radius of arc A_2B_2 is R .

In $\triangle A_1B_1C$, in terms of Tangent law, the following equation can be obtained that

$$\frac{B_1C - A_1C}{B_1C + A_1C} = \frac{\tan \frac{A_1 - B_1}{2}}{\tan \frac{A_1 + B_1}{2}} \quad (6)$$

Next, the analysis of angles and distance involved in **equation (6)** will be made.

Figure 2 indicates that

$$B_1C - A_1C = Hb + v - Ha - i = h \quad (7)$$

$$B_1C + A_1C = R + Ha + i + R + Hb + v = 2R + Ha + i + Hb + v \quad (8)$$

$$\frac{A_1 + B_1}{2} = \frac{180^\circ - \theta}{2} = 90^\circ - \frac{\theta}{2} \quad (9)$$

If A_1 and B_1 are represented by zenith distance Z , refraction angle ε and angle θ , just as that

$$A_1 = 180^\circ - Z - \varepsilon, B_1 = Z + \varepsilon - \theta \quad (10)$$

$A_1 - B_1$ can be represented by the equation that

$$A_1 - B_1 = 180^\circ - Z - \varepsilon - (Z + \varepsilon - \theta) = 180^\circ - 2(Z + \varepsilon) + \theta \quad (11)$$

the angle can be represented by the equation that

$$\frac{A_1 - B_1}{2} = 90^\circ - (Z + \varepsilon) + \frac{\theta}{2} \quad (12)$$

Substituting **equation (12)** in Tangent law results in the equation that

$$h = B_1C - A_1C = \frac{\tan \frac{A_1 - B_1}{2}}{\tan \frac{A_1 + B_1}{2}} (B_1C + A_1C)$$

$$= \frac{\tan \frac{\theta}{2}}{\tan(Z + \varepsilon - \frac{\theta}{2})} (2R + Ha + Hb + i + v) \quad (13)$$

Then, simplify **equation (13)**:

The observation distance is always within 5 km with the use of total station instrument, so $\theta = \frac{S}{R} \leq \frac{S_0}{R} < \frac{A_1B_1}{R} = \frac{5}{6371} \text{ rad} = 2.7'$, which is so small

a value that $\tan \frac{\theta}{2} \approx \frac{\theta}{2} = \frac{S}{2R}$.

Because $\varepsilon - \theta$ is a small angle, the denominator $\frac{1}{\tan(Z + \varepsilon - \frac{\theta}{2})} = \cot(Z + \varepsilon - \frac{\theta}{2})$ can be considered as

a function of Z . Therefore, within the neighborhood $(Z - \varepsilon + \frac{\theta}{2}, Z + \varepsilon - \frac{\theta}{2})$, there is the $(n+1)$ th derivative of function Z , which can be in accordance with the application condition of Maclaurin equation, so expanding the former equation into series with the first term reserved produces that

$$\cot(Z + \varepsilon - \frac{\theta}{2}) = \cot Z + \cot' Z (\varepsilon - \frac{\theta}{2}) =$$

$$\cot Z - \frac{\varepsilon - \theta}{\csc^2 Z} = \cot Z - (\varepsilon - \frac{\theta}{2}) \sin^2 Z$$

According to the definition of atmospheric refraction coefficient [4], it can be known that

$$\text{Refraction angle } \varepsilon = K \frac{S_0}{2R} \quad (14)$$

Compared with $2R$, S_0 is small enough that $\frac{S_0}{R} \approx \frac{S}{R}$.

Zenith distance Z is represented by α and substituting equations that $\theta = \frac{S}{R}$ and $2\varepsilon = \frac{S_0}{R_0}$ in **equation (13)** gives that

$$\cot Z - (\varepsilon - \frac{\theta}{2}) \sin^2 Z = \tan \alpha - (K \frac{S}{2R} - \frac{S}{2R}) \cos^2 \alpha \quad (15)$$

$$\text{and } h = \left[\tan \alpha - \frac{S}{2R} (K-1) \cos^2 \alpha \right] (2R + Ha + Hb + i + v) \quad (16)$$

Supposing that $Hm = \frac{Ha + Hb + i + v}{2}$ and ignoring the term of $1/R^2$ yields that

$$h = (1 + \frac{Hm}{R}) S \tan \alpha + \frac{S^2(1-K)}{2R} \cos^2 \alpha \quad (17)$$

Actual observation value S' can be replaced by S through methods of inclination correction, the correction from arc to chord and the correction from chord to arc. To be simple, setting the condition that $S = S' \cos \alpha$ produces the height difference between points A and B that

$$Hb - Ha = \left(1 + \frac{Hm}{R}\right) S' \sin \alpha + \frac{S'^2(1-K)}{2R} \cos^4 \alpha + i - v \quad (18)$$

If special biprisms are used or the distance between two prisms keeps unchanged by adjusting the height of prisms or the changing height maintains consistent during the observation process, twice trigonometric leveling with different heights of prisms is carried out for each observation. So the following equations can be obtained that

$$Hb - Ha = \left(1 + \frac{Hm}{R}\right) S'_1 \sin \alpha_1 + \frac{S_1'^2(1-K)}{2R} \cos^4 \alpha_1 + i - v_1 \quad (19)$$

$$Hb - Ha = \left(1 + \frac{Hm}{R}\right) S'_2 \sin \alpha_2 + \frac{S_2'^2(1-K)}{2R} \cos^4 \alpha_2 + i - v_2 \quad (20)$$

Consolidating these two equations leads to the following equation that

$$K = 1 - \frac{2R(S'_2 \sin \alpha_2 - S'_1 \sin \alpha_1 + E)}{S_1'^2 \cos^4 \alpha_1 - S_2'^2 \cos^4 \alpha_2} - \frac{2Hm(S'_2 \sin \alpha_2 - S'_1 \sin \alpha_1)}{S_1'^2 \cos^4 \alpha_1 - S_2'^2 \cos^4 \alpha_2} \quad (21)$$

Compared with the second term, the last one is small enough to be ignored and the simplification of equation (21) is that

$$K = 1 - \frac{2R(S'_2 \sin \alpha_2 - S'_1 \sin \alpha_1 + E)}{S_1'^2 \cos^4 \alpha_1 - S_2'^2 \cos^4 \alpha_2} \quad (22)$$

E represents the changing height of prisms and $E = v_1 - v_2$.

4 CONCLUSION

On earth surface, atmospheric refraction coefficient may change quickly and constantly and affected by terrain and landmarks and atmospheric factors on survey line are not in uniform distribution, so the operation to simulate atmospheric factors at each point on survey line through feature points is complicated and easy to cause errors. There is no atmospheric factor in equation (22) and only conventional measured values of trigonometric leveling including altitude angles measured by two observations, and slope distance and the changing height of prisms are needed to produce the real-time atmospheric refraction coefficient, so the accuracy of atmospheric refraction coefficient improves.

REFERENCES

- [1] Han Junsheng, Study on key technologies of large holing through survey with an example of Tongxi mine (2011).
- [2] Chen Zhenjie, Xin Guoli, Zhang Shushou, Trial research on temperature model of measurement route in juxtaterrestrial atmosphere, J. Sci. Commun. 10(2001) 45-47.
- [3] Jiang Lilong, The inversion of atmospheric refraction for the dissymmetric topography, J. Sci. Commun. 12(2003) 3-5.
- [4] G. Bomford, Geodesy, second volume, Surveying and Mapping Press, Beijing, 1981.
- [5] Kong Xiangyuan, Guo Jiming, GEO-spatial information science, Wuhan University Press, 2006.
- [6] He Liheng, Bao Qisheng, Wei Haohan, Characteristics, application and progress on triangle height measurement of total station instrument, J. Sci. Commun. 3(2009) 22-26.

Cerium-induced genotoxicity in Arabidopsis at environmentally relevant doses

L.Z. Sun

Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China

P.D. Tai, W. Liu, X.J. Li & C.Y. Jia

Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China

Key Laboratory of Pollution Ecology and Environmental Engineering, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China

Z.H. Liu

Key Laboratory of Pollution Ecology and Environmental Engineering, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China

Warrnambool campus, Deakin University, Warrnambool, Vic 3280, Australia

X. Wang

Key Laboratory of Pollution Ecology and Environmental Engineering, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China

ABSTRACT: Random amplified polymorphic DNA (RAPD) test is a feasible method to evaluate the toxicity of environmental pollutants on vegetal organisms. Herein, Arabidopsis thaliana (Arabidopsis) plantlets following Cerium(Ce) treatment for 21 days were screened for DNA genetic alterations by DNA fingerprinting. Three primers amplified 12 mutated RAPD fragments in 0-100 $\mu\text{mol}\cdot\text{L}^{-1}$ Ce-treated Arabidopsis plantlets. Cloning and sequencing analysis of 6 representative mutated fragments revealed 99-100% homology with genes of MADS-box family protein, ABC transporter B family, RING-finger domain-containing protein, phosphoglucosmutase, and noncoding regions, respectively, in Arabidopsis genome. Therefore, these altered loci are the early events in Ce-treated Arabidopsis seedlings and would greatly improve environmental risk assessment.

1 INTRODUCTION

The rare earth elements (REEs) comprise a group of 15 trivalent metallic elements with similar chemical properties. The dispersion of Ce in the environment has drastically increased over the past decades through anthropogenic activities, such as mining deposits and aerial fallout from industrial processes. The REEs normally occur as basic oxides and phosphate complexes in nature (Das, et al., 1988; Kabata-Pendias & Pendias, 1992). Numerous studies have shown that genotoxicity of Ce is directly related to its effect on structure and function of DNA, which may be determined by using a number of laboratory methods (Hui, et al., 2004; Chengrun, et al., 2011). However, there have been a few direct experimental demonstrations of the wider relationships between DNA effects and their subsequent consequences at higher levels of biological organization in water and soil (Wolf, et al., 2004). Therefore, the principal objectives of

the present study were to (1) identify several markers which revealed genetic differences between the control and Ce-stressed Arabidopsis seedlings exposed to relevant environmental doses by RAPD-PCR; (2) determine the molecular nature of RAPD bands to search for homology with any previously described loci in the model plant Arabidopsis genome.

2 MATERIALS AND METHODS

2.1 *Plant materials, growth conditions, and determination of growth, chlorophyll and total soluble protein levels of Arabidopsis plantlets*

Arabidopsis seeds were sown hydroponically in sterile flasks kept on a rotary shaker at 50 rpm (300 mL flasks containing 200 mL of Murashige-Skoog medium with 0.5% Sucrose (pH 5.8)), at about 20 plants per flask containing a final concentration

of 0, 10 and 100 $\mu\text{mol}\cdot\text{L}^{-1}$ $\text{Ce}(\text{NO}_3)_3$, respectively, in three replicates.

After 21 days following germination, seed germination rate and growth of the plantlets were measured and then pooled to measure chlorophyll and soluble protein content in shoots (i.e. stems and leaves) described in the previous study.

2.2 DNA isolation, RAPD-PCR, and Cloning and sequencing of RAPD fragments

For both the control and Ce treatments, shoots were collected after 21 d following germination, and flash frozen in liquid nitrogen prior to storage at $-70\text{ }^\circ\text{C}$. Total DNA was extracted from approximately 200 mg of fresh shoots by using a CTAB protocol and RAPD-PCR was performed on DNA from control and Ce-exposed shoots. All analyses were performed with 3 independent replicates.

RAPD reactions were performed by using the following decamer oligonucleotides: P1(5-TCCGATGCTG-3), P2(5-CTGGGGCTGA-3) and P3(5-AAAGTGC GGC-3). RAPD bands of interest were excised from the agarose gel and then purified with a commercial DNA purification kit. The RAPD insert was sequenced by the DNA Sequencing Core Facility, Beijing Genomics Institute, Beijing, PR China. Sequences obtained from clones were compared with sequences in the GeneBank database using the BLASTn and BLASTx programs. Chemicals were ordered from TaKaRa Biotechnology Ltd. (Dalian, China)

2.3 Statistical analysis

The software SPSS for Windows (version 10.0) was used for the statistical evaluation of the results. Statistical significance was confirmed by analysis of variance (ANOVA) and by the least significant differences (LSD) tests corrected for the number of comparisons. The two-sided probability level of 0.05 was used as the criterion for significance in all procedures. All results were expressed as mean and standard deviation of the mean.

3 RESULTS

3.1 Effect of Ce on seedling growth, chlorophyll level and total soluble protein level

Table 1 showed that there were statistically significant differences between the control and Ce-treated plantlets after 21 d ($P < 0.05$). The results showed that the root length, fresh weight, chlorophyll and total soluble protein content first increased and then decreased, while no significant changes occurred in leaf number in Arabidopsis seedlings under Ce stress of 0-100 $\mu\text{mol}\cdot\text{L}^{-1}$.

3.2 RAPD loci with DNA damage and mutations induced by Ce stress

Using the 12 oligonucleotide primers (10 bp), genomic DNA of the shoots of Arabidopsis seedlings cells was amplified by PCR. We observed 12 mutated loci amplified only by primers 1, 2 and 3 from 0 - 100 $\mu\text{mol}\cdot\text{L}^{-1}$ Ce-treated plantlets as determined through comparison of Ce-treated RAPD fingerprints with the control (Figure 1). The sizes of these RAPD bands are indicated in Table 2. Also, we observed alterations in the fluorescence intensity of the RAPD fragments in the DNA of Ce-treated shoots as comparison to the control.

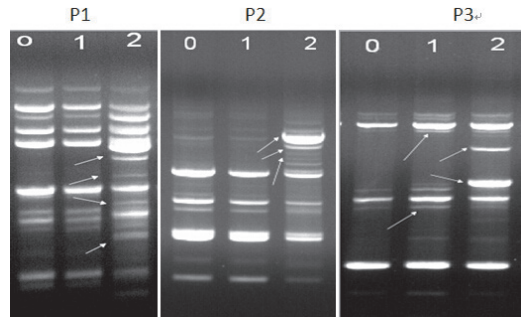


Figure 1. RAPD profiles of genomic DNA from shoots of Arabidopsis seedlings exposed to varying Ce concentration for 21 d. Lanes 0, 1, 2 denote seedlings exposed to a range of 0, 10 and 100 $\mu\text{mol}\cdot\text{L}^{-1}$ Ce, respectively.

Table 1. Effects of Ce stress on leaf number, fresh weight, root length, chlorophyll and total soluble protein content in Arabidopsis thaliana seedlings.

| Ce ($\mu\text{mol}\cdot\text{L}^{-1}$) | leaf number | fresh weight(mg) | root length (cm) | chlorophyll content ($\text{mg}\cdot\text{g}^{-1}$) | total soluble protein content($\text{mg}\cdot\text{g}^{-1}$) |
|--|-------------|------------------|------------------|---|--|
| 0 | 6 | 22.33±0.62 b | 7.21±0.53 b | 0.1924±0.02 b | 73.547±1.88 b |
| 10 | 6 | 25.26±0.72 c | 8.53±0.35 c | 0.2373±0.03 c | 84.262±2.28 c |
| 100 | 6 | 19.08±0.54 a | 5.24±0.32 a | 0.1710±0.02 a | 50.276±0.95 a |

3.3 Sequencing of RAPD fragments and sequence homology

The selected six representative RAPD bands were P1₁₃₄₅, P1₈₉₈, P2₁₄₄₅, P2₁₂₃₅, P3₁₀₈₇ and P3₇₈₀ in 0, 10 and 100 $\mu\text{mol}\cdot\text{L}^{-1}$ Ce-treated shoots, respectively. The sequences of the 6 bands were deposited in the GenBank, and the molecular lengths were fairly close as determined by DNA sequencing as compared to gel electrophoresis (Table 2). The results of BLAST analysis revealed that long stretches of the RAPD bands were homologous to the following known sequences of the Arabidopsis genome in the GenBank database at the DNA level:

- 1 The fragment of 898 bp (P1₈₉₈) only in the control had 99.89% homology with stretch of hypothetical protein from nt 17534419 to 17535316 on chromosome 1. Also, the 1345 bp band (P1₁₃₄₅) in the control and 10 $\mu\text{mol}\cdot\text{L}^{-1}$ Ce-treated shoots had 99.9% homology with hypothetical protein from nt 4354609 to 4355933 on chromosome 4. These 2 bands were from noncoding regions in Arabidopsis genome (Table 2).
- 2 The 1235 bp band (P2₁₂₃₅) showed 100% homology with MADS-box family protein gene from nt 13452076 to 13451331 on chromosome 2 in Arabidopsis genome. Also, the band of 1445 bp (P2₁₄₄₅) had a significant DNA base sequence

- similarity (99%) with phosphoglucomutase (PGM) gene from nt 8222503 to 8223283 on Arabidopsis chromosome 1.
- 3 The 1087 bp band (P3₁₀₈₇) in the control and 10 $\mu\text{mol}\cdot\text{L}^{-1}$ Ce-treated shoots had 100% homology with RING-finger domain-containing protein gene from nt 4298166 to 4299245 on chromosome 3. Likewise, gain of 780 bp band (P3₇₈₀) in the Ce-treated shoots showed 99.88% (824/825) homology with ABC transporter B family from nt 12563272 to 12564096 on chromosome 4 in Arabidopsis genome (Table 2). A search for the conserved domain using NCBI conserved domain database revealed that the ABC transporter B family had a high degree of functionally conserved domains among plants. In order to further characterize and identify the cause for the gain of the 780 bp fragment in the RAPD fingerprints of the Ce-treated shoots, the comparison of this sequence from the GenBank revealed 4 point mutations in the DNA of the Ce-treated shoots. These small mutations (i.e. CT for forward primer and AG for reverse primer) were the new priming sites for primer 3 used for RAPD-PCR amplification, thus generating amplification of the 780 bp fragment from the Ce-treated shoots, and its absence in the normal DNA was due to the absence of the priming sites in the corresponding sequence.

Table 2. Genebank accession numbers, molecular lengths by sequencing and homologous segments of six representative RAPD markers in shoots of Ce-contaminated Arabidopsis seedlings for 21 days.

| RAPD marker* | Accession No. | Molecular length | | Homology (%) | Homologous segments |
|--------------------|---------------|------------------|-------|--------------|---------------------------------------|
| | | A(bp) | B(bp) | | |
| P1 ₁₃₄₅ | NC_003075.7 | 1345 | 1340 | 99.9 | Noncoding region |
| P1 ₈₉₈ | NC_003070.9 | 898 | 880 | 99.89 | Noncoding region |
| P2 ₁₄₄₅ | NC_003070.9 | 1445 | 1440 | 99 | Phosphoglucomutase gene |
| P2 ₁₂₃₅ | NC_003071.7 | 1235 | 1230 | 100 | MADS-box family protein |
| P3 ₁₀₈₇ | NC_003074.8 | 1087 | 1050 | 100 | RING-finger domain-containing protein |
| P3 ₇₈₀ | NC_003075.7 | 780 | 778 | 99.88 | ABC transporter B family |

*P1₈₉₈ denotes primer 1 and molecular length of the 898 bp of RAPD band, respectively. The same below. A and B indicate molecular length by sequencing and electrophoresis, respectively.

4 CONCLUSIONS

The identification of 6 representative RAPD fragments showed 99-100% homology with genes of MADS-box family protein, ABC transporter B family, RING-finger domain-containing protein, phosphoglucomutase, and noncoding region in Arabidopsis genome under Ce stress by DNA fingerprinting, which plays important roles in maintaining genomic integrity by interaction with DNA base excision repair proteins and in rapid cellular activation responses to oxidative stress for ubiquitination in Ce contaminated habitats. The ability to detect significant genetic alterations in Arabidopsis seedlings at Ce levels of 10-100 $\mu\text{mol}\cdot\text{L}^{-1}$ for 21 days, prior to the onset of significant health effects, would greatly improve environmental risk assessment.

ACKNOWLEDGMENTS

This research was financially sponsored by National Science and Technology Major Project (2012ZX07505-001) and National Natural Science

Foundation of China (No. 21347007, 41271336 and 41201310)

REFERENCES

- [1] T. Das, A. Sharma, G. Talukder. 1988. Effect of lanthanum in cellular systems. *Biol. Trace Element Res.* 18: 201–228.
- [2] A. Kabata-Pendias, H. Pendias. 1992. Trace Elements in Soils and Plants, 2nd edn. CRC Press, Florida, pp.166–178.
- [3] Chengrun Wang, Mei He, Wen Shi, et al. 2011. Toxicological effects involved in risk assessment of rare earth lanthanum on roots of *Vicia faba* L. seedlings. *Journal of Environment Science*.23(10): 1721–1728.
- [4] Hui Zhao, Wei-dong Hao, Hou-en xu, et al. 2004. Gene expression profiles of hepatocytes treated with $\text{La}(\text{NO}_3)_3$ of rare earth in rats[J]. *World J. Gastroenterol.* 10(11): 1625–1629.
- [5] Li N., Wang S. S., Liu J., et al. 2010. The oxidative damage in lung of mice caused by lanthanoids[J]. *Biol. Trace Element Res.* 134: 68–78.
- [6] Wolf, H.D., Blust, R., Backeljau, T., et al. 2004. The use of RAPD in ecotoxicology. *Mutat Res.* 566, 249–262.

Determination of the cycle duration for 155 years in the Nikolaevsk-on-Amur meteorological station

E.E. Kholoden

Far East Federal University, 8, Sukhanova, Vladivostok Russian Federation

O.M. Morina

Pacific National University: 136, Tikhookeanskaya, Khabarovsk, Russian Federation

S.A. Lobanov

Far East Federal University, 8, Sukhanova, Vladivostok, Russian Federation

ABSTRACT: There are many hypotheses about the causes of global change of climate and landscapes. The study of these fluctuations is associated with an increase of economic, social, environmental, and political value of climatological information. A source was archival materials of hydrometeorological service of the Khabarovsk region by Nikolaevsk-on-Amur meteorological station in the period 1855–2010. The search of common factors of such changes is one of the most important tasks. A monthly determination of this dynamic is essential, as annual average values neutralize a picture of established trends. Duration of cycles was found by the graphs based on 5-years moving-average method as the most optimal number for a half century of observation. The average value was calculated from the 8–12 determined ones of regularly and rapidly changing air temperature. It was determined that the annual average cycle duration is comparable with the solar cycle of 11–12 years. However, the minimum number of years in the cycle may vary from 7 years to the maximum of 28 years, which also corresponds solar activity. During the year, the cyclical amplitude changes from 10 years in May to 20 years in March. The complexity and diversity of cycle forming in different months were showed.

KEYWORDS: Climate dynamic, cycles, mean monthly values of cycle duration.

1 INTRODUCTION

To date, researchers of dynamics of climate and landscape developed three possible variants of climate changes: warming, cooling, and stable development of process. The greatest number of scientific publications is devoted to the consideration of scenarios of global warming, which is associated with natural cycles and anthropogenic factors. According to some researchers, we are moving from one ice age to another, but the rate of changes is very small—about 0.02°C over the century. All this phenomena do not contradict the concept of anthropogenic climate change but relate to different time scales (Kokorin 2005).

Climate changes were not the same in all parts of the globe for a given period, which had differential nature. Most natural zones as with warming and with cooling are marked. There are vast areas in which the decrease in air temperature is observed with global hemisphere warming or without noticeable changes of annual surface temperature. The obtained data of annual averages allowed us to determine that in the Khabarovsk region and in the Jewish autonomous

region, 67% of the territory is subject to warming, 22% cooling, and 11% have a relatively smooth running of temperatures. Vectors of temperature changes throughout the year are often in the different directions. The result is that it is necessary to make a detailed temperature analysis for all nature-use types (Morina 2006, Morina et al. 2008).

The underlying reason for the rhythm and the cyclicity of physiographic process is pulse-coupled and wave fluctuations in the Earth's crust. Small rhythms occur against a background of longer rhythms. In genetic terms, grouping of rhythms and cycles appears from bottom to higher ranks—solar, planetary, stellar, and galactic. Basic rhythm of solar activity is on average 11–12 years.

In the study of climate change, it should be taken into account not only the anthropogenic but also the natural component of climate variety; because of the impact superposition of environmental factors on climate, anthropogenic warming wave may be damped by backward wave of natural climate cooling.

If natural process leads to the climate warming, the anthropogenic factors only enhance this process.

On the contrary, in the case of regular climate cooling, anthropogenic factors reduce the cooling intensity.

In recent decades, the synchronicity of solar activity is changing, and recent climate fluctuations is established, upon which it was concluded that there are causal relationships between these phenomena. Fluctuations in solar activity have polycyclic character. There are cycles of varying duration superordinate to each other (5–6, 10–12, 22–23 years, etc.). Changes in solar activity are accompanied by changes in the quantitative and qualitative composition of the solar radiation. The inflow of ultraviolet radiation to the Earth increases in cycles of amplification, coinciding with the formation of active sunspots. This entails a change in the geomagnetic field of the Earth and global process in the atmospheric circulation.

2 RESEARCH METHOD

As a source of information archival hydrometeorological data was used, this was processed by moving-average method of 5-year period as the most optimal one for these observations. The study was carried out in the coastal weather station “Nikolaevsk-on-Amur” in the Khabarovsk region from 1855 to 2010.

Based on the constructed graphs, cycle times are primarily calculated. Cycle in Greek means circle, and its essence is to establish the fact that all occurring fluctuations lead to reversion to the top of these change processes. In this case, it is clear that in the nature not only the 12-year solar cycle acts, which is also the average from 8 to 14 years of incident waves, but also in the individual areas and in some months, there are shorter or longer change periods. Cycle duration is calculated as the year difference between maximum or minimum temperatures, from which the process of temperature changes began.

Moving-average (overlapping) method was proposed at the end of the XIX century for smoothing of curves. It is a mathematical filter allowed to mark fluctuations with longer wavelength significantly repaid short-period fluctuations. Dynamical and statistical (probabilistic) regularities of nature are discernible by studying the interaction of external (exogenous) and internal causes (Sverlova 2004).

3 RESULTS AND DISCUSSION

It was found that the nature of the cycle forming for one and half a century is quite complicated and non-linear. Monthly averages of cycle times are shown in Figure 1. It can be noted that the duration of the winter cycles is relatively uniform and is of 12–15 years. In spring, fluctuations are enough violent—from 10 years in May to 20 years in March. Summer cyclicality

is close to 15 years, which may indicates the instability of natural processes and consistently requires stricter of doing any type of the nature use.

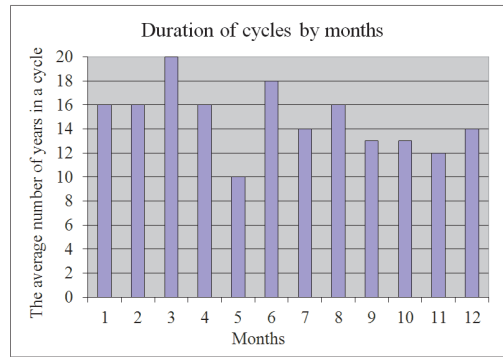


Figure 1. Distribution of average cyclicality periods during the year in the Nikolaevsk-on-Amur meteorological station.

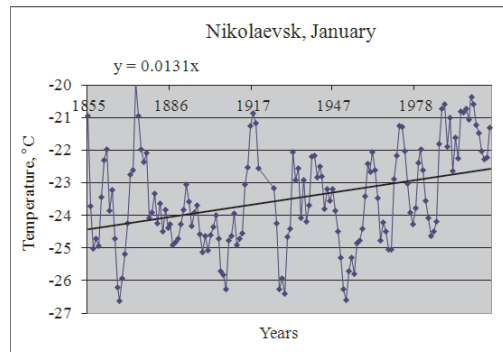


Figure 2. The formation of cyclicality and dynamic direction of the air temperature in the Nikolaevsk-on-Amur meteorological station in January.

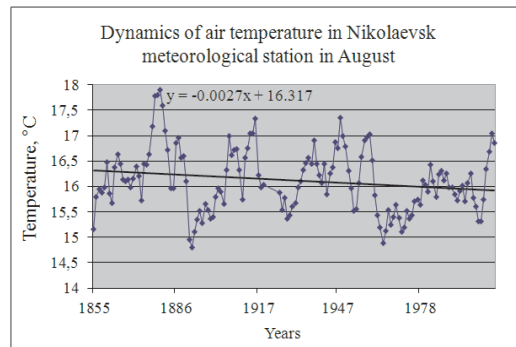


Figure 3. Cycle times and dynamic direction of air temperature in Nikolaevsk-on-Amur meteorological station in August.

Unequal of cycle lengths is complicated by multi-directional formation of climate change patterns. Hence, it is discovered that in January, air temperature increased to 2°C (Figure 2); in August over the study period, there is a tendency of temperature reducing to 0.4°C (Figure 3). Such a distribution of climate change is a characteristic of the northern hemisphere.

4 CONCLUSION

The necessity of monthly research of main climatic dependencies is due to highly irregularity of climate forming at each point of any territory. This statement is due to the fact that the annual average values of the temperature dynamics and cyclicity ones hide special aspects of distribution of these characteristics, not only from season to season but also for several months. However, for a different rate of nature use, it is necessary for us to take the dynamic character of climatic variables for different periods into account. Hence, for industrial and civil engineering, winter periods are important in view of reducing the risk of soil heaving. For agriculture and forestry, competency

to predict the dynamics of the climatic characteristics of the vegetation period is necessary.

It was determined that the duration of annual average cycle is comparable with the solar parameters of 11–12 years. However, the minimum number of years in the cycle may vary from 7 years to a maximum of 28 years, which also corresponds to the type of solar activity. During the year, the rate of cyclicity forming varies from 10 years in May to 20 years in March.

REFERENCES

- [1] Kokorin, A.O. 2005. *Climate changing: Review of research information about anthropogenic climate change*. Moscow: RREC, GOF, www-Russia.
- [2] Morina, O.M. 2006. Method of determination of the vulnerability of ecosystems as part of the environmental monitoring of abiotic factors. *Ecological Systems and Devices*. 12. 11–14.
- [3] Morina, O.M., Kholoden, E.E., Lobanov S.A., Derbentseva, A.M. 2008. *Dynamics of soil temperature under anthropogenic impact*. Vladivostok: FESU.
- [4] Sverlova L.I. 2004. *System analysis of natural process rhythms*. Moscow: Megalion.

Soil erosion spatial distribution simulation of the Jinghe river basin in the Loess Plateau of China

G.S. Wu, D.L. Wang, Y.C. Wu, X.Y. Chen & W.S. Wei

Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China

ABSTRACT: Watershed soil erosion model is an important tool to support watershed soil erosion control and ecological engineering planning. In this article, a region soil erosion assessment model is advanced. The model includes five independent factors such as rainfall erosivity, soil erodibility, vegetative cover, land use, and the topographical factor. And each factor is calculated by an independent equation, using such remote sensing data as vegetation, the DEM and land uses, rainfall characteristics, and other spatial data of Jinghe river basin on GIS platform, where the distribution of soil erosion factor is estimated. The spatial distribution of annual soil erosion during 1997–2008 over Jinghe river basin is assessed with the developed model. Historically recorded hydrological data are used to verify the assessment. The results indicate that the model can assess the spatial distribution of soil erosion and total watershed soil erosion. The water soil erosion model proposed in this article is of great significance to soil erosion assessment over the Loess Plateau.

KEYWORDS: Soil erosion, Jinghe river, Loess Plateau, Spatial distribution, GIS.

1 INTRODUCTION

Watershed soil erosion model is an important method for studying soil erosion processes and mechanisms, by monitoring and evaluating the soil conservation environment. This model is of important significance to watershed environment protection, ecological and agricultural planning, and monitoring of vegetation, soil, and water environment.

Researchers have developed a lot of statistical models for testing watershed and regional scale, mainly focusing on the improvement and application of USLE (Ahmet I et al., 2007; Yang Zisheng, 1999). By referring to American Soil Loss Equation, Liu Baoyuan (1994) developed the Chinese soil loss equation, to calculate the average annual soil loss of slope in many years^[3]. China has developed many small watershed soil erosion empirical models for the Yellow River Basin and the Loess Plateau. Mou Jingze model can be used to forecast the total annual sediment yield with few factors (Mu Jingze et al., 1980). Based on the results of analyzing the factors of 10 typical channels in the loess and gully region including rainfall runoff, vegetation, soil, and topography characteristics, Jiang Zhongshan (1980) developed the watershed soil erosion empirical model to predict the amount of soil erosion caused by rainstorm (Jiang Zhongshan et al., 1980).

With the development of geographic information systems and computer technologies in recent

years, distributed hydrological model has been widely developed and applied. Many physical parameters need to be calibrated for the distributed physical model, which has limitations in predicting the soil erosion of large watersheds on the Loess Plateau (Hu Liangjun et al., 2010). Meanwhile, with the application of geographic information systems and remote sensing technologies, quantitative evaluation of spatial distribution of soil erosion through GIS has become an effective means for rationally planning and using regional land and water resources and has been widely used in many countries (Vente J et al., 2011). This article intends to develop the watershed distributed statistical model with few parameters using GIS, RS technology, and empirical statistical model, to achieve low-cost data acquisition and easy calibration. This model can estimate the changes of watershed soil erosion in spatial distribution and time sequence and can directly provide reference figure objectives for ecological monitoring and planning of the Loess Plateau watershed.

This article has studied, analyzed, and summarized the previous empirical statistical models for watershed, developed the soil erosion evaluation model for Loess Plateau watershed, analyzed the factors of influencing soil erosion, obtained calculation models for all erosion factors, analyzed and processed the spatial data of Jinghe river basin, calculated and obtained spatial distribution of all factors, and assessed the time and process of annual soil erosion of Jinghe river basin by applying the soil erosion evaluation model.

2 METHODS

2.1 Model structure

On the basis of referring to the current Regional Soil Erosion Model, this article develops the soil erosion model for Loess Plateau watershed (Yu Tian, et al., 2011; J K Lørup et al., 2008; Pallavi Behera et al., 2005). This is a distributed empirical statistical model operating under GIS support. Main principles of developing the model include (1) limited and independent factors, (2) reliable and continuous regional data sources, and (3) applicability: rainfall, climate, and topography of the Loess Plateau. Factor independence means only considering influence of all factors on soil erosion rather than the interaction among factors. Limited and independent factors are conducive to improve the applicability of model. In case of change of certain region factor, it is only necessary to calibrate the changed factor equation, instead of changing model structure and other erosion factor equations; continuous and reliable data sources make this model possess the ability to continuously monitor soil erosion conditions of watershed for a long term; to achieve applicability, topography, rainfall, and other characteristics of the Loess Plateau, it should be considered to develop the watershed soil model. Through repeated revision, the structure of watershed soil erosion model is determined finally. Its structure is shown in [equation \(1\)](#):

$$SE = R \cdot K \cdot LC \cdot V \cdot T, \quad (1)$$

where SE is the average soil erosion amount per unit area, $t \text{ hm}^{-2} \text{ a}^{-1}$; R is the rainfall erosivity factor, $\text{MJ mm hm}^{-2} \text{ h}^{-1} \text{ a}^{-1}$; K is the soil erodibility factor, $t \text{ h MJ}^{-1} \text{ mm}^{-1}$; LC is the land use factor, dimensionless; V is the vegetation factor, dimensionless; and T is the topography factor, dimensionless.

2.2 Erosion factors

2.2.1 Rainfall erosivity factor

Erosive rainfall on the Loess Plateau is characterized by short duration and high intensity. The erosive storm rainfall generally lasts for about 1 h (Zhang Hanxiong et al., 1982). Annual rainfall R of the Loess Plateau region should have the intensity of minimum 10 min (I_{10}) and maximum 60 min (I_{60}) (Wang Wanzhong et al., 1982). Through the stepwise regression method, annual rainfall erosivity calculation equation is obtained, which is as shown in [equation \(2\)](#) (Wang Wanzhong et al., 1987).

$$R = 0.008P^{0.776}I_{10}^{0.965}I_{60}^{0.732}, \quad (2)$$

where R is the annual rainfall erosivity, $\text{MJ mm hm}^{-2} \text{ h}^{-1} \text{ a}^{-1}$; P is the annual rainfall, mm ; and I_{10} and I_{60} , respectively, denote the maximum rainfall intensity at 10 min and 60 min, mm min^{-1} .

2.2.2 Soil erodibility factor

Erodibility (soil erodibility) is an indicator of evaluating the difficulty level of soil separation, erosion, and transfer by rainfall erosivity and an internal factor of affecting soil erosion. In the erosion-productivity impact calculator (EPIC), Williams (1990) developed the equation of calculating soil erodibility factor K into an equation only related to soil sand content SA , silt content SIL , clay content CIA , and organic carbon content OC (Williams J R et al., 1990). [Equation \(3\)](#) is as follows:

$$K = \left\{ 0.2 + 0.3 \exp \left[-0.0255 \left(1 - \frac{SI}{100} \right) \right] \left(\frac{SI}{CL + SI} \right)^{0.3} \right. \\ \left. \left(1.0 - \frac{0.25OC}{OC + \exp(3.72 - 2.95OC)} \right) \left(1.0 - \frac{0.75SV1}{SV1 + \exp(-5.51 + 22.95SV1)} \right) \right\}, \quad (3)$$

where K is the soil erodibility factor, $t \text{ h MJ}^{-1} \text{ mm}^{-1}$; SA is the sand mass fraction, %; SI is the powder mass fraction, %; CL is the clay mass fraction, %; OC denotes the organic carbon mass fraction, %; $SV1 = I - SA/100$.

In 1999, FAO released the Harmonized World Soil Database (HWSD). HWSD adopted the 1:1,000,000 Soil Figure drawn by ISSCAS according to the results of second soil census (Shi X Z et al., 2004; Fu B J et al., 2004). For the physical and chemical parameters of soil provided by HWSD, soil erodibility factor can be directly calculated by [equation \(3\)](#).

2.2.3 Land use cover factor

During land use, factors of vegetation coverage, soil properties, and runoff rate are changed to influence the occurrence and development of soil erosion (Zhang Wen Jie et al., 2006). Soil coverage LC_p denotes the proportion of surface substances that cannot be eroded (such as crop residues, rocks, artificial buildings, and water body) to the surface area, namely the soil coverage ratio, with % as the unit. Coverage of the surface with straws can significantly reduce the soil erosion amount^[18]. Soil erosion amount is proportional to the area percentage of soil exposed to the surface, that is, the soil-surface area ratio and soil erosion amount are linear (Rajan, B et al., 2006). The surface utilization factor can be estimated by [equation \(4\)](#):

$$LC = 1 - LC_p, \quad (4)$$

where LC is the land use cover factor and LC_p is the soil coverage.

Soil coverage is estimated by land use data. This article adopts the classification system of the United States Geological Survey (Anderson J R et al., 1976) and uses the land use coverage data provided by USGS to estimate soil coverage. Land use coverage factor is estimated by [equation \(4\)](#).

2.2.4 Vegetation factor

Vegetation canopy and surface vegetation coverage can intercept rainfall, prevent surface soil from direct impact of raindrops, and weaken the scouring impact of runoff on soils, so as to reduce soil erosion (Liu Baoyuan et al., 2001). Normalized difference vegetation index (NDVI) as the most commonly used index has shown very good correspondence relationship with productivity of vegetation, and its changes, to a certain extent, can represent the surface vegetation changes. NDVI is calculated by [equation \(5\)](#):

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}, \quad (5)$$

where NDVI is normalized difference vegetation index and ρ_{nir} and ρ_{red} , respectively, denote the near-infrared reflectivity and infrared reflectivity.

Vegetation coverage denotes the proportion of vegetation's (including leaves, stems, and branches) vertical projection area to unit area (Zhang Wenbo et al., 2001). NDVI is used for estimation (Gutman G et al., 1998), which is shown in [equation \(6\)](#):

$$f = \frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}}, \quad (6)$$

where f is the vegetation coverage; NDVI is normalized difference vegetation index; $NDVI_{\max}$ and $NDVI_{\min}$, respectively, denote the maximum and minimum of NDVI in Jinghe river basin.

Vegetation's role of intercepting rainfall is also related to vegetation type. Different types of vegetation have different canopy heights and different roles of intercepting rainfall, thus influencing the occurrence of soil erosion (Wischmeier W H and Smith D D, 1965). The impact of vegetation factor on soil erosion is calculated by [equation \(7\)](#):

$$V = 1 - f \cdot \exp(-0.323 \cdot CH) / 100, \quad (7)$$

where V is the vegetation factor; f is the vegetation coverage, %; and CH is the raindrop landing height after vegetation canopy interception, m. Vegetation canopy height can be estimated according to the type of surface vegetation coverage.

From 1997 to 2002, satellite images of Landat5 Landsat7 were used; from 2003 to 2008, MODIS remote sensing data were used. Moderate resolution imaging spectroradiometer (MODIS) on the Terra satellite almost can conduct global hyperspectral measurement of day resolution and is used to manufacture land vegetation coverage products, such as leaf area index (LAI) and NDVI (Shabanov N V et al., 1997). This article adopts the MODIS land product MOD13A1, namely global 16d vegetation index synthetic product of 500 m resolution. After preprocessing of two kinds of remote sensing data, annual NDVI can be obtained through the maximum value composite (MVC) method. The canopy height of vegetation by combining the land use coverage data was estimated, and the vegetation factor according to [eqs. \(6\) and \(7\)](#) was also calculated.

2.2.5 Topography factor

Topography is an important factor in influencing the surface runoff and soil erosion (Yang Qinke et al., 2009). Driven by rainfall erosivity, soil erosion occurs mostly on the slope and rill (Cai Qiangguo et al., 1988). Watershed soil erosion probability and potential erosion can be measured through topography surface area and local elevation range, respectively, and topography factor can be calculated through the multiple regression analysis of soil erosion amount and topography parameters of typical watershed, which is shown in [equation \(8\)](#):

$$T = (-21.13 + 22.30 \times SD) \times \left(\frac{ER - ER_{\max}}{ER_{\max} - ER_{\min}} \right)^{-0.43}, \quad (8)$$

where T is the topography factor, SD is the topography surface density, dimensionless; ER is the elevation range, m. ER_{\max} and ER_{\min} , respectively, denote the maximum and minimum elevation ranges in the $1 \text{ km} \times 1 \text{ km}$ grid of Jinghe river basin, namely 92.5 and 6.3 m, respectively.

Digital elevation models (DEMs) are very convenient to represent the continuously varying topographic surface of the earth (Thompson et al., 2001). The DEMs used for this study is ASTER GDEM Version 2 released by the National Aeronautics and Space Administration, with the horizontal resolution of about 30 m. Three-dimensional (3D) area and elevation range in the $1 \text{ km} \times 1 \text{ km}$ scope are, respectively, extracted through the Zonal Statistics tool on the geographic information system platform, and the ratio between 3D area and two-dimensional area (1 km^2) is the topography surface density.

3 MODEL APPLICATION

3.1 Study area

It is located in the southern part of the Loess Plateau (east longitude of $106^{\circ}14'$ – $108^{\circ}42'$, and north latitude of $34^{\circ}46'$ – $37^{\circ}19'$), Jinghe river basin stretches over parts of areas of Ningxia, Gansu, and Shanxi Provinces. With a watershed area of $45,421 \text{ km}^2$, the vast majority of watersheds of Jinghe river basin belong to Longdong Loess Plateau. Based on geomorphological features, the Jinghe river basin can be divided into northern loess hilly area, central loess terrace, and gully areas, southwestern mountain forest area, and southeastern mountain river area. The overall topography of Jinghe river basin is that the eastern, northern, and western parts tilt toward the southeastern direction. Jinghe river basin is high in northwest and low in southeast. With rich river sources, Jinghe river basin has 13 main tributaries with a catchment area of more than $1,000 \text{ km}^2$ and 26 tributaries with a catchment area of more than 500 km^2 . Loess is widely distributed in the Jinghe river basin. Except for some areas of Liupan Mountain, the other areas have all been covered by loess. The loess is very thick, and its thickness is generally ranged from

50 to 80 m and can be as thick as 100 m in some parts of the Loess Plateau. The Jinghe river basin, located in the transition zone between the temperate semihumid climate and the temperate semiarid climate, has a typical temperate continental climate. With few rains, it is dry in winter but rainy in summer. The average temperature is 8°C . The annual precipitation is $350\text{--}600 \text{ mm}$. With a big precipitation intensity, the precipitation in summer and fall can account for 72%–86% of the total annual precipitation. The differentiation of climates in different areas leads to the differentiation of vegetation patterns, which lie in loess mountains, loess gullies, and loess hills, which are forests, bush woods, and grassland, respectively. For decades, the natural vegetation in Jinghe river basin has been seriously damaged. The current forest coverage rate is only 6.5%, with even 3% in some areas. The grass cover rate is 25%–65%. Soil erosion area of the Basin is 3.32 km^2 , accounting for 73% of the total watershed area. The average sediment runoff had been $2.458 \times 10^8 \text{ t}$ from 1950 to 2000, which is one of the main sources of the sediments in the Yellow River (Cao Ying et al., 2010). Soil erosion is an important reason for causing the deterioration of ecological environment in Jinghe river basin.

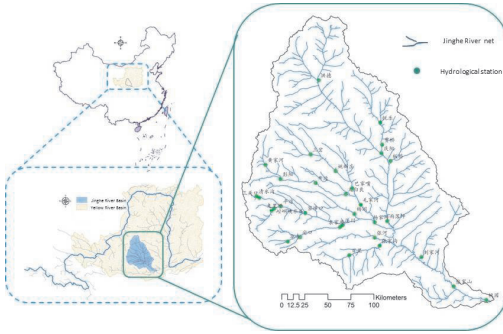


Figure 1. The location of Jinghe river basin.

3.2 Erosion factor

This article uses various kinds of watershed data sources, and relevant information is shown in Table 1. Remote sensing data are processed through the ENVI software, and all data from the hydrological and meteorological sites are processed into raster data through the optimal interpolation method Kriging. Rainfall data are the annual rainfall data and rainfall characteristics statistical data of 87 hydrological stations in the Jinghe river basin. All data are processed into the unified coordinates on the geographic information system platform, for the purpose of model calculations.

The distribution of erosion factors of Jinghe River basin from 1997 to 2008 is calculated through the

Table 1. The data sources and information.

| Data set | Factors | Data type | Data source |
|------------------------------------|--------------------|-----------|---|
| HWSD soil database | Soil Erodibility | Static | Food and Agriculture Organization of the United Nations |
| Land use Coverage data | Land use coverage | Dynamic | University of Maryland |
| DEM | Topography | Static | United States Geological Survey |
| MODIS, LandSat remote sensing data | Vegetation | Dynamic | National Aeronautics and Space Administration |
| Rainfall data from 1982 to 2009 | Rainfall erosivity | Dynamic | China Meteorological Administration, Jinghe River Data Center |

above factor calculation model. Among them, topography factor and soil erodibility factor are regarded as static factors, and only airborne distribution is considered. Other factors are dynamic factors. The distribution of soil erosion factors of Jinghe river basin in 2008 is shown in Figures 2–6:

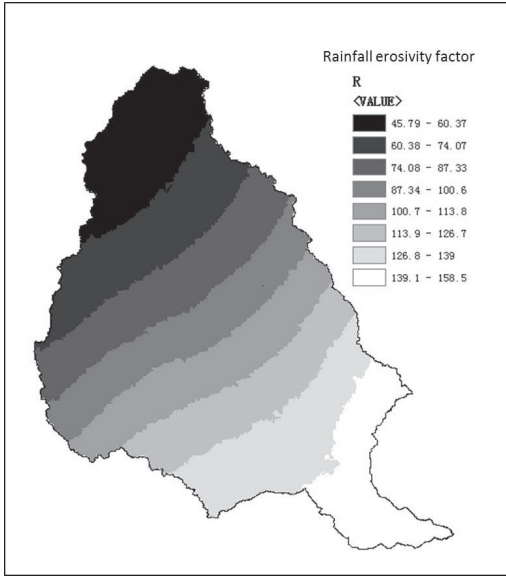


Figure 2. Rainfall erosivity factor.

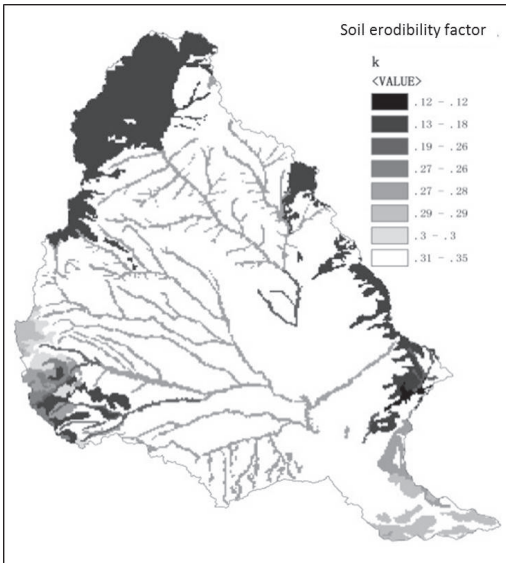


Figure 3. Soil erodibility factor.

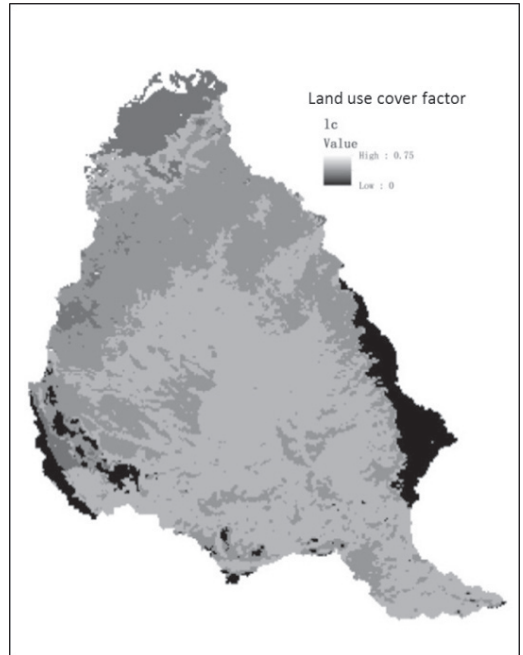


Figure 4. Land use coverage factor.

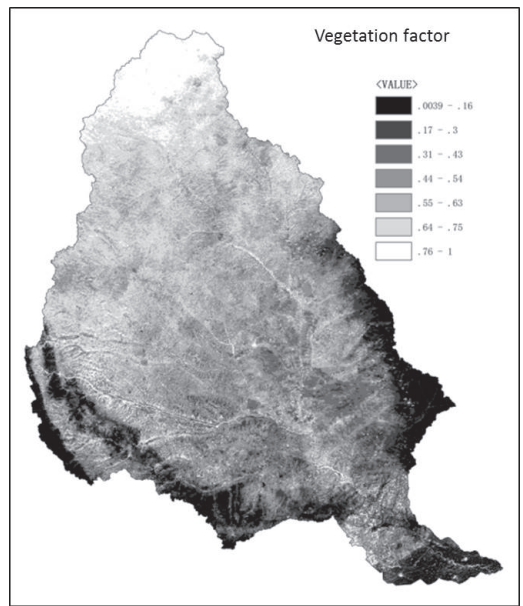


Figure 5. Vegetation factor.

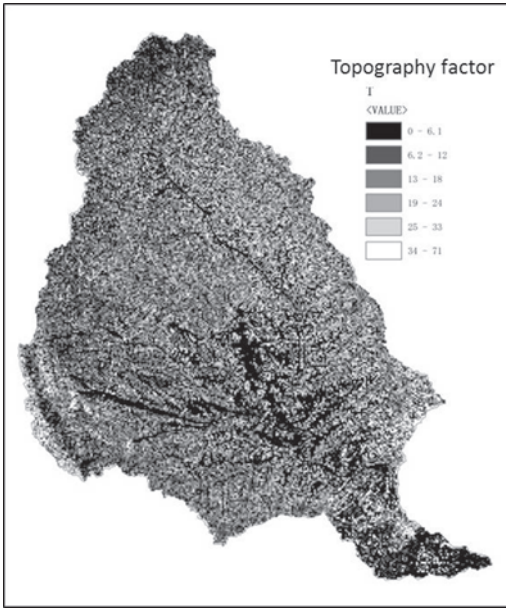


Figure 6. Topography factor.

4 RESULTS AND DISCUSSIONS

4.1 Spatial distribution simulation

By calling the calculation tool on the GIS platform and using [equation \(1\)](#), the spatial distribution of annual soil erosion of the Jinghe river basin from 1997 to 2008 is calculated. The soil erosion distribution of 1997 and 2008 is shown in [Figure 3](#). Results of soil erosion distribution show that the soil erosion amounts of southwestern and eastern parts of the Basin are low because of good forest coverage conditions. The areas suffering from the serious erosion are concentrated in the southeast part of the Basin and both sides of the main channel of Jinghe river. Spatial assessment accuracy inspection needs the support of small-watershed soil erosion data. Under the conditions of lacking these data, this article only makes qualitative analysis of spatial distribution accuracy.

4.2 Model performance

Based on the soil erosion distribution in the Jinghe river basin from 1997 to 2008, the total annual simulated erosion can be calculated. It must be noted that the total soil erosion amount of the Jinghe river basin is the total average sediment yield, and the sediment runoff of the basin outlet Zhangjiashan station refers to the total suspended runoff load. However, this article refers to the sediment runoff of basin outlet for

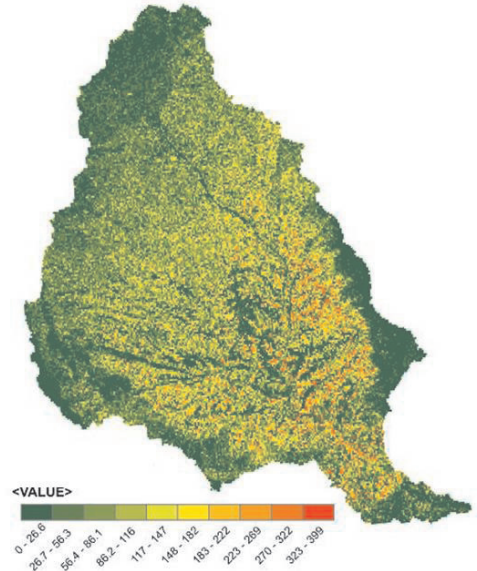


Figure 7. The soil erosion distribution over Jinghe river basin of 1997.

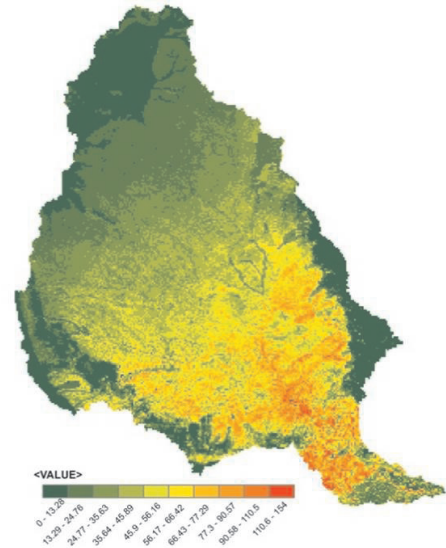


Figure 8. The soil erosion distribution over Jinghe river basin of 2008.

model building, and the sediment delivery ratio of many watersheds in the Loess Plateau region is 1 (Liu Jigen et al., 2004). The sediment yield period of the Yellow River basin is mainly concentrated in the flood season, and the sediment runoff in the flood season generally accounts for over 80% of the total throughout the year. The sediment yield of the basin

is mainly closely related to limited rainstorms. Based on the above reasons, the comparison between total soil erosion amount and sediment runoff has certain reference significance. The measured value and the assessed value of annual soil erosion of the Jinghe river basin are shown in Figure 9.

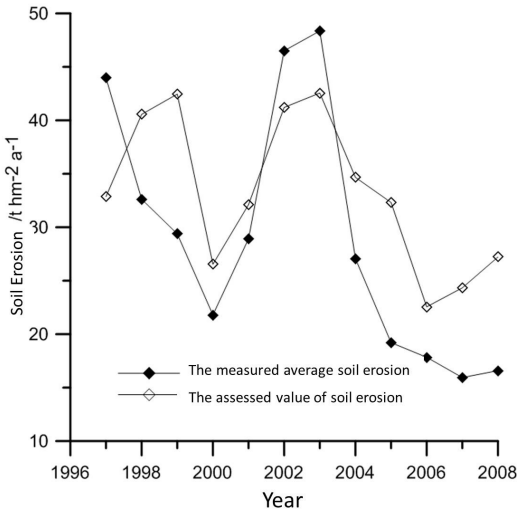


Figure 9. The measured annual soil erosion and the assessed value.

Figure 9 shows the fitting relationship for analysis of linear correlation between the measured value and the assessed value of soil erosion of the Jinghe river basin. We can see that the measured annual soil erosion and the assessed value of soil erosion over Jinghe river basin have the same trends at the annual time series and, to some extent, show the feasibility and reliability of the proposed model.

The correlation between the measured annual soil erosion and the assessed value is tested through the statistical software SPSS. Test result (Figure 10) shows that the correlation between total assessed amount of soil erosion and the measured sediment runoff is significant, and the determination coefficient R^2 is 0.71.

Assessment results show that total assessed amount of soil erosion has been higher than the exported sediment runoff since 2004, mainly because the runoff volume of Jinghe river basin has been reducing significantly in recent years, sediment transport capacity of larger tributary and trunk river channels has been declining, and sand generated by the rainfall has sediment in the trunk river channels, and also because of decreased suspended quality ratio of the Jing River basin and the increased proportion of bed-load sediment. Further research on these reasons is necessary.

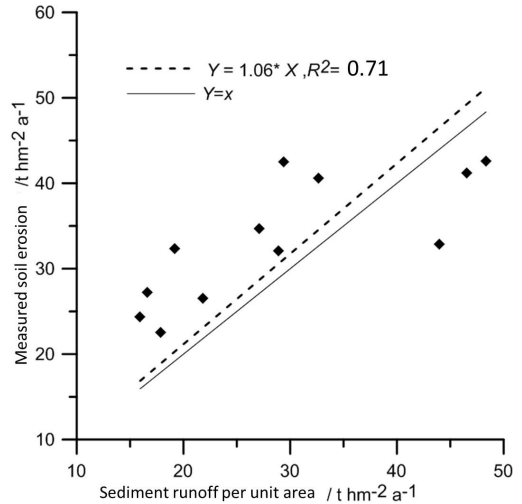


Figure 10. The correlation between the measured and the assessed value of annual soil erosion.

5 CONCLUSION

Through analysis of the watershed soil erosion mechanism, this article summarizes the factors affecting soil erosion and analyzes the erosion factor calculation method, to develop the soil erosion assessment model based on the watershed distributed empirical statistics. With support of GIS, this article, using spatial data about remote sensing, rainfall, soil and land use, taking Jinghe river basin, for example, and applying the soil erosion model, assesses the spatial distribution of annual soil erosion of the Jinghe river basin from 1997 to 2008. The empirical model developed can quickly assesses the spatial distribution of soil erosion and provides scientific basis and decision support for governance of watershed soil erosion.

ACKNOWLEDGMENT

This research is supported by 2014 Agricultural Resources Planning Project: Full-dimensional information management and application of Agriculture Regional Planning, and Central-Level Nonprofit Research Institutes Special Fundamental (301-3,302), National Science and Technology Support Program (2011BAJ07B05). First, hereby extend gratitude to Dr. Wu Yongchang, and the Institute of Ecology Studies, Landscape Ecology and Sustainable Science Research Center, and Jinghe river basin Data Center of Beijing Normal University for providing Jinghe river basin Data Download Service, and National Climatic Data. Second, I want to thank

China Meteorological Data Sharing Service Network for providing the rainfall data. And thanks to global data service provided by NASA.

REFERENCES

- [1] Ahmet I, Fatih T, Veli U. Estimating spatial distribution of soil loss over Seyhan River Basin in Turkey, *Journal of Hydrology*, 2007, 336:30–37.
- [2] Yang Zisheng. The topographic factor of soil erosion of sloping cultivated land in the northeast mountain region of Yunnan province[J]. *Journal of mountain research*, 1999, (17): 16–18(in Chinese with English abstract)
- [3] Liu B Y, Nearing M A, Risse L M. Slope gradient effects on soil loss for steep slopes. *Transactions of the ASAE*, 1994, 37(6): 1835–1840.
- [4] Mu Jingze, Xiong Guiqu. small watershed sediment yield prediction in Northern Shaanxi and soil conservation measures sediment retention calculation, *The 1st International Colloquium Proceedings*[C], Beijing: Guanghai Press, 1980. (in Chinese)
- [5] Jiang Zhongshan, Song Wenjing. Sediment yield calculations of the middle reaches of the loess hilly area watershed on Yellow River[C]. *The 1st International Colloquium Proceedings* [C], Beijing: Guanghai Press, 1980. (in Chinese)
- [6] Hu Liangjun, Yang Haijun, Yang Qinke, et al. A GIS-based Modeling Approach for Fast Assessment of Soil Erosion by Water at Regional Scale, Loess Plateau of China. *Chin. Geogra. Sci*, 2010, 20(5): 423–433.
- [7] Vente J, Poesen J, Verstraeten G, et al. Spatially distributed modeling of soil erosion and sediment yield at regional scales in Spain[J]. *Global and Planetary Change*, 2008, 60(3/4): 393–415.
- [8] Yu Tian, Jihui Fan, Feipeng Xiao, et al. A New Soil Erosion Model for Hilly Region Based on Information Technology[J]. *Advanced Research on Computer Science and Information Engineering*, 2011, (153): 473–483.
- [9] J. K. Lørup, M. Styczen. *Soil Erosion Modeling* [J]. *Water Science and Technology Library*, 1996, (22): 93–120.
- [10] Pallavi Behera, K. H. V. Durga Rao, K. K. Das. Soil erosion modeling using MMF model -A remote sensing and GIS perspective [J]. *Journal of the Indian Society of Remote Sensing*, March 2005, 33(01): 165–176.
- [11] Zhang Hanxiong, Wang Wanzhong. Storm Characteristics and distribution characteristics on the Loess Plateau[J]. *Bulletin of Soil and Water Conservation*, 1982(01): 19–25. (in Chinese)
- [12] Wang Wanzhong. Study on the relations between-rainfall characteristics and loss of soil in loess region. Correlation between rainfall parameters and soil loss[J]. *Bulletin of Soil and Water Conservation*, 1983(04): 7–13. (in Chinese with English abstract)
- [13] Wang Wanzhong. Study on index of erosivity R of rainfall in loess area[J]. *Soil and Water Conservation In China*, 1987(02): 34–37. (in Chinese with English abstract)
- [14] Williams J R, Sharply A N. EPIC—Erosion Productivity Impact Calculator I. Model Documentation. US Department Of Agriculture Technical Bulletin. Number 1768[R]. Washington D. C. USDA ARS, 1990.
- [15] Shi X Z, Yu D S, Warner E D, et al. Soil Database of 1: 1,000,000 Digital Soil Survey and Reference System of the Chinese Genetic Soil Classification System[J]. *Soil Survey Horizons* 2004, 45: 129–136.
- [16] FAO, IIASA, ISRIC, et al. *Harmonized World Soil Database (version 1.1)*[R]. FAO, Rome, Italy and IIASA, Laxenburg, Austria. 2009.
- [17] Fu B J, Zhao W W, Chen L D, et al. Multiscale soil loss on evaluation index[J]. *Chinese science Bulletin*. 2006, 51(4): 448–456.
- [18] Zhang WenJie, Zheng Yi, Fullen M A, et al. Effects of Rice Straw Covering Net on Soil Erosion and Wheat Yield under Simulated Rainfall. *Chinese Journal of Soil Science*, 2006, 37(5): 969–974. (in Chinese with English abstract)
- [19] Rajan B, Khera K L. Effect of tillage and mode of straw mulch application on soil erosion in the submontaneous tract of Punjab, India[J]. *Soil&Tillage Research*, 2006, 88: 107–115.
- [20] Anderson J R, Hardy E E, Roach J T, et al. *A land use and land cover classification system for use with remote sensor data*[R]. Geological Survey Professional Paper, Washington, DC. 1976: 28–34.
- [21] Liu Baoyuan, Xie yun, Zhang Keli. *Soil Erosion Prediction Model*[M]. Beijing: China Science and Technology Press, 2001. 72–75. (in Chinese)
- [22] Zhang Wenbo, Fu Suhua, Liu Baoyuan. Error assessment of visual estimation plant coverage[J]. *Journal of Beijing Normal University(Natural Science)* 2001, 37(3): 402–408. (in Chinese with English abstract)
- [23] Gutman G, Ignatov A. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models[J]. *International Journal of Remote Sensing*, 1998, 19(8): 1533–1543.
- [24] Wischmeier W H, Smith D D. *Predicting rainfall-erosion losses from cropland east of the Rocky Mountains*[R]. USDA Agricultural Hand book No. 282. 1965.
- [25] Shabanov N V, Huang D, Yang W, et al. Analysis and optimization of the MODIS leaf area index algorithm retrievals over broadleaf forests[J]. *IEEE Transactions on Geoscience and Remote Sensing*. 1997, 43, 1855–1865.
- [26] Yang Qinke, Zhao Mudan, Liuyongmei, et al. Application of DEMs in Regional Soil Erosion Modeling[J]. *Geomatics World*, 2009. 02(1): 25–31. (in Chinese with English abstract)
- [27] Cai Qiangguo. Slope erosion and sediment yield model[J]. *Geographical Research*, 1988, 7(4): 94–122. (in Chinese with English abstract)
- [28] Thompson, J.A., Bell, J.C., Butler, C.A., 2001. Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma* 100, 67–89.
- [29] Cao Ying, Zhang Guanghui, Luo Rongting. Response of runoff and sediment discharge to global climate change in Jinghe River Basin[J] *Science of Soil and Water Conservation*. 2010, 8(2): 30–35. (in Chinese with English abstract)
- [30] Liu Jigen, Cai Qiangguo, Fan Liangxin, et al. Methods of scale transfer in modeling of soil erosion and sediment yield in catchments[J]. *Journal of Sediment Research*, 2004(3): 69–74. (in Chinese with English abstract)

Predicting sinking resistance of jacked piles with CPT in a multilayer soil at the pearl river delta alluvial plain

Y.Q. Hu

Engineering school, Sun Yat-Sen University, Guangzhou, China.
School of civil engineer, Guangzhou University, Guangzhou, China

L.S. Tang & H.T. Sang

Guangdong Provincial Key Laboratory of Mineral Resources & Geological Processes, Guangzhou, China
Department of earth science, Sun Yat-Sen University, Guangzhou, China

Z.Z. Li

Guangdong Yongji building foundation Co. Ltd, Shunde, China

ABSTRACT: Comparative tests between a new type of cone penetration, which an auxiliary total force sensor has installed, and model piles in same site are done in the Pearl River delta region typical strata. The relationships of the shaft resistance and total resistance between the model piles and the new type of CPT are analyzed, and the characteristics and working mechanism of the shaft and tip resistance of pile in the pile penetration process are investigated. Through analyzing the phenomenon of penetration of the model pile and the CPT, a novel empirical method is proposed, and a new simple empirical formulation that can be used to estimate the total resistance of jacked pile is presented at last. The influence of sliding friction and the share of tip resistance should be considered when the shaft resistance of pile is calculated. The continuous injection mobilization coefficient η , the share ratio n , and its correction coefficient ξ of the total shaft resistance are introduced to calculate the total resistance.

1 INTRODUCTION

Jacked piles are more and more widely used in recent years because pile jacking technology allows displacement piles to be installed without noise and vibration. The sinking resistance of the jacked pile is the installation resistance at the end of jacking. (Jackson 2007) reveals that the sinking resistance is a short-term axial force that is different from the long-term ultimate capacity. The selection of the piling machine and the possibility of the pile installation depend on the sinking resistance, and even the long-term ultimate capacity can be predicted by using the sinking resistance too. Hence, attention is paid to the sinking resistance of jacked pile, and several prediction methods based on the CPT data have been proposed (Chu and Wang 2000; Fan 2007; Han 1996; Jackson 2007; Li et al. 2010; Li et al. 2011; Zhang and Deng 2003; Zheng and Gu 1998; Zhu 1992). The difference between the sinking resistance and the long-term ultimate capacity is mainly caused by the fatigue of shaft resistance (Gavin and O Kelly 2007; Reddy and Stuedlein 2014; Samson and Authier 1986; Skov and Denver 1988). These methods adopted empirical coefficient (Han

1996; Zhu 1992; Jackson 2007), sensitivity coefficient (Zheng and Gu 1998; Fan 2007; Li et al. 2010), and sliding friction coefficient (Li et al. 2011), and the total shaft resistance is added up layer by layer.

However, the shaft resistance is variable during the whole installation, and the variation is concerned with the base resistance. Hence, the empirical coefficient of each layer is difficult to be determined. In this article, a new prediction method of the total shaft resistance is proposed and applied on the prediction the sinking resistance.

2 RESISTANCE TEST

2.1 *The test instruments for the CPT and model pile*

In the CPT, if the total force P on the top of rod could be measured, the total shaft resistance of rod corresponding to the any depth of penetration would be acquired. As a result, the share ratio between the shaft and the total resistance could be naturally found. Therefore, a new CPT instrument, which is fixed an auxiliary sensor of total resistance force P , is adopted and developed by Foshan Shunde survey and design limited

company (Chinese patent number: 200420044127.5), to collect the data of shaft and end resistance of CPT instrument. These data are used to accomplish the comparative study with the model pile.

As for the model pile (Fig. 1), a closed-end steel pipe pile is chosen, which the diameter is 300 mm. To obtain the test data of base-bearing capacity, shaft resistance, and lateral pressure on the model pile, two pressure plates are installed on the side and bottom of the model pile separately, and the distance from the end of pile to the lateral compression plate is 55 cm. In addition, the upper and lower friction tubes are coordinated on the model pile, where the distance between these two friction tubes is 2 m. Considering the above model pile, a static driver is adopted to press the pile into the strata, and then, the required test results are collected by the data acquisition system.

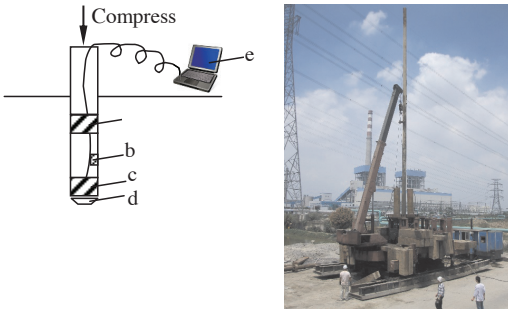


Figure 1. The structure of model pile and the test in the field: a. upper friction tube; b. lateral pressure plate; c. lower friction tube; d. tip pressure plate; and e. data collection system.

Finally, the Pearl River Delta alluvial plain is selected as measurement ground, which belongs to fluvial landform, and the geologic information is given in Table 1. Since the subsoil profile is in non-homogeneous, the calculation of the shaft resistance requires division of the entire soil profile into several layers.

3 THE TESTING RESULTS AND ANALYSIS

3.1 Base resistance analysis

As mentioned before, the base resistance q_b is usually supposed to equal CPT resistance q_{c_s} , that is, $q_b = \beta q_{c_s}$, where $\beta < 1$ is a reduction factor (Bustamante and Ganeselli 1982; Kuitert and Beringen 1979). This reduction factor reflects the effect of size difference between the model pile and the rod of the CPT instrument. Therefore, the factor β is named as the

correction coefficient of size effect. As a result, the following equation can be used to calculate the end bearing capacity:

$$Q_p = \beta A_p q_p \quad (1)$$

where Q_p is the end resistance force of pile, A_p is the area of pile section, q_p is the tip resistance of CPT, and β is the correction coefficient of size effect.

According to the result of test, the base resistance curve of model pile is very close to the tip resistance curve of CPT except at the sudden change of resistance, especially the sudden increase (Fig. 2). The difference may be ascribed to the size effect in these two measurements. The intersection area of model pile is 707 cm², and that of CPT probe is 15 cm², so the cone penetrator is more sensitive to the variation of resistance than the pile. Therefore, the sudden increase

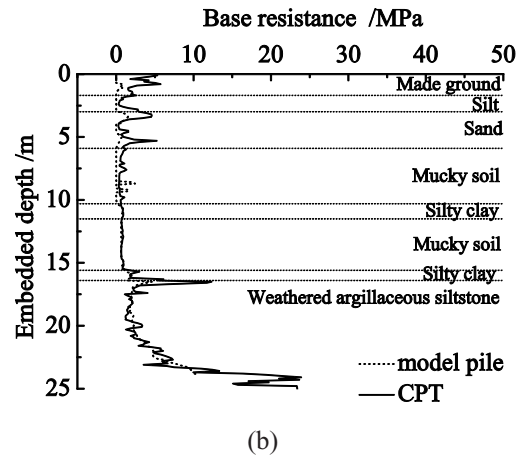
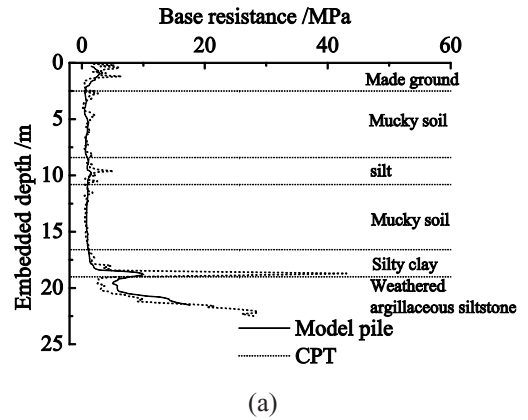


Figure 2. Comparison of the share ratio curve of total shaft resistance between the model pile and the corresponding CPT: (a) pile A and (b) pile B.

should be smoothed down. In the Pearl River Delta area, as shown in Figure 2, the correction coefficient of size effect β is very close to 1.

3.2 Shaft resistance analysis and the continuous installing mobilization factor

Due to the extension of auxiliary total resistance sensor in the new CPT instrument, the data of shaft resistance corresponding to any penetration depth of model pile can be obtained. Therefore, these data can be compared with the computation results based on the total ultimate shaft resistance. This comparison has been shown in Figure 3. The comparative study has found that the ultimate total shaft resistance is apparently larger than the shaft

resistance corresponding to a specific penetration depth. This means that the actual shaft resistance is less than its ultimate value. This phenomenon is caused by the increasing of pore water pressure and structure disturbance of lateral soils (Reddy and Stuedlein 2014), fatigue (Basu et al. 2011), or sliding friction (Li et al. 2010; Li et al. 2011; Zhang and Deng 2003).

Between the piles are the soils during installation. It is also the reason of setup effect in pile installation.

The present study shows that the time-related increase in capacity (i.e., setup effect) results from the shaft resistance mainly (Gavin and O Kelly 2007; Reddy and Stuedlein 2014; Samson and Authier 1986; Skov and Denver 1988). The test also reveals that the actual total shaft resistance is variable with the change of base resistance. Therefore, the shaft resistance is the key to prediction sinking resistance.

To calculate the decrease of total shaft resistance, a new parameter to characterize this phenomenon is introduced. That is, the continuous injection mobilization factor. The definition of this factor is as follows:

$$\eta = \frac{P - A_p q_c}{u \sum l_i f_{si}} \quad (2)$$

That is,

$$Q_{cs} = \eta u \sum l_i f_{si}, \quad (3)$$

where Q_{cs} is the total shaft resistance of rod in CPT, P is the total force of penetration, A_p is the area of tip, f_{si} is the shaft resistance of i th layer, q_c is the end resistance of rod, and u is the perimeter of probe intersection. Similarly, the injection mobilization factor is also available for the model pile, for example, as to the case illustrated in Figure 3(a), the factor $\eta = 0.671$ for CPT and the factor $\eta = 0.495$ for the model pile. If the size effect is not considered, the mobilization factor is 0.455. The reason of difference is so-called size effect.

Furthermore, considering the reduction factor between the shaft resistance of model pile and CPT, the shaft resistance of model pile can be estimated by the following equation:

$$Q_s = \alpha \eta u \sum l_i f_{si}, \quad (4)$$

where Q_s is the total shaft resistance of model pile and α is the coefficient of size effect.

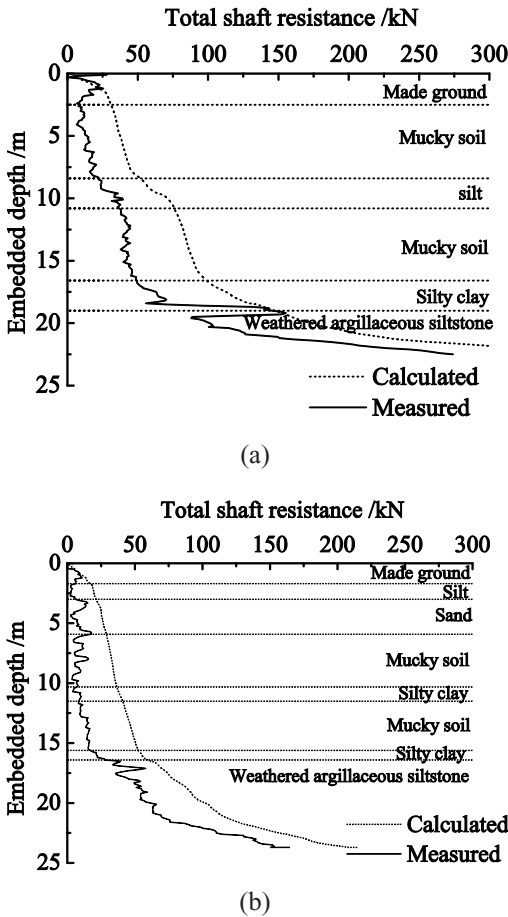


Figure 3. Comparison between the measured total shaft resistance and its corresponding ultimate total shaft resistance calculated using CPT data during installation: (a) pile A and (b) pile B.

3.3 Prediction of sinking resistance

The test reveals that the actual total shaft resistance is variable with the change of base resistance; therefore, the actual total shaft resistance is concerned with the base resistance.

Therefore, a new share ratio of total shaft resistance in sinking resistance is defined. For CPT, the share ratio of total shaft resistance can be described as follows:

$$n_0 = \frac{\eta u_0 \sum l_i f_{si}}{\eta u_0 \sum l_i f_{si} + A_{p0} q_p} \quad (5)$$

It has been illustrated that the total shaft resistance share ratio of CPT in sinking resistance is correlative closely to that of jacked piles, as shown in Figure 4. Hence, the share ratio can be used to predict the sinking resistance of pile.

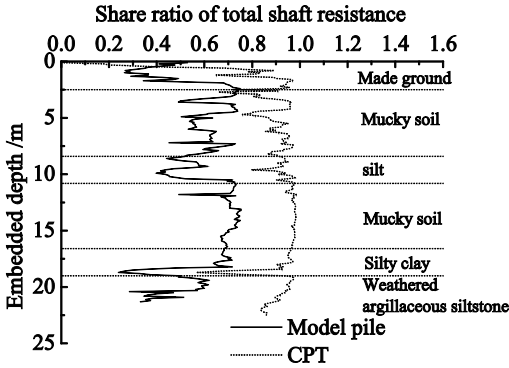


Figure 4. Comparison of the share ratio curve of total shaft resistance between the model pile and the corresponding CPT (pile B).

For the pile, the share ratio can be calculated with CPT and is defined as following:

$$n = \frac{\alpha \eta u \sum l_i f_{si}}{\alpha \eta u \sum l_i f_{si} + \beta A_p q_p}, \quad (6)$$

where A_{p0} and A_p are the areas of rod in CPT and model pile, respectively, and u_0 and u are the circumference of rod intersection and model pile, respectively.

Combining the equations (5) and (6), the flowing equation is obtained.

$$\left(\frac{\frac{1}{n} - 1}{\frac{1}{n_0} - 1} \right) = \frac{\beta}{A_{p0}} \cdot \frac{A_p u_0}{\alpha u} \quad (7)$$

ζ is introduced as flowing

$$\xi = \frac{\beta}{\alpha} \cdot \frac{A_p u_0}{A_{p0} u} \quad (8)$$

Then, the share ratio of jacked piles can be defined as following:

$$n = \frac{n_0}{\xi + (1 - \xi) n_0} \quad (9)$$

As a result, the total shaft resistance of model pile can be estimated using the following equation:

$$Q_s = \frac{n}{1 - n} \beta A_p q_p = \frac{n_0}{(1 - n_0) \xi} \beta A_p q_p \quad (10)$$

Finally, the total resistance of jacket pile may be predicted according to the following empirical formulation:

$$Q = Q_s + Q_p = \frac{\xi(1 - n_0) + n_0}{\xi(1 - n_0)} \beta A_p q_p \quad (11)$$

In this formulation, the total resistance could be estimated considering the size effect and the share ratio.

4 CONCLUSIONS

A novel empirical method is proposed based on a new CPT instrument and comparative study between the model pile static penetration test and CPT in this article. Using the auxiliary total penetration force sensor in CPT, the shaft resistance is measured along the model pile corresponding to the any penetration depth. Three new parameters are also introduced to describe the effects of size, mobilization, and share ratio on the total resistance. At last, a new simple empirical formulation that could be used to predict the bearing capacity of jacket pile is obtained.

However, this formulation derives from the data of model pile static penetration test, in which the diameter is 300 mm. As to different sizes of model pile, the

empirical equation is not investigated yet. In a word, this empirical method should be identified in further research works and practical applications.

REFERENCES

- [1] Basu, P., Loukidis, D., Prezzi, M., and Salgado, R. (2011). "Analysis of shaft resistance of jacked piles in sands." *International Journal for Numerical and Analytical Methods in Geomechanics*, 35(15), 1605–1635.
- [2] Bustamante, M., and Gianceselli, L. "Pile bearing capacity prediction by means of static penetrometer CPT." *Proceedings of the 2-nd European Symposium on Penetration Testing*, Amsterdam, 493–500.
- [3] Chu, W., and Wang, N. (2000). "Analysis and calculation of sinking-pile resistance of jacked piles." *Geotechnical Engineering Technique*(01), 25–28+46.
- [4] Fan, X. (2007). "Study of sinking resistance and compaction effect of static-press pile," PHD, Tongji University.
- [5] Gavin, K. G., and O Kelly, B. C. (2007). "Effect of friction fatigue on pile capacity in dense sand." *Journal of Geotechnical and Geoenvironmental Engineering*, 133(1), 63–71.
- [6] Han, X. (1996). "Experimental research on static pressure and bearing capacity of statically pressed piles." *Journal of building structures*, 17(6), 71–77.
- [7] Jackson, A. (2007). "Pile jacking in sand and silt." *Winner of the ICE's national Graduates and Students Papers Competition 2008, Feb 2008*.
- [8] Kuiter, J. D., and Beringen, F. L. (1979). "Pile foundations for large North Sea structures." *Marine Georesources & Geotechnology*, 3(3), 267–314.
- [9] Li, Y., Li, J., and Zhao, Z. (2010). "Estimation of Resistance of Static Pressure Pile Sinking Based on Cone Penetration Test." *Subgrade Engineering*(3), 67–69.
- [10] Li, Z., Tang, L., Hu, Y., Li, Z., and Lin, P. "A new estimation method for penetration depth of displacement pile based on a new type cone penetration test.", Jinan, China, 234–244.
- [11] Reddy, S. C., and Stuedlein, A. W. "Time-Dependent Capacity Increase of Piles Driven in the Puget Sound Lowlands." *From Soil Behavior Fundamentals to Innovations in Geotechnical Engineering@sHonoring Roy E. Olson*, 464–474.
- [12] Samson, L., and Authier, J. (1986). "Change in pile capacity with time: case histories." *Canadian Geotechnical Journal*, 23(2), 174–180.
- [13] Skov, R., and Denver, H. "Time-dependence of bearing capacity of piles." *Proc. Third International Conference on the Application of Stress-Wave Theory to Piles. Ottawa*, 25–27.
- [14] Zhang, M., and Deng, F. (2003). "A spherical cavity expansion-sliding friction calculation model on penetration of pressed-in piles." *Rock and Soil Mechanics*(05), 701–704+709.
- [15] Zheng, G., and Gu, X. (1998). "Analysis of some problems of statically pressed piles in soft soils." *Journal of building structures*, 19(4), 54–60.
- [16] Zhu, X. (1992). "The prediction of pile jacked installation resistance and the installation possibility." *Theory and Practice on Soft Soil Foundations*, D. Gao, ed., China building industry publishing house, Publishing house of Tongji University, Beijing, 125–128.

Numerical simulation of oil spill caused by ship collision in Dongjiakou Port

R.J. Hu, N. Wang, S.H. Jiang, L.H. Zhu & J.Z. Wu

College of Marine Geoscience, Ocean University of China, Qingdao, Shandong, China

Key Lab of Submarine Geosciences and Prospecting Techniques, Ministry of Education, Ocean University of China, Qingdao, Shandong, China

F. Ma

North China Sea Branch of the State Oceanic Administration, Qingdao, Shandong, China

ABSTRACT: On the basis of oil particle model, the trajectories of oil spill caused by ship collision were simulated. The spill model is based on Lagrangian discrete parcel method, which has the characteristics of high stability and high efficiency. When oil spill happens at the high water time, oil slick mainly moves parallel to the coastline with the interaction of NW wind and tidal current. While the oil slick mainly spreads toward southwest with the interaction of NE wind and tidal current. The spill trajectory is dominated by winds rather than currents. Oil spill will bring serious influence to sensitive objects such as Nature Reserve, Zhaitang Island, and National Ocean Park. This study can provide scientific basis for the relevant departments to formulate oil spill emergency plan.

1 INTRODUCTION

Oil spill accidents bring about much ecological damage to local environment, resulting in huge economic loss (Andrade et al. 2010, Xu et al. 2012). Dongjiakou port is an important part of the port of Qingdao, located in Langyatai bay, south of Qingdao (Fig. 1). Dongjiakou port is a natural fine deep-water port with the average depth about 15 m. The length of quay line is about 35.7 km, planned berth is 112, and the total handling capacity will reach 370 million tons after its completion. Dongjiakou port is a national bulk cargo distribution center and an important transportation center. The types of goods are mainly composed of bulk cargo and groceries. The coastal water of Qingdao is not only the traditional area of fishery production but also the spawning and feeding grounds (Ren et al. 2005). In addition, there are lots of aquaculture area and salt pan around the port. With the use of the port and the increase in the number of ships, Dongjiakou port and nearby sea area will become the key area of oil spill risk (Sun et al. 2012). Therefore, the research on drift trajectory and diffusion area of oil spill in the Dongjiakou port has important practical significance for the protection of fishery resources and the water quality environment. At the same time, it can provide the scientific basis of marine oil spill emergency response measures for the relevant departments.

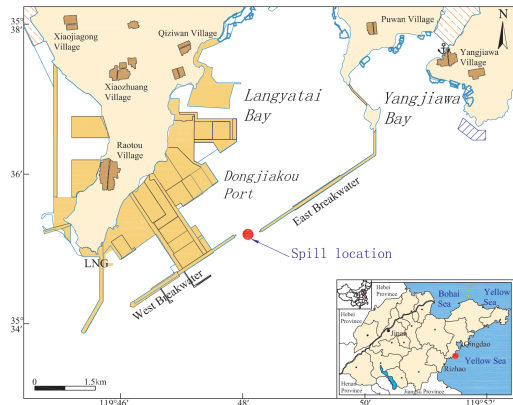


Figure 1. The oil spill location (located at the entrance of Dongjiakou Port).

2 CHARACTERISTICS OF TIDAL CURRENT FIELD

The tidal property of the sea nearby Dongjiakou port belongs to regular semidiurnal tide. The tidal current property is regular semidiurnal tidal current, and the tidal current motion is the reciprocating flow. Current velocity inside the breakwater is small, while current velocity outside the breakwater is larger. At the

middle time during flood tide, tidal currents mainly flow from NE to SW, and the current velocity is generally between 35 and 80 cm/s (Fig. 2). At the middle time during ebb tide, tidal currents mainly flow from SW to NE, and the current velocity generally ranges from 35 to 80 cm/s (Fig. 3).

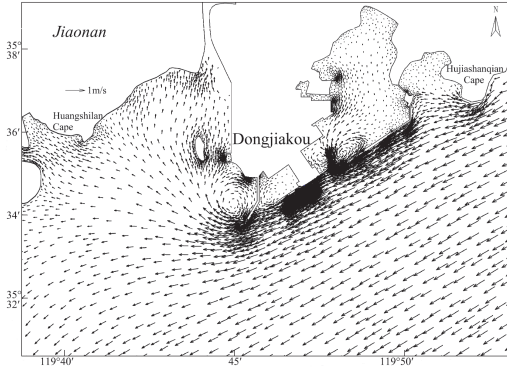


Figure 2. The distribution of flood tide field.

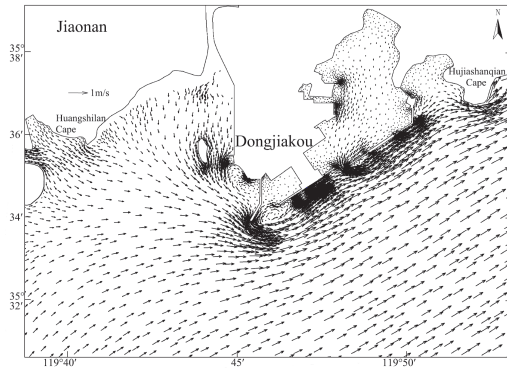


Figure 3. The distribution of ebb tide field.

3 BASIC PRINCIPLE OF OIL SPILL MODEL

Spill modeling has been recognized as an effective tool to hindcast and forecast the trajectory and oil properties evolution during an oil spill accident (Janeiro et al. 2012, Vethamony et al. 2007). The results are used by decision makers not only to manage pollution crises but also in programming and setting up contingency plans (Catarina Frazão et al. 2013). This article adopts oil particle model for numerical simulation of oil spill trajectory and the diffusion range. The model takes consideration of the decaying processes due to heat transform (evaporation), dissolution, and entrainment. The model is based on Lagrangian discrete parcel method, which has the characteristics

of high stability and high efficiency. Oil spill is discretized into oil particles in the model, and each particle represents a certain amount of oil. Oil slick is composed of these large amounts of oil particle. First, changes of each oil particles in position, composition, and water content are calculated. Second, oil particle number and each grid content are calculated for simulating the temporal and spatial distribution of concentration and composition changes of oil slick. Third, the change of oil slick temperature is simulated through the heat balance calculation. Finally, changes in physical and chemical properties of the oil slick are calculated according to the component change and temperature change of oil slick.

4 PARAMETERS OF SIMULATION MODEL

4.1 Oil type and oil quantity

Oil spill risk mainly comes from the ship collision when they are out of port, and the oil type is the fuel oil that ships carry their own. About 100,000 ton cargo ship is selected as the typical ship for simulating the drift trajectory and diffusion area of oil spill. According to the type of ship, it is determined that the single cabin-spilled oil quantity is 500 tons, and the spilled oil leakage is completed within 1 h.

4.2 Oil spill point

The entrance of breakwater is the only way through which the ships of Dongjiakou port must pass. However, the water body of the entrance is relatively small, and with the increase in the number of ships out of the port, the probability of ship collision is the largest at the entrance of breakwater. Therefore, the oil spill point is set at the entrance of breakwater (Fig. 1).

4.3 Setting for the model parameter

The related parameters of model are given in Table 1.

Table 1. The settings for model parameter.

| Coefficient | Process | Value |
|---------------------------|---------------|--------------------|
| Wind friction coefficient | Advection | 0.035 |
| Maximum water content | Emulsion | 0.85 |
| Emulsification constant | Emulsion | 5×10^7 |
| Emulsification constant | Emulsion | 1.2×10^5 |
| Mass transfer coefficient | Dissolution | 2.36×10^6 |
| Emissivities of oil | Heat transfer | 0.82 |
| Emissivities of water | Heat transfer | 0.95 |
| Emissivities of air | Heat transfer | 0.82 |
| Diffusion coefficient | Heat transfer | 0.1 |

4.4 Environmental conditions of the accident

According to the local natural environment condition and the sensitive target position, the relevant environmental conditions are determined (Table 2).

Table 2. Environmental conditions of the accident.

| Environmental conditions | Value |
|----------------------------------|---|
| Wind direction | Static wind, northeast and northwest wind |
| Wind speed | 11.0 m/s |
| Annual average air temperature | 12.2°C |
| Annual average water temperature | 13.3°C |
| Occurring time | High water time |

5 RESULTS

5.1 No wind condition

Oil slick mainly moves toward northeast when it occurs at the high water time. The oil slick can arrive in Yangjiawa Bay after 5 h and arrive in the waters near the Zhaitang Island 8 h later. Oil slick primarily diffuses along the Northeast-Southwest under the action of the flood and ebb flow. After 72 h, the maximum dispersal distance to the northeast of the oil slick is about 23.7 km, to the southwest is about 6.6 km, to the north is about 4.3 km, and to the south is 3.3 km. The trajectories of oil spill are shown in Figures 4 and 5.

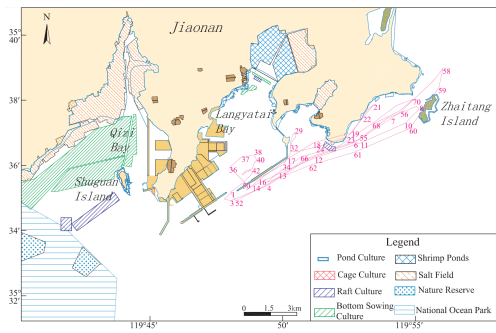


Figure 4. Simulated trajectories of oil spill (no wind condition, 72 h after the oil spill at the high water time, single particle).

5.2 The northwest wind condition

Oil slick mainly moves toward east with the action of NW wind and tidal current when it occurs at the high water time. The oil slick spreads to the southern sea area of Yangjiawa Bay with the distance about 2.0 km away after 4 h. It arrives at the southern sea area of

Zhaitang Island with the distance about 1.3 km away 6 h later. After 7 h, the maximum dispersal distance to the northeast of the oil slick is about 19.2 km and to the south is about 1.6 km. The trajectories of oil spill are shown in Figures 6 and 7.

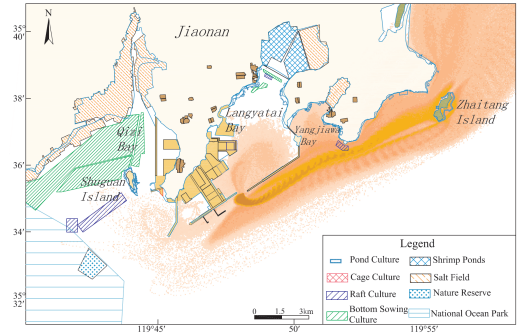


Figure 5. Simulated trajectories of oil spill (no wind condition, 72 h after the oil spill at the high water time, multiparticle).

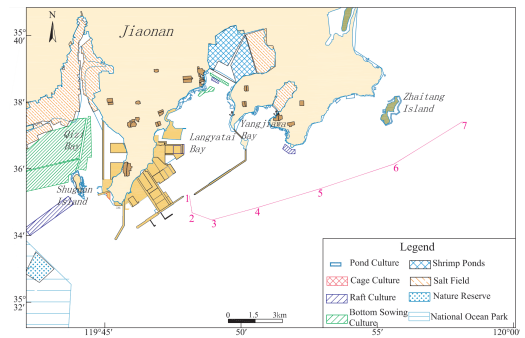


Figure 6. Simulated trajectories of oil spill (northwest wind condition, 7 h after the oil spill at the high water time, single particle).

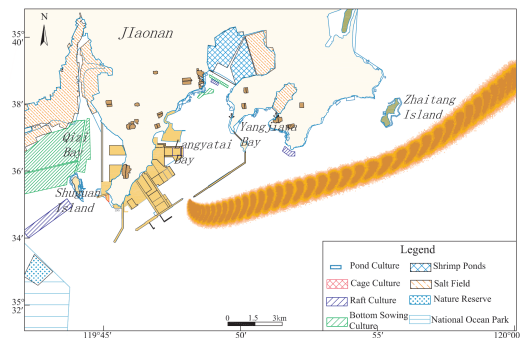


Figure 7. Simulated trajectories of oil spill (northwest wind condition, 7 h after the oil spill at the high water time, multiparticle).

5.3 The northeast wind condition

The oil slick mainly spreads toward the southwest with the action of NE wind and tidal current when it occurs at the high water time. The oil slick arrives in the waters near the ore terminal after 2 h. It arrives at the southern sea area of Zhaitang Island 8 h later and reaches Rizhao National Ocean Park. After 24 h, the maximum dispersal distance to the northeast of the oil slick is about 13.1 km and to the southeast is about 14.7 km. The trajectories of oil spill are shown in Figures 8 and 9.

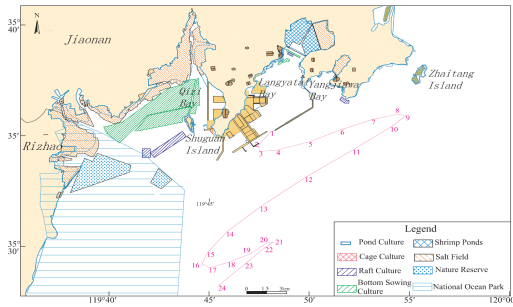


Figure 8. Simulated trajectories of oil spill (northeast wind condition, 24 h after the oil spill at the high water time, single particle).

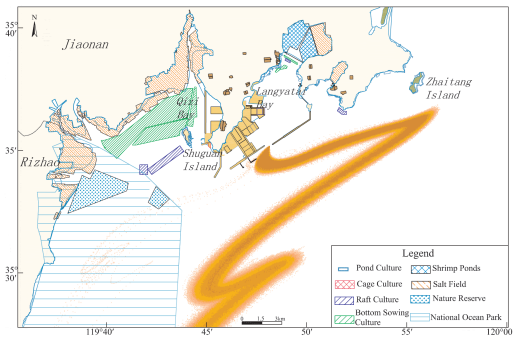


Figure 9. Simulated trajectories of oil spill (northeast wind condition, 24 h after the oil spill at the high water time, multiparticle).

6 CONCLUSIONS

When the oil spill happens at the high water time, the results show that the oil slick mainly moves parallel to the coastline with the interaction of NW

wind and tidal current or the only action of tidal current, while the oil slick mainly spreads toward southwest with the interaction of NE wind and tidal current. The spill trajectory is dominated by winds rather than currents. Once the oil spill occurs in the Dongjiakou port, the oil spill would move quickly, and it could travel up to about 300–3400 m within 1 h. Therefore, the oil spill would bring serious influence to the sensitive objects such as Muguan Island, Nature Reserve, Zhaitang Island, National Ocean Park. At the same time, it will also affect normal operation of the port. Therefore, the development organization should formulate detailed preventive measures to avoid oil spill accident.

REFERENCES

- [1] Andrade, M.M.N., Szlafsztein, C.F., Souza-Filho, P.W.M., Araújo, A.R., Gomes, M.K.T. 2010. A socio-economic and natural vulnerability index for oil spills in an Amazonian harbor: a case study using GIS and remote sensing. *Journal of Environment Management* 91: 1972–1980.
- [2] Xu, H.L., Chen, J.N., Wang, S.D., Liu, Y. 2012. Oil spill forecast model based on uncertainty analysis: A case study of Dalian Oil Spill. *Ocean Engineering* 54:206–212.
- [3] Ren, Y.P., Xu, B.D., Ye, Z.J., Liu, Y.G. 2005. Preliminary Study on Community Structure of Fishery Resources during Spring and Autumn in the Coastal Waters of Qingdao. *Periodical of Ocean University of China* (in Chinese) 35(5): 792–798.
- [4] Sun R.R., Lv Z.F., Wang, X.C. 2012. The Numerical Simulation of Oil-Spill Drifting in the Dongjiakou Port. *Coastal Engineering* (in Chinese) 31(2): 7–13.
- [5] Janeiro, J., Martins, F., Relvas, P. 2012. Towards the development of an operational tool for oil spills management in the Algarve coast. *Journal of Coastal Conservation* 16 (4): 449–460.
- [6] Vethamony P., Sudheesh, K., Babu, M.T., Jayakumar, S., Manimurali, R., Saran, A.K., Sharma, L.H., Rajan, B., Srivastava, M. 2007. Trajectory of an oil spill off Goa, eastern Arabian Sea: Field observations and simulations. *Environmental Pollution* 148:438–444.
- [7] Catarina Frazão Santos, Jaqueline Michel, Mário Neves, João Janeiro, Francisco Andrade, Michael Orbach. 2013. Santos Marine spatial planning and oil spill risk analysis: Finding common grounds. *Marine Pollution Bulletin* 74:73–81.

Statistics and analysis on causes for well blowout accident

W.F. Ge, L.F. Qian & F.B. Song

CNOOC Safety Technology Service Co. Ltd, Tianjin, China

F. Zhang

China National Petroleum Offshore Engineering Co. Ltd, Tianjin Branch, Tianjin, China

H.G. Liu, X. Zhang & G.Y. Zhao

CNOOC Safety Technology Service Co. Ltd, Tianjin, China

ABSTRACT: According to operating procedures, accident levels, and accident causes, this article summarized and analyzed 134 well blowout accidents occurring at home and abroad during 1970–2013. According to frequency of blowout causes, it calculated the contribution rate of different blowout causes using support percentage of correlation analysis methods and compared the importance of various blowout causes. It indicated that well blowout occurred more frequently during drilling operation; once well blowout occurred in offshore platforms, the proportion of serious environmental pollution and casualties was higher than onshore, due to the special environment nature in offshore operations; the contribution rates of management fault and equipment defect were highest for different blowout causes. Based on the contribution rate of blowout causes, it proposed some targeted countermeasures to prevent well blowout. Complying with significance rank of accident prevention could reduce well blowout.

1 INTRODUCTION

Well blowout is a phenomenon that formation fluid ejects out the ground or flows into other formation. With the development of petroleum exploration day by day, the occurrence frequency of well blowout accidents also increases gradually. According to incomplete statistics of 134 well blowout accidents from Petro-China, Sinopec, CNOOC, and foreign companies during 1970–2013, the accidents that have caused serious casualties and environmental pollutions account for 35%, while drilling blowout accounts for 95%.

There have been many researches on blowout risks. The importance of human factors in well blowout accidents was discussed by Zhang et al. 2010^[1]. The latest security requirements from the perspective of American studies and American regulatory classification societies were described by Harish et al. 2011^[2], and the recertification, maintenance, and testing of BOP, in addition to the conjugation between standard-based behavior of offshore drilling operations and current regulatory regime, were discussed. The design, staff, equipment, management, supervision, and other causes for well blowout accidents were described by Jiang (2011)^[3]. He also proposed corresponding countermeasures, without detailing specific reasons and importance of various blowout causes.

For the enormous economic and environmental problems caused by explosion of well blowout accidents, well site measures to prevent explosions were proposed by Zhang et al. 2012^[4]. By taking the well blowout accidents in the Norwegian continental shelf during 2003–2010 as main line and departing from the human factors, technical factors, and organizational factors, the causes of the blowout and possible mitigation measures were investigated by Elisabeth et al. 2013^[5]. By considering the BOP failure, throttle manifold failure, casing failures, and flames after the blowout, the main factors leading to uncontrolled blowout and failure modes were studied by Du et al. 2014^[6], without considering other factors but the equipment factors.

According to different well types, operating procedures, and accident loss levels, this article summarized and analyzed 134 well blowout accidents occurring at home and abroad during 1970–2013. It considered six accident causes that were design defects, defects in process operation, process defects, equipment defects, management deficiencies, and personnel training. According to frequency of blowout causes, it calculated the contribution rate of different blowout causes using support percentage of correlation analysis methods and compared the importance of various blowout causes. Based on contribution rate of blowout causes, it proposed some targeted countermeasures to prevent well blowout.

2 STATISTICS OF WELL BLOWOUT ACCIDENTS

This article selected blowout case, offshore blowouts, and well control and good operating practices^[7-11] as the accident data sources. It summarized 134 well blowout accidents occurring at home and abroad during 1970–2013, which referred to Petro-China, Sinopec, CNOOC, and foreign companies. According to the statistics, different well types and accident loss levels were compared.

2.1 Well blowout accidents for different operation stages

The operation stages that referred to well control safety could be divided into drilling, completion, production operation, work-over operation, and well abandonment during whole well life cycle. According to different operation stages, well blowout accidents were analyzed as shown in Figure 1.

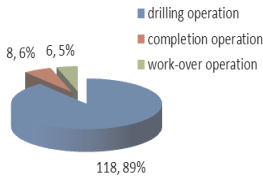


Figure 1. Percentages of well blowout accidents in different operations.

From Figure 1, among 132 well blowout accidents, well blowout occurred more frequently during drilling operation, which had 118 well blowout accidents and accounted for 45%; there were eight well blowout accidents occurring during completion operation that accounted for 6%; there were six well blowout accidents occurring during work-over operation that accounted for 5%. Thus, drilling operation stage was the critical stage for well blowout prevention.

2.2 Well blowout accidents for different accident loss levels

The accident loss levels that were described in this article were mainly divided into three kinds: casualties, environmental pollution, and other losses. Among them, the casualty referred to the well blowout accidents that could cause injuries or death; environmental pollution referred to the well blowout accidents that could cause environmental pollution; the casualties and environmental pollution referred to the well blowout

accidents that could cause both casualty and environmental pollution at the same time; and other losses referred to no casualties and environmental pollution in well blowout accidents.

According to different accident loss levels, well blowout accidents were analyzed, as shown in Figure 2.

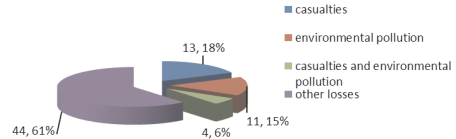


Figure 2. Comparison of accident loss levels.

From Figure 2, serious accidents that could cause casualties and environmental pollution were 35%. Once well blowout occurred in offshore platforms, the proportion of serious environmental pollution and casualties was higher than onshore, due to the special environment nature in offshore operations, as shown in Figure 3.

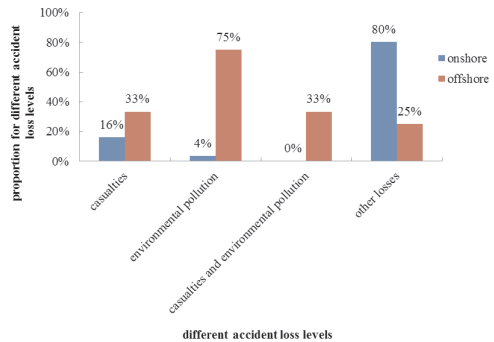


Figure 3. Comparison of accident loss levels between onshore and offshore platforms.

3 ANALYSIS ON WELL BLOWOUT CAUSES

There are many causes that can induce well blowout, such as inaccurate formation pressure, lower mud density, drilling suction, and other improper measures^[12]. Usually, these blowout causes can be divided into human factors, equipment factors, and management factors or be divided into human factors, equipment factors, environmental factors, and management factors in Management Science^[13, 14].

According to statistical situation for well blowout accidents, this article analyzed well blowout causes.

The well blowout causes were divided into six kinds, including design defects, operation defects, process defects, equipment defects, management defects, and personnel training. According to frequency of blowout causes, it calculated the contribution rate of different blowout causes using support percentage of correlation analysis methods and compared the importance of various blowout causes.

3.1 Frequency for different blowout causes

Frequency refers to the times for each object appearance. It is also named as times. This section summarized and analyzed 108 well blowout accidents occurring at home and abroad during 1970–2013 using the conception of frequency. The frequency for different blowout causes and the percentages of each blowout causes were shown in Figure 4.

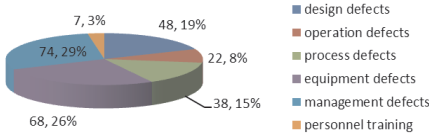


Figure 4. Percentages of blowout causes.

As shown in Table 1 and Figure 4, management defects and equipment defects were the most in statistical blowout causes with the percentage of 28% and 26%. Among them, management defects occurred 74 times and equipment defects occurred 68 times, respectively. Management defects mainly included incomplete installation of well control equipments, unstandard wellhead installation, mismatch of pressure levels, no timely reporting after overflow, violation of operation procedure and well control technique, unstrict well control measures, awareness for well control, and the implemented personal responsibility. Equipment defects mainly included the lower pressure test or no pressure test, no detection for the defective equipment, obsolete equipment, insufficient emergency materials (weighting material, mud), and so on.

3.2 Contribution rate for different blowout causes

To compare the importance for different blowout causes further, according to the frequency that each blowout cause appeared, it calculated the contribution rate of different blowout causes using support percentage of correlation analysis methods.

Correlation rule describes the influence degree for one event's appearance to another event's appearance.

The importance of correlating value could be evaluated quantitatively using support percentage^[15]. It assumes that a complete set $I = \{a, b, c, d, e, f\}$ exists. Among them, a, b, \dots, f were called items. Then, $T_1 = \{a, b, c\}$, $T_2 = \{a, c\}$, ..., $T_5 = \{e, f\}$ were called item sets. The item set refers to a collection that contains a number of items. The collection $D = \{T_1, T_2, \dots, T_5\}$ that contains all item sets is called routine database. The mathematical meaning of support percentage is that assume D as a routine database and X as an item set. Compare numbers of item sets that exist both in D and X with all item sets in D and calculate the proportion, and this proportion is called support percentage. The formula is as follows:

$$Support(X) = \frac{\|\{d \in D | X \subseteq d\}\|}{\|D\|} \quad (1)$$

The level of support percentage represents the representativeness of an item set in all item sets and the higher support percentage for the item set stands for the more importance. The contribution rate of a blowout cause stands for the dependence on this blowout cause if well blowout occurs. The higher contribution rate stands for the higher dependence on this blowout cause, and this blowout cause should be paid more attention to when the accident prevention policy was drafted. Therefore, the support percentage could properly represent the contribution rate of blowout cause.

By making each accident of 108 well blowout accidents as a routine, these 108 routines form a routine database, and the collection formed by each accident cause is called item set. According to the frequency of each accident cause in Table 1, it calculated the support percentage (contribution rate) that caused well blowout as shown in Figure 5.

As shown in Figure 5, support percentages of management defects and equipment defects were the

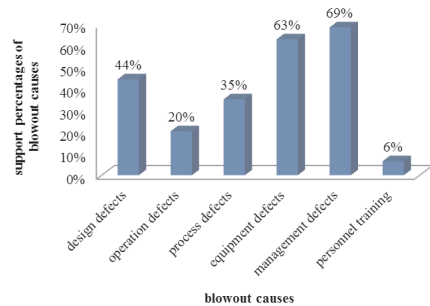


Figure 5. Support percentages of blowout causes.

most in 108 well blowout accidents with the percentage of 69% and 63%, respectively. Thus, we should strengthen design, installation, inspection, maintenance, and supervision of well control equipment and improve well control management system.

4 WELL BLOWOUT PREVENTIVE MEASURES

To summarize and analyze blowout causes, the following well blowout preventive measures were proposed:

- 1 It should comply with the significance rank of accident prevention. More attention paid to the higher contribution rate could improve blowout prevention effect and save security investments.
- 2 Well blowout is not accidental and could be caused when several technical and managerial aspects are out of control. It should strengthen the implementation of management system, criterion, and procedures.
- 3 BOP is the last barrier for ensuring well security. It must be reliable. We should consider the allocation of BOP and its control system. They should have the ability to shut off at a critical moment.
- 4 It is fatal for well control facilities if there are some design defects. Therefore design review is very important. A risk assessment should be conducted for existing well control equipments, and appropriate risk control procedures should be made.
- 5 Control skills of operators are essential, and it should strengthen well control training.
- 6 It should handle the relationship among job schedule, cost, and safety correctly and adhere to the principle "prevention first and safety first." It could not catch up operating schedule blindly to save cost.

5 CONCLUSIONS

- 1 Well blowout occurred more frequently during drilling operation with the percentage of 89% in 134 well blowout accidents when divided by different operation stages.
- 2 Serious accidents that could cause casualties and environmental pollution were 35%. Once well blowout occurred in offshore platforms, the proportion of serious environmental pollution and casualties was higher than onshore, due to the special environment nature in offshore operations.
- 3 Support percentages of management defects and equipment defects were the most with the

percentage of 69% and 63%, respectively. Thus, we should strengthen design, installation, inspection, maintenance, and supervision of well control equipment and improve well control management system.

- 4 Based on contribution rate of blowout causes, it proposed some targeted countermeasures to prevent well blowout. Complying with significance rank of accident prevention could reduce well blowout.

REFERENCES

- [1] Zhang, H. et al. 2010. Feature analysis and countermeasures of human error in drilling blowout incidents. *Natural Gas Industry*, 30(8):98–100,104.
- [2] Harish, P. et al. 2011. Safety enhancement to offshore drilling operations. OTC 22758.
- [3] Jiang, X. B. 2011. Analysis and countermeasures of well control incidents. *Technique Accouter*, (3):34.
- [4] Zhang, X. Q et al. 2012. Study on explosion-proof technology of drilling blowout. *Journal of Safety Science and Technology*, 8(6):129–133.
- [5] Elisabeth, L. et al. 2013. Risk of major accidents: causal factors and improvement measures related to well control in the petroleum industry. SPE 163775.
- [6] Du, G. et al. 2014. Analysis on factors and preventive action for drilling uncontrollable blowout. *Journal of Safety Science and Technology*, 10(2):120–125.
- [7] Xu, Y. Q. 2010. Analysis of causes of blowout accident, simulation of gas diffusion, and lash - up countermeasure research. Qingdao: Qingdao Technological University.
- [8] Blowout Case. 2006. China National Petroleum Corporation Engineering Technology Marketing, Petroleum Technology & Engineering Contractors Association, Beijing: Petroleum Industry Press. 2010. Blowout accidents and dangerous case analysis (Part II), China National Petroleum Corporation Engineering Branch.
- [9] Wang, S. P. & He, B. S. 2011. Offshore blowouts and well control. Beijing: Petroleum Industry Press.
- [10] Good operating practices. 2013. China National Offshore Oil Corporation Health Quality Safety and Environmental Protection Department.
- [11] Zhao, Q. M. 2012. Quantitative risk evaluation of human factors and failures in the gas drilling industry. *Natural Gas Industry*, 32(7):80–83.
- [12] Wang, Y. C. 2010. Safety analysis and evaluation for the process of underbalanced drilling. Qingdao: China University of Petroleum (East China).
- [13] Fu, Z. M. & Chen, Y. G. 2007. Risk analysis on direct property damage for major fire incidents in towns based on association rules. *Science & Technology Review*, 25(15):35–39.
- [14] Hu, K. Y. 2008. Data mining theory and application. Beijing: Tsinghua University Press.

Water quality and growth trend of aquatic plant duckweed: Management and benefits

M.W. Shammout & H. Zakaria

Water, Energy and Environment Center, The University of Jordan, Amman Jordan

ABSTRACT: This paper focuses on the irrigation water quality and on growth trend of aquatic plant so called duckweed (*Lemna sp.*) in farm ponds. It also focuses on duckweed growth management utilizing its benefit at farm level. The results showed that the water quality analysis in mg/l of Ca is between 120-113, Mg 48-40, K 22-18, Cl 349-340, SO₄ 220-188, PO₄ 7.6-6.3, Na 212-205, and NO₃ 38-31. The average protein value in dry matter of duckweed is 26%. At laboratory, duckweed's growth showed an approximately exponential trend that continues until fronds face scarcity of nutrients. This requires the involvement of local farmers towards plant growth management. In this study, 50% of duckweed growth should be harvested every four days to be utilized as future promising feed source for broilers. Further studies will be conducted on the duckweed's benefit values.

1 INTRODUCTION

Duckweed plants (*Lemna sp.*) belong to the Lemnaceae family. They are small free-floating aquatic plants that are often seen growing on a thick blanket, like a mat. They are the smallest and fastest growing plants in the world (Rejmankova 1981), frequently doubling their biomass in 2 or 3 days under optimum conditions of nutrient availability, sunlight and water temperature. Duckweed reproduction is primarily vegetative. An individual leaf may go through 10 divisions over a period of 10 days to several weeks before the original plant senesces. Duckweeds can double their mass in between 16 hours to 2 days under the optimal conditions. This is faster than almost any other higher plant (Leng et al. 1995). It is indicated that a few handfuls of duck weed plant quickly turn into a lot more as long as there are light and nutrients.

Duckweed plants are found worldwide without any agronomic care. Shammout et al. (2008), Shammout & Zakaria (2014) showed that the quality of water that supports duckweed with nutrients is the main issue for the environment that integrates the existence of duckweed plants.

Duckweed plants tolerate a wide range of pH which varies from 3.0 to 10.0 but display an optimum growth in a medium of 5.0-7.0. Doubling of biomass in 2 to 4 days happens at pH levels between 7.0 and 8.0; yet, biomass is inhibited at a pH greater than 10. Optimum temperature for maximum growth ranges from 17.5–34° C, and as light increases growth rate increases. They can survive at water temperatures as low as 5–7° C (Buddhavarapu & Hancock 1991). Although duckweed can survive extreme temperatures,

it is best suited for warm and sunny conditions. Duckweeds can be found on almost any water surface because they are not affected by depth. In winter they sink to the bottom where the warmer water protects them. In addition, Duckweeds are advantageous plants that remove nutrients from water and can dominate over algae (Shammout & Zakaria 2014). They have a very high protein content of up to 50% of its dry mass (Haustein et al. 1988, Skillikorn et al. 1993 & Leng et al. 1995), as well as high levels of vitamins and pigments. Average growth rates tend to decrease from: nutrient scarcity; extreme pH levels; temperature, and crowding by overgrowth of the colony (Skillikorn et al. 1993).

Not to mention, in all over the world and especially in Jordan, the crucial problem with poultry production is the feed cost which has a percentage range of 60–70% from the total cost. Very limited number of feed ingredients is available to choose for the formulation of balanced diet and the chronic scarcity and high cost of animal protein supplements has increased the work to seek alternative protein source for feeding. The non-availability and higher prices of feed ingredients are main constrains of poultry production (Ahhammad et al. 2003). From economical point of view, the poultry should be supplied with cheaper feed to get maximum return with minimum cost.

In Jordan, farmers complain about the duckweed's growth phenomenon in their farm irrigation ponds and they get rid of it. They assume duckweed is a useless plant and are not aware of its importance around the world. Therefore, this paper aims to examine the irrigation water quality, growth trend of duckweed (*Lemna sp.*), and pond management via participatory approach of local farmers based on the benefit of its protein value.

2 METHODOLOGY

Interviews with farmers were conducted to identify the information with regard to the duckweed's growth phenomenon and its related problems on farm ponds.

The water source for irrigation ponds that feeds duckweed with nutrients was analyzed. The analysis was conducted at different intervals to cover wet and dry periods taking into consideration the duckweed vegetation period (spring season). Water samples were analyzed according to the Standard Methods of the American Public Health Association (2012). The parameters of the analysis are Ca, Mg, K, Cl, SO₄, PO₄, Na, NO₃, Zn, Cd, Pb, and Cu.

Fresh duckweed plants were collected carefully from the irrigation pond and transferred to the laboratories of the University of Jordan for several studies on duckweed within this project. From these studies were the growth trend of the duckweed plants and the protein value.

At the laboratory, an experiment was conducted to study the growth characteristics of duckweed plants during the vegetative period. At room temperature (20°C), 10 healthy fronds of *Lemna sp.* were seeded in four dishes containing 50 ml of irrigation water. All the dishes were maintained at 50 ml level by the addition when necessary of irrigation pond water. Duckweed fronds growth was monitored daily for 10-days by counting the number of fronds in the four dishes to calculate the average number of new growing fronds.

Involvement of local farmers to manage the growth of duckweed at farm level.

3 RESULTS AND DISCUSSIONS

3.1 Interviews with Jordanian farmers

Interviews with Jordanian Farmers indicated that duckweed plants have been seen to grow naturally in their farm irrigation ponds. Their explosive reproductive capacity can cause a dense cover of this plant and pose a serious problem to irrigation pipes lines that are pumping irrigation water from irrigation ponds to irrigate the field crops. This problem is representing the closure of the irrigation pipes lines. Farmers used to add herbicides to irrigation ponds to kill duckweed plant but this treatment is useless. Farmers complain about its growth and seek solution to overcome it.

3.2 Water quality and protein value in duckweed dry matter

Table 1 shows the analysis of water quality that feed farm irrigation ponds with nutrients. The results showed that the water quality range values in mg/l of Ca is between 120-113, Mg 48-40, K 22-18, Cl

349-340, SO₄ 220-188, PO₄ 7.6-6.3, Na 212-205, and NO₃ 38-31. As for the heavy metals; Zn, Cd, Pb, and Cu are within Ideal Detection Limits, where average value of Zn is <0.02, Pb is <0.01, Cd is <0.002 and Cu is <0.01. The average protein value in dry matter of duckweed is 26%.

Table 1. Analysis of water quality that feed farm irrigation ponds with nutrients.

| Parameter | Water analysis in mg/l |
|-----------------|------------------------|
| Ca | 120–113 |
| Mg | 48–40 |
| K | 22–18 |
| Cl | 349–340 |
| SO ₄ | 220–188 |
| PO ₄ | 7.6–6.3 |
| Na | 212–205 |
| NO ₃ | 38–31 |
| Zn | <0.02 |
| Pb | <0.01 |
| Cd | <0.002 |
| Cu | <0.01 |

3.3 Duckweed growth trend

Table 2 shows the results of the average growth of duckweed fronds in the four seeded dishes. The average number of duckweed growth varies according to growth days, where 10 fronds produced 6 new fronds within two days and total number of fronds was 23 after four days, with healthy appearance of all the seeded fronds. The fronds number was 56 after 9 days but two fronds out of the 56 appeared yellow. Average healthy duckweed fronds' growth versus the growth days is shown in Figure 1, and Figure 2; a doubling of biomass was obtained in approximately three days. Figure 1 shows the average growth of duckweed (*Lemna sp.*) in irrigation water. Figure 2 shows the growth trend of duckweed (*Lemna sp.*) in irrigation water; approximately an exponential growth. The exponential growth continues until the duckweed fronds face scarcity of nutrients.

Table 2. The average growth of duckweed fronds in the four seeded dishes.

| Date / 2012 | Average No. of duckweed fronds of the four dishes |
|-------------------|---|
| 15 April = 0 day | 10 |
| 17 April = 2 days | 16 |
| 19 April = 4 days | 23 |
| 22 April = 7 days | 38 (2 yellow) = 40 |
| 24 April = 9 days | 54 (2 yellow) = 56 |

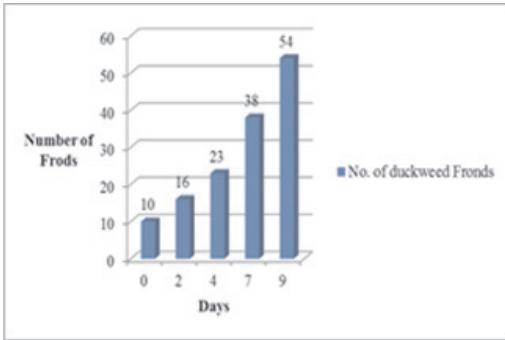


Figure 1. Average growth of duckweed (*Lemna sp.*) in irrigation water.

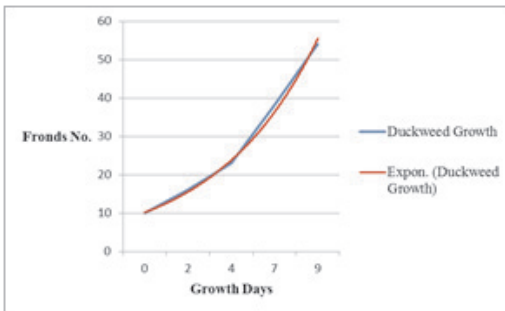


Figure 2. Growth trend of duckweed (*Lemna sp.*) in irrigation water.

4 CONCLUSIONS

In Jordan, the water quality for irrigation farm ponds supports duckweed growth with nutrients to be a highly reproductive plant. Under laboratory conditions, Duckweeds can double their mass in about 3 days. This results in an exponential growth that continues until the duckweed fronds face scarcity of nutrients.

Duckweed growth requires serious management, attention, and frequent harvests on a regular basis throughout the year. This study proposes 50% of

duckweed growth should be harvested every four days by the involvement of local farmers to be utilized as new source of forage. Further studies will be conducted on its benefits to be used as a broiler feeds.

ACKNOWLEDGMENTS

This paper is based on the results of the project that was carried out with financial support from Abdul Hameed Shoman Fund for Supporting Scientific Research. Amman- Jordan.

REFERENCES

- [1] Ahammad, M.U., Swapon, M.S.R., Yeasmi, T., Rahman, Ms. & Ali, M.S. 2003. Replacement of sesame oil cake by duckweed (*Lemna Minor*) in broiler diet. *Pakistan Journal of Biological Sciences* 6(16): 1450–1453.
- [2] American Public Health Association 2012. *Standard Method for the Examination of Water and Wastewater*. 22 ND Edition, New York.
- [3] Buddhavarapu, L.R. & Hancock 1991. Advanced treatment for lagoons using duckweed. *Water Environment and Technology, Water Pollution Control Federation*: 41–44.
- [4] Haustain, A.T., Gilman, R.H., Skillicorn, P.W., Vergara, V., Guevara, V. & Gastanaduy 1988. Duckweed, a useful strategy for feeding chickens: Performance of layers fed with sewage-grown *Lemnaceae* species. *Poultry Sci.* 69:1835–1844.
- [5] Leng, R.A., Stambolie, J.H. & Bell, R. 1995. Duckweed a potential high-protein feed source for domestic animals and fish. *Livestock res. Rural Dev. (LRRD)* 7:1.
- [6] Rejmankova, E. 1981. On the production ecology of duckweeds. Presented at the international workshop on aquatic macrophytes. Illmitz, Austria.
- [7] Shammout, M.W., Oran, S. & Fayyad, M. 2008. The application of duckweed (*Lemna sp.*) in wastewater treatment in Jordan. *International Journal of Environment and Pollution (IJEP)* 33(1):110–120.
- [8] Shammout, M.W. & Zakaria, H. 2014. Wild water lentils plant (Duckweed, *Lemna sp.*) in nutrients removal of Jordan's irrigation ponds. *Biomedical Engineering and Environmental Engineering, WIT Transactions on the Built Environment* 145:525–531.
- [9] Skillicorn, P., Spira, W. & Journey, W. 1993. Duckweed aquaculture-a new aquatic farming system for developing countries. *The World Bank* 76pp. Washington DC.

Estimation of influence of mine waters of cassiterite-sulfide deposits on hydrosphere in Komsomol'sk and Kavaleroovo districts of the far east

V.P. Zvereva

Far East Federal University, Vladivostok, Russian Federation
Far East Geological Institute, FEB of RAS, Vladivostok, Russian Federation

E.A. Yastremskaya & A.D. Pyatakov

Far East Federal University, Vladivostok, Russian Federation

ABSTRACT: The authors estimated the influence of the mine waters of the cassiterite-sulfide deposits on the hydrosphere in Komsomol'sk and Kavaleroovo districts of the Far East using the method of physicochemical modeling. We show that the hypergene processes, proceeding in the deposit ore bodies, result in the formation of the highly concentrated solutions of mine waters, from which a wide set of hypergene minerals are crystallized: chalcantite, wroewolfeite, posnjakite, antlerite, goethite, fibroferrite, scorodite, anglesite, plumbojarosite, bayldonite, and olivenite. We have obtained new data on Eh-pH parameters of the equilibrium solutions in a wide temperature interval from 0 to +45°C, and the quantitative composition of elements in them has been established. Such waters, not purified the year round over decades, enter the surface and ground waters and pollute them.

1 INTRODUCTION

Komsomol'sk and Kavaleroovo districts are rich in the deposits of cassiterite-sulfide ores, which are in Khabarovsk and Primorsky regions in the South Far East of Russia. In Komsomol'sk district, three types of mineralization have been recognized: purely tin ores, copper-tin, tin-polymetal (Zvereva 2008a). In Kavaleroovo district, there are ore-bearing zones: western, eastern, and latitudinal and are conjugated tin-bearing deposits rich in sulfide ores (Zvereva 2008a). The deposits were developed by both open (quarries) and closed (adits) ways to a depth of 200–300 m from the surface. Mining industry here has been developed for more than 70 years. In 2001, the mining industry stopped its activity, although the ore reserves were not worked out. Reclamation measures are not carried out in the districts. While the ore bodies, composed of sulfides or enclosing them, are in the natural setting, and they are practically not changed, but the deposit exploitation with the open (quarry) or closed (adits) ways favors the access of the weathering agents (water, oxygen, carbon dioxide, and others) to them. As a result, the hypergene processes in the mine workings are activated, and the system passes to the technogene stage contaminating the hydrosphere. The processes of the present technogene mineral formation are observed ubiquitously on the walls and roof of the mine workings. They are represented by the sintered and stalactite-stalagmite formations of

dark-blue, blue, green, yellow, and brown color of different tinges about 0.5 m thick (Yakhontova et al. 1981, 1984, Postnikova & Yakhontova 1984, Zvereva 2008a, b). This work is aimed at the modeling of the chemical composition of interpore solutions of the being oxidized sulfide ores in a wide temperature interval from 0 to +45°C that is together with the surface and ground waters form the mine waters. To perform the modeling, it is necessary to solve the following tasks: to determine the composition of possible hypergene and technogene mineral phases crystallized from the highly concentrated solutions, their ion composition, and Eh-pH parameters.

The modeling is carried out with the account for the database including the mineral composition of ores of each district. Oxidation of minerals is studied in their isolated state and in the contact with each other with regard to the most common sulfides (Postnikova & Yakhontova 1984, Zvereva 2008a).

2 METHODS OF INVESTIGATION

The modeling is carried out with the help of the «Selektor» program complex (developers: I.K. Karpov, K.V. Chudnenko, V.A. Bychinsky) allowing the determination of the element volumes to evacuate into the hydrosphere from the mine workings. The method is based on the minimization of Gibbs energy and is a universal program device whose list of the

potential phases is mostly completed compared with others. It corresponds to the requirements imposed upon the programs of calculations of chemical equilibriums. The authors first carry out the investigation of the mine water formation in the districts under consideration.

We choose for the modeling the unified thermo-barometric conditions – T of 0 to +45°C and $P = 1$ atm. Water–rock ratio is taken as 10:1 (100 g of tailings per 1 kg of water). The rain water composition is taken from the literature (Karpov 1981). The models are open to the atmosphere. The chemical composition is calculated after R. Horne (Horne 1972). The composition of 10 kg of the atmosphere includes (in moles): Ar – 3.2, C – 0.10, N – 539.48, O – 144.85. In the modeling calculations in accordance with each variant of the system, we took account of 8–12 independent components (Ar, C, N, Fe, S, As, Cu, Pb, Zn, Ag, Sb, H, O, e), 120–285 dependent components, 90–222 of which are the dissolved particles, 18 gases (Ar, H₃N, CO₂, CO, C₂H₆, H₂, H₂S, CH₄, NO, N₂, NO₂, N₂O, O₂, C₃H₈, S₂, O₂S, O₃S, H₂O), and 3–44 hypogene and hypergene minerals.

The following original minerals from the sulfide class are introduced into the models being created: sphalerite, galena, pyrite, pyrrhotite, chalcocopyrite, and arsenopyrite, and those of the second sulfide concentration: chalcocine, covellite, and bornite, as the ores are stripped to a very big depth. In the case of the sulfide oxidation in contact with each other, the original minerals are used as the base, and chalcocine, covellite, and bornite each comprised 5%. Sulfide ratio in the models is determined on the basis of literature data (Yakhontova et al. 1981, 1984).

3 RESULTS AND DISCUSSION

Sulfide ores can be both monomineral and polymineral. At first, we will consider the results of the modeling of the monomineral sulfide ore oxidation when minerals occur in the isolated state (Table 1).

The summary of ion composition of the modeled solutions is O₂⁰, H⁺, H₂O, HSO₄⁻, SO₄²⁻, N₂⁰, NO₂⁻, NO₃⁻, HNO₂⁰, CO₂⁰, HCO₃⁻, As⁵⁺, H₂AsO₄⁻, H₃AsO₄⁰, HAsO₄²⁻, Cu²⁺, CuHCO₃⁺, CuO⁰, CuOH⁺, CuSO₄⁰, CuCO₃⁰, CuO₂²⁻, Fe²⁺, Fe³⁺, FeO⁺, FeOH²⁺, Pb(SO₄)₂²⁻, Pb²⁺, PbHCO₃⁺, PbNO₃⁺, PbOH⁺, PbSO₄⁰, Zn²⁺, ZnHCO₃⁺, ZnOH⁺, ZnSO₄⁰, and Zn(SO₄)₂²⁻, and HFeO₂⁰, HNO₃⁰, and As³⁺ are found in small concentrations.

The chalcocine oxidation in both districts results in the crystallization of the micropore solutions of wroewolfeite Cu₄[SO₄](OH)₆·2H₂O (to 30°C), posnjakite Cu₄[SO₄](OH)₆·H₂O (only at 30°C), antlerite Cu₃[SO₄](OH)₄ (35–45°C), and chalcocanthite Cu[SO₄]·5H₂O, and in the case of the bornite presence in the model, goethite (Komsomolsk ores) and fibroferrite

Fe⁺³[SO₄](OH)·5H₂O (Kavalerovo ores) are added to these minerals. Both systems being modeled have the Eh–pH parameters of 1.04 V and 3.3, respectively. The covellite oxidation results in chalcocanthite precipitation in the systems of both districts. The parameters of the solutions are 1.13 V and 1.62–1.64, and their mineralization is 64.7–224 g/l. In the oxidation of chalcocine, covellite, and bornite in contact with each other, the solution parameters and the precipitated mineral composition are close to those of the previous modeled systems (singly for chalcocite and bornite). The element concentration in the solution is 65.2–187 g/l.

Now we will discuss the results obtained in the models of oxidation of hypogene sulfide minerals: pyrite, pyrrhotite, chalcocopyrite, arsenopyrite, galena, and sphalerite in the isolated state (Table 1). In the systems being modeled, the Eh–pH parameters are within 0.92–1.18 V and 0.6–5.6. The sphalerite oxidation at all deposits generates the highly concentrated solutions, but technogene minerals are not crystallized from them, and zinc and sulfur remain in full in the solution. Their mineralization is persistently high and is 172–455 g/l. The galena disintegration results in the precipitation of zinc sulfate—anglesite PbSO₄ in amounts of 114.5–136.3, and the solution mineralization is 1.45–14.9 g/l. Pyrite and pyrrhotite in the oxidation process promote the formation of fibroferrite from 184 to 215 g (Kavalerovo ores) or goethite a-FeOK-OH from 75 to 294 g (Komsomol'sk ores). Mineralization of the solutions is 94–260 g/l and 0.01–129 g/l (pyrite and pyrrhotite, correspondingly). The chalcocopyrite oxidation shows for both districts that from the modeled solutions, chalcocanthite is added to goethite and fibroferrite. In the process of chalcocopyrite oxidation, the mineralization is 72 to 191 g/l. The arsenopyrite oxidation generates scorodite Fe⁺³[AsO₄]·2H₂O (Komsomol'sk ores) and fibroferrite (Kavalerovo ores). In this case, the solution mineralization changes from 64.5 to 298 g/l. Consequently, in the process of the hypogene sulfide oxidation at the deposits of Komsomol'sk district, from the concentrated solutions of mine waters, the iron hydroxide—goethite—is crystallized, and from Kavalerovo ores, the iron sulfate—fibroferrite—is precipitated. It should be noted, that in the first variant, the pore waters practically do not contain any iron, and in the second variant, the solution is enriched in it and contains about 260 g/l.

As in the veins of sulfide ores, the simultaneous presence of all primary minerals is not necessary, and then in subsequent models, we consider different possible natural combinations, that is, the minerals are removed from the system alternately (Table 1). In Komsomol'sk district, the joint oxidation of the most common primary sulfides (sphalerite, galena, pyrite, pyrrhotite, chalcocopyrite, and arsenopyrite as well as with addition of chalcocine, bornite, or covellite) results in precipitation from the

solution of the following minerals (g): plumbojarosite $\text{PbFe}_6^{+3}[\text{SO}_4]_4(\text{OH})_{13}$ (47–71), scorodite (11–21), and goethite (1–14). The system parameters are within 0.76–1.17 and 1.15–1.18 V, and the solution mineralization changes from 104 to 131 g/l. Then chalcocine, covellite, and bornite together are introduced into the system. At the next stage, different sulfides (chalcopyrite, pyrite, pyrrhotite, and sphalerite) are removed one by one from the system. In all these variants at the deposits of Komsomol'sk district, the following minerals are formed from the solutions (g): chalcantite (0.02–19 at temperature from 0 to 10 °C), plumbojarosite (47), scorodite (12–14), and goethite (1–13). The parameters of the modeled solution are 0.87–1.29 and 1.15–1.17 V, and their mineralization is 110–130 g/l. When only arsenopyrite is removed from the model, then the oxidation of the listed above sulfides together with all minerals of the second sulfide concentration allows the following minerals to be crystallized from the solutions (g): chalcantite (0.03–17, at 0–10°C), plumbojarosite (47), and goethite (6–15). The parameters of these systems are within 1.2 V and 1.15, and the solution mineralization changes from 123 to 135 g/l. When removing galena from the model, we obtain the solution allowing the following minerals to be formed (g): chalcantite (2–19, at 0°C to 10°C), scorodite (10–14), and goethite (24–36). Its parameters are 1.15 and 1.12 V.

Now we will consider the same systems of the sulfide oxidation models for the deposits of

Kavalerovo district (Table 1). In the case of the oxidation of all hypogene sulfides together with all minerals of the second sulfide concentration and with each singly, as well as removing chalcopyrite from the system, only anglesite (19 g) and fibroferrite (95–144) are precipitated. All modeled solutions have the Eh–pH parameters of 1.16–1.18 V and 0.76–1.24. The solution mineralization changes from 85 to 293 g/l.

Removal of arsenopyrite, pyrite, and sphalerite one by one from the association of sulfides being oxidized results in the crystallization of anglesite (19 g), fibroferrite (74–117), and chalcantite (0.02–7g at 0–15°C). When galena is absent in the system, only fibroferrite (117 g) and chalcantite (to 0.15 g at 0°C–5°C) are precipitated. If the pyrrhotite is removed from the model, then the following hypogene minerals are formed: anglesite (17–19 g), fibroferrite (84 g), chalcantite (0.001 g), bayldonite $\text{PbFe}_3^{+3}[\text{AsO}_4][\text{SO}_4](\text{OH})$ (4.4–4.7 g at 0–25°C), and olivenite $\text{Cu}_3[\text{AsO}_4](\text{OH})$ (3.8–5.2 g at 25–45°C C). The parameters (Eh–pH) of all models under consideration are 1.1–1.16 V and 1.23–2.23, and the solution mineralization is 88 to 337 g/l.

The presence of all hypogene minerals, precipitated in the modeled systems (Table 1), is verified by the literature data. They are found on the walls and roof of mine workings in the districts under consideration and in other mining-ore districts (Yakhontova et al. 1981, 1984, Postnikova & Yakhontova 1984).

Table 1. Oxidation of Komsomol'sk and Kavalerovo sulfide ores.

| # | Hypogene sulfide minerals (%) | Komsomol'sk district | | | Kavalerovo district | | |
|----|---|--|--------|--------|--|--------|--------|
| | | Technogene minerals | pH | Eh | Technogene minerals | pH | Eh |
| 1 | Chalcocine Cu_2S (100) | Chalcantite Wroewolfeite Posnjakite Antlerite | 3.3019 | 1.0435 | Chalcantite Wroewolfeite Posnjakite Antlerite | 3.3045 | 1.0433 |
| 2 | Bornite Cu_3FeS_4 (100) | Chalcantite Goethite Wroewolfeite Posnjakite Antlerite | 3.2987 | 1.0436 | Chalcantite Fibroferrite Wroewolfeite Posnjakite Antlerite | 3.3041 | 1.0433 |
| 3 | Covellite CuS (100) | Chalcantite | 1.6398 | 1.1334 | Chalcantite | 1.6249 | 1.1342 |
| 4 | Chalcocine, bornite, covellite | Chalcantite Goethite Wroewolfeite Posnjakite Antlerite | 3.2992 | 1.0436 | Chalcantite Fibroferrite Wroewolfeite Posnjakite Antlerite | 3.3030 | 1.0433 |
| 5 | Sphalerite ZnS (100) | — | 1.3475 | 1.1495 | — | 1.3309 | 1.1504 |
| 6 | Galena PbS (100) | Anglesite | 1.6999 | 1.1299 | Anglesite | 1.6982 | 1.1299 |
| 7 | Pyrite FeS_2 (100) | Goethite | 0.7877 | 1.1807 | Fibroferrite | 0.3197 | 1.2055 |
| 8 | Pyrrhotite FeS (100) | Goethite | 0.9689 | 1.1704 | Fibroferrite | 5.5990 | 0.9188 |
| 9 | Chalcopyrite CuFeS_2 (100) | Goethite Chalcantite | 1.1788 | 1.1587 | Chalcantite Fibroferrite | 1.6183 | 1.1345 |
| 10 | Arsenopyrite FeAsS (100) | Scorodite | 0.5823 | 1.1910 | Fibroferrite | 1.2034 | 1.1570 |

(continued)

Table 1. Oxidation of Komsomol'sk and Kavalerovo sulfide ores. (continued)

| # | Hypogene sulfide minerals (%) | Komsomol'sk district | | | Kavalerovo district | | |
|----|--|---|--------|--------|--|--------|--------|
| | | Technogene minerals | pH | Eh | Technogene minerals | pH | Eh |
| 11 | Sphalerite (15), galena (15), pyrite (20), pyrrhotite (20), chalcopyrite (15), arsenopyrite (15) | Plumbojarosite Scorodite Goethite | 0.7614 | 1.1813 | Anglesite Fibroferrite | 0.7615 | 1.1810 |
| 12 | Sphalerite (10), galena (10), pyrite (20), pyrrhotite (20), chalcopyrite (15), arsenopyrite (10), covellite (5), bornite (5), chalcocine (5) | Chalcanthite Plumbojarosite Scorodite Goethite | 1.0804 | 1.1640 | Anglesite Fibroferrite | 0.9896 | 1.1687 |
| 13 | Sphalerite (10), galena (10), pyrite (20), pyrrhotite (20), chalcopyrite (15), arsenopyrite (15), covellite (10) | Plumbojarosite Scorodite Goethite | 1.1549 | 1.1600 | Anglesite Fibroferrite | 0.8621 | 1.1756 |
| 14 | Sphalerite (10), galena (10), pyrite (20), pyrrhotite (20), chalcopyrite (15), arsenopyrite (15), bornite (10) | Plumbojarosite Scorodite Goethite | 1.1704 | 1.1591 | Anglesite Fibroferrite | 0.9454 | 1.1711 |
| 15 | Sphalerite (10), galena (10), pyrite (20), pyrrhotite (20), chalcopyrite (15), arsenopyrite (15), chalcocine (10) | Chalcanthite Plumbojarosite Scorodite Goethite | 1.2077 | 1.1571 | Anglesite Fibroferrite | 1.0163 | 1.1672 |
| 16 | Sphalerite (15), galena (15), pyrite (20), pyrrhotite (20), chalcopyrite (15), covellite (5), bornite (5), chalcocine (5) | Chalcanthite Plumbojarosite Goethite | 1.2130 | 1.1568 | Anglesite Fibroferrite Chalcanthite | 1.2392 | 1.1552 |
| 17 | Sphalerite (15), galena (15), pyrite (20), pyrrhotite (20), arsenopyrite (15), covellite (5), bornite (5), chalcocine (5) | Chalcanthite Plumbojarosite Scorodite Goethite | 1.2131 | 1.1568 | Anglesite Fibroferrite | 1.2079 | 1.1569 |
| 18 | Sphalerite (15), galena (15), pyrite (20), chalcopyrite (20), arsenopyrite (15), covellite (5), bornite (5), chalcocine (5) | Chalcanthite Plumbojarosite Scorodite Goethite | 0.8744 | 1.1752 | Anglesite Fibroferrite Chalcanthite Bayldonite olivinite | 2.2339 | 1.1013 |
| 19 | Sphalerite (15), galena (15), pyrrhotite (20), chalcopyrite (20), arsenopyrite (15), covellite (5), bornite (5), chalcocine (5) | Chalcanthite Plumbojarosite Scorodite Goethite | 1.2943 | 1.1523 | Anglesite Fibroferrite Chalcanthite | 1.3450 | 1.1495 |
| 20 | Sphalerite (15), pyrite (20), pyrrhotite (20), chalcopyrite (15), arsenopyrite (15), covellite (5), bornite (5), chalcocine (5) | Chalcanthite Scorodite Goethite | 1.1346 | 1.1612 | Fibroferrite Chalcanthite | 1.2337 | 1.1555 |
| 21 | Galena (15), pyrite (20), pyrrhotite (20), chalcopyrite (15), arsenopyrite (15), covellite (5), bornite (5), chalcocine (5) | Chalcanthite Plumbojarosite Scorodite Goethite | 1.2147 | 1.1567 | Anglesite Fibroferrite Chalcanthite | 1.3265 | 1.1504 |

Note: "—" means uncrystallized mineral.

4 CONCLUSION

The results obtained allow consideration of the formation process of the highly concentrated mine waters, which before and after the precipitation of hypergene and technogene minerals from them enter the surface (river) and ground waters. The physicochemical modeling performed by the authors demonstrates the possibility of the iron and lead precipitation in the form of sulfates—fibroferrite, anglesite, and plumbojarosite in Kavalerovo district or in the form of the iron mineral—goethite of the oxide and hydroxide class in Komsomol'sk district. The presence and oxidation of the minerals of the second sulfide concentration (chalcocite, bornite, and covellite) result in crystallization of abundant sulfates of copper: chalcantite, wroewolfeite, posnjakite, and antlerite in both districts. Through the sphalerite oxidation, zinc enters into the solution. Disintegration of arsenopyrite proper and together with other sulfides allows the precipitation of hypergene minerals of iron, copper, and lead: scorodite in Komsomol'sk district and bayldonite and olivenite in Kavalerovo district. The solution mineralization is high and reaches 445 g/l, but it should be taken into account the technogene minerals are already precipitated from them. Before the moment of crystallization of hypergene and technogene minerals and after this process, all mine waters run and enter the river waters.

In the districts under consideration, from the Silinka River (Komsomol'sk district) and the Vysokogorka River (Kavalerovo district), the drink water is taken. At the place where the mine waters enter the Silinka River, the river water pH decreases to 3.5.

REFERENCES

- [1] Horne, R. 1972. *Marine chemistry*. Moscow: Mir.
- [2] Karpov, I.K. 1981. *Physicochemical modeling with electronic computers in geochemistry*: 372–383. Novosibirsk: Nauka.
- [3] Postnikova, V.P. & Yakhontova, L.K. 1984. *Mineralogy of hypergenesis zone of tin deposits in Komsomol'sk district*. Vladivostok: FESC of the USSR Acad. Sci.
- [4] Yakhontova, L.K., Postnikova, V.P., Vlasova, E.V., Sergeeva, N.E. 1981. New data on posnjakite, serpierite, and woodwardite. *Reports of the USSR Acad. Sci.* 256(5): 1221–1226.
- [5] Yakhontova, L.K., Postnikova, V.P., Sergeeva, N.E. & Yudin, R.N. 1984. New data on rare sulfate of copper. *Bulletin of Moscow State University*(4) 3: 41–46.
- [6] Zvereva, V.P. 2008a. *Ecological consequences of hypergene processes at tin-sulfide deposits of the Far East*. Vladivostok: Dal'nauka.
- [7] Zvereva, V.P. 2008b. Present-day mineral formation in hypergenesis zone of the exploited tin deposits of the Far East. *Transactions of Mineralogical Society* 6: 28–37.

Effect of sodium sulfite and humic acid on Cr(VI) content in artificial wastewater with or without Cr-hyperaccumulator *Leersia hexandra* Swartz

Z.G. Cheng, X.F. Zhang & C. Tian

Guangxi Scientific Experiment Center of Mining, Metallurgy and Environment, Guilin University of Technology, Guilin, PR China

College of Environmental Science and Technology, Guangxi Key Laboratory of Environmental Pollution, PR China

ABSTRACT: A hydroponic experiment was conducted to study the effect of sodium sulfite and humic acid on the Cr(VI) concentration of the nutrient solution. For the given treatment, in the group without *L. hexandra*, the concentration of Cr(VI) in the solution was invariant with the increasing time in the control treatment, while increased with the increasing time in sodium sulfite and humic acid treatments. In the group with *L. hexandra*, the concentration of Cr(VI) in the solution decreased with the increasing time in the control, sodium sulfite, and humic acid treatments. For the same period, in the group without *L. hexandra*, sodium sulfite and humic acid significantly decreased the concentration of Cr(VI) in the solution compared with the control. In the group with *L. hexandra*, sodium sulfite and humic acid had no significant effect on the concentration of Cr(VI) in the solution compared with the control.

KEYWORDS: Cr(VI), sodium sulfite, humic acid, *L. hexandra*

1 INTRODUCTION

Chromium, a kind of heavy metals in water and soil, mainly exists as Cr(III) and Cr(VI). Compared with Cr(III), Cr(VI) has a stronger toxicity and can cause human skin acute allergic reaction [1]. Sodium sulfite and humic acid has a strong reduction, but the governance effect of sodium sulfite and humic acid on chromium-polluted water are still to be studied. *Leersia hexandra* Swartz was a Cr-hyperaccumulator [2]. The plant exists or not may affect the oxidation–reduction of Cr(VI) in the solution. Therefore, the present study was to investigate the effect of sodium sulfite and humic acid on Cr(VI) content in artificial wastewater with or without Cr-hyperaccumulator *Leersia hexandra* Swartz.

2 MATERIALS AND METHODS

The experiment with 14 treatments and 3 replicates was conducted in greenhouse of Guilin University of Technology (Tables 1 and 2). *L. hexandra*, collected from Guilin suburb, was washed with deionized water, preincubation for 7 days, Then 20 seedlings (20 cm height) of *L. hexandra* were transplanted into each 1.5-L plastic barrels. Each plastic barrel has 1 L of 50% Hoagland nutrition liquid. Chromium

was added to the solution as potassium dichromate ($K_2Cr_2O_7$). Electric aeration (ACO-001) was used through the experiment. About 2 ml of water samples in each barrel was corrected at 1h, 4h, 8h, 1d, 2d, and 3d. The pH of water samples were used nitrate to regulate to 1–2, and kept in 4°C. Cr(VI) content was determined by diphenylcarbohydrazide spectrophotometric method [3,4].

The experimental data were analyzed using SPSS 18 software by one-way ANOVA using the least significant difference (LSD) to compare significant differences between treatments ($P < 0.05$).

Table 1. Experimental settings with *L. hexandra* (mg/kg).

| Treatments | Concentration of $K_2Cr_2O_7$ | Concentration of Na_2SO_3 | Concentration of humic acid |
|------------|-------------------------------|-----------------------------|-----------------------------|
| CK1 | 20 | | |
| S1 | 20 | 10 | |
| S2 | 20 | 50 | |
| S3 | 20 | 100 | |
| HA1 | 20 | | 5 |
| HA2 | 20 | | 10 |
| HA3 | 20 | | 50 |

Table 2. Experimental settings without *L. hexandra* (mg/kg).

| Treatments | Concentration of K ₂ Cr ₂ O ₇ | Concentration of Na ₂ SO ₃ | Concentration of humic acid |
|------------|--|--|-----------------------------|
| CK2 | 20 | | |
| S11 | 20 | 10 | |
| S21 | 20 | 50 | |
| S31 | 20 | 100 | |
| HA11 | 20 | | 5 |
| HA21 | 20 | | 10 |
| HA31 | 20 | | 50 |

3 RESULTS AND DISCUSSION

Effects of different treatments on Cr(VI) concentration were shown in Tables 3 and 4. For the given treatment, in the group without *L. hexandra*, the concentration of Cr(VI) in the solution was invariant with the increasing time in the control treatment (CK2), while increased with the increasing time in sodium sulfite and humic acid treatments. In the group with *L. hexandra*, the concentration of Cr(VI) in the solution decreased with the increasing time in the control, sodium sulfite, and humic acid treatments.

Table 3. Cr(VI) concentrations of solution in different treatments 1 h, 2 h, and 3 h (mg/kg).

| Treatment | 1 h | 4 h | 8 h |
|-----------|----------------|---------------|----------------|
| CK1 | 6.48 ± 0.093d* | 6.10 ± 0.093d | 5.73 ± 0.582d |
| CK2 | 16.78 ± 0.07a | 15.44 ± 0.1a | 16.4 ± 0.2a |
| S11 | 6.78 ± 0.072d | 7.25 ± 0.1c | 6.81 ± 0.017b |
| S21 | 5.82 ± 0.056e | 6.88 ± 0.22d | 6.96 ± 0.7b |
| S31 | 4.93 ± 0.172f | 5.93 ± 0.064e | 7.31 ± 0.49b |
| HA11 | 7.33 ± 0.25c | 7.7 ± 0.031b | 7.56 ± 0.082b |
| HA21 | 7.23 ± 0.09c | 7.62 ± 0.3b | 7.53 ± 0.59b |
| HA31 | 6.22 ± 0.03e | 6.92 ± 0.05c | 7.61 ± 0.047b |
| S1 | 6.48 ± 0.093d | 6.05 ± 0.161d | 6.26 ± 0.372c |
| S2 | 6.48 ± 0.093d | 6.32 ± 0.186d | 6.32 ± 0.186c |
| S3 | 5.99 ± 0.246e | 6.64 ± 2.0d | 7.93 ± 0.406c |
| HA1 | 7.66 ± 0.559b | 7.18 ± 0.582b | 7.28 ± 0.246bc |
| HA2 | 6.26 ± 0.652d | 4.97 ± 1.343f | 6.10 ± 0.493c |
| HA3 | 6.64 ± 0.093c | 6.69 ± 0.322d | 6.86 ± 0.888c |

*Mean ± SE; “—” means no data; data followed by the same letters in the same column for the same metal indicate no significant difference at $P = 0.05$ level by LSD test.

Table 4. Cr(VI) concentrations of solution in different treatments at 1, 2, and 3 days (mg/kg).

| Treatment | 1 d | 2 d | 3 d |
|-----------|-----------------|---------------|---------------|
| CK1 | 5.89 ± 0.427cd* | 5.67 ± 0.246e | 5.63 ± 0.285c |
| CK2 | 15.9 ± 0.0982a | 16.8 ± 0.074a | 16.6 ± 0.1a |
| S11 | 7.24 ± 0.066c | 7.49 ± 0.09c | 7.8 ± 0.08b |
| S21 | 7.28 ± 0.062c | 7.56 ± 0.124c | 7.95 ± 0.23b |
| S31 | 7.51 ± 0.153b | 7.84 ± 0.069b | 8.2 ± 0.35b |
| HA11 | 7.71 ± 0.049b | 7.96 ± 0.6b | 8.3 ± 0.09b |
| HA21 | 7.43 ± 0.081b | 7.96 ± 0.5b | 8.35 ± 0.6b |
| HA31 | 7.0 ± 0.09c | 7.16 ± 0.07c | 7.6 ± 0.23b |
| S1 | 6.21 ± 0.427cd | 5.4 ± 0.279e | 5.46 ± 0.189c |
| S2 | 5.78 ± 0.745cd | 5.56 ± 0.161e | 5.95 ± 0.285c |
| S3 | 5.67 ± 0.372cd | 5.24 ± 0.582c | 5.21 ± 0.071c |
| HA1 | 6.75 ± 0.466c | 6.59 ± 0.493d | 7.31 ± 1.285b |
| HA2 | 5.35 ± 0.336d | 5.3 ± 0.372e | 5.42 ± 0.428c |
| HA3 | 4.81 ± 1.655c | 5.62 ± 0.406e | 6.08 ± 0.433c |

*Mean ± SE; data followed by the same letters in the same column for the same metal indicate no significant difference at $P = 0.05$ level by LSD test.

For the same period, in the group without *L. hexandra*, sodium sulfite, and humic acid significantly decreased the concentration of Cr(VI) in the solution compared with CK2. This results indicated that sodium sulfite and humic acid showed reduction, which were in consistent with the previous studies [5, 6]. In the group with *L. hexandra*, sodium sulfite and humic acid had no significant effect on the concentration of Cr(VI) in the solution compared with the control (CK1). There may be two reasons for this result. Sodium sulfite and humic acid may show no reduction in this condition, and the other may due to *L. hexandra*, which can regulate the oxidation–reduction of Cr(VI) according to the absorption and release of Cr in the solution by plant.

ACKNOWLEDGMENTS

The project was supported by National Natural Science Foundation of China (41273142), Natural Science Foundation of Guangxi (2013GXNSFBA019026), and Scientific Research Foundation of Guilin University of Technology.

REFERENCES

- [1] S CG. Chromium. Encyclopedia of Toxicology (Third Edition), 2014, Pages 952–954.
- [2] Zhang XH, Liu J, Huang HT, et al. Chromium accumulation by the hyperaccumulator plant *Leersia hexandra* Swartz. *Chemosphere*, 2007, 27: 1138–1143.
- [3] Ministry of Environmental Protection of People's Republic of China (MEP). 2002. Monitoring and analysis methods for water and wastewater (4th Ed). Environmental Science Press, 346–349. Beijing:China.
- [4] Ministry of Environmental Protection of People's Republic of China (MEP). Water quality-Determination of hexavalent chromium GB7466–87. Beijing:China, 1987.
- [5] Jing GW, Jie MY, Xiang ZK et al. Spectrophotometric determination of EDTA in aqueous solution through ferriox formation using sodium sulfite as the reducer. *Chemosphere*, 2013, 91:351–357.
- [6] Barbara S, Fulvia T, Fabrizio A. Cr(VI) reduction capability of humic acid extracted from the organic component of municipal solid waste. *Journal of Environmental Sciences*, 2013, 25:Pages 487–494.

Low cost adsorbent to reduce disinfection by-products from drinking water in small communities

T. Husain

Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL, Canada

M. Ahmad

Muskeg River Mine, Shell Canada Energy, Fort McMurry, Alberta, Al, Canada

ABSTRACT: Although there are different types of treatment methods such as membrane filtration, coagulation, reverse osmosis and other adsorption processes available to reduce trihalomethanes (THMs) and haloacetic acid (HAAs), and the two major groups of disinfection by-products (DBPs) in the drinking water systems, most of these methods require trained personnel to operate and maintain the system and high capital investment and operating cost. Due to lack of skilled personnel to operate and maintain water systems and also high capital cost, it is not feasible to install such systems in small communities. The main objective of this study is to identify a cost effective adsorbent to remove DBPs and their precursors. The adsorbent developed from oil fly ash residues is available in abundance at very low cost. The adsorbent to be developed by extracting clean carbon can be easily installed and operated in small systems. In this research, the extracted adsorbent is tested with a series of experiments using raw and treated water from the Pouch Cove community near St. John's. The results show a significant removal of DBPs from tap water and precursors such as natural organic matters from intake source causing formation of DBPs after chlorination. The extracted carbon from oil fly ash shows significant removal of THMs and HAAs from drinking water and from the Pouch Cove drinking water supply system. The clean carbon is also used to remove natural organic matter and turbidity from raw water before chlorination.

1 INTRODUCTION

Water utility managers in small communities are facing numerous challenges to protect drinking water supplies from water-borne diseases. To supply safe drinking water, chlorine is the most common oxidant and disinfectant used to eradicate and inactivate pathogens. Chlorine also acts as an oxidant in water treatment in oxidizing iron and manganese. It also removes taste and color and prevents biological re-growth in the water distribution systems. Due to these characteristics and relatively lowcost, more than 90% of the world water supply systems use chlorine. This disinfectant in residual form, however, reacts with the natural organic matter (NOM), forming disinfection by products (DBPs) in the water supply systems.

NOM consists of numerous organic materials such as proteins, humic substances, hydrophilic acid, lipids, carbonates and bicarbonates. Reaction between the NOM and the residual chlorine causes the formation of more than 600 DBPs, among which Trihalomethanes (THMs) and Haloacetic acids (HAAs) are the most common groups. Some of the compounds within these groups are suspected as carcinogen and are harmful to human health at chronic

dose and long-term exposure. To protect public from exposure to these DBPs, regulatory agencies have placed limitations on their levels in the drinking water. Among various methods, activated carbon is one of the effective filtration technologies to reduce THMs and HAAs in the drinking water supplies but due to its high cost, small communities cannot afford such technology. The objective of this study is to develop a low cost adsorbent using extracted carbon from oil fly ash and assess its feasibility by applying it to a small community near St. John's, Newfoundland.

2 DRINKING WATER AND HEALTH

The removal of pathogens in drinking water causing water-borne diseases such as cholera, typhoid and dysentery is a global issue and water distribution supply systems are unable to provide safe water supply to consumers unless these diseases-causing pathogens are inactivated in the system. Even countries such as USA and Canada with advanced water treatment technologies are facing problems to control the outbreaks of water-borne diseases.

Approximately 3.4 million people and young children, under the age of five, die every year in Asia,

Africa and Latin America due to the water-borne diseases (WHO, 2002). More than 100 people died and 400,000 were affected in 1993 in Milwaukee city of Wisconsin state in USA as a result of presence of cryptosporidium in the drinking water source (Mackenzie et al., 1994). In May 2000, in Walkerton (Canada), 2,300 people suffered from E-coli contamination in the drinking water and seven died (MOE, 2002).

Chlorine is now commonly used as disinfectant in North American water systems (USEPA, 2006; Health Canada, 2007). Although other disinfectants such as chloramines, chlorine dioxide, ozone and ultraviolet (UV) radiation are also used to disinfect water but chlorine is the most effective and economical. However once water is disinfected with chlorine, some by-products of DBPs, in the presence of natural organic matters (NOM) in the water supply system, are formed. The formation of DBPs, however, depends on parameters such as level and characteristics of NOM, pH value, temperature of water, and chlorine dose and bromine concentration in the water (Zhang et al., 2008). Most common DBPs are Trihalomethanes (THMs), Haloacetic acids (HAAs), Haloacetonitriles (HANs), Haloketones (Hks), Carboxylicacids, Aldoketoacids, Aldehydes, Oxyhalides, and Cyanogen halides (Krasner,1999). More than 600 DBPs have been found in water systems (Richardson et al., 2007).

Drinking water is supplied to consumers through the distribution systems and most of these systems have DBPs. The Canadian regulatory agencies have developed guidelines for THMs as 100µg/L and for HAAs as 80 µg/L and other countries follow their own guidelines for drinking water. The individual THM compounds are Chloroform (CHCl₃), Bromodichloromethane (CHCl₂Br), Dibromochloromethane (CHClBr₂) and Bromoform (CHBr₃). Nine compounds of HAAs are Monochloroacetic acid (MCAA), Dichloroacetic acid (DCAA), Trichloroacetic acid (TCAA), Bromochloroacetic acid (BCAA), Bromodichloroacetic acid BDCAA, Dibromochloroacetic acid (DBCAA), Monobromoacetic acid (MBAA), Dibromoacetic acid (DBAA), and Tribromoacetic acid (TBAA).

3 DPB IN NEWFOUNDLAND WATER SUPPLY SYSTEM

The Department of Environment and Conservation in Newfoundland regularly monitors THMs and HAAs in the water supply systems (MOENL, 2011) in the province. The results are plotted in Figure-3.2 and Figure-3.3. As demonstrated in these Figures, THMs and HAAs in various communities in the province are higher than the Health Canada guidelines of 100 µg/L for THMs and 80 µg/L for HAAs;

the THMs level in few communities even exceeds 400 µg/L. More than one third communities in the province have higher HAAs level than Health Canada guidelines, and 27 communities have even higher than 300 µg/L HAAs in the water supply systems.

St John's is the capital city of Newfoundland and Labrador with a population of approximately 125,000. The province has hundreds of smaller communities. This research focuses on Pouch Cove community with a population of 1,900 and is located about 30km northeast of St. John's. The main source of raw water is a pond. This community does not have water treatment plant and relies solely on the chlorination of intake water and pH adjustment. This community has high level of THMs and HAAs in the drinking water supply system.

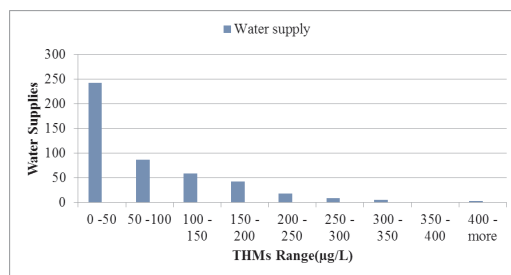


Figure 1. Distribution of THMs in Newfoundland & Labrador Communities (MOENL,2011).

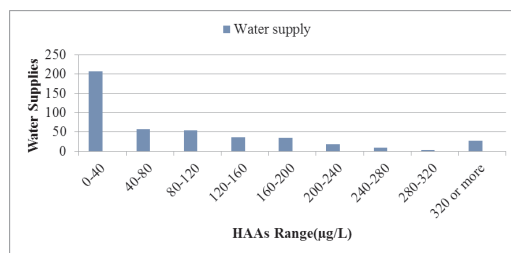


Figure 2. Distribution of HAAs in Newfoundland & Labrador Communities (MOENL, 2011).

St John's is the capital city of Newfoundland and Labrador with a population of approximately 125,000. The province has hundreds of smaller communities. This research focuses on Pouch Cove community with a population of 1,900 and is located about 30km northeast of St. John's. The main source of raw water is a pond. This community does not have water treatment plant and relies solely on the chlorination of intake water and pH adjustment. This community has high level of THMs and HAAs in the drinking water supply system.



Figure 3. Location map of Pouch Cove Community near St. John's, Canada.

4 SAMPLE CHARACTERIZATION

Tap water samples from Pouch Cove community were collected from different locations and were analyzed for nine compounds of HAAs and 4 compounds of THMs.

The micro-electron capture detector (μ -ECD) was used in the GC/ECD to analyse all nine compounds of HAAs which are Monochloroacetic acid (MCAA), Dichloroacetic acid (DCAA), Trichloroacetic acid (TCAA), Bromochloroacetic acid (BCAA), Bromodichloroacetic acid (BDCAA), Dibromochloroacetic acid (DBCAA), Monobromoacetic acid (MBAA), Dibromoacetic acid (DBAA), and Tribromoacetic acid (TBAA). To measure the concentration of four THM compounds in drinking water (i.e., Chloroform (CHCl_3), Bromodichloromethane (CHCl_2Br), Dibromochloromethane (CHClBr_2) and Bromoform (CHBr_3)), the EPA Method 501.3 with purge and trap procedure was used. GC/MS was used.

The extracted carbon from heavy oil fly ash (HOFA) was used as a filter media. HOFA is collected directly by the electrostatic precipitators of the power plants which use heavy fuel oil or crude oil as a fuel source.

Within Atlantic Canada Newfoundland and Labrador uses 3.4%, Nova Scotia and Prince Edward Island uses 15.8% and New Brunswick uses 36.1% of heavy fuel oil or mixture of heavy fuel oil and petroleum coke for power generation (Paul and Caouette, 2007). This indicates huge amount of HOFA is being generated by provincial power industries.

Major power generation and desalination facilities in the oil rich countries use heavy fuel oil. Saudi Arabia uses about 320 million barrels of heavy fuel oil and crude oil annually for power generation (Breakbulk online news, 2010-07-20) and as a result,

more than 400 tons HOFA is generated each day which is collected from the flues gas using electrostatic precipitators installed in most of the major facilities and dispose into landfills. Annual production of HOFA in Italy is about 27,600 tons (Rapporto Ambientale, 2000) while in Taiwan, it is 45,000 tons per year. (Hsieh and Tsai (2003). There are many other countries in the world which use heavy fuel oil for their facilities and generate significant amount of HOFA.

Since HOFA contains heavy metals and organic and inorganic contaminants, the conventional land disposal of HOFA can cause serious problems in the air, water, and soil on a long-term basis. The poor management and uncontrolled disposal of HOFA is hazardous to human health and environment. The toxic metals from HOFA dumping can leach into the ground water. The airborne particles from the OFA landfill can be a potential risk source to the surrounding population and has long term effects on human health

The chemical composition of HIFA varies from plant to plant but it contains significant amount of unburned carbon which can be used as adsorbent for water and wastewater treatment processes.

The impurities from HOFA obtained from a power plant were removed using the following cleaning procedures.

- 1 HOFA was washed with distilled water with the ratio of 1gm of fly ash with 10 ml of water and stirred at 10 rpm using Birds & Philips stirred Model number 7790-400 for 12 hours with the magnetic bar dip in the solution to capture the contamination at room temperature.
- 2 After filtration, the washed raw HOFA was thoroughly mixed with aqueous acid solution in the ratio of 1gm of fly ash with 5 ml distilled water, with 28% nitric acid solution at 60°C for 2 hours and then rinsed several times with the distilled water to remove nitrate ions.
- 3 The washed HOFA was then treated with HCL in the ratio of 1gm of fly ash with 5 ml distilled water with 15% HCL at 60°C for one hour and rinsed several time with the distilled water to remove the chloride ions.
- 4 After filtration the wet clean carbon was dried in the oven for 24 hours at 105°C.

A detailed analysis was conducted for raw fly ash and treated carbon for heavy metals obtained from Shoaibah power plant in Saudi Arabia as listed in Table 1. This power plant alone generates more than 50 tons of fly ash on daily basis. As shown in the Table, most of the metals are removed by washing with acid and water. The research is, however, in progress to recover metals such as nickel and vanadium from the fly ash and makes clean carbon free from the trace metals and other impurities at low cost.

5 EXPERIMENTAL SETUP

Column test experiment was designed using Pyrex glass column with dimension of 30 cm length and 40 mm internal diameter. 64 g of extracted clean carbon was used in the column. In order to prevent leaching of the fine particles into filtered water, glass wool and a 1 μ m filter at the outlet of the column were used. The column test was conducted using the same tap water which was used in the batch test. A flow rate of 3.5 ml/min was maintained using peristaltic pump. The advantage of using a peristaltic pump is that there is no cross-contamination since flowing water through the column does not touch any mechanical part of the pump and water passes through the internal tubing system. The filtered sample was collected and analyzed for the concentration of THMs and HAAs in the Pouch Cove water supply system.

The particle size of HOFA was measured with the HORBIA PARTICLE laser scattered particles size

Table 1. Metal in heavy oil fly ash (HOFA) and cleaned carbon.

| Metal | Raw HOFA (mg/kg) | Cleaned HOFA (mg/kg) | US EPA Limit (mg/kg) |
|---------------|------------------|----------------------|----------------------|
| Arsenic(As) | ND | ND | 0.01 |
| Cadmium(Cd) | ND | ND | 0.005 |
| Cobalts (Co) | 1.23 | ND | - |
| Copper(Cu) | 3.6 | ND | 1.3 |
| Mercury(Hg) | .094 | ND | 0.002 |
| Nickel (Ni) | 2176 | 332 | - |
| Lead (Pb) | 2.593 | ND | - |
| Selenium (Se) | ND | ND | - |
| Vanadium(V) | 5251 | 221 | - |
| Zinc(Zn) | 47 | 15 | 5.0 |
| Carbon | 86 | 90% | |

*ND = Not detectable

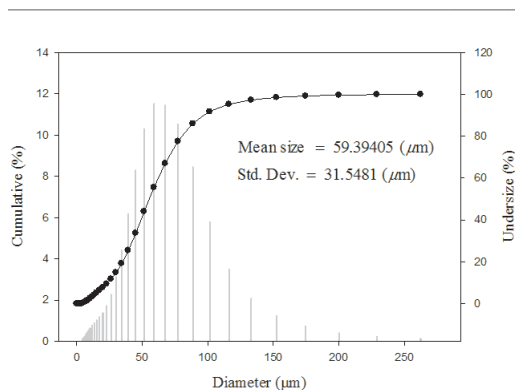


Figure 4. Particle size graph.

analyzer model LA-950 with wet dispersion methods in NaPO_3 solution. It shows the mean diameter of the collected washed fly ash is 59.39.5 μm with standard deviation 31.54 μm (Figure 4).

6 RESULTS AND ANALYSIS

The results are very encouraging to remove chloroform, BDCM, and bromoform. The adsorbent used in this experiment helped in significant removal of THMs. As shown in Figure 5 the total THMs in the reference tap water was 288 $\mu\text{g/l}$. While running water through filter media for 1440 minutes, the concentration of the total THM was reduced to 70 $\mu\text{g/l}$. After 2880 minutes of filtration, the concentration of total THMs was reduced to 117 $\mu\text{g/l}$.

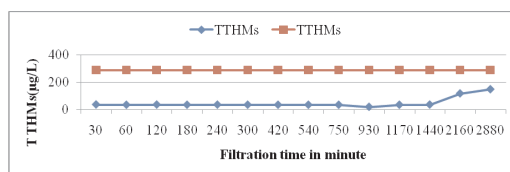


Figure 5. TTHMs concentration in filtered water with time of filtration.

Analysis of HAAs

The adsorbent was found very effective in the removal of HAA compounds in the reference tap water as shown in Figure 4.

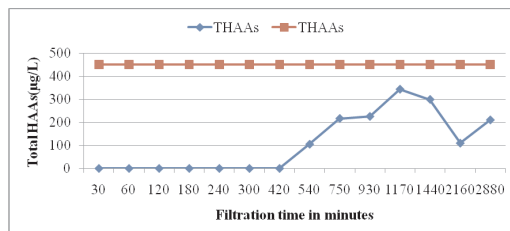


Figure 6. THAA concentrations in filtered water.

Even after filtration for up to 480 minutes, the total HAAs TBAA were removed by more than 70% in the filtered water supply system even for 2880 minutes filtration.

Although this research focuses mainly on the removal of DBPs from drinking water, in order to make the technology sustainable and affordable for small communities, column tests were also conducted on intake water to remove precursors causing the formation of DBPs in this water.

Water samples from the intake in Puch Cove Community were analysed for pH, total organic carbon, turbidity, and UV (254 nm), and column tests were performed to determine the removal efficiency of TOC

and dissolved matter. As shown in Table 2, extracted carbon from OFA shows promising results in reducing TOCs and other related parameters which are precursors in DBP formation (Ahmad, 2013). In the Pouch Cove system, a TOC level of 13.64 mg/l was reduced by more than 70% by AC (Table 2). There is also a considerable improvement in the turbidity reduction in the filtered water. UV254, a measure of UV absorption due to the presence of dissolved organic carbon, was reduced from 0.13 to 0.007; this indicates that the filter barrier can remove most of the organics from the water. Once the filtered water was chlorinated, there was a significant reduction in the formation of THMs and HAAs in the drinking water (Table 3) with much reduced DBP levels (Ahmad, 2013).

Table 2. Intake water quality before filtration and reduction in TOC, turbidity and UV adsorbance after filtration.

| | pH | UV254 (nm) | Turbidity (NTU) | TOC (mg/l) |
|------------------------|------|------------|-----------------|------------|
| Raw water quality | 6.61 | 0.13 | 1.45 | 13.64 |
| Filtered water in ml | | | | |
| Filtered water quality | pH | UV254 (nm) | Turbidity (NTU) | TOC (mg/l) |
| 240 | 6.14 | 0.007 | 0.35 | 1.16 |
| 540 | 6.7 | 0.01 | 0.34 | 1.04 |
| 900 | 6.63 | 0.03 | 0.34 | 2.10 |
| 1620 | 6.43 | 0.01 | 0.36 | 5.7 |
| 2940 | 6.4 | 0.005 | 0.4 | 6.73 |
| 3720 | 6.51 | 0.01 | 0.38 | 4.90 |
| 4020 | 6.71 | 0.05 | 0.34 | 5.49 |
| 4440 | 6.72 | 0.005 | 0.33 | 5.64 |
| 5100 | 6.65 | 0.001 | 0.36 | 5.45 |
| 6000 | 6.72 | 0.001 | 0.35 | 6.84 |
| 7500 | 6.74 | 0.001 | 0.37 | 5.56 |

Table 3. THM and HAA formation potential in raw and filtered water in µg/l with various contact times after chlorination.

| Contact time after chlorination in hours | Total THM Concentration, µg/l | | Total HAA Concentration, µg/l | |
|--|-------------------------------|----------------|-------------------------------|----------------|
| | Raw Water | Filtered Water | Raw Water | Filtered Water |
| | 4 | 258 | 115 | 192 |
| 12 | 270 | 114 | 226 | 34 |
| 18 | 298 | 149 | | |
| 24 | 281 | 158 | 232 | 40 |

7 CONCLUSIONS AND RECOMMENDATIONS

The main conclusions drawn from this study are as follows:

- 1 From tap water, more than 50% of the THMs were removed by the carbon filter with a continuous run for 48 hours.
- 2 The extracted clean carbon removed more than 95% of the HAAs from the tap water with a continuous filtration for seven hours.
- 3 The carbon also has potential to remove turbidity, dissolved organic carbon and total organic carbon significantly thus reducing DBP formation potential in filtered water

Although the adsorbent used for this research has the potential to remove DBPs from drinking water supply systems and their precursors, it requires further in-depth investigation with a summary for future research as follows.

- 1 It is recommended that the adsorption capacity of the carbon be improved by employing different physical and activation processes. Further research work is needed on the preparation of activated carbon with different temperatures, activation times, and loss rates.
- 2 More research is required on the regeneration of the adsorbent.

ACKNOWLEDGMENT

The financial support provided by the Harris Centre of Memorial University through the Royal Bank of Canada (RBC) Water Research and Outreach Fund is highly appreciated.

REFERENCES

- [1] Ahmad, M. 2013. Affordable Filtration Technology of Safe Drinking Water for Rural Newfoundland, M.Eng. Thesis, Memorial University, St. John's, Canada.
- [2] Break-bulk online news (2010-07-20).
- [3] CDW. (2006). Federal-Provincial-Territory Committee on Health and the Environment. Canadian Drinking Water Guidelines.
- [4] Health Canada, (2007). Guidelines for Canadian Drinking Water Quality. http://www.hcsc.gc.ca/ewh-semt/pubs/water-eau/guideline_sixth-rec_e.html.
- [5] Hsieh, Y.-M., and Tsai, M.-S. (2003). Physical and chemical analyses of unburned carbon from oil-fired fly ash. Carbon, 41, 2317–2324.
- [6] Krasner, S. W.(1999). Chemistry of disinfection by product formation. In P.C. Singer(Ed), Formation and Control of Disinfection By- Products in Drinking water (pp. 27–52). American Water Works Association, Denver, CO.

- [7] Mackenzie, W. R., Hoxie, N. J., Proctor, M. E., Grandus, M. S., and Blair, K. A. (1994). A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the public water supply. *The new England Journal of Medicine*, 331, 161–171.
- [8] MOE. (2002). The Ontario Ministry of the Environment, A summary: Report of the Walkerton inquiry: The events of May 2000 and related issues. Part One: Ontario Ministry of the Attorney General.
- [9] MOENL. (2011). Department of Environment and Conservation of New found land and Labrador. http://www.env.gov.nl.ca/env/waterres/quality/drinkingwater/haa_graph.html.
- [10] Paul, M., and Caouette, A. (2007). Heavy fuel oil consumption in Canada. Canada: Analytical Paper Statistics.
- [11] Richardson, S. D., Plewa, M. J., Wagner, E. D., Schoeny, R., and DeMarini, D. M. (2007). Occurrence of Genotoxicity and Carcinogenicity of regulated and emerging disinfection by-product in drinking water, A review and road map for research. *Mut. Res.*, 636, 178–242.
- [12] Rapporto Ambientale 2000. ENEL, Italy.
- [13] USEPA (2006). National primary drinking water regulation, disinfectants and disinfection by product rule. *Federal Register*, 71(2).
- [14] WHO. (2002). The World Health report: Reducing Risk, promoting healthy life, 1211 Geneva 27, Switzerland.
- [15] Zhang, H., Qu, J. H., Liu, H. J., and Zhao, X. (2008). Proportion of bromo-DBP in total DBPs during reclaimed-water chlorination and its related influencing factors. *Science in China Series B- Chemistry*, 51, 1000–1008.

A submersible automatic monitoring system of water quality employing a spectral scanning sensor

Y.Q. Zhao, X. Li, J. Gu & F. He

College of precision instrument and opto-electronics engineering, Tianjin University, Tianjin, P.R. China

Y.J. Fang

Institute of Hygiene and Environmental Medicine, Academy of Military Medical Science, Tianjin, P.R. China

ABSTRACT: Optical absorption is related to the content of most kinds of organisms in water. This article introduces a submersible water quality measurement sensor embedded in a spectral scanning system ranging from 200 to 720 nm. With the help of the sensor, multiple indexes, such as COD, $\text{NO}_{3,-}$, temperature, turbidity and so on, can be recorded in 1 minute. First, this article presents the entire prototype for the measurement system, including the optical entrance, diffraction grating, and opening cell that can be directly immersed in water, and a wireless data remote transportation is used. Then, the sensor was installed at several spots for longtime monitoring in river, lake, and sewage processing plants. Experimentally, there is a very interesting phenomenon observed in our experiments. Those indexes like COD are waving day by day in rhythm and have different patterns when it is sunny and rainy. This sensor is mostly suitable for the in situ monitoring of water quality with the virtue of all through spectrum information, which can be carried out simple parameters to describe the changes of water quality in real time.

1 INTRODUCTION

Water is the source of life and is also an important resource in the industrial and agricultural production. However, production and the sewage discharged by production contain many harmful factors, leading the environmental pollution to be more and more serious. Water quality monitoring has been the most important foundation for water resources management and protection by government and plants around the world.

Conventional methods of water quality event detection detect events directly by the national standard chemical methods or its modifications, which can generate many problems such as many kinds of chemical reagents consumed, secondary pollution presence, complex operation, time consuming, slow speed of analysis, poor stability, and so on.

Ultraviolet-visible (UV-VIS) spectroscopy using absorption of the spectrum in ultraviolet ranges to detect chemicals in water body appeared with the advantages such as no chemical reagents, simple operation, rapid detection (D.B. Hou, 2014). The UV absorption spectra of the water from the sea, lake, rain, rivers, and other natural water, which were described by Norio Ogura, showed that there was a certain correlation between the UV absorbance of the organic compounds in water and COD at 220 nm. However, because of the interference to the UV

absorption of $\text{NO}_{3,-}$, it will be more suitable for the determination of organic compounds in water containing a large number of NO_3 , at 250 nm (Norio Ogura, 1965). Mizunok et al. found that a good correlation exists between COD concentration and absorbance at 254 nm (A_{254}) using the UV absorption spectrophotometry detecting the content of organic pollutants in the river (Mizunok, 1978). To exclude the influences of turbidity and particles in water sample, additional wavelengths such as 546 nm were introduced as a sensitivity references (H. Zhang, 2008).

Traditionally, the major water quality monitoring system analyzed in laboratory is normally related to chemical reactions after collecting sample on site, which is not capacity of alarming those events like burst pollution accident. That is to say the existing laboratory-based analytical methods cannot meet the needs of real-time, multiple parameters for the water quality monitoring. As a submersible multiparameter water quality monitor, HydroLab analyzer, in which a probe is equipped synchronously with chroma, total phosphorus, total nitrogen, turbidity, and other kinds of parameter probes, is a multiparameter automatic online monitor that can measure 5–6 parameters for one time (K. Di, 2007). The method of in situ monitoring of water quality, which has high integration, is applicable to water quality monitoring of flowing water far away from the

shore and is used to monitor movingly in buoys or shipboard way.

This article introduces a submersible water quality measurement sensor embedded in a spectral scanning system ranging from 200 to 720 nm. With the help of the sensor, multiple indexes, such as COD, NO₃, temperature, turbidity, and so on, can be recorded in 1 minute. The data are transferred by ZigBee wireless transmission module to the rear monitoring platform. Then, the sensor is installed at several spots for long-time monitoring in river, lake, and sewage processing plan, achieving the in situ and unattended online monitoring of water quality.

2 PRINCIPLE OF MEASUREMENT

Substances with different molecules can absorb light with different certain wavelengths, leading to the jump of the electron between the two energy levels in atom to occur. In the same substances, the intervals between energy levels are worlds apart, while the energy needed for electron jumping in the same substances is equal. Therefore, with the selective light absorption characteristics of the material, a specific absorption spectrum can be formed by a specific substance, which can be used for qualitative analysis of water quality parameters.

Lambert–Beer law is the theoretical basis for the quantitative analysis of material when the light exposure to the material is strict monochromatic light and no interaction occurs among absorbing particles of the substance to be determined. The optical transmission is shown in Figure 1. Its definition can be expressed as follows:

$$I = I_0 e^{-KCL}, \quad (1)$$

where K is the absorption coefficient, C is the concentration of the objective substance, L is the optical path, and here it is the length of the cuvette. For a particular sample slot, the length L is constant, and for a given optical wavelength, the absorption coefficient K is essentially the same. (Y.Q. Zhao, 2011). In summary, Lambert–Beer law can be explained like this: the absorbance of a colored solution with homogeneous medium, which is passed through by a beam of

parallel monochromatic light, is proportional and to the product of the concentration of the solution and the optical path (Y.Ch. Li, 2005).

The assessments of the total amount of organic pollutants in water pollution focus on the chemical oxygen demand at present. The organic compounds that can absorb ultraviolet light at 254 nm is mainly conjugated with double bonds or aromatic substances, which is not present in the inorganic compounds, leading to no absorption in ultraviolet region (W. Yin, 2009). As a representative of the inorganic salt, nitrate nitrogen is one of the important indexes for evaluating the water quality. The absorption peak of nitrate is around 220 nm with no absorption at 254 nm. The intervention brought by the organic compounds in this band is relatively small (H.M. Wang, 2010).

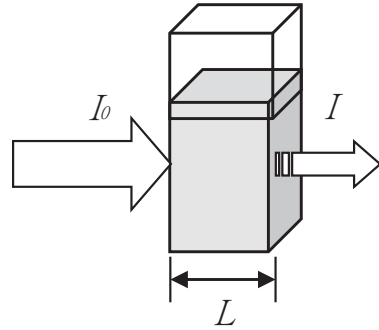


Figure 1. Schematic diagram of the changes of light through the solution.

3 EXPERIMENT AND DISCUSSION

The experiment of the submersible automatic monitoring system of water quality employing a spectral scanning sensor was conducted in the modern optical analysis technology and instrumentation laboratory, Tianjin University, Tianjin, China.

3.1 The experimental system

The structure diagram of the submersible automatic monitoring system of water quality is shown in Figure 2. The system with a submersible water quality

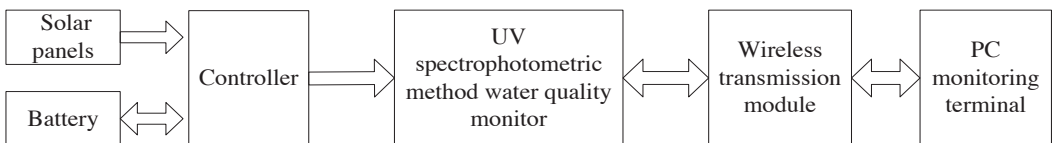


Figure 2. The submersible automatic monitoring system diagram.

measurement sensor is embedded with spectrometer, ranging from 200 to 720 nm, and spectrum width is 3 nm, with a motor to drive the system for scanning. This probe can be directly thrown into the target water body like lakes near our campus, multiple indexes such as COD, NO₃-, and turbidity can be recorded in 1 minute in one measurement. The absorbance of a single wavelength can be quickly returned in 1 second. For a loop of spectrum scanning, the total time cost is about 90 seconds. The various technical indexes of the UV water quality monitor are in conformity with the relevant provisions of the country. The physical diagram of UV water quality monitoring instrument discussed in this article is shown in Figure 3.

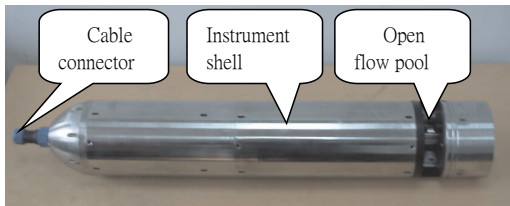


Figure 3. Physical diagram of UV water quality monitoring instrument.

3.2 Experimental design

The study can be split into two steps. As the first step, it is to work out the possibility to use UV absorbance as a method to assess the contents of COD and nitrate nitrogen, that is to say, it is to find out the relationship between the absorbance and the content of COD and nitrate nitrogen. The second step is using the absorbance to predict the concentration of COD and nitrate nitrogen according to the relationship.

The detected water samples are from a certain river in Tianjin. The absorbance spectrum of the river is shown in Figure 4. The results indicate that the water quality is stable and has no large fluctuations. The absorbance with a peak at 220 nm and a smaller one at 260 nm is small on the whole, which proves that the concentration of organic matter is small. No absorption in the visible area proves that the water is transparent and the turbidity content is low.

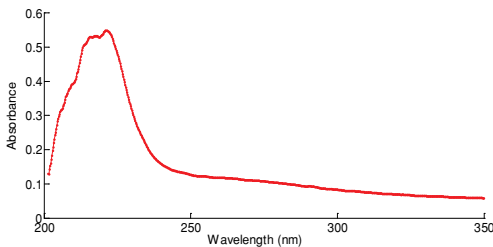


Figure 4. Absorbance spectra of a river.

3.3 Experiment of COD and the data analysis

According to the GB11914-89 “dichromate method for the determination of chemical oxygen demand (COD) of water quality,” the COD data are obtained with the water samples in same site of the river and so is the permanganate data according to the GB11892-89 “the determination of permanganate index of water quality.” To get rid of the influence of turbidity on COD, A_{546} is chosen for compensation. The fitted curve of the water absorbance and COD is obtained and shown in Figure 5. The relation is as follows:

$$\{COD\}_{mg/l} = 432.5(A_{254} - A_{546}) - 7.054$$

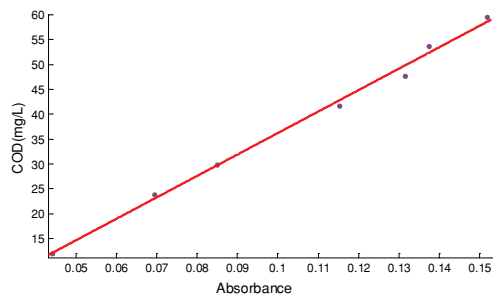


Figure 5. The standard curve of COD of the river.

Five consecutive days monitoring data are selected for analysis from April 17, 2010, 0 to April 22, 2010, 0. The measurement is done every 5 minutes, and the monitoring results are shown in Figure 6, the above of which is the scatter plot of the COD data during this period and the below is the trend chart. It shows that the COD concentration ranging from 40 to 80 mg/L, with two absorption peaks and two absorptions, is different at different times and waving day by day in cycle in Figure 6.

The compared value in every day between the UV method and the national standard method is shown in Table 1. The relative errors within 5% indicate quite good consistency between them. To be more intuitive to compare the measurement results of two methods, the line chart is set up as shown in Figure 7.

Table 1. The compared value between the UV method and the national standard method.

| Date | 4-17 | 4-18 | 4-19 | 4-20 | 4-21 |
|--------------------------|-------|-------|-------|-------|-------|
| COD _{Cr} (mg/L) | 51.04 | 55.74 | 59.25 | 61.25 | 64.81 |
| COD _{UV} (mg/L) | 49.35 | 54.57 | 59.08 | 63.83 | 63.69 |
| Relative error (%) | 3.31 | 2.10 | 0.29 | 4.20 | 1.73 |

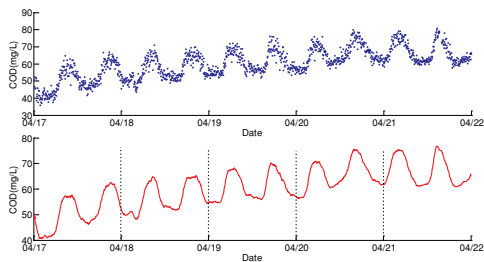


Figure 6. The variation of COD content of a river.

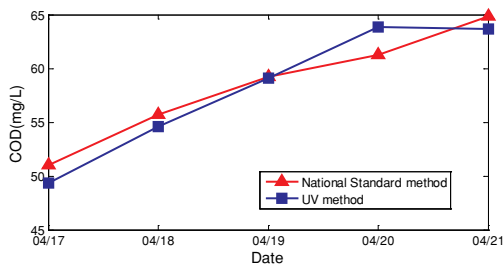


Figure 7. The compared diagram of COD concentration between the national standard method and the UV method.

3.4 Experiment of nitrate nitrogen and the date analysis

According to GB5750.5-2006, the national standard method, the concentration of nitrate nitrogen is detected every day. The absorbance at 220 nm is monitored with UV method at the same time. The absorbance at 275 nm can be used to remove the influence on the determination of nitrate nitrogen produced by organic matter. The fitting curve of the water absorbance and the concentration of nitrate nitrogen are shown in Figure 8. The relation is as follows:

$$\{\rho_{\text{nitrate nitrogen}}\}_{\text{mg/L}} = 4.643 \times A - 0.6434$$

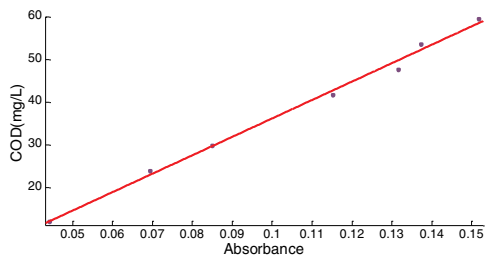


Figure 8. The standard curve of nitrate nitrogen concentration of the river.

The monitoring data for five consecutive days are selected for analysis from April 17, 2010, 0 to April 22, 2010, 0. The measurement is done every 5 minutes, and the monitoring results are shown in Figure 9. The above is the scatter plot of the nitrate nitrogen data during this period, and the below is the trend chart.

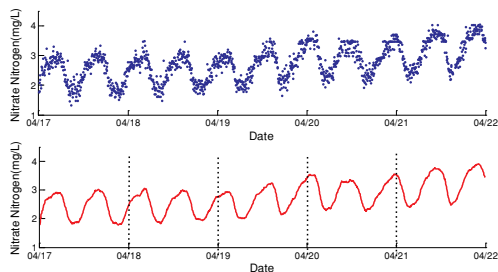


Figure 9. The variation of the nitrate nitrogen concentration in the river.

Similar to the changes of COD, the nitrate nitrogen concentration waving from 1.5 to 4 mg/L, with two absorption peaks and two absorptions, is different at different times in Figure 9.

It can be discovered from Figure 6 and Figure 9 that the COD concentration is just the reverse of nitrate nitrogen. When the COD concentration is in the peak, the nitrate nitrogen concentration is in the trough and when the COD concentration is in the trough, the nitrate nitrogen concentration is in the peak.

The compared value between the UV method and the national standard method is shown in Table 2. The relative errors within 5% present quite good consistency. To be more intuitive to compare the measurement results of two methods, the line chart is set up as shown in Figure 10.

As to national standard method, while it is possible to accurately reflect the content of water quality parameters, its detection cycle is long, resulting in failing to obtain the real-time online reflection of the change of the water quality, which proves the

Table 2. The compared value between the UV method and the national standard method.

| Date | 4-17 | 4-18 | 4-19 | 4-20 | 4-21 |
|---------------------------------------|------|------|------|------|------|
| Nitrate Nitrogen _{Cr} (mg/L) | 2.22 | 2.20 | 2.28 | 2.38 | 2.41 |
| Nitrate Nitrogen _{UV} (mg/L) | 2.21 | 2.23 | 2.31 | 2.43 | 2.39 |
| Relative error (%) | 0.45 | 1.36 | 1.32 | 2.10 | 0.83 |

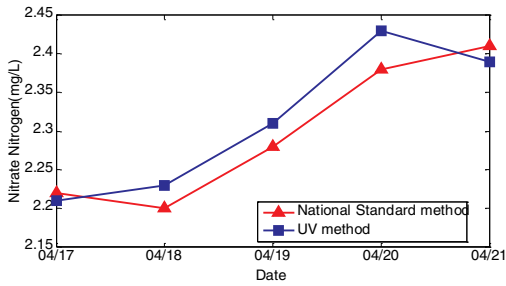


Figure 10. The compared diagram of nitrate nitrogen concentration between the national standard method and the UV method.

superiority of the ultraviolet water quality monitoring technology. Under the guarantee accuracy, Figure 6 and Figure 9 clearly reflect the change trend of COD and nitrate nitrogen every day, which is unmatched by the national standard method.

4 CONCLUSION

This article introduces a submersible automatic monitoring system of water quality employing a spectral scanning sensor, which can monitor many water quality parameters and be used with long-term unattended. Meanwhile, analysis of differences between methods monitored by the system and National Standard method for a river displays that a good consistency exists with each other. What is more, the system can work in a good condition when monitoring in a bad climate and can keep good communication. This automatic monitoring system really realizes in situ, real-time continuous monitoring of water

quality of lakes, reservoirs, rivers, and remotely data monitoring, which is going to play a significant role in the field of automatic monitoring of water quality in the future.

REFERENCES

- [1] K. Di, G.L. Peng, J.K. Yang, 2007. The major technical indicators and test methods of UV-Vis spectrophotometer, *University Physics Experiment*, 2007, 12 (4): 1–6.
- [2] D.B. Hou, Sh. Liu, j. Zhang, F. Chen, P.J. Huang, G.X. Zhang, 2014. Online Monitoring of Water-Quality Anomaly in Water Distribution Systems Based on Probabilistic Principal Component Analysis by UV-Vis Absorption Spectroscopy. *Journal of Spectroscopy*, 2314–4920.
- [3] Ch.H. Li, 2005. UV-Vis spectrophotometer. Beijing: Chemical Industry Press.
- [4] Mizuno Masaru, Suzuki Akira, et al. 1978. An application of an ultraviolet spectrophotometric method for the evaluation of matter in river water. *Aichiken Kogai Chosa Senta Shoho*, 6: 81–85.
- [5] Norio Ogura, 1965. Ultraviolet absorbing materials in natural water. *Nippon, Kagaku Zasshi*, 86(12):1286–1288.
- [6] H.M. Wang, 2010. The on-line water quality monitoring technology research of UV method: [Master thesis], Tianjin University, Tianjin.
- [7] W. Yin, 2009. The water quality monitoring equipment of UV absorbance and its application[J]. *Modern Scientific Instruments*, 02(1):125–129.
- [8] H. Zhang, J.Z. Wang, R. Li, J. Bai, et al, 2008. Attenuated effects of peptides derived from porcine plasma albumin on in vitro lipid per oxidation in the liver homogenate of mice. *Food Chemistry* 111, 364–369.
- [9] Y.Q. Zhao; Y.Ch. Li; Y. Zhen; Y.J. Fang, 2011. A Novel Monitoring System for COD Using Optical Ultraviolet Absorption Method, *Procedia Environmental Sciences*, 10: 2348–2353.

Characteristics of dissolved trace elements in the Jinshui River of the South Qinling Mts., China

H.M. Bu

Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

Q.F. Zhang

Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China

ABSTRACT: Dissolved As, Cu, Fe, Cr, Sr, Si, V and Zn and associated physiochemical parameters were analyzed to investigate their characteristics from October 2006 to March 2008 in the Jinshui River in the Qinling Mts., China. These dissolved trace elements displayed significant temporal trends rather than spatial patterns except Sr. Correlation and regression analysis showed the dissolved trace elements were controlled by physicochemical processes. The results provide a reference for ecological protection of river in the Qinling Mts., China.

1 INTRODUCTION

Trace elements in the natural environment, especially in rivers, are one of the primary pollutants due to their toxicity, persistence and bioaccumulation problems (Pekey et al., 2004). Dissolved trace elements have formed pollutions in many rivers all over the world and greatly threatened human health. The Han River in the South Qinling Mts., a major tributary of the Yangtze River, is the water source of the middle line of the south-to-north water diversion project (Zhang et al., 2009). However, with the economic development and intensive human activities, the water quality of the Han River has been deteriorated by dissolved trace elements. The study has chosen the Jinshui River, a major tributary of the Han River, to carry out an environmental monitoring program for water and soil conservation. The main objectives of the paper are to 1) investigate the spatiotemporal distributions and concentrations of dissolved trace elements, 2) recognize the relationships among dissolved trace elements and physicochemical parameters and 3) identify the physicochemical processes of dissolved trace elements.

2 MATERIALS AND METHODS

2.1 Study area

The Jinshui River (33°16'-33°45'N, 107°40'-108°10'E) of the upper Han River locates in Shaanxi Province of the south Qinling Mts., China. It drains

an area of 730 km² with a total length of 87 km and includes four major tributaries, *i.e.*, Xi River, Xindianzi River, Dong River and Lvguan River (Fig. 1). The river basin is in a humid monsoon climate zone. The annual average temperature in the catchment is approximately 11.5 °C, while the highest is 21.9 °C in summer and the lowest -0.3 °C in winter. The rainfall varies between 924 mm and 1240 mm per year, and most of the rainfall occurs from June to August. The Jinshui River basin is a typical mountainous catchment in the South Qinling Mts. Forest and shrub lands, as major land use types in the river basin covers 96.4% of the drainage area. Arable land is scattered in the midstream of the river. 51.39% of total populations of 19.21 thousand and 94.06% agricultural lands in the catchment are distributed in the lower reaches of the river, especially in Jinshui Town, but no industrial activities exist there.

2.2 Water sampling and analytical methods

Water samplings have been conducted during high-flow periods (October 2006 and August 2007), low-flow periods (December 2006, March 2007 and March 2008) and mean-flow periods (May and November 2007), respectively. Twelve sampling sites along the river are selected (Fig. 1). Of which, sites 1–5, sites 6–9 and sites 10–12 are located in the upper, middle and lower reaches of the river, respectively. Site 10 is located close to the Jinshui Town owning population of 9, 837, while site 12 is the confluence of the two rivers between the Jinshui River and the Han River.

Physicochemical parameters including water temperature (Temp), dissolved oxygen (DO), pH, total dissolved solids (TDS) and nitrate nitrogen ($\text{NO}_3\text{-N}$) were measured *in situ* using multiparameter water quality monitoring instrument (YSI Incorporated, Yellow Springs, Ohio, USA). Calibration of sensors was performed before every survey. For analysis of dissolved trace elements and dissolved phosphate (DP), water samples were filtrated through membrane filters with pore size of 0.45 μm (Millipore, USA), and then stored in the acid-cleaned polyethylene bottles after acidified to $\text{pH}<2$ with high-purity nitric acid for further analysis.

Dissolved trace elements and DP were quantified by inductively coupled plasma atomic emission spectrometer (ICP-AES) (IRIS Intrepid II XSP DUO, USA). Method accuracy and quality control were verified by triplicate analysis of standard reference material (SRM, SPEX CertiPrep, Inc, USA). A good agreement was observed between the data analyzed and those certified by the SRM, and the recoveries ranged from 99.8% (DP) to 105.4% (Zn). Concentrations of Cd, Co, Ni, Mo and Mn were below the detection limits and not shown here. Thus, dissolved trace elements of As, Cu, Fe, Cr, Sr, Si, V and Zn were selected for further analyses.

2.3 Statistical analysis

In this study, two-factor ANOVA was performed to evaluate the spatiotemporal variability of dissolved trace elements and physicochemical parameters. Homogeneity of variances was also evaluated using Levene's tests ($p>0.05$), while transformations were required for TDS, As, Fe, Cr and Sr. Multiple comparisons were conducted using Student-Newman-Keuls (SNK). Correlation and regression analyses were used to explore the relationships among the variables. All statistical analyzes were finished by performing SPSS 13.0.

3 RESULTS AND DISCUSSION

3.1 Concentrations of dissolved trace elements and their spatiotemporal variations

The concentrations of dissolved Cu, Fe, Cr, V, Sr and Zn (Fig. 2) are all below WHO recommendations for drinking water and, with the exception of Fe and Sr, generally within the Chinese standards for drinking water as well. Cu, Fe, Cr and Sr show peak concentrations at site 10 closing to Jinshui Town (Fig. 1), which suggests their common origin from anthropogenic sources (Martin et al., 1993; Kumaresan & Riyazuddin, 2008). Dissolved As has a larger concentration in high-flow period, implying that it may be induced into river by soil erosion in rainy season (Crosa et al., 2006). Dissolved Sr exceeds 1–2 orders of magnitude of the Chinese standards for drinking water, which indicates some unusual conditions needing further study. During three sampling periods, the maxima of dissolved V all occur at site 12, revealing the hydrological and chemical differences due to the confluence of the Jinshui River and Han River (Ravichandran, 2003). Mean values of As and Si are much higher in high-flow period than in the other two periods. Meanwhile, their peak values occur at sites 6 and 2, respectively. Cu has larger concentrations in high-flow and low-flow periods, while the maximum value appears in high-flow period at site 10. Maxima of dissolved Fe and Sr are observed at site 10 in low-flow period. Dissolved Cr, however, has the maximum at site 10 in mean-flow period.

The spatiotemporal variations of dissolved trace elements are shown in Fig. 2. Dissolved As, Cu, Fe, Cr, Sr and Si demonstrate no significant spatial patterns except Sr ($p<0.001$), but significant temporal trends ($p<0.05$). The results of V and Zn are not displayed because their data are insufficient to examine the spatiotemporal variability, although their mean values are illustrated in Fig. 2.

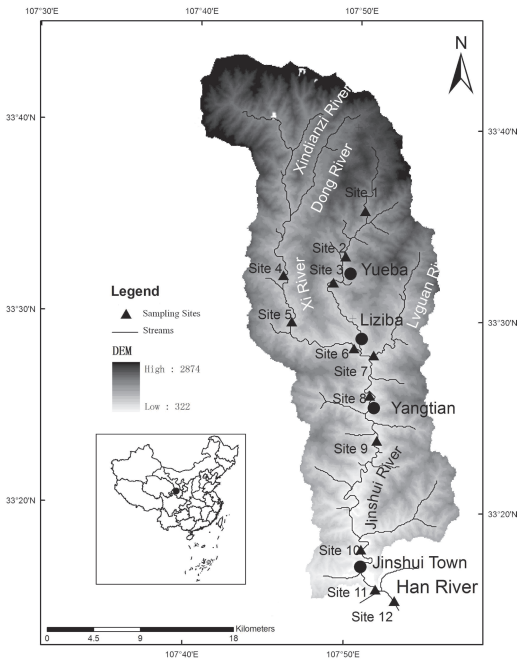


Figure 1. The location and 12 sampling sites of the Jinshui River in the South Qilinging Mts., China.

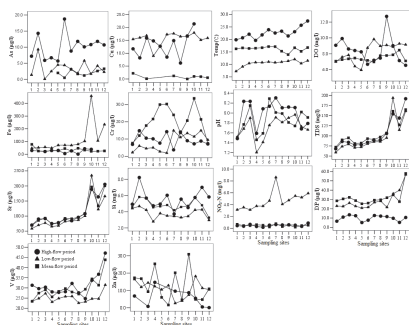


Figure 2. Spatiotemporal values (mean) of dissolved trace elements and associated physicochemical parameters in the Jinshui River, China.

Spatiotemporal distributions of physicochemical parameters are also plotted in Fig. 2, showing no significant spatial variabilities except TDS ($p < 0.001$). Water temperature, DP and $\text{NO}_3\text{-N}$ present significant seasonal variations ($p < 0.001$, 0.001 and 0.05, respectively). Changes in pH values (7.5–8.3) are relatively small among the different sampling sites and periods. Concentrations of TDS in sites 10 and 12 are much higher than those in other sampling sites. High value of $\text{NO}_3\text{-N}$ appears in low-flow periods, significantly different from those in high-flow and mean-flow periods. DP displays the same trend as $\text{NO}_3\text{-N}$, but its high concentrations appear in mean-flow periods.

3.2 Correlations between dissolved trace elements and physicochemical variables

Correlations between dissolved trace elements and physicochemical variables are presented. Water temperature is significantly correlated with As, Cr and V ($p < 0.01$). This may suggest a temperature effect on a series of chemical and physical processes including adsorption (Shiller, 1997), complexes formation and ion exchange (Kuppusamy & Giridhar, 2006) in the streams. Dissolved Cu is negatively related to DO ($r = -0.390$, $p < 0.05$), and its concentrations may be determined by oxidation-reduction process due to its oxidability (Hatje et al., 2003). As pH values in the river vary over a relatively narrow range (*i.e.*, 7.5–8.3), the dissolved trace elements do not show any significant pH-related change (Shiller, 1997) except Cr ($r = 0.477$, $p < 0.05$). Concentrations of dissolved Cu, Fe Sr and V are positively correlated to TDS ($p < 0.05$, 0.01, 0.01 and 0.05, respectively), respectively. The correlations indicate that the concentrations of dissolved trace elements might be increased by desorption from the TDS or sedimentary inputs (Shiller & Boyle, 1985). The dissolved Sr concentrations are strongly bound by the TDS contents

and linearly increase with them ($y = 18.08 + 0.778x$, $r^2 = 0.969$, $p < 0.001$). The dissolved V takes on a cubic trend with TDS ($y = 227.9 - 12.938x + 0.343x^2 - 0.002x^3$, $r^2 = 0.460$, $p < 0.001$), suggesting a relationship of regulation effect, to some degree.

The relationships among dissolved trace elements and nutrients are also discussed. Concentrations of $\text{NO}_3\text{-N}$ are positively associated with dissolved Cu ($p < 0.01$) and Fe ($p < 0.05$) since dissolved Cu and Fe display similar temporal patterns with $\text{NO}_3\text{-N}$ (Fig. 2), indicating the same sources. Dissolved Cu displays a cubic increase ($y = 2.743 - 0.452x + 0.02x^2 + 0.01x^3$, $r^2 = 0.440$, $p < 0.001$) with $\text{NO}_3\text{-N}$ as well. This could be explained by the influences of redox chemistry since nitrate is a fair measure of the redox condition in the river (anoxic) water column (Zwolsman et al., 1997). The inverse relationships between $\text{NO}_3\text{-N}$ and Si ($p < 0.05$), and V ($p < 0.01$) indicate possible uptake by floating algae (Ramessur et al., 2001), especially by diatom.

The DP values have strong negative affiliations with As and Cu ($p < 0.01$). Furthermore, dissolved As and Cu follow an exponential decrease ($y = 69.696 \times \exp(-0.18x)$, $r^2 = 0.644$, $p < 0.001$) and a cubic decrease ($y = 55.6 - 9.695x + 0.71x^2 - 0.016x^3$, $r^2 = 0.665$, $p < 0.001$) with DP, respectively, displaying removal trends for uptake of phytoplankton (Elbaz-Poulichet et al., 2006) as substitutes for DP in the growing periods, especially in May. Meanwhile, Dissolved phosphate is positively related to Si and V ($p < 0.01$), Sr and Zn ($p < 0.05$), indicating chemical similarity between them (Caccia & Millero, 2003).

3.3 Interrelationships between dissolved trace elements

The relationships among dissolved trace elements also exist. Dissolved As and Cu are significantly related with each other ($r = 0.995$, $p < 0.01$), suggesting a common origin probably from anthropogenic sources (Kumaresan and Riyazuddin, 2008). Most likely, dissolved Cu and Fe concentrations are determined by local redox rate, but the silicate input can change this condition, leading to the precipitation of Cu-Fe oxides (Elbaz-Poulichet et al., 2006). Therefore, significant negative correlations are observed between the two elements and dissolved Si ($p < 0.01$ and $p < 0.05$, respectively). With the exception of As, dissolved V is correlated to all the studied trace elements (Cu and Fe, $p < 0.01$; others, $p < 0.05$), which implies that these related elements are likely driven by similar hydrologic and geochemical processes (Neal et al., 1997). Dissolved Zn demonstrates no significant correlations with any other dissolved trace elements, although it is prone to be adsorbed and oxidized in rivers (Shiller & Boyle, 1985).

4 CONCLUSIONS

In this study, dissolved trace elements (As, Cu, Fe, Cr, Sr, Si, V and Zn) and physiochemical parameters (Temp, DO, pH, TDS, NO₃-N and DP) were analyzed from October 2006 to March 2008 at 12 sampling sites of the Jinshui River in the South Qinling Mts, China. The two-factor ANOVA indicated significant temporal variations of the dissolved As, Cu, Fe, Cr, Sr and Si ($p < 0.05$), but without significant spatial variations except Sr ($p < 0.001$). Correlation and regression analysis detected that the concentrations of dissolved trace elements were controlled by adsorption, desorption, complexes formation, ion exchange and redox processes. The analytical results for the Jinshui River provide an important basis for the ecological restoration and protection of the river system in the South Qinling Mts. and useful information on the hydrology and water quality of river environment for the middle line of the south-to-north water diversion project.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (Grant No. 41103069) and the Open Funding Project of the Key Laboratory of Aquatic Botany and Watershed Ecology, Chinese Academy of Sciences.

REFERENCES

- [1] Caccia, V.G. & Millero, F.J., 2003. The distribution and seasonal variation of dissolved trace metals in Florida Bay and adjacent waters. *Aquatic Geochemistry* 9, 111–144.
- [2] Crosa, G., Froebrich, J., Nikolayenko, V., Stefani, F., Gallid, P., Calamari, D., 2006. Spatial and seasonal variations in the water quality of the Amu Darya River (Central Asia). *Water Research* 40, 2237–2245.
- [3] Elbaz-Poulichet, F., Seidel, J. L., Casiot, C., Tusseau-Vuillemin, M.H., 2006. Short-term variability of dissolved trace element concentrations in the Marne and Seine Rivers near Paris. *Science of the Total Environment* 367, 278–287.
- [4] Hatje, V., Apte, S. C., Hales, L. T., Birch, G. F., 2003. Dissolved trace metal distributions in Port Jackson estuary (Sydney Harbour), Australia. *Marine Pollution Bulletin* 46, 719–730.
- [5] Kumaresan, M. & Riyazuddin, P., 2008. Factor analysis and linear regression model (LRM) of metal speciation and physico-chemical characters of groundwater samples. *Environmental Monitoring and Assessment* 138, 65–79.
- [6] Kuppasamy, M. R. & Giridhar, V. V., 2006. Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore. *Environment International* 32, 174–179.
- [7] Martin, J.M., Guan, D.M., Elbaz-Poulichet, F., Thomas, A.J., Gordeev, V.V., 1993. Preliminary assessment of the distribution of some trace elements (As, Cd, Cu, Fe, Ni, Pb and Zn) in a pristine aquatic environment: the Lena River estuary (Russia). *Marine Chemistry* 43, 185–199.
- [8] Neal, C., Robson, A. J., Jeffery, H. A., Harrow, M. L., et al., 1997. Trace element inter-relationships for the Humber rivers: inferences for hydrological and chemical controls. *Science of the Total Environment* 194/195, 321–343.
- [9] Pekey, H., Karakaş, D., Bakoğlu, M., 2004. Source apportionment of trace metals in surface waters of a polluted stream using multivariate statistical analyses. *Marine Pollution Bulletin* 49, 809–818.
- [10] Ramesur, R.T., Parry, S.J., Ramjeawon, T., 2001. The relationship of dissolved Pb to some dissolved trace metals (Al, Cr, Mn, and Zn) and to dissolved nitrate and phosphate in a freshwater aquatic system in Mauritius. *Environment International* 26, 223–230.
- [11] Ravichandran, S., 2003. Hydrological influences on the water quality trends in tamiraparani basin, south India. *Environmental Monitoring and Assessment* 87, 293–309.
- [12] Shiller, A. & Boyle, E.A., 1985. Dissolved zinc in rivers. *Nature* 317, 49–52.
- [13] Shiller, A., 1997. Dissolved trace elements in the Mississippi River: seasonal, interannual, and decadal variability. *Geochimica et Cosmochimica Acta* 61, 21–30.
- [14] Zhang, Q., Shen, Z., Xu, Z., Liu, G., Wang, S. 2009. The Han River Watershed Management Initiative for the South-to-North Water Transfer Project (Middle Route) of China. *Environmental Monitoring and Assessment* 148, 369–377.
- [15] Zwolsman, J., Van Eck, B., Van Der Weijden, C., 1997. Geochemistry of dissolved trace metals (cadmium, copper, zinc) in the Scheldt estuary, southwestern Netherlands: impact of seasonal variability. *Geochimica et Cosmochimica Acta* 61, 1635–1652.

Progress and prospect of migration character about insoluble sedimentary pollutants

L.X. Han

Key Laboratory of Shallow Lakes of Ministry of Education, Hohai University, Nanjing, China
College of Environment, Hohai University, Nanjing, China

F.F. Yan & B. Chen

College of Environment, Hohai University, Nanjing, China

ABSTRACT: Until now, it is insufficient to research the migration characteristics and mathematical prediction method of insoluble sedimentary organic pollutants discharged accidentally. Hence, the study on the migration characteristic of insoluble sedimentary organic pollutants and quantify simulation can provide theoretical basis and technical reference for predicting the water quality risk, when this kind of pollutant is discharged into the river under emergency. This article shows the research progress of migration characteristic and mathematical simulation in four different stages, namely process of adsorption and settlement, process of pollutants transportation along the bed, process of resuspension from bed to the water body, and process of transportation in water body. It also directs the future research contents and methods.

1 INTRODUCTION

With the development of chemical industry, the frequency of sudden accident has been increasing in China. The chemical discharged under emergency not only can cause a succession of serious economic losses but also could lead to water pollution, which would have an effect on water supply security and social stability. For the reason of the explosion of CNPC Jilin petrochemical double benzene plant in November 2005, about 100 tons of benzene homologs and derivatives flowed into the Songhua River, which had polluted the water so that residents' lives are influenced by this issue [Jin 2007]. On December 31, 2012, Tianji coal chemical company limited by shares, located in Changzhi Shanxi province, leaked 38.7 tons of aniline by accident of which 8.7 tons flowed into the Zhuozhang river, causing the concentration of aniline in the Wangjiazhuang monitoring section of the Shanxi provincial boundary, which reached 720 times as high as the national standard limit. This issue led a threat to drinking water safety of the millions of people in Hebei and Henan Province along the polluted river, which has drawn the wide attention from different social fields. Except the discharge from the chemical factories under emergency, it is also the potential risk source resulting in the leakage from the transportation of chemicals [Liu 2009]. When this kind of accident happened, it is significant to predict the space distribution character and transportation changes by time. Also, it can provide

important technical support for the government departments to take response timely, rapidly, accurately, and effectively.

However, there is a wide variety of dangerous chemicals. According to the characters of their solubility in water, chemicals can be divided into soluble and insoluble. They can also be divided on the basis of their density against water [Liu 2009]. Solubility and density are the two important physical indicators to impact transformation characteristic of contaminants when they flow into the natural water body. So far, most of technologies of water quality simulation and prediction are aimed at soluble pollutants. For the hydrophobic pollutants insoluble in water, the studies are mainly confined to the floating petroleum contaminants whose density is less than water [Han 2012]. The research about the hydrophobic organic compounds and density that is greater than the water is insufficient so that the related technical specifications are inadequate, which is tested by the system of environmental impact assessment. Therefore, it can be applied to improve the basic theory of different kinds of pollutant diffusion and methods of water environmental impact prediction technology, which can carry out the physical experiments and numerical simulation about the diffusive characteristics of the hydrophobic organic compounds, density greater than water, in the natural water body and also provide the basic theory and technical reference for water quality forecast of pollutants and risk prewarning.

2 RESEARCH PROGRESS

Sedimentary hydrophobic organic compounds are referred to the liquid organic matters that the density of which is greater than water, insoluble, or soluble poorly in the water. In general, they are all high toxic, persistent, volatile, lipophilic, hydrophobic, and so on, belonging to the field of persistent organic pollutants focused widely. Hence, the research about their transportation in the water under emergency is very significant. At present, mainly two kinds of methods, physical experiment and mathematical model simulation, are used to study the migration and transformation after the pollutants flowed into the water.

Considering the domestic and foreign existing research results synthetically, the process of migration and transformation in the surface water can be summarized as four processes: process of adsorption and settlement, process of pollutants transportation along the bed, process of resuspension from bed to the water body, and process of transportation in water body. They are shown in Figure 1.

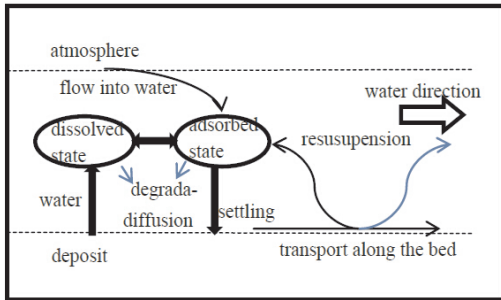


Figure 1. Process of transportation and transformation in the surface water.

Now, narrate the research progress of these four processes as follows.

2.1 Process of adsorption and settlement

The large number of sedimentary hydrophobic organic compounds would settle when they are discharged into the water accidentally, because of the available gravitation generated by the greater density. Settlement time and position in the water and the distribution area are closely related to the factors, when the matters sink into the bottom, such as the physico-chemical properties of the matters, the hydrodynamic conditions, the water depth, and the characteristics of suspended particles in the water. Priya et al. reported that a certain number of nitrobenzene would sink to bottom in the state of yellow-green oily matter after discharging into the still water [Priya 2006, Li 2009].

Li studied sedimentation and diffusion of sedimentary insoluble chemicals onboard in the sea believed that the effect of the main factors on the sedimentation process of the discharged pollutants accidentally is the initial bulk density and grain size of chemicals [Li 2009]. This kind of chemicals always hold small initial bulk density and then forms the so-called cloud after mixing with the surrounding seawater in the sea. Finally, they conclude that the depth of water has a great influence on the setting process of the liquid cloud due to the fact that the settling velocity of the cloud is slower than that of the conglomerate.

Chen thought that part of organic compounds would be adsorbed by suspended particles in water environment and then would settle with the large suspended particles. Their settle velocity is usually 0.2m/s [Chen 2004]. Wang reported the fact that the group of liquid chemicals settles itself is one of the main movement processes. Most of small droplets would accumulate near the collision point of the bed, also a little part of small droplets settle the area far away the collision point because of wave at the bottom [Wang 2009].

For the hydrophobic organic compounds discharged accidentally, Guo thought that they are mainly adsorbed by suspended particles, and most of them settle with suspended particles into the sedimentary organic matter [Guo 2009, Chevreuril et.al. 1998].

Sedimentary hydrophobic liquid, once leaks into the water, would be affected by turbulence effect in the process of sedimentation [Zhang 2003]. The matters, similar to the oily mass, would gradually become broken. The final point of sedimentation and drifting can be estimated by the water flow, settling velocity and water depth.

2.2 Process of pollutants transportation along the bed

Nitrobenzene as the persistent toxic substance has high stability in the water. It can settle in the bottom of water because of the greater density than water and not change for a long time. After the water pollution accident happened in the Songhua River, there are about 50–100 tons of nitrobenzene settled to the bottom of the Songhua River even though 44 tons had been flowed away [Han et al. 2008]. Then, they migrated to the downstream with the flow constantly or suspend into the water again through the turbulence effect of water flow. Song et al. showed that pollutants settle and enrich gradually into the deposition from the water. Under certain conditions, the organic matters absorbed in sediment phase that can produce various conversions to return into the water and migrate to the far place through repeated sedimentation and suspension [Song et al. 2012].

The study about migration of carbon tetrachloride, whose density is 1.597, showed that the motion

of carbon tetrachloride settled down to the bottom of the sea is a combination of migration and diffusion [Zhang 2003]. When it extended the maximum radius, the pollutants group broke into pieces, then suspended again to the overlying water, or adsorbed by the vegetation under sea [Del Vento & Dachs 2002].

The accident happened in the Songhua River as an example, Yu et al. when we apply the unsteady multimedial environment model to study the change of concentration in the air, water, soil, and deposition, the conclusion expresses that nitrobenzene, settled in the bottom of the river, suspends again into the water partly, and most of the rest flows to the next section.

2.3 *Process of resuspension from bed to the water body*

Until now, there are many researches on the laws of the organic matters releasing from deposition to the water. However, it is rare when it comes to resuspension characteristics of contaminants from the deposit on the bed surface. The hydrodynamic effect is one of the important factors that affect the pollutants released from the sediment in the aquatic environment. Flow and wave can produce shear force in the depositional interface. It could lead to suspend again when the force is greater than the critical value of the critical shear force [Dai & Liu 2011]. Numerous studies have verified that the density and rate of the sail have a large impact on the resuspension. Garrad et al. analyzed the release of pollutants from the deposition under the powerboat disturbance and found the phosphorus element having increased significantly by monitoring the water [Garrad 1987, Nedohin 1997]. Qin found that the sediment particles in the Taihu would suspend again once the surface wind velocity and the shear force reached 4 m/s and 0.4 dyne/cm², respectively, the higher speed and the more suspension [Qin et al. 2003]. For the PCBs of sediment from the Hudson river release simulation research showed that 2 dyne/cm² shear force continued for 3 days; the release flux would have one hundred times than that under stationary state [Schneider et al. 2007]. Among the study of sediment resuspension, the devices, like Annular Flume and Shaker, were used to simulate the surface sediment resuspension under the action of the low shear force, and the test apparatus like Sedflume was used to simulate the resuspension under the high shear force [McNeli et al. 1996, Poberts et al. 1998, Lick & McNeil 2001]. Another kind of research method is in view of the fate research of pollutants in the large scale of time and space. In 1979, Donald Mackay from Canada firstly proposed the fugacity method of chemical multimedial environment migration [Mackay 1979]. China began the studies about this field in 2001. The study areas cover the Hai River, the Yellow River, the Liao River, the Yangtze River, and

the Songhua River. However, most of them focused on the large-scale fugacity model without spatial resolution. It cannot be applied to describe the exchange between the deposition and the pollutants in the water [Liu et al. 2011] and the dynamic change process of contaminant resuspension from the bed surface.

Li adopted the “lift coefficient” to calculate resuspension flux for contaminants resuspension from the bed surface under accidentally discharging [Li 2009].

2.4 *Process of transportation in water body*

The process of transportation in water body includes the interaction between particles and pollutants, migration and diffusion along with the flow, volatilization of water–gas interface, sedimentation and resuspension of deposition–water interface, degradation of pollutants, and so on.

Li studied the contents of 10 kinds of nitrobenzene class organic pollutants in the river located in the middle and lower reaches the Yellow River. The results showed that the content distributed in the deposition is higher than that in the water and less than that in the suspended particles [Li et al. 2006]. The existing research results of transportation for sedimentary hydrophobic organic compounds are acquired using the same mathematical method with the soluble pollutants, and the convection–diffusion equation is suitable.

In the past, scientists carried out wide study about the adsorption and desorption of hydrophobic organic compounds in the water and also put forward a variety of theories, models, and empirical formulas [Shu et al. 2009, Chen & Peng 2001, Rao et al. 2004]. Researchers mainly made a lot of work in two ways to study the mechanism of organic–particle interaction. One is about the distribution coefficient of organics between the adsorption state and solubility state in the view of pollutants’ physical and chemical properties. The other one is particles’ composition and structure effect on adsorption [Meng et al. 2000, Zhang & Zhang 2008]. Liu showed adsorption rate constantly depended on the stirring intensity strongly supported by the semidynamic experiment [Liu 2006]. Chao demonstrated that suspended particles have a strong ability to adsorb nitrobenzene. In the high turbidity rivers, nitrobenzene was adsorbed by suspended matters and transported along with them [Chao et al. 1998]. Zhao founded that adsorption isotherm of trace nitrobenzene adsorbed by fine sediment was in accordance with Henry formula and C type of Giles’ adsorption isotherm models [Zhao 2006].

Steady nonequilibrium fugacity model is used to calculate the transportation and distribution of pc-DDT in deposition–water interface of the Haihe trunk stream; they concluded that the main factors affected the transportation of pc-DDT in deposition–water

interface are the biodegradation rate constants, the distribution coefficient of pollutants between suspended particles and water, and particles sedimentary flux [Chi & Zhang 2009]. Nie founded the transient discharged PCBs would distribute into all kinds of mediums immediately. It would take 2 hours to reach the balance [Nie et al. 2004]. Chen et al. studied the environmental factors, including temperature, volatilization, and light, which have influence on the transportation of nitrobenzene in the overlying water by the laboratory simulation device. Then they expressed that the temperature has the greatest effect and the light has the least effect [Chen et al. 2008, Zepp et al. 1987].

At present, the research about the settlement regularity of sedimentary hydrophobic organic compounds is relatively insufficient. Li calculated the deposition flux based on the settling rate is counted by the particle settlement formula under the still water environment [Li 2009].

2.5 Numerical simulation of pollutant transportation–transformation

At present, it is relatively insufficient about the physical mechanism of transportation–transformation for sedimentary hydrophobic pollutants. Therefore, there is a restriction for its mathematical simulation and model research.

Zhang insisted the source of the mathematical model could be treated as instantaneous source from the perspective of diffusion model with neglecting the duration of leakage [Zhang 2003]. Jurado et al. analyzed turbulence diffusion has an influence on vertical distribution of the persistent organic pollutants by firstly establishing a coupled one-dimensional hydrodynamic pollutants model in vertical direction [Jurado et al. 2007]. Peng established one-dimensional flow-transformation model according to an emergency event of nitrobenzene pollution. Based on measured concentration of the accident section, considering photodegradation, volatilization, adsorption, deposition, and so on, Peng forecasted the nitrobenzene concentration of time and space distribution in the downstream section [Peng 2011]. Wang et al. set up one-dimensional water quality model considering the effects of pollutant resuspension and analyze the concentration change along the flow and time [Wang et al. 2012].

3 PROSPECT

3.1 Process of adsorption and settlement

We did a lot of physical tests to research the process and characteristics of settlement about the sedimentary hydrophobic organic compounds discharged into the water under emergency. According to the hydrological

and hydrodynamic of surface water in the river, the physical model is established. And then we set up the corresponding model after a series of parallel tests. These tests are used to study what effect the density of pollutants, solubility, water depth, water velocity, the content of suspended particles will have on sedimentary process, respectively. Due to the flow of the turbulent effect, parts of the pollutants diffuse into the water body in the sedimentary process. Therefore, it is meaningful to analyze the response relationship quantitatively through identifying the main factor that affects the diffusion amount in the tests.

3.2 Process of pollutants transportation along the bed

In the leakage accident, the hydrophobic pollutants will deposit in certain area and show thickness in the bed surface. The deposition will extend and migrate along the bed surface under the synthetic action of shear force and topography underwater. After we analyze the motion law and dynamic characteristics of deposition on the bed surface through the physical model test and study the migration characteristics of different types of pollutants, which determined the differences in density and viscosity preliminarily, by designing the scene with different topography underwater and hydrodynamic environment, combining with the density flow theory of fluid mechanics, the mathematical physics equation that described the movement of the deposition is established to research and analyze the migration features.

3.3 Process of resuspension from bed to the water body

Due to the turbulence effect of overlying water, depositions are suspended again from the bed surface into the water. The strength of resuspension, which can be defined as the flux of pollutants suspended again into the water per unit time and per unit area, is related to the physical characteristics and the hydrodynamic characteristics of organic matters after analyzing the response relationship between the resuspension strength and the density of pollutants, solubility, viscosity, water velocity, and shear force through the physical model, such as annular flume.

3.4 Process of transportation in water body

The sedimentary hydrophobic organic compounds in the water will flow into the downstream along with diffusion, self-purification, evaporation, and photodecomposition. In the high sediment river, due to the effect of sediment's adsorption, the settlement of sediment will bring the settlement of pollutants with the adsorption and desorption between the sediment and

pollutant. Therefore, it is meaningful to focus on the adsorption–desorption relationship between sediment and organic matters, and the vertical distribution characteristics of dissolved and adsorbed pollutants are also significant.

3.5 Numerical simulation of pollutant transportation–transformation

The research needs to notice the adsorption and desorption between the sediment and pollutant for the heavy sediment-laden rivers. The vertical distribution of sedimentary pollutants is inhomogeneous due to the fact that the densities of this kind of pollutants are all greater than water and the sediment with serious inhomogeneous vertical distribution. Then, we establish the mathematical model of three dimension for dissolved state and adsorbed state, respectively. And we verify these models through the physical model tests of typical pollutants and simulate the plane and vertical distribution characteristics quantitatively in the downstream under emergency according to the typical cases.

4 CONCLUSION

At present, scientists have carried out a serious of related studies about the transportation of sedimentary hydrophobic organic compounds for the marine and surface water at home and abroad. Overall, the research in view of the transportation of sedimentary hydrophobic organic compounds under emergency is relatively insufficient. Summarization in detail is as follows:

- 1 The study on sedimentary hydrophobic pollutants settling from the water surface to the bed surface is found only in the marine. And the study found related to the surface water is relatively short. Also, due to the different depths and hydrodynamic characteristics between the surface water and marine, the existing research achievements cannot be applied to the surface water.
- 2 The studies on migration and extension of pollutants along with the bed surface are found only in the marine. Almost of the simulations are qualitative and simply quantitative. There are a lot of studies about the pollutants suspended again from the deposition to the overlying water. However, the achievements cannot be applied to describe quantitatively the resuspension process of the large number of depositions under emergency.
- 3 Various kinds of theories, models, and empirical formulas have been put forward to show the characteristics of adsorption–desorption, degradation, biological toxicity, and sedimentation for

the hydrophobic organic compounds in the water body. They are also applied to guide environmental pollution control and environmental remediation. However, in the accident, the sedimentary hydrophobic organic compounds have the characteristic of high concentration. And on one hand, the settling rate is related to the concentration of sediment and other suspended matters. On the other hand, its adsorption–desorption characteristic is different from the normal. Currently, the research about the law of organics sedimentation, which is combined with adsorption–desorption of the effects of sediment and other suspended matters, is absent.

- 4 The mathematical simulation should be quantitatively, comprehensively, and systematically based on the each process above. For the reason of the shortage of current research on the physical mechanism, the mathematical simulation method is lack of the system coupling of the entire transportation process.

REFERENCES

- [1] Andy, L. Sandy, J.G.R.J. 2012. Mass transfer coefficients for volatilization of polychlorinated biphenyls from the Hudson River, New York measured using micrometeorological approaches. *Chemosphere*. Article in Press.
- [2] Chao, K.P. Say Kee Ong. Angelos, Protopapas. 1998. Water-to-air mass transfer of VOCs: laboratory-scale air sparging system[J]. *J. Environ. Eng.* 124: 1054–1060.
- [3] Chen, D.Y. Peng, Y. 2001. The adsorption and desorption of hydrophobic organic compounds in water-soil-sediment system[J]. *Chemical Industry Times*(4): 11–14.
- [4] Chen, G. Jin, X.C. Jiang, X. Lv, X.J. 2008. Research of nitrobenzene's translocation and transformation rule in simulative section of stream[J]. *Research of Environmental Sciences* 21(1): 78–84.
- [5] Chen, X.M. 2004. Research on Decision-making Support System for Sinkable Chemicals Spilled Shipboard[D]. Dalian Maritime University.
- [6] Chevreuil, M. Blanchard, M. Teil, M.J. et al. 1998. Polychlorinated biphenyl behavior in the water /sediment system of the Seine River, France[J]. *Water Research* 32(4):1204–1212.
- [7] Guo, H.W. 2009. Advance in transportation and transformation of PCBs in water[J]. *Journal of Meteorology and Environment* 25(4): 48–53.
- [8] Chi, J. Zhang, X. 2009. DDTs transport between sediment and water in the mainstream of Haihe River in the urban area[J]. *Environmental Science* 30(8): 2376–2380.
- [9] Dai, G.H. Liu, X.H. 2011. Factors affecting the migration of persistent organic pollutants across the sediment-water interface of aquatic environment[J]. *Environmental Chemistry* 30(1): 224–228.

- [9] Garrad, P.N. Hey, R.D. 1987. Boat traffic, sediment resuspension and turbidity in a Broad land river[J]. *J Hydrol* 95: 289–297.
- [10] Han, H.J. Xue, M. 2008. The law research of nitrobenzene pollution's migration in the Songhua River. <http://www.paper.edu.cn>.
- [11] Jin, Z. Feng, J.P. Zhang, L. Peng, J. 2007. Monitoring and countermeasure of sudden water pollution accident[J]. *Water Resources & Hydropower of Northeast China* 25(270): 56–58.
- [12] Jurado, E. et al. 2007. Fate of persistent organic pollutants in the water column: Does turbulent mixing matter? [J] *Marine Pollution Bulletin* 54(4): 441–451.
- [13] Laura, J. Steinberg, K.H.R.R. 1997. Characterization of parameters in mechanistic models_ A case study of a PCB fate and transport model[J]. *Ecological Modelling*(97): 35–46.
- [14] Lick, W. McNeil, J. 2001. Effects of sediment bulk properties on erosion rates[J]. *Sci Total Environ* 266: 41–48.
- [15] Li, J.S. Xu, J. Luo, J.W. Luo, Z.L. 2009. A review on the research of environmental effects of nitrobenzene[J]. *Ecology and Environmental Sciences* 18(1):368–373.
- [16] Li, L.F. 2009. Study on Spread of Insoluble and Conservative Bulk Liquid Chemicals: 33–34[D]. Dalian Maritime University. Liu, H.F. 2009. The causes and countermeasures of frequent water pollution accidents[J]. *Environmental Protection and Circular Economy*(5).
- [17] Li, X.R. He, M.C. Sun, Y. Xia, X.H. Yun, Y. 2006. Nitrobenzene compounds distribution in the river from Xiaolangdi to Gaocun[J]. *Environmental Sciences* 27(3): 513–518.
- [18] Liu, R.P. 2008. Characterization of the Songhua River sediments and evaluation of their adsorption behavior for nitrobenzene[J]. *Journal of Environmental Science*(20): 796–802.
- [19] Liu, S.J. Lv, Y.L. Shi, Y.J. 2011. Advances in spatially explicit environmental multimedia fate models of POPs[J]. *Asian Journal of Ecotoxicology* 6(2): 129–137.
- [20] Liu, T. 2006. Nitrobenzene's transformation law between gas-water-sediments phases in the river water[D]. Harbin Institute of Technology.
- [21] Mackay, D. 1979. Finding fugacity feasible[J]. *Environmental Science & Technology* 13(10): 1218–1223.
- [22] McNeil, J. Taylor, C. Lick, W. 1996. Measurement of the erosion of undisturbed bottom sediments with depth[J]. *J Hydraulic. Eng.* 122: 316–324.
- [23] Nedohin, D.N. Elefsiniotis, P. 1997. The effects of motorboats on water quality in shallow lakes[J]. *Toxicol Environ* 61(1/4): 127–133.
- [24] Nie, X.P. Wei, T.L. Lan, C.Y. 2004. Distribution, accumulation and transfer dynamics of PCB1254 in a simulated micro-cosmos system[J]. *Acta Hydrobiologica Sinica* 28(5): 478–483.
- [25] Peng, H. Qi, D. Zhang, W.S. 2011. Study of fate model of nitrobenzene in rivers[J]. *Yangtze River* 42(24): 81–84.
- [26] PRIYA, M. H. MADRAS, G. 2006. Photocatalytic degradation of nitrobenzenes with combustion synthesized nano-TiO₂[J]. *Chemistry* 178: 1–7.
- [27] Qin, B.Q. Hu, W.G. Gao, G. et al. 2003. The dynamic mechanism of sediment suspension and conceptual model of endogenous release in Taihu[J]. *Science Bulletin* 48: 1822–1831.
- [28] Rao, P.S.C. Ran, Y. Xing, B.S. 2004. Importance of adsorption (hole-filling) mechanism for hydrophobic organic contaminants on an aquifer kerogen isolate[J]. *Environ Sci. Technol.* 38(16):340–348.
- [29] Roberts, J. Jepsen, R. Gotthard, D. et al. 1998. Effects of particle size and bulk density on erosion of quartz particles[J]. *J Hydraulic. Eng.* 124: 1261–1267.
- [30] Schneider, A.R. Porter, E.T. Backer, J.E. 2007. Polychlorinated biphenyl release from re-suspended Hudson River sediment[J]. *Environ Sci Technol* 41: 1097–1103.
- [31] Song, X.X. Wang, Z.W. Hu, X.F. 2012. Research progress of persistent organic pollutants in different media[J]. *North Environment* 24(4): 135–139.
- [32] Shu, Y.H. Wu, H.H. Huang, X.R. Jia, X.S. 2009. The adsorption and desorption behavior of CBs in the sediment[J]. *Geochemistry* 38(2): 153–158.
- [33] Wang, C. et al. 2012. A dynamic contaminant fate model of organic compound: A case study of Nitrobenzene pollution in Songhua River, China[J]. *Chemosphere* 88(1): 69–76.
- [34] Wang, C. 2009. The early warning and simulation technology of the unexpected chemicals discharge risk in the plain river network[D]. Tongji University.
- [35] Yoshinori, N. Tameo, O. 1995. Determination of nitrobenzenes in river water, sediment and fish samples by gas chromatography-mass spectrometry[J]. *Analytica Chimica Acta* 312(1):45–55.
- [36] Yu, X.F. Chen, J.W. Qiao, X.L. Wang, Z. Yang, F.L. 2006. Unsteady multi medium environment model in the application of environmental impact prediction of sudden environment pollution accident, taking the nitrobenzene pollution accident of Songhua River as an example. Environmental science society academic conference of China in 2006.
- [37] Zepp, R.G. Braun, A.M. Hoign, J. et al. 1987. Photoproduction of hydrated electrons from natural organic solutes in aquatic environments[J]. *Environ Sci Technol.* 21: 485–490.
- [38] Zhang, J. Zhang, L.J. 2008. The transportation research of organic matter in estuary[J]. *Journal of Ocean University of China* 38(3): 489–494.
- [39] Zhang, L.F. 2003. The decision system research of bulk liquid chemicals leakage at sea[D]. Dalian Maritime University.
- [40] Zhao, L. 2006. The transportation research of nitrobenzene pollutant in the mixed process of water and sewage[D]. East China Normal University.

Toward a more sustainable future: A critique on the mainstream of coastal tourism development since the 1960s

H. Zhang, Y.B. Wei & C.H. Fu

Hainan College of Software Technology, Qiong Hai, Hainan, China

ABSTRACT: Coastal tourism is considered as a crucial means of achieving sustainable development in many developing countries today. This article conducts a systematic review on the global mainstream of coastal tourism development since the 1960s, which illustrates its evolutionary stages, ideological and theoretical underpinnings, and most representative patterns and cases of each stage. Based on such an investigation, implications for coastal tourism researchers, developers, and policy makers are achieved to assist sustainable development in the global south.

KEYWORDS: Coastal tourism, sustainable development, literature review

1 INTRODUCTION

In 2010, the United Nations World Tourism Organization (UNWTO) stated that 940 million tourists had generated 693 billion Euros, and further forecasted that this number would reach 1.6 billion by 2020 [1]. Tourism not only is one of the largest industries in the world today but also has been considered as a means of development, specially in the developing world. During the past, various bodies related tourism with poverty reduction in less economically developed countries (LEDCs), due to its potential regarding generating economic growth at a large scale [2–4]. Along with this trend, as one of the most common and important types of tourism, coastal tourism (CT) has rapidly developed in the past decades. Nevertheless, studies also indicate the harmful impacts that tourism could have on the physical, economic, and social environments of those destinations [5–8].

As a response to this problem, some scholars argue that tourism has the ability and ought to benefit the vulnerable groups, not only generating net profits for them but also improving their living conditions thus attacking poverty, ultimately creating a sustainable future [3–4]. Among the most significant initiatives introduced are pro-poor tourism (PPT) [3] and sustainable tourism-eliminating poverty (ST-EP) [2]. However, these initiatives have also received criticisms in terms of both theoretical objections and practical issues [9–11], which retain themselves at best a microsolution to what is essentially a macroproblem.

2 KEY ISSUES TO BE ADDRESSED

Hence, the above facts led the author to believe that it was urgent to conduct a critical review on the global mainstream of coastal tourism development (CTD) since the 1960s, which advocated primarily by the north. Indeed, there have been substantial literatures on this topic, focused on theories, practices, effectiveness and deficiency, lessons from the past, and suggestions for the future [12–15]. However, the author insists that more systematic reviews on CTD are necessary, including its evolution, the ideological underpinnings, representative cases of each period, and the implications for those destinations in the south today, which are eager for a sustainable future via developing coastal tourism. Unfortunately, such literatures are still scarce up to date.

3 EVOLUTIONAL STAGES OF GLOBAL CTD

3.1 *1960s—1970s: Modernization and mass tourism*

Although the history of coastal tourism can be traced back to Roman times, the generally accepted modern coastal tourism firstly appeared in the Mediterranean islands after the Second World War. As Andriotis [16] stated, coastal tourism on these islands emerged as an unplanned activity due to the unexpectedly rapid growth. Hence, coastal tourism then has been introduced to the developing world in the late 1960s, as its

potential of accelerating economic growth via export promotion while exerting developing countries' comparative advantages [17]. Nevertheless, the proposition of CTD in the LEDCs is a "must" rather than an "unexpected incident" due to the prevailing ideology of development at the time, namely the modernization theory.

Academic circle considers modernization as a theory, which is used to explain the process of human societies' development based on Social Darwinism. Dominated by such an ideology, coastal tourism in this era has been inevitably assumed as a means that not only generates foreign exchange and employment but also a "[...] catalyst for modernization, economic development and prosperity in emerging nations in the third world" [18]. As a result, the "Fordism" and mass tourism have become the most representative pattern of CTD in this period. Mass tourism was first developed in the second half of the 20th century within a few developed countries in the north, along with improvements in technology, incensement in disposable personal income (DPI), and greater convenience of mobility, altogether allowed the transport of massive people in a short space of time to destinations of leisure interest [19]. As a product of the prevailing ideology, mass tourism precisely represents the characteristics of capitalistic mode of production and social structure of this era. As Tourism Intelligence International [20] concluded, the pattern of mass tourism was designed "[...] no differently from the way in which Ford made automobiles [...]" everything is standardized and payable, based on an industrialized and standardized form of mass production.

However, modernization theorists has been questioned for a range of reasons [21–24], primarily due to its nature as an ideology spread by the western world for justifying its involvement and domination of the developing world. As a matter of course, attentions of mass tourism development in those LEDCs were merely paid to increase the economic scale, while the negative impacts on environment and society were largely ignored by developers [24].

First of all, according to Holder's "Self Destruction Theory of Tourism" [25], mass tourism would eventually lead to the environmental degradation of a remote and exotic coastal destination. The most representative example that is corroborated with this theory was Cancun Quintana Roo of Mexico, where Wiese [26] observed how island had been devastatingly developed from an "escape paradise" to a heavily polluted region with a series of solid waste and social friction. Meanwhile, the construction of freeways and golf courses in the Mediterranean resulted in obvious degradation of the wetland ecological systems, numbers of species such as *Chelonia mydas* and *Dermochelys coriacea* are now in peril of extinction [27].

In addition to its negative impacts on the coastal environment, some scholars argued that mass tourism had played a role of supporting a new form of colonialism to a certain extent in the south. Contrarily, coastal tourism was an industry dominated by MNCs or local elites [28], which exploited local workforce and resources and perpetuated inequity and poverty [14–15]. For instance, Cevat stated that coastal tourism has not made an obvious contribution to the national development in Turkey since 1963. Due to interventions from the western MNCs, CTD of Turkey primarily focused on how to shape the country as a destination with "cheap sun and sand" to attract massive tourists from the north. The economic growth was offset to a certain extent by intensification of poverty, imbalanced development among different regions, and friction among various social classes.

Therefore, under such a pattern, power has been inevitably transferred to those MNCs and local elites who dominate the coastal tourism market [28], while the local nonprivileged people have little control. At the same time, the most representative international development agencies, namely the World Bank Group (WBG) and the International Monetary Fund (IMF), as the leading bodies of CTD of this era, also intervene in the coastal tourism policy making of developing countries, thus advancing the interests of the north in the south. For instance, the deforestation of the Brazilian Amazon is due to the investments of the Federal government, which is in fact ill advised by international agencies [29]. Besides, the WBG-funded "master plans" for Fiji and Sri Lanka are also ill proposed, since environmental issues are hardly discussed while social issues "regarding community participation are not even on the agenda" [30].

3.2 1980s–1990s: *Dependency theory as a response*

Since the 1980s, although mass tourism and Fordism still have dominated the market, the dependency paradigm has become popular as a critical response to this situation. Adherents of dependency theory maintains that "[...] the metropolises that initiates international tourism have exercised control over developing countries through provision of expertise and capital, trapping the recipients into a global system, over which they have minimal control. Therefore, receipts are retained in the metropolitan countries and foreign exchange is leaking away from the peripheries and into the metropolises [...]" [24]. Thus, in the developing coastal destinations, political, institutional, and economic structures have kept them in a dependent position in developed countries.

However, the dependency theory also faces challenges since it emphasizes LEDCs' external conditions over internal factors, while it does not provide

explicit policy recommendations for newly independent nations [23]. For CTD, it is vague on suggesting concrete plans practical issues the LEDCs faced, such as increase in economic benefits or mitigation of environmental issues [13]. Moreover, it views tourism as a pure neocolonial project while ignores “[...] the wants and ambitions of those about to be modernized [...]” [22].

3.3 1990s—Neoliberal paradigm and sustainable tourism development

Also in the 1990s, the Neoliberal economic paradigm became fully enshrined in the approaches taken by global developers and was strongly supported by the leading multilateral agencies such as the United Nations (UN), Asian Development Bank (ADB), WBG, and IMF. The Neoliberal paradigm advocates economic rationalism and efficiency, market liberalization, globalization, and a slight roll for the state [21]. Accordingly, tourism is considered as an invisible export industry in the tertiary sector with respect to the entire international markets [21]. Meanwhile, the emergence of sustainable development leads to a new school of thought, which believes that tourism should be recognized as a means to benefit the poor, not only generating net profits for them but also improving their living conditions toward a better life [31–34]. The strong alignment between the two schools of thought has led to a new pattern, namely the sustainable tourism development (STD).

The UNWTO (2004) states that guidelines and management practices of STD are applicable to all forms of tourism in all types of destinations including coastal tourism. A long-term sustainability cannot be achieved unless an integrated planning has been made before the actual development to guarantee a suitable balance among the environmental, economic, and sociocultural dimensions. Such an approach ensures a “continuous, flexible, adaptive, and transparent process” of CTD, thus to generate multiple benefits for national (or regional) development.

3.3.1 Patterns of sustainable coastal tourism development

Governments, international organizations and associations, and scholars have devoted to establishing various patterns of STD to assist coastal tourism since the 1990s; among the most significant patterns proposed are Planning for Sustainable Tourism Development (PSTD), Strategic Integrated Sustainable Tourism Planning (SISTP), Natura 2000 Tourism Planning Approach, and Integrated Coastal Zone Management (ICZM). These patterns demonstrate the core of STD, namely maximizing the positive benefits of tourism while minimizing the negative impacts in

a sustainable manner, through good planning and management.

3.3.2 The issues of STD

Is STD a panacea? The answer of this article is negative. Apart from the vagueness and uncertainty of its definition [19; 35], academic circle has also criticized STD due to its theoretical and practical flows. First, scholars argue that the uncertainty of the definition has led to the problem of how sustainability could be usefully measured [36–37]. The concept itself only highlights a “suitable” balance among three dimensions, however, but often fails to illustrate the criterion to those developers. Second, it appears that the proposition of STD is based on a “tourism-centric” assumption, while ignores the opportunity cost of tourism development, which is in fact inappropriate. Because of the fact that, as Williams [38] argued, tourism cannot comprehensively represent all the needs of regional development, since its competing relationships with other industries over regional scarce resources. Moreover, it is also hard to determine the intensity of sustainability, thus to inform that the precise relationships between various modes of STD and their beneficiaries [39–40]. Finally, underpinned by the Neoliberal paradigm along with the trend of globalization, the global STD is “culture-disparaged” in nature, which actually harms cultural integrity and traditional ways of life in developing destinations.

4 CONCLUSION

This article has systematically reviewed the mainstream of global CTD since the 1960s. It appears that the evolution of the global mainstream is divided into three stages, with distinct theories, patterns, and representative cases of development. This implies that the research of CTD is never an isolated island but a significant component of global development community. The constant evolution of ideologies of the development of human being will keep shaping and transforming the ideologies, theories, and patterns of CTD, like the way it always does. Therefore, at this point, the author claims that tourism studies can never separate from development research. For a future research agenda of CTD, more knowledge and methodologies from the development research domain should be encouraged.

The investigation of global mainstream of CTD illustrates that although CTD is a global phenomena, however, it has long been dominated by western ideologies and schools of thought. Hence, the western ethnocentrism is inevitably embedded in the prevailing theories and patterns of CTD. As a result, the mass tourism, as a product of the modernization

paradigm that massively developed after the Second World War, fails to reduce poverty at large but causes a numbers of environmental and social issues in the developing world.

STD has become close to the centre of global CTD since the 1990s, advocated by the leading multilateral agencies. Dominated by the Neoliberal paradigm, STD focuses on eliminating the global poverty while ensuring a sustainable future via tourism development. Nevertheless, the dependency nature of global tourism is not diminished, since globalization and contractual agreements have empowered those MNCs in the north even more [24]. Evidence from practical studies implies that even STD may still issue in exploitation on workforce or displacement of communities, which is far from a realistic answer to sustainable development in the south.

After all, what we learned from STD as a crucial means for achieving sustainable development in the south today? This article suggests that STD should be recognized as a “constructive strategy” in the developing world, rather than a bible that can be followed blindly. On one hand, STD emphasizes that a successful CTD ought to be market driven, sophisticated, and planned while utilizing multidisciplinary expertise, attracting diverse capital from private sectors, and creating long-term economic, environmental, and sociocultural benefits for the host community. These ideas could be positively drawn into the policy making of CTD in developing countries. On the other hand, the function of the invisible hand should not be overestimated. The policy makers should establish a balance between the market force and government intervention. Attention should be paid not only to keep the integrality of the local culture but also to increase the cultural competitiveness within a global system.

ACKNOWLEDGMENTS

This article is funded by (a) the Hainan Federation of Social Science Circles (Project ID: HNSK 14–23), (b) Hainan Provincial Department of Science and Technology (Project ID: ZDXM2014084), and (c) Commission of Tourism Development, Hainan Gov, Research Project: <coastal tourism development in a long-term and sustainable manner - a comparison study on coastal tourism development between china and the west>. Corresponding author: C.H. Fu.

REFERENCES

[1] UNWTO, 2013. *Global Code of Ethics for Tourism: General information*. Available from: <http://ethics.unwto.org/en/content/global-code-ethics-tourism> [Accessed 27th Oct 2013]

[2] UNWTO, 2003. *Tourism with a Human Face*. Available from: www.world-tourism.org/liberalisation [Accessed 15th Aug 2013].

[3] Ashley, C., Boyd C. and Goodwin, H., 2000. *Pro-poor tourism: Putting poverty at the heart of the tourism agenda*. March, no. 51. ODI: London.

[4] Roe, D. & Urquhart, P., 2004. *Pro-poor tourism: Harnessing the world's largest industry for the world's poor; turning the rhetoric into action for sustainable development and poverty reduction*. In T. Bigg (ed.) London: Earthscan.

[5] Beekhuis JV., 1981. Tourism in the Caribbean: impacts on the economic, social and natural environments. *Ambio* 1981; 10:325–31.

[6] Archer E., 1985. Emerging environmental problems in a tourist zone: the case of Barbados. *Caribbean Geography*, 2:45–55.

[7] Baines GBK, 1987. Manipulation of islands and men: sand-cay tourism in the South Pacific. In: Britton S, Clarke WC, editors. *Ambiguous alternative: tourism in small developing countries*. Suva: University of the South Pacific, 1987. p. 16–24.

[8] Hanna N, Wells S., 1992. *Sea sickness*. In Focus (Tourism Concern) 1992; 5:4–6.

[9] Komilis, P., 1994. Tourism and sustainable regional development. In A.V. Seaton (ed.) *Tourism: The State of the Art* (pp. 65–73). Chichester: John Wiley.

[10] Hunter, C. & Green, H. (1995) *Tourism and the Environment: A Sustainable Relationship?* London and New York: Routledge.

[11] Harrison, D., 1996. Sustainability and tourism: Reflections from a muddy pool. In L. Briguglio, B. Archer, J. Jafari and G. Wall (Eds) *Sustainable Tourism in Islands and Small States* (pp. 69–89). London: Pinter.

[12] Khan, M. M., 1997. Tourism Development and Dependency Theory: Mass Tourism vs. Ecotourism. *Annals of Tourism Research*; 1997, 24(4):988–991

[13] Brown, D. O., 1998. In search of an appropriate form of tourism for Africa: lessons from the past and suggestions for the future. *Tourism Management*; 19(3), 237–245.

[14] Schilcher D., 2007. Growth Versus Equity: The Continuum of Pro-Poor Tourism and Neoliberal Governance. *Current Issues in Tourism*. 2007(10), 2/3, p166–193. 28p.

[15] Spenceley A. & Meyer D., 2012. Tourism and poverty reduction: theory and practice in less economically developed countries. *Journal of Sustainable Tourism*. 2012(10), 3, p297–317. 21p.

[16] Andriotis, K., 2003. Coastal resorts morphology: the Cretan experience. *Tourism Recreation Research*, 28 (1), 67–75.

[17] Brohman, J., 1996. New directions in tourism for third world development. *Annals of Tourism Research*. 23(1), 48–70.

[18] Williams, S., 1998. *Tourism geography*. London: Routledge.

[19] Yu Xiangyang, Sha Run & Hu Shanfeng, 2009. Criticism and Research on Sustainable Tourism, *ECONOMIC GEOGRAPHY*, 12(29), 2090–2095.

[20] Tourism Intelligence International, 2008. *Sustainable tourism development: a practical guide for*

- decision makers*. Available from: <http://www.tourism-intelligence.com> [Accessed 27th Oct 2014]
- [21] David J.T., 2002. The Evolution of Tourism and Development Theory. In: Sharpley R., & David J.T. (Eds) *Tourism and Development: Concepts and Issues*. Clevedon • Buffalo • Toronto • Sydney: Channel View Publications. Available from: http://web.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzgwNTQxX19BTg2?sid=bd59f783-0967-4026-90b7-906a90974e77@sessionmgr112&vid=1&format=EB&lpid=lp_v&rid=0 [Accessed 6th Nov 2013].
- [22] Harrison, D., 1992. International tourism and the less developed countries: The background. In: D. Harrison (ed.) *Tourism and the Less Developed Countries* (pp. 1–18). Toronto: Belhaven.
- [23] So, A., 1990. *Social Change and Development Modernisation, Dependency, and World-System Theory*. London: Sage.
- [24] Koutra, C., 2013. *More than simply corporate social responsibility, Implications of corporate social responsibility for tourism development and poverty reduction in developing countries, A political economy perspective*. New York: Nova Science Publishers, Inc.
- [25] Holder J., 1988. Pattern and impact of tourism on the environment of the Caribbean . *Tourism Management*, 1988, 9 (2) : 119–127.
- [26] Wiese, P V., 1996. *Environmental impact of urban and industrial development a case history: Cancun Quintana Roo, Mexico*. Available from: <http://www.unesco.org/csi/wise/cancun1.htm>. [accessed 5th Nov 2014]
- [27] Davenport J., 1998. Temperature and the life history strategies of sea-turtles. *Journal of Thermal Biology*, 1998, 22(6): 479–488.
- [28] orres, R., 2002. *Cancun's tourism development from a fordist spectrum of analysis*. London: Tourist Studies, Sage.
- [29] Moran, E., 1983. *The dilemma of Amazonian development*. Boulder, CO: Westview Press.
- [30] Burns, P., 2004. The 1990 Solomon Islands tourism plan: a critical discourse analysis. *Tourism and Hospitality Planning and Development*, 1(1), 57–78.
- [31] UNEP, 1999. UNEP principles on the implementation of sustainable tourism. 20th Session of UNEP's Council. New York: UN.
- [32] UNCSD, 1999. Decisions of the General Assembly and the CSD. 7th Session, Decision. Number 7 & 3: *Tourism and Sustainable Development*.
- [33] UNLDC III, 2001a. The Canary Islands declaration on tourism in the LDCs. In: *High Level Meeting on Tourism and Development in the Least Developed Countries*, UN 3rd Conference, 26–29 March 2001, Las Palmas, Spain. Spain: United Nations.
- [34] UNLDC III, 2001b. Summary of sessions. In: International trade, commodities and services/tourism A/conference 191/L.12. 3rd United Nations Conference on the Least Development Countries. 14–20 May 2001, Brussels. Belgium: United Nations.
- [35] Liu Z., 2003. Sustainable tourism development: a critique. *Journal of Sustainable Tourism*, 2003, 11 (6): 459–474.
- [36] McCool, S.F., Moisey, R.N. & Nickerson, N.P., 2001. What should tourism sustain? *Journal of Travel Research* 40 (2), 124–31.
- [37] McCool, S.F., & Lime, D.W., 2001. Tourism carrying capacity: Tempting fantasy or useful reality? *Journal of Sustainable Tourism* 9 (5), 372–88.
- [38] Clarke J., 1997. A framework of approaches to sustainable tourism. *Journal of Sustainable Tourism*, 1997, 5 (3): 224–234.
- [39] Hunter C., 1997. Sustainable tourism as an adaptive paradigm. *Annals of Tourism Research*, 1997, 24 (4) : 850–867.

Design of earthquake-data processing platform

S.Y. Zhang

Key Lab of Mathematics and Applied Mathematics, School of Mathematical Sciences, Peking University, China

Y. Li & Y. Zhao

School of Mathematical Sciences, Peking University, China

ABSTRACT: Earthquake often caused extremely high damages with wide range, so how to efficiently acquire, analyze and process the disaster information is of important significance. Nevertheless, integrating collected data and generating a preliminary analysis are not that easy. In this paper, we designed and developed an earthquake disaster information process platform with groovy language under the Grails framework, which is able to collect disaster information and give visual analysis at the first time, helping professionals get a precise picture and sending them proper assistance.

1 INTRODUCTION

In recent years, natural disasters happened frequently both at home and abroad [1]. In china, the Wenchuan earthquake brought nearly seventy thousand deaths [2]. Earthquake always happens instantaneously, and the aftermath of an earthquake is expected to be difficult. Earthquake often caused extremely high damage with wide range, and how to quickly collect, analyze and process the information on the disaster situation is of vital significance to earthquake emergency commanding and disaster mitigation.

Therefore, we designed and developed earthquake disaster data processing system based on the public service number 12322. The system was mainly developed with groovy language under the Grails framework, and was written by C#, by using ArcGIS Desktop, and several other functions are supplied by other systems. These components interact with each other via SOAP style web services.

2 EARTHQUAKE DISASTER DATA PROCESS PLATFORM

2.1 Function framework

According to the characteristics of earthquake disaster, for acquiring seismic disaster information analysis and emergency needs, combined with computer technology and geographic information technology, we designed the framework of the overall function of the system which is shown in Figure 1.

Earthquake Disaster Information Process System function includes the following 5 parts.

Guiding the disaster information input and inquiries according to rules and knowledge provided by the knowledge base.

Obtaining Global Position System (GPS) coordinates using 3rd-party GIS plug-ins.

Data filtering with Bayes method to remove some false information (according to experience in Wenchuan Earthquake, these false information are very large).

Drawing a statistical chart by different axis such as time, intensity, etc.

Making interpolation analysis on earthquake intensity and drawing.

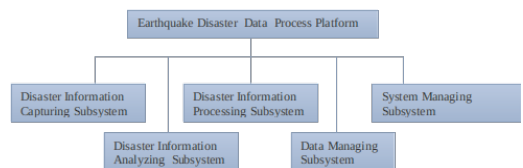


Figure 1. Framework of earthquake disaster information system.

2.2 System structure

The hierarchical structure design method is adopted in our system, and it is divided into four layers (shown in Figure. 2).

Data Layer

Data Layer includes database software, Hibernate the data persistence plug-in, and a series of uniform data operating interface including ArcSDE.

Application Layer

The main functions of our system are coded in this layer, including disaster information capturing module, disaster information analyzing module; data managing module and system managing module.

Interface Layer

The interface layer includes a series of standard SOAP web-services and clients for external modules.

External Module Layer

Several function modules are realized as external modules, such as GIS analyzing module. Additionally, the presentation function is also included in this layer.

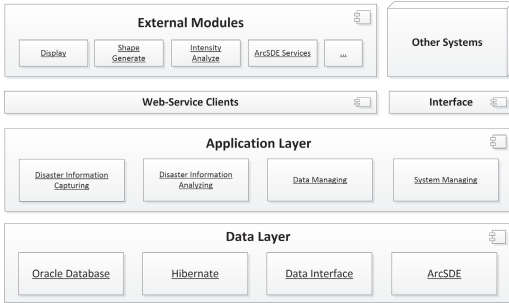


Figure 2. Hierarchical structure.

3 DISASTER INFORMATION ACQUISITION

3.1 Disaster information report

In this part, our system provides the data acquisition services such as the emergency, emergency inquiry, daily report, daily inquiry, and daily consultation. Telephone access or agents active click links can trigger disaster information reporting interface.

The questionnaire parts of this interface are generated by the knowledge base automatically. The layout of the questionnaire is expansion, which is convenient to do guided inquiry. Select to open one class, and you can see the details of data.

Input data in the disaster site can trigger the disaster location window. The system can search the data user filled and user can locate the disaster spot.

3.2 Disaster information processing

Information preprocessing is conducted every time the disaster information is reported. Information preprocessing includes screening and identifying processes. Screening is to check whether there is data in the database close to the disaster information. If the answer is yes, then the new data would be deleted. Identifying is correction of false disaster information. Using Bayes method, we can calculate the probability of some events in particular earthquake felt. So we can get a conclusion whether the disaster information is reliable.

3.3 Disaster information searching and statistics

Figure 3 is a line chart according to the event time sequence of the intensity statistics respectively. The x-axis represents the event time while the y-axis represents the intensity.

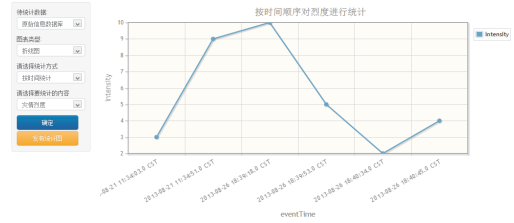


Figure 3. Line chart of intensity (changing over time).

3.4 Disaster information analysis

The scatter plot of disaster information is plotted based on all the screening disaster information. The pictures are saved as shape layers. The flowing two figures are examples of scatter plot with and without map respectively.

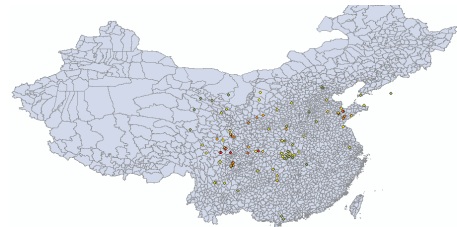


Figure 4. The scatter plot of disaster information.

3.5 Disaster information interpolation analysis

We use Thiessen polygonal method to do the interpolation analysis.

Define the discrete disaster point to be Dot (x, y, z) , where x is the longitude of the discrete disaster points, y is the latitude of the discrete disaster point, and Z is the information attribute of the discrete disaster point. We use $Dot(x, y, z)$ to build Thiessen polygonal through the following four steps.

The discrete points construct a triangulation network automatically.

Recode the id of all the adjacent triangle of each point.

Calculate the circumcircle center of each triangle.

Connect the circumcircle centers of the adjacent triangles to obtain the Tyson polygon.

After building the Thiessen polygonal, the update of the outer and inner boundary is needed to get smooth results.

In our system, in the early earthquake, we fit the disaster intensity for the first time by using the natural neighbor method combined with the packet disaster information and message records obtained from B4. Fitting results are rounded to 1-12 degree. And results would be provided for other packages for reference.

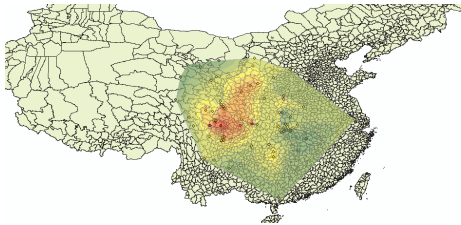


Figure 5. Result of interpolation analysis.

Thematic map is divided into 5 categories. The building damage thematic map.

Secondary disasters of thematic map include collapse, landslide, debris flow, flood, and fire for 5 map layers.

The lifeline thematic map includes water supply system, drainage system, oil system, fuel gas system, power system, communication system and water conservancy project map layers.

Special engineering damage thematic map includes space base, aboveground segment, nuclear power station, reservoir, and oil for 6 map layers.

The rescue requirements thematic map include tents, food, water, clothes map layers.

Each thematic map draws thematic information according to the disaster information description by using different legend.



Figure 6. The disaster thematic map.

4 DESCRIPTION OF KEY TECHNIQUES

4.1 Query information knowledge base

Because of the objective situation of different provinces, the earthquake reports in the project are also different. So we add the query knowledge into the system in the form of a knowledge database. Besides,

users with administrator permissions are allowed to manage the knowledge base.

The query information knowledge base includes common state knowledge base and emergency state knowledge base. The two knowledge bases are different in datasets, but similar in database structure, such as,

Item name: only used to display.

Item types: single choice, multiple choice, number, text, etc.

Item description: combination of string according to certain rules. For example, in single choice the item description is Option 1 | Option 2 | Option 3|...|Option n.

Item notes.

When the users enter the emergency report, emergency query or daily report, and daily query interface, the system will generate a table for users automatically according to the current query information knowledge base state. Then the data will be integrated into a string for storage when users submit.

4.2 The unified interface management

With WebService + XML, the data is transferred through the unified B6Report interface. The data transmission process includes the following three parts.

Earthquake triggering or releasing the trigger. Information transferred from bus can show our system that the earthquake happened. Then our system is able to switch to emergency state automatically. At the same time, all telephone access can pop up the emergency window while the reported data automatically marks the number of the corresponding earthquake.

Query of System Status. Bus periodically sends messages to each subsystem to check the working state of each subsystem.

Data Synchronization and Data Transmission. Each province synchronizes data with the State General Administration. The system receives short message data from other systems.

In the traditional system, all the services use the same interface which makes the larger distributed systems very complicated. In the traditional system, each subsystem needs to know exactly the details and the call mode of the other subsystem interface. So we use the unified interface management, in which each subsystem uses the unique WebService interface. And the different service is reflected through the parameters of XML.

At the same time, each subsystem only communicates with the bus unilaterally. If one subsystem tries to communicate with other subsystems, the information will be switched by the bus through an XML parameter. In this way, we just need the bus know the details of each subsystem interface. The WebService is built by using SOAP and CXF plug-in.


```

This WSDL file does not appear to have any style information associated with it. The document tree is shown below.
<?xml version="1.0" encoding="utf-8"?><wsdl:definitions xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/" name="UnifiedInterface" targetNamespace="http://schemas.xmlsoap.org/wsdl/unifiedinterface/"><wsdl:import namespace="http://schemas.xmlsoap.org/wsdl/unifiedinterface/" location="http://schemas.xmlsoap.org/wsdl/unifiedinterface/"></wsdl:import><wsdl:import namespace="http://schemas.xmlsoap.org/wsdl/unifiedinterface/" location="http://schemas.xmlsoap.org/wsdl/unifiedinterface/"></wsdl:import><wsdl:import namespace="http://schemas.xmlsoap.org/wsdl/unifiedinterface/" location="http://schemas.xmlsoap.org/wsdl/unifiedinterface/"></wsdl:import></wsdl:definitions>

```

Figure 7. The WSDL Description of Unified Interface.

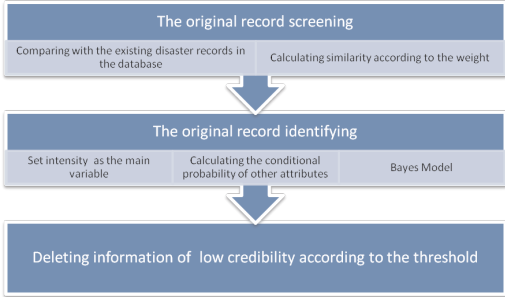


Figure 8. The flow chart of screening and identifying process.

The following is the web description of the interface (WSDL).

4.3 Disaster screening and identifying model

Disaster screening is processing the obtained disaster information, removing duplication, deleting obvious error information and getting effective disaster information. Disaster identifying is calculation of the credibility of disaster information after the screening.

Disaster Information Screening Process

Compare the main attributes (including time, position, intensity, etc.) of disaster information

If the disaster information has the same main attribute to the information in knowledge base, it can be treated as repeated information and be deleted.

If the main attribute is consistent and the reporter is the same, the new information would update the old one.

If the system cannot judge the similarity of the main attribute, then the detailed description of disaster information would be considered.

Make a similarity test for each description field of disaster information and scoring. Scoring is based on the type of description field. For example,

$$\text{DetailMark}(a, b, \text{Property}) =$$

$$\frac{\text{GetDelta}(a.\text{Property},b.\text{Property},\text{Property.type})}{\text{GetMaxDelta}(\text{Property})}$$

Calculate weighted average value of similarity score according to the similarity score in the database.

$$\text{Mark}(a,b) = \frac{1}{\sum i.Weight} \sum_{i \in \text{Properties}} \text{DetailMark}(a,b,i) \times i.Weight$$

If the similarity score is higher than a threshold, then the information would be determined to repeated information.

Disaster Information Identifying Process

Disaster information is made up of time, position, intensity and the description information.

Set $x[i]$ to be the i -th description field, and R to be the intensity which is one of the main attributes. An item of disaster information can be described as

$$\text{Inf} = (x[1],x[2],\dots,x[n],R).$$

According to the Bayes conditional probability model, the credibility of Inf can be described as $P(x[1],x[2],\dots,x[n]|R)$.

$$P(x[1],x[2],\dots,x[n]|R) = \frac{\text{Amount}(x[1],x[2],\dots,x[n], R)}{\text{Amount}(R)}$$

Here, **Amount(I)** presents the number of records in the database of I events. Considering the limited amount of data in our system, we simplified the Bayes model.

In Native Bayes, value attributes are assumed to be conditionally independent given the target value. So the last equation can be simplified as

$$P(x[1],x[2],\dots,x[n] | R) = P(x[1] | R) \times P(x[2] | R) \times \dots \times P(x[n] | R)$$

This makes the calculation much simpler.

4.4 The application of GIS system

We use the GIS function frequently in the system in a different form.

In the data entry interface, we call Baidu map API. Using Baidu map, we search the spot the user input and then mark the near positions automatically.

After the user selects the correct position, we will show the latitude and longitude information. The following figure is one example of using Baidu map.

In the scatter plot, the interpolation analysis and thematic map, we use the Open-Layers to read the map, and then indicate the plotting or analysis results on the map. The system supports WMS map layers provided by provincial or national center or a free map network directly provided by Open-Layers as a base map. These parameters can be changed in our system. Figure 9 is an example.



Figure 9. WMS layers (loaded by OpenLayers Plugin).

In the disaster interpolation analysis, since the GIS (in this system is JavaScript/Flex) is unable to complete the interpolation analysis, we need to call the ArcGIS Desktop to complete calculation. We use a computer with ArcGIS Desktop calculate the interpolation analysis through C#. Then the system will call the interpolation results and display them on the map.

5 CONCLUSION

In this paper, we raised an on-line hot-line system which is able to acquire and arrange disaster-data, make preliminary analysis and intuitively display the result.

During the development period, we learned a lot from the last generation of the disaster collecting system (the 12322 platform), and made considerable progress according to the advice from Seismological Bureau of various regions. These make our platform more practical. We are happy to see if the platform would contribute to earthquake relief efforts and help decrease casualties.

REFERENCES

- [1] Abuhadrous, S. Ammoun, F. Nashashibi, F. Goulette, and C. Laugeau, "Digitizing and 3D modeling of urban environments and roads using vehicle-borne laser scanner system," Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 76–81, 2004.
- [2] S.K. Lodha, E.J. Kreps, D.P. Helmbold, and D. Fitzpatrick, "Aerial LiDAR Data Classification Using Support Vector Machines (SVM)," Proc.

- IEEE Symp. 3D Data Processing, Visualization, and Transmission, IEEE Press, 2006, pp. 567–574, doi: 10.1109/3DPVT.2006.23.
- [3] X. Huang, H. Li, X. Wang, and F. Zhang, "Filter algorithms of airborne LiDAR data: review and prospects," Acta Geodaetica et Cartographica Sinica, vol. 38(5), 2009, pp. 466–469.
- [4] H. Zhao and R. Shibasaki, "Reconstructing a textured CAD model of an urban environment using vehicle-borne laser range scanners and line cameras," Machine Vision and Applications, vol. 14(1), 2003, pp. 35–41.
- [5] H. Zhao and R. Shibasaki, "A vehicle-borne urban 3-D acquisition system using single-row laser range scanners," IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 33(4), pp. 658–666, 2003.
- [6] M. Dinesh and S. Ryosuke, "Auto-extraction of urban features from vehicle-borne laser data," Symposium on Geospatial Theory, Processing and Applications, Ottawa, 2002.
- [7] B. Grinstead, A. Koschan, D. Page, A. Gribok, and M.A. Abidi, "Vehicle-borne scanning for detailed 3D terrain model generation," SAE 2005 Transactions Journal of Commercial Vehicles, 2005, pp. 196–204.
- [8] T. Ou, X. Geng, and B. Yang, "Application of vehicle-borne data acquisition system to power line detection," Journal of Geodesy and Geodynamics, vol. 29(2), 2009, pp. 149–151.
- [9] X. Lu and L. Huang, "Grid method on building information extraction using laser scanning data," Geomatics and Information Science of Wuhan University, vol. 32(10), 2007, pp. 852–855.
- [10] B. Li, Z. Fang, and J. Ren, "Extraction of building's feature from laser scanning data," Geomatics and Information Science of Wuhan University, vol. 28(1), 2003, pp. 65–70.
- [11] D. Manandhar and R. Shibasaki, "Vehicle-borne laser mapping system (VLMS) for 3-D GIS," Proc. IEEE Symp. Geoscience and Remote Sensing, IEEE Press, 2001, pp. 2073–2075, doi: 10.1109/IGARSS.2001.977907.
- [12] C. Früh and A. Zakhor, "An automated method for large-scale, ground-based city model acquisition," International Journal of Computer Vision, vol. 60(1), 2004, pp. 5–24.
- [13] H. Zhao and R. Shibasaki, "Reconstructing urban 3D model using vehicle-borne laser range scanners," Proc. IEEE Conf. 3-D Digital Imaging and Modeling, pp. 349–356, 2001.
- [14] C. Früh and A. Zakhor, "Constructing 3D city models by merging aerial and ground views," Computer Graphics and Applications, vol. 23(6), 2003, pp. 52–61.
- [15] B. Yang, Z. Wei, Q. Li, and Q. Mao, "A classification-oriented method of feature image generation for vehicle-borne laser scanning point clouds," Acta Geodaetica et Cartographica Sinica, vol. 39(5), 2010, pp. 540–545.
- [16] W. Shi, B. Li, and Q. Li, "A method for segmentation of range image captured by vehicle-borne laserscanning based on the density of projected points," Acta Geodaetica et Cartographica Sinica, vol. 34(2), 2005, pp. 95–100.
- [17] H. Zhao and R. Shibasaki, "Updating a digital geographic database using vehicle-borne laser scanners and line cameras," Photogrammetric Engineering & Remote Sensing, vol. 71(4), 2005, pp. 415–424.

Activated sludge model parameters calibration based on multi-objective PSO

R.R. Wang, L.L. Cao & M.D. Liu

Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, China

ABSTRACT: The model of an Orbal oxidation ditch activated sludge process was set up based on ASM3 and Takacs's double index settlement rate of secondary sedimentation tank model in this paper. According to the condition of the multi-objective output simulation model, multi-objective particle swarm optimization algorithm (MOPSO) is used to calibrate part of the high sensitivity parameters in the activated sludge process model. Simulation results show that the effectiveness of MOPSO for parameter calibration is obvious, which can further improve the accuracy of the model.

1 INTRODUCTION

Mathematical model plays a very important role in the study of sewage treatment in activated sludge process. A reliable mathematical model can predict the actual water composition accurately in wastewater treatment plant. International Association of Water Quality (IAWQ) put forward the activated sludge model no.3 (ASM3) in 1999, forming an ASM series model based on a process description of microbial reactions with ASM1 and ASM2. But the promotion and application of the activated sludge mathematical model still have many problems, especially when model parameters are involved. The ASM series model involves a large number of chemometrics and kinetic parameters which have recommended values given by IAWQ. But due to different environment, using these default values will result in reduced accuracy of the model. At present, the widely used method of parameter correction are experimental assays, expert experience method, sensitivity analysis, process engineering method and system parameter identification method.

Typical values of chemometrics parameter are not affected by temperature, but for different wastewater composition, the values will be different. Hence researched readily biodegradable and nitrogen degradation process of wastewater, especially the process of biodegradation in activated sludge model, and proposed breathing measurement to measure the value of chemometrics parameter.

On behalf of biochemical reactions activity, the sensitive reaction kinetics parameter can be corrected based on sensitivity analysis. Experimental measurement, the experimental data fitting method and

system parameter identification method are the main methods of Sensitive parameter correction. For example Torrijos considered using breathing measurement measured maximum respiratory rate of autotrophic bacteria.

Mathematical methods also have a wide range of applications in the activated sludge model correction: Ayesa adopted the Kalman filter algorithm to calibrate activated sludge model parameters; Kabouris used the maximum likelihood method in activated sludge model parameters identification; Li W. Z. summarized the basic commonly used parameters checking methods of the activated sludge model, and introduced the application of genetic algorithm in the optimization of these parameters. Yao C. used genetic algorithm to determine part of the parameters of the simplified SFBR model;

Orbal oxidation ditch pond activated sludge simulation model is established in this paper based on ASM3 and Double index sedimentation rate model of the secondary sedimentation tank, and multi-objective particle swarm optimization (MOPSO) algorithm has been used to revise part of the high sensitivity parameters in the activated sludge process model.

2 THE ESTABLISHMENT OF THE ACTIVATED SLUDGE MODEL

2.1 *The design parameters of activated sludge process*

An activated sludge process using the Orbal oxidation ditch craft can process 25,000 m³ water one day. Process flow diagram is shown in [Figure 1](#). The activated sludge process is made up of three concentric

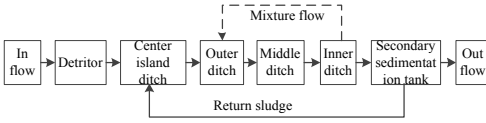


Figure 1. The flow chart of Orbal Oxidating ditch.

elliptical channels. After though the center island ditch first, the waste water inflows into the outer ditch, middle ditch in turn and finally outflow from the inner ditch.

2.2 Inflow composition structure

The conventional monitoring data of this activated sludge treatment process are COD, TN and S_{NH_4} , which need to be converted into the thirteen components' data of ASM3 correspondingly. The conventional water quality data of the activated sludge treatment process is obtained from literature [10]. Choose on August 6~29, continuous monitoring data of 6 AM (twenty four groups totally) as the basic data of dynamic simulation and correction in this article as shown in Figure 2.

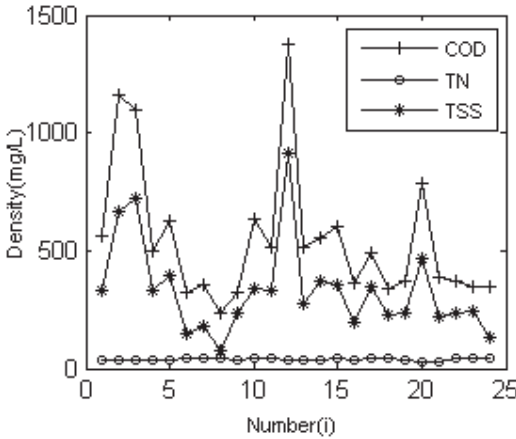


Figure 2. The influent water curves.

2.3 The establishment of the activated sludge process models

Compared to ASM1, the emphasis of the reaction process in ASM3 is changed from hydrolysis to the storage in organic substrate, also including intracellular storage product.

Through the analysis of oxidation ditch process, four completely mixed bioreactors (CSTR) compose anaerobic-anaerobic- anoxic-aerobic reactor in series system. Twelve biochemical reaction sub

processes and thirteen wastewater components are defined in ASM3. Each reaction has several kinds of components involved in each process; meanwhile each component is involved in several sub processes. For a single steady reaction pool, the law of material conservation exists in the system. The secondary sedimentation tank adopts Takacs double index sedimentation rate model. Combining the above component's material balance equation of four biological reaction pool and the second pond can get 62 differential equations, because there are 13 equations in each biological reaction pool and 10 equations in second pond. ODE, which is a system function provided by MATLAB can be used to solve the differential equations and can work out the effluent component concentration according to the concentration of influent.

Transform the data in figure 2 into the input components of ASM3 model correspondingly. Then convert the model calculation results obtained by the dynamic simulation reversely into conventional water component concentration, and compare it with the actual monitoring values.

As a parameter measuring the error of dynamic simulation, the relative average error between monitoring value and simulation value is shown in Equation 1 below.

$$e_j = \frac{1}{n} \sum_{i=1}^n \left| \frac{Q_{ji} - Q_{jio}}{Q_{jio}} \right| \quad (1)$$

where e_j = relative average error; Q_{jio} = monitoring value of effluent; Q_{ji} = simulation value of the model; n = number of data; and j = index of monitoring variable.

According to the results of dynamic simulation, COD's error is 31.66%; TN's error is 29.22%, which indicates better results of dynamic simulation and model effluent quality change trend is closer to reality; S_{NH_4} 's error is 106.19%, that is to say the actual water quality change trend is bad, the dynamic simulation result is poor, therefore it needs to be dynamic calibrated for sensitive parameters in the model.

3 MOPSO INTRODUCTION

Particle Swarm Optimization (PSO) is an optimization algorithm put forward by Kennedy and Eberhart in 1995. Due to easy understanding and implementation, PSO has been widely used in engineering.

Usually a multi-objective optimization problem can be expressed as Equation 2 below:

$$y = f(x) = (f_1(x), f_2(x), \dots, f_n(x)), \quad (2)$$

s.t. $g_i(x) \leq 0;$

where x = decision vector; y = target vector; $f_i(x)$ = objective function; and $g(x)$ = system constraints.

In most cases, the objective function may conflict with each other, so the multi-objective optimization problem probably do not possess the only global optimal solution to meet all objective function optimal at the same time, however, there can be such a special solution: for one or a few objective functions they could not be further optimized, and for the others they keep away from further degradation. Such solution is called the Pareto optimal.

3.1 Updating of Population update

In MOPSO, particle update itself includes the current speed and position based on the individual optimal particle and global optimal particle which was selected in non-inferior set randomly. Particles update formula is as follows:

$$V^{k+1} = \omega * V^k + c1 * r1 * (P_{id}^k - X^k) + c2 * r2 * (P_{gd}^k - X^k) \quad (3)$$

$$X^{k+1} = X^k + V^{k+1} \quad (4)$$

where k = current iteration; w = inertia weight; P_{id}^k = individual optimal particle; P_{gd}^k = global optimal particle; V = speed of the particle; and X = position of the particle. $c1$ and $c2$ are constant; $r1$ and $r2$ are random number between 0 and 1.

3.2 Update individual optimal particle

According to the dominance relationship of new particle generated by updating of population and the individual optimum particle at present, the individual optimal particle updates in every iteration.

When the new particle is superior to the current individual optimal particle in all objectives, select the new particle as a current individual optimal particle. When it can't decide which particle is dominated, then select a particle randomly as the new individual optimal particle.

3.3 Filtrate non-inferior set

The non-inferior set filtration mainly contains initial screening and updating of the non-inferior set.

Initial screening of the non-inferior set: after the particle initialized, put the particle into the non-inferior set when the particle is not controlled by the other particles, and select a particle in non-inferior set randomly as global optimal particle before updating of population.

Update non-inferior set is divided into two steps. First, obtain a merged set by combining the new individual optimal particle and non-inferior set of previous generations. Second, generate the new non-inferior set based on the dominance relations in the new merged set.

4 PARAMETER SENSITIVITY ANALYSIS

Sensitivity is used to measure the influence on target when relevant parameter is changed.

Define the sensitivity as below:

$$S_j^i = \left| \frac{Y_{i1} - Y_{i0}}{Y_{i0}} / \frac{P_{j1} - P_{j0}}{P_{j0}} \right| = \left| \frac{\Delta Y_i}{Y_{i0}} / \frac{\Delta P_j}{P_{j0}} \right| \quad (5)$$

where Y_i = concentration of outflow; P_j = relevant parameter; and S_{ji} = sensitivity of Y_i to P_j .

Assume initial P_{j0} is a given value, Y_{i0} is the corresponding value of P_{j0} , getting P_{j1} by giving parameter P_{j0} a 10% increase, Y_{i1} changes from Y_{i0} to Y_{i1} correspondingly. Calculate each sensitivity of concentration target respectively. The experimental results show that each target selected corresponding to the first five large sensitivity parameter calibration can get more precise mathematical model. The biggest five parameters to COD are: $f_{SI} > b_{H, O2} > Y_{H, O2} > Y_{STO, O2} > K_H$. The biggest five parameters to TN are: $b_{H, O2} > b_{H, NO} > Y_{STO, NO} > \mu_A > Y_{H, O2}$. The biggest five parameters to S_{NH} are: $K_{A, O} > \mu_A > Y_A > b_{A, O2} > K_{A, NH}$, as shown in Table 1. From discussed above, parameters required correction are: K_H , $b_{H, O2}$, $b_{H, NO}$, μ_A , $K_{A, NH}$, $K_{A, O}$, $b_{A, O2}$, f_{SI} , $Y_{STO, O2}$, $Y_{STO, NO}$, $Y_{H, O2}$, Y_A .

Table 1. Parameter sensitivity analysis results.

| Parameter | COD | TN | S_{NH} |
|---------------|--------|--------|----------|
| K_H | 0.0110 | 0.0066 | 0.0877 |
| $b_{H, O2}$ | 0.1065 | 0.0712 | 1.1109 |
| $b_{H, NO}$ | 0.0022 | 0.0482 | 0.2926 |
| μ_A | 0.0058 | 0.0173 | 5.9710 |
| $K_{A, NH}$ | 0.0005 | 0.0008 | 1.8313 |
| $K_{A, O}$ | 0.0028 | 0.0049 | 6.0987 |
| $b_{A, O2}$ | 0.0021 | 0.0042 | 2.7599 |
| f_{SI} | 0.1471 | 0.0010 | 0.0304 |
| $Y_{STO, O2}$ | 0.0171 | 0.0126 | 0.2656 |
| $Y_{STO, NO}$ | 0.0010 | 0.0355 | 0.1962 |
| $Y_{H, O2}$ | 0.0454 | 0.0152 | 0.8277 |
| Y_A | 0.0020 | 0.0034 | 3.7318 |

5 PARAMETER CALIBRATION BASED ON MOPSO

The MOPSO algorithm procedure used in activated sludge process dynamic calibration is given in Figure 3.

Population size was set as 30. w , a adaptive inertia factor was designed as below:

$$w = wmax - (wmax - wmin) \frac{iter}{MaxIt} \quad (6)$$

where $MaxIt = 100$, means maximum iterations; $c1 = c2 = 0.8$, means particle learning factor; $wmax = 1.2$, means a maximum of inertia factor; $wmin = 0.1$, means a minimum of inertia factor; and $iter =$ current evolution number.

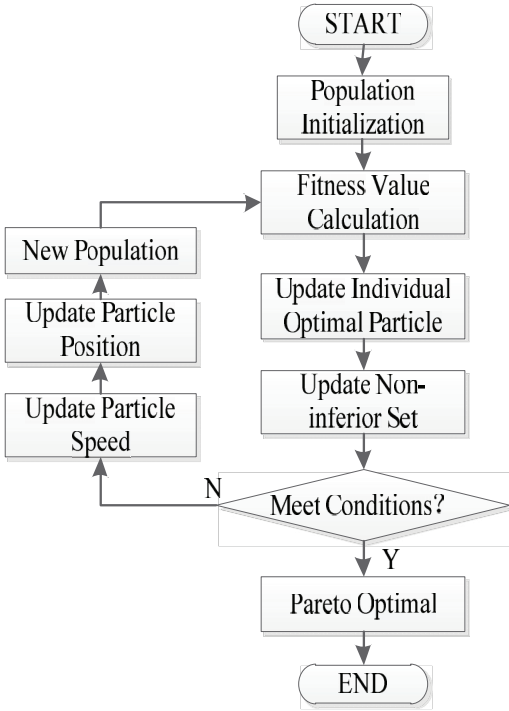


Figure 3. Flow chart of parameter calibration.

After calibration, a final non-inferior set will contain many non-inferior solutions. Because lacking of global optimal solution and for the convenience of analysis algorithm optimization performance, the average of the non-inferior set in each generation is calculated and assumed to be the contemporary global optimal solution.

The definition ϵ is the total weighted error of relative average error of the contemporary global optimal solution as below:

$$\epsilon = \sum_{j=1}^m a_j e_j \quad (7)$$

where $m = 3$, means the number of monitoring index; $a_j =$ weight coefficient. In process, $a_1 = a_2 = a_3 = 1$.

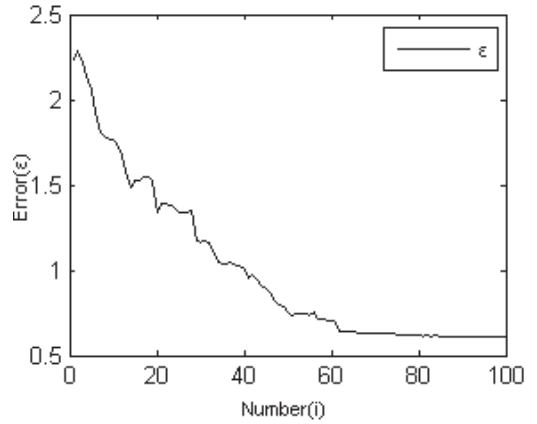


Figure 4. The average evolution curve.

Execute algorithm calibration independently 50 times. The final result is shown in Table 2 and the average curve trend of evolution is shown in Figure 4.

Table 2. The final results of MOPSO executed 50 times.

| Optimal value | Average value | Variance |
|---------------|---------------|----------|
| 55.86% | 61.17% | 0.0164 |

Continue to carry on dynamic model simulation with parameters which have been calibrated. Error of model outflow target before and after parameters calibrated is shown in Table 3.

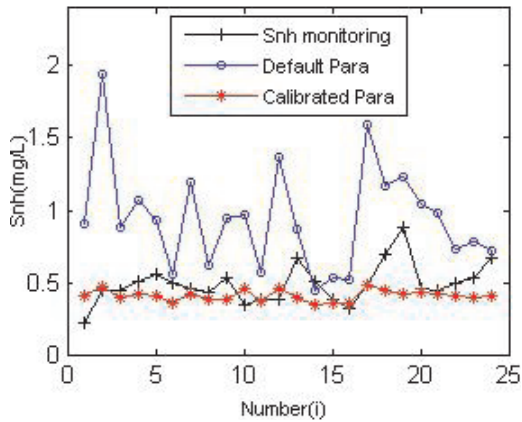
Table 3. The comparison of outflow error.

| Parameter | e (COD) | e (TN) | e (S_{NH}) | ϵ |
|------------|-----------|----------|------------------|------------|
| Default | 31.6% | 29.2% | 106.1% | 167% |
| Calibrated | 14.9% | 18.8% | 23.1% | 57% |

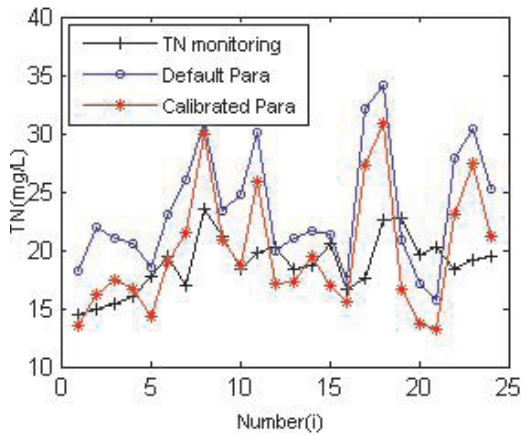
Because the model error of TN with default parameters is small, the decline space of error is limited. But after calibration the result is still lower than the default parameter error. With a difference, the error of COD and S_{NH} fall a great deal after calibration.

Figure 5 contains an outflow value curve comparison between monitoring, the simulation with parameters of default and calibrated.

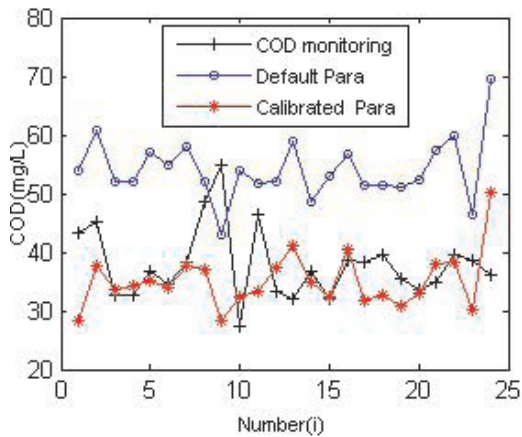
From Fig. 5, the model outflow value trend basically is consistent with the monitoring value after calibration, especially the COD and S_{NH} . Simulation effect is improved greatly through MOPSO parameter calibration according to the simulation result.



(a) COD



(b) TN



(c) S_{NH}

Figure 5. The outflow value curve comparison.

Many reasons can result in the relative error, including the flaws of the model itself, saltation of temperature or pH, hydraulic load mutation, uncontrollable external factors, etc. The uneven dissolved oxygen in aeration pool also can produce certain error by affecting nitrification reaction.

6 CONCLUSION

This paper established an activated sludge model based on Orbal oxidation ditch process and calibrated the sensitivity parameters through multi-objective particle swarm optimization algorithm. Simulation results show that the algorithm introduced into the activated sludge model parameter calibration is feasible, after correction the model outflow value can well approximate monitoring, further improved the accuracy of the model. The work in this paper can also be extended to other complex biochemical reaction model parameter calibration.

ACKNOWLEDGMENTS

Supported by “the fundamental research funds for the central universities”.

REFERENCES

- [1] Henze M, Gujer W, et al. Activated Sludge Model No.3 [J]. Water Science Technology, 1999, 39(1).
- [2] Gan L. J., Liu J. Y., Gu G. W., et al. Measurement and Calculation for Stoichiometric Coefficients in ASM1[J]. China Water & Wastewater, 2003, 19(1): 80–81.
- [3] Duan H. F., Yu G. P. A study on method of determining basic parameters in activated sludge models[J]. J. of University of Science and Technology of Suzhou, 2005, 18(3): 41–43.
- [4] Henze M. Characterization of wastewater for modeling of activated sludge process[J]. Water Science and Technology. 1992, 25(6): 1–15.
- [5] M. Torrijos, R.M. Cerro, B. Capdeville, et al. Sequencing Batch Reactor: A tool for wastewater characterization for IAWPRC model[J]. Water Science and Technology. 1994, 29(7): 81–90.
- [6] E. Ayesa, J. Florez, J.L. Garcia-Heras, et al. State and coefficients estimation for the activated sludge process using a modified Kalman filter algorithm[J]. Water Science and Technology. 1991, 24(6): 235–247.
- [7] E. Ayesa, J. Florez, J.L. Garcia-Heras, et al. State and coefficients estimation for the activated sludge process using a modified Kalman filter algorithm[J]. Water Science and Technology. 1991, 24(6): 235–247.
- [8] Li W. Z., Jin X. Y., Niu L. B. Basic parameters and calibration of activated sludge model[J]. Journal of North China Institute of Water Conservancy and Hydroelectric Power, 2010, 31(2): 87–90.

- [9] Yao C., Shao Y. Y., Yang Z. L. Application study on the activated sludge model parameters determination[J]. *Guangdong Chemical Industry*, 2009, 36(3): 60–62.
- [10] Jie D. L. Application research of ASM1 to Wuxi Chengbei WWTP[D]. Nanjing: Hohai University, 2007
- [11] Takacs I, Patry G G, Nolasco D. A dynamic model of the clarification-thickening process [J]. *Wat, Res.*, 1991, 25(10): 1263–1271.
- [12] R Eberhart, J Kennedy. A new optimizer using particle swarm theory. In: Proc of the 6th Int'l Symposium on Micro Machine and Human Science. Piscataway, NJ: IEEE Service Center, 1995. 39–43.
- [13] J Kennedy, R Eberhart. Particle swarm optimization. IEEE Int'l Conf on Neural Networks, Perth, Australia, 1995.
- [14] Cheng S. T., Chen Y. L. Environmental System Analysis[M]. Beijing: Higher Education Press, 1990.

Optimization of hexavalent chromium (Cr(VI)) reduction by granulated nano zero-valent iron (nZVI)

Y.J. Shih, C.W. Chen, K.F. Hsia & C.D. Dong

Department of Marine Environmental Engineering, National Kaohsiung Marine University, Nanzih District, Kaohsiung City, Taiwan

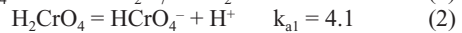
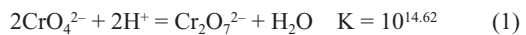
ABSTRACT: The granulated nZVI is used to totally remove Cr(VI) from solution. The kinetics indicates that the rate of Cr(VI) reduction is optimized at a specific nZVI dose even though Cr(VI) decreases with increasing nZVI dose. The increment of aggregation and the surface precipitation which inhibits the efficacy of nZVI is explained through pH effect and SEM/EDS analysis.

1 INTRODUCTION

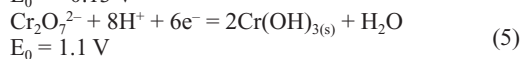
Zero-valent iron (ZVI), Fe^0 , is a highly reactive material used for remediating a variety of pollutants. Because ZVI has much lower standard reduction potential ($\text{Fe}^{2+} + 2e^- = \text{Fe}^0$, $E^0 = -0.447 \text{ V}$; $\text{Fe}^{3+} + 3e^- = \text{Fe}^0$, $E^0 = -0.037 \text{ V}$), elemental iron is thermodynamically stable in oxide forms under normal conditions. ZVI is capable of directly reducing different types of metal ions such as As(III), Pb(II), Cu(II), Ni(II), Co(II) and Cr(VI) during getting oxidized very easily (Zhang, 2003; Sun, 2006; Noubactep, 2008). The pathways for reducing contaminants involve: (a) adsorption of pollutants, (b) direct reduction at the surface of ZVI, and (c) catalytic reduction through Fe(II) at the interface of ZVI and target products. The nanoscale ZVI particles (Cao, 2005) demonstrate high surface-to-volume ratios and high surface activity. However, the adoption of nZVI displays a drawback which is the aggregation in time finally decreases their activity (Noubactep, 2012). In the past few years, researchers have focused on stabilizing nZVI and reduction of the cost, thus expanding their utilization in environmental technologies. As reported recently, the reduction efficiency of nZVI ranked in an order: $\text{As} > \text{Cr} > \text{Cu} > \text{Hg} > \text{Pb} > \text{Zn} > \text{Cd} > \text{Ni}$ (O'Carroll, 2013). Cr(VI) can especially be reduced with ZVI by a series of mechanisms, reduction, adsorption, co-precipitation.

The transformation, solubility, mobility, and consequent toxicity of hexavalent chromium (Cr(VI)) can be governed by redox reactions, precipitation/dissolution reactions, and adsorption/desorption phenomena. The mobility and toxicity of Cr in the environment are strongly dependent on its oxidation

states. Cr(VI) is very toxic and highly soluble and mobile in soils, whereas Cr(III), an essential nutrient, is relatively non-reactive and forms relatively insoluble oxide and hydroxide compounds (Jacobs, 2005). Due to the mobility and toxicity of Cr being subject to its redox state, remediations using ZVI to reduce Cr(VI) are of great interest. Dichromate ($\text{Cr}_2\text{O}_7^{2-}$) and chromate (CrO_4^{2-}) salts predominate as the Cr(VI) species in the soil, groundwater, and rocks. The equilibrium and hydrolysis reactions of dichromate and chromate are:



When pH is lower than 7, $\text{Cr}_2\text{O}_7^{2-}$ is the major species of Cr(VI), which is converted to CrO_4^{2-} as pH is higher than 7. Both Cr(VI) species which are commonly found at an industrial waste site can be therefore immobilized by precipitation as $\text{Cr}(\text{OH})_3$ or incorporated into metal hydroxides.



ZVI has proven to be feasible for remediation of Cr(VI) in the groundwater in terms of batch, column experiments and pilot-scale trials (Blowes, 1992, Powell, 1995, Blowes, 1997).

The best known-strategy for removing Cr(VI) from water is to reduce it to the less toxic and immobile Cr(III). This work demonstrated the reduction of dichromate (Cr(VI)) by a commercial granulated nZVI in aqueous solution. Kinetic was carried out

for determining the effects of initial pH, and the dose ratio of Cr(VI) to ZVI on the reduction of Cr(VI). According to the proposed reaction mechanism, the crucial condition that could efficiently treat Cr(VI) was therefore optimized.

2 MATERIAL AND METHODS

2.1 Chemicals

The nZVI material used was obtained in granular form (bimodal size, 10nm and 1 μ m, purity > 99%) from GeoNano Co., Taiwan. Analytical grade potassium dichromate (K₂Cr₂O₇) for preparing Cr(VI) solution was purchased from Ferak Berlin GmbH, Germany. Diphenylcarbazone (C₁₃H₁₂N₄O) as a colorimetric agent to monitor Cr(VI) concentration was supplied by Acros Organics, USA. The deionized water used to prepare the solutions was purified with a laboratory-grade RO-ultrapure water system (resistivity > 18.3 M Ω cm).

2.2 Batch experiments

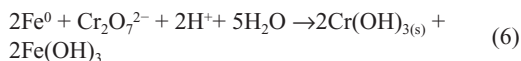
All the batch experiments were performed in three-neck flasks at 30°C keeping in the water bath incubator. 500 mL solution containing nZVI nanoparticles and certain concentration of potassium dichromate were mixed in the flasks. The suspension was mechanically stirred by a Teflon-made stirrer for 60 min with a specific initial pH. ORP (oxidation reduction potential) (WTW GmbH Co., Germany) and pH (Sensorex, USA) probes were installed for the measurement of redox potential and pH changes during batch experiments. The reaction samples withdrawn at given intervals were filtered through a 0.45 μ m membrane filter. The kinetics of the reaction was investigated under different initial Cr(VI) concentrations of 20–100 mg-Cr L⁻¹ (3.85 \times 10⁻⁴–1.92 \times 10⁻³ M) and different nZVI dose of 0.03–0.07 g L⁻¹. The effect of initial pH from 2 to 10 were also studied. The Cr(VI) concentration was analyzed by 1,5-diphenylcarbazide of 540 nm colorimetric method (3500B of Standard Methods, APHA) by using a visible spectrophotometer (Hitachi U-2001, Japan).

3 RESULTS AND DISCUSSION

3.1 Effect of nZVI dose

Figure 1 (a) depicts the Cr(VI) (3.85 \times 10⁻⁴ M) can be reduced by nZVI (initial pH 2), and its efficiency is significantly improved by increasing nZVI dose; 90% of Cr(VI) reduces rapidly in 30 min by using 0.07 g

nZVI. The redox reaction between dichromate and zero-valent iron is as below.



In addition to precipitation of Cr(OH)₃, the co-precipitation of Cr_xFe_(1-x)(OH)₃ as a shell structure of nZVI may occur resulting from the reduction of Cr(VI) (Zhang, 2012). The iron substrate is consumed in the redox reaction by oxidation to Fe(III) with the subsequent formation of mixed precipitates with Cr(III). As Fe⁰ is supposed to be a majority and unchanged during redox reaction, the kinetics of a second rate order can be approximately related to the concentration changes of Cr₂O₇²⁻:

$$-\frac{dC}{dt} = k\frac{A}{V}C^2 = k'C^2 \quad (7)$$

where C is the concentration of Cr₂O₇²⁻, A is the total surface of ZVI and V is the volume of reaction solution; therefore, the rate constant k can be an observed k' by which A and V are controlled. The integration of the pseudo second order rate equation between C at time t and initial C₀ leads to:

$$C^{-1} - C_0^{-1} = k't \quad (8)$$

Figure 1 (b) presents the observed rate constants by fitting the data of Fig. 1 (a) with the second order rate equation. The rate constant increases with increasing nZVI dose, and however there is a maximum rate constant at 0.05 g dose as the observed rate constant is normalized by the weight of nZVI (k*). This result implies that the rate of Cr(VI) reduction is moderately influenced by the amounts of nZVI use even though the extent of Cr(VI) reduction is maximized at highest nZVI dose. SEM observation (Fig. 2(a)) reveals that the particle size distribution of nZVI is of multi-model; parts of them are granulated in 1 μ m particles and others are 10 nm. It means nZVI would easily form aggregates at high dose, leading to the decline in its active surface area. Besides, the Cr compound detected by EDS elemental analysis in Fig. 2 (b) indicates the co-precipitation of iron oxide with chromium hydroxide onto the surface of nZVI.

3.2 Reaction rate of Cr(VI) reduction by nZVI

Effect of initial Cr(VI) concentrations (3.85 \times 10⁻⁴~1.92 \times 10⁻³ M) on Cr(VI) reduction efficiency are

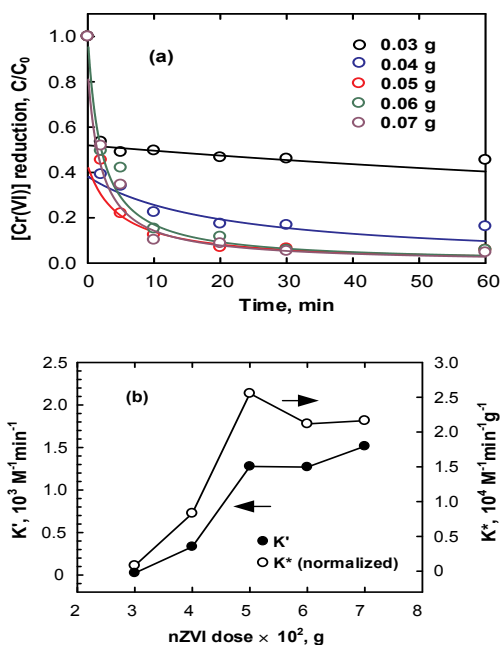


Figure 1. (a) Effect of nZVI dose (0.03–0.07 g) on Cr(VI) reduction (pH 2, [Cr(VI)] = 3.85×10^{-4} M). (b) Reaction rates (with and without normalized by nZVI dose) as a function of nZVI dose.

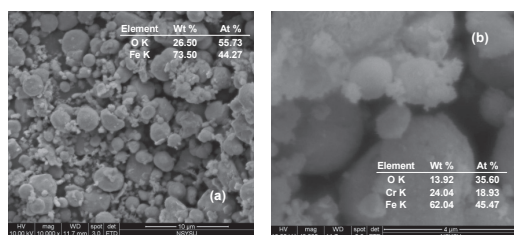


Figure 2. SEM observation and elemental analysis of nZVI before and after Cr(VI) reducing.

presented in Fig. 3(a). Cr(VI) could be absolutely removed at a concentration of 3.85×10^{-4} to 1.15×10^{-3} M using 0.05 g nZVI, while around only 60% reduces as the concentration is above 1.15×10^{-3} M. It is expectable that once the quantity of nZVI is fixed, there is limited active sites for adsorption, which then leads to low removal efficiency under a high Cr(VI) concentration. The data in Fig. 3 (a) is also described via second order reaction at constant ZVI dose as is shown in Fig. 3 (b); the observed rate constant (k') sharply decreases from around 1200 to $20 \text{ M}^{-1} \text{ min}^{-1}$ as

the initial concentration increases to 1.15×10^{-3} M. When more Cr(VI) ions approach the surface of nZVI, the metallic iron oxidizes to Fe(II)/(III) quickly, and the reduced reactivity of nZVI due to the occupation of surface site by the precipitates subsequently results in a decrease in k' . The explanation of Cr(VI) reduction strongly affected by the decrease in the active surface site should be also suitable for considering the nZVI dose as a variable. However, the observed rate constant as a function of molar ratio of initial Cr(VI) to nZVI declines much more sharply at a fixed initial Cr(VI) than at a fixed nZVI dose (Fig. 3 (c)). In other words, in addition to the competition of Cr(VI) adsorption on active sites on nZVI, the aggregation of nZVI is a factor that critically influences the reduction efficiency of dichromate.

Accordingly (Eq. (6)), the co-precipitation of chromium hydroxide and iron hydroxide occurs at a neutral pH or mild acid condition. The oxidation of nZVI might be as a cation state at relatively low pH after reducing the dichromate or chromate (Cr(VI)).

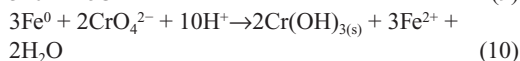
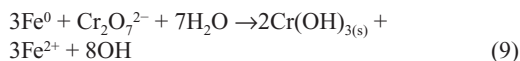


Figure 4 indicates that the efficiency of Cr(VI) reduction is obviously inhibited by increasing the initial pH. The rate constant of around $1200 \text{ M}^{-1} \text{ min}^{-1}$ at pH 2 goes steeply downward to lower than $10 \text{ M}^{-1} \text{ min}^{-1}$ as pH is higher than 4. Actually, the Cr(VI) reduction could depend mainly on adsorption and reactivity of nZVI. The pH_{pzc} of ZVI is normally around 7, which is higher than that of $\text{Fe}(\text{OH})_3$ (pH 4–5) (Lv, 2011). As the charge profile is totally altered through the surface coverage of $\text{Fe}(\text{OH})_3$ and $\text{Cr}(\text{OH})_3$, the dichromate/chromate has less chance to get close to the surface of nZVI and therefore it was difficult to be reduced. On the other hand, due to insufficient proton to prevent ferrous cation from co-precipitation with $\text{Cr}(\text{OH})_3$, nZVI has limiting fresh surface sites to react with Cr(VI), thus leading to low efficiency of reduction.

3.3 Variations in ORP during Cr(VI) reduction

Because the critical mechanisms of dichromate removal are the redox reactions between nZVI and Cr(VI), the monitoring of ORP can verify the

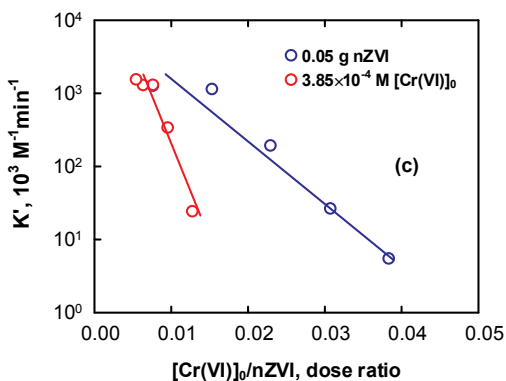
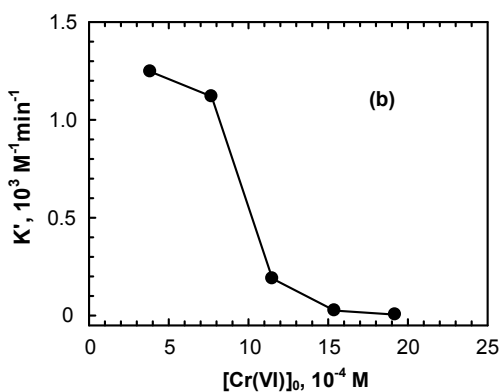
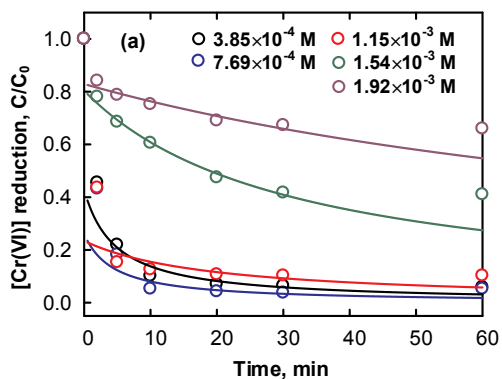


Figure 3. (a) Effect of initial dichromate concentration on Cr(VI) reduction and (b) rate constants of Cr(VI) reduction as a function of initial dichromate (3.85×10^{-4} – 1.92×10^{-3} M, nZVI = 0.05 g, pH 2); (c) comparison of rate constants as a function of the molar ratios of Cr(VI) to nZVI of between fixed nZVI dose (0.05 g) and fixed initial dichromate (3.85×10^{-4} M).

efficacy of nZVI to reduce Cr(VI). As is shown in Fig. 5 (a), there is a clear tendency that the removal of Cr(VI) is almost proportional to the ORP value; ORP indicates that the whole reaction system goes a linear trend from a more oxidation state to a reduction state for all the nZVI doses. The ORP changes normalized with the extent of Cr(VI) reduction, namely the slopes in Fig. 5 (a), are plotted against with nZVI dose in Fig. 5 (b). Notably, the normalized ORP decreases with increasing nZVI even though 0.05 g nZVI has been evidenced to most efficiently remove the dichromate of 3.85×10^{-4} M. This result implies that the largest specific surface of iron exposing to solution will result in higher reactivity to the target compound by eliminating the competitive effect of aggregation. Eventually, 0.05 g nZVI adopted for treating 3.85×10^{-4} M Cr(VI) could reach a balance between the sufficient active sites for reacting with Cr(VI) and the less probability to form an aggregation.

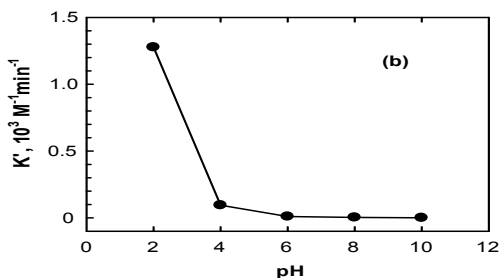
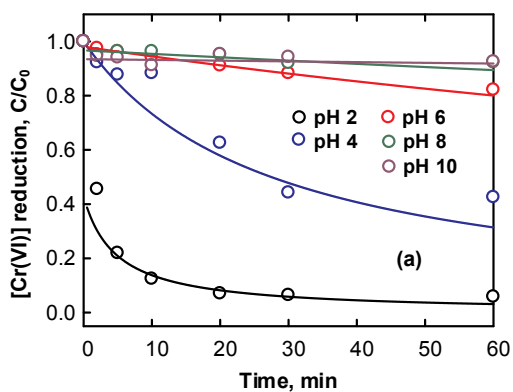


Figure 4. (a) Reduction of Cr(VI) using nZVI as a function of initial pH (Cr(VI) = 3.85×10^{-4} M, nZVI = 0.05 g, pH 2); (b) effect of pH on the rate constants of reduction of Cr(VI).

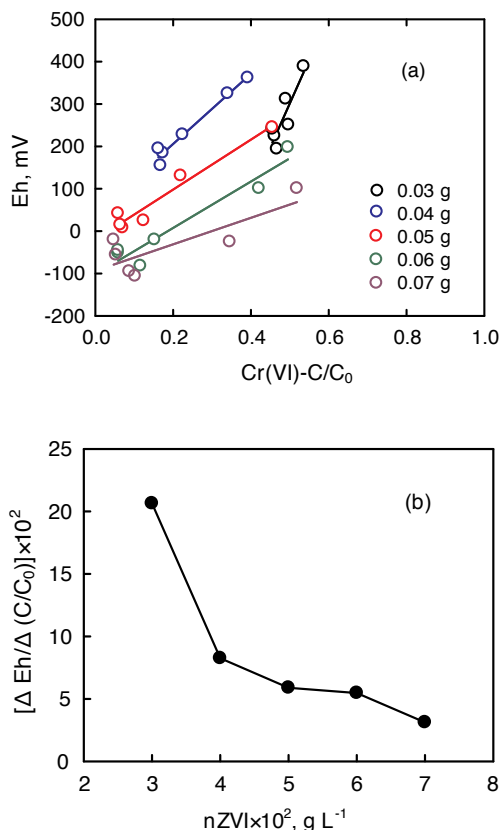


Figure 5. (a) Correlation of Cr(VI) reduction efficiency and measured ORP for various nZVI dose; (b) Eh normalized with Cr(VI) reduction efficiency as a function of nZVI dose (Cr(VI) = 3.85×10^{-4} M, nZVI = 0.05 g, pH 2).

4 CONCLUSIONS

A granulated nanoscale zero-valent iron (nZVI) is utilized to reduce Cr(VI) from wastewater. The results show that either the level of initial Cr(VI) or nZVI dose potentially influences the efficiency of Cr(VI) removal. Kinetic studies are evaluated by a pseudo second-order rate model. nZVI can reduce totally

Cr(VI) and however the rate constant decreases at high nZVI dose due to the increment of aggregation. A surface precipitation which inhibited the capability of nZVI to reduce Cr(VI) is proposed through pH effect and SEM/EDS analysis. Such granulated nZVI offers a promising alternative material for the removal of Cr(VI) ions from wastewater.

REFERENCES

- [1] Blowes, D.W., Ptacek, C.J., Jamber, J.L., 1997. In-situ remediation of Cr(VI) contaminated groundwater using permeable reactive walls: Laboratory studies. *Environ. Sci. Technol.* 31, 3348–3357.
- [2] Cao, J., Clasen, P., Zhang, W.X., 2005. Nanoporous zero-valent iron. *J. Mater. Res.* 20: 3238–3243.
- [3] Jacobs J.A., Testa S.M., 2005. Overview of chromium(VI) in the environment: background and history: Chromium(VI) handbook. New York: CRC Press: 1–21.
- [4] Lv, X., Xu, J., Jiang, G., Xu, X., 2011. Removal of chromium(VI) from wastewater by nanoscale zero-valent iron particles supported on multiwalled carbon nanotubes. *Chemosphere* 85, 1204–1209.
- [5] Noubactep, C., Caré, S., Crane, R., 2012. Nanoscale metallic iron for environmental remediation: prospects and limitations. *Water Air Soil Pollut.* 223:1363–1382.
- [6] Noubactep, C., 2008 A critical review on the process of contaminant removal in Fe⁰-H₂O systems. *Environ. Technol.* 29: 909–920.
- [7] O'Carroll, D., Sleep, B., Krol, M., Boparai, H., Kocur, C., 2013. Nanoscale zero valent iron and bimetallic particles for contaminated site remediation. *Adv. Water Resources* 51: 104–122.
- [8] Powell, R.M., Puls, R.W., Hightowr, S.K., Sabatini, D.A., 1995. Coupled iron corrosion and chromate reduction: Mechanisms for subsurface remediation *Environ. Sci. Technol.* 29, 1913–1922.
- [9] Sun, Y.P., Li, X.Q., Cao, J., Zhang, W.X., Wang, H.P., 2006. Characterization of zero-valent iron nanoparticles. *Adv. Colloid Interf. Sci.* 120: 47–56.
- [10] Zhang, W.X., 2003. Nanoscale iron particles for environmental remediation: an overview. *J. Nanoparticle Res.* 5: 323–332.
- [11] Zhang, Y., Li, Y., Li, J., Sheng, G., Zhang, Y., Zheng, X., 2012. Enhanced Cr(VI) removal by using the mixture of pillared bentonite and zero-valent iron. *Chem. Eng. J.* 185–186, 243–249.

Effect of calcination temperature and time of thermo-hydrolytic TiO₂ on photocatalytic degradation of methylene blue

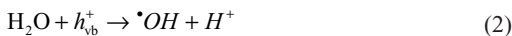
C.W. Chen, C.C. Su, Y.C. Chen & C.D. Dong

Department of Marine Environmental Engineering, National Kaohsiung Marine University, Kaohsiung, Taiwan

ABSTRACT: This study investigates the photocatalytic degradation of methylene blue (MB) in synthetic TiO₂ by a thermal hydrolysis process with titanium (IV) chloride as the precursor. The influence of calcination temperature and time on the degradation of MB was examined. SEM and XRD are used to characterize the morphology and structure of synthetic TiO₂. Results showed that TiO₂ catalyst are rutile and anatase phases with weaker crystallinity is at a low calcination temperature (300°C for 1 h), whereas a mixture of anatase and rutile phases with strong crystallinity existed in TiO₂ at a high calcination temperature (500°C for 0.5 and 1 h). The optimal operation conditions are obtained in 15 mg/L of MB, 0.37 g catalyst/L and TiO₂ calcination at 300°C for 3 h. Kinetic study illustrated that photocatalytic degradation of MB fits well in the pseudo-first order model.

1 INTRODUCTION

Titanium dioxide (TiO₂) has been studied extensively because it is well known as a semiconductor with photocatalytic activities and has a great potential for application in environmental remediation. When the photocatalyst is irradiated by UV radiation, electrons transited from the valence band to the conduction band. An electron-hole pair is generated (Rauf et al. 2011). $h_{\nu b}^+$ can react with H₂O to generate hydroxyl radical (\cdot OH); e_{cb}^- react with O₂ to generate superoxide radical anion of oxygen.



TiO₂ can be synthesized in three distinct crystalline structures: anatase, rutile, and brookite. Titania with anatase phase is the most widely used photocatalysts for oxidative decomposition of organic compounds, and an excellent photocatalyst for photodecomposition and solar-energy conversion due to its high photoactivity (Li & Zeng 2011). Thus, it is very important to develop synthetic methods in which the crystalline forms.

The promising method for the synthesis of TiO₂ is thermal hydrolysis, which requires solution chemistries involving in organic or inorganic titanium as a precursor (Chen et al. 2013). The particle size and morphology of TiO₂ can be controlled in this method, so as to manipulate the formation of TiO₂ according to the target applications. In the case of organic hydrolytic reactions TiO₂ nanoparticles obtained are crystallized primarily in the anatase phase but a minor phase of brookite could not be eliminated by tuning the reaction conditions (Yin et al. 2001, Li et al. 2004). The presence of trace amounts of brookite might have side effects on the application of anatase nanoparticles in the photocatalytic reactions and many other chemical processes. The literature on the use of powders produced by hydrolysis of TiCl₄ productions for photocatalysis is very limited. Besides, calcination is a common treatment that can be used to increase the photocatalytic activities, although it is far from clear that how it affects the physicochemical parameters which control the photoactivity.

In this study, the preparation of titania by thermal hydrolysis of TiCl₄ in a mixed solution of hydroxypropylcellulose (HPC, polymer dispersing agent) and water is studied. The objectives of this study are to investigate the effect of calcination temperature and time on the photocatalytic activity of TiO₂ in terms of the degradation of methylene blue (MB). The kinetic of MB degradation at different operating conditions is also examined. SEM and XRD are used to characterize the surface morphology and structure of TiO₂.

2 MATERIALS AND METHODS

2.1 Materials

All chemicals are purchased from various companies and used as received without further treatment: titanium chloride (TiCl_4 , purity > 99.9%, Showa Co., Japan), hydroxypropylcellulose (molecular weight ~ 100000, 99%, Sigma-Aldrich) and methylene blue ($\text{C}_{16}\text{H}_{18}\text{N}_3\text{ClS}\cdot 3\text{H}_2\text{O}$, 85%, Fisher Scientific). All the solutions are prepared with high-purity deionized water.

2.2 Synthesis of TiO_2 powder and photocatalytic performance

In the procedure, TiCl_4 is used as precursor. 3.3 mL of TiCl_4 solution is slowly dosed at 1 mL/min into 100 mL deionized water at 10°C in an ice-water bath with vigorously stirring; the solution gradually turns into semi-transparent for 10 min. Then, this solution is mixed with 1 g of HPC. After 10 min of mixing process, the mixed solution is placed into a constant temperature (70°C) bath maintain for 1 h under stirred condition. Subsequently, the solution is rested at room temperature for 2 days. Afterward precipitated titanium oxide ($\text{TiO}_2\cdot n\text{H}_2\text{O}$) is obtained and dried at 85°C under oven for overnight. Finally, TiO_2 powders are grounded into fine powder. In some case, TiO_2 powders are calcined at 300, 400, and 500°C for 0.5, 1, 2 and 3 h.

The photocatalytic reaction is carried out using 10 mL solution of pre-determined concentrations of MB and TiO_2 in 15 mL glass vials, which are placed in the Rayonet RPR-100 (USA) reactor, with the content continuously mixed by rotator. Four side 254 nm UV immersion lamps (1 cm i.d. \times 25 cm long) are set up around in the UV reactor the irradiation intensity is determined by an UVP radiometer (model UVX-25) at wavelength of 254 nm. After equilibrating the catalyst and methylene solution for 20 min by mixing, the UV lights turns on to initiate the reaction. Samples are periodically collected for 1 h from the reactor, centrifuged and analyzed. Methylene blue is identified and quantified by UV-2700 spectrometer at 660 nm.

2.3 Analysis and characterization

X-ray diffraction spectroscopy (XRD, Bruker D8) and environmental scanning electron microscopy (ESEM, FEI Quanta 2000) with EDS (Genesis XM 4i) are used to characterize the synthesized photocatalysts.

3 RESULTS AND DISCUSSION

3.1 Characterization of synthetic photocatalysts

Figure 1(a) shows the XRD patterns of the synthetic TiO_2 photocatalysts and Degussa P25. The results indicate that peak centers at 25.3°, 37.8° and 48.0° 2 θ can be assigned as (101), (004) and (200) orientations of the anatase phase. In addition, rutile phase at 27.4°, 36.1° and 54.3° 2 θ can be assigned as (110), (101) and (211) orientations. The synthetic TiO_2 are in good agreement with the reference patterns for anatase (JCPDS 71-1166) and rutile (JCPDS 73-1232). TiO_2 particles obtained at calcination 300°C for 1 h are rutile and anatase phases with weaker crystallinity than at calcination 500°C for 1 h and 0.5 h. Besides, for both samples calcined at 300 and 500°C, the rutile phase is the majority accompanying with

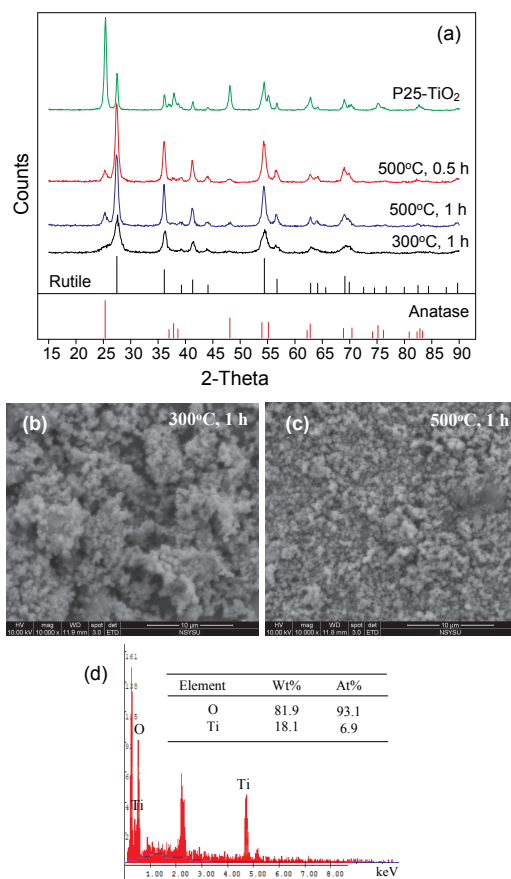


Figure 1. (a) XRD pattern, (b) and (c) ESEM images, (d) EDS of synthesized photocatalysts.

little anatase phases. Figures 1(b)–(d) show the morphology and EDS of the synthetic photocatalysts. All photocatalysts show that the aggregated consists of smaller particle of less than 1.0 μm .

3.2 Photocatalytic activity of synthesized TiO_2

Figure 2 shows the background tests of synthetic TiO_2 are obtained from 300°C for 3 h. An MB degradation of 10% is achieved, when 15 mg/L of MB is irradiated by UV light in the reactor without the presence of photocatalyst. The degradation of MB is only 11%, when 15 mg/L of MB and photocatalyst (synthetic TiO_2) are used in the reactor. This result indicates that the adsorption of MB onto the surface of photocatalyst is negligible.

Figure 3 shows that the degradation of MB under various calcination temperature and time of TiO_2 . The degradation efficiency of MB increases with calcination time from 1 to 3 h at 300°C and 400°C (Figs. 3(a)–(b)). At the end of reaction, the degradation efficiencies of MB are 90%, 100% and 99% at 300°C for 1, 2 and 3 h. Additionally, the degradation efficiencies of MB are 93%, 95% and 97% at 400°C for 1, 2 and 3 h.

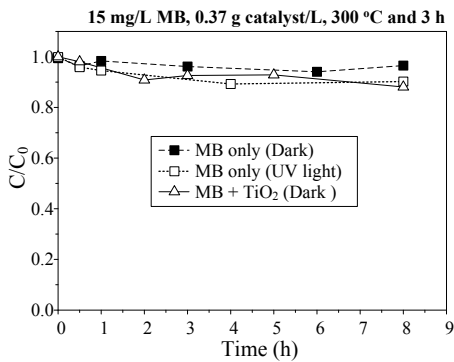


Figure 2. Background tests. Experimental conditions: reaction volume = 10 ml, at pH 6.5 and room temperature.

Figure 3 shows that as the degradation of MB under various calcination temperature and time of TiO_2 , the degradation efficiency of MB increases with calcination time from 1 to 3 h at 300°C and 400°C (Figs. 3(a)–(b)). At the end of reaction, the degradation efficiencies of MB are 90%, 100% and 99% at 300°C for 1, 2 and 3 h. Additionally, the degradation efficiencies of MB are 93%, 95% and 97% at 400°C for 1, 2 and 3 h.

At 500°C, MB degradation efficiency increases to 98% when the calcination time for 1 h, but further increase of calcination time would slightly decrease the MB degradation efficiency (Fig. 3(c)). Hadavi et al. (2012) showed that the nanoparticles of TiO_2 increases with the temperature. The large photocatalyst particles decrease available photocatalyst surface for photon adsorption and thereby cause slight reduction on the corresponding photocatalytic degradation. According to the result of Fig.3, MB degradation efficiency using TiO_2 from 500°C for 1 h is higher than that from 300°C for 1 h. Presumably, the anatase phase with strong crystallinity in TiO_2 lead to improvement of the MB degradation efficiency (Fig. 1). In addition, the literatures show that the anatase/rutile mixtures could improve photodegradation than the pure phases (Clemente et al. 2014).

3.3 Kinetic study

In this study, photocatalytic degradation kinetics of the photocatalyst could follow pseudo-first order kinetic (Tolosa et al. 2011):

$$-\frac{d[C]}{dt} = k[C] \quad (4)$$

And thus,

$$\ln \frac{[C]}{[C]_0} = -kt \quad (5)$$

where k is the apparent first-order rate constant, and $[C]$ and $[C]_0$ are the concentrations of MB in the solution at any time t and time zero, respectively. Figure 4 and Table 1 show the kinetic results of MB degradation at various experimental conditions. The degradation of MB follows the first-order reaction. This is obvious with the straight line relationship on the linearized plot of $\ln(C/C_0)$ vs. which is shown in Figure 4. The calculated R^2 values exhibits the tendency of the reaction which fits best with the pseudo-first order reaction (Table 1). The rate constant (k) can be calculated using the given slope of the plot. Figures 4(a) and (b) show that k values increases with calcinations time at 300°C and 400°C. On the other hand, the k values and r_i (initial rate) increases with calcination time at 500°C, but further decreases (Fig.4(c)). According to the results of Table 1, the highest rate constant ($k = 1.13 \text{ h}^{-1}$), r_i and degradation efficiency are achieved at optimum conditions: initial MB concentration of 15 mg/L, 0.37 g/L TiO_2 calcination 300°C for 3 h.

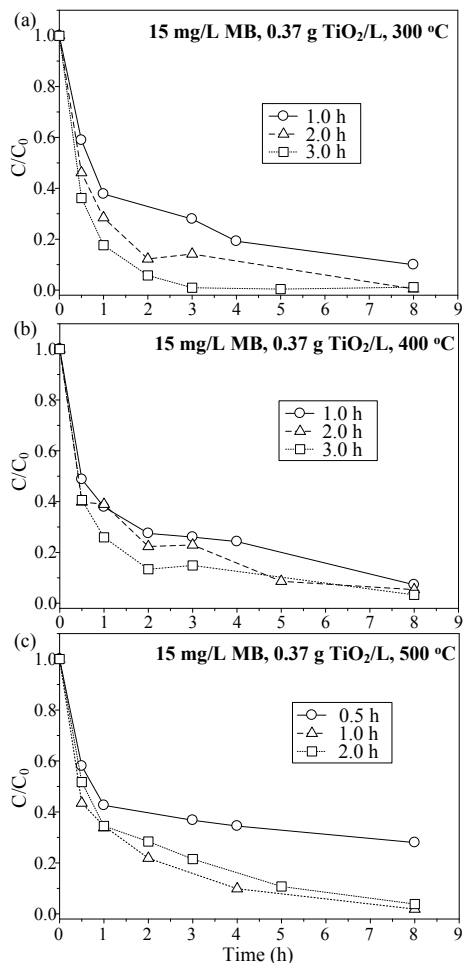


Figure 3. Shows the effects of calcination temperature and time of TiO_2 on MB degradation. Experimental conditions: reaction volume = 10 ml, UV wavelength at 254 nm at pH 6.5 and room temperature.

4 CONCLUSION

Photocatalytic degradation of MB in aqueous solution is investigated using synthetic TiO_2 as the catalysts. The degradation efficiency of MB and k values increases with calcination time from 1 to 3 h at 300°C and 400°C. At 500°C, MB degradation efficiency, initial rate and k value increase as the calcination time for 1 h; but further increase of calcination time would slightly decrease the MB degradation efficiency, initial rate and k values. About 100% of MB degradation is obtained in the 0.37 g/L TiO_2 aqueous solution with an initial MB

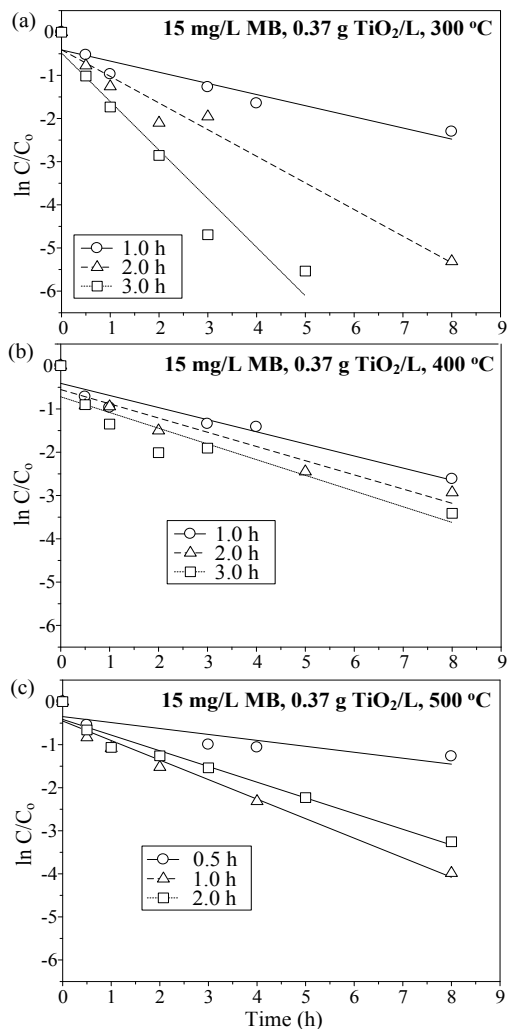


Figure 4. Linearized plot for the photodegradation of MB at various calcination temperature and time of TiO_2 .

Table 1. Initial rate, rate constant and MB degradation efficiency at various photocatalyst systems.

| Temp. (°C) | Tim c (h) | MB conc. (mg/L) | TiO_2 dosage (g/L) | r_i ($\times 10^{-2}$ mM/h) | Degradation (%) at 8 h | k value (h^{-1}) | R^2 |
|------------|-----------|-----------------|----------------------|--------------------------------|------------------------|------------------------|-------|
| 300 | 1.0 | 15 | 0.37 | 2.45 | 90 | 0.26 | 0.898 |
| 300 | 2.0 | 15 | 0.37 | 2.32 | 100 | 0.62 | 0.969 |
| 300 | 3.0 | 15 | 0.37 | 3.27 | 99 | 1.13 | 0.944 |
| 400 | 1.0 | 15 | 0.37 | 2.45 | 93 | 0.28 | 0.921 |
| 400 | 2.0 | 15 | 0.37 | 2.27 | 95 | 0.33 | 0.904 |
| 400 | 3.0 | 15 | 0.37 | 3.41 | 97 | 0.36 | 0.855 |
| 500 | 0.5 | 15 | 0.37 | 2.22 | 72 | 0.14 | 0.756 |
| 500 | 1.0 | 15 | 0.37 | 2.53 | 98 | 0.45 | 0.969 |
| 500 | 2.0 | 15 | 0.37 | 2.41 | 96 | 0.36 | 0.959 |

concentration of 15 mg/L after 180 min irradiation. As the calcination time of 1 h, both the MB degradation and k value increase with calcination temperature. Therefore, the structural characteristic of the synthesized catalysts influences the photocatalytic performance under UV light illumination.

REFERENCES

- [1] Chen, C.W., Liao, Y.L., Chen, C.F., Dong, C.D. 2013. Phenol degradation by photocatalysis on synthesized nano-TiO₂: Evolution of intermediates, organic acids, end-products, and toxicity. *Bionanoscience* 7: 1–8.
- [2] Clemente, Z. Castro, V.L.S.S. Moura, M.A.M. Jonsson, C.M. Fraceto, L.F. 2014. Toxicity assessment of TiO₂ nanoparticles in zebrafish embryos under different exposure conditions. *Aquatic Toxicology* 147: 129–139.
- [3] Hadavi, M.S. Sharifi, S. Jafari, S. 2012. The size effect of TiO₂ anatase nanoparticles on photon correlation spectroscopy. *Soft Nanoscience Letters* 2: 77–80.
- [4] Li, W., Zeng, T. 2011. Preparation of TiO₂ anatase nanocrystals by TiCl₄ hydrolysis with additive H₂SO₄. *Plos One* 6(6): e21082.
- [5] Li, Y., Lee, N.H., Hwang, D.S., Song, J.S., Lee, E.G., Kim, S.J. 2004. Synthesis and characterization of nano titania powder with high photoactivity for gas-phase photo-oxidation of benzene from TiOCl₂ aqueous solution at low temperature. *Langmuir* 20: 10838–10844.
- [6] Rauf, M.A. Meetani, M.A. Hisaindee, S. 2011. An overview on the photocatalytic degradation of azo dyes in the presence of TiO₂ doped with selective transition metals. *Desalination* 276: 13–27.
- [7] Tolosa, N.C. Lu, M.C. Mendoza, H.D. Rollon, A.P. 2011. The effect of the composition of tri-elemental doping (K, Al, S) on the photocatalytic performance of synthesized TiO₂ nanoparticles in oxidizing 2-chlorophenol over visible light illumination. *Applied Catalyst A*: 401: 233–238.
- [8] Yin, H., Wada, Y., Kitamura, T., Kambe, S., Murasawa, S., Mori, H., Sakata, T., Yanagida, S. 2001. Hydrothermal synthesis of nanosized anatase and rutile TiO₂ using amorphous phase TiO₂. *Journal of Material Chemistry* 11: 1694–1703.

2. Remote sensing, GIS and application

Simulation of soil moisture fluctuation based on the signal-to-noise ratio of GPS

R.H. Hu

Institute of Seismology, China Earthquake Administration, Wuhan, China

Y.J. Hu & X.Y. Yu

Faculty of Information Engineering, China University of Geosciences, Wuhan, China

ABSTRACT: Soil moisture and its fluctuation play an important role in monitoring the water resources recycling within a region. Based on the analysis of the multipath reflection model of Global Positioning System (GPS), the principle and method of simulating soil moisture using multipath error amplitude extracted from the signal-to-noise ratio (SNR) data of GPS is described in this paper. According to the principle and method, a simulation algorithm is designed, and a function package is developed to obtain soil moisture fluctuation. In addition, a comparison between the soil moisture fluctuation obtained by the simulation method and that directly measured by a soil moisture meter is performed. The results indicate that the variation of multipath error amplitude obtained by SNR data can reflect the fluctuation of the soil moisture. The results also indicate that the algorithm of obtaining soil moisture using SNR data is effective.

1 INTRODUCTION

Precipitation is held in the soil of the earth surface at a certain humidity before the moisture evaporates, flows to the river or infiltrates into the groundwater. The water in the soil is an important part of the land water cycle (Robock A et al.,2000).With the changes of climate and environment, dynamic changes of the soil water content affect the entire ecological system. Therefore, the fluctuation of soil moisture plays an important role on weather, climate and flood disaster forecast, as well as water resources circulation research.

Currently, the main method to measure soil moisture is to use a soil moisture meter or a remote sensing satellite. The former method is mainly adopted in soil moisture detection with high precision in a small region. Although the soil moisture measurement using a soil moisture meter can achieve high precision and large range data sharing networks are also being built to extend their range, there are still some problems. For example, the installation of a moisture meter will impact or damage the observed object and the surrounding environment and the existing regional soil moisture observation network generally has low coverage. It is difficult to assimilate the data from different networks because of the influence of soil type, the type of instrument and other relative factors; therefore, performing soil moisture measurements with high precision over a large range is difficult to achieve. Soil moisture detection based on

satellite remote sensing is generally used for moisture measurement in a large area, usually in the context of climate change detection.

The research on the extraction method of GPS multipath error provides a new method for obtaining soil moisture. In the 1990s, Comp (1997) initially proposed that the GPS signal-to-noise ratio (SNR) data can be used to correct the GPS carrier phase by establishing a model according to the frequency and amplitude characteristics of the SNR. Bilich (2008) proposed a method by which the SNR amplitude-frequency characteristic is used to establish a model and extract the dynamic, time-varying multipath parameters, including the amplitudes of the direct and reflected wave signals as well as the phase difference between the two waves. Subsequently, Bilich (2008) established a mapping relationship between the SNR data and the multipath environment by improving the multipath parameters in the model. Larson (2008) proposed a new method of using GPS multipath to measure the soil moisture fluctuation. Compared to the traditional soil moisture measurement methods, the new method can further extend the function of the existing permanent GPS networks and lacks the problem of determining how to assimilate the soil moisture data affected by different soil moisture meter networks.

Based on the analysis on the GPS multipath reflection model, the principle and method of simulating the soil moisture by using the multipath error amplitude extracted from the GPS SNR data is described

in this paper. According to the described principle and method, a simulation algorithm is designed and a function package (GPS_mois) is developed to conveniently obtain the soil moisture fluctuation. In addition, a comparison between the soil moisture fluctuations obtained by the simulation method and that directly measured by the soil moisture meter is performed.

2 PRINCIPLE AND METHOD

GPS SNR data involves the characterization of the signal strength received by GPS receiver antenna. The data are mainly affected by the antenna gain parameter, the multipath effect and the random noise of the GPS receiver. When the elevation angle is high, the antenna gain becomes higher and the SNR can be effectively improved. When the elevation angle is low, the antenna gain becomes lower and the multipath effect enhances, which makes the SNR decrease significantly. When only one multipath reflection from a smooth surface exists, the multipath effect model is straightforward, which is as shown in [Figure 1](#). Under the simplified model of GPS signal tracking, the SNR at any instant is described by:

$$SNR^2 = A_d^2 + A_m^2 + 2A_d A_m \cos \phi \quad (1)$$

Where A_d and A_m are the amplitudes of the direct and multipath signals, respectively, and ϕ is the phase difference between the two signals.

As a GPS satellite passes overhead, the reflection geometry and the phase difference change, creating oscillations in the SNR magnitude. The frequency of these SNR oscillations depends on the satellite-reflector-antenna geometry. To simplify the analysis, the reflective surface are restricted to an approximately horizontal surface (i.e., $\gamma \approx 0$). Thus, the optical path difference (Δ) and the phase difference (ϕ) between the direct and reflected signal can be expressed as [equations \(2\)](#) and [\(3\)](#), respectively:

$$\Delta = 2h \sin \theta \quad (2)$$

$$\phi = \Delta \cdot \frac{2\pi}{\lambda} = 4\pi \frac{h}{\lambda} \sin \theta \quad (3)$$

Where λ is the wavelength of the carrier wave.

From [equations \(1\)](#) and [\(3\)](#), the frequency of the SNR oscillation depends on the antenna height (h), the elevation angle (θ) and its time rate of change ($d\theta/dt$):

$$\frac{d\phi}{dt} = 4\pi \frac{h}{\lambda} \cos \theta \frac{d\theta}{dt} \quad (4)$$

This equation is further simplified by performing a change of variable ($x = \sin \theta$) in [equation \(4\)](#):

$$\frac{d\phi}{dx} = 4\pi \frac{h}{\lambda} \quad (5)$$

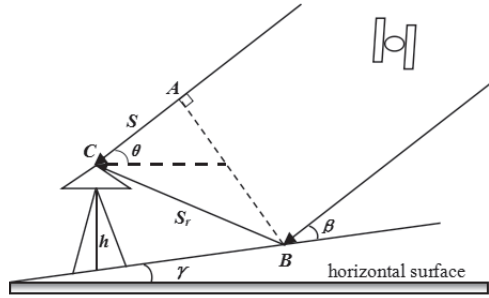


Figure 1. The multipath effect model under the condition of only one reflective surface. S and S_r represent the direct and reflected signals from the GPS satellite, respectively; h is the vertical height between the antenna and the earth surface; θ is the angle between the satellite signal and the horizontal surface, which is called the elevation angle; γ is the angle between the reflective surface and the horizontal surface; β is the angle between the satellite signal and the reflective surface.

In [equation \(5\)](#), there is a linear relationship between the phase difference (ϕ) and $\sin \theta$ for a horizontal reflector. From [equation \(1\)](#), there is a sine or cosine relationship between SNR and ϕ . Therefore, there is also approximately a sine or cosine relationship with a certain frequency between SNR and $\sin \theta$. The amplitude of the SNR oscillation depends primarily on the surface reflectivity. At GPS frequencies (1.57542 GHz and 1.2276 GHz), this effective surface dielectric constant is a strong function of the soil moisture content in the upper few centimeters, with a lesser dependence on the vegetation and the soil type. Thus, when a fixed area of the ground with unchanging soil composition is probed, the variation in the amplitude of SNR oscillations serves as a proxy for the changes in the near surface soil moisture.

Based on the above principle, a method is outlined here by which the soil moisture can be simulated using the SNR data from the standard GPS receivers, including the following three steps:

- 1 The direct signal component is removed and the multipath signal component is extracted from the SNR. During the processing of the extracting

multipath component, when the elevation angle is low, we generally consider that the amplitude of the multipath reflection signal is far less than the amplitude of direct signal, i.e., $A_m \ll A_d$. Therefore, the direct signal component in the SNR has a lower frequency and can mirror the changing trend. Generally, we use a polynomial formula to eliminate the direct component and extract the multipath reflection component.

- 2 Resample the extracted multipath reflection component. The extracted multipath signal component is expressed as A_m ; the relationship between A_m and t (time) can be easily converted into the relationship between A_m and θ . Based on equation (5), the resampling aims to convert the relationship between A_m and θ to the relationship between A_m and x ($\sin \theta$).
- 3 Use the spectral analysis method to assess the amplitude and frequency characteristic parameters A_m and ω_m (i.e., the frequency estimation corresponding to the maximum A_m) of the multipath reflection component. In the simulation of the soil moisture and its change trend, we mainly use the multipath reflection component amplitude characteristic A_m , which is estimated from many days of SNR data to measure the soil moisture content in one day; determining A_m over many days reveals the soil moisture changing trend. In general, under ideal conditions, when the soil moisture is higher, the reflected signal exhibits lower attenuation, and the multipath error signal amplitude A_m is larger; conversely, when the soil humidity is lower, the reflected signal exhibits greater attenuation, and the value of A_m is small.

3 ALGORITHM, EXAMPLE AND ANALYSIS

3.1 Algorithm

Based on the multipath effect model under the condition of only one reflective surface and the method by which the soil moisture fluctuation is simulated, we produced a function package (GPS_mois) in the Matlab environment according to the flow diagram shown in Figure 2. This package can be conveniently used for the inversion of soil moisture parameter using the SNR data.

As it is shown in Figure 2, the function package provides two modes: interactive processing and batch processing. The basic unit in processing is a project. The function package mainly includes the following four functional modules:

- 1 Data read-in module, by which the SNR data, elevation and azimuth angles are fetched and stored in the form of a matrix. The data stored are used

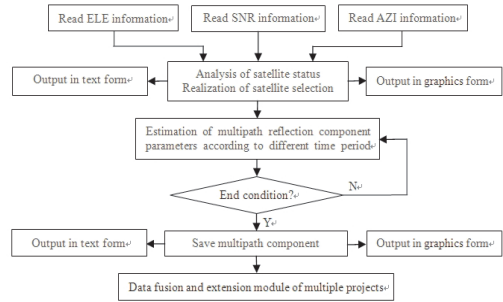


Figure 2. Flow and main modules for the inversion of the soil moisture parameter using SNR data.

to provide the basis for the satellite selection, cutoff elevation angle and other related parameters set and calculate soil moisture through data processing.

- 2 Satellite status statistics and analysis module, the main function of which is to provide the basis for the selection of the satellite and to set the other related parameters through the statistical analysis of the data, including SNR, elevation and azimuth angles, etc. Statistical analyses of the data are divided into two categories: single day and multiple day statistical analyses. The single day statistical analysis mainly determines the visual situation information of the satellites (including the number of consecutive epochs and the beginning and ending epochs), SNR data, elevation and azimuth angles, etc. within one day during the observation time period, the main statistical values of which include the maximum, minimum, average, variance and other information. The multiple day statistical analysis primarily obtains information over multiple days of some satellite during the observation time period, the main statistical values of which are the same as those of the single day case. Different from the statistical analysis on the satellite situation of a single day, the statistical analysis of the correlations between the SNR data and the elevation and azimuth angles of different days are made in the statistical analysis on the satellite situation over multiple days.
- 3 The parameter estimation module of the multipath reflection component, the function of which is to calculate, save and output the frequency and amplitude of the reflection signal over multiple days using the circulation mode. In the process of parameter estimation, the algorithm to estimate the soil moisture fluctuation can be divided into two categories: based on the amplitude and the relative phase of multipath reflection component. The algorithm based on the amplitude of multipath

reflection component is used in the function package, because the former is more intuitive and its computational process is also relatively simple.

The data fusion and extension module of multiple projects are used to achieve fusion, interpolation and filtering for the calculation results of different observation time periods at the GPS station and form new result file when a portion of the GPS observation data is missing. The “project” is used as a unit in the data processing. In one project, there is only one station along with its corresponding SNR data, elevation and azimuth angles as well as sampling rate etc. The function package provides two data processing module: interactive processing and batch processing. The interactive processing module sets the corresponding parameters, such as Project Name, Site ID, Antenna Height, Wave Length and Polynomial Order. The batch processing module uses Control File to set these parameters. Some of the control files are shown in Figure 3. The Control File mainly includes three pieces of information: basic project information, output setting and other parameters setting. The function package sets the parameters required in data processing through reading the parameter setting line in the control file.

```
#####
# CONTROLFILE
#
# This is the control file of the code package GPS_mois. In this control file,
# the basic control parameters can be set to realize the batch process. The
# rows which are started with symbol '#' are comments that will not be read
# and processed by GPS_mois software.
#####

# PART I
# ----- Basic Project Input -----
Project Name   :      P484   ,%%
Site ID        :      p484   ,%%
Year           :      2006   ,%%
First day of year:      1     ,%%
Last day of year:      70    ,%%
PRN Number     :      10     ,%%
Antenna Height :      1.2    ,%%
Wave Length    :      2      ,%%
# -----

# PART II
# ----- Output File Setting -----
Single Day Satellite Report :      Y      ,%%
Multiple Day Satellite Report:      Y      ,%%
Final Report                 :      Y      ,%%
Satellite Sky Plot           :      Y      ,%%
# -----
```

Figure 3. Some of the control files in the GPS_ mois software.

3.2 Example and analysis

To verify the effectiveness of the simulation results of the soil moisture using the above algorithm and the SNR data, the GPS station P484 is used for

the experimental study, which is one of the GPS stations of the American Continental Plate Boundary Observatory Network (PBO, <http://pbo.unavco.org/station/maintenance/P484>). The environmental conditions of this GPS station is shown in Figure 4. The elevation, longitude, and latitude of the GPS station are 1417 m, 116.620762°W and 33.375430°N, respectively. The station was constructed using the Trimble NetRS receiver with a choke-ring antenna and a SCIGN radome, which operates at one sample/15 seconds. We choose this GPS station because its observation data are of high quality and there is a weather observation station (Sky Oaks Young) nearby, located approximately 200 m from the GPS station. This weather observation station is one of the weather observation stations of the American-Flux meteorological environmental monitoring network (<http://www.americanflux.net/>). The elevation, longitude and latitude of the weather observation station are 1429 m, 116.6227°W and 33.3772°N, respectively. The environmental conditions of the weather observation station are shown in Figure 5. The station provides important meteorological observation data, including the soil moisture, the precipitation and the relative humidity of the atmosphere.

The comparison between the simulation results and the direct observation values of soil moisture should meet the following three preconditions: (1) the change of the soil moisture parameter is primarily caused by precipitation events; (2) the characteristics of water absorption of the soil at the GPS station is similar to that at the weather observation station; (3) the elevation, topography and weather of the two stations are nearly identical. Because the GPS station and the weather observation station selected basically meet the above preconditions, the simulation result of the soil moisture from the SNR data can be compared with that from the weather observation station.

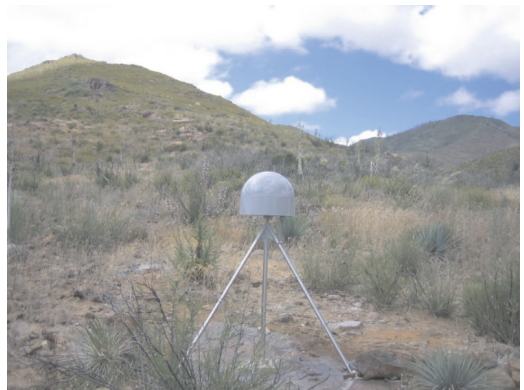


Figure 4. Environmental conditions of GPS station P484.

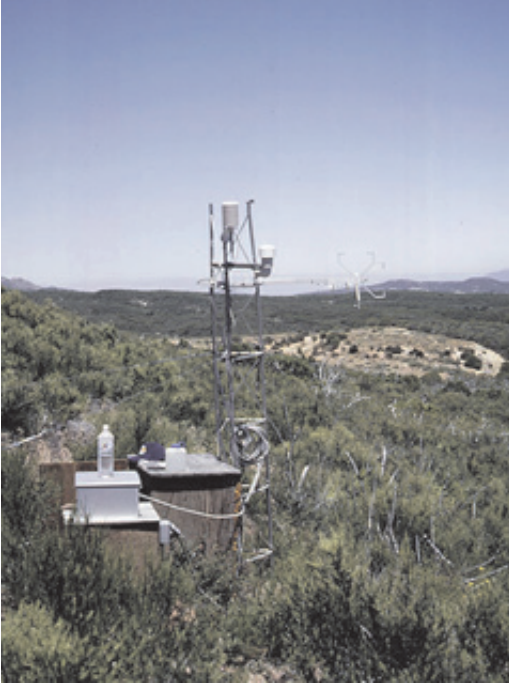


Figure 5. Environmental conditions of the weather observation station-Sky Oaks Young.

The SNR data of the nineteenth satellite from the first day to seventieth day of 2006 is used for the estimation of the multipath error amplitude A_m . We choose these days because during this period, both the SNR and meteorological data of each day are available. In the data processing, we increase the sampling interval of the GPS data from the original 15 s to 30 s, and we assign one day as a computation unit. In addition, the cut-off elevation angle of the satellite is set to 30° . The estimation results of A_m in the GPS station and the recorded data of soil moisture and precipitation in the weather observation station are shown in Figure 6.

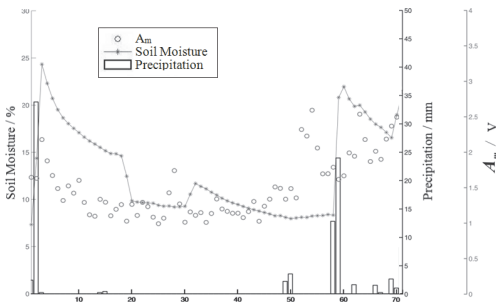


Figure 6. Comparison between the multipath error amplitude A_m and the meteorological data.

In Figure 6, the hollow and solid points represent A_m and soil moisture obtained by soil moisture meter, respectively, and the rectangular bars represent precipitation. Figure 6 shows that there are four main rainfall events within this area from the first day to seventieth day of year, which approximately occurs in the second day (30 mm), the forty-ninth and fiftieth day (4 mm), the fifty-seventh and fifty-eighth day (25 mm) and the sixty-ninth day (3 mm). The two heavy rainfall events of over 25 mm directly lead to a rapid increase of the soil moisture. And the soil moisture slowly decreases after every rainfall event and increases rapidly again when new rainfall event occurs.

When the soil moisture decreases gradually with the time, the A_m value also gradually decreases after initially increase. Conversely, when the soil moisture increases rapidly, the A_m value also increases rapidly. For example, after the rainfall event in the second day, the A_m value increases from 1.5 V during the second day to 2 V during the third days, by 25%. After the precipitation event of the second day, the A_m value decreases gradually until it returns to the normal state (1 V). Figure 6 indicates that the variation of the A_m value can generally reflect the fluctuation of the soil moisture, although there is some difference between them. The difference may be caused by many factors, such as the soil type difference between the soil moisture monitoring station and the GPS station, as well as the topography difference between the two sites. As we know, different types of soil exhibit different water sorption capacities and evaporation speeds, and a different topography will influence the storage of water. These factors can cause the difference between the fluctuation of soil moisture obtained by the SNR data and that directly measured by the soil moisture meter. It can be inferred that the change of multipath error amplitude A_m obtained by the SNR data can better reflect the fluctuation of soil moisture if the GPS and weather observation stations are located in the same place.

4 CONCLUSIONS

Multipath error is a harmful source of error in GPS navigation and positioning. However, multipath error also provides the possibility of extending the application of GPS technology in other areas, such as the simulation of soil moisture fluctuation, the determination of the thickness of ice or snow, and crop growth. The results of this study indicates that the variation of the multipath error amplitude obtained by the GPS SNR data can reflect the fluctuation of the soil moisture. The results also indicates that the algorithm for obtaining soil moisture using SNR data is reasonable and effective. Although

these results are encouraging, some problems, such as determining how to choose the proper observation satellite used for the estimation of the multipath error amplitude and how to evaluate the validity and reliability of the simulation results, warrant further study.

ACKNOWLEDGMENTS

The authors would like to thank Key Laboratory of Precise Engineering and Industry Surveying (National Administration of Surveying, Mapping and Geoinformation) (No.PF2012-21) and National Engineering Research Center for Geographic Information System (No.20111110) for their financial support.

REFERENCES

- [1] Bilich A, Larson K M, Axelrad P.2008.Modeling GPS Phase Multipath with SNR: Case Study from the Salar de Uyuni, Bolivia. *Geophys Res*113 (B04401): 1–12.
- [2] Bilich A, Larson K M.2008.Mapping the GPS Multipath Environment Using the Signal-to-Noise Ratio (SNR) .*Radio Sci* 43 (RS2006): 1.
- [3] Comp C, Axelrad P.1997.Adaptive SNR-based Carrier Phase Multipath Mitigation Technique.*IEEE Trans Aerosp Electron Syst*34 (1): 64–276.
- [4] Guan Zhi, Zhao Kai, Song Dongsheng, etc.2006. Measuring soil moisture Using Reflected GPS Signal. *Advances in Earth Science*21 (7): 747–750.
- [5] Larson K M, Small E E, Gutmann E D, et al.2008.Using GPS Multipath to Measure Soil Moisture Fluctuations: Initial Results. *GPS Solut*12 (3): 173–177.
- [6] Larson K M, Gutmann E D, Zavorotny V U, et al.2009. Can We Measure Snow Depth with GPS Receivers?. *Geophys R L*36 (L17502): 1–5.
- [7] Li Guozhong, Li Junzheng, Li Jianwen, et al.2011. Analysis of Availability of GPS in China Territory Volume Based on Satellite Service Intermittence. *Journal of Geomatics Science and Technology* 28 (1): 1–4.
- [8] Robock A, Vinnikov K, Srinivasan G, et al.2000.The Global Soil Moisture Data Bank. *Bull Am Meteor Soc* 81(6): 1281–1299.
- [9] Xiong Bo, Wan Weixing.2006.Statistical study on Signal to Noise ratio of GPS beacon. *Chinese Journal of space science* 26 (5): 326–330.

Sorted run length coding-application to Meteosat image compression

M. Cherifi, M. Lahdir & S. Ameur

*Department of Electronic, Faculty of Electrical Engineering and Computing, LAMPA laboratory,
University of Tizi-Ouzou, Tizi-Ouzou, Algeria*

ABSTRACT: Image compression is the heart of several multimedia techniques. It is used to reduce the number of bits required to represent an image. Meteosat Second Generation (MSG) satellite allows the acquisition of 12 image files every 15 minutes. Which result a large databases sizes. In this paper, an image compression method based on the arithmetic coding and a novel coding technique called Sorted Run Length Coding (SRLC) for MSG images is defined. The SRLC allows us to find the occurrence of the consecutive pixels of the original image to create a sorted run. The arithmetic coding allows the encoding of the sorted data of the previous stage to retrieve a unique code word that represents a binary code stream in the sorted order to boost the compression ratio. Through this paper, we show that our method can perform the best results concerning compression ratio and bit rate unlike the method based on the Run Length Coding (RLC) and the arithmetic coding. Evaluation criteria like the compression ratio and the bit rate allow the confirmation of the efficiency of our method of image compression.

1 INTRODUCTION

The image compression is to reduce the volume of data required for the description of the image. The exploitation of spatial and spectral redundancy and the weaknesses of the psychovisual system have enabled to develop efficient image compression methods. Digital images require large capacity for memory storage and bandwidth for transmission. These images come from different sources (satellites, radars, photography, television, medical imaging).

There are two major classes of image compression: lossless compression and lossy compression.

In the case of lossless compression, the reconstructed image after decompression is identical to the original image. No disruption is introduced on the pixel values. We distinguish in this category the methods of compression based on entropy coders such as Huffman coding [1], arithmetic coding, RLC coding, Tunstall coding [2] and Golomb-Rice coding [3]. We distinguish also in this class of image compression those based on predictive coders like the linear predictive coding LPC.

These encoders differ from the lossy compression, which introduce an irreversible degradation on the original image but allow a high compression ratio than that obtained by lossless compression methods. There are three types of lossy compression: the first type is the compression of images in the field of linear transformations. Among these transformations, we distinguish the Karhunen-Loève transform (KLT), Discrete Cosine Transform (DCT) [4], Discrete Sine Transform (DST), the discrete Walsh Hadamard

transform (DWHT), the Radon transform [5], the Mojette transform [6] and the fractal transform. The second type is the compression of image in the field of the Sub-bands where the Wavelet transform is applied to the entire image as the case of the EZW (Embedded Zerotrees on Wavelets coefficients) [7] and SPIHT (Set Partitioning in Hierarchical Trees) algorithms [8]. The third type is the lossy predictive coding like the DPCM (Differential Pulse Code Modulation) where the overall error during the reconstruction of pixels depends on the design and performance of the quantifier in addition to that of the predictor [9].

Scalar or vector quantization is at the heart of most lossy compression methods. It is used to represent a result of an output of a transform by a reduced number of coefficients. These coefficients are subjected to an entropy coding to generate the codewords. Note that in most cases the lossless compression is the basis of the lossy compression.

Arithmetic coding is a variable length lossless encoding method that has known a growing success since the last 20 years [10]. In the case of arithmetic coding, there is no reason to store the codebook as the case of Huffman coding. It is used to encode a sequence of symbols to generate at the end one codeword. Arithmetic coding is based essentially on the calculation the Cumulative Distribution function (CDF) of the probabilities of the symbols. It is the best choice for several standardization and image compression methods such JBIG, JPEG, teleconferencing video standard H. 263 [11], Advanced Video Coding H. 264. [12]. Le Gall & Tabatabai [13] have presented a new method of

digital image encoding in the sub-bands field by using symmetric short kernel filters and arithmetic coding techniques. In order to achieve a high compression ratio, Le Gall & Tabatabai have changed the approach proposed by Gharavi & Tabatabai [14] that consists to use DPCM coding of lowest band signal, PCM of upper band signal and RLC coding of the address of non-zero PCM values for channel coding of sub-bands. They have introduced the arithmetic coding in combination with the DPCM and PCM coding. The image is first divided into four bands by applying 1-D filters along rows and columns. An adaptive DPCM in combination with arithmetic coding is applied for the lowest band. A non-zero PCM values together with their positional information is transmitted by using the arithmetic coding. The simulation results demonstrate that the use of this method ensures a good quality of images reconstruction in the range of 0.7 to 0.8 bits/pixel. Abdmouleh et al. [15] provide a new method that combines arithmetic coding with RLC for lossless image compression. This combination ensures a high degree of adaptation and achieves a best compression efficiency. This technique consists to encode successively the bit planes of high weight by RLC coding whereas the other bit planes are encoded by the arithmetic coding. This method is compared to both static and adaptive model.

The RLC is one of the most simple compression techniques. It consists to replace a sequence (run) of identical symbols by a pair containing the symbol and the length of the sequence. It is used as a primary compression technique in standard 1D CCITT Group 3 fax and by combination with other techniques in the JPEG standard [16]. It is used in the case of compression in the transformation field because the coefficients quantified in each block can contain long sequences of zeros [17].

In this paper, we propose a new method of coding called the Sorted Run Length Coding (SRLC). This encoding will be combined with the arithmetic coding to develop a novel lossless image compression in order to achieve a high compression ratio. The proposed technique is compared to the basic compression method that uses the arithmetic coding with RLC coding.

The rest of this paper is organized as follows: Section 2 is devoted to describe the proposed method for lossless image compression and to present the difference between the RLC and the SRLC coding with an example. Experimental analysis of the comparative study is discussed in Section 3. Finally, Section 4 summarizes the entire paper with a conclusion.

2 PROPOSED METHOD

The diagram block of our method is illustrated in Figure 1.

Our algorithm begins first by the SRLC encoding. This step assures us a strong compression of the

image. As is shown in Figure 2, the SRLC coding generates a sequence of pixels without repetition of values. This sequence is in sorted order which allows to create a small sequence compared to that generated by the RLC coding. The SRLC also provides a table of occurrence of pixels which is not large compared to that obtained by the RLC coding. Thus, the coding of the table of the occurrences of the pixels and the sequence by the arithmetic encoding allows to generate a highly compressed file compared to the method that use the RLC coding. This encoding SRLC is especially effective when we have long sequences of identical pixels but separable by others pixels.

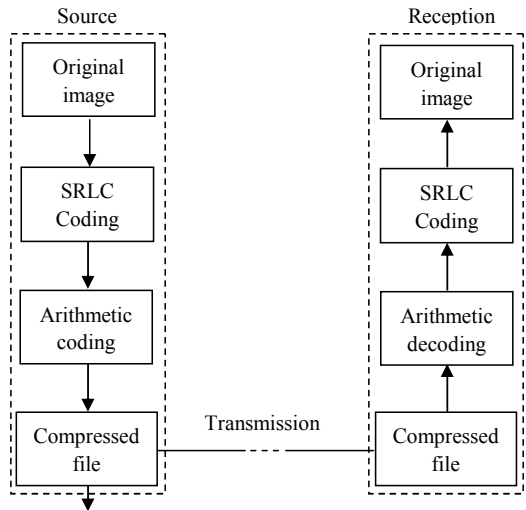


Figure 1. Image compression based on arithmetic coding with SRLC.

The second step of our algorithm consists to apply a lossless entropic coding that is the arithmetic coding. This encoding ensures a good performance compared to that of the Huffman coding because there is no reason to save the codebook [17]. This step ensures a high compression ratio.

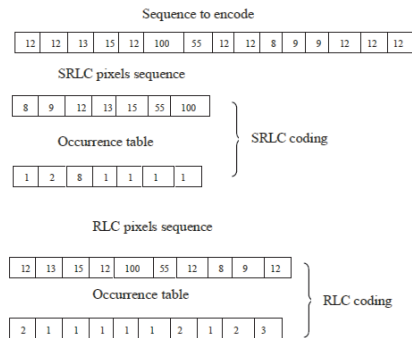


Figure 2. Difference between RLC and SRLC coding.

The step of coding at the source and that of decoding at the reception are summarized by Algorithm 1 and 2.

Algorithm 1: At the source.

A. Encoding

- Input : Image
 - Implement the SRLC coding :
 - Generate a sorted sequence of the pixels
 - Generate the sorted table of the occurrences
 - Implement the arithmetic's coding
 - Output : Compressed image
-

The first step of the decoding starts with the application of the arithmetic decoding to recover the sorted sequence of occurrences of pixels with their table. The second step consists to apply the SRLC decoding to reconstruct the original image starting from the unique values of the pixels sorted containing in the sequence with their occurrences.

Algorithm 2: At the reception.

B. Decoding

- Input : Compressed image
 - Implement the arithmetic decoding :
 - Restore the sorted sequence of the pixels
 - Restore the sorted table of the occurrences
 - Implement the SRLC decoding
 - Output : Decompressed image
-

3 RESULTS AND DISCUSSIONS

In this study, we use a database consisting of four Meteosat image. MSG images are provided by the meteorological station of the National Office of Meteorology (ONM) of Dar El Beida, Algeria recorded on 22 October 2007. The resolution of MSG images is 512×512 with 24 bpp. As is shown in Figure 3, our choice is fixed on the following MSG images: visible 0.6 (VIS 0.6), near infrared 1.6 (NIR 1.6), infrared 3.9 (IR 3.9) and water vapor 6.2 (WV 6.2).

The MSG satellite monitors a quarter of the Earth and its atmosphere from a fixed position in geostationary orbit at a longitude of 0 degree, 35,800 km above the Gulf of Guinea, off the West coast of Equatorial Africa. This satellite provides an essential contribution to the weather forecasting and to the study of environmental and climatic phenomena around the world. The MSG satellite allows to collect 12 files image that correspond to its 12 channels every 15 minutes for different resolutions.

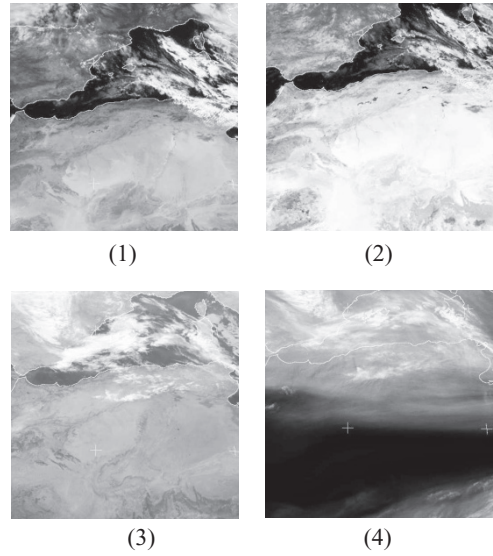


Figure 3. Images database.
(1) VIS 0.6; (2) NIR 1.6; (3) IR 3.9; (4) WV 6.2.

We used two criteria to evaluate our method and to compare it with that which uses the RLC coding with the arithmetic coding. These criteria are: the compression ratio T_c in percentage and the rate in bits per pixel (bpp).

The compression ratio provides to measure the performance of the applied image compression method. The redundancy of data between the original and compressed image CR is given by the following equation:

$$CR = \frac{\text{Number of bits in the original image}}{\text{Number of bits in the compressed image}} \quad (1)$$

The compression ratio in percentage is defined as follows:

$$T_c = \left(1 - \frac{1}{CR}\right) \times 100 \quad (2)$$

The compression ratio CR is linked to the rate that is used to measure the average number of bits required to represent a single sample [18]. The rate is defined by the following equation:

$$\text{Rate} = \frac{\text{Number of bits per pixel in the original image}}{CR} \quad (3)$$

The results of the evaluations criteria for both methods are listed in Table 1 and represented as graphs in Figure 4 and 5.

Table 1. Results of evaluation criteria.

| Method | Image | Tc (%) | Rate (bpp) |
|---|---------|---------|------------|
| The proposed method | Vis 0.6 | 99.8181 | 0.0466 |
| | NIR1.6 | 99.8066 | 0.0464 |
| | IR 3.9 | 99.8447 | 0.0373 |
| | WV 6.2 | 99.8062 | 0.0465 |
| The RLC with the arithmetic coding method | Vis 0.6 | 7.7816 | 22.1321 |
| | NIR1.6 | 16.8690 | 19.9518 |
| | IR 3.9 | 16.3600 | 20.0736 |
| | WV 6.2 | 23.4798 | 18.3655 |

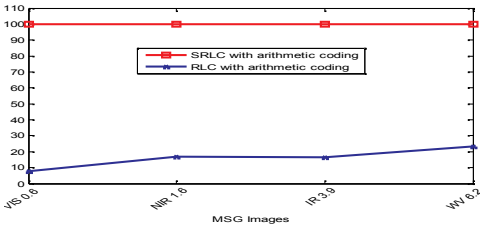


Figure 4. Tc (%) values of compressed MSG images.

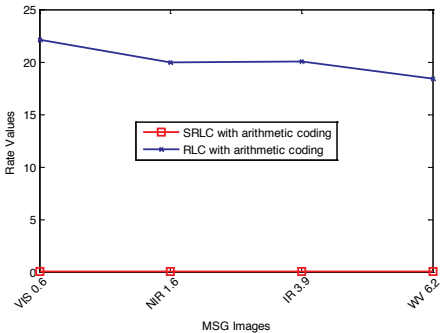


Figure 5. Rate (bpp) values of compressed MSG images.

We note from the figure 4 that our compression method provides a very high compression ratio for a lossless compression method. This is achieved thanks to the SRLC coding which has advantage over the RLC coding. This advantage lies in the fact that the sequences and their occurrences tables resulting from the SRLC coding are small compared to those resulting from the RLC coding. Thus, the application of an arithmetic coding gives a small codeword compared to that obtained by the compression method that uses the RLC coding with the arithmetic coding.

We note also from the figure 5 that our method of image compression provides a low rate compared to that obtained by the method using the RLC coding.

This result gives a great advantage because it allows us to represent each image by a reduced number of bits: less than 1 bit per pixel without degrading the images after their decompression.

We take the example of the VIS 0.6 image. The sequence of the unique values of the pixels generated by the SRLC coding contains only 256 values with a table of occurrence of 256 values only. On the other hand, the RLC coding generates a sequence of 234307 with a table of occurrence of the same size. This difference justifies the high values of the compression ratio that we have achieved and the major advantage of our method compared to that based on the RLC coding.

Of course we are not interested in this study to the quality of the reconstructed images after decompression because both methods are lossless image compression methods.

4 CONCLUSION

The MSG satellite can collect 12 images files that correspond to its 12 channels every 15 minutes for different resolutions. This process generates large databases and creates transmission and memory storage problems. These difficulties lead us in this paper to devise a new encoding called the SRLC coding.

From there, we develop a new lossless image compression method based on the SRLC coding with the arithmetic coding. Our method is simple and effective. Simplicity in its implementation and effectiveness in its application on MSG images and its power are shown in the results of the different evaluation criteria. These results clearly show that our method is qualified to be a good choice for archiving and transmission.

REFERENCES

- [1] Huffman, D. A. 1952. A method for the construction of minimum redundancy codes. IRE Proc 40 (9): 1098–1101.
- [2] Tunstall, B. P. 1967. Synthesis of noiseless compression codes. PhD thesis, Georgia Institute of Technology: 108.
- [3] Rice, R. F. & Plaunt, R. 1971. Adaptive Variable-Length Coding for Efficient Compression of Spacecraft Television Data. IEEE Transactions on Communications 16(9): 889–897.
- [4] Ahmed, N., et al. 1974. Discrete Cosine Transform. IEEE Computers Transactions 23(1): 90–94.
- [5] Radon, J. 1917. On the determination of functions from their integrals along certain manifolds. translation of Radon's 1917 paper by R. Lohner, The Radon transform and some of its applications, Annexe A. John Wiley & Sons, 1983: 204–217.

- [6] Guédon, J. 1995. Psychovisual image coding via an exact discrete Radon transform. Proc. Visual Communications & Image Processing (VCIP), Lance T. Wu editor 2501: 562–572.
- [7] Shapiro, J. M. 1993. Embedded image coding using zerotrees of wavelet coefficients. IEEE transactions on signal processing 41(12): 3445–3462.
- [8] Pearlman, W. & Said, A. 1996. A new fast and efficient image codec based on set partitioning in hierarchical trees. IEEE transactions on Circuits and Systems for Video Technology 6 (3): 243–250.
- [9] Chapin, C. 1952. Differential quantization of communication signals. Filed June 29, 1950, issued July 29, 1952.
- [10] Rissanen, J. J. & Langdon, G. G. 1979. Arithmetic coding. IBM J. Res. Dev 23(2): 149–162.
- [11] Mitchell, J. L et al. 2002. MPEG video compression standard. Kluwer Academic Publishers: 509.
- [12] Shi, Y. Q. & Sun, H. 2007. Image and video compression for multimedia engineering. CRC Press, Taylor & Francis Group: 565.
- [13] Le Gall, D. & Tabatabai, A. 1988. Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques. Proc. ICASSP-88. 2: 761–764.
- [14] Gharavi, H. & Tabatabai, A. 1986. Sub-band coding of digital images using two-dimensional quadrature mirror filtering. Proc., SPIE. 707: 51–61.
- [15] Abdmouleh, M. K., et al. 2012. A New Method Which Combines Arithmetic Coding with RLE for Lossless Image Compression. Journal of Software Engineering and Applications 5(1): 41–44.
- [16] Marques, O. 2011. Practical image and video processing using Matlab. John Wiley & Sons, publication and IEEE Press: 696.
- [17] Thyagarajan, K. S. 2011. Still image and video compression with Matlab. John Wiley & Sons, publication: 442.
- [18] Sayood, K. 2012. Introduction to data compression. 4th Edition, Morgan Kaufmann: 743.

Application of GIS to Tibetan Buddhist monastery data platform

Z.M. Caiji, L. Guo, D.Y. Xue & Y.J. Wang

College of Life and Environmental Sciences, Minzu University of China, Beijing, China

ABSTRACT: The Tibetan Buddhist Monastery Data Platform (TBMDP) is an underrepresented area of research on culture, information communication and ecological environment in China. GIS provides data for the collection, organization, storage and retrieval of Tibetan Buddhist Monastery (TBM) information, which can play an important role in the management and preservation of TBM. In this article, we integrate GIS data with TBM data to assemble multiple forms of visible data and maps to better geographically understand TBM and create new forms of communication.

1 INTRODUCTION

TBM is the center of Tibetan religious culture; there are thousands of TBM in China and most of them are distributed in Qinghai-Tibet region. As one of the most ecologically sensitive region on the earth, the Qinghai-Tibet Plateau has a very unique and fragile environment which can be easily affected by global climate change and socioeconomic activity (Wang et al, 2008; Guo et al, 2007). Thus, there has been a growing concern for the effective governance, protection and building of such sacred place, so that we can find some related traditional knowledge and culture and have a positive effect on local eco-environmental conservation. The geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. GIS can play an important role in the management and conservation of TBM by providing resources and expertise for the collection, organization, storage and retrieval of TBM. The success of TBMDP in Taiwan should be an example of its benefits for developing countries like China. This paper introduces how the TBMDP can be used as a new tool in the protection and management of the TBM of China's ethnic groups. We collect information on the monastery and build the Tibetan Buddhist monastery data platform from there.

2 GIS APPLIED IN RELIGIOUS RESEARCH

2.1 *Application of GIS to in Tibetan Buddhism*

The acronym GIS, sometimes used for geographical information science or geospatial information studies, is used to refer to the academic discipline

or career of working with geographic information systems and is a large domain within the broader academic discipline of Geo-informatics. In many faith traditions, location holds both functional and theological significance, and neighborhood worship-houses are associated with sacred spaces (Pauline et al, 2009). To participate meaningfully in contemporary “information society,” faith-based organizations are faced with pressures and challenges related to the adoption of information and communication technology (ICT) (Cukier & Middleton, 2003). GIS offers a powerful set of tools for analyzing spatial data (Alexis et al, 2008). TBM represent the continuation of Tibetan Buddhism culture. They factor heavily in the daily religious functions of many people. The monastery serves as not only a place for religious activity, but also as a building base for the stability of the social structure of the Tibetan people. Meanwhile, Tibetan people's traditional culture and living ideas originated from Tibetan Buddhism and acquired from TBM. For instance, Tibetan people have maintained traditional ecological conservation idea and behavior.

From this, one can see the importance of careful managing the monasteries. Conventionally, TBM have been mapped as a community asset, as they are popularly perceived to exemplify the best forms of social capital (Putnam, 2001). A few religious ministries are also using GIS to support their work. For example, the Christian ministry Global Mapping International uses the Global Ministry Mapping System for Arcview software to track the ratios of churches to population in each neighborhood, to identify locations and responses to literature distribution, to map project locations, and the like (<http://www.gmi.org/>).

2.2 GIS application in the protection for Tibetan Buddhist monastery

With how pronounced effect globalization has had on all aspects of human life, religious organizations have begun trying new forms of bonding and bridging of social capital, so as to better integrate themselves in this new network of globalization. The local and international flow of religious information as well as goods and services all benefit from this (Cheong at al, 2008). Consequently, ICTs may facilitate changes in notions of place through the reconfiguration of how religious information is assembled and communicated online. Such changes reflect the multimodality of the Internet as a new medium that facilitates representation, communication, and expression. Thus, Rusted (2004) observes that new media not only convert objects and texts into numerical or digital codes, but also allow users to engage in spatial encounters virtually.

3 GIS DATA PLATFORM FOR TIBETAN BUDDHIST MONASTERY IN CHINA

3.1 GIS data for Tibetan Buddhist monastery

a. The natural geography data

In order to learn more about the spatial distribution patterns of TBM, we employ the GIS to obtain particular spatial data. This includes: area, ordinate positions, terrain, slope gradient and high pass filtering of TBM, as well as the distance between the TBM and the surrounding land. By this we mean, the relative locations of the TBM too certain types of land use. The collection of these geographical data is important to analyze the spatial distribution patterns of TBM as well as to evaluate the ecological environment of TBM surroundings.

b. Tibetan Buddhist Monastery data

We should be increasingly concerned with the development of TBM. To do this, we must get much data on TBM such as: history (including building age), monastery rules, monastery sects, monastery activities/community programs and in general how a monastery provides for the local people. Finally, we use these monastery data as one part of TBMDP to provide more statistics and analysis information for doing research on the development of monasteries in culture, society, economy, ecological environment.

3.2 The analysis platform for GIS data

Based on these GIS Data and TBM Data, the platform is built for analysis of the spatial interactions among different monasteries and between the monasteries and the environment, population, economy and culture (Fig.1). Moreover, the collection of this data

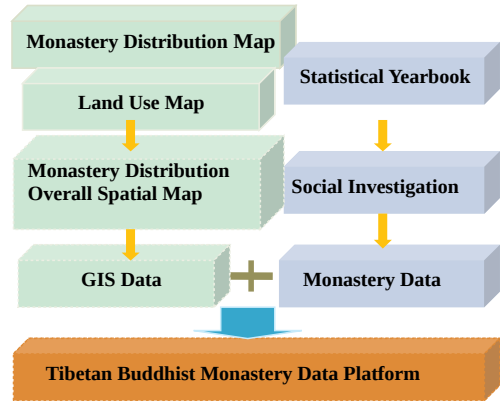


Figure 1. New structure of TBMDP in China.

is used to assess the local functionality of TBM and provide more information for research and management. For example, using the information to do some related research on the relationship between Tibetan Buddhist monasteries and its surrounding ecological environment.

3.3 A case study on application of GIS to TBM in guoluo tibetan autonomous prefecture

Guoluo Tibetan Autonomous Prefecture is located in the hinterland of the QingHai-Tibet plateau. Located in the southeast of QingHai province, this place serves as the source of the Yellow River, and is one of six Tibetan Autonomous Prefecture in QingHai. Tibetan Buddhism is the religion of 99% of this region's population and 66 monasteries are distributed amongst its six counties. Using ArcGIS as the work platform, relations between monastery distribution and terrain and between monastery distribution and slope gradient are analyzed according to the results of map layers intersection and statistics analyzing. Main research results are as follows:

a. Terrain data analysis

DEM is analyzed by ArcGIS, the results of this analysis exhibited the terrain characteristics of Guo Luo is high altitude to low from northwest to southeast. This region was located nearly 4, 500 m, 65.56%, 21.07% and 8.85% of the total area distributed in altitude 4 000 ~ 4 500 m, 4 500~ 5 50 00 m and 3 500~ 4 000 m respectively. The little proportion of the area is located below 3, 500 m and above 5000 m. According to the topographic characteristics of Guo Luo, the region was divided eight altitude levels: 0 ~ 1000 m, 2500 ~ 3000 m, 3000 ~ 3500 m, 3500 ~ 4000 m, 4000 ~ 4500 m, 4500 ~ 5000 m, 5000 ~ 5500 m, >5500 m. Distribution of monasteries in different altitudes is shown in Table 1 using ArcGIS software.

Table 1. The elevation and distribution of monastery in guoluo.

| Elevation (m) | Area (km ²) | Number of monastery | Density of monastery (/km ²) |
|---------------|-------------------------|---------------------|--|
| 0-1000 | 2812.1376 | 0 | 0 |
| 2500-3000 | 2.5515 | 0 | 0 |
| 3000-3500 | 344.7930 | 1 | 0.0029 |
| 3500-4000 | 6809.6100 | 22 | 0.003231 |
| 4000-4500 | 50454.7000 | 38 | 0.000753 |
| 4500-5000 | 16212.8000 | 5 | 0.000308 |
| 5000-5500 | 285.3140 | 0 | 0 |
| >5500 | 33.3072 | 0 | 0 |
| Total | 76955.2133 | 66 | |

The most monasteries are distributed in altitude between 3 500~4 500 m. 57.6% of them distributed altitude range from 4 500 to 5 000 m. Only 5 monasteries were located above 5 000 m. There were no monasteries in altitude above 5 000 m. Therefore the altitude 5 000 m can be regarded as the highest altitude distribution of Guoluo Prefecture monasteries (Fig. 2)

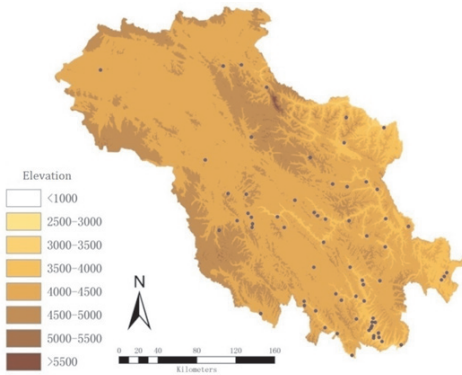


Figure 2. TBM spatial distribution of guoluo.

b. Slope gradient data analysis.

According to the data, the distribution density of monasteries increased with the increasing of the slope (0°-25°) (Table 2). This result showed that most

Table 2. Slope gradient and monastery distribution density.

| Slope gradient (°) | Total area (km ²) | Monastery number | Monastery density (/km ²) |
|--------------------|-------------------------------|------------------|---------------------------------------|
| 0-7 | 30811.5 | 5 | 0.0002 |
| 7-15 | 20306.6 | 18 | 0.0009 |
| 15-25 | 17702.9 | 34 | 0.0019 |
| >25 | 8134.25 | 9 | 0.0011 |

monasteries of Guoluo prefecture were located in steeper slope gradient (15°-25°).

We obtained GIS data from the 1987, 1997, 2007 land use maps. Firstly, the remote sensing image was dealt with and classified. (Fig.3). We transcribed visual data into a numerical form in the GIS analysis (Schuurman, 2000). The study found that ecological underlying surface of 66 TBM had changed by 24% in last 20 years (22% of TBM land cover types have varied in land degeneration and 2% of them have increased), and 76% of them have no change.

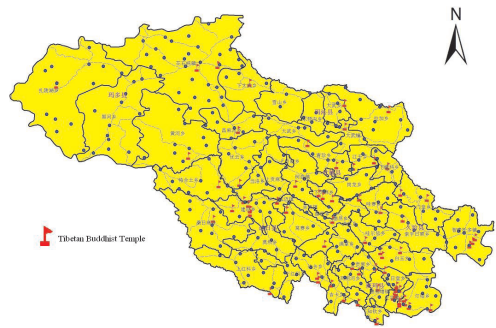


Figure 3. TBM spatial distribution of guoluo.

In addition, we compared the land use status of surrounding TBM with settlements (Radius is 4 km). The result is that the fragmentation index of the surrounding settlements is higher than that of the TBM from 1987 to 2007 (Table 3). It is speculated that religious sacred sites have played a positive role in the protection of the surrounding ecological environment. This is a result of strict Tibetan Buddhism rules banning the killing of living things as well as a traditional belief in protecting the living environment. Therefore, we need to pay more attention to effectively protecting and managing the TBM.

Table 3. Change of landscape index of monastery and settlement from 1987 to 2007.

| Years | FN (Monastery) | FN (Settlement) |
|-------|----------------|-----------------|
| 1987 | 0.00485 | 0.00765 |
| 1997 | 0.00492 | 0.00762 |
| 2007 | 0.00486 | 0.00762 |

4 CONCLUSION

TBM has always been a center piece in Tibetan culture and as such, plays an important role in the research of the culture and environment surrounding it.

This article has attempted to understand the nature of TBM and their interactions with the surrounding ecological environment through the use of the GIS Data and TBM Data. The analysis platform for TBM uses GIS not only to transmit information about TBM and religious influence, but also to analyze TBM geographic sensibilities and the effects of TBM on the ecological environment. There have also been attempts to map virtual geographies by visualizing individual activities in both physical and virtual space. For example, using a multi scale GIS environment to portray three different spatial ranges of physical and telepresence (Kwan, 2001) and the creation of a spatiotemporal GIS design to identify different patterns of human interaction on space time paths (Yu, 2006). Our study's final aim of establishing TBMDP in China is to adopt an interdisciplinary and multi-method data collection and analysis approach that is recommended for the study of media and religion.

ACKNOWLEDGMENT

The work presented in this paper was supported by the National Natural Science Foundation of China (No. 31370480) and 111 Project (B08044). We also acknowledge all people who help us built the library.

REFERENCES

- [1] Wang X.D., Zhong X.H., Liu S.Z. et al.2008.Regional assessment of environmental vulnerability in the Tibetan Plateau: Development and application of a new method. *Journal of Arid Environments* 72:1929–1939
- [2] Cheng Guodong and Wu Tonghua.2007. Responses of permafrost to climate change and their environmental significance, Qinghai-Tibet Plateau. *Journal of geophysical research*.112.
- [3] Pauline Hope Cheong, Jessie P. H. Poon, Shirlena Huang. 2009.The Internet Highway and Religious Communities:Mapping and Contesting Spaces in Religion-Online. *The Information Society*, 25: 291–302.
- [4] Cukier,W., and C. A. Middleton. 2003. Voluntary sector organizations on the Internet: The Canadian experience. *IT and Society* 1(3):102–130.
- [5] Alexis Combera, Chris Brunsdona, Edmund Green. 2008.Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. *Landscape and Urban Planning*, 86:103–114.
- [6] Putnam, R. D. 2001. Bowling alone: The collapse and revival of American community. *New York: Simon and Schuster*.
- [7] <http://www.gmi.org/>
- [8] Cheong, P. H., A. Halavais., and K. Kwon. 2008. The chronicles of me: Understanding Blogging as a religious practice. *Journal of Media and Religion* 7:107–131.
- [9] Rusted, B. 2004. Editor's introduction: Visual studies, digital imaging, and new media. *Visual Studies* 19(1):1–5.
- [10] Schuurman, N. 2000. Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography* 24(4):569–590.
- [11] Kwan, M.P. 2001. Cyberspatial cognition and individual access to information: The behavioral foundation of cyber geography. *Environment and Planning* 28(1):21–37.
- [12] Yu, H. 2006. Spatio-temporal GIS design for exploring interactions of human activities.*Cartography and Geographic Information Science*33(1):3–19.

The application of GIS in electric vehicle charging/battery swap infrastructure and operation

D.S. Bai, G. Chen, X.Q. Li, J. Liu & W.F. Li

Beijing Puri High-voltage Transmission Technology Co., LTD of State Grid, Beijing, China

W.N. Yuan

China Electric Power Research Institute. Haidian, Beijing, China

ABSTRACT: Electric vehicle charging/battery swap infrastructure is the premise and basis of large-scale promotion and application for the electric car. In order to work out a reasonable layout for the infrastructure to be fully used and orderly scheduled, this paper discusses the combination of GIS and GPS technology in specific application in the electric vehicle charging/battery swap infrastructure location, the electric vehicle dynamic positioning, the near charging/battery swap infrastructure search and schedule the way to go there. It provides a new concept in the Electric Vehicle Charging/Battery Swap Infrastructure and Operation. And the paper finally analyzes the heart of the matter in using GIS in the infrastructure construction, operation and management, and then it summarizes the paper.

1 INTRODUCTION

As a new kind of transport, Electric vehicles have played an active role in solving the energy crisis, improving the environment. During recent years, countries around the world have stepped up efforts to develop electric vehicles, and China has also introduced a series of measures to promote the development of electric vehicles^[1-2]. The basis for development of electric vehicles is an electric vehicle charging/battery swap infrastructure, as an important part of smart grid construction, and its construction and operation is important to promote the development of electric vehicles to protect^[3-5].

GIS (geographic information system), is an integrated processing and analysis technology systems for geospatial data. GIS mapping measurement is based on the database as a data source to the storage and is to use the data for computer programming as a platform for the analysis of real-time global space technology. In recent years, a constant volume in the substation siting and construction of information technology in power plants has been widely used^[6].

GPS (global positioning system), is a system which uses satellites to real-time positioning, navigation worldwide. GPS function must have the GPS terminals, transmission networks and monitoring platform with three elements. Because of all-weather, high precision and automatic measurement, etc., GPS technology has been widely applied to the vehicle

anti-theft, anti-robbery, driving route monitoring, call command, and other fields of social^[7-8].

This paper presents the application of GIS technology in siting analysis of charging for electric facility, and prospects the application of GIS in finding the nearest filling station. Combined with application of GPS navigation system, GIS can make it possible for vehicle dynamic positioning, track playback, route planning and other functions. It can offers technology-oriented for the design and construction of high-tech level of electric vehicle charging for electric facilities, rational distribution of filling station, orderly scheduling and operational management of resources charging facilities to meet the needs of electric vehicle charging, and building intelligent traffic harmonious.

2 FUNCTION INTRODUCTION OF GIS AND GPS

GIS has four elements, which are: data acquisition and editing attribute, data editing and analysis, graphing capabilities, and spatial database management functions. GIS technology is based on components of software systems and architecture, object-relational database management, system management and graphic and attribute data. GIS has many functions, such as spatial data management, multi-scale multiple data sources, editing, spatial query, spatial analysis and statistics, decision-making and support.

GPS is a space satellite navigation and positioning system, which consists of space segment, ground control section, and the user segment. GPS dynamic positioning is to GPS signals to measure the Earth's motion relative to the user's state of antenna parameters, and these parameters include the three-dimensional coordinates of the state, three-dimensional velocity and time seven parameters. GPS navigation measures the motion vector of state parameters and motion vectors to accurately guide the movement to a predetermined subsequent position.

3 APPLICATION OF GPS AND GIS IN CHARGE OF FACILITY OPERATIONS AND MANAGEMENT SYSTEM

3.1 Application of GIS in location of charge of facility

The growing popularity of electric car requires charging facilities for electric coverage area should be broad, and covering density should be larger. Reasonable siting and layout are the primary issues of charging for electric facilities.

Whether the location is reasonable directly affects the convenience of charging for electricity for the electric vehicle users and the effective use for power facilities and its economic benefits and other issues. The purpose of the location analysis is to charge for electricity facility to find an appropriate position, according to the geographical characteristics of the region, by considering the allocation of resources, market potential factors, traffic conditions, terrain characteristics, environmental impact, etc., within the region to choose the best location.

Use GIS to achieve charging for electric facility siting study area to be selected first to get geographical data region, including multi-scale general maps, thematic maps, remote sensing data, the use of GIS software to build a spatial database. And the condition indicates the location to facilitate the use of computer storage and data structures. Widely used representation methods are first-order predicate logic notation, production representation, frame representation, and object-oriented representation method. According to siting factors, it can determine the spatial charge for electricity facility layout and scope of services, through spatial analysis, such as buffer analysis, Thiessen polygon analysis, resource allocation analysis, layout guidelines for the development of facilities, and the establishment of the objective function and constraints. Finally, the development of rational strategies is to complete the site.

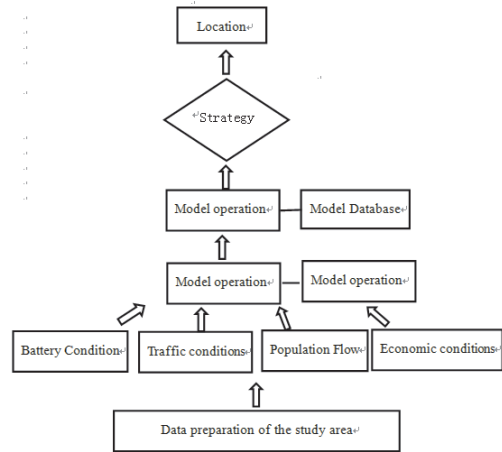


Figure 1. Charge for electricity facilities location.

3.2 Applications of GIS in electric vehicle dynamic positioning

The existing problem of an electric car battery is the driving range not long enough. For electric car, during the run, you need to understand the operational status in real-time power of the battery. When the battery power is low, the owners need to first determine the car's driving position, and then find the case for electric charging facilities nearby, and finally find the optimal path to charge for electricity and the facility is located, or to charge for electricity. When the electric car cannot move forward, the owners can use GPS navigation and positioning system for positioning themselves in the position, and then send the information to help the charging facility operations management systems, and operations management system by scheduling maintenance vehicles will take the car to the charging station or repair station.

Car navigation GPS receiver can receive the satellite signal and obtain location information of the current electric vehicles. GPS technology can quickly and accurately obtain location information, but the signal is weak or does not need to rely on the case of continuous dead reckoning continuous navigation. State Kalman filter-based GPS / DR Integrated GPS and DR mode can be used for navigation and positioning information simultaneously solving process, so that the DR system in the filtering process has been continuously amended portfolio positioning output and can provide a more accurate initial position and orientation information, even when the GPS fails, with alone DR positioning projections. It is also for a long time to maintain a relatively high positioning accuracy. The vehicle position information or the running track GPS / DR integrated navigation system

is provided by the navigation system with electronic map to match the high-precision road data to calculate the vehicle which is currently traveling the roads and the specific location on the road, and it accurately displays on the electronic map. In avoiding positioning map matching, logic errors may occur, to ensure the car on the road.

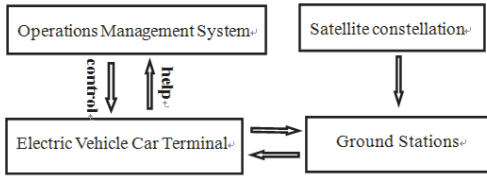


Figure 2. Electric vehicle dynamic positioning.

3.3 Application of GIS in finding in electric vehicle charging facilities for electric

Charge for electricity facilities using GIS to find you first need to navigate the required charging for electric facilities of various types of spatial information and attribute data related to organizational management, the establishment of road network topology, and the symbolic expression of various types of spatial data, and electronic map content form contains the exact road data, configuration information charging for electric facilities, house information, and land use, the surrounding environment, and the display text information. Electric car users not only can easily check the charge for electricity facilities nearby construction, but also can implement the query charge for amenities such as bus station near the power station, hotels, ATM machines and other information. Electronic map as a visual carrier provides intuitive, clear and comprehensive information to the user of expression.

3.4 Application of GIS in planning electric vehicle route

To determine the location information and information of filling station nearby for electric vehicles, GPS navigation feature helps the user choose an optimal path when departure and destination are selected. It is also based on road congestion, dynamic programming running lines.

Path planning needs to build in navigation geographic data in detailed road network information on the topology of the road network to run the policy formulated by the calculation and select the optimal running path. The navigation system is connected with the traffic management system; the optimal path can be changed at any time based on current traffic conditions. Real-time traffic information center will

link travel time, intersection delay time of travel time and transport to the car terminal, automotive terminal to choose the path of travel time as weights, according to pre-established shortest path algorithm chooses to run the line.

With the position of the vehicle to be guided as a function of time and the best route to continue the path planning module for comparison, send a signal to the mobile terminal. Because the travel guide must be clear, which can completely tell the driver where the exit is, where the roads are, where the steering is and other road information, therefore, the guide must be based on a detailed travel guide road network layer.

Infrastructure search

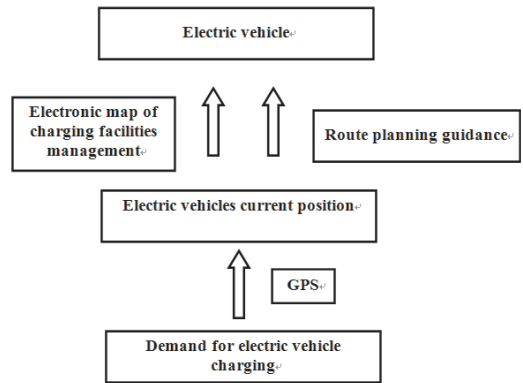


Figure 3. Electric vehicle charging/battery swap.

3.5 Application of GIS in electric cars running track playback

The trajectory of the vehicle playback in favor of management and security sector departments can grasp vehicle running track. Track playback function is based on the historical location coordinates stored in the database and the historical trajectory of the vehicle on a journey to re-map playback. After the completion of a first exercise of the vehicle, its route will be saved and, if necessary, the trajectory of the vehicle and geographic information on a variety of electronic map layers will be redisplayed. By setting the time, you can adjust the playback time and sections [20]. This is very important for correct assessment of the performance of electric vehicles and batteries, vehicle mileage and electric consumption statistics to assess the economics of electric vehicles as well as appearing in unexpected run to analyze.

4 KEY ISSUES IN APPLICATION

Detailed, complete and accurate geodatabase is the premise and basis for charging for electric to establish

facilities siting, and for finding and planning path. How to quickly and efficiently build geographic data of charging facilities for electric and how to dynamically update the data in a timely manner are the key technology issues in applications. The special nature of geospatial data charging facilities around the content, structure, etc., should be fully taken into account, ready to fully charge for electricity around the facility map data, text data and statistical information related to the establish a database [19].

The problems are that the car terminal screen is small and computing ability is slow. For car navigation, the location information, the location and the destination information can be obtained through wireless communications to be uploaded to the operations management system and management system by the operator after running dynamic route planning vehicles issued to the vehicle terminal, which can effectively solve the above issues.

Construction and operation of electric cars with GIS still need to effectively support information sharing on the Internet of Things, asset management and other technologies. Things to electric vehicle charging facilities for electric are interconnected through a variety of wireless or wired long distance or short-distance communications network, providing real-time online monitoring of charging for electric facilities in operation to look for the idle charge for electricity facilities, development of operational decision-making, safe and controlled dispatching idle equipment, implementation management and service functions

5 CONCLUSION

This article describes a vision which GIS can be used in construction and operation management for electric vehicle charging facilities for electric. GIS geographic information combined with high-speed precision GPS positioning function, will be widely used in operations management and orderly

scheduled with vehicle electric, vehicle charging for electric facilities siting, layout, and charging infrastructure resources. This article can provide new ideas for the construction and development of intelligent transportation.

ACKNOWLEDGMENTS

This work was financially supported by The National High Technology Research and Development of China 863 Program (2011AA05A109). National Energy Application Technology Research and Engineering demonstration project(NY20110703-1)

REFERENCES

- [1] CC Chan TRANSACTIONS Zhu Jianguang modern electric vehicle technology [M]. Beijing: Beijing Institute of Technology Press, 2002.
- [2] Zhang Wenliang, Wu Bin, Li Wufeng, to the well-off and explore the development direction of China's energy supply mode of pure electric vehicles [J] Power System Technology, 2009, 33 (3): 1-5.
- [3] Xu Fan, Yuguo Qin, Gu temporary peak, and other electric car charging stations Layout Analysis [J] East China Electric Power, 2009, 37 (10): 1678-1682.
- [4] Q / GDW478-2010 electric vehicle charging facilities Technical Guidelines [S].2010.
- [5] Wu Chunyang, Li Canbing, Du Li, Cao is an electric vehicle charging facilities planning method [J] Automation of Electric Power Systems, 2010, 34 (24): 36-45.
- [6] Xin Jianbo, Wenyu Bin, Li Rui electric vehicle charging facilities Methods of demand forecasting [J] Jiangxi, 2010, 34 (5):1-5.
- [7] Wang Shan, Wei ocean, Xiao Jun, and other two-stage method for optimal planning of substation locating and sizing [J] power system automation, 2005, 29 (4): 62-66.
- [8] Liang, Zhong ear design Cadastral Parcel Information System database [J].China Symposium GIS, Beijing, 1999.

Validation of LSMA and SAM for geological mapping: Comparative study among BGIS-2000, TM and ASTER

A. Bannari

Geoinformatics Department, Arabian Gulf University, Manama, Kingdom of Bahrain

G. Girouard

Remote Sensing and Geomatics of the Environment Laboratory, Ottawa-Carleton GeoScience Center, University of Ottawa, Ottawa (Ontario) K1N 6N5 Canada

A. Saquaque

MANAGEM Mineral ONA Group, Twin Center Casablanca, Morocco

ABSTRACT: The aims of this study are to map and validate the linear spectral mixture analysis (LSMA) and spectral angle mapper (SAM) algorithms for geological mapping and compare the results among three optical sensors: BGIS-2000 of Quickbird, Landsat-TM and ASTER-TERRA. Combined corrections of atmospheric effects and sensor radiometric calibration drift were used. Geometric distortions and relief displacement effects were corrected using digital elevation models (DEM). The rock classes' end members were extracted from each image using PCA. The results show that SAM applied to TM data provides the best geological maps that compare favourably with ground truth. Considering SAM and LSMA methods, ASTER results are less superior comparatively to TM. The BGIS-2000 data does not provide good results either for SAM or for LSMA methods because the number of spectral bands is limited and the SWIR bands are absent.

1 INTRODUCTION

Multispectral remote sensing techniques have been widely used in the past decades to discriminate different materials based on the dissimilarity that exists among their spectral properties [1]. Geological applications could greatly take advantage of this technology because it allows to observe and map the Earth surface over large areas, but generally low spectral resolution of multispectral sensors obstruct geological research. Digital image classification is a technique that is widely used to classify each individual pixel in an image based on the spectral information in order to create geological maps. Various classification algorithms have been used in the past decades in a variety of applications for mapping: forestry, agriculture, geology, etc. The main disadvantage of these algorithms is that each pixel in the image is compared to the training site signatures identified by the analyst and labelled as the class it most closely "resembles" digitally [2]. SAM is different from these standard classification methods because it compares each pixel in the image with every pure spectral signature (endmember) for each class and assigns a ponderation value between 0 (low resemblance) and 1 (high resemblance). However, SAM

method does not consider the sub-pixel value. The spectral mixture problem can become problematic because most of the Earth's surface is heterogeneous, thus for a medium space resolution sensor like TM or ASTER, false assumptions can be made. Differently, LSMA approaches adopt various strategies optimizing uncertainty on the fractional abundance of the end members in the mixed pixels according to their spectral contrast and their variation.

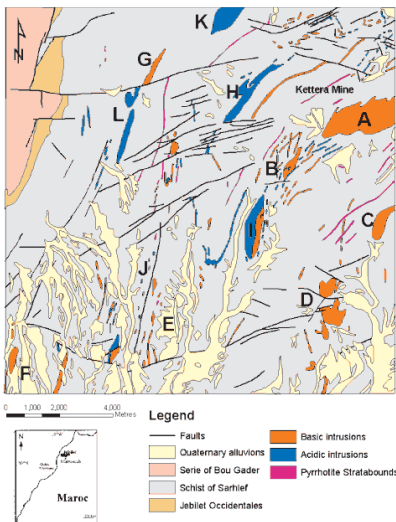
The objectives of this study are to map and validate LSMA and SAM algorithms results for geological mapping and compare the results among three different optical sensors: BGIS-2000, TM and ASTER. The spatial and spectral characteristics of these sensors are opposite to each other. TM has a medium spatial resolution (30 meters) versus the high spatial resolution (2.5 meters) for BGIS-2000. The ASTER image was re-sampled to 15 meters pixels size to harmonize between SWIR and VNIR bands. Nonetheless, the spectral dimensions of TM and ASTER are richer for geological mapping than BGIS-2000. Indeed, with their 6 and 9 bands, respectively, for TM and ASTER covering the spectrum between 0.45–2.35 μm in opposition to 4 bands for BGIS-2000 that only covers the 0.45–0.90 μm range. Combined corrections of atmospheric effects and sensor radiometric calibration drift

were used to transform the images “DN” to surface reflectances. Then geometric distortions and relief displacement effects were corrected using GCP and DEM to derive ortho-images.

2 MATERIALS AND METHODS

2.1 Study site and geological setting

The study site is located in the Paleozoic terrain of central Jebilet in Morocco, which is about 35 Km north of Marrakech. The topography is formed by small-elongated hills, which vary in elevation from 350 to 750 meters, and it is characterized by a lithological diversity, semi-arid environment, and significant mineral potential. The geology of this site is composed of four main rock classes of which the schist is the dominant rock unit. It is intensively deformed and intruded by numerous acidic and basic magmatic rocks. The economic interest of the area is focused on lens-shaped pyrrhotite stratabound massive deposits (iron caps), which appear on the surface as rust-color “oxidized minerals” caused by the oxidation of iron sulphides. The Figure 1 shows the simplified geological map of Central Jebilet established by Huvelin in 1970 [3] using aerial photo-interpretation and fieldwork. For validation preprocess, this map was used and the fundamental classes were identified by alphabetic letters and their corresponding geographic and local names.



| | | |
|-----------------------------|---------------------------|-------------------------------|
| A - K ^{at} Kettera | E - J. Hadit | I - K ^{at} Mirouga |
| B - S ⁱ Mahjoub | F - Mena el Kahla | J - K ^{at} Aicha |
| C - K ^{at} Ahril | G - J. Ahmar | K - K ^{at} Hamra |
| D - Gour es Sefra | H - K ^{at} Dalaa | L - K ^{at} Bouzelfar |

Figure 1. Simplified geological map of Central Jebilet (Huvelin, 1970).

2.2 Images data

Three images were used in this study: TM image was acquired on June 26th 1986, ASTER image on December 10 2000 and the BGIS-2000 image on February 26th 2003. These images were acquired over the same area and at the same solar time, thus have similar illumination characteristics. They also show very low vegetation distribution, which is good for geological and lithological mapping.

2.3 Pre-processing data

For accurate analysis, radiometric sensor drift, atmospheric, geometric and topographical effects corrections are necessary [4,5]. The three considered images were corrected for atmospheric effects using the updated Herman transfer radiative code (H5S) [6]. Sensors radiometric calibration drift were done with the post-calibration dynamics ranges (gain and offset) delivered with the data by NASA. To preserve the images radiometric integrity, the atmospheric effects and sensor calibration were combined in one step. Geometric distortions and relief displacement effects were corrected on the three images in one step using GCP and a DEM derived from topographic map. A 30 x 30, 15 x 15 and 2.5 x 2.5 meters grids were created for TM, ASTER and BGIS-2000, respectively. The images ortho-rectifications were achieved using the “Rational Function” model implemented in PCI-Geomatics [7].

2.4 Spectral angle mapper (SAM)

The SAM has been used successfully for geological mapping and for identifying potential mineral exploration sites [8,9]. It was used to map alteration minerals associated with mine waste [10] and to map different ophiolite lithologies in Cyprus [11]. SAM is a classification method based on the spectral similarity calculation in n -dimensions space between the image spectrums to references reflectance’s spectra [12,13,14]. Small angles indicate high similarity and high angles indicate low similarity. This method is not affected by solar illumination factors, because the angle between the two vectors is independent from the vectors length [12,15]. It is also a very powerful classification method because it represses the influence of shading effects to accentuate the target reflectance characteristics [16]. The main disadvantage of this method is the spectral mixture problem. In general, the spectral mixture problem should decrease with higher resolution images like BGIS-2000. But in some cases, it can also increase the mixture problem because more local variations in the surface spectral property become more apparent, such as humidity, elevation and shading effects [17].

2.5 Linear spectral mixture analysis (LSMA)

The LSMA approach has been extensively used in many applications: geology, soil, land use, forestry, etc. [18,19,20]. The fundamental assumption of LSMA is that generally each pixel on the surface is a physical mixture of several constituents weighted by surface abundance, and the mixture spectrum is a linear combination of the endmember reflectance spectra. In spite of the omission of the multiple scattering effects among different targets in the pixel [21], the errors associated with the linear assumptions have been found insignificant [22]. As a result, LSMA generates a series of fraction maps, which determine the relative contribution of each endmembers to the total reflectance recorded for each pixel, in which the values range from 0 to 1, where 0 indicates a low abundance and 1 a high one. However, the validity and accuracy of the LSMA rest heavily on the "user-supplied" set of endmembers spectra. LSMA was performed on the images using the full spectral bands relative to each sensor.

2.6 Endmembers selection

The endmembers for TM, ASTER and BGIS-2000 were manually selected and extracted through PCA transformation [23]. A total of 8 endmembers were extracted from the TM and ASTER images, which are: basic and acidic magmatic rocks, iron caps, schists, carbonates, quaternary deposits, sandy loams and vegetation. All these endmembers were significantly different from each other, thus could be used with confidence for classification. The carbonate endmember was selected in the TM and ASTER images where it was observed in the field using a GPS with ± 3 m accuracy. For BGIS-2000, the same endmembers were chosen for SAM, but only basic and acidic magmatic rocks and vegetation endmembers were considered for LSMA methods because the spectral bands number is limited.

3 DISCUSSION AND CONCLUSIONS

3.1 SAM

Fractions maps for each endmember class were retrieved considering the three different images. The significant and representative eight endmembers fractions maps are, then, grouped and compiled to create geological maps for TM, ASTER and BGIS-2000 sensors (Figure 2). Considering the eight endmembers classes, the TM and ASTER results maps are relatively accurate compared to the geological map (Figure 1). First, the basic rocks class emerges well in different zones of the study site, such as: A (K^{at} Kettera), B (Sⁱ Mahjoub), C (K^{at} Ahril),

D (Gour es Sefra), E (J.Hadit), F (Mena el-Kahla) and G (J. Ahmar). The acidic rocks class also demonstrates good results; they were identified in the major zones, such as: H (K^{at} Dalaa), I (K^{at} Mirouga), J (K^{at} Aicha), K (K^{at}Hamra) and L (K^{at} Bouzefar). In addition, some basic and acidic rocks were identified by SAM and these do not figure in Huvelin geological map. Indeed, in the south-west of L (K^{at}Bouzefar) area, some small basic and acidic rocks, and iron caps were discovered. Some large acidic rocks were also found in the west of K (K^{at}Hamra) area and they came out very well in the TM and ASTER maps. These regions have successfully validated in the field.

The iron caps class was harder to discriminate with SAM because they are relatively narrow in thickness (few meters), but they usually extend for a few kilometres long. For TM, which has a medium spatial resolution, and SAM, which does not consider the sub-pixel, the mixture problem for this class can cause problems. Also, there is a similarity between the iron caps and the basic rocks because they are both rich in ferromagnesian minerals, thus their altered surfaces (patina) are similar due to the dark reddish color caused by the iron oxides alteration and hydroxides. However, this class was correctly identified and mapped with TM in the areas where they were known to occur, such as: Kettera Mine, the Gour-El-Hokkane area localised west of I (K^{at}OuledHar) region, the E (J.Hadit) area and the west of L (K^{at}Bouzefar) region. Unfortunately, these zones do not stand out so well in the ASTER map because the pixels are often not classified. Furthermore, the sandy limestone class did not provide good results. Indeed, when we validate this class in the field, some pixels were classified as sandy limestone. However, several other pixels were classified in the derived maps as limestone-sandstone. Unfortunately, it was not possible to validate correctly this class because it is absent on the Havelin geological map, and it was impossible to have accessibility to this area in the field.

As it is shown from the geological map (Figure 1), the schist class covers most of Central Jebilet. This class was satisfactory and similarly classified in the ASTER and TM maps. Unfortunately, the quaternary deposits distribution was over-estimated in TM and ASTER maps comparatively to the geological map because the remote sensing technology can only observe the Earth surface without any auxiliary data. The carbonate endmember is not presented on the geological map, but it was selected in the TM image where the unit was observed in the field using the GPS (± 3 m); the output of this class was verified in different points in the field and it shows an accurate representation result. Finally, the vegetation class show an excellent result due to the semi-arid environment of the study site and sparse vegetation cover, which is very contrast with rocks and bare soils.

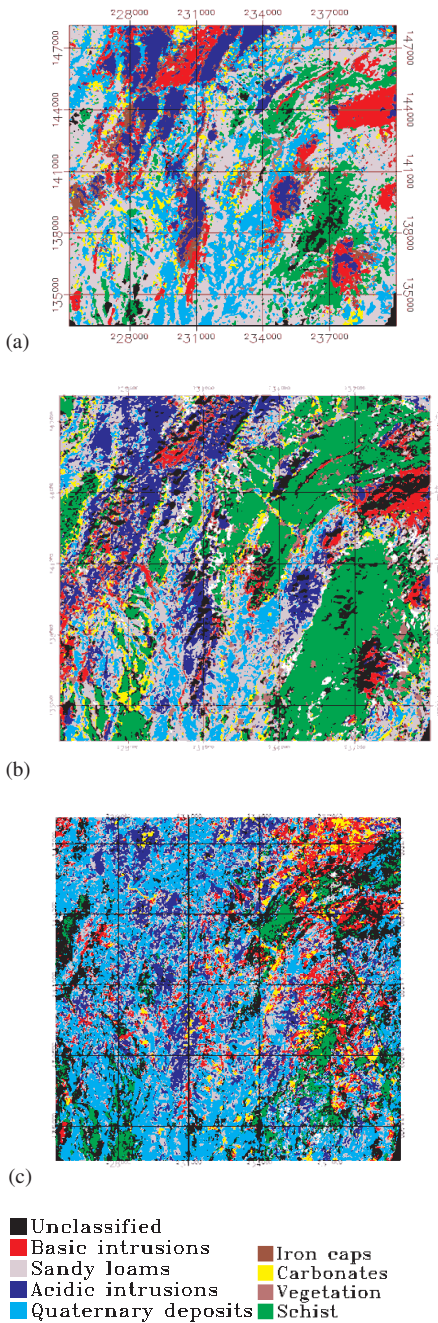


Figure 2. SAM maps output for TM (a), ASTER (b) and BGIS-2000 (c).

The derived map from BGIS-2000 image (Figure 2c) using SAM approach was not able to provide satisfactory results for different rock classes' identification

and the differentiation. The basic and acidic rocks and iron caps were not clearly recognized, as they were with TM and ASTER maps. BGIS-2000 did however identify zones where basic rocks occur, such as the following zones: A (K^{at} Kettera), C (K^{at} Ahril), D (Gour es Sefra), E (J.Hadit) and F (Mena el Kahla); and where acidic intrusions occur such as H (K^{at} Dalaa), J (K^{at} Aicha), K (K^{at} Hamra) and L (K^{at} Bouzefar) as well as the iron cap of Kettera Mine. The quaternary deposits and sandy loams were widely spread over the study area compared to the schist unit. These classes were not classified properly because of the BGIS-2000 spectral limits. Finally, vegetation is the class that provide the best results because it's reflectance high in the NIR band. Nevertheless, overall the SAM results were scattered and most of the geology classes were misclassified due to the low spectral dimensions of BGIS-2000.

After analyzing the three geological maps generated with SAM method, we can conclude that the TM map (Figure 2a) is the most similar to Huvelin geological map (Figure 1) and the field observations and validation. The produced ASTER map has not been able to provide superior results despite the best spatial resolution and spectral richness. Several factors have contributed to these results. The radiometric integrity of ASTER image could be modified and generating errors during there-sampling operation when the SWIR bands were geo-referenced from 30 to 15 meters with the VNIR bands. Moreover, even after the ASTER image ortho-rectification, some regions were still shaded (micro-topography artefacts), which has increased the number of unclassified pixels. Indeed, even the used DEM pixel size was similar to the re-sampled ASTER image (15 m), this DEM was generated from 1 / 50 000 topographic map with 20 m interval of contours lines. To correct perfectly this micro-topography shadow effects the DEM pixels size should be very low than de used image pixels size. Finally, it is possible that the ASTER raw radiometric quality was probably lower than TM.

3.2 LSMA

The Figure 3 illustrates the derived geological maps from TM, ASTER and BGIS-2000 images using LSMA approach. The TM and ASTER maps are relatively comparable to the Huvelin geological map. All basic rocks were identified with TM map (Figure 3a), such as: A (K^{at} Kettera), B (Sⁱ Mahjoub), C (K^{at} Ahril), D (Gour es Sefra), E (J. Hadit), F (Mena el Kahla) and G (J. Ahmar) regions. This class was also identified with ASTER; unlike B (Sⁱ Mahjoub) area which was classified as acidic rocks. In addition, G (J. Ahmar) area has not been classified correctly with ASTER data because it consists of a mixture of pixels classified as basic-acidic rocks, schist's and quaternary deposits. However, this "G" region

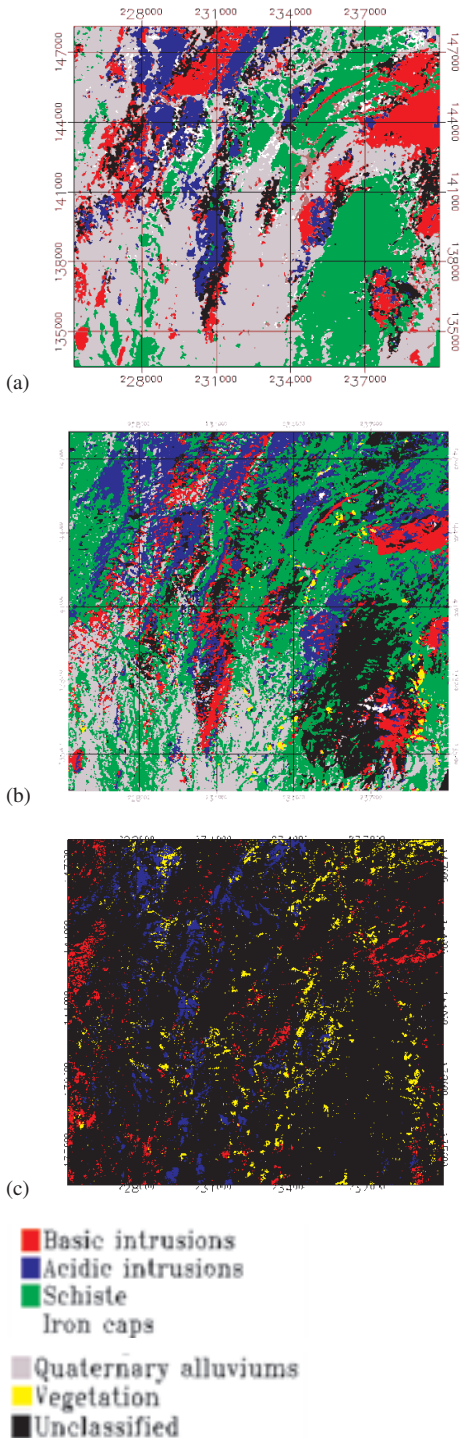


Figure 3. LSMA maps output for TM (a), ASTER (b) and BGIS-2000 (c).

was classified as a homogeneous basic rocks class with TM. The obtained map using SAM method and ASTER image (Figure 2b) had given the best results for this class than LSMA (Figure 3b). Acidic rocks were also well classified with TM and ASTER images in the following regions: H (K^{at} Dalaa), I (K^{at} OuledHar), J (K^{at} Aichaa), K (K^{at} Hamra) and L (K^{at} Bouzelfar).

As being successful with SAM, some new basic and acidic rocks were discovered with LSMA approach. Indeed, in the southwest of L (K^{at} Bouzelfar) region a few small basic-acidic rocks and iron caps were discovered. Some large acidic rocks were also found in the west of K (K^{at} Hamra) area and have been successfully validated in the field. As with SAM method, these areas stand out well in the TM and ASTER maps. Some iron caps are visible on the map derived from TM data, especially around the Kettera mine and Gour-El-Hokkane region west of I (K^{at} OuledHar) area. However, this class was more accurately classified with SAM approach. Unfortunately, compared to the field and Huvelin geological map, the iron caps in the E(J.Hadit) and west of L (K^{at} Bouzelfar) regions were absent in the TM and ASTER maps. Schist's class was classified on the TM map similarly to that of SAM, with extensive area in the north of D (Gour esSefra) and in the Kettera mine regions. On the other hand, schist's have not been classified correctly with ASTER map in the northern of D (Gour es Sefra) area due to pronounced shadow in this region. However, this zone had been correctly classified with TM and ASTER images using SAM method, because this approach is less affected by solar illumination variations [12,24]. The major difference between the obtained maps using LSMA and those derived with SAM, is only six classes which were used for TM image. Consequently, the sandy loams and quaternary alluvial deposits classes were merged in one class which did not help to improve the accuracy of the derived TM and ASTER maps. Finally, despite the strong contrast between the vegetation cover and other rocks units, this class was not correctly mapped with LSMA.

Only three classes were used to process BGIS-2000 image with LSMA, which supplied very poor map results (Figure 3c). The limitation of spectral band number and the absence of the SWIR did not help for mineral and geological mapping. The SAM method was relatively suitable for BGIS-2000 image, because the used classes' number was independent of the bands number. Certainly, the BGIS-2000 high spatial resolution is more helpful for visual interpretation to extract information with accurate details and facilitate structural mapping. This demonstrated that for geological mapping and mineral exploration it was much more important to have a high spectral resolution rather than a high spatial resolution.

Obviously, after analyzing the three obtained maps using the LSMA method, we can conclude that the derived map from TM was relatively comparatively to the Huvelin geological map and the field truth. The derived map from ASTER had once again been unable to provide superior results to TM, despite its higher spatial resolution and spectral richness. Several factors have contributed to these results, as it is explained before.

REFERENCES

- [1] G.R. Hunt, J.W. Salisbury and G.J. Lenhoff, 'Visible and Near Infrared of Minerals and Rocks III'. Oxides and hydroxides. *Modern Geology*, vol. 2, pp. 195–205, 1971.
- [2] R. Schowengerdt, 'Remote Sensing, Models and Methods for Image Processing. Academic press, Toronto, 522 pages, 1997.
- [3] P. Huvelin, 'Carte Géologique et des Minéralisations des Jebilet Centrales au 1/100 000'. Sous-secrétariat d'état au commerce, à l'industrie, aux mines et de la géologie. Royaume du Maroc. Notes et mémoires, no 232a, 1970.
- [4] A. Bannari, P.M. Teillet, et G. Richardson, 'Nécessité de l'étalonnage radiométrique et standardisation des données de télédétection'. *Canadian Journal of Remote Sensing*, vol. 25, no.1, pp. 45–59, 1999.
- [5] A. Bannari, D. Morin, G.B. Bénéï and J.F. Bonn, 'A Theoretical Review of Different Mathematical Models of Geometric Corrections Applied to Remote Sensing Images'. *R.S. Reviews*, vol. 13, pp. 27–47, 1995.
- [6] P.M. Teillet and R.P. Santer, 'Terrain Elevation and Sensor Altitude Dependence in Semi-Analytical Atmospheric Code'. *Canadian Journal of Remote Sensing*, vol. 17, no 1, pp. 36–44, 1991.
- [7] PCI Geomatics (2006) Using PCI software. Richmond Hill, ON. 540 pages.
- [8] A.P. Crosta, C. Sabine and J.V. Taranik, 'Hydrothermal Alteration Mapping at Bodie, California, using AVIRIS Hyperspectral Data'. *Remote Sensing of Environment*, vol. 65, pp. 309–319, 1998.
- [9] G. Girouard, A. Bannari, A. El-Harti and A. Desrochers, 'Validated Spectral Angle Mapper Algorithm for Geological Mapping: Comparative Study between Quickbird and Landsat-TM'. *Geo-Imagery Bridging Continents Istanbul, Turkey*, July 12–23, pp. 599–604, 2004.
- [10] I. McCubbin, R. Green, H. Lang and D. Roberts, 'Mineral Mapping Using Partial Unmixing at Ray Mine, AZ'. *Summaries of the 8th JPL Airborne Earth Science Workshop*, JPL Publication 97–21, 4 pages, 1998.
- [11] F. Van-der-Meer, M. Vasquez-Torres and P.M. Van Dijk, 'Spectral Characterization of Ophiolite Lithologies in the Troodos Ophiolite Complex of Cyprus and its Potential in Prospecting for Massive Sulphide Deposits'. *Int.J.of R.S.*, vol. 18, no. 6, pp. 1245–1257, 1997.
- [12] F.A. Kruse, J.W., Boardman, A.B., Lefkoff, K.B., Heidebrecht, A.T., Shapiro, P.J., Barloon, and A.F.H. Goetz, 'The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data'. *R.S. of Env.* vol. 44, pp.145–163, 1993.
- [13] J. Schwarz, and K. Staenz, 'Adaptive Threshold for Spectral Matching of Hyperspectral Data'. *Canadian Journal of Remote Sensing*, vol. 27, no 3, pp. 216–224, 2001.
- [14] L.C. Rowan, and J.C. Mars, 'Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborn' Thermal Emission and Reflection Radiometer (ASTER) data. *R.S. of Env.*, vol. 84, pp. 250–266, 2003.
- [15] A.P. Crosta, C., Sabine, and J.V. Taranik, 'Hydrothermal Alteration Mapping at Bodie, California, using AVIRIS Hyperspectral Data'. *Remote Sensing of Environment*, Vol. 65, p. 309–319, 1998.
- [16] O.A. De-Carvalho, and P.R. Meneses, 'Spectral Correlation Mapper (SCM); An Improvement on the Spectral Angle Mapper (SAM)'. *Summaries of the 9th JPL Airborne Earth Science Workshop*, JPL Publication 00–18, 9 pages, 2000.
- [17] M.S.K. Gebbinck, 'Decomposition of Mixed Pixels in Remote Sensing Images to Improve the Area Estimation of Agricultural Fields'. Ph.D. thesis, Katholieke Universiteit Nijmegen, Veenendaal Universal Press, 165 pages, 1998.
- [18] J.B. Adams, M.O. Smith and A.R. Gillepsie, 'Imaging spectroscopy: Interpretation based on spectral mixture analysis'. In *Remote Geochemical Analysis: Elemental and Mineralogical Composition*. (C.M. Pieters and P.A.J. Englert, Eds). Cambridge University, Cambridge. pp. 145–166, 1993.
- [19] K. Staenz, C. Nadeau, J. Secker and P. Budkewitsch, 'Spectral unmixing applied to vegetated environments in the Canadian Arctic for mineral mapping'. *Proceedings of XIX ISPRS Congress*. Amsterdam, The Netherlands. July 15–23, CD-ROM, 8 pages, 2000.
- [20] A.F.H. Goetz, G. Vane, J.E. Solomon and B.N. Rock, 'Imaging spectrometry for earth remote sensing'. *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.
- [21] L. Zhang, D. Li, Q. Tong and L. Zheng, 'Study of the spectral mixture model of soil and vegetation in PoYang Lake area, China'. *Int. J. of Remote Sensing*, vol. 19, no. 11, pp. 2077–2084, 1998.
- [22] R.B. Myneni, S. Maggion, J. Jaquinta, J.L. Privette, N. Gordon, B. Pinty, D.S. Kimes, M.M. Verstraete and D.L. Williams, 'Optical remote sensing of vegetation: modeling, caveats, and algorithms'. *Remote Sensing of Environment*, vol. 51, no. 1, pp. 169–188, 1995.
- [23] C. Wu, and A.T. Murray, 'Estimating Impervious Surface Distribution by Spectral Mixture Analysis'. *R.S. of Environment*, vol. 84, pp. 493–505, 2003.
- [24] R.H. Yuhas, A.F.H., Goetz, and J.W. Boardman, 'Discrimination Among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm'. *Summaries of the 4th JPL Airborne Earth Science Workshop*, JPL Publication 92–41, pp.147–149, 1992.

Monitoring of disaster and restoration of eastern Japan tsunami by satellite images

H. Hashiba

Department of Civil Engineering, College of Science and Technology, Nihon University, Tokyo, Japan

ABSTRACT: The earthquake and tsunami that occurred on March 11, 2011 affected the entire Eastern Japan. Several urban areas and environmental habitats, such as tidal flats and forests that serve as protection against storm surges, were extensively damaged by the tsunami. The state of these areas immediately after the tsunami struck was observed by using satellite remote sensing. In addition, these techniques will be encouraged for future long-term monitoring of the damaged areas. In this study, the damage around a typical wildlife park in a coastal area was investigated by land cover classification using the ALOS satellite image data before and after the tsunami along with the ground research. In addition, the restoration process was continuously monitored with the ALOS/AVNIR-2, THEOS, EO-1/ALI, and Landsat-8/OLI for three years after the tsunami's impact. The inundation caused by the tsunami immediately after its impact was evaluated, and the characteristics of the restoration process were revealed.

1 INTRODUCTION

A large-scale disaster caused by an earthquake and tsunami occurred across the entire region of eastern Japan on March 11, 2011. More than 300 km of the Tohoku coastal regions in eastern Japan suffered extensive damage from the tsunami that struck the coast. Many urban areas, farmlands, and fishing ports were inundated by the tsunami (MLITT 2011). In addition, a lot of damage occurred in large areas of natural environments and a wildlife park in the coastal area. Satellite remote sensing can provide important information for the understanding and assessment of impacted areas after a large-scale natural disaster. For example, many observation missions were executed after the tsunami disaster that occurred off the coast of Sumatra, and diverse strike information was used for disaster investigations (Singh et al. 2007, Kumar et al. 2007, etc.). Many of the Sumatran earthquake disaster studies used high-resolution satellite images and were able to provide detailed reports (Wikantika et al. 2007, Yang et al. 2007, Liu et al. 2007, Anandan & Sasidhar 2011). In addition, satellite remote sensing has been used to analyze the effectiveness of tsunami disaster prevention measures for protecting forests against storm surges in coastal regions (Olwig et al. 2007). A lot of research has focused on the tsunami hazard from the 2011 East Japan earthquake. As an example, the tsunami disaster that occurred on the coast in the eastern region of Japan was widely investigated by using multi-spectral data with a 10×10 m spatial resolution provided by the ALOS satellite over

Japan. Using the ALOS image data, the entire image of the strike area and its relationship to geographical features were examined (Hashiba & Sugimura 2011, Rao & Lin 2011, etc.). In addition, many other satellite observations were utilized by a research consortium in which the author participated (Scientific research working group for high resolution satellite remote sensing /RSSJ 2011). The scale and features of the 2011 tsunami have been reported in various measurement results (Dunbar et.al. 2011 etc).

In the future, it will be necessary to continuously monitor the changes after a disaster through remote sensing to support restoration. However, few studies have used long-term remote sensing to monitor the restoration process of the stricken areas. In addition, it is necessary to use two or more different satellite images for more effective, long-term time series investigations.

The present research focused on a coastal wildlife park and an adjacent natural environment located within the area struck by the 2011 tsunami. The surrounding Matsukawa Bay, located in the Souma City in the Fukushima Prefecture, was selected as a typical site in which the coastal wildlife park and the natural environment were located. The area inundated by the tsunami was determined by land cover classification processing by using the ALOS satellite images from immediately before and after the tsunami strike. The extension of the inundation was evaluated from this classification. Moreover, the long-term changes in the land cover of the park zone were interpreted from a time series data set of ALOS/AVNIR-2, THEOS,

EO-1/ALI, and Landsat-8/OLI images ranging from before the tsunami to three years after the tsunami. The long-term characteristics of the changes in land cover caused by the tsunami strike in the wildlife park were detected.

2 TEST SITE AND SATELLITE IMAGE DATA

2.1 Test site

The Matsukawa Bay Wildlife Park, located in the coastal region of Souma City, Fukushima Prefecture, is composed of tidal flats and wooded areas (Fig. 1).

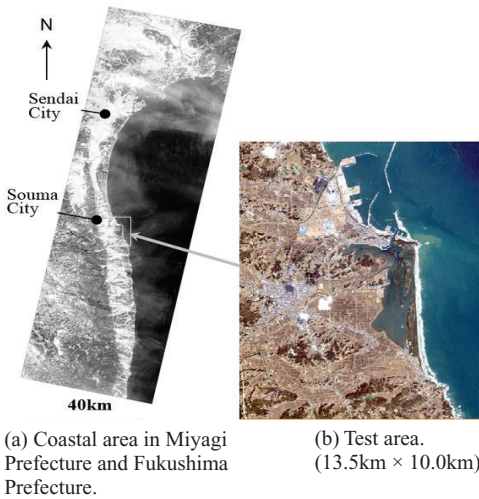


Figure 1. Research area (Matsukawa Bay Wildlife Park and surroundings).

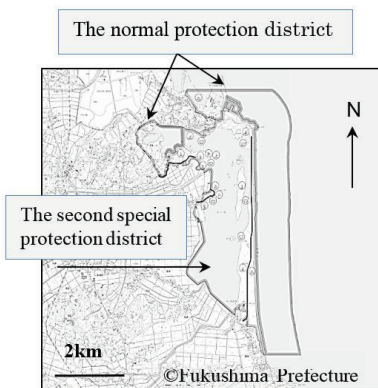


Figure 2. The Protection planning map of Matsukawa Bay Wildlife Park planned by Nature Conservation Bureau in Fukushima Prefecture (Fukushima Prefecture 2011).

The area of the park is 979 ha. Pine forests grow thickly on the sand bar facing the Pacific Ocean, and many living species are endemic to the tidal flats in the bay area (Aoki et al. 2003). The average depth of the tidal flat is about 1 m and the maximum depth is about 4 m. Agricultural fields are widely distributed in the plains around the park. This study examined a rectangular area of 10.5×12.8 km that includes the wildlife park (Fig. 1). The park administration recognizes two different protection plan districts, as shown in Figure 2 (Fukushima prefecture 2011). One is a range of coast, usually referred to as the plan district, and the other is a range referred to as the special district region, which contains woods and tidal flats.

2.2 Satellite image data

Image data from two or more different satellites was used to investigate the long-term changes of the tsunami-stricken area. The parameters of the satellite data are given in Table 1.

Table 1. Used Satellite image data in this study.

| Satellite/Sensor | Used Bands in this study (unit:µm) | Spatial Resolution | Acquisition Date |
|---------------------|------------------------------------|---|-------------------------|
| ALOS-1/AVNIR-2 | B1(blue, 0.42-0.50) | 10m×10m | 2011/2/23 |
| | B2(green, 0.52-0.60) | | 2011/3/14 |
| | B3(red, 0.61-0.69) | | 2011/4/1 |
| | B4(near infrared, 0.76-0.89) | | |
| THEOS/Multispectral | B0(blue, 0.45-0.52) | 15m×15m | 2011/6/11 |
| | B1(green, 0.53-0.60) | | |
| | B2(red, 0.62-0.69) | | |
| EO-1/ALI | B3(near infrared, 0.77-0.90) | Pan:10m×10m, Multispectral:30m×30m, Pan-sharpen(B1,B2,B3): 10m×10m | 2011/11/22 2013/4/23 |
| | Pan(0.480-0.690) | | |
| | B1(blue, 0.450-0.515) | | |
| | B2(green, 0.525-0.605) | | |
| Landsat-8/OLI | B3(red, 0.630-0.690) | Pan:15m×15m, Multispectral:30m×30m, Pan-sharpen(B2,B3,B4): 15m×15m | 2014/4/13 |
| | B4(near infrared, 0.775-0.805) | | |
| | Pan(0.50-0.68) | | |
| | B2(blue, 0.45-0.51) | | |
| | B3(green, 0.53-0.59) | | |
| B4(red, 0.64-0.67) | | | |
| | B5(near infrared, 0.85-0.88) | | |

3 METHOD

3.1 Classification analysis to assess the inundation by the tsunami

Eight-band ALOS/AVNIR-2 images with accurately matched coordinates from before (from band 1 to 4) and after (from band 1 to 4) the tsunami's impact was composed. The inundated area and the area of disappeared trees caused by the tsunami were classified by the maximum likelihood classification method by using all the eight bands from the image data before and after the strike. The sampling site for several kinds of training data for classification was determined based on the location of the inundated area and the area of disappeared trees. Ten kinds of training data were sampled for classifying the inundated area. Ten locations within the sampling area were selected in which the difference between the color and brightness of the surface of the inundation

was distinguishable by the use of image interpretations before and after the impact. Training data for classifying the disappeared tree region was measured in two locations in which the tree disappearance was clearly distinguishable from image interpretations before and after the impact. The validity of the sampling area was confirmed by image interpretations and field investigations.

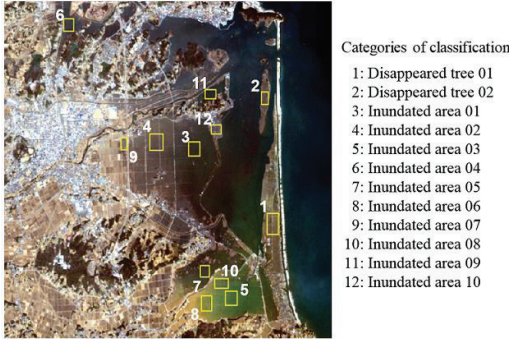


Figure 3. Sampling areas of training data for image classification.

3.2 Site investigations immediately after the tsunami strike

Ground investigations were executed in the target region on May 14, 2011. This detailed investigation confirmed the local environmental condition, and the condition after the actual strike was validated by the satellite images and analytical results.

3.3 Image interpretations for change detection of the tsunami strike's effects and restoration

The short-term changes were interpreted by using the ALOS/AVNIR-2 images from before the disaster and approximately three weeks after the disaster. The long-term change was interpreted using the THEOS, EO-1/ALI, and Landsat-8/OLI images from about four months to three years after the disaster. For more detailed interpretations, the multispectral images of EO-1/ALI, and Landsat-8/OLI were pan-sharpened. The widely used technique developed by Zhang (2002) was used for the pan-sharpening process.

4 RESULTS AND DISCUSSION

4.1 Land cover changes in the Wildlife Park after the tsunami strike

The disappearance of the forest in the wildlife park due to the tsunami strike and the inundation of the

surrounding areas were determined from the image analysis by classification processing. The sampling area of the classification training data for the classification processing is shown in Figure 3. The sampling area of ten kinds of inundations (indicated by No.3–No.12 in Fig. 3) was determined from the level of surface conditions determined by the image interpretations. The classification of the distribution of the disappearance of trees in the coastal woods for two other sampling areas (indicated by No.1, 2 in Fig. 3) was determined from the image interpretations. The large scale of this disaster allowed the sampling area of the classification standard to be set sufficiently from the observation images at a spatial resolution of 10×10 m, based on the image interpretations. The classification results are shown in Figure 4.

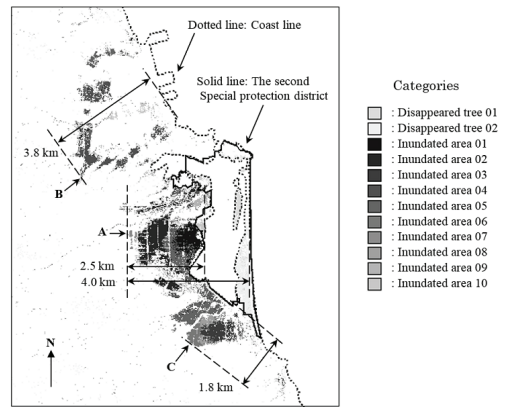


Figure 4. The Classification result of area of disappeared trees and inundated area by the tsunami disaster.

The inundated area widely extended into the outlying farmlands beyond the park. The tsunami extended inland from the bay by approximately 4 km from

Table 2. The Classification result of the whole of test site.

| Category | Code | Pixels | Area (km \times km) | Ratio (%) |
|-----------------------------------|------|------------------|-----------------------|-----------|
| Disappeared trees 01 | 1 | 5312 | 0.53 | 0.39 |
| Disappeared trees 02 | 2 | 11527 | 1.15 | 0.85 |
| Inundated area 01 | 3 | 6042 | 0.60 | 0.45 |
| Inundated area 02 | 4 | 16729 | 1.67 | 1.23 |
| Inundated area 03 | 5 | 5142 | 0.51 | 0.38 |
| Inundated area 04 | 6 | 8657 | 0.86 | 0.64 |
| Inundated area 05 | 7 | 17814 | 1.78 | 1.31 |
| Inundated area 06 | 8 | 2461 | 0.25 | 0.18 |
| Inundated area 07 | 9 | 7164 | 0.71 | 0.53 |
| Inundated area 08 | 10 | 3144 | 0.31 | 0.23 |
| Inundated area 09 | 11 | 935 | 0.09 | 0.07 |
| Inundated area 10 | 12 | 10287 | 1.03 | 0.76 |
| Other(Non inundated area) | 0 | 1261284 | 125.88 | 92.98 |
| Total | | 1356498 | 135.38 | 100 |
| Total area of disappeared trees : | 1.68 | (km \times km) | | |
| Total area of inundated area : | 7.82 | (km \times km) | | |

the coastline on the boundary of the Pacific Ocean, as indicated by point A on Figure 4. The tsunami reached a position of approximately 2.5 km from the coastline of the bay in the inland direction, and then extended approximately 3.8 km from the coastline of the Pacific Ocean to point B. The tsunami extended approximately 1.8 km from the coastline of the bay inland to point C. Most of the forest belt along the coastline of the boundary with the Pacific Ocean disappeared, and most of the trees in the second protection plan district were damaged.

This inundated area and the area of disappeared trees were assessed in Table 2, which showed the classification resulted from the entire test site. The total area of disappeared trees in this object area was 1.68 km², and the total inundated area was 7.82 km². A confusion matrix was used to assess the classification accuracy (Table 3). The average accuracy of the classification processing was 96.29%, and the overall accuracy was 95.65%. The extraction result was confirmed to be close to excellent, although a few error margins were admitted for the image interpretation.

Table 3. The assessment of classification result by confusion matrix.

| Category | Code | Pixels | Code | | | | | | | | | | | | | | |
|-------------------|------|--------|------|------|------|------|-----|------|------|------|------|----|------|------|------|------|------|
| | | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | | |
| Disappeared trees | 01 | 1 | 322 | 1.86 | 93.2 | 4.04 | 0 | 0.62 | 0 | 0 | 0 | 0 | 0.31 | 0 | 0 | 0 | |
| Disappeared trees | 02 | 2 | 1504 | 3.66 | 1.6 | 94.3 | 0 | 0 | 0 | 0 | 0 | 0 | 0.4 | 0 | 0 | 0.07 | |
| Inundated area | 01 | 3 | 1015 | 1.08 | 0 | 0 | 97 | 0 | 0 | 0 | 1.97 | 0 | 0 | 0 | 0 | 0 | |
| Inundated area | 02 | 4 | 1221 | 3.69 | 0 | 0 | 0 | 95.5 | 0 | 0 | 0 | 0 | 0.41 | 0 | 0 | 0.41 | |
| Inundated area | 03 | 5 | 930 | 2.47 | 0 | 0 | 0 | 0 | 93.6 | 0 | 0 | 0 | 0 | 3.98 | 0 | 0 | |
| Inundated area | 04 | 6 | 480 | 1.25 | 0 | 0 | 0 | 0 | 0 | 98.3 | 0.21 | 0 | 0 | 0 | 0 | 0.21 | |
| Inundated area | 05 | 7 | 504 | 1.59 | 0 | 0 | 0.2 | 0.79 | 0 | 0.4 | 97 | 0 | 0 | 0 | 0 | 0 | |
| Inundated area | 06 | 8 | 630 | 2.7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 97 | 0 | 0.32 | 0 | 0 | |
| Inundated area | 07 | 9 | 128 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | |
| Inundated area | 08 | 10 | 459 | 1.96 | 0 | 0 | 0 | 0 | 3.7 | 0 | 0 | 0 | 0.22 | 0 | 94.1 | 0 | 0 |
| Inundated area | 09 | 11 | 248 | 2.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 98 | 0 |
| Inundated area | 10 | 12 | 210 | 1.9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.48 | 0 | 0 | 0 | 97.6 |

Average accuracy = 96.29 %
Overall accuracy = 95.65 %

(code 0: other)

4.2 Analysis of the strike situation through site investigations

Figure 5 showed local photographs taken in the tsunami struck area by a local survey about two months after the strike on May 14, 2011. Most areas looked much like what they had immediately after the tsunami receded. The breakwater and the shore protection woods, which faced the Pacific Ocean, essentially disappeared in some places (Photographs from No.3 to No.6). Many forests inside the breakwater disappeared, as indicated by uprooted and fallen trees (Photographs No.1 and No.2). This local investigation confirmed the inundation of farmland in the coastal plain (Photographs No.7 and No.8). Salt damage was visually confirmed in the areas where seawater had receded (Photograph No.9). Most of the shore protection woods were knocked down. Investigation of several areas where trees had disappeared showed the ground was covered with marine sand (Photograph No.10). The damage to areas containing dwellings and farmlands around the wildlife park was extensive (Photographs No.11 and No.12).

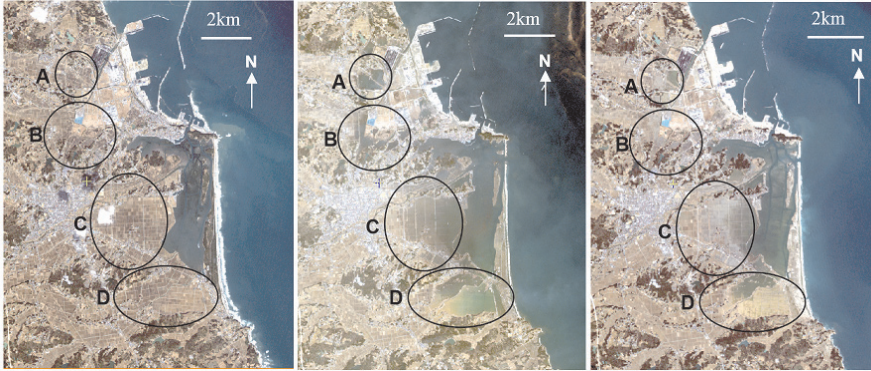


Figure 5. Ground photographs after 2 months from disaster (Investigated date: May 14, 2011) and ALOS/AVNIR-2 image observed 3 weeks after the disaster.

4.3 *Detection of the inundation and restoration process through a time series*

The disappearance of most storm surge protection forests was determined from the image immediately

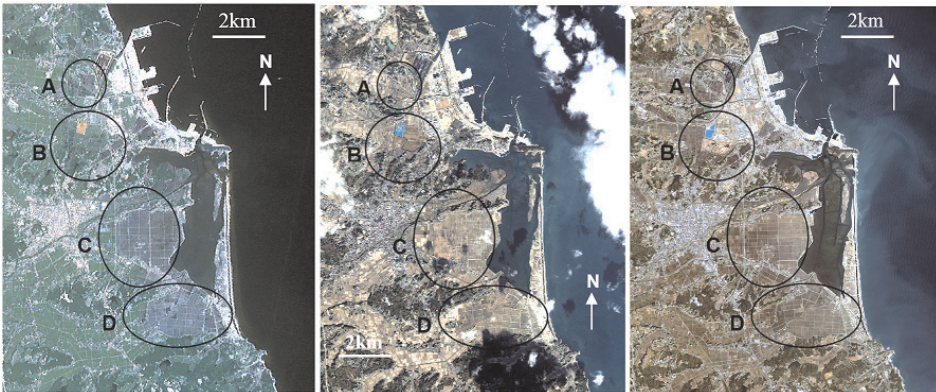
after the tsunami's impact. The seawater inundation of the plains was interpreted from the observation images in the areas indicated as A–D in [Figure 6](#).



(a) Before disaster observed on Feb. 23, 2011 (ALOS/AVNIR-2).

(b) 3days after the disaster observed on Mar.14, 2011 (ALOS/AVNIR-2).

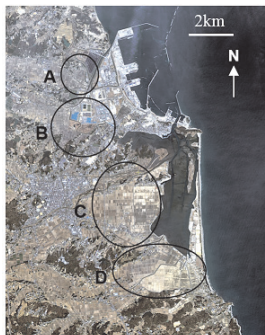
(c) 3weeks after the disaster observed on Apr.1, 2011 (ALOS/AVNIR-2).



(d) 4months after the disaster observed on Jul.11, 2011 (THEOS).

(e) 8months after the disaster observed on Nov.22, 2011 (EO-1/ALI).

(f) 2 years after the disaster observed on Apr.23, 2013 (EO-1/ALI).



(g) 3 years after the disaster observed on Apr.13, 2014 (Landsat-8/OLI).

Figure 6.(a)–(g). The change in land cover by using several kind of satellite images. (ALOS/AVNIR-2: Included ©JAXA and Distribution RESTEC, THEOS: ©2011 GISTDA ALL RIGHT RESERVED, EO-1/ALI, Landsat-8/OLI: These data are courtesy of the U.S. Geological Survey.).

The inundation was especially extensive in the farmlands (area C in Figs. 6(a)-6(c)) around the wildlife park and in the area containing dwellings (area D in Figs. 6(a)-6(c)). The recession of inundation in these areas could be confirmed after one week on the basis of image interpretation. This indicated that the progress of the natural drainage of the flooded area was difficult immediately after the strike. The drainage progressed gradually in the areas A–D three weeks after the tsunami strike. However, many spots in area C were still inundated.

The inundation was especially extensive in the farmlands (area C in Figs. 6(a)-6(c)) around the wildlife park and in the area containing dwellings (area D in Figs. 6(a)-6(c)). The recession of inundation in these areas could be confirmed after one week on the basis of image interpretation. This indicated that the progress of the natural drainage of the flooded area was difficult immediately after the strike. The drainage progressed gradually in the areas A–D three weeks after the tsunami strike. However, many spots in area C were still inundated.

The THEOS image taken about four months after the tsunami confirmed that the inundation of areas A and B was mostly receded (Fig. 6(d)). The image revealed that in many of the inundated areas, many linear road structures of the farmland had not been greatly destroyed. Such defensive features were considered to be one of the effect of defense of the storm surge protection forests and the terrain condition of the Wildlife Park basin against the tsunami. Areas where the land surface was damp or wet because of consecutive rainfall during the rainy season were also considered.

The recession of the inundation throughout the areas in A–D was observed from the EO-1/ALI image that was taken about eight months after the tsunami (Fig. 6(e)). In addition, the progress of the re-leveling of farmland was assessed. The division of the farmland was identified more distinctly compared with what was observed after four months. For the division of the farmland, a tendency towards restoration was interpreted.

The remarkable change in condition of land cover was not so interpreted in the EO-1/ALI image taken about two years after the tsunami compared with the one taken after eight months (areas C and D in Fig. 6(f)). Based on these images, the delay of progress of the social infrastructure and farmland maintenance and restoration were detected.

On the other hand, the Landsat-8/OLI image that was observed after about three years confirmed the progression of the new leveling of land (Fig. 6(g)). The continual development of farmland aiming at its reproduction was identified in areas C and D. From these analyses, the time taken for the restoration of the affected areas to begin to progress was evaluated.

5 CONCLUSION

This study focused on the long-term environmental changes caused by the tsunami in the Matsukawa Bay Wildlife Park, Fukushima Prefecture and the surrounding area as determined by using the ALOS/AVNIR-2, THEOS, EO-1/ALI, and Landsat-8/OLI images. The analyses of the ALOS/AVNIR-2 images and field investigations revealed that the tsunami almost eradicated the wildlife park. Image analyses revealed that the tsunami caused extensive inundation of the farmlands around the park on the inland side. Most inundated areas were not completely drained up to three weeks after the strike, and the restoration of the environment will still take a long time. The THEOS image observed four months after the tsunami showed that much of the inundation had receded. In addition, some defensive features against the tsunami were considered in this coastal Wildlife Park and surrounding area. Based on the EO-1 /ALI image observed after eight months, the initial motion of restoration was interpreted in many inundated farmland. However, the delay of restoration was analyzed by the EO-1/ALI image observed after two years. Based on the Landsat-8/OLI image observed after three years, the tendency that the new restoration started was evaluated by interpreting the change in surrounding farmland in the Wildlife Park.

From these results, the effectiveness of the continuous remote sensing in the time series investigation of the affected area was shown. In the future, more effective spatial information will be extracted and continual restoration assessments will be made to support decision-making regarding future reconstructions and disaster mitigation.

ACKNOWLEDGMENTS

The ALOS/AVNIR-2 data used to investigate this research received a distribution from JAXA (Japan Aerospace Exploration Agency) and RESTEC (Remote Sensing Technology Center of Japan). The THEOS data were provided by the Geo-Informatics and Space Technology Development Agency (GISTDA) in Thailand and RESTEC in Japan. The EO-1/ALI data were courtesy of the U.S. Geological Survey (The USGS home page is <http://www.usgs.gov>).

REFERENCES

- [1] Anandan, C. & Sasidhar, P. 2011. Changes in coastal morphology at Kalpakkam, East Coast, India due to 26 December 2004 Sumatra tsunami. *Geomatics, Natural Hazards and Risk* 2(2): 183–192.

- [2] Aoki, K., Horii, Y., Horikawa, Y., Isogai, F., Kubota, Y., Kurita, Y., Matsuki, T., Mizuno, T. & Toyooka, A. 2003. Study on the lebensspuren of benthic fauna on the tidal flats of the Matsukawa-ura, Soma City, Fukushima Prefecture. *Earth Science* 57(1/2): 31–48.
- [3] Dunbar, P., Mccullough, H., Mungov, G., Varner, J. & Stroker, K. 2011. 2011 Tohoku earthquake and tsunami data available from the National Oceanic and Atmospheric Administration/ National Geophysical Data Center. *Geomatics, Natural Hazards and Risk* 2(4): 305–323.
- [4] Fukushima Prefecture. 2011. 2010 the park plan map of Matsukawa bay prefectural Wildlife Park. http://www.cms.pref.fukushima.jp/download/1/shizen_h22_matukawaurakenritushizenkouenkouenkeikakuzu.pdf: (accessed June 10, 2011).
- [5] Hashiba, H. & Sugimura, T. 2011. Research of tsunami disaster caused by the 2011 off the Pacific coast of Tohoku Earthquake by using ALOS/ AVNIR-2 image and DEM. *Proceedings of the 50th Spring Conference of the Remote Sensing Society of Japan*: 263–264.
- [6] Kumar, A., Chingkhel, R.K. & Dolendro, T.H. 2007. Tsunami damage assessment: a case study in Car Nicobar Island, India. *International Journal of Remote Sensing* 28(13–14): 2937–2959.
- [7] Liu, C.C., Liu, J.G., Lin, C.W., Wu, A.M., Liu, S.H. & Shieh, C.L. 2007. Image processing of FORMOSAT-2 data for monitoring the South Asia tsunami. *International Journal of Remote Sensing* 28(13–14): 3093–3111.
- [8] Ministry of Land, Infrastructure, Transport and Tourism (MLIT). 2011. Great East Japan Earthquake. http://www.mlit.go.jp/page/kanbo01_hy_001411.html: (accessed June 10, 2011).
- [9] Olwig, M.F., Sorensen, M.K., Rasmusen, M.S., Danielsen, F., Selvam, V., Hansen, L.B., Nyborg, L., Vestergaard, K.B., Parish, F. & Karunagaran, V.M. 2007. Using remote sensing to assess the protective role of coastal woody vegetation against tsunami waves. *International Journal of Remote Sensing* 28(13–14): 3153–3169.
- [10] Rao, G. & Lin, A. 2011. Distribution of inundation by the great tsunami of the 2011 Mw 9.0 earthquake off the Pacific coast of Tohoku (Japan), as revealed by ALOS imagery data. *International Journal of Remote Sensing* 32(22): 7073–7086.
- [11] Scientific Research Working Group for High Resolution Satellite Remote Sensing/RSSJ. 2011. High Resolution Satellite Remote Sensing concerning the 2011 off the Pacific Coast of Tohoku Earthquake and Tsunami Disaster. *Journal of the Remote Sensing Society of Japan* 31(3): 344–367.
- [12] Singh, R.P., Cervone, G., Kafatos, M., Prasad, A.K., Sahoo, A.K., Sun, D., Tang, D.L. & Yang, A, R. 2007. Multi-sensor studies of the Sumatra earthquake and tsunami of 26 December 2004. *International Journal of Remote Sensing* 28(13–14): 2885–2896.
- [13] Soma City, Fukushima Prefecture. 2011. Matsukawa bay prefectural Wildlife Park. http://www.city.soma.fukushima.jp/kanko/matukawaura_index.html: (accessed June 10, 2011).
- [14] Wikantika, K., Sinaga, A., Hadi, F. & Darmawan, S. 2007. Quick assessment on identification of damage building and land-use changes in the post-tsunami disaster with a quick-look image of IKONOS and Quickbird (a case study in Meulaboh City, Aceh). *International Journal of Remote Sensing* 28(13–14): 3037–3044.
- [15] Yang, M.D., Su, T.C., Hsu, C.H., Chang, K.C. & Wu, A.M. 2007. Mapping of the 26 December 2004 tsunami disaster by using FORMOSAT-2 images. *International Journal of Remote Sensing* 28(13–14): 3071–3091.
- [16] Y. Zhang. 2002. Problems in the fusion of commercial high-resolution satellite as well as Landsat 7 images and initial solutions. In ISPRS, Vol. 34, Part 4, *Geo Spatial Theory, Processing and Applications*, Ottawa, 2002. Canada.
- [17] Y. Zhang. 2002. A new automatic approach for effectively fusing Landsat 7 as well as IKONOS images. *Proceedings of IEEE IGARSS 2012*, Toronto, Canada: 2429–2431.

Application of PCA and KLE in high-rate GPS positioning

X.Y. Yu, Y.J.Hu & K.H. Ding

Faculty of Information Engineering, China University of Geosciences, Wuhan, China

M.S. Ao

School of geosciences and info-physics, Central South University, Changsha, China

ABSTRACT: Multipath error and random noise are two important error sources in high-rate GPS positioning. Some methods, including those based on sidereal day filtering and digital signal processing, have been developed to mitigate multipath errors. To improve the accuracy of high-rate GPS positioning, this paper uses Principal Component Analysis (PCA) and Karhunen-loeve Expansion (KLE) to evaluate the random noise level and mitigate the multipath error in the coordinate time series of GPS stations. The experiment results show that the response ratio of the principal component coefficients of PCA can effectively reflect the impact of strong random noises on coordinate time series, and the multipath error can be significantly mitigated by PCA transformation. Compared to PCA, the response ratio of the principal component coefficients of KLE is less sensitive to random noise and its ability to reduce multipath error is also weaker. However, KLE can preferably suppress random noises and local anomalies.

1 INTRODUCTION

As high-rate GPS receivers that can be operated at 1Hz, 20Hz, or 50Hz sampling rate are increasingly used in GPS networks, a large amount of high-rate GPS observation data are being obtained and used in the analysis of geophysical signals, such as seismic wave observation and causes of earthquake.

To improve the accuracy of high-rate GPS positioning in geophysical applications, it is important to mitigate the multipath error and random noise. In the early 1990s, Genrich and Bock (1992) recognized that the multipath error can be mitigated by utilizing the sidereal day filtering. Many scholars have studied the correlations between GPS coordinates and the sidereal day repetition cycle to mitigate the multipath errors (Elosequi et al., 1995; Ding et al., 1999; Radovanovic, 2000;). Choi et al. (2004) proposed a modified sidereal day filtering method to mitigate the multipath effect, which used the real repetition cycles of observed satellites. Larson et al. (2007) proposed the Aspect Repeat Time Adjustment (ARTA) algorithm based on coordinate domain to improve the positioning accuracy of high-rate GPS.

Besides the methods based on sidereal day filtering, some digital signal processing methods have also been utilized in the mitigation of multipath effect. Xia and Liu (2001) used the Finite Impulse Response filter to extract the multipath error of pseudo-range code. Satirapod and Rizos (2005) used wavelet analysis to extract the multipath error from the residuals

of double differential observation solutions. However, digital signal processing methods are based on the prior knowledge of the characteristic frequency of multipath signals. Therefore, improving our understanding and description of the multipath effect, and adjusting the parameters for filtering are still worth investigating.

Principal Component Analysis (PCA) and Karhunen-loeve expansion (KLE) have been widely used in geophysical signal processing. Dong et al. (2006) used PCA and KLE to describe and remove the Common Mode Error that widely existed in regional GPS networks. It was demonstrated that PCA could reduce the error by 87.6%, 73.1% and 74.9%, respectively in north, east and vertical directions.

Inspired by Dong's work, we used PCA and KLE to evaluate the random noise level and mitigate the multipath error in high-rate GPS positioning in this study.

2 THE PRINCIPLES AND METHODS OF PCA AND KLE

The core idea of PCA or KLE is to transform original variables into a series of new variables which are irrelevant to each other. These new variables are used to represent the original variables and reflect back enough information in the original ones. Ideally, the GPS coordinate time series on many days in the short baseline situation should be distributed on a straight line.

However, due to the multipath effect and random noises, the coordinates are fluctuated as a function of time. When PCA is used to process a coordinates time series of GPS positioning of n days, if there are m epochs each day, the error matrix can be described as where x_{ij} represents the error value at epoch i of the day j .

$$X(i, j), (i = 1, \dots, m, j = 1, \dots, n) \quad (1)$$

We assume $m > n$. Then, the covariance matrix of X can be defined as

$$B = \frac{1}{m-1} X^T X \quad (2)$$

where covariance matrix B is symmetric, the size of which is $(n \times n)$. B can be decomposed as

$$B = V \Lambda V^T \quad (3)$$

where, the matrix V , whose size is $(n \times n)$, is an orthonormal matrix with orthonormal columns. The matrix Λ , which is composed of k nonzero eigenvalues, is a diagonal matrix. Actually, in most case, the rank of matrix B is full, which means $k = n$. According to the linear algebra theory, the error matrix X can be expanded with the orthonormal function basis V :

$$X = AV \quad (4)$$

where, the size of A is $(m \times n)$. If we arrange the feature vectors in the descending order according to the eigenvalues, the matrix A can be composed of n components calculated as

$$A = XV^T \quad (5)$$

Conversely, the approximate error matrix X' of the original error matrix X can be represented as

$$X' = A_{m \times l} V_{l \times n} \quad (6)$$

where, $A_{m \times l}$ is composed of l components and $V_{l \times n}$ is the coefficients matrix corresponding to the components.

As mentioned above, the descending order of the eigenvalues indicates that the first component represents the greatest contribution to the covariance of the error time series of multiple days. The contribution ratio of each component is defined as the according eigenvalue divided by the sum of all eigenvalues. That is to say, if the coordinate error of each day is

mainly the multipath error, considering the repetition of multipath error, we can select one component or a combination of several components ranked in the top to represent the effect of multipath error on coordinate series of each day.

If we use the covariance vector σ to standardize the covariance matrix, a correlation matrix C can be obtained, for example,

$$c_{ij} = b_{ij} / (\sigma_i \sigma_j)$$

The covariance vector σ is defined as follows

$$\sigma_j = \sqrt{b_{jj}} = \sqrt{\frac{1}{m-1} \sum_{k=1}^m X^2(k, j)} \quad (7)$$

Similarly, the correlation matrix C can be decomposed as

$$C = W \Lambda_c W^T \quad (8)$$

The error matrix X , component matrix A and approximate error matrix X' can be described respectively by equations as follow

$$X = AW \quad (9)$$

$$A = XW^T \quad (10)$$

$$X' = A_{m \times l} V_{l \times n} \quad (11)$$

Equation (11) is the result of Karhunen-loeve expansion. It is obvious that the only difference between PCA and KLE is that the former uses the covariance matrix and the latter uses the correlation matrix to calculate the orthonormal matrix.

If we rescale the error matrix X as

$$X_k(i, j) = X(i, j) / \sigma_j \quad (12)$$

and σ_j in the formula (12) is still defined according to the formula (7), the result of KLE of the original time series is equivalent to the one of PCA of the rescaled time series.

3 COLLECTION AND PROCESSING OF RESEARCH DATA

3.1 Data collection

The GPS data for this research was collected on the roof of Building No.64 on the campus of China University of Geosciences (Fig.1). Two GPS stations

(REFF and MP01) were set up on the roof of the building, and the linear distance between them was 40.95m. There was no obvious multipath source around station REFF, so it was used as a reference station. There was an obvious multipath source around station MP01, a white wall, which was 3.15m high and about 2 meters next to the station. The observation time started on the 27th and ended on the 30th in December (361st to 364th day of year), 2010.



Figure 1. Locations and Environments of MP01 and REFF stations.

The two stations were both equipped with the same type of GPS receiver (TOPCON NET-G3A) and dual-frequency choke-ring antenna (TOPCON CR-G3). The heights of the two antennas was about 1.57m. The data were collected at 1s recording rate and the cut-off elevation angle was set as 10 degree.

The multipath effect indexes (MP1 and MP2) of the two stations obtained by using the software TEQC are shown in Tab.1. It can be seen from Tab.1 that MP1 and MP2 of station MP01 are more than 0.3m while those of station REFF are not more than 0.03m, which means that the multipath effect around station MP01 is much greater than station REFF. The kinematic module TRACK of GAMIT/GLOBK was used to calculate the baseline length between station MP01 and REFF. To find out whether different satellite-receiver geometry will influence the effectiveness of the method in this paper, we take REFF station as a reference station to process observation data of two different time periods. The satellite sky trajectories of two different time periods of station MP01 are shown in Fig 2.

Table 1. Multipath effect indexes of station MP01 and REFF (unit:m).

| station | 361 | | 362 | | 363 | | 364 | |
|---------|------|------|------|------|------|------|------|------|
| | MP1 | MP2 | MP1 | MP2 | MP1 | MP2 | MP1 | MP2 |
| MP01 | 0.33 | 0.32 | 0.35 | 0.33 | 0.35 | 0.41 | 0.33 | 0.34 |
| REFF | 0.01 | 0.02 | 0.02 | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 |

3.2 Data processing

The precise ephemeris was introduced to calculate the coordinate series of station MP01. Since the errors

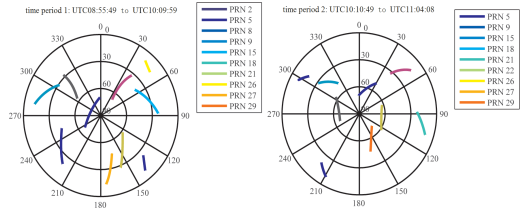


Figure 2. Satellite sky trajectories of two different time periods of station MP01.

from satellites, receiver, ionosphere and troposphere can be well eliminated in the short baseline situation, we use separately L1 and L2 frequency solution strategy to fix the ambiguity. The results showed that more than 95% of the ambiguity had been fixed, which means the dominant errors in the solutions are multipath errors and random noises. Usually, the random noises are much smaller than multipath effects. But according to the observation environments in Fig.1, we think that the errors of station MP01 were primarily caused by the nearby wall.

We used the cross-correlation function to estimate the shift seconds of repetition periods within the time span from UTC08:30:00 to UTC11:29:29. The results showed that the “repeat lags” of the 362nd, 363 and 364th day of the year are respectively 250s, 249s and 249s. After the repeat lag and the coordinate mean are eliminated, the coordinate time series of the four days of station MP01 extracted respectively from time period 1 and 2 are shown in Fig.3.

In Fig 3, the red and blue curves represent respectively the time series in north and east directions. Fig.3 shows that the coordinate time series are fluctuant due to the multipath effect and random noises. However, their shape is similar to each other. For time period 1, the fluctuation in north direction is smaller but relatively acute in east direction.

4 EVALUATION OF RANDOM NOISE POLLUTION

Besides the multipath effect, random noises are also important error sources. PCA and KLE are used to evaluate the influence of random errors on positioning in this paper.

If the coordinate time series is influenced mainly by the multipath and random noises, the principal components generated by PCA or KLE transformation on the error matrix should reflect the multipath effect. Thus, the response of coordinate time series of an individual day to the principal components can be treated as the influence of multipath error. If the multipath is the main source of error, the responses to the principal components should be nearly the same for

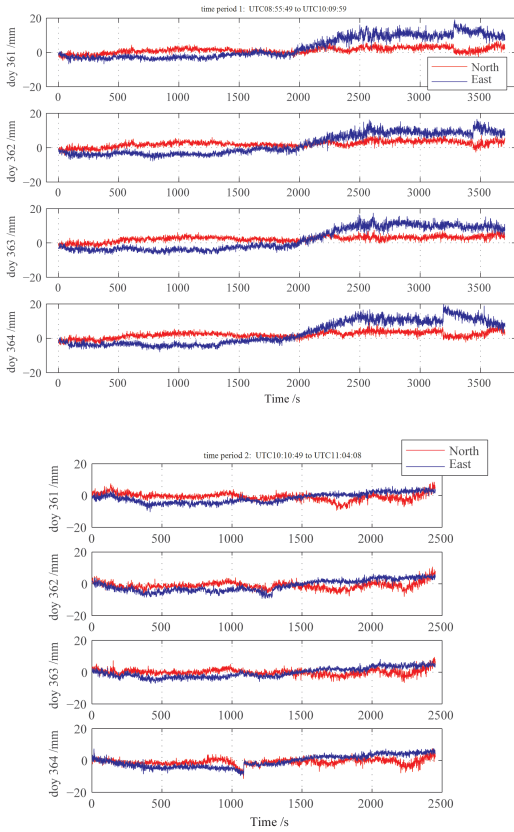


Figure 3. Coordinate time series of station P01 withremoving repeat lag and mean.

multiple days. Contrarily, if the effect of random error is larger on a certain day, the response to the principal components would be inconsistent with other days. It can be seen from equation (6) that the coefficients corresponding to the principal components can be used to evaluate the response. The response coefficients of PCA and KLE in the north and east directions are shown in Tab.2 and Fig.4.

Table 2. Principal component coefficients of time period 1 and 2 in N and E directions from PCA and KLE.

| Model | Direction /time period | 361/mm | 362/mm | 363/mm | 364/mm | |
|-------|------------------------|--------|---------|---------|---------|---------|
| PCA | N | 1 | -0.3697 | -0.5313 | -0.5612 | -0.5192 |
| | | 2 | -0.5940 | -0.6194 | -0.3662 | -0.3598 |
| | E | 1 | -0.5026 | -0.4503 | -0.4960 | -0.5464 |
| | | 2 | -0.4859 | -0.5466 | -0.4627 | -0.5321 |
| KLE | N | 1 | -0.4810 | -0.5068 | -0.5060 | -0.5057 |
| | | 2 | -0.5320 | -0.5664 | -0.4866 | -0.3994 |
| | E | 1 | -0.4995 | -0.4990 | -0.5008 | -0.5007 |
| | | 2 | -0.4995 | -0.5077 | -0.4998 | -0.4979 |

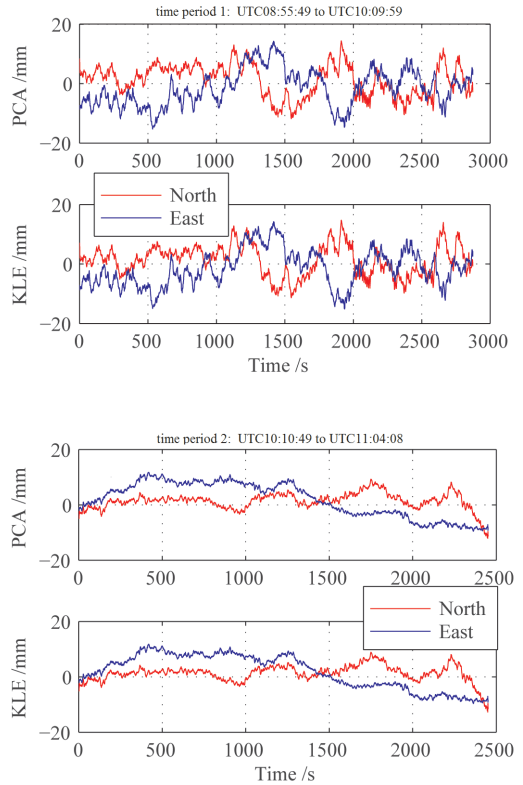


Figure 4. First principal components in N and E directions from PCA and KLE.

As is shown in Tab.2, the differences between the responsive coefficients of first principal components obtained by PCA and KLE are quite small when the coordinate time series are not seriously contaminated by random noises. Compared to PCA, the responsive coefficients obtained by KLE are even more similar to each other. This is because the standardization process of KLE can suppress the abnormal coordinate time series.

5 EXTRACTION AND MITIGATION OF MULTIPATH REEOE

Sidereal filtering includes three steps. The first step is the low-pass filtering to remove the high-frequency noises. In the second step, the filtered results from the first step are superposed according to the sidereal day cycle and the principle of equal weight. The final step is to correct the positioning results using the multipath error model extracted from the second step. Taking the previous data as an example, the low-pass filtering with a sampling window length of 7 epochs was used to remove the high-frequency random noises. For time period 1, only the first two

principal components in the north or east direction are selected. The cumulative contribution rates of PCA transformation are 93.9%, 98.3% respectively in the north and east directions and that of KLE transformation are 93.2%, 98.3% respectively. For time period 2, only the first principal components in the north or east directions are selected. The cumulative contribution rates of PCA transformation are 87.1%, 93.7% respectively in the north and east directions and that of KLE transformation are 86.3%, 93.6% respectively. After the multipath errors were removed, the error distributions of the positioning results are shown in Fig.5.

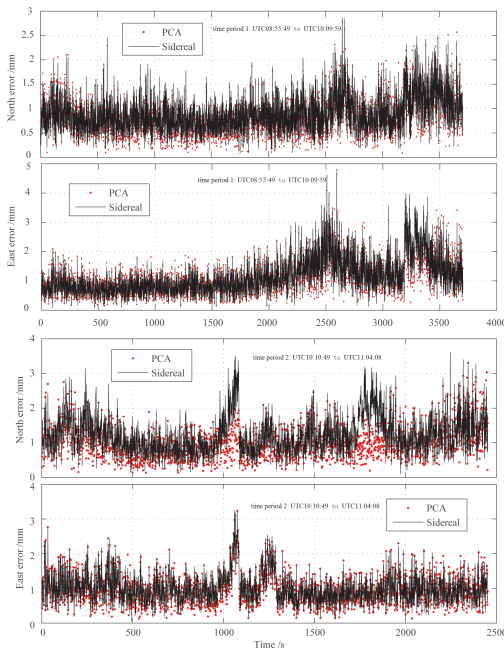


Figure 5. L1 norm scatter of coordinate time series in N and E Directions from PCA and Sidereal filtering.

As is shown in Fig.5, the differences of the residual errors between PCA and sidereal day filtering are not remarkable. But overall, the L1 norm distribution of PCA is slightly better than that of sidereal day filtering. This phenomenon can be explained by the satellite sky trajectories (Figure 3). Due to almost no low altitude satellites which can easily cause multipath effect in east direction, the multipath error in this direction is originally small. Under this situation, the effect of sidereal day filtering is basically the same as that of PCA method. The mean norm scatter and

Table 3. Mean norm scatter and running time of sidereal day filtering, PCA and KLE.o

| Norm | Direction /time period | original | sidereal | PCA | KLE | |
|---------|------------------------|----------|----------|-------|-------|------|
| L1 (mm) | N | 1 | 2.25 | 0.89 | 0.83 | 0.88 |
| | | 2 | 1.70 | 1.23 | 1.03 | 1.05 |
| | E | 1 | 5.76 | 1.15 | 1.11 | 1.15 |
| | | 2 | 1.99 | 1.44 | 1.21 | 1.24 |
| L2 (mm) | N | 1 | 2.44 | 1.05 | 0.98 | 1.04 |
| | | 2 | 3.02 | 0.99 | 0.97 | 0.99 |
| | E | 1 | 5.92 | 1.35 | 1.30 | 1.35 |
| | | 2 | 3.23 | 1.17 | 1.13 | 1.17 |
| time(s) | 1 | 0 | 0.765 | 0.775 | 0.776 | |
| | 2 | 0 | 0.512 | 0.540 | 0.536 | |

running time of the coordinate time series of 3703 epochs obtained from sidereal day filtering, PCA and KLE transformations are shown in Tab.3.

It is obvious from Tab.3 that all three filtering methods can effectively mitigate the multipath error in GPS positioning.

6 CONCLUSIONS

This paper introduces PCA and KLE transformations in high-rate GPS positioning for the evaluation of random noise level and the mitigation of multipath errors in coordinate time series. The results show that the response ratio of principal component coefficients of PCA can effectively reflect the impacts of strong random noises on coordinate time series and multipath errors can be significantly mitigated by PCA transformation. Compared to PCA, the response ratio of principal component coefficients of KLE is less sensitive to random noise pollution and its ability to eliminate multipath errors is also weaker. However, KLE can preferably suppress random noises and local anomalies. Thus, in most practical applications, PCA transformation can be used to evaluate the random noise level and filter the coordinates. If we know that the random noise level of the observation data on a certain day is higher, then KLE transformation can be used for filtering to improve the positioning accuracy.

ACKNOWLEDGMENTS

This research work is supported by Key Laboratory of Precise Engineering and Industry Surveying, National Administration of Surveying, Mapping and Geoinformation (PF2012-21).

REFERENCES

- [1] Choi, K., Bilich, A., Larson, K. M. and Axelrad, P. (2004) Modified sidereal filtering: Implications for high-rate GPS positioning, *Geophysical Research Letters*, vol. 31, no. L2260822.
- [2] Ding, X., Chen, Y., Zhu, J. and Huang, D. (1999) Surface deformation detection using GPS multipath signals, Paper presented at ION GPS-99, Satell. Div. of the inst. of Navig., Nashville, Tenn, pp. 53–62.
- [3] Dong, D., Fang, P., Bock, Y., Webb, F., Prawirodirajo, L., Kedar, S. and Jamason, P. (2006) Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis, *Journal of Geophysical Research - solid earth*, vol. 111, no. B03405B3.
- [4] Elósegui, P., Davis J.L., Jaldehag, R.T.K., Johansson, J.M., Niell, A.E. and Shapiro, I.I. (1995) Geodesy using the Global Positioning System: The effects of signal scattering on estimates of site position, *Journal of Geophysical Research - solid earth*, vol. 100, no. B6, pp. 9921–9934.
- [5] Genrich, J.F. and Y. Bock. (1992) Rapid Resolution of Crustal Motion at Short Ranges with the Global Positioning System, *Journal of Geophysical Research-solid earth*, vol. 97, no. B3, pp. 3261–3269.
- [6] Larson, K.M., Bilich, A. and P. Axelrad. (2007) Improving the precision of high-rate GPS, *Journal of Geophysical Research - solid earth*, vol. 112, no. B05422B5.
- [7] Radovanovic, R.S. (2000) High accuracy deformation monitoring via multipath mitigation by day-to-day correlation analysis, Paper presented at ION GPS 2000, Satell. Div. of the Inst. of Navig., Salt Lake, Utah, pp. 35–44.
- [8] Satirapod, C. and Rizos. C. (2005) Multipath mitigation by wavelet analysis for GPS base station applications, *Survey Review*, vol. 38, no. 295, pp. 2–10.
- [9] Xia, L. (2001) Theoretical research and numerical results regarding multipath in GPS observation, Ph.D thesis, Wuhan University.

Spatio-temporal processing of geoinformation for decision making in critical situations

K. Janecka

University of West Bohemia, Pilsen, Czech Republic

L. Berardi

Sapienza University of Rome, Rome, Italy

A. Cimbelli

Italian Institute of Statistics, Rome, Italy

ABSTRACT: The effective spatio-temporal processing of geoinformation is crucial in the area of civil protection. The article covers essential aspects including security side related to the deployment of a complex spatial data infrastructure for decision making in critical situations. Nine civil protection and risk management pilots have been deeply analysed in order to find and solve common issues for accessing and processing the available spatial datasets. The key emphasis was placed on temporal aspect of the spatial data, as the time dimension is crucial in civil protection applications. The consideration of time dimension, as the key aspect for the management of particular emergency situations (i.e. earthquakes) has brought to the novel use of the granularity and the update frequency in the set of metadata elements. The final software architecture uses a modular methodology based on a federated approach to data collection, processing and distribution of spatio-temporal data.

1 INTRODUCTION

The processing of spatio-temporal information in case of civil protection and risk management is crucial in order to obtain an efficient plan of the resources before the event, and also in case of quick response in dangerous situations. Spatial Data Infrastructure (SDI) is increasingly considered a critical aspect of decision-making in disaster management (van Oosterom et al. 2006, Scholten et al. 2008). First, to build a robust and usable solution it is necessary to collect and analyse the user (pilots related) requirements. In order to implement value added spatio-temporal services based on real used needs and practices the project had a relevant number of user partners, such as Agencies and Public Administrations, engaged in urban planning, environmental management, risk management, which needed processing of spatio-temporal geoinformation to support decision making in critical situations. The objective was to clearly identify the services and functionalities that best comply with the needs of the users and their perceived requirements with regards to their daily activities in the field of civil protection and risk management. The following nine pilots have been considered in the BRISEIDE project:

Emergency Management. The purpose of the pilot was to coordinate responses to a wide variety

of emergencies and urgencies occurring in Navarra, Spain. This coordination extends along the whole life of the emergency from the call taking to the closing of the incident, and it includes incident classification, resource assignment and dispatching, and resource management. Within this context a resource is meant as an emergency vehicle being driven by an operator on site. The dispatching is done according to predefined coverage areas and actual availability of the resources.

Earthquake Risk. The pilot dealt with the evaluation of geological effects caused by earthquakes, on transport networks and critical facilities. The specific goals of the pilot are the identification of risk areas, i.e. most prone to the occurrence of coseismic geological effects (particular attention will be paid to surface faulting and slope instabilities as rock falls) and to point out the vulnerable strategic elements located within the risk areas, including transport networks (roads and railways), lifelines (gas and water pipelines) and critical facilities (high-risk industrial plants). The Pilot addresses both the emergency conditions and the “silent” periods. Immediately after an earthquake, the Pilot shows the vulnerable elements located in the “risk areas” within an affected area whose size and shape are a function of earthquake focal mechanism and magnitude (i.e. max ground

acceleration). During “silent” periods, the Pilot shows the distribution of vulnerable elements located in the risk areas at NUTS 3 (Nomenclature of Territorial Units for Statistics) level, which corresponds, in Italy, to province level.

Landslide Risk, dealt with landslide risk assessment. There were two specific areas of intervention: a) short-term intervention aimed at reducing landslide impact on social and economic systems, mainly addressed at safeguarding people’s lives and security, that are typically managed by Civil Protection at central and local levels; b) medium-long term analysis that includes activities such as data collection, data elaboration and analysis, from slope to small-scale monitoring systems, land use planning and management, that are mainly issued by research and policy makers at various levels. Specifically, the objectives of the pilot have been the classification of the landslides in two main categories and the implementation of simplified landslide risk mapping procedure by overlapping landslide areas with available data on potential exposed elements.

Monitoring of hydrogeological disturbances. The pilot was focused on a concept of interoperable management of multiple sensors within a landslide-monitoring scenario. In this context the target is of this scenario is to provide: a) IT tools based on geoinformation that can manage the information needed to monitoring, forecast and assessment of hydrogeological disturbances, taking into account the dynamic information coming from monitoring sensor networks and the meteorological condition; b) a system that, on the basis of previous analysis, identifies the risk level and the evolution of the disturbance to assess if the deformation patterns are within a normal trend or if they represent a risk for local population.

Flood Management pilot. Aimed at enhancing the capabilities of the Liberec Flood Portal developed in the framework of the “Flood Plan of Liberec Region” (Czech Republic), a set of policies for managing floods events issued in 2004 by the Liberec Region Council. The specific objectives of the pilot were to: simplify the communication among the stakeholders, simplify maps interpretation, improve and extend data and information retrieval, implement functionalities for generating data and information in relation to time as the flood occurs.

Sustainable Forest Management pilot. This pilot aimed at formalizing expert knowledge about sustainable forestry principles in terms of territorial development. The Pilot provides information on forests cover state and changes, and about the health state and changes in order to support forest policy. In particular, the goals are: a) to provide signal information on forest cover state and changes based on satellite and aircraft images; b) to monitor changes in forest health state.

Environmental quality. The purpose of the pilot was to evaluate urban growth and its impact on environmental quality in order to support a more effective land use and spatial planning. Through a set of indicators, the scenario intends to better understand to what extent spatial planning has/is relevant in promoting quality life and welfare in urban areas. There are two specific system products: the relation between the spatial planning and the urban growth on a temporal scale and the impact of urban growth on environmental quality.

Fire risk and emergency management. The purpose of the pilot was to develop and test a web application that enables tracking, managing and accessing aerial photos and land cover data in order to detect land cover changes, as well as to classify fire risk areas to support emergency management in the context of major fires. The web applications take into account also dynamic information such as meteorological conditions.

Forest fire management. This pilot aimed at supporting decision making among civil protection operators, public administration, and forest fires experts during the four life cycle phases of forest fires management (prior to fire, during the fire, after the fire, ex-post analysis). The specific goals of the spatio-temporal processing tools of the pilot are: a) to identify areas of high risk for fire ignition taking into account meteorological daily conditions; b) to model the forest fire propagation; c) to assess accurately the burnt areas; d) to provide access to statistic information to the fire history of the area of interest.

2 USE CASES

Comparing the pilots’ user requirements, collected and described in detail in de Amicis et al. (2010), six common Use Cases were defined – System Access, Management of Information, Information Visualization, Data Creation, Processing of information, and Generation of Alerts. Furthermore, many Pilot-driven Use Cases and Requirements were collected. Use Cases have been fully explained and represented by means of UML (Unified Modelling Language) models and flowcharts. In the UML Use Case (UC) diagrams the actors are considered as the roles played by the users interacting with the subject. A UC could be articulated in a number of sub Use Cases further specialising the description. A solid line between two actions represents a “uses” association (the perform of the first action always requires what specified by the second action), while a line ending with an arrow represents an “extend” relationship (assuming the arrow is drawn from Use Case B to A, this means that B is a special case of the more general A case).

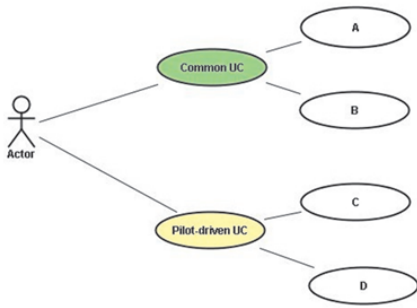


Figure 1. Example of general (green) as well as Pilot-driven (yellow) Use Cases (de Amicis 2010).

Each use case has been expressed also with a flowchart in order to give exactly a graphic representation of the sequence of actions needed for the solution of a given problem. The different boxes are related to actions (rectangles), data repositories (cylinders), choices (hexagons or diamonds). Input and output dataset are represented with a rounded or smoothed rectangles.

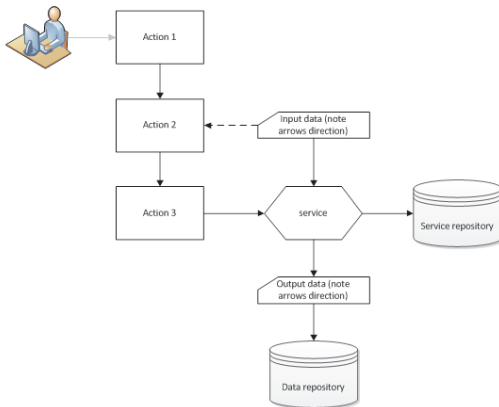


Figure 2. Example of flowchart of use case (de Amicis 2010).

3 FROM USE CASES TO SERVICE REQUIREMENTS

User requirements have been analysed in order to map each Use Case to Service Requirements. Services were classified according to standard taxonomy, inspired by ISO 19119 (ISO 2005), and are mainly focused on spatial and temporal processing

of geographic information related to civil protection issues. At service level, requirements are structured as follows:

General: requirements related to general conditions within BRISEIDE context (e.g. relationship and/or connections to other services)

Functional: requirements related to functionalities available server-side (e.g. coordinate conversion)

Non-functional:

Performance: requirements related to response time for a given operation and/or high availability of the system

Interface: requirements related to service standards and/or technical specs to be implemented (e.g. OGC CSW 2.0.2 AP ISO)

Operational: requirements related to functionalities available client-side (e.g. zoom in/out)

Spatial-temporal: requirements related to spatial and temporal queries to be implemented (e.g. intersection).

3.1 Catalogue services

Spatial and temporal searches shall be implemented as the priority task: either spatial or temporal filters are needed to query metadata repository and retrieve data and services. Indexed and searchable metadata provide a disciplined vocabulary against which intelligent geospatial search can be performed. Therefore a catalogue service allowing user to perform search and retrieval of geographic data and services has been realised. Catalogue services provide the functionality both to manage and to search catalogues; they are a special case of geospatial processing service. They can be considered the heart of any SDI. The BRISEIDE Catalogue Services relies on the CSW (Catalogue Service for the Web) ISO Application Profile as defined in INSPIRE (Infrastructure for Spatial Information in the European Community) Discovery Service, so extending additional aspects not covered by the CSW ISO Application Profile. The main purposes of this service are according to de Amicis et al. (2010b):

- to collect all resource metadata (dataset and service) related to BRISEIDE topics,

- to allow a Client to perform spatio-temporal queries,

- to allow a Client to invoke data access services on the basis of temporal parameters.

3.2 Ingestion services

The proposed BRISEIDE infrastructure is able to ingest data in near real-time as they are produced or made available by the external stakeholders. Taking this into account means that the infrastructure shall comply with a number of requirements that ensure

that it will be ready to react to external events, like file arrival on a specific FTP (File Transfer Protocol) site. The infrastructure should support a sequence of operations that should range from data integrity verification to format transcoding, ingestion into dissemination services, as well as metadata cataloguing. For this reason, the Ingestion Services (IS) are an essential component of the infrastructure, since they represent the main entry point for data ingestion. Furthermore, they are responsible for correct and meaningful archival and cataloguing of data received. The IS are responsible for ingesting, pre-processing and storing data received from external sources of information, external stakeholders (users and services), as well as from the other services of the proposed architecture.

Specifically, according to de Amicis et al. (2010b), being an archiving and ingestion component, its responsibilities shall comprehend:

Transparent retrieval of data from external sources regardless of the employed protocol (FTP, HTTP (Hypertext Transfer Protocol), SOAP (Simple Object Access Protocol), OWS (OGC Web Services), etc.).

Transparent Ingestion of the retrieved data, which means:

- verification of data integrity,
- pre-processing for successive data,
- archiving (via interaction with the dissemination services),
- cataloguing (via interaction with the cataloguing services).

Besides these duties, the IS shall ensure that all the tasks it performs during the various ingestion stages will be logged. This is necessary in order to allow users to analyse the evolution of the various processing performed. Moreover the IS shall expose administration and monitoring interfaces in order to allow system administration and monitoring.

3.3 Dissemination services

Dissemination services allow users to view, access and manipulate spatially related data and/or graphical representation of them, also considering temporal dimension. Data source are GIS (Geographic Information System) and Remote sensing data, sensors observations and measurements, and (if needed) non-geographical information. In BRISEIDE, Dissemination services include four types of services: Feature Data Access, Coverage Data Access, Sensor Data Access, and Portrayal. Feature Data Access Services focus on access and filter operations to allow clients to perform search and retrieval of geographic features (vector) on the basis of spatial and temporal parameters. Coverage Data Access Services focus on access and filter operations

to allow clients to perform search and retrieval of grid coverage on the basis of spatial and temporal parameters. Sensor services allow the user to access to sensor data using standard protocol. The service interface are based on OGC's Sensor Observation Services standards, and the information model is based on the SWE (Sensor Web Enablement) and Observation & Measurements standards. Portrayal Services are related to the graphical representation of geographical information. Graphical representation means the display of geographical information through the use of maps (choropleth, dots) and/or diagrams (pie, bar).

3.4 Processing services

The BRISEIDE geoprocessing services provide generic geoprocessing functionality on geographically referenced vector and raster datasets. It is possible to remotely trigger them by using the standardized Web Processing Service Interface specification. This allows providing advanced GIS functionality, supporting also time as a dimension, within spatial data infrastructures. The BRISEIDE topological services perform operations in order to obtain computed features from the input features based on topological relationship between them. A topology is a set of rules and behaviours that model how points, lines and polygons share their geometries. The Raster algebra processing services perform a subset of algebraic and logical operation on the raster layer. The operators shall be applied to a single raster layer both at multiple layers. The statistical service shall perform a series of basic statistic operator on feature attribute and raster (min, max, mean, median, standard deviation). The process shall calculate the statistic index based on field selection, time series of retrieved features. The feature extraction process service shall perform algorithms in order to extract some features from raster or satellite images. They can use supervised or unsupervised methods using the training data set as input. The process shall give the layer containing the classified feature class as input. The change detection process service shall compare two layers with different temporal extension in order to find some changes occurred during the time. The service requires two different layers containing the same information referred to different epoch as input. The results shall be a layer with the changed elements. The statistic time-driven process shall perform some operations such as interpolation, geostatistic, spatial analysis etc. The services shall receive one or more layers and parameters based on the specific analysis as input. The transformation process shall perform the transformation of geographic data either in terms of Coordinate Reference System (CRS) and format.

3.5 Location services

BRISEIDE Location Services are used to perform geocoding and routing operations and to elaborate spatial-temporal request parameters in order to satisfy general and pilot-specific requirements. The service interface is based on OGC's Open Location Service (OLS) standards and the information model. Geocoding Services are used to perform travel directions and routing operations and to elaborate spatial-temporal request parameters in order to satisfy general and pilot-specific requirements. Routing Services are used to perform travel directions and routing operations, also taking into consideration spatial-temporal request parameters in order to satisfy general and pilot-specific requirements. Furthermore, BRISEIDE operates with Security Level Services and Pilot-Specific Services. The detailed information about BRISEIDE services can be found in Cipriano et al. (2010).

4 METADATA REQUIREMENTS

The purpose of the analysis in terms of temporal metadata requirements, derived from the user requirements, was to consider which temporal elements are important in order to characterize the datasets used in BRISEIDE pilots. BRISEIDE was not oriented to a specific theme contained in the INSPIRE annex, but the used pilot datasets belong to several different themes. For these reasons, the BRISEIDE metadata addressed to satisfy some general aspects, common to all the datasets, and extend the profile adding the required spatio-temporal information. The analysis underlined the need to include some common metadata information regarding the temporal dimension.

ISO standard foresees 4 mandatory elements for the management of the temporal information in the ISO 19115 core set (ISO 2005). The first is the detection date, meant as the date when the dataset has been generated. Other two elements refer to the date of last revision and the date of publication of the dataset. The time extent represents the period covered by the resources. It defines, in particular, the time period covered by the content of the resource. This time period could be expressed as an individual date, an interval of dates, expressed through the starting date and end date of the interval, and a mix of individual dates and intervals of dates.

In general INSPIRE Implementing Rules (IR) do not completely satisfy the need of a complete temporal description necessary for the development of web services for the management of emergencies. This is particularly true for two reasons. First, INSPIRE focuses on the resource, considered as the geometric

element, and even the time properties are related to it. It does not consider an event-oriented approach, but a resource-oriented one. The difference is crucial because in BRISEIDE the interest was not the initial date of the dataset, but the date when an event has started. Secondly, the IR elements do not guarantee to filter out the resources by the needed time coverage in applications areas as environmental risk management or civil protection. BRISEIDE proposed a solution that adds to the temporal information already foreseen in IR other necessary parameters. As stated in Dekkers and Craglia (2008), there is a need of specific temporal information in INSPIRE.

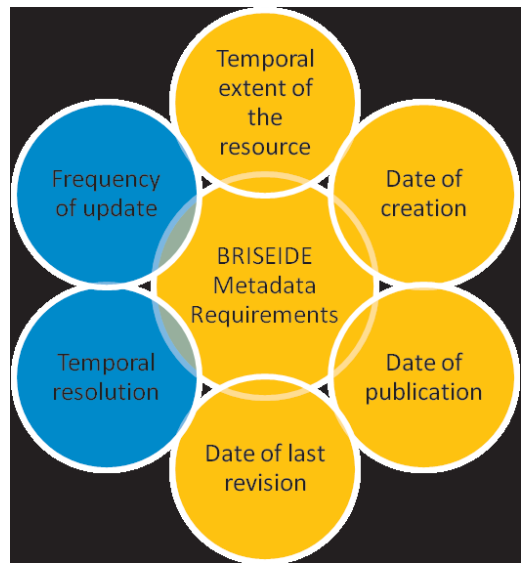


Figure 3. Metadata elements in BRISEIDE (Cimbelli et al. 2010).

The problem of lack of temporal information in INSPIRE metadata is relevant, especially when the phenomena under observation vary during time. As an example the measurements of an environmental variable from different ground stations could be used. During the whole period of measurements of ten years (extent) the stations record daily (temporal resolution) for 1 hour (temporal extent of an event). Currently the ISO standard doesn't allow to describe such a situation. On the other hand, the need of a better expression of temporal attributes related to an event clashes with the system implementation of completely new metadata elements. In addition, when it is not necessary to analyze a very fast evolving event (i.e. earthquake) the new elements are not so important.

In conclusion, BRISEIDE has considered two other temporal attributes important: the temporal

resolution and the frequency of the update. The first attribute, related to the temporal granularity of the dataset and already foreseen in INSPIRE for grid dataset, has been considered in BRISEIDE even for vector ones. This could be a date with day resolution in case of satellite images or year/months resolution in case of an orthophoto or digital map. Frequency of the update can be viewed such as the time period after which a new data is available. The element is not included in the minimum set of mandatory elements and it could be expressed with the element `maintenanceAndUpdate` Frequency foreseen in the class `MD_Identification`. It is possible to create a resource that register its position with a resolution of seconds but the frequency on which transmits the data is hourly. In case of risk analysis or when creating a risk maps it is not only important to know the date when the map or analysis is performed but also the time accuracy of the dataset used in the process. In this way, it is possible to create a risk map generated in a particular date although the set of facilities used to generate the map is older. Detailed information about BRISEIDE metadata requirements can be found in Cimbelli et al. (2010).

5 SURVEY AND ADAPTATION OF AVAILABLE GEOINFORMATION AND SERVICES

The diversity of available pilot datasets, mainly regarding structure, format, geometry, and temporary resolution (their update), requires facing the issue of their harmonisation in order to maximise the possibility to be re-used by several actors.

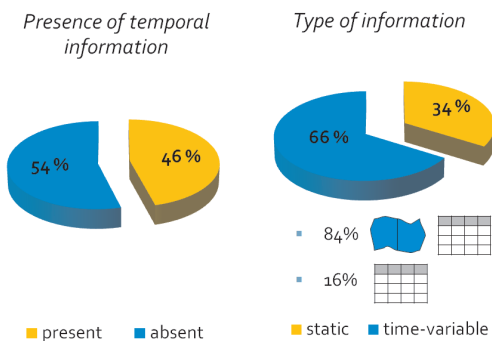


Figure 4. Time information in BRISEIDE datasets (Prandi et al. 2010).

Regarding the adaptation of existing data several aspects for achieving the harmonized data sets have been considered, such as data

format, storage, transformation, data re-projection, specifically underlining the adaptation needed for including temporal extension as the principal aspect considered within the project. The harmonization into a common data model has been already addressed by quite a number of projects, e.g. Plan4all or Humboldt, for which the final objective was to be able to provide INSPIRE compliant datasets and metadata. They intend to harmonize information from different themes, according to INSPIRE specifications, also considering the compliance with the formats for data (GML 3.2.1) and metadata (XML). In BRISEIDE, some of them have been reviewed in order to avoid effort duplications in providing methods or guides for adaptation of the data model, and this information has been used in next steps of the project. For evaluating the adaptation of existing services, it has been initially focused on the OWS defining the aspects required for adapting WMS (Web Map Service), WFS (Web Feature Service), WCS (Web Coverage Service), and WPS (Web Processing Service). Special attention has been devoted to the Sensor-related services as they are essential to ensure adequate monitoring of the environment, but require a more in deep analysis and detailed explanation of the related standards and the adaptation required in BRISEIDE. The experience obtained from previous co-financed projects regarding the services is that a minimum adaptability for publishing the information has been required; however, in some cases, additional services needed to be created. From the technological perspective, some issues raised in services preparation, especially those related to the adaptability to the specific versions required by INSPIRE, are not yet considered by some commercial or open source software. Additionally, the adaptation to current infrastructures from the security-related perspectives has been reviewed, finally offering an integrated view of the Web Services whose orchestration relies on BRISEIDE services to provide access to information in an SDI, as a simple exchange of raster data or either complex processes that perform a series of operations. The services are, usually, distributed over a network of peers, so that each provider could have its own services maintained on its servers. The detail information about analysis of available vector, raster, alphanumeric, and services data sources are stated in Prandi et al. (2010) and Cimbelli et al. (2010b).

6 SOFTWARE ARCHITECTURE

The design has followed a modular methodology based on a federated approach to data collection, processing, and distribution of spatio-temporal data. This allows creating a logical layer between existing services, exposing data and functionalities via SOAP

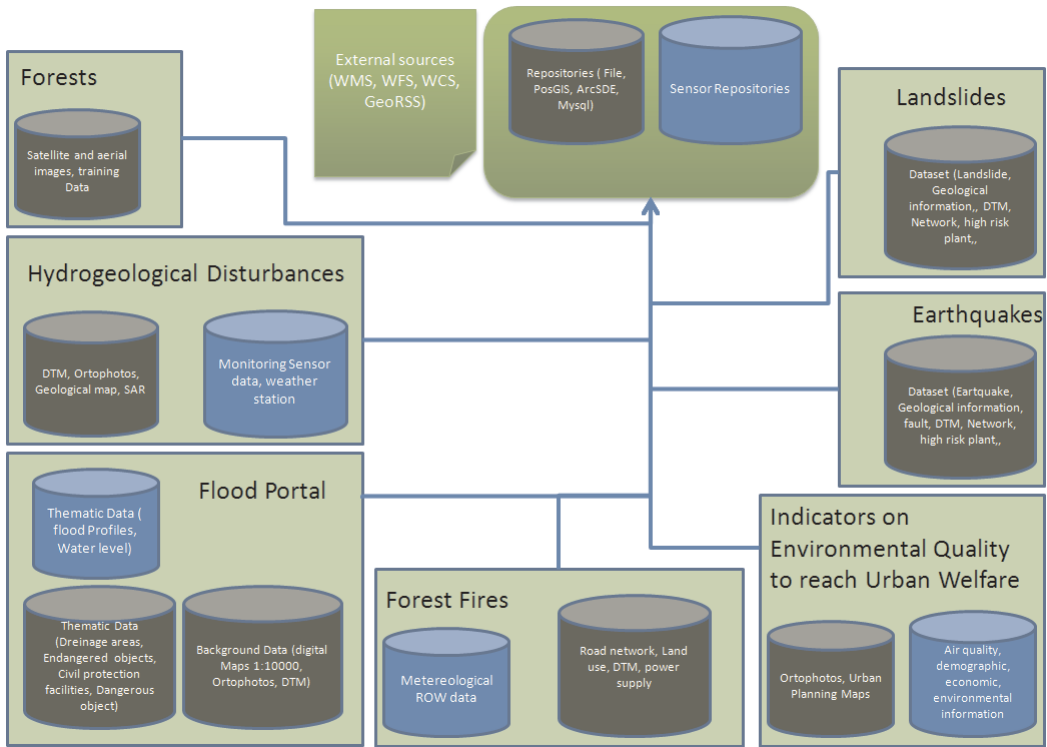


Figure 5. Overview of the data layer (Prandi et al. 2010).

or through OGC communication protocols such as WMS, WFS, WPS, etc. The framework followed a multi-level structure. At the lowest level there are the data repositories available through OWS. The middleware level then builds on top of these repositories a number of geo-processing services deployed as WPS. The different software components can be eventually packaged within a software toolbox that allow data and service providers to create plug-and-play software services capable to expose spatio-temporal extending BRISEIDE's services.

6.1 Application level – BRISEIDE 3D client

The developed BRISEIDE 3D client is a desktop application based on uDig and Eclipse RCP (Rich Client Platform) that allows accessing to the services and perform operation on dataset as well as the 3D globe visualization. The extreme flexibility provided by the Eclipse plug-in approach has led to the definition of the BRISEIDE platform in terms of a set of logical components, all of them wrapped inside one or more plug-ins for the uDig framework and exposed through an xml file following the OSGi paradigm. OSGi provides a mechanism for defining and running

separate components as well as a services mechanism to support inter-bundle collaboration.

The user starts the application and logs into the system by using the authentication module which, in turns, queries the underlying authentication service to check if the user is registered on a remote database. This verification is done with the help of a persistency manager module that abstracts the requests from the underlying remote database. Once the user is logged in, he can access the functionalities provided by several modules each one with a correspondent graphical menu. The main menu and TOC module are both based on the layer manager module that handles the set of available layers and it provides them to the renderer module in order to graphically display their related information. In addition to that, the main module accesses the data filter and deals with the upload/download manager, in charge of managing the file transfer from and to the web processing and map services, passing the data through the orchestration and the persistence manager respectively. Similarly, the search module leans on the catalogue component in order to have a list of all the available services, while the RSS feed module leans on a RSS feed service. Each of these

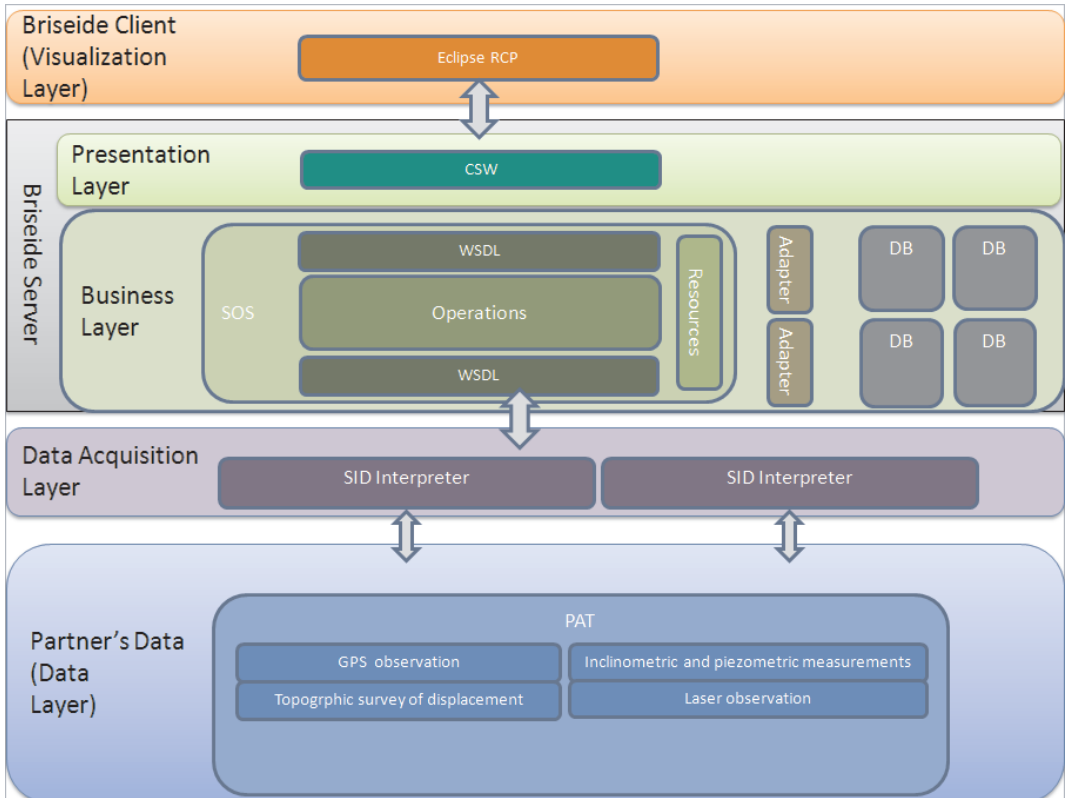


Figure 6. Example of BRIDE architecture (de Amicis et al. 2010b).

components, with particular focus on the front-end modules, is dynamically handled by the uDig RCP (Eclipse Rich Client) through on-demand loading of features as long as the user requests them (de Amicis et al. 2010b). The main extensions provided by BRIDE 3D client to existing client functionalities are as follows (Maroni et al. 2011):

- 3D globe visualization.
- Accessing WMS supporting time dimension.
- Accessing WFS.
- Accessing WCS including time dimension.
- Accessing SOS (Sensor Observation Service).
- Open Location Services client.
- WPS client and workflow manager.
- Geo Tiff manager.
- GML to shapefile converter.
- Resource manager (FTP client).

Each of these extensions is deployed as Java .jar files plug-ins for uDig. The installation of the BRIDE 3D client requires two steps - downloading and installation of uDig software and downloading and installation of the plug-ins. The plug-ins and the needed libraries for BRIDE components are

downloadable on BRIDE web site after registration at link . The client requires the installation of the Java virtual machine. The deployment is multi-platform but some specific functionality could be not available into Mac Os/x or Linux due to the use of some specific Windows libraries.

6.2 Middleware level

The middleware part is the central core of BRIDE architecture, this component is devoted to provide services to the client. The component of the middleware is the Mapserver , the batch system, the catalogue, the SOS and WPS server. Further components of the middleware level are the ESB (Enterprise Service Bus), the Web Server and the Servlet. The used components are in detail described in de Amicis et al. (2010b).

Security Level - from the security point of view BRIDE architecture relies on 52North's Security system suite. This is a comprehensive solution that delivers both authentication and authorization of users as well as the security of communications between web services. The solution, available as an

open source, is able to work correctly during the authentication and authorization phases because it relies on a proprietary User database, as well as Lightweight Directory Access Protocol (LDAP) Directory and Certificate Store. With the notable exception of the User DB, all the other components are included in every standard compliant Public Key Infrastructure (PKI).

Furthermore, the User DB could be considered as a middleware that requests authorization from the LDAP Directory. Due to this the PKI infrastructure guarantees both a full compatibility with 52North's solution and the undeniable advantage of referencing the digital identities only once in a unique repository, ensuring that the person in charge of management has control over every user's „lifespan“. In the standard approach all OWS are channeled through standard unsecure HTTP.

The approach followed, as illustrated in Figure 7, basically creates a secure channel through two software interfaces acting as proxies, respectively at the client and middleware level. This approach is very beneficial because the communication between client and server remains essentially unchanged. The details are illustrated in Figure 8.

6.3 Web service orchestration framework

The platform selected to create BRISEIDE's orchestration framework is Apache ODE. ODE is a platform, composed of high level APIs, which is capable to execute BPEL described processes in a reliable and performance wise way. It can be integrated via JBI (Java™ Business Integration) or it can be deployed within a servlet container. In the specific case of BRISEIDE the system also uses Axis2 on Apache Tomcat (7.0.2). Axis2 is the SOAP engine that succeeds the former Apache Axis SOAP stack, it features a more flexible and efficient architecture, it provides support for SOAP 1.1, 1.2 and REST and is more XML oriented than its predecessor. As more advanced characteristics, it supports hot deployment of services, asynchronous non-blocking invocation, WSDL 1.1 and basic 2.0 functionalities as well as a module system that allows the introduction of new components, such as WS-Security, WS-Policy, etc. It must be noted that Apache ODE does not provide a native graphical designer, but just a management interface for its APIs (de Amicis et. al. 2010b).

The implementation is modular and extensible by third-party partners, so that it is possible to add new windows, options, constructs, palettes and especially new run-time environments or engines.

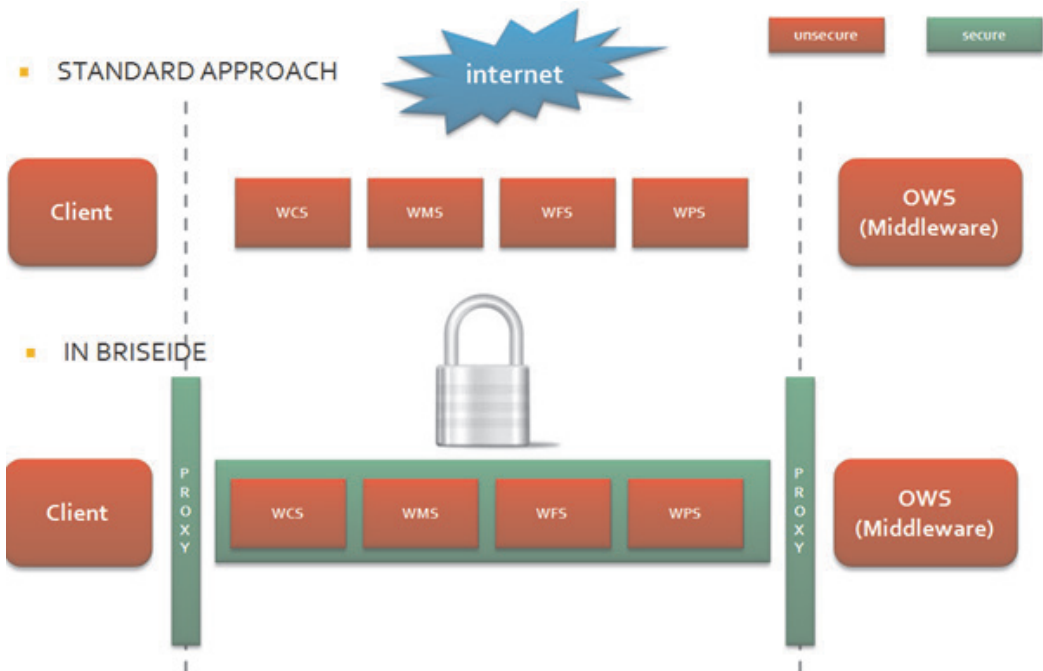


Figure 7. Standard (unsecure) vs BRISEIDE secure approach (de Amicis et al. 2010b).

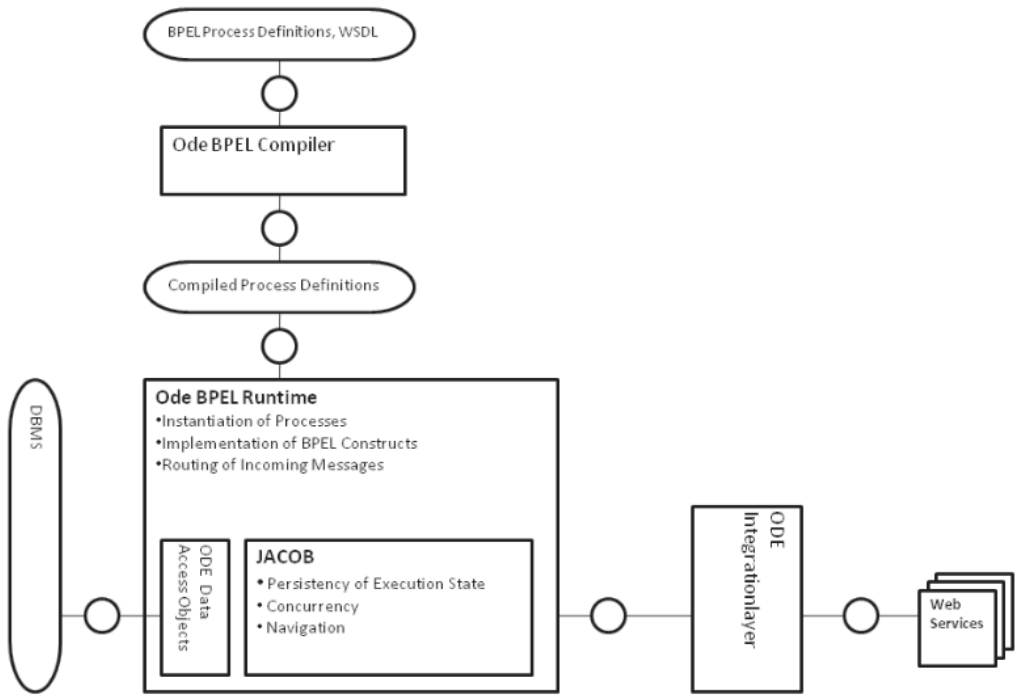


Figure 8. ODE architecture Overview (Apache 2014).

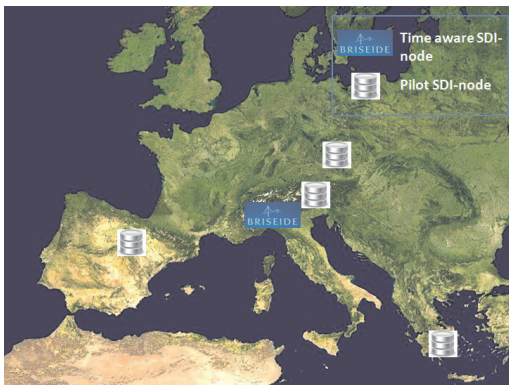


Figure 9. General overview of the geographic location of Time-aware SDI node and Pilot SDI nodes (de Amicis et. al. 2010b).

7 CONCLUSION

BRISEIDE has clearly shown the heterogeneity among the different pilots, regarding not only existing data and services, but also potential requirements for the pilots. The first raised issue has been

the adaptation of existing data. Several actions have been performed for achieving harmonized data sets, like data format and storage, data transformation, and data re-projection. In all of these actions, the adaptation needed for including temporal extension in the harmonised dataset has been specifically considered. For evaluating the adaptation of existing services, the focus was on the OWS, by means of the standardization of the available WMS, WFS, WCS, and WPS. Special attention has been devoted to the sensor-related services, as they are essential to ensure adequate monitoring of the environment. The design of the software architecture of the BRISEIDE toolkit has followed a modular methodology based on a federated approach to data collection, processing and distribution of spatio-temporal data. This allows creating a logical layer between existing services, exposing data and functionalities via SOAP or through OGC communication protocols such as WMS, WFS, WPS, etc. The framework follows a multi-level structure. At the lowest data level data repositories are placed including those from other EU projects, which are already available through OWS. Then, the middleware level builds, on top of these, low-level OWS a number of geo-processing services deployed as WPS.

ACKNOWLEDGMENTS

This work was supported by the European Regional Development Fund (ERDF), project “NTIS – New Technologies for the Information Society”, European Centre of Excellence, CZ.1.05/1.1.00/02.0090 and eContentPlus project “BRISEIDE – Bridging Services, Information and Data in Europe”, CIP-ICT-PSP-2009-3.

REFERENCES

- [1] Apache Software Foundation. 2014. *Apache ODE Developer Guide*. WWW document, <http://ode.apache.org/developerguide/architectural-overview.html>.
- [2] Cimbelli A, Prandi F, Conti G, and De Amicis R. 2010. Metadata Requirements. Trento, Italy, Fondazione GraphiTech - Centre for Advanced Computer Graphics Technologies, BRISEIDE Technical Report D1.1.03.
- [3] Cimbelli A, Prandi F, Conti G, and De Amicis R. 2010b. Survey on existing GI and Services. Trento, Italy, Fondazione GraphiTech - Centre for Advanced Computer Graphics Technologies, BRISEIDE Technical Report D1.1.04.
- [4] Cipriano P, Prandi F, Conti G, De Amicis R, Giannecchini S, and Pezzi S. 2010. Service Requirements. Trento, Italy, Fondazione GraphiTech - Centre for Advanced Computer Graphics Technologies, BRISEIDE Technical Report D1.1.02.
- [5] De Amicis R et al. 2011. Time dimension in e-government GI driven services. In Proceedings of Agile 2011 conference. Salt Lake City, Utah.
- [6] De Amicis R, Conti G, Prandi F, and Di Donato P. 2010. User Requirements and use cases. Trento, Italy, Fondazione GraphiTech - Centre for Advanced Computer Graphics Technologies, BRISEIDE Technical Report D1.1.01.
- [7] De Amicis R, Conti G and Prandi F. 2010b. Software Architecture. Trento, Italy, Fondazione GraphiTech - Centre for Advanced Computer Graphics Technologies, BRISEIDE Technical Report D.1.1.06.
- [8] Dekkers M and Craglia M. 2008. Temporal Metadata for Discovery. WWW document, <http://inspire.jrc.ec.europa.eu/reports/TemporalMetadataforDiscovery>.
- [9] International Organization for Standardization. 2005. ISO 19119 Geographic Information – Services.
- [10] Maroni L, Caldon S, and Prandi F. 2011. Portal and Web-Client Deployment. Trento, Italy, Fondazione GraphiTech - Centre for Advanced Computer Graphics Technologies, BRISEIDE Technical Report D2.2.07.
- [11] Prandi F, Conti G, Cabello M, and De Amicis R. 2010. Analysis of adaptation of existing data/services. Trento, Italy, Fondazione GraphiTech - Centre for Advanced Computer Graphics Technologies, BRISEIDE Technical Report D1.1.05.
- [12] Scholten H, Fruijter S, Dilo A, and van Borkulo E. 2008. Spatial Data Infrastructure for Emergency Response in Netherlands. Remote Sensing and GIS Technologies for Monitoring and Prediction of Disasters Environmental Science and Engineering: 177–197.
- [13] van Oosterom P, Zlatanova S, Fendel E. 2006. Geoinformation for Disaster Management. Springer Science & Business Media.

Spatial distribution of ethnic villages in Qiandongnan mountainous region

D. Wang & L. Guo

College of Life and Environmental Sciences, Minzu University of China, Beijing, China

B.S. Wu

Humanities and Social Study Education Academic Group, National Institute of Education, Singapore

L. Lv

Geographic Information Center of Zhejiang Province, Hangzhou, Zhejiang, China

ABSTRACT: This paper analyzed the spatial distribution of ethnic villages of two groups, Miao and Dong, in the Qiandongnan mountainous region based on field survey data, DEM, and vector map. Five variables: elevation, slope, aspect, distance between villages and rivers, and distance between villages and roads were considered as key factors affecting spatial distribution of ethnic villages. Results showed that Miao group villages mainly settled on the southern aspect with elevation between 273m and 1256m, and the range of slope was 1.14-36.11°. The distribution of Dong group villages was on the eastern aspect with elevation between 244m and 915m and slope between 1.64° and 30.44°. This research concluded that the terrain had a greater influence on the distribution of the Miao group villages than Dong group villages. A large part of ethnic inhabitants were far away from water sources and the motorway.

1 INTRODUCTION

Human settlement is formed based on certain environment conditions, such as elevation, topology and natural resources (Ma et al., 2012, Yan et al., 2011). Nowadays, transportation becomes more and more important for development (Zhu et al., 2009, Tan et al., 2010). It is also a key factor affecting settlements distribution. Studies have been done to explore relationships between natural features and settlement in rural areas through spatial analysis (Jiang et al., 2007, Liu et al., 2011). Cheewinsiriwat (2013) has adopted similar approaches to analyze settlement patterns of ethnic groups in Thailand. However, little attention has been paid in settlement in mountainous regions and minorities settlements. In China, ethnic minorities often inhabit in mountainous areas, especially those ecologically vulnerable belts with backward economy and an isolated cultural system (Chen et al., 2005). Recently the Chinese government has proposed several projects to develop western mountainous regions and stimulate economic growth. In order to make appropriate development plans, spatial distribution analysis and adjustment is needed to decide the locations of new investment and the clustering services and facilities (Misra, 1996, Grossman, 1983).

Qiandongnan prefecture is clustered by 33 minorities. The Miao people and Dong people are the most

populous ethnic groups. Most of the minorities are living in mountain with low living quality and less production material. In order to enhance the economy in ethnic areas and improve the minorities' living quality, adjustment and optimization of ethnic residential area should be done. Spatial analysis of the settlement can help decision makers figure out the key factors affect the distribution of ethnic villages.

Taking Miao group villages and Dong group villages as examples, this paper quantitatively analyzes the spatial distribution characterizations of ethnic settlements in the area from the perspectives of elevation, slope, aspect, roads and rivers.

2 METHODOLOGY

2.1 Study area

Qiandongnan Prefecture is located in southeast of Guizhou province, China (Fig.1). This region has a northern subtropical monsoon climate. The altitude over there varies between 137 m and 2179 m above sea level with an approximately average slope of 22°. Most of this region is over 1000 m above sea level. This prefecture contains 16 counties. The Miao and the Dong groups are the two dominant

groups(42.22% and 29.75% of the population) in this region, respectively(Wu, 2012).

There are 70 ethnic villages, including 36 Miao group villages and 34 Dong group villages, selected based on the statistical year book of Qiandongnan Prefecture, published in 2012. Those villages are classic and traditional ethnic villages. The Miao group villages mainly are located in southwest area and the Dong group villages are mainly distributed in southeast area (Fig.1).

2.2 Data and methods

The location of each ethnic village is measured by GPS devices. Spatial data such as elevation, slope, aspect and river network are extracted from DEM (1:250,000) based on the software ArcGIS 10. Road data are obtained from national transportation vector map (1:5,000,000). Five variables: Elevation, slope, aspect, distance between villages and rivers, and distance between villages and roads, are chosen as the topology, resource and transportation restrictions.

According to the topology variance in Qiandongnan, the elevation is divided into 6 grades from an altitude of 200 m, taking 200 m as spacing. The slope is divided into 4 grades: $0 - 7^\circ$, $7 - 15^\circ$, $15 - 25^\circ$, $>25^\circ$. The aspect is divided into 4 grades. According to the convenience to water source, the distance between villages and water is divided into 6 grades: $<100m$, $100-200m$, $200-400m$, $400-600m$, $600-800m$ and $>800m$. The distance between villages and roads is also divided into 6 grades: $< 1 \text{ Km}$, $1 - 3 \text{ Km}$, $3 - 5 \text{ Km}$, $5 - 7 \text{ Km}$, $7 - 10 \text{ Km}$, and $10 - 15 \text{ Km}$.

The spatial analysis of point pattern is done based on Getis-ord General G to analyze the distribution pattern of ethnic villages on different variables(Yan et al., 2009).

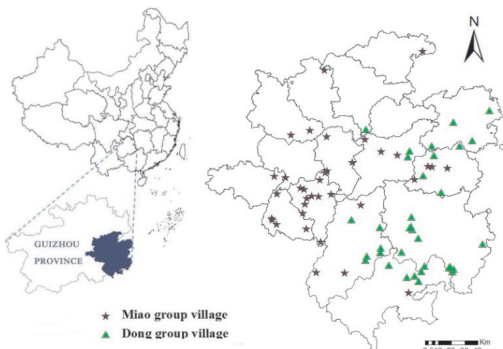


Figure 1. Distribution of Miao group villages and Dong group villages in Qiandongnan, Guizhou province, China.

3 RESULTS

3.1 Impacts of elevation on the distribution of ethnic villages

The Miao group villages are mainly settled within the altitude range from 800 m to 999 m (Fig. 2). There are five villages located in the area above an altitude of more than 1000m. The average altitude is 826m. However, the Dong group villages are mainly distributed within the altitude from 600–799m (Fig.3), and no village is found in the area over 1000m. The average altitude is 630m. In addition, the average altitude for the Miao in each elevation gradient is higher

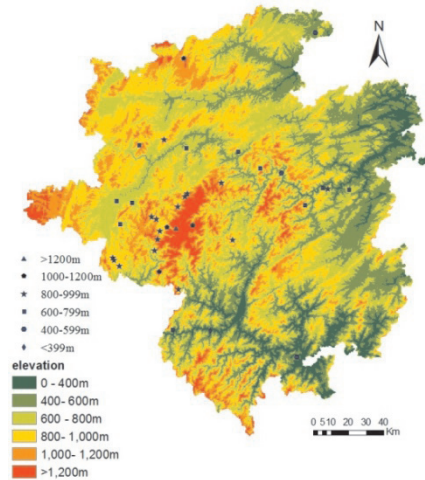


Figure 2. Distribution of Miao group villages on elevation gradients.

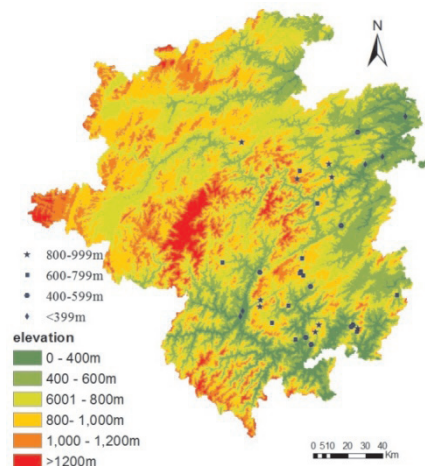


Figure 3. Distribution of Dong group villages on elevation gradients.

than the Dong. That means the Miao settled in higher elevation regions than the Dong.

3.2 Impacts of slope on the distribution of ethnic villages

The Miao group villages mainly distribute on the slope from 7–25 degrees (Table 1), and the Dong group villages mainly distribute on the slope from 7–15 degrees. There are 2 Dong group villages distributed over 25degree on the slope, and 4 Miao group villages over there. The reason for this distribution phenomenon is that the Miao group villages are situated in the mountainous region, and the Dong group villages are in the hilly area. Less than 25% of ethnic villages have settled in a flat area with the slope less than 7°.

Table 1. The distribution of Miao group villages and Dong group villages on every slope gradient.

| Slope(°) | | 0–7 | 7–15 | 15–25 | >25 |
|----------|------------------|------|-------|-------|-------|
| Miao | Count | 8 | 12 | 12 | 4 |
| | Percentage (%) | 22.2 | 33.3 | 33.3 | 11.2 |
| | Average slope(°) | 4.69 | 10.08 | 19.19 | 30.42 |
| Dong | Count | 9 | 16 | 7 | 2 |
| | Percentage (%) | 26.5 | 47.1 | 20.6 | 5.9 |
| | Average slope(°) | 3.87 | 11.3 | 17 | 30.26 |

3.3 Impacts of aspect on the distribution of ethnic villages

The Miao group prefers to live in the southern and western aspect and Dong people like to live in the eastern aspect (Table 2). There are only 18% of ethnic villages settled in northern aspect. This distribution pattern is related to the climate. The wind direction in Qiandongnan is mainly northern direction all the year round.

Table 2. The distribution of Miao group villages and Dong group villages on four aspects gradient.

| Aspect | Western | Southern | Eastern | Northern |
|--------|---------|----------|---------|----------|
| Miao | 11 | 13 | 6 | 6 |
| Dong | 8 | 7 | 12 | 7 |

3.4 The distance between ethnic villages and rivers

Water is an essential part in people’s daily life and important material for production. However, a large

part of ethnic villages have difficulty in getting water. The straight-line distance between ethnic villages and water is shown in Table 3. It shows that a large part of ethnic inhabitants are far away from water sources. Two-fifth of the Miao villagers has to walk for more than 400 meters to get water. This situation is more serious for Dong group. Even ten percent of ethnic villagers have to walk more than 800 meters to get water. That means more water facilities are needed to help ethnic groups to get water.

Table 3. The distance between Miao and Dong group villages and rivers.

| Distance to the river (m) | <100 | 100–200 | 200–400 | 400–600 | 600–800 | >800 |
|---------------------------|------|---------|---------|---------|---------|------|
| Miao | 6 | 6 | 9 | 6 | 6 | 3 |
| Dong | 8 | 4 | 5 | 3 | 9 | 5 |

3.5 The distance between ethnic villages and roads

Transportation is extremely important to the development of a region. There are a few villages in

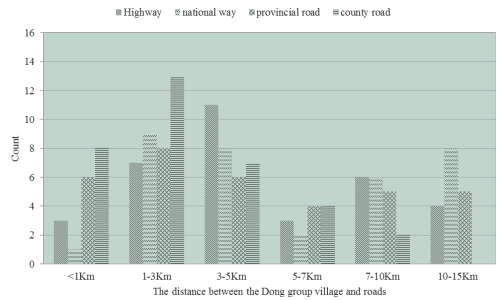


Figure 4. The distance between Miao group villages and roads.

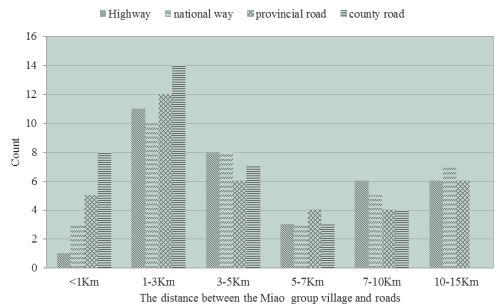


Figure 5. The distance between Dong group villages and roads.

Qiandongnan that can hardly reach a motorway. Most ethnic villages located in the region are 1–3km far from roads (Fig.6 and Fig.7). The nearest road type that ethnic villages can get is county roads. The national way and highway are more difficult to reach than provincial road. There are still 7 Miao group villages and 6Dong group villages located in the area over 5Km far away from county roads.

3.6 *The spatial distribution pattern of ethnic villages*

The results of the spatial pattern analysis show that the distribution of the Miao group villages on elevation and slope is concentrated on high altitude and steep slope, and the distribution of the Dong group villages on aspect is concentrated. That means the terrain has a greater influence on the distribution of the Miao group villages than Dong group villages. The main reason is that the Miao people, mainly distribute in Leigong mountain region with high and steep mountains and Dong people settle in lower and flatter area. The Miao group villages are clustered in the area near to county road and highway. The distribution of Miao group villages on aspect, the distance between Miao villages and national road, provincial road and rivers is at random. The Dong group villages are clustered in the area near to the highway, national way and provincial way. The distribution of Dong group villages on slope, elevation, the distance between Dong villages and county road and rivers are random.

4 CONCLUSION

There are a large number of ethnic villages located in an unsuitable environment. They have difficulties on travel, cultivation and getting water. They are suffering high risks of geological disaster because many ethnic villages located in areas with high altitude and steep slope. Those conditions also limit the development of those villages. As a result, more water and transportation facilities are needed to improve the ethnic people's living condition.

The spatial pattern of those two type, ethnic villages indicates that Miao group villages need to improve the transportation more urgently than Dong group villages. In addition, the Miao group villages undertake a higher risk of geological disaster, such as landslide, because Miao group villages settled in high and steep area.

According to the research, we propose the government pay more attention to this situation and help the minorities solve the problem.

The ethnic villages in this study are chosen based on the yearly statistical books published by the Chinese government. Those ethnic villages are just samples of the Miao and Dong; other minorities do not take into consideration. We are expecting to collect more data and more ethnic villages in the future research.

In the further research, the calculation of distance between villages and rivers and the distance between villages and roads should consider the influence of topology. Straight-line distance is less convincing. In addition, the ethnic villages' area is ignored in this study. Further research should take the area of each village into consideration.

ACKNOWLEDGMENT

The work presented in this paper is supported by the National Natural Science Foundation of China (No. 31370480) and 111 Project (B08044). We also acknowledge all people who help us collected data.

REFERENCES

- [1] CHEEWINSIRIWAT, P. 2013. The use of GIS in exploring settlement patterns of the ethnic groups in Nan, Thailand. *Asian Ethnicity*, 14, 490–504.
- [2] CHEN, Y., CHEN, G. J., LIU, S. Q. & WANG, Q. 2005. Ecological Studies of Human Settlements of Ethnic Groups in Southwestern Sichuan—A Case Study of Maidi Village, Miyi County. *Journal of Mountain Research*, 23, 108–114.
- [3] GROSSMAN, D. 1983. Settlement patterns in Judea and Samaria. *GeoJournal*, 7, 299–312.
- [4] JIANG, G. H., ZHANG, F. R., CHEN, J. W., DUAN, Z. Q. & SU, Z. Y. 2007. Analysis of the driving forces of change of rural residential areas in Beijing mountainous areas based on Logistic regression model. *Transactions of the CSAE*, 23, 81–87.
- [5] LIU, X. Q., BI, R. T. & GAO, Y. 2011. The analysis of spatial distribution and optimization on rural settlement in upland half a mountainous using GIS technology. *Economic Geography*, 31, 822–826.
- [6] MA, L. B., GUO, X. D. & ZHANG, Q. Y. 2012. Spatial pattern of rural settlements in Loess Hilly area—A case study of Tongwei County, Gansu Province. *Journal of Mountain Science*, 30, 408–416.
- [7] MISRA, P. 1996. Modelling Settlement Patterns for Metropolitan Regions: Inputs from Remote Sensing. *Journal of the Indian Society of Remote Sensing*, 24, 287–289.
- [8] TAN, X. L., DUAN, J. N., BAO, C. H. & ZHU, H. M. 2010. A GIS-Based Study on Optimized Spatial Distribution of Rural Settlements in Mayang County. *Research of Soil and Water Conservation*, 17, 177–185.

- [9] WU, M. G. 2012. *The yearbook of Qiandongnan Prefecture*, Kunming, Yunnan Arts Press.
- [10] YAN, L. J., SHI, Y. S., YANG, R. X., LU, P. & FU, Q. 2011. The analysis of the spatial pattern of prehistoric settlement sites based on GIS and spatial data mining. *International Conference on GeoInformatics*. Shanghai, China.
- [11] YAN, Q. W., BIAN, Z. F. & WANG, Z. 2009. A spatial analysis on patterns of settlements distribution in Xuzhou. *Science of Surveying and Mapping*, 34, 160–163.
- [12] ZHU, X. X., WANG, H. M., YUAN, X. J., GUAN, L. M. & LIAO, L. J. 2009. Study on Optimizing Spatial Distribution of Residential Land in Fogang County, Guangdong Province. *China Land Science*, 23, 51–57.

Study of AIS beacon control system based on ARM

Q. Hu, X.J. Li & S.F. Zhang

Dalian Maritime University, Dalian, Liaoning, China

ABSTRACT: Beacon is an important guarantee for the navigation safety, and if its equipments are damaged or lost, it will directly affect the safety of ships at sea. AIS beacon designed in this paper is based on ARM7 and AIS receiver. The system achieves to monitor and display the beacon voltage. It designs the algorithm to predict potential collision risk before the accident, and reports to the AIS center so as to assist finding the incident ship. The test results show that AIS beacon can completely receive all the information from the around ship and realize collision avoidance.

1 INTRODUCTION

The beacon is an important sailing safe symbol to assist ship in the waters and the river domain and other regions and its navigational information is necessary for maritime traffic safety. Now in the vessel traffic safety applications, automatic identification system called AIS is already quite mature and it is becoming more and more effective. The AIS system plays an increasingly important role in the ship navigation safety [1]. AIS beacon technologies, namely the installation of AIS equipment among traditional beacons that the buoys were involved in the AIS system messaging, have important implications for the safe navigation of ships at sea. Currently this technology is maturing. AIS beacon can be controlled and measured on the remote and the staff can perform unified management and beacon collision prediction work by AIS base station system. An AIS beacon in this paper not only achieves universal functionality of existing aids, to navigation, but also predicts whether a ship may hit the beacon and promptly sends dangerous vessel information to the AIS center. Therefore, AIS-based virtual beacon and beacon monitoring technology are becoming an important means of buoy information technology and AIS beacon technology will be a new hot research field.

AIS system is used by vessels at sea primarily for identification and exchange of information with other vessels as an aid in collision avoidance [2]. AIS system can provide identifications, including MMSI (Maritime Mobile Service Identity), call sign, and the name [3]. This system designed the algorithm based on these identifications.

The rest of the paper is organized as follows. Section states the overall design of the AIS beacon controller including the hardware and the software while Section introduces the key technology of the

system. In Section IV the test results are presented. Finally, this paper concludes with Section V.

2 OVERALL SYSTEM DESIGN

The overall structure of the AIS beacon controller is shown in Figure 1.

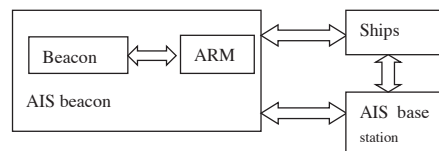


Figure 1. Overall system configuration diagram.

The system can achieve intelligent control by adding an ARM7 chip in the ordinary beacon. It can communicate with AIS shore station and the surrounding ships via messages so as to achieve the purpose of navigation aids. Beacon voltage can be measured on the remote and its parameter information can be put into a data message which AIS beacon will send out. In communications, the system can predict the risk of information and timely feedback to the AIS center to prevent the collision. In turn, the system can achieve real-time beacon control and monitoring based on the received message data.

2.1 Hardware design

AIS beacon controller mainly consists of LPC2378 development platform in the series of ARM7, LCD, keyboard and AIS transceiver component of the hardware device. Its core parts are the AIS transceiver

and ARM7 processor chip. When AIS transceiver receives messages around the beacon at all times, ARM will continuously identify and decode the received message and store the considerable useful data. With external LCD and keyboard, AIS buoy realizes remote control.

2.2 Software design

The software design of the system is divided into four sections (see Figure. 2):

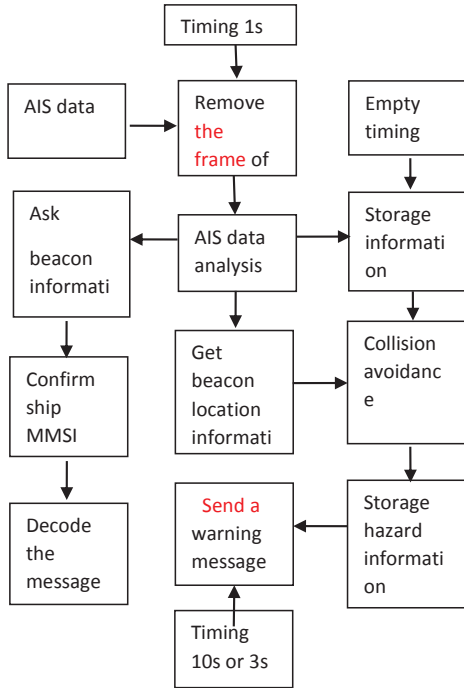


Figure 2. Overall structure of the software.

- Beacon voltage measurement.
- AIS data processing includes data decoding and storage. The system receives the message from the serial port to store them in the queue and decode them in accordance with ITU-RM.1371-4 and IEC61162 protocols. So AIS can receive and process messages in real-time so as to prevent most information from all around the ships.
- Implement of the collision avoidance algorithm by calculating the distance and orientation between the ship and buoy, instantly monitoring ship tracks to divide into four different levels of danger zones. The design uses four queues to storage vessel MMSI numbers. Finally the dangerous ship's MMSI number will be packaged into the 6th

message in an ABM format to send out to achieve collision avoidance and early warning of the ship.

3 SYSTEM KEY TECHNOLOGIES

The underlying idea of overall system design is: first, initializing the underlying system such as various peripherals including timers, serial ports, AD converter, I2C, external interrupt 3 and the queues and lists needed to store data; Second, entering the loop section of the program. When serial ports receive data successfully, the system sets interrupt backstage to accept data, and uses the queue to accept AIS data. After the data is received, the design will process into the decoding links, including data identification and data CRC XOR parity. If the data meet the requirements, parse the message according to specific ITU1371-4rules. Finally a good list of messages is stored in the lists to meet the demand for data in the later algorithms. When receiving the asking from AIS message 6, beacon self-voltage parameter which is tentatively measured in the voltage pin once every 2 seconds with a timer will be packaged into the ABM format. In addition, the system designs to use timers to empty the lists tentatively scheduled for 6 minutes to empty list 1 and 1 minutes to empty list 2.

3.1 Serial software design

Because of the high real-time and high quantity of AIS information, it is necessary to find a method to process and analyze these current data and historic data [4]. Therefore, taking into account the large amount of AIS data, sometimes the system needs to simultaneously receive a large number of information from several boats. So the serial port is designed to be the form of a queue to receive AIS data. Queue uses an array of structures *g_qRcvData*. The size of the array is defined as *queue size*. Firstly, the structure is built, including the array storing data (length is *queue size + 1*), the head pointer of the queue *front* and the end point of the queue *rear*. The system initializes the queue (*front* and *rear* are 0). When there are the new data added in the queue, tail pointer plus one is assigned. When getting the data out, head pointer plus one is assigned. Judging whether the queue is empty, the system can calculate whether the head of the queue *front* is equal to the end of the queue *rear*. If we want to know whether the queue is full, it can be calculated according to the calculation that $(rear + 1) \% queue\ size$ is equal to the front. Timer sets 1s to detect the queue. As long as there is a data, it will be removed. The array stores a full frame of AIS data.

3.2 Ship data storage design

The system designs lists to storage VDM messages previously put in a temporary structure. The latitude and longitude from VDO No. 21 beacon message directly are defined as the global variables $g_BuoyLat$ and $g_BuoyLon$. A data storage portion of the list is a structure, setting up MMSI, vessel name arrays, IMO, latitude and longitude, UTC, the actual course and so on. In order to facilitate the insertion, deletion and query of the list later, the list is designed to lead the list of nodes that the data is not stored in the data field of the head node pointer and pointer field is only used to point to the location of the first node of the list. In this design, the program has established two lists. The first list is used to store information about all of the ship, but the same ship is stored in only one node updating the contents of the ship. It is set to empty the list once every six minutes. The second list is used to record the ship's track which has a lot of dynamic data at different times. It needs to store more content, so it is emptied once every minute.

3.3 Timer design

Timer 0 designs a few second timing of receiving and processing the messages from AIS base stations and AIS ship stations to timely update data in lists.

Timer 1 is used to achieve buoy voltage measurements with AD converter on LPC2378 chip. It indicated a dynamic voltage display and good real-time performance. At the same time timer 1 sets 3s timing to send warning messages of different ships.

Taking into account the distribution and rational use of ARM memory, the system could not indefinitely open up new space for the storage vessel information. So timer 2 sets six minute timing to empty the list 1. Thus facilitate the next set of data storage.

Timer 3 sets 10s timing to send the same ship warning messages.

3.4 Collision avoidance algorithm

The most important section about AIS buoy is to achieve early warning zoning to meet the need of collision avoidance. Warning area is divided into four levels: the first warning area is that there is going to collide between ship and buoy; the second warning area increases appropriate redundancy, considering the length of the ship; the standard of the third warning area is that ships directly traveled at the buoy is within 5 kilometers; in the forth warning area the task is recording the trends of ships sailing to the beacon between 5 km and 20 km. Hence, according to the tracks, the system can get a comprehensive judgment to warn dangerous ships. The first three warnings are very important. After the warning information, there

is the need to respond to each other in a timely manner. The forth is mainly trending judgment. After the warning, the other party could not respond.

Known to the buoy chain length l and maximum depth h , the available anchor chain is generally twice as longer than the maximum depth according to actual experience. It is considered that the buoy will move as the surface water itself dose [5]. Buoy swings circle radius r_0 can be drawn from formula (1). R_0 is set a warning range.

$$r_0 = \sqrt{l^2 - h^2} \quad (1)$$

Cyclotron radius multiplies by 20 (the appropriate redundancy range). Then there obtains radius of the circle r_1 . R_1 is equal to $20 * r_0$ as a secondary warning range; outside this range, the algorithm considers the ship heading to the issue within 5 miles. In this area those cases that a ship travels directly to the beacon are defined as three warnings. The algorithm is that there gets two tangents according to the ship's position and the point of the secondary warning range circle. As long as the actual heading line is located within these two tangents, the case will be set for the three warnings. S in formula (2) is the distance between the ship and buoy. If the actual ship heading angle T_{cog} is between the calculated two corners (see formula (2)), you can judge for three warning.

$$\alpha = \arcsin\left(\frac{r_1}{S}\right)$$

$$\alpha_1 = dVal - \alpha$$

$$\alpha_2 = dVal + \alpha \quad (2)$$

Ship's latitude and longitude (lat_0, lon_0), (lat_1, lon_1) obtained from AIS message can be used to calculate the distance S (unit: km) between the buoy and ship and the angle $dVal$ (unit: radian) between connecting line and the positive north. The unit of latitude and longitude needs convert degree and minute to radian.

Next the angle $dVal$ needs to do further processing. Supposing that the beacon is the origin of a Cartesian coordinate system, the quadrant point of the ship can be defined based on the actual latitude and longitude position. If the ship is in the first quadrant, the angle will be not treated. If in the second or third quadrant, $dVal$ is changed to $(PI - dVal)$. If in the forth quadrant, $dVal$ is changed for $(2 * PI + dVal)$.

After the relative position of the ship and the beacon is set, in order to facilitate the calculation of the subsequent coordinates, the angle $dVal$ between the two connections is converted to one between the ship anchored north direction and angle. If $dVal$ is less than $3 * PI / 2$, $dVal$ is changed for $(3 * PI / 2 - dVal)$. Otherwise, $dVal$ is changed for $(3 * PI / 2 - dVal + 2 * PI)$.

The algorithm idea of the forth warning: the algorithm design for ships between 5 km and 20 km area. If it travels directly at the buoy in this area, the danger will be judged as the same way as the third warning. Ship message will be stored in a new ship list (list 2). Unlike the first list, list 2 does not cover the same vessels MMSI number storing dynamic information in consecutive time. Due to the different speed of ship, the system stores the information at different times. Otherwise the storage points are too close to lead to misjudgment and false warning. List 2 sets a certain fixed number of nodes. When the number of list nodes exceeds the quota, list 2 views information. Empty the list 2 once the investigation finished. The dynamic information of ship is stored with 5 array $uxxx[k]$. Xxx represents a record of the ship speed, latitude, longitude, heading or UTC time. K is the ship track point recording 50 tracks points. With the calculation of two adjacent track points, the system gets the angle $dAngleTrack$ stored in $dAngleTrack[k-1]$. This angle can help to determine the trend that the ship is away from the buoy or near buoy.

Firstly determine the relative position of in $uxxx[k]$ the first heading line and the connecting line between ship and buoy. That is judging whether the heading angle is greater or less than the angle between the line and the north direction.

If greater than the angle and the $dAngleTrack$ is getting smaller and smaller, the trend is judged to be close to the buoy; if less than and the $dAngleTrack$ is getting bigger and bigger, the trend is judged to be close to the buoy. The system issues the forth warning. The standard of $dAngleTrack$ judgment briefly defined as recording all of the dot tracks of the ship in the list. Firstly contrast the size of the two adjacent $dAngleTrack$, and then summarizes the number of the bigger case that the later angle is bigger than the front angle. If the result is greater than the half number of the track points, it can judge that $dAngleTrack$ angle increases, on the contrary, it can judge that $dAngleTrack$ angle is smaller.

After completely judging warning information of the ship, the system will store MMSI numbers with different warning levels in four different queues. AIS itself can promptly extract information from the queue to send the 6th message in an ABM format (considering AIS transmit gap and timing 10s).

Establish function *CalcPriWarn* (double dh) to achieve specific design algorithm. The inlet parameter is the maximum depth. Extract the ship's dynamic data and the latitudes and longitudes of the buoy from the lists and global variables. Finally, use data to achieve collision avoidance warning based on the theory of algorithms. The system calculates the distance of two points based on latitude and longitude and establishes function *CalcDistance* (float $fLat1$, float $fLong1$, float $fLat2$, float $fLong2$). The angle between two points and the north is calculated based on latitude and longitude. Build function *CalcAngle* (double $dLon1$, double $dLat1$, double $dLon2$, double $dLat2$). The inlet parameters of the two functions are the latitudes and longitudes of two points, the return values are the distance and angle, each unit is respectively kilometer and radian. Function *CalcPriWarn* divides into four warning areas based on the above specific algorithm design.

3.5 Sending a warning message

Considering the AIS device sends the package statements out with the interval between each frame of the statement, the system design queues to storage dangerous ship MMSI numbers and there is a risk that too much data could not be sent successfully once detecting data in the queue. So MMSI number and warning level sent off will be stored in the list. Detecting the arrival of the new early warning information, the system judges whether it belongs to the same ship. And if yes, send the latest vessel movement every 10 seconds; if not, store it in the list and then send out information. However, the new problem is that if a number of ships are in the warning area, early warning information will be sent at the same time, but not successfully. Therefore, warning information on different ships is directly packaged in different arrays and sent every 3 seconds.

4 TEST RESULTS

As shown in [Figure 3](#), it is one interface menu of AIS beacon. Display the latitude and longitude of voltage beacon in real time on the right top of the menu, the right bottom is menu selection, display the menu contents on the left side of the screen. [Figure 3](#) shows a beacon with all around the ship, at the same time you can zoom. AIS buoy voltage is designed to test in real-time and displayed on the LCD. The system is access to relevant information and owns information of the ships around the AIS beacon. Display the surroundings of the aid to navigation, predict the risk of the ship and record them.

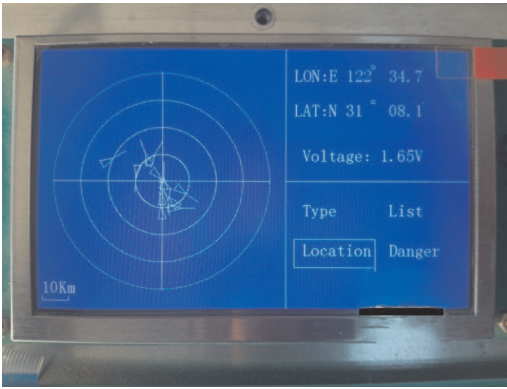


Figure 3. The menu of ship location.

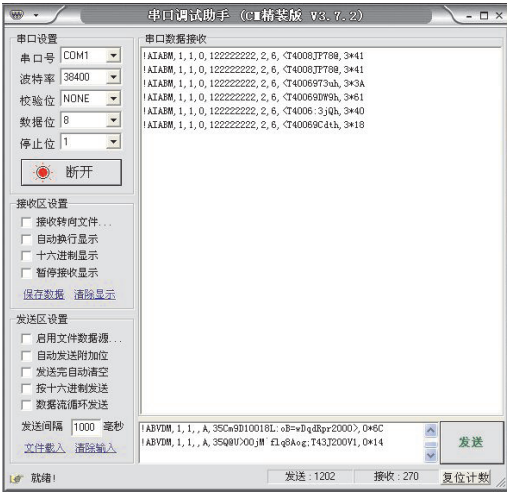


Figure 4. Serial port to send hips information out.

Test serial sent packaged dangerous information out (see Figure. 4). When there is the danger forecast of ship information, beacon will send the ship MMSI number and Hazard Class No. packaged in ABM statement foreign AIS station center.

5 CONCLUSIONS

This design is set up on the ARM7 development platform to achieve the AIS beacon controller system. First, the system completes the AIS buoy voltage

monitoring function. And then it uses a queue to receive AIS data to realize the acceptance of large amounts of data without loss of data. After saving the ship data in the list of the parsed information, appropriate algorithm is designed to display the relative position of the ship and buoy. Collision avoidance warning and hazard information can be sent to the AIS center in real time, so as to achieve a beacon collision avoidance function. Sending a dangerous message to predict the risk of the ship can effectively prevent the beacon from being destroyed, and timely stop the occurrence of dangerous events. The system greatly protects the maritime security of the beacon, thereby contributing to the safety of navigation at sea.

ACKNOWLEDGMENTS

This work was supported in part by the China National Science and Technology Infrastructure Program under Grant 2012BAH36B02, The State Key Program of National Natural Science of China under Grant 61231006. The Fundamental Research Funds for the Central Universities is NO.3132014329.

REFERENCES

- [1] Hu Wenhua; Xu Kaiyu; Zhou Weina, "AIS Information Display System Design Based-on the Embedded System," Electronic Measurement and Instruments, 2007. ICEMI '07. 8th International Conference on, vol., no., pp.3-191,3-195, Aug. 16 2007-July 18 2007.
- [2] Lessing, P.A; Bernard, L.J; Tetreault, C.B.J.; Chaffin, J.N., "Use of the Automatic Identification System (AIS) on Autonomous Weather Buoys for Maritime Domain Awareness Applications," OCEANS 2006 , vol., no., pp.1,6, 18-21 Sept. 2006.
- [3] Chang, S.J., "Development and analysis of AIS applications as an efficient tool for vessel traffic service," OCEANS '04. MTTs/IEEE TECHNO-OCEAN '04, vol.4, no., pp.2249,2253 Vol.4, 9-12 Nov. 2004.
- [4] Zheng Pan; Shujun Deng, "Vessel Real-Time Monitoring System Based on AIS Temporal Database," Information Management, Innovation Management and Industrial Engineering, 2009 International Conference on , vol.3, no., pp.611,614, 26-27 Dec. 2009.
- [5] Zhengxu Gian, "A Wave Sensing Drift Buoy and its Homodyne Receiver," OCEANS 81, vol., no., pp.437,441, 16-18 Sept. 1981.

3. Power electronics and energy engineering

Stochastic optimal dispatch model considering wind turbine shut down events under extreme wind condition

Y.X. Zhang & Z.Y. Dong

School of Electrical & Information Engineering, The University of Sydney, Australia

H.M. Yang

School of Electrical and Information Engineering, Changsha University of Science & Technology, China

ABSTRACT: The wind power output is affected primarily by external meteorological factors. Under extreme wind condition, individual wind turbine might automatically shutdown once the wind speed is near or exceeds the cut-out speed for self-protection. This results in the output power reduction of a wind farm, which raises the security concern due to the imbalance between generation and demand, particularly for power systems with high wind power penetration. In this paper, the Type I Extreme Value Theory is adopted to depict the extreme wind speed. A stochastic optimal dispatch model considering wind turbine unexpected shutdown events based on chance-constrained programming theory is established. An improved Genetic-algorithm is developed to solve the non-convex optimization problem. The results imply that the dispatch model will give full consideration of such emergency situation, and provide a more reasonable dispatch plan for power system disaster prevention and reduction.

1 INTRODUCTION

Wind power, one of the most appealing renewable energy resources, grew rapidly during the last decades. Currently, an increasing number of wind turbines have been installed all over the world. One of the major challenges is to adopt appropriate strategies to smooth the wind power output due to the intermittent and stochastic characteristics of wind resources. The power output increases linearly as the wind speed increases from the cut-in wind speed to rated wind speed. However, when the strong wind speed exceeds the cut-out wind speed, wind turbines are forced to shut down to avoid damage to the structure due to excessive mechanical loads. The wind power output fluctuation due to the volatile and stochastic characteristic of wind speed may gradually threaten the secure and stable operation of the power system. The previous study mainly focuses on the probability analysis when the WTs functions normally and the output wind power is available, however, the operation status of WTs and wind power output fluctuation under adverse condition are seldom discussed. In wind resources abundant area, the extreme weather event is not unusual, the power reduction caused by such extreme wind condition is worth studying.

Thus, in this work, firstly, an appropriate wind speed estimation method is adopted to explicitly analyze the extreme wind speed. Furthermore, a stochastic

optimal dispatch model based on chance constrained programming theory is established and the composite economic effects are analyzed. The proposed method is tested in IEEE-14 systems. The numeric results indicate that the model is, in some extent, effective and accurate.

2 WIND POWER OUTPUT PROBABILITY MODEL

2.1 *The type i extreme value theory distribution model*

The impact of variable wind speed influence on wind power fluctuation requires more specific and reasonable simulation methods. Various probability distribution models were proposed for the statistical analysis of recorded wind speed. Among them, The Weibull distribution is the most widely used classical distributions for fitting history wind speed especially under normal condition. The Extreme Value Theory is also well known in simulating extreme events such as: flood disaster; storm surge, extreme wind speed, etc. In this study, the Gumbel distribution model is applied to estimate the probability density function of extreme wind speed.

The Gumbel distribution is known as the Type I extreme value distribution, and its PDF is:

$$f(x) = (1/\sigma) \exp\{-\exp[-(x-\mu)/\sigma] - (x-\mu)/\sigma\} \quad (1)$$

The CDF can be obtained as:

$$F(x) = \exp\{-\exp[-(x-\mu)/\sigma]\} \quad (2)$$

Where μ and σ are location and scale parameters, respectively.

The commonly used methods for the determination of Gumbel parameters include least square method, method of moment and maximum likelihood method. The methods of parameter estimation are not what this work is about, all the parameters are calculated by the maximum likelihood estimation method.

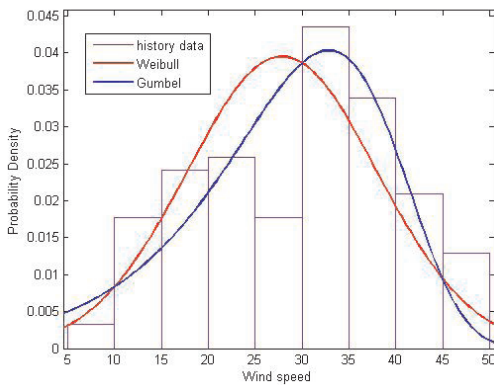


Figure 1. Cumulative probability function plot of extreme wind speed for station A.

Fig.1 shows a comparison of the cumulative distributions empirical of the three distributions of the recorded extreme wind speed. It can be seen that both Weibull and Gumbel are closer to the empirical cumulative distribution that was obtained on the basis of the history wind speed from certain station. However, by comparison, the Gumbel cumulative probability density distribution is more consistent with the calculated result with the wind speed data.

2.2 Wind energy conversion model

The general calculation form of wind power output is as Eq. 3, where ρ is the density of air in kg/m^3 , A is the blades area in m^2 , and v is the velocity of wind speed in m/s :

$$P_w = \begin{cases} 0 & v_w \leq v_{cut-in} \text{ or } v_w \geq v_{cut-off} \\ 1/2 \rho A v^3 & v_{cut-in} \leq v_w \leq v_{rated} \\ P_{rated} & v_{rated} \leq v_w \leq v_{cut-off} \end{cases} \quad (3)$$

The outage failure events of wind turbine generator may be affected by multiple external factors, temperature, wind speed, etc.. In this work we suppose the operation states of wind turbines are independent of each other, the wind power output is only influenced by the wind speed variation. Meanwhile, in an actual situation, a contingency may include the tripping of a branch, generator, or load. For calculation simplicity, assuming the single wind generator shut down events which do not lead to multiple generator/ load failures or any cascading failures. Random forced outage events of thermal/wind turbine generator, unexpected increases in the load, etc., thus, all the other uncertain factors as load and wind speed forecast error are neglected.

3 OPTIMAL DISPATCH FUNCTION FORMULATION

The optimal dispatch in a wind-thermal system involves the allocation of generation among wind plants and thermal plants as to minimize the total production costs while satisfying various constraints. Without the loss of generality, it's assumed that wind power generators are owned by the utility power producer, thus the operating cost is neglected. Each thermal generation unit has a limitation and can only carry a limited amount of spinning reserve. Spinning reserve refers to the unloaded generating capacity that is on-line and synchronized to the system and all of which can be made available within 10 minutes. Adopting spinning reserve strategy can effectively mitigate the influence when there is sudden voltage fluctuation caused by the disconnected wind turbine generator.

A contingency in this work is defined as unexpected shut down of wind turbine generator. In this paper we mainly focus on the extreme scenario: when the spinning reserve is inadequate to compensate for the unavailable wind power output, the system has to curtail certain amounts of load in order to maintain secure operation.

For clarity, the main assumptions of the proposed model are summarized below:

- Loads are assumed to be inelastic and do not participate in the reserve market, the active power losses are neglected.
- The minimum power output of each generator unit is assumed to be zero: this way, non-convexity issues are sidestepped.

The probability of shedding loads due to an un-forecasted wind power failure under some extreme wind condition mainly comprises of two situations:

- Not having a shortage of generation: the total invalid wind power is no greater than the previous system spinning reserve level.

- The available spinning reserve is less than the unavailable wind turbine generator, the system has to shed a certain amount of load to satisfy the power flow constraints.

The objective function involves two parts: the total cost of thermal generation cost; the value of load-shedding cost:

$$\min F_p \left\{ \sum_{i=1}^M (a_i P_{Gi}^2 + b_i P_{Gi} + c_i) + \sum_{j=1}^D \Delta c \bullet P_{Dj} \right\} \quad (4)$$

Commonly, there are two states of system generation units: on and off status. A two state Markov model is often used to depict the availability of wind power turbine. We define a set of controllable binary variables $u_i \in \{0,1\}$ ($i=1,2\dots N$) denoting the discrete distribution function of the i^{th} wind power generator operation status.

which subject to:

3.1 Real power balance constraints:

$$\sum_{j=1}^D P_{Dj} = \sum_{i=1}^M P_{Gi} + \sum_{i=1}^N P_{wi} \bullet u_i \quad (5)$$

Since P_{wi} is random variable, Eq.5 could be written as probability form:

$$\Pr \left\{ \sum_{j=1}^D P_{Dj} = \sum_{i=1}^M P_{Gi} + \sum_{i=1}^N P_{wi} \bullet u_i \right\} \geq \beta_1 \quad (6)$$

3.2 Thermal and wind power unit generation constraints

$$0 \leq P_{Gi} \leq P_{Gi, \max} \quad i \in M \quad (7)$$

$$0 \leq P_{wi} \leq P_{wi}^{\max} \quad i \in N \quad (8)$$

3.3 Spinning reserve constraints

$$P_{Gi} + R_i \leq P_{Gi, \max} \quad (9)$$

$$R_i = \min \left\{ d\% \times P_{Gi}^{\max}, P_{Gi}^{\max} - P_{Gi} \right\} \quad (10)$$

R_i is the spinning reserve provided by thermal unit i . $d\%$ usually takes 5%. During a sudden decrease in wind power output, the thermal units should be able to ramp up to compensate for the reduction in wind power.

3.4 System power flow constraints

$$-k_l^{\max} \leq k_l \leq k_l^{\max} \quad l \in L \quad (11)$$

Eq.11 represents the system power flow constraints which cannot be volatile under any circumstance.

3.5 Load shedding constraints

$$0 \leq \Delta P_d \leq P_{Dj} \quad (12)$$

When the spinning reserve cannot compensate for the system shortage, the system has to curtail some loads to maintain stable operation. ΔP_d represents the total shedding amount of load to satisfy system power flow constraints.

$$\Pr \left\{ \sum_{i=1}^n h_{it} \times (P_{wi,t} \bullet u_i + P_{Gi,t} - P_{Dj,t}) \geq f_{i, \max} \right\} \leq 1 - \beta \quad (13)$$

4 CASES STUDIES

In this paper, an IEEE 14-bus system is applied to verify the proposed model and solving method that considers the invalid generation due to wind turbine shut down events. The wind-thermal system consists of three thermal units, one wind farm located and bus 6, which consists of 14 wind turbines; The generator parameters are shown in Tables 1. There are eleven loads, and their shedding loads cost coefficient is shown in Table 2. Different confidence levels are considered to study the stability and reliability of the power system under extreme wind condition.

The Monte Carlo simulation generates a large number of scenarios subject to a Gumbel distribution, each with the same probability.

The solution procedure is described as follows:

Step 1: Initialize algorithm parameters and group individuals;

Step 2: Apply Monte-Carlo simulation technique to generate lots of different scenarios with the same probability;

Step 3: In each scenario, the amount of shedding load is determined according to operation strategy;

Step 4: Check whether the chance constraints are satisfied. If so, continue GA reproduction process to generate the next generation and go to step 5. Otherwise, generate new random group individuals and return to step 2;

Step 5: Repeat step 2 to 4 until the given iteration step is reached or other termination is satisfactory.

The simulation results in Table 3 and 4 imply that the total system has to provide a spare spinning reserve in case of wind turbine shut down events,

especially in extreme wind events prone area; meanwhile the scheduling schemes are correspondingly adjusted. To enhance system reliability, the total generation cost rises with the increase of confidence level β .

Table 1. Main Parameters of thermal units.

| Bus | a (\$/MW ²) | b (\$/MW) | c (\$) | P_{Gi}^{\min} (MW) | P_{Gi}^{\max} (MW) |
|-----|------------------------------|----------------|-------------|-------------------------|-------------------------|
| 1 | 0.043 | 20 | 765.4 | 0 | 332.4 |
| 2 | 0.25 | 20 | 654.3 | 0 | 140 |
| 3 | 0.05 | 25 | 843.2 | 0 | 100 |
| 6 | 0.05 | 25 | 956.7 | 0 | 100 |

Table 2. Cost coefficient of shedding loads.

| Bus | ΔC | | ΔC | | ΔC |
|-----|------------|-----|------------|-----|------------|
| | (\$/MW) | Bus | (\$/MW) | Bus | (\$/MW) |
| 2 | 54.3 | 6 | 28 | 12 | 45.3 |
| 3 | 235.5 | 9 | 73.8 | 13 | 33.8 |
| 4 | 119.5 | 10 | 22.5 | 14 | 37.3 |
| 5 | 39 | 11 | 29 | - | - |

Table 3. Comparison of total cost under different confidence level.

| β | Wind power Generation | Shedding load cost | Total cost |
|---------|--------------------------|--------------------|------------|
| | (MW) | (MW) | (\$) |
| 0.80 | 116.4 | 18.23 | 5845.6 |
| 0.85 | 112.7 | 23.56 | 5857.0 |
| 0.90 | 108.5 | 27.88 | 5863.4 |
| 0.95 | 103.2 | 29.76 | 5864.4 |
| 0.99 | 103.1 | 31.23 | 5864.4 |

Table 4. Comparison of spinning reserve amount under different confidence level.

| Different confidence level | 0.80 | 0.85 | 0.90 | 0.95 | 0.99 |
|-----------------------------|-------|-------|-------|-------|-------|
| Spinning reserve amount(MW) | 17.46 | 19.23 | 21.46 | 23.24 | 26.31 |

5 CONCLUSION

This paper proposes a stochastic optimal dispatch model considering wind turbine shut down events under extreme wind condition, which is successfully tested in IEEE 14-bus system. The simulation results imply that the presented model is, to some extent, effective and accurate. It can effectively consider the uncertain status of wind turbine generator, therefore, ensures a safe and stable operation of the system.

REFERENCES

- [1] C. Opathella. & B.Venkaresh. 2013. Managing Uncertainty of Wind Energy With Wind Generators Cooperative. In IEEE on Power systems 28(3):2918–2928.
- [2] Gundolf, Danny. 2001. Power Reserve in Interconnected Systems with High Wind Power Production. Proc. Porto IEEE Power Tech Conference, Porto, Portugal, 10th–13th, September 2001.
- [3] Jianhui, Wang. & Mohammad, Shahidehpour. 2008. Security-Constrained Unit Commitment with Volatile Wind Power Generation, In IEEE on Power systems 23(3):1319–1327.
- [4] N. J,COOK. 1982. Towards Better Estimation of Extreme Wind Speed. In Journal of Wind Engineering and Industrial Aerodynamics(9):295–323.
- [5] Oscar E, Moya. 2005. A spinning Reserve, Load Shedding, and Economic Dispatch Solution. In IEEE on Power systems 20(1):384–388.
- [6] T. Manco,COOK. &A. Testa.1982. A Markovian Approach to Model Power Availability of a Wind Turbine. Proc. Lausanne IEEE PowerTech Conference, Lausanne, Switzerland, 2007.
- [7] Y.Q, Xiao.& Q.S, Li.2006.Probability Distributions of Extreme Wind Speed and its Occurrence Interval. InEngineering Structures (28):1173–1181.

A machine learning approach to increase energy efficiency in district heating systems

F. Dalipi & S.Y. Yayilgan

Gjøvik University College, Faculty of Computer Science and Media Technology, Gjøvik, Norway

A. Gebremedhin

Faculty of Technology and Management, Gjøvik, Norway

ABSTRACT: Heat demand prediction is an important part of increasing system efficiency within district heating. To achieve this efficiency, the energy provider companies need to estimate how much energy is required to satisfy the market demand. In this paper, we propose a method to investigate the application of online machine learning algorithm to achieve energy efficiency and optimization in District Heating (DH) systems by predicting the heat demand on the consumer side. To accomplish this, we are planning to use operational data from a Norwegian company (EffektivEnergi AS, Hamar) for a group of buildings that are connected to DH in other places.

1 INTRODUCTION

District heating system is an optimal way of supplying heat to different sectors of the society such as residence, premises and industries. It gives us the possibility to use a variety of fuels and surplus heat for heat production. District heating is also instrumental in reducing the global and local CO₂ emissions. It offers an enormous flexibility to combine different types of energy sources efficiently. Among other things, the prominent advantage of district heating is that it enables the deployment of combined heat and power (CHP) production thereby enabling the most efficient use of primary energy [1].

The EU Commission's landmark Roadmap 2050 report indicates that to reach 80% emissions reductions in 2050 across the European economy as a whole, the power sector will need to be almost completely decarbonized (95-100%). Thus, the EU Commission's Power Program aims to decarbonize European electricity generation by 2050, by replacing unabated coal-fired power with renewables and flexible low-carbon back-up supplies. This requires policies that will encourage significant investments in new generation capacity and grids over the next two decades [2].

DH plays an important role in the energy system of the North European countries where the share of DH in the total heat market is significantly high. Though its share in the heat market is low, the introduction of DH has contributed to reducing consumption of oil and also electricity in Norway. Thus, due

to use of district energy networks, the use of oil and hydroelectricity has dropped about 10% in Norway [3]. Recently, Norway has built the world's largest heat pump district heating plant, providing 13 MW for Drammen, near Oslo [4].

In a DH system, heat is delivered via a network of pipes from heat production plants to end users for purposes of space heating and domestic hot water demand. The increasing degree of deregulation of the energy market and the increasing focus on energy efficient buildings place a demand on DH systems to be more efficient than ever. Moreover, the application of new technology and innovation in district heating is therefore considered to be essential to improve energy efficiency. Several factors can affect the operation of a DH system, such as energy demand patterns in consumer's side, delay in DH network etc. The time delay in a DH network is usually large compared with time delays in other parts of the DH system (e.g. building level substation). The time delays appearing in the network are mostly a result of the transportation time of the DH water from the production plant to consumers and back again. It can take several hours before a consumer registers a change in the supply temperature from the plant [5].

Moreover, there is usually response delay at the CHP plant to changes at the consumer's side. In bigger systems, it can take up to 10–12 hours before the change has reached the most distant consumer [6].

In addition, heat load prediction can also affect the DH operation by enhancing the performance of techniques developed for demand side management and

load control. Such techniques are applied in managing the heat utilization at the consumer's side [7], [8], [9]. In order to obtain an accurate prediction of the heat load, there is a need for computational models that can be incrementally updated to capture changes in the heat load over time and hence can adapt to the predicted heat load changes.

The main objective of this paper is to present a suitable machine learning method for enhancing energy efficiency at the consumer side of DH system and to discuss possible related parameters, which influence the operation of the targeted DH system. Our work is in progress and aims to propose a framework for the application of supervised learning as online method in heat demand prediction.

This paper is organized as follows: section 2 describes state of the art in the field; section 3 provides a brief overview of machine learning algorithms and heat demand prediction; section 4 discusses the proposed approach and section 5 concludes the paper.

2 STATE OF THE ART

So far, various prediction techniques have been applied to heat demand prediction. In general, these methods are machine learning and mathematical models while the heat demand pattern is usually a nonlinear, stochastic function.

District Heating (DH) related works such as [6], [10], [11] have approached DH systems by applying mathematical models for optimal control. Samuel Idowu et al. proposed a complex model, in which they use a combination of reinforcement and supervised learning to investigate how existing knowledge captured by smart city systems could improve heat load predictions from consumers [5]. Another combination of regression models to investigate heat load forecasting is presented in [12].

Nevertheless, dealing with massive amount of real-time data on one hand, and applying simultaneously different or a combination of techniques for prediction analytics purposes on the other hand makes this process complex and difficult to use and manage. There is another interesting model, which address the utilization of a grey-box that combines physical knowledge with mathematical modeling [13]. The main purpose indicated in this work is to find a model linking the heat consumption to climate and calendar information. The most widely used approach to predict the heat load in DH systems is Neural Networks [14], [15], [16].

Building an accurate and integrated model of a DH system would demand more than a simple approach. This is because of its magnitude, setup and multipart interaction between various units of the DH

network, including pipes and heat exchanger. By taking into account the DH's dynamic properties, which require a large amount of details, its model would ask high computation times and intensive computational resources for solving optimization and knowledge extraction problems. Thus, there is a need for focusing on the application of fast and efficient algorithms for real-time processing of data as a main goal to provide accurate predictions. Accordingly, machine-learning techniques are able to deliver highly scalable predictive analytics. Considering the fact that machine-learning algorithms permit to capture nonlinear variations of the time-delays in the network and nonlinear behavior of the heat demand, they are suitable to predict the energy efficiency enhancement at the consumer side of a district heating system.

3 MACHINE LEARNING AND HEAT LOAD PREDICTION

Machine Learning (ML) is a branch of Artificial Intelligence, with its focus on building models that learn from data. In general, ML can be clearly defined as a set of methods that can automatically detect patterns in empirical data, such as sensor data or databases and then use the discovered patterns to predict future data, or execute other types of decision making under uncertainty. ML is divided into three principal groups: Supervised Learning (predictive learning approach), Unsupervised Learning (descriptive learning approach) and Reinforcement Learning [17].

In supervised learning, a model using training data that consists of N data records $(x_1, y_1), (x_2, y_1), \dots, (x_N, y_N)$, called instances is built. Each instance is described by an input vector x_i - which consists of a set of attributes $A = \{A_1, A_2, \dots, A_m\}$ - and a label y_i of the target attribute that denotes the desired output. The purpose of this method is to train a model that can afterwards predict the label of new instances. Based on the type of output variable y_i , supervised learning tasks are divided as classification and regression. In cases where the output variable is categorical or nominal, the ML tasks are called as classification problems or pattern recognition, whereas in regression problems the output variable is a real valued scalar or takes continuous values. The most common algorithms used for classification tasks include Naïve Bayes, C4.5 decision trees and kNN (k Nearest Neighbor) while most widely used algorithms for regression include linear regression, decision trees etc.

The heat load in a DH system is defined as the amount of energy that would need to be added for space-heating and heating tap water at any given moment in time together with the distribution losses.

Generally, the heat demand is dependent on outdoor temperature regarding heating. Time of day, time of week and season of the year are also important as this gives indication concerning user related consumption pattern.

Because the modern district heating management systems collect large amounts of data during the heating operations, there is a need for memory efficient and high-speed models that will be able to process real time data. In this work, we will benefit machine learning for its capabilities of continuous learning or incremental feedback; especially we focus on the adaptive online methods of machine learning.

4 PROPOSED APPROACH

The innovations and emergence of new technology impose energy systems to face a challenge in adapting to customers' more flexible and individual solutions. Consequently, one needs to know more about what drives customers' requirements, choices and priorities.

The next generation district heating system promotes the idea of Total Thermal Comfort. In this case, the customer facility is equipped with cutting-edge technology (installation, operation and service) by the energy suppliers and puts the customer in a position to take full responsibility on heating adjustments and optimization with the help of visually simple and inspiring user interface in the form of apps to smartphones, tablets and PCs. In line with these objectives, we motivate the use of optimal control of energy distribution in a district heating system as a mean to increased energy efficiency. We postulate that with a well-learned optimal control of some parameters, higher energy efficiency will be achieved in the district energy network.

In this study, we aim to learn the optimal control in each building. Heat load prediction shall be approached with the supervised machine-learning algorithm.

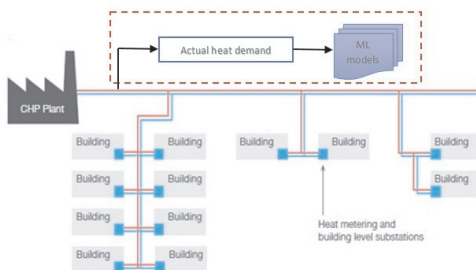


Figure 1. Proposed approach.

We plan to use the data collected from buildings at several locations for a set period (middle of October 2014 to April 2015) to assist in generating the required prediction model.

The following parameters will be considered in relation to the machine learning approach:

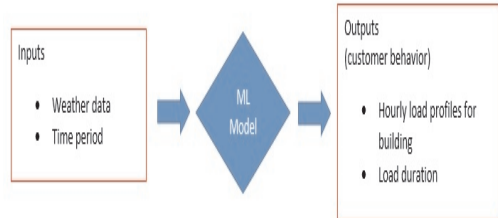


Figure 2. Simplified schematic representation of the Machine Learning approach.

Some of the currently identified weather parameters include indoor and outdoor temperature, humidity and wind speed, whereas the time parameters include time, day of the week, month and season of the year. The first stage of our approach will be to train the machine learning model for certain input values to certain output values in a supervised manner by using training data. At the next stage of our approach, we will test our prediction model with test data, which is different from the training data. Once the model learns the patterns from the inputs to the outputs in the training data, it will be able to predict corresponding outputs for provided input in the test data.

We are currently at the stage of data collection and once our experiments are completed, we will evaluate the performance of the proposed algorithm in terms of accuracy in predicting the heat load. For this purpose, we intend to use common evaluation metrics, such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

5 CONCLUSIONS AND FUTURE WORK

Nowadays, the research community is largely focused on the area of heat load prediction, since it can assist in energy efficiency of the district heating systems, which also leads to cost reduction for heat suppliers.

This paper introduces a work in progress project to achieve increased energy efficiency at the consumer side of the district heating system and identifies the supervised machine learning method as a suitable method to predict the heat demand. As future work, we intend to install modern smart sensors to

collect certain parameters about heat consumptions. This would stimulate the effective integration of smart city components.

ACKNOWLEDGMENT

The authors would gratefully acknowledge the support and collaboration of EffektivEnergi AS Hamar, Norway, for helping us to access measurement data at the user side from their measurement system.

REFERENCES

- [1] Gebremedhin A., Zinko H., Seasonal Heat Storages in District Heating Systems, *Effstock 2009 - Thermal Energy Storage for Energy Efficiency and Sustainability*, 14–17 June 2009, Stockholm, Sweden.
- [2] Persson J., Larsson M., Innovations in District Heating, Report produced by InnoHeat Project, Partly funded by EU, February 2014.
- [3] Pavlas M., Stehlik P., Oral J., Sikula J., Integrating renewable sources of energy into an existing combined heat and power system, *Energy*, 31 (2006), pp. 2499–2511.
- [4] <http://www.districtenergy.org/blog/2013/07/08/08/norway-starts-world's-largest-heat-pump-district-heating-plant/> [Cited 17 November 2014].
- [5] Idowu S., Åhlund C., Schelen O., Machine Learning in District Heating System Energy Optimization, *IEEE International Conference on Pervasive Computing and Communication Work in Progress*, 24–28 March 2014, Budapest, Hungary.
- [6] Benonysson A., Bhm B., Ravn H.F., Operational optimization in a district heating system. *Energy Conversion and Management*, 36(5):297–314, 1995.
- [7] Johansson C., Wernstedt F., Davidsson P., “A Case Study on Availability of Sensor Data in Agent Cooperation”, in *Intelligent Distributed Computing III 2010*, Vol. 7, pp. 111–120.
- [8] Johansson C., Wernstedt F., Davidsson P., “Deployment of Agent Based Load Control in District Heating Systems.” in *First International Workshop on Agent Technologies for Energy Systems 2010*.
- [9] Wernstedt F., Davidsson P., Johansson C., “Demand Side Management in District Heating Systems”, in *Sixth International Conference on Autonomous Agents and Multiagent Systems 2007*.
- [10] Choi S.Y., Yoo K.Y., Lee J.B., Shin C.B., M.J. Park, Mathematical modeling and control of thermal plant in the district heating system of Korea. *Applied Thermal Engineering*, 30(14–15):2067–2072, October 2010.
- [11] Chramcov B., Heat Demand Forecasting for Concrete District Heating System, *International Journal of Mathematical Models and Methods in Applied Sciences 2010*, Vol. 4, pp. 231–239.
- [12] Provatas S., Lavesson N., Johansson C., An Online Machine Learning Algorithm for Heat Load Forecasting in District Heating Systems, *The 14th International Symposium on District Heating and Cooling*, September 7–9 2014, Stockholm, Sweden.
- [13] Nielsen H.A., Madsen H., Modeling the heat Consumption in District Heating Systems Using a Grey-Box Approach, in *Energy and Buildings 2006*, Vol. 38, pp. 63–71.
- [14] Grzenda M., Macukow B., Demand Prediction With Multi-Stage Neural Processing, *Advances in Natural Computation and Data Mining 2006*, pp. 131–141.
- [15] Kato K., Sakawa M., Ishimaru K., Ushiro S., Shibano T., Heat Load Prediction Through Recurrent Neural Network in District Heating and Cooling Systems, *Systems, Man and Cybernetics 2008*, pp. 1401–1406.
- [16] Tae C.P., Ui S.K., Hyun K.L., Byung W.J., Yeong K.Y., Heat Consumption Forecasting Using Partial Least Squares, Artificial Neural Network and Support Vector Regression Techniques in District Heating Systems, *Korean Journal of Chemical Engineering 2010*, Vol. 27, pp. 1063–1071.
- [17] Murphy K.P., *Machine Learning: A Probabilistic Perspective*, Adaptive Computation and Machine Learning series, MIT Press, 2012.

Modeling of hypergene processes proceeding in the tailing dumps of Krasnorechenskaya concentrating mill of Dalnegorsky district in the temperature range of 0 to 45 °C using new thermodynamic data

V.P. Zvereva & A.M. Kostina

Far East Federal University, Vladivostok, Russian Federation
Far East Geological Institute FEB of RAS, Vladivostok, Russian Federation

A.I. Lysenko

Far East Geological Institute FEB of RAS, Vladivostok, Russian Federation

ABSTRACT: The authors have performed the modeling of hypergene processes proceeding in the tailing dumps of the Krasnorechenskaya concentrating mill in the positive temperature range of 0 to 45 °C. New data are obtained on the Eh-pH parameters of equilibrium solutions, quantitative composition of elements in the solution, and possible composition of hypergene and technogenic minerals crystallize in the mining-industrial technogenic system. The modeling is done with maximal, minimal, and average contents of sulfide minerals that characterize the tailings. In the calculations, we use new thermodynamic data for some hypergene minerals taken from the literature sources and obtained by the program founders.

KEYWORDS: hypergenesis, technogenesis, hypergene minerals, physicochemical modeling, tailings, tailing dump.

1 INTRODUCTION

Due to the development of thin-polymetallic deposits of the Dalnegorsky district (Russian Far East) many technogenic mining objects (ditches, adits, tailing dumps, and open pits) are formed. In 1956 within the Krasnorechenskaya village two tailing dumps – old and new – were formed, in which up to 1995 the tailings of the Krasnorechenskaya concentrating mill (KCM) were stockpiled. The tailings dump area is 27.4 hectares, and the amount of the accumulated tailings is 14.3 million tons. Old tailing dump was almost completely drained, and the new one is covered with a slime lake approximately for 1/20 part.

Hypergenesis and technogenesis are the main processes that through the mining activity resulted in the ecologically negative consequences in the atmo-, litho-, hydro-, and biosphere. To study these processes, proceeding in the wastes of the mining industry (tailings), and to determine the crystallization of possible hypergene and technogenic mineral phases, ion composition of solutions, and their parameters we have performed the modeling for the most common sulfide minerals of Dalnegorsk deposits in the contact with each other and with the enclosing rocks.

The modeling is done with the use of the “Selektor” program complex (elaborators I.K. Karpov, K.V. Chudnenko, V.A. Bychinsky). This program makes it possible to determine the volumes of elements evacuated into the hydrosphere from mining workings and tailing dumps. The method is based on the minimization of Gibbs energy and allows the calculation of full and metastable equilibrium in the multi-component heterogeneous systems (Avchenko et al. 2009, Borisov 2000, Bychinsky et al. 2008, Golovnykh et al. 2007, Grichuk 2000, Karpov 1971, 1977).

To create the physicochemical models it is necessary to set thermodynamic parameters of the components of water, gaseous, and solid phases. They're searching, as a rule, involves difficulties because of the imperfect study of hypergene and technogenic minerals. The physicochemical modeling of the processes of mineral formation by using the method of minimization of thermodynamic potential began to develop rather recently (80s of the last century) and was applied mainly to magmatic, hydrothermal, and metamorphic processes (Avchenko et al. 2009, Grichuk 2000, Avchenko & Chudnenko 2005, Artimenko et al. 2012, Bessonova et al. 2012, Daniel 2005). For the hypergene processes, modeling is used very rarely. Researchers

and elaborators must steadily supplement and improve the available bases of thermodynamic data. In our previous works (Zvereva et al. 2013a, b) we obtained interesting results, because after selection and calculation of new thermodynamic parameters (Tupitsin et al. 2009) we continue the work on modeling of hypergene processes proceeding in the tailings. In this paper the problem is considered from the standpoint of temperature effect on the hypergene process taking into account new thermodynamic data for a number of hypergene minerals existing in the given technogenic system.

It is clear that the elaboration of the computer models of the sulfide oxidation processes it is impossible to take it into account that all things influencing the direction and rate of chemical reactions that define what elements and in what quantities are present in pore waters and will be evacuated by drainage and mine waters.

The oxidation rate of sulfides depends on many factors: tailings humidity, their friability, temperature, pressure, redox potential, and other parameters. The tailings oxidation occurs mainly on the surface of tailing dumps because the oxygen content harshly decreases with depth. According to the experimental data, up to 35% of sulfides are in constant contact with air and at the sufficient amount of moisture may be oxidized for 6 months (Vishnyak 2005).

Relative rates of reaction of the mineral interaction with water have not been imposed into the calculations since the kinetics of changes in mineral phases under the processes of weathering and interaction with the atmosphere (silicates in particular) which has not been sufficiently studied yet. As a result, any rate ratios of the water – solid phase interaction can always be viewed as subjective, because during the process, the change of pH and Eh values, the component concentrations, and other records of the system result in the change of the kinetic reaction parameters.

Therefore, all data, obtained in modeling, on the quantitative ratio of mineral phases formed in models and on the element concentration in solutions may be considered possible and maximally approached to the real situation in the natural systems. The models are able to demonstrate the hypergenesis and technogenesis processes that proceed in all constituents of the mining-industrial technogenic system. They can show the qualitative and quantitative aspects of the potential ecological threat of wastes that is stockpiled in the tailing dumps which are not processed and been opened to access of weathering agents.

The aim of this work is modeling of the oxidation processes in the dried tailings in the old and new tailing dumps of KCM in the wide temperature

range from 0 to +45 °C. To achieve this purpose, it is necessary to solve the following tasks: to simulate the hypergene processes for different ratios of sulfide minerals and host rocks; to obtain Eh-pH system parameters; to determine the element contents in the solutions and possible mineral phases crystallized from them; to assess the impact of the process on the hydrosphere of the district.

2 RESEARCH METHOD

Since the annual temperature in the area changes from -40 to +40 °C (Vitvitsky 1961), and increases with sulfide oxidation, it is competent to consider the features of hypergene mineralization under different temperature conditions. For this reason physicochemical models in the positive temperature range from 0 to +45 °C in increments of 5 °C at 1 atm. are created to consider more objectively the processes occurring in the system. The water – rock ratio is taken as 10:1 (100 g tailings per a liter of water). Models are open to the atmosphere. The chemical composition of the atmosphere, calculated by R. Horne (Horne 1972), includes (%): N₂ – 78.088, O₂ – 20.949, Ar – 0.93, CO₂ – 0.03.

In the calculations in modeling of hypergene processes we take into consideration of 21 independent (Ag-Al-Ar-As-B-C-Ca-Cu-Fe-K-Mg-N-Na-Pb-S-Si-Sb-Zn-H-O-e) and 372 dependent components, among which 322 components are the dissolved particles, 18 gases, 32 minerals, and solid solutions (probable hypogene and hypergene minerals).

Into the newly formed models the following hypergene minerals are entered: hydrogoethite, fibroferrite, hisingerite, jamesonite, argentine, and acanthite with their thermodynamic parameters that are crystallized in a given mining-industrial technogenic system (Bulavko 2000, Zvereva 2001, Tarasenko & Zinkov 2001). They enrich the hypergene solutions with silver, antimony, lead, sulfur, and corresponding ions.

Chemical compositions and host rock - sulfide ratios are defined on the basis of literature data (Bulavko 2000, Tarasenko & Zinkov 2001). Ratio of mineral content in the host rocks and sulfides in the models of tailing oxidation are given in Table 1. The authors take maximal, minimal, and average concentrations of sulfide minerals in the tailings of KCM, so three variants of the tailing mineral composition are considered (Table 1). Sulfide ratios are determined in accordance with a possible mineral composition of wastes (Bulavko 2000, Tarasenko & Zinkov 2001), and the host rock – sulfides ratio for each of three variants is as follows: 95:5 and 60:40.

Table 1. Mineral composition of oxidized tailings, %.

| Minerals | Variants | | |
|--|----------|------|------|
| | 1 | 2 | 3 |
| Sphalerite ZnS | 29.9 | 34.9 | 19.9 |
| Galena PbS | 20 | 10 | 30 |
| Pyrite FeS ₂ | 5 | 10 | 10 |
| Pyrrhotite FeS | 10 | 30 | 15 |
| Chalcopyrite CuFeS ₂ | 5 | 2,5 | 5 |
| Arsenopyrite FeAsS | 5 | 2,5 | 5 |
| Argentite Ag ₂ S | 0.05 | 0.05 | 0.05 |
| Acanthite Ag ₂ S | 0.05 | 0.05 | 0.05 |
| Jamesonite Pb ₄ FeSb ₆ S ₁₄ | 25 | 10 | 15 |
| Axinite Ca ₂ FeAl ₂ [BO ₃] [Si ₄ O ₁₂](OH) | 20 | 20 | 20 |
| Diopside CaMg[Si ₂ O ₆] | 17 | 17 | 17 |
| Calcite CaCO ₃ | 10 | 10 | 10 |
| Albite Na[AlSi ₃ O ₈] | 20 | 20 | 20 |
| Clinochlore Mg ₅ Al[AlSi ₃] O ₁₀ (OH) ₈ | 13 | 13 | 13 |
| Feldspar K[AlSi ₃ O ₈] | 20 | 20 | 20 |

3 RESULTS AND DISCUSSION

Model results indicate a gradual reduction of pH from 9.73 to 8.74 (at the ratio of rock – sulfides 95:5) and 6.41-7.54 (60:40) and a slight change in Eh potential – 0.66-0.70 and 0.80-0.81 V (respectively) with temperature increasing. It can be explained by the increasing of intensity of oxidative reactions in the system. The temperature here acts as a catalyst.

The similar composition of aqueous solutions is formed containing the following ions:

- atmospheric gases are in the form of ions CO₂⁰, CO₃²⁻, HCO₃⁻, HNO₂⁰, HNO₃⁰, N₂⁰, NO₂⁻, NO₃⁻, O₂⁰, OH⁻, H⁺;
- host rock elements – Al(OH)²⁺, Al³⁺, AlO⁺, AlO₂⁻, AsO₄³⁻, B(OH)₃⁰, BO₂⁻, Ca(HCO₃)⁺, CaHSiO₃⁺, Ca²⁺, CaCO₃⁰, CaCl⁺, CaOH⁺, CaSO₄⁰, H₂P₂O₇²⁻, H₂PO₄⁻, H₃PO₄⁰, HAlO₂⁰, HP₂O₇³⁻, HPO₄²⁻, HSiO₃⁻, K⁺, KHSO₄⁰, KOH⁰, MgCO₃⁰, Mg(HCO₃)⁺, MgHSiO₃⁺, Mg²⁺, MgCl⁺, Mn²⁺, MnOH⁺, MnSO₄⁰, Na⁺, NaAsO₄²⁻, NaCl⁰, NaHSiO₃⁰, NaOH⁰, NaSO₄⁻, P₂O₇⁴⁻, PO₄³⁻, SiO₂⁰;
- sulfide elements – Cu(CO₃)₂²⁻, Cu²⁺, CuCO₃⁰, CuO⁰, CuOH⁺, CuSO₄⁰, H₂AsO₄⁻, H₃AsO₄⁰, HAsO₄²⁻, HCuO₂⁻, HFeO₂⁰, FeO⁺, FeO₂⁻, FeOH²⁺, HPbO₂⁻, HSO₄⁻, HZnO₂⁻, Pb(CO₃)₂²⁻, Pb(SO₄)₂²⁻, Pb²⁺, PbHCO₃⁺, PbNO₃⁺, PbO⁰, PbOH⁺, PbSO₄⁰, Zn(CO₃)₂²⁻, Zn(SO₄)₂²⁻, Zn²⁺, ZnCO₃⁰, ZnHCO₃⁺, ZnO⁰, ZnOH⁺, ZnSO₄⁰, SO₄²⁻ (sulfur amount in the solutions is stable and almost invariable).

The mineralization of solutions of two system types with host rock – sulfides ratios 95:5 and 60:40 according to the content of all (and sulfide) elements varies in the range of (g/l): 7.52 to 8.01 (2.88-3.04) and 27.28-27.49 (12.7-12.9), respectively, and it slightly grows up with temperature increasing.

For greater clarity of the temperature effect on the mineralization processes the diagrams of hypergene minerals content depending on the temperature (0 – +45 °C) for different ratios of host rock and sulfides – 95:5 (Fig. 1) and 60:40 (Fig. 2) are constructed.

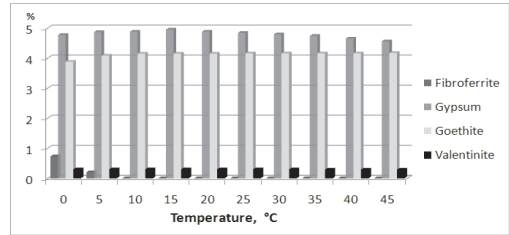


Figure 1. The content of hypergene minerals in the system at the temperature range of 0 to +45 °C, ratio – 95:5.

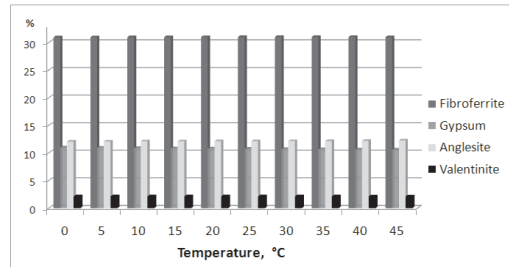


Figure 2. The content of hypergene minerals in the system at the temperature range of 0 to +45 °C, ratio – 60:40.

The diagrams show that a dependency of mineralization processes on the temperature is as follows:

- At the ratio of 95:5 the mineral percentage in the system changes insignificantly with temperature increasing. An exception is iron sulfate fibroferrite, crystallized at 0 to +5 °C (its amount is reduced). We can see low changes in the content of iron hydroxide – goethite α-FeOOH in the systems. The total amount of crystallized minerals is about 120 g;
- At the ratio of 60:40 amount of calcium sulfate – gypsum CaSO₄ · 2H₂O decreases, lead sulfate – anglesite PbSO₄ and iron sulfate – fibroferrite (Fe³⁺)[SO₄](OH) · 5H₂O increases slightly. Amount of crystallized antimony oxide – valentinite Sb₂O₃ does not depend on the temperature.

The total quantity of minerals in the system reaches 105 g.

In the modeling of the sulfide oxidation with temperature change the alkaline (60:40) and neutral (95:5) solutions are formed, oxidative conditions persist throughout the temperature range.

The behavior of the content of main ore elements in the modeled solutions at the ratios 60:40 and 95:5 under temperature changing from 0 to +45 °C is the same and varies slightly. In the limits of the solutions of each system, the dynamics of main sulfide elements in the models is different: amount of copper, zinc, arsenic, and silver is low and constant, and antimony and sulfur increases with temperature and reaches high concentrations. Lead concentration rises slightly. Iron in the systems is significantly less than other elements, because it is almost completely used in the mineral crystallization – goethite and fibroferrite.

Under sulfide oxidation their main components in the solution can be arranged in decreasing order in the following series: Zn > S > Pb > As > Cu > Ag > Sb > Fe (ratio 60:40) and Pb > Zn > S > As > Cu > Sb > Ag > Fe (95:5).

With temperatures increasing in the modeled solutions the ion composition varies insignificantly and is represented mainly by the same particles typical for standard conditions.

The mineral composition in the models is close enough: valentinite, gypsum, goethite, and fibroferrite. Anglesite formation depends directly on the amount of sulfur in the system; it crystallizes at the host rock – sulfide ratio of 60:40 in all three variants in the considered whole temperature range. It should be noted that the content of some minerals (anglesite and fibroferrite) in the system increases with temperature constantly or within a certain time interval. Other minerals (gypsum) tend to decrease their amount in the system or exist in a small temperature range and at a certain host rock – sulfide ratio (fibroferrite and anglesite).

Modeling with the widened composition of hypergene minerals and new thermodynamic data on minerals and dissolved particles carried out for KCM tailing dumps showed that the process of hypergene mineral formation is more intense. More complete mineral composition obtained in the modeling is fully verified by literature data (Bulavko 2000, Zvereva 2001, Tarasenko & Zinkov 2001).

Obviously, the completeness of the thermodynamic data provides a more accurate result in the modeling with the highest degree of compliance with the natural process. In variants with high sulfide content, as in nature, the system has a more acidic environment and the solution becomes more mineralized.

Verification, carried out with ionic and mineral composition (Bulavko 2000, Zvereva 2001, Tarasenko & Zinkov 2001) by comparing the content of the elements in the simulated solutions with data of the slime and drainage water of hydrochemical samples of slime and drainage water, confirmed the correctness of the results.

Physicochemical modeling shows how the process of sulfide oxidation occurs in the mining wastes. It allows to determine what minerals and in what quantities are crystallized from the saturated pore fluids. In addition, modeling makes it possible to estimate the concentration of the elements and their ionic forms in the solution and to establish its total mineralization. These data show the amount of toxic elements which are washed out into the surface and ground waters.

ACKNOWLEDGMENT

The authors are thankful to PhD Valeriy Bychinsky and PhD Alexey Tupitsin (Institute of Geochemistry SB of RAS) for their help in modeling.

REFERENCES

- [1] Artimenko, M.V., Chudnenko, K.V. & Koroleva, G.P. 2012. Physicochemical modeling of hydrothermal fluid flow in zones of hydrothermal magmatic system. *Proceedings of the Irkutsk State University "Earth Science"* 5 (1). 64–82.
- [2] Avchenko, O.V. & Chudnenko, K.V. 2005. Physicochemical modeling of mineral assemblages in metamorphic rocks. *Doklady Earth Sciences* 401(3). 378–383.
- [3] Avchenko, O.V., Chudnenko, K.V. & Alexandrov, I.A. 2009. *Fundamentals of physicochemical modeling of mineral systems*. Moscow: Nauka.
- [4] Bessonova, E.P., Chudnenko, K.V. & Cherepanov, E.K. 2012. Modeling of physicochemical transformation of modern volcanic-hydrothermal systems taking into account the dynamics of temperature and pressure in the deposit (using a program Selektor). *Geological evolution of the interaction of water with rocks: Proceedings of the All-Russia conf. involving foreign scientists*. 58–61.
- [5] Borisov, M.V. 2000. Thermodynamic models of Pb-Zn vein ore bodies formation. *Geochemistry* (8). 829–851.
- [6] Bulavko, N.V. 2000. *Mineralogy of skarn deposits of the Dal'negorsk ore field (Primorsky Krai)*. Vladivostok: Far East Publishing House.
- [7] Bychinsky, V.A., Didenkov, Y.N., Golovnykh, N.V., Shepelev, I.I., Tupitsin, A.A. & Chudnenko, K.V. 2008. Comprehensive measures for pollution reducing of natural waters in the slime storages of alumina refineries. *Geoecology. Engineering geology. Hydrogeology. Geocryology* (3). 22–231.

- [8] Daniov, B.S. 2005. Effect of fluid composition on the metamorphic assemblages in the basic rocks. *Petrology* 13 (1). 93–104.
- [9] Golovnykh, N.V., Bychinsky, V.A. Shepelev, I.I., Didenkov, Y.N., Chudnenko, K.V. Tupitsin, A.A. & Altynnikova, M.A. 2007. Physicochemical features of comprehensive utilization of solid and liquid wastes from alumina production. *Chemistry for sustainable development* (15). 417–425.
- [10] Grichuk, D.V. 2000. *Thermodynamic models of submarine hydrothermal systems*. Moscow: Scientific World.
- [11] Horne, R. 1972. *Marine chemistry*. New York: Wiley.
- [12] Karpov, I.K. 1971. Physicochemical modeling of equilibrium and non-equilibrium processes of mineral formation used the PC with methods of linear and nonlinear programming. *International Geochemistry Congress. Proceedings* (2). 511–513.
- [13] Karpov, I.K. 1977. Development of theoretical basis of physicochemical modeling of natural processes of mineral formation with the PC. *Basic Research. Earth Science*. 91–94.
- [14] Tarasenko, I.A. & Zinkov, A.V. 2001. *The environmental effects of mineralogical-geochemical transformation of the tailings of Sn-Ag-Pb-Zn ores*. Vladivostok: Dal'nauka.
- [15] Tupitsin, A.A., Mukhetdinova, A.V. & Bychinsky, V.A. 2009. *Preparation of the thermodynamic properties of individual substances to the physicochemical modeling of high-temperature processes*. Irkutsk: Publishing House of ISU.
- [16] Vitvitskiy, G.N., 1961. *The Far East. Physiographic characteristic. Ch. Climate*. Moscow: Publishing House of the USSR Academy of Sciences. 93–115.
- [17] Zvereva, V.P. 2001. Morphology and mineralogy of hypergenesis zone of polymetal deposits of Dal'negorsk ore district (Primorye). *Collection: Ore deposits of continental margins*. 180–190.
- [18] Zvereva, V.P., Kostina, A.M., Pyatakov, A.D., Frolov, K.R. & Lysenko, A.I. 2013a. *Physicochemical modeling of hypergene processes in the sulfur-containing mining technogenic systems of the south Far East*. Vladivostok: Publishing House of FEFU.
- [19] Zvereva, V.P., Lysenko, A.I. & Kostina, A.M. 2013b. Estimation of effect of hypergene processes proceeding in the mining-industrial technogene system on the hydrosphere of Dal'negorsk district using physicochemical modeling method (Primorsky Region). *Advanced Materials Research* 813. 246–249.

Research on steel industry, energy-saving efficiency evaluation of Tangshan

X.H. Wan, H.L. Wang & L.H. Zhou

College of Science, Hebei United University Tangshan, P.R. China

ABSTRACT: In this paper, we followed the relevant principles of establishing indexes and set up the efficiency evaluation system for Tangshan steel industrial energy conservation and emissions reduction. By analyzing the relevant date of the last ten years, we obtained scale efficiency, technical efficiency, the comprehensive efficiency and malmquist index. As shown by them, we found, in the iron and steel industry, the city scale reward had a decreasing trend, and the efficiency of energy conservation and emissions reduction was not stable. Finally, some suggestions on saving energy and reducing emissions had been made, that is, the municipal government should take certain measures to guide the development of iron and steel enterprise following the theory of circular economy and sustainable development so as to provide an impetus to energy conservation and emissions reduction.

1 INTRODUCTION

Since following the policy of reform and opening up, China has made steady progress quickly in the fields of economy, science and technology. However, along with obtaining a huge economic benefits, the phenomenon of environmental deterioration has become increasingly apparent and turned into one of the factors which restrict the development of our country's economy. In order to realize the harmonious development of man and nature, the government of China has attached great importance to the work of energy conservation and emissions reduction and regarded it as the main breakthrough to change the existing circumstances.

Tangshan is a traditional heavy industry city. The iron and steel industry is the first pillar of economic development and the main source of revenue of the city. Since 1998, the scale of iron and steel industry has expanded rapidly, and has formed regional iron and steel industry, in which the Tangshan steel & iron plays the primary role and local industry are the assistance that are mainly distributed in the north with abundant iron, Kaiping and Fengnan adjacent Tangshan city. In recent years, the productive forces of steel industry have developed rapidly and the scale has been expanded gradually, so far, Tangshan has become one of the important iron and steel bases. With the constant development of iron and steel industry, waste gas, waste water and various large amounts of waste emission have caused serious damage to Tangshan residents of ecological environment. On the other hand, since the iron and steel industry has huge energy resource consumption, the

low carbon economy would be the most development potential of the industry in Tangshan city.

Based on the relative theory abroad, the energy conservation and emissions reduction in our country mainly focuses on the research from the following aspects, such as, how to carry out the energy conservation and emissions reduction, the way of energy conservation and emissions reduction, the efficiency evaluation of energy conservation and emissions reduction, etc. In the literature, some authors also carry out their study on energy-saving emission reduction by introducing more advanced policies, making a summing up of the loopholes and deficiencies, and then putting forward a series of policies and suggestions, which have guided our country's energy conservation and emissions reduction work and have a great realistic meaning and practical value.

There are amounts of papers in the literature which focused their attention on energy conservation and emissions reduction of iron and steel industry. For example, Shu (2004) analyzed the situation of the overall steel operation of the year in our country, and put forward a series of measures to conserve energy and reduce emissions for iron and steel industry in China, including closing down outdated production facilities, using of mature technology, reducing emissions, establishing the evaluation index system of energy conservation and emissions reduction and so on. Yan and Chen (2009) made an estimation of China's steel production in 2020, and calculated the emissions ^[4]; Along this line, Chen (2008) further analyzed the present situation of energy conservation and emissions reduction, and suggested China should develop the steel industry circulation economy ^[5];

Some researchers deemed that the country's tax policy limited the export of steel products. They claimed to reduce the export tax rebate by raising tariffs, hence achieving the effect of energy conservation and emissions reduction [6] through effective tax incentive policy.

To sum up, the studies on energy conservation and emissions reduction in our country had their methods mainly focus on qualitative ones, not much work have been done by means of quantitative methods. In this paper, we base on the development of Tangshan iron and steel industry in recent years, on the basis of the Tangshan statistical yearbook, we establish a suitable system for Tangshan iron and steel industry of energy conservation and emissions reduction efficiency index system, and use the DEA model to evaluate the relevant data of energy saving and emission reduction. The main results will provide Tangshan government with a practical and theoretical basis to deal with energy conservation and emissions reduction of Tangshan iron and steel industry and is meaningful in reality and practical to help the municipal government make energy saving and emission reduction policy and achieve the goal of energy conservation and emissions reduction.

2 THE ESTABLISHMENT OF TANGSHAN STEEL INDUSTRY EVALUATION INDEX SYSTEM OF ENERGY CONSERVATION AND EMISSIONS REDUCTION

The index system should be established on the premise of being scientific and practical. In view of the specific characteristics of the iron and steel industry, it must be comprehensive, terse, reflects all aspects of the iron and steel industry energy conservation and emissions reduction, sets up the iron and steel industry energy conservation and emission reduction evaluation index to reflect the energy efficiency and environmental indicators. The data must be easy to be obtained and simple to be named. When we choose the selection of energy conservation and emissions reduction efficiency index, we have to consider three aspects, energy efficiency index, ecological environmental index and economic index. In terms of energy saving, we must bring the iron and steel industry, energy into consideration, because the steel industry is the largest coal consumption and electric power, the coal consumption and power consumption are two kinds of energy for the evaluation index system; In terms of emissions, we must consider the emission pollution which will affect the environment. According to the actual situation of Tangshan iron and steel industry of various data in this paper, the evaluation index of waste water and waste gas emissions is elected; And more important

on the one hand is to put economic factors into consideration, and emission reduction commitment needs the support of economy and science. In this study, because it is the study of the Tangshan Iron and steel industry energy conservation and emissions reduction, we put the net value of fixed assets at the end, the number of employees and industrial gross domestic product for an evaluation index into the model. The final selection of indicators is shown in table 2-1:

Table 2-1 Input and output index selection.

| | | indicators |
|--------|---------------------|---|
| Input | | At the end of the net value of fixed assets (X_1) |
| | | The number of employees (X_2) |
| | | The coal consumption (X_3) |
| | | The power consumption (X_4) |
| | The expected output | Waste water (X_5) |
| Output | | Emissions (X_6) |
| | expect output | Gross value of industrial output (X_7) |
| | | Steel production (X_8) |

The index is based on the actual situation of Tangshan iron and steel industry to be considered and achieved, from the size of these index values can be seen the city's iron and steel industry development and the efficiency of energy conservation and emissions reduction.

This article selects Tangshan related data from 2003 to 2012, as a decision-making unit, uses the BBC model.

3 RESEARCH ON TANGSHAN STEEL INDUSTRY ENERGY CONSERVATION AND EMISSIONS REDUCTION EFFICIENCY OF EMPIRICAL

3.1 Data sources and processing

All the data we used in this article comes from the Tangshan in 2003-2012 statistical yearbook, then we pick the conform to the requirements of the data related to steel.

Dealing with the date, firstly, in the selection of physical input index, because the net value of fixed assets and the number of employees data are accessible, namely choosing the two index as input indicators; Secondly, since coal and electricity are the main energy, the selection of the coal in energy consumption and the power of these two kinds of energy as input indicators; Third, the expected output is referring to the production process of waste water, waste gas, and other products, so be careful to handle the expected output. Many scholars have discussed the research at home and abroad, and summarized a series of methods, mainly including curve measure evaluation method, the pollutant input processing method, data conversion function processing method and processing method of distance function and so on^[7]. This article selects pollutants to make input processing method, because this processing method is the expected output as input indicators into the DEA model, which is consistent with the principle of reducing output unchanged inputs, and more convenient when processing.

3.2 Correlation test

Using a DEA model for evaluation of the efficiency, the input variable and output variable must conform to the principle which is the input variable increase, the output variable increase too. Otherwise, the evaluation model has no meaning. In this paper, we test the input variable and output variable in the Pearson correlation test, test results are shown in [table 3-1](#):

The test results show that the input variables and output variables in DEA model have a strong relationship. Therefore, in the models of present paper, as the inputs increase, output will increase. That is to say, the evaluation model in this paper is scientific and reasonable.

Table 3-1 presents Pearson correlation coefficient between the input and output variables.

| | X ₁ | X ₂ | X ₃ | X ₄ | X ₅ | X ₆ | X ₇ | X ₈ |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| X ₁ | 1 | | | | | | | |
| X ₂ | 0.95 | 1 | | | | | | |
| X ₃ | 0.82 | 0.73 | 1 | | | | | |
| X ₄ | 0.88 | 0.93 | 0.85 | 1 | | | | |
| X ₅ | 0.44 | 0.5 | 0.68 | 0.75 | 1 | | | |
| X ₆ | 0.82 | 0.94 | 0.75 | 0.58 | 0.48 | 1 | | |
| X ₇ | 0.87 | 0.95 | 0.71 | 0.95 | 0.68 | 0.59 | 1 | |
| X ₈ | 0.94 | 0.98 | 0.81 | 0.95 | 0.62 | 0.61 | 0.97 | 1 |

3.3 Model analysis and calculation results

3.3.1 Tangshan efficiency analysis of energy saving and emission reduction

By making use of DEAP2.1, we analyse the data of the Tangshan area from 2003 to 2013, and present the results in [Table 3-2](#).

As seen from the above results, the efficiency values of energy conservation and emissions reduction from 2002 to 2012 are, in order, 1, 0.935, 0.901, 0.917, 0.901, 1, 1, 1, 0.842, 0.887. It can be seen that in the past two years the comprehensive efficiency of energy conservation and emissions reduction of the city was lower than 0.9, but the comprehension efficiency of energy conservation and emissions reduction average value is 0.938. Energy conservation and emissions reduction of iron and steel industry in recent 10 years is in a relatively stable state, and even the comprehensive efficiency for a few year's trends toward 1. Both observations show that the iron and energy conservation and emissions reduction in the steel industry has done a lot of work and has obtained certain achievements. But on the other hand, we can see the size remuneration of the steel industry has changed. Namely, in 2002, 2007-2009, the steel industry scale has a little change; in 2011, the city of steel and steel industrial scale increases; in the rest of the few years, the city scale is in a state of decline. On the whole, the size of the steel industry is in a state of decline in the past ten years, which shows that in iron and steel industry development of our city has

Table 3-2 Tangshan energy conservation and emissions reduction efficiency calculation results.

| EFFICIENCY SUMMARY | | | | |
|--------------------|-------|-------|-------|-----|
| firm | crste | vrste | scale | |
| 1 | 1 | 1 | 1 | - |
| 2 | 0.935 | 1 | 0.935 | drs |
| 3 | 0.901 | 1 | 0.901 | drs |
| 4 | 0.917 | 1 | 0.917 | drs |
| 5 | 0.901 | 0.936 | 0.963 | drs |
| 6 | 1 | 1 | | - |
| 7 | 1 | 1 | 1 | - |
| 8 | 1 | 1 | 1 | - |
| 9 | 0.842 | 0.856 | 1 | irs |
| 10 | 0.887 | 1 | 0.887 | drs |
| mean | 0.938 | 0.979 | 0.959 | |

Note: crste: comprehensive efficiency; vrste: pure technical efficiency; scale: scale efficiency; drs: diminishing scale; irs: diminishing size; -: size remains the same.

been curbed in recent years, and presented the trend of decreasing of size. This requests the municipal government to find and solve the problem in time, and vigorously promote the development of Tangshan iron and steel. Meanwhile, we should also grasp the development of the energy conservation and emissions reduction work, for fear that the efforts spent in the steel industry become a mirage.

3.3.2 Malmquist index analysis efficiency

Through the data of Tangshan, we use the DEAP2.1 to calculate Malmquist index, get the results presented in table 3-3. And we can further decompose the Malmquist efficiency into comprehensive efficiency value and technological progress. Specific functional relations are shown below:

$$\text{Comprehensive efficiency (effch)} = \text{scale efficiency(sech)} \times \text{pure technical efficiency(pech)}$$

$$\text{Malmquist value (tfpch)} = \text{technology changes (techch)} \times \text{integrated efficiency (effch)}$$

As showed in Table 3-3, Tangshan Malmquist efficiency value is 0.972. That is to say, our city has made significant achievements in saving energy and reducing emissions.

3.4 Conclusion and suggestion

Firstly, the DEA results show that the study in the steel industry is insufficient and changes in technology efficiency is not high, hence the technology of energy conservation and emissions reduction contribution rate is not high. In this case, our city should pay attention to energy conservation and emissions reduction technology research, and improve the effect of the science and technology so as to provide technical support to the steel industry energy conservation and emissions reduction work. Technology research and development require a lot of talents, so we should cultivate the energy conservation and emissions reduction technology personnel and institutions, establish the advanced science and technology research team, increase investment

Table 3-3 Tangshan Malmquist efficiency energy saving and emission reduction.

| MALMQUIST INDEX SUMMARY OF FIRM MEANS | | | | | |
|---------------------------------------|-------|--------|------|------|-------|
| firm | effch | techch | pech | sech | tfpch |
| 1 | 1 | 0.972 | 1 | 1 | 0.972 |
| mean | 1 | 0.972 | 1 | 1 | 0.972 |

in research and development funds, and study some advanced production technology, thereby improving the contribution of technology for energy conservation and emissions reduction.

Secondly, we can see from the results, energy conservation and emission reduction efficiency of Tangshan presents an unstable trend without a stable value, varying from 0.8 to 1 as the passage of time. Therefore, the government should step up efforts in supervising the iron and steel industry energy conservation and emission reduction, strengthen the consciousness of energy conservation and emissions reduction along with iron and steel enterprise development, consolidate and governance some low resource utilization, serious environmental pollution, safe hidden trouble, poor technology and productivity of small and medium-sized iron and steel enterprise, and guide, encourage and support the healthy development of the iron and steel enterprises. Through the cooperation between the government and the iron and steel enterprises, the stable development of energy conservation and emissions reduction is guaranteed.

Thirdly, we should develop the circular economy vigorously. In today's society, developing circular economy of the enterprise has become an inevitable trend. Tangshan city as a resource-based city should develop circular economy as the main work of energy conservation and emissions reduction. Only increasing the recycling of the resources can effectively reduce the consumption of resources and waste, and promote energy conservation and emissions reduction work. The developing circular economy is not only the oral advice, but also should be put into the actual development. The government should take some mandatory measures to increase the degree of iron and steel enterprise's emphasis on energy conservation and emissions reduction. Iron and steel enterprises should make effective promotion and use resources, saving techniques to reduce the consumption of energy; Recycle secondary energy, improve energy efficiency; Recycle waste water, waste gas, waste through the relevant recycling technology. Waste recycling reduces the pollution to the environment, and what cannot be recycled is reasonable processing, real have security work to put into practice the idea of energy conservation and emissions reduction.

Finally, the government should strengthen and improve macroeconomic regulation and control the iron and steel enterprises, assessment to the state-owned and private enterprises are carried out in accordance with the provisions. Meanwhile, in order to realize the sustainable development of the whole steel industry, the government should also establish new rules and regulations to encourage and support the private enterprises and small and medium-sized steel enterprises energy conservation and emissions reduction work, speed up the adjustment of industrial

structure and product structure, eliminate backward production capacity and complete the industrialization of new technology.

ACKNOWLEDGMENTS

This project is supported by the mayor of Tangshan City Special Projects (No. 13110203A) and the National Natural Science Fund (No. 11401160). Thanks.

REFERENCES

- [1] TM alone. 2002. European Steel Industry and Climate Change. Legislation[J]. Steel Times International, (4): 34–36.
- [2] UK Energy White Paper. 2003: Our future-creating low carbon economy.
- [3] [3]Trabajos de Estadística.1953. Malmquist S.Index numbers and indifference curves[J]. 4: 209–242.
- [4] Huawei Lei. 2012. in various areas of our country iron and steel industry energy efficiency and energy conservation and emissions reduction potential analysis [D]. Guangzhou: jinan university master degree thesis: 5–8.
- [5] Bing Li. 2008. Iron and steel industry energy conservation and emissions reduction and recycling economy system [J]. Journal of comprehensive utilization of resources in China, (6): 12–14.
- [6] Guang Zhong. 2010. To promote energy conservation and emissions reduction of tax policy studies [D]. Shijiazhuang: hebei university master degree thesis: 34–39.
- [7] Jianling Zhang. 2008. Production enterprise ecological morphological evaluation efficiency research [D]. Changsha: central south university master degree thesis:28–32.
- [8] Yizi Yang. 2012. Research on steel industry energy conservation and emissions reduction efficiency evaluation of Chinese[D]. Harbin: Harbin industrial university economics Ph.D. Dissertation: 19–43.
- [9] Jun Kawal. 2000. The Development of the Environment Friendly Steel Products at Nippon Steel [J]. Journal of Nippon Steel Technical Report(1) : 1–4.
- [10] Lihong Song, Huizhong Dong. 2007. The thinking and countermeasures of energy conservation and emissions reduction iron and steel industry [J]. Industrial technology economy(11) : 82–84.
- [11] Pei Fang. 2009. To promote energy-saving and emission reduction countermeasures study [D]. Qingdao: Qingdao university, master thesis :2–3.

Selected aspects of wind- power plants building influence on the operation of radio communications system

M. Džunda & N. Kotianová

Technical University of Košice, Faculty of Aeronautics, Slovakia

ABSTRACT: This paper introduced one of the possible approaches to influence examination of wind - power plant installation in some locality on operational parameters of radio communications system (RCS). During influence examination at least two negative factors were identified. These factors includes loss of the signal caused by shielding, and devaluation of the signal by multi path wave propagation.

1 INTRODUCTION

The contribution presents one of the possible approaches to the evaluation of the influences of windmill construction (VE) in a given location exerted upon the operational parameters of the radio communication type of VOR (RNS). In order to ensure the required coverage, the RNS are protected by the protective zone defined in terms of distances and heights of the planes of the zone. (Grewing, 2004, Szabo, 2013).The planes of protective zone lines must not exceed the artificial and natural obstacles. The VE form obstacles which are, as a rule, as high as 100 m above the terrain.

2 DEFENCE LINE FOR VOR RNS VOR

The defense line of the VOR navigation system is determined as an area of the circular cone with its top located in the point where the VOR system antenna is fixed. The angle of ascent of this plane is 2°. The planes of the protective zone must not exceed the artificial or natural obstacles. If in the zone there are some obstacles, which transit over the determined planes, it may then result in the deformation of the area coverage by VOR RNS specific signals and thereby in the degradation of the operational parameters of these systems. At the propagation of the electromagnetic wave in meter band, diffraction of the electromagnetic wave is experienced on the top of the obstacle. Thereby, the electromagnetic waves are moving to the area of reception and reduction, the intensity of the electromagnetic wave don't need to be so articulated. When evaluating the effect of the signal loss as a result of shadowing, making use of the method of situation modeling is of advantage.

3 THE INFLUENCE VE EXERTED ON THE VOR RNS

When the influence of the VE exerted on the operation of VOR RNS is inspected, a minimum of three factors have been found, negative affecting their operation. Among those factors are (Grewing, 2004, Bloudíček&Jeřábek, 2007) errors in measuring the target azimuth, loss of signal resulting from shadowing and degradation of the signal owing to multi-way propagation. When all the three factors are evaluated, a key parameter is attributed to the distance of the windmill from the VOR radio navigation system. With the distance increasing, the impact of construction of VE on the VOR RNS is reducing substantially. Further important factor is the line of sight between the VOR RNS and the VE. For VOR systems, the errors of azimuth measurements are influenced by the distribution of the electromagnetic field. The errors in the azimuth must be simulated for the smallest height of coverage, which is defined for each VOR station. In simulation of error in azimuth measurement for the VOR system was conducted for the altitude of 1000 m and

$$F_T = 2Vr/\lambda = 2 \cdot 98,95/2,772 = 71,39 \text{ s}^{-1}$$

The results showed that the maximal error in the azimuth is about 0,7° for the DVOR and as much as 7,0° for the VOR. Modeling showed that the Doppler effect does not seem to be a problem for the VOR/DVOR. (Grewing, 2004)

4 LOSS OF VOR RNS SIGNAL RESULTINGFROM SHADOWING

Among the most significant effects of windmills on VOR RNS are the losses of signal due to shadowing.

The loss of signal as a result of shadowing may cause stoppage in measuring the navigation parameters and degradation of orientation of pilots at IFR flights. The obstacles exceeding the connecting line between the transmitting and receiving antenna are shadowing the location of the place of reception and cause degradation of the signal. Thereby, behind the windmill, an area with low intensity electromagnetic field may be formed. In case of having a VE located in the vicinity of a VOR RNS, shadowing occurs. Therefore, behind the VE, radio shadow is generated. See Figures 1a and 1b. The dimensions and distances between the VE, their number, distances from the VOR RNS and the surrounding terrain.

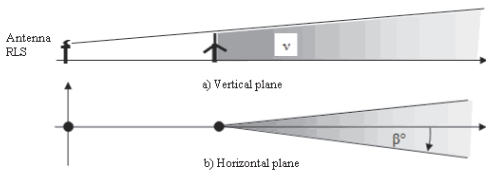


Figure 1. Mechanism of shadowing.

If the height of the obstacle, at short distances from the VE, is higher than the height of the VE, then the effect of the VE on the direct line of sight is to be neglected. If the height of the obstacle in front of the VE is higher than that of the VE, then the VE has no effect on the direct line of sight of the VOR RNS, which is in full extent determined by the dimensions of the obstacle. The consequences of the shadowing can be successfully predicted by using the method of modelling and simulation. Of advantage is the application of such software which enables entering into the digital model of the terrain and modelling obstacles in it. As the VOR system operates in the frequency band about

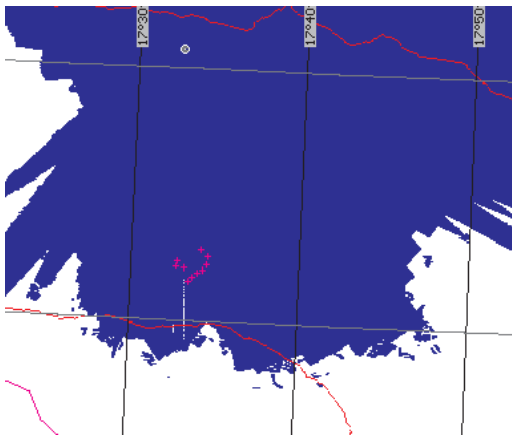


Figure 2. Simulation of VOR signal shielding.

118 MHz, simulation of ranges at various altitudes of flying for the VOR system have been made. Based on the results of the simulation, we can ascertain which of the wind mills cause losses of the specific signal as a result of shadowing and drop them from the plan of construction. The minimum losses of signal of the system as well as the minimum acceptable shadowing that must not occur in the areas of interest have been chosen as the criteria of success. The example of simulation of the VOR signal shadowing is in Fig. 2.

5 MODELING THE INTENSITY OF THE RNS VOR ELECTROMAGNETIC FIELD

Potential effects of the shadowing are to be neglected for real distances of wind generator from VOR/DVOR station. At the propagation of the electromagnetic wave in the meter band diffraction of the electromagnetic wave on the top of the obstacle occurs. Thereby, the electromagnetic waves are moving into the area of reception and degradation in the intensity of the electromagnetic wave need not be so substantial. When we evaluate the effect of signal losses resulting from shadowing, use of modeling and simulation is of advantage.

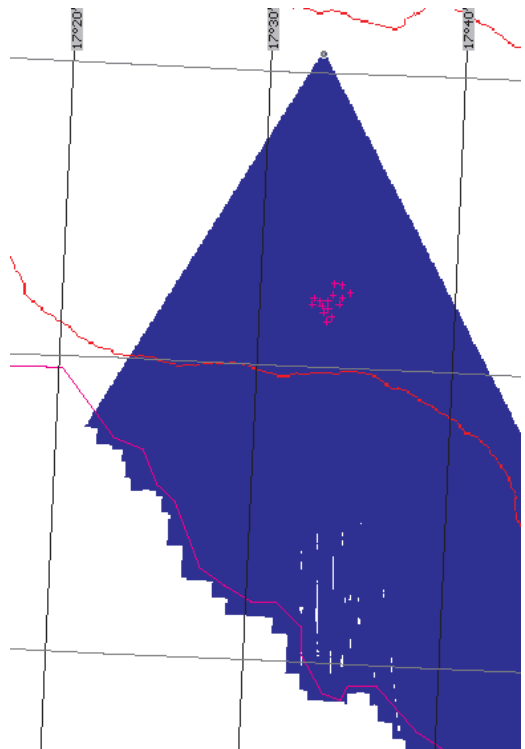


Figure 3. Simulation of shielding of the electromagnetic field of VOR system.

Comparing the results of simulation, we have found out how the distribution of the intensity of field changes for various altitudes of flying. As success criteria, we have chosen the minimal losses of the VOR system signal. Figure 3 shows the example of simulating shadowing of the electromagnetic field intensity of the VOR system.

6 DEGRADATION OF THE SIGNAL RESULTING FROM MULTIWAY PROPAGATION

Among the next negative influence of constructing VE exerted upon the VOR system is the degradation of the signal due to multi way propagation of signals. As a result, escape of signals and reflected signals can become sources of narrow-band interference. This effect is manifested if there is a direct line of sight between the VOR RNS VOR and the VE. If the VE were located close to the ground antennas of the VOR, then substantial degradation of the radio channel quality occurs. When we evaluate this influence, it is necessary to proceed from the mutual positions of the transmitting antennas of the VOR system and the VE, the modulation applied, performances of transmitters and the efficient radar cross-sections of the obstacles. Further, it is a must to be knowledgeable of the VOR receiver parameters. As a basic criterion for the evaluation we can use the comparison of output of potential reflections and that of the effective signal. The influence of interference exerted upon the operation of the VOR system has already been published, eg. (Džunda, 2006). The result of research in this field demonstrates that the greatest extent of interference to the VOR system may come from pulse jamming. As a success criterion is the minimum performance of the interference signal on the input of the VOR RNS (the required ratio of the effective signal output to that of the interference) if there is no direct line of sight between the VOR RNS and the VE.

7 CONCLUSION

The operation of the VOR RN is adversely affected by three main factors. Among them are the errors in measuring the target azimuth, loss of signal due to shadowing and degradation of the signal as a result of multi way propagation. In evaluating all the factors, the profile of the terrain along the connecting line between the VE and the VOR RNS is considered as a key factor, and, as the distance between them is increasing, the impacts of the VE construction to the operation of the VOR RNS are reducing substantially.

REFERENCES

- [1] Bálint, J.2006.Rizikovéfaktory v bezpečnosti letovej prevádzky.*Proc. of the International Conference "Air Force 2006"*.Brno:University of defence.
- [2] Bloudíček, R & JERÁBEK, J. 2007.New Technologies Implementation in the Air-Ground Radio Communications Systems.*Proc. of the International Scientific Conference "Modern Safety Technologies in Transportation"*.1st edition.ZlataIdka : Robert Breda. p. 30–35. ISBN 978-80-969760-2-7.
- [3] Džunda, M. 2006. Simulation of measuring distance using a DME system under conditions of interference. *Trans & MOTAUTO '06 : 13. international scientific - technical conference*. Bulgaria:Varna, N.Y. Vaptsarov. Naval Academy,4 p.
- [4] Greving, G. 2004.Modern Threats to Precision Approach and Landing - The A380 and Windgenerators and their Adequate Numerical Analysis.*Paper ISPA 2004 Intern. Symposium on Precision Approach and Landing, Munich 10/2004*.
- [5] Jeřábek, J. 2007.Informačnísystém v radiotechnickémzabezpečeníletectva AČR.*Proc. of the International Scientific Conference "Měření, diagnostika, spolehlivostpalubníchsoustavletadel"*.Brno :Univerzity of defence, s. 57-63. ISBN 978-80-7231-281-8.
- [6] Szabo, S. &Ferencz,V. &Pucihar, A. 2013. Trust, Innovation and Prosperity.*Quality Innovation Prosperity* vol. 17, issue 2, s. DOI: 10.12776/qip.v17i2.224., ISSN 1335–1745. Online <http://www.qip/>> <http://www.qip/journal.eu/index.php/QIP/article/view/224>.

Development of a solar cell robot optimization program GUI

M.J. Kim, T.W. Park, J.H. Lee, K.S. Sim & I.K. Hwang

Graduate School of Mechanical Engineering, Ajou University, woncheon-dong, Yeongtong-Gu, SuwonCity, Korea

D.I. Park

Korea institute of Machinery & Materials, Sinseongno, Yuseong-Gu Daejeon City, Korea

ABSTRACT: Recently, using the computer technology before development and application of the product is becoming more common. The entire development process tends to rely on computer analysis. Especially, analysis technology for various mechanical systems, such as solar cell transfer robotic, has dramatically developed. Now, the optimization design is proposed by using dynamic analysis without real product. In this study, optimization design modules are developed based on the theory of sensitivity analysis and optimal design. The program is based on multi-body dynamic analysis techniques. Methods of Plackett-Burman are used for sensitivity analysis. DOE and response surface analysis are used in the optimization process. The platform using Drag & Drop method is available to be modified. On optimization process, the auto-run is provided so that users can solve problems more easily. Through developed analysis platform, it is expected to develop the robot showing more effective and stable performance.

1 INTRODUCTION

The petroleum resource which we usually use is getting exhausted. So in order to find alternative energy researches are actively processed preparing exhaustion of petroleum energy. Solar Cell is one that is being studied actively in the field.

The solar Cell robot is used to handling the solar cell in each process. The Solar cell is a large and thin plate, which enables the solar cell to move stably. Therefore, it is very important to include the Hand deflection and stress analysis, to design a robot for Solar Cell transfer. In this study, we developed a program for the optimal design module based on the Solar Cell Analysis Program. The Solver is based on flexible Multi-body Dynamics Analysis technique.

2 DYNAMIC ANALYSIS PROGRAM

The Solar Cell Program provides a GUI that you can easily proceed with the modeling and analysis. Solver of the analysis program is based on the multi-body dynamic analysis techniques. The GUI is composed of a pre-processor, solver and post-processor as shown in Fig 2.1.

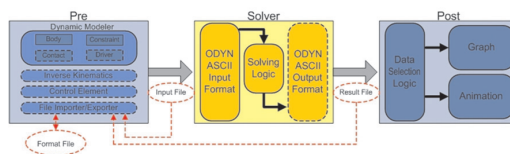


Figure 2.1. Configuration of the GUI.

2.1 Main program

Fig.2.2 shows the main screen of the program.

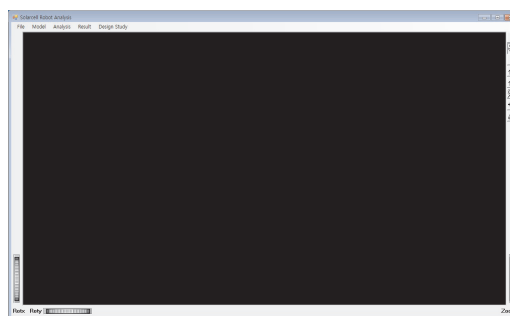


Figure 2.2. Main Screen of Program.

The detailed description of each menu is configured as shown in Table 2.1.

Table 2.1. Program Menu.

| | | | | |
|------|------------|----------|-----------|-----------------------------------|
| File | Model | Analysis | Result | DesignStudy |
| New | New | Solve | Graph | Case Study |
| Exit | Open | | Animation | Sensitivity Analysis Optimization |
| | Cad Import | | | |

2.2 Modeler

Modeler is configured as Fig 2.3.

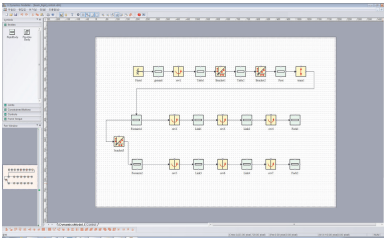


Figure 2.3. Modeler.

Each Block contains Body, Flexible Body, Joint, Control, Force and Motion information. The platform using Drag & Drop method is available to be modified.

2.3 Solve

The Solver can be executed in a program by [Analysis]-[Solver]. When calling the solver, let's set the direction of gravity. Click the [Run] after entering Analysis conditions. We can see the confirmation of the

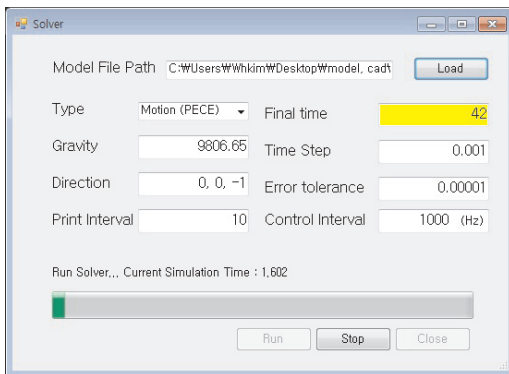


Figure 2.4. Solver Form.

analysis results as a graph. If we have CAD files, we can view the animation.

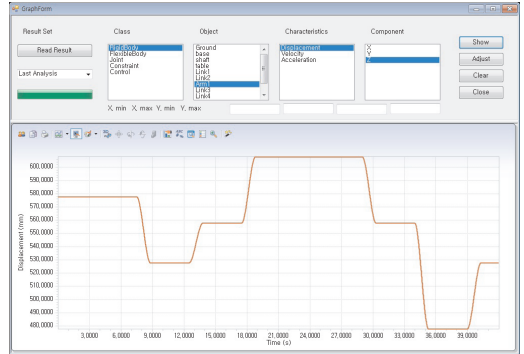


Figure 2.5. Graph of the Result.

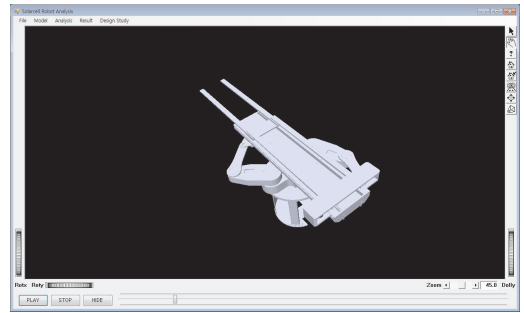


Figure 2.6. Animation of the Result.

3 DESIGN STUDY

3.1 Sensitivity analysis

The sensitivity analysis method for the mechanical system varies. In this study, we used the experimental method. Plackett-Burman method' from the experimental method is used. User can set the design parameters up to 11. By setting the range of

| Run | Fork2.density | Fork2.length | Fork2.Element_Young's_Modulus | Analysis result |
|-------------|---------------|--------------|-------------------------------|-----------------|
| Run 1 | 0.0000165 | 1600 | 12000000 | 643.557 |
| Run 2 | 0.0000165 | 1700 | 12000000 | 646.021 |
| Run 3 | 0.0000165 | 1700 | 14000000 | 648.947 |
| Run 4 | 0.0000195 | 1700 | 14000000 | 648.529 |
| Run 5 | 0.0000165 | 1600 | 14000000 | 648.75 |
| Run 6 | 0.0000195 | 1700 | 12000000 | 645.023 |
| Run 7 | 0.0000195 | 1600 | 14000000 | 649.36 |
| Run 8 | 0.0000195 | 1600 | 12000000 | 645.536 |
| Sensitivity | -0.147 | 0.165 | 1.891 | |

Figure 3.1. Sensitivity Analysis Table.

the design parameters, sensitivity Analysis Table as Fig 3.1. are generated.

The table was made with ‘Plackett-Burman method’. Use the table and drives the Solver. After solving is complete, set the objective function. Ultimately, the sensitivity analysis can be found in the objective function.

3.2 Optimization

When Sensitivity Analysis is complete, select the design parameters again. Optimization proceedings are similar to the initial proceedings of Sensitivity Analysis. After the selection of parameters, DOE table is created. Then the solver run and the objective function are set. Then, the ANOVA table is created as Fig 3.2.

| | S | Phi | V | F0 | F(0.01) |
|----------------------|-------------------|-----|-------------------|------------------|---------|
| Regression Variation | 43.3333333333333 | 3 | 16.1731938204444 | 218.623315222169 | 4.43 |
| Residual Variation | 0.813751672000012 | 11 | 0.073977442903982 | | |
| Sum | 48.5195814613333 | 14 | | | |

Figure 3.2. ANOVA Table.

In this result, when F(0) is greater enough than F(0.01), model function is appropriate in the 0.01 significance level. Finally, the optimum value is obtained as shown in Fig 3.3.

| Parameter | Optimized |
|-------------|-----------|
| Parameter 1 | 12.7 |
| Parameter 2 | 44.45 |
| Parameter 3 | 145 |

Figure 3.3. Optimization Result.

3.3 Numerical example

In order to verify the validity, Sensitivity Analysis and Optimization was performed targeting the 4 degrees of freedom Flexible Robot. Objective function was set to Deflection of Flexible Fork. The design parameters are listed in the table 3.1.

Table 3.1. Design Parameter.

| Parameter | Min | Max |
|------------------------------|------------|------------|
| Density(kg/mm ³) | 0.00000155 | 0.00000165 |
| Width(mm) | 50 | 70 |
| Height(mm) | 15 | 25 |
| Length(mm) | 1600 | 1700 |
| Young's Modulus(Pa) | 120000000 | 140000000 |

As we can check the results from Fig3.4, The sensitivity of the height is the largest and density is the smallest.

| Run | Fork2_Density | Fork2_Element_Width | Fork2_Element_Height | Fork2_Length | Fork2_Element_Young's_Modulus | Analysis Result |
|-------------|---------------|---------------------|----------------------|--------------|-------------------------------|-----------------|
| Run 1 | 0.00000155 | 50 | 15 | 1700 | 120000000 | 641.937 |
| Run 2 | 0.00000155 | 70 | 15 | 1600 | 140000000 | 646.021 |
| Run 3 | 0.00000155 | 70 | 25 | 1600 | 120000000 | 646.147 |
| Run 4 | 0.00000155 | 70 | 25 | 1700 | 120000000 | 648.529 |
| Run 5 | 0.00000155 | 50 | 25 | 1700 | 140000000 | 648.75 |
| Run 6 | 0.00000155 | 70 | 15 | 1700 | 140000000 | 646.021 |
| Run 7 | 0.00000155 | 50 | 25 | 1600 | 140000000 | 648.19 |
| Run 8 | 0.00000155 | 50 | 15 | 1600 | 120000000 | 645.139 |
| Sensitivity | -0.1471 | 0.195 | 1.931 | -0.901 | 0.323 | |

Figure 3.4. Sensitivity Analysis Result.

Based on the results, height, length Young's Modulus was chosen as the design variable. Fig 3.5 is ANOBATABLE.

| | S | Phi | V | F0 | F(0.01) |
|----------------------|--------------------|-----|---------------------|------------------|---------|
| Regression Variation | 64.8348733333333 | 3 | 21.593710379227 | 4419.81975149734 | 4.43 |
| Residual Variation | 0.0537421947516818 | 11 | 0.00489565406633562 | | |
| Sum | 64.7811311385811 | 14 | | | |

Figure 3.5. ANOVA Table.

It can be seen that the value of F(0) is greater than the value of F(0.01). This shows that the model function has a confidence. Finally, the results as Fig. 3.6 is obtained.

| Parameter | Current | Optimized |
|-------------------------------|-----------|-----------|
| Fork2_Element_Height | 20 | 25 |
| Fork2_Length | 1650 | 1600 |
| Fork2_Element_Young's_Modulus | 130000000 | 140000000 |

Figure 3.6. Optimization Result.

4 CONCLUSION

In this study, the optimization module is added to the Solar Cell transfer program. Using the module, users can conveniently obtain Optimization results. This study result is expected to have a positive effect on improving the performance of the Solar Cell Transfer Robot.

REFERENCES

- [1] HAUG, E.J., 1989, *Computer Aided Kinematics and Dynamic of Mechanical System*, Vol.1 : Basic Method, Prentice-Hall, Inc., 1989.
- [2] H. S. Han, J.H. Seo, *Design of a multi-body dynamics analysis program using the object-oriented concept*, *Advances in Engineering software*, Vol. 35, pp.95–103,
- [3] Shabana, A.A., *Dynamics of Multibody Systems*, 2nd Edition, Cambridge University Press, Cambridge, 1998.
- [4] M. Otter, H. Elmqvist, F.E. Cellier, *Modeling of multibody systems with the object-oriented modeling language Dymola*, *Nonlinear Dynamics*, Vol.9, 1996, pp.91–112.
- [5] O. M. Oshinowo, J.J. McPhee, 1997, *Object-oriented implementation of a graph-theoretic formulation for planar multibody dynamics*, *International Journal for numerical methods in engineering*, Vol. 40, pp.4047–4118.
- [6] D. L. Kunz, 1998, *An object-oriented approach to multibody systems analysis*, *Computers and Structure*, Vol. 69, pp. 209–217.

Research and design of a portable phase detector with power source based on the phase difference

T.K. Sun & D. Zhang

Zhecheng County Power Supply Co., Ltd Zhecheng County, China

W. Liu

Suixian Power Supply Co., Ltd Suixian, China

W. Liu

State Grid Electric Power Company in Henan province Shangqiu Power Supply Company Shangqiu, China

ABSTRACT: At present, the phase detector is made between two high voltage switchgears and the voltage signal used to detect the phase coming from the capacitive sensor. However, the indicators of each capacitive sensor are inconsistent, so some problems such as failure to detect the phase because of the large voltage difference and misjudgment of phase sequence detection can occur. In order to solve these problems, the phase difference method is proposed for the phase sequence detection between two high voltage switchgears. Not affected by voltage levels, this method has the versatility to judge the phase detection accurately. Actual test shows it can fully meet the phase detection requirements of high voltage switchgear.

KEYWORDS: Phase difference, Phase detector, High voltage switchgear.

1 INTRODUCTION

At present, the phase detection test for the high voltage switchgear is limited to the traditional voltage difference method, which could accurately detect the phase when the voltages of switchgears at two sides are basically the same; when the voltages are in-phase, the voltage difference is minimum, and it's usually considered as zero value, when out-phase the voltage difference is maximum. But when the voltages of switchgears at two sides are different, the difference is great even when they are in-phase, and it's impossible to judge whether it's in-phase or out-phase when using the voltage difference method. Especially in the actual application, when the capacity of capacitance type sensor is different, the coupled voltages could not be absolutely consistent, so the traditional voltage difference phase detection method could lead to serious problems such as misjudgment. In order to solve the problems mentioned above, this paper proposes the phase detector for high voltage switchgears based on the phase difference.

2 SELECTION OF PHASE DETECTION

At present, there are many phase detection methods, which can be divided into simulation method and

digital method. The simulation method transfers the time difference of signals of two circuits' zero-crossing time to voltage or current to calculate the frequency. Application of simulation method for phase detection requires a set of special hardware circuit which is not sensitive to the change of phase difference signal with poor accuracy. The digital method directly uses the count pulse to fill the phase difference pulse in the phase difference pulse, calculate the number of count pulse and get the average to calculate the phase. Therefore, the higher the count pulse frequency is, the higher the measurement accuracy is. Various electric instruments and mechanic instruments can be used; the vector method, multiplier method, diode discrimination method and sampling mixing method and the testing system by computers can be applied. During the actual application, different methods can be used according to specific conditions and the common applied methods are: quick phase detection method, correlation analysis method, zero-crossing method and wave transformation method, etc.

The measurement accuracy of quick phase detection method greatly depends on the selection of bipolarity sawtooth waveform, while the determination of the bipolarity saw tooth wave form depends on the peak value of input sinusoidal voltage signals. However, due to the fluctuation of network voltage, it can greatly affect the measurement accuracy.

Therefore, the quick phase detection method is not suitable for this design.

Although the correlation analysis method is the optimum choice in theory, its accuracy actually depends on the number of signal sampling points, which has a very high requirement on analog-digital conversion of sinusoidal signals; on the other hand, with the increase of sampling points, the algorithm complexity increases in quadratic relation. So, if the correlation analysis method is applied, it will have a strict requirement in the design of hardware and software, especially when it's required to improve the accuracy. It's necessary to apply the stricter hardware and more complicated procedure, which will cost a lot.

Therefore, this paper proposes and designs a phase detection method based on the combination of zero-crossing method and wave transformation method. This method transfers the signal to square wave through the wave transformation treatment of zero-crossing circuit and inputs the square wave signals of two circuits to the corresponding testing circuit to obtain the phase difference pulse signals, and then it can obtain the phase difference through the duty ratio of detecting pulse signal. The detection of duty ratio is similar to the detection of frequency and it's only necessary to detect the width of the high level of square wave. In this method, not only the hardware circuit is easily achievable, but also the software algorithm is simple and easy.

The zero-crossing method calculates two or more of the same frequency signals, and then converts the time difference to phase difference.

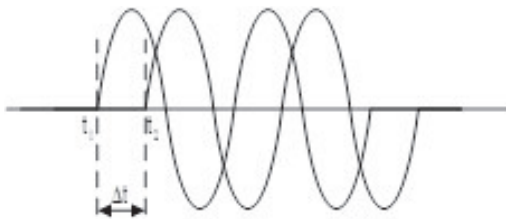


Figure 1. Zero-crossing principle diagram.

In order to determine the time interval Δt between t_1 and t_2 of two signals crossing the zero in Figure 1, it's possible to convert the time difference to phase difference. The signal sampling cycle is τ , n means the number of sampling points for t_1 and t_2 of two signals crossing the zero, and the signal cycle is T . The calculation formula of phase difference is:

$$P = 2\pi * \Delta t / T = 2\pi * \tau * n / T$$

In view of the distortion factor of network voltage due to nonlinear factor (such as loading), the detector filters the signals first to make the sampling signals close to sinusoidal wave and change it to square wave

through wave transformation. If the two square waves make exclusive-or processing, the output pulse width is the corresponding phase difference. As is shown in Figure 2, through the detection of pulse width, it's possible to calculate the corresponding phase difference.

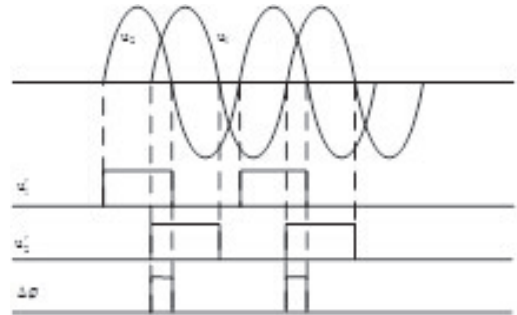


Figure 2. Wave transformation method schematic diagram.

3 THE PRINCIPLE OF GETTING ELECTRICITY FROM THE HIGH VOLTAGE ELECTRIC SIGNAL EXTRACTION

3.1 Power frequency electric field calculation principle [1-5]

The frequency of power frequency alternating electric field is 50(or 60) Hz, the wave is greater than the geometry of the study field, and then we can use the general concept of electrostatic field to analyze it. The algorithm on transmission lines and the surrounding of power frequency electric field commonly adopts the equivalent charge method based on the theory of mirror image method to calculate power frequency electric field intensity of high voltage transmission line space according to the method recommended by "International Conference on Power Grid 36.01 Working Group".

The calculation of high voltage line electric field adopts the equivalent charge method, and the equivalent electrical charge of the high voltage line is the line charge. As the radius r of the power line is far less than the erect height h , the position of equivalent electric charge can be considered in the geometric center of the electric wire. Assuming that the length of transmission line is infinite and it is parallel to the ground, and considering the ground as a good conductor, then we can calculate the charge of the unit length of wire.

3.2 The calculation of equivalent charge on unit length of wire

Assuming that the length of transmission line is infinite and it is parallel to the ground, and considering the ground as a good conductor, we can calculate

the charge of unit length of wire by adopting the mirror method. In order to calculate the (1) equivalent charge of multiple wires, by using Maxwell's potential coefficient method to calculate, we can get the following matrix equation: $[U]=[\lambda][Q]$ that is

$$\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \cdots & \\ \vdots & & & \\ \lambda_{n1} & \lambda_{n2} & \cdots & \lambda_{nn} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_n \end{bmatrix} \quad (1)$$

In the equation: $[U]$ is the unit matrix of voltage to ground on each wire; $[Q]$ is the unit matrix of equivalent charge on each wire; $[\lambda]$ is then-order phalanx of its potential coefficient and mutual potential coefficient on each wire (n means the number of wires). Matrix $[U]$ can be determined by the voltage and phase of transmission line, by taking the rated voltage as voltage calculation. It can calculate the voltage to ground of each wire by the line voltage of the three phase line, and the voltage is:

$$|U_a| = |U_b| = |U_c| = \frac{U_l}{\sqrt{3}}$$

The matrix $[\lambda]$ can be determined by the principle of the image. The ground is a plane whose potential is equal to zero, and the induced charge can be replaced by the image charge of image charge. i, j, \dots represent the actual parallel wires, while i', j', \dots represent their mirror.

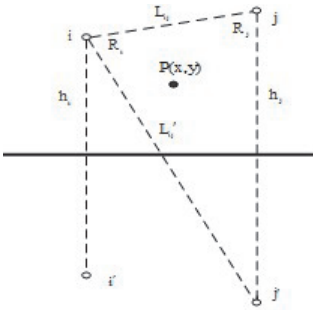


Figure 3. Potential coefficient calculation chart.

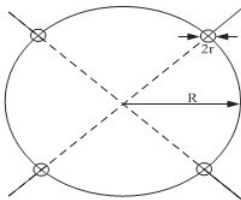


Figure 4. The equivalent radius calculation chart.

The potential coefficient can be written as:

Potential coefficient:

$$\lambda_{ii} = \frac{1}{2\pi\epsilon_0} \ln \frac{2h_i}{R_i} \quad (2)$$

Mutual potential coefficient:

$$\lambda_{ij} = \frac{1}{2\pi\epsilon_0} \ln \frac{L'_{ij}}{L_{ij}} \quad (3)$$

$$\lambda_{ij} = \lambda_{ji}$$

In the equation, ϵ_0 is the air dielectric constant, $\epsilon_0 = \frac{1}{36\pi} \times 10^{-9} \text{F/m}$;

R_i is power supply wire radius, and the conductor can be replaced with a single wire equivalent, and R_i calculation equation is:

$$R_i = R * \sqrt[nr]{R} \quad (4)$$

In the equation: R is split wire radius (as is shown in figure 2); n is the sub wire root number; R is the sub conductor radius.

It can solve $[Q]$ matrix by $[Q]$ and $[\lambda]$ matrices by using equation (1):

$$[Q] = [\lambda]^{-1}[U] = [\tau] [U]$$

Due to the fact that the AC voltage varies with time, according to the above principles, it can generate a moving electric charge that will generate current. For the known fixed arrangement of three phase or multiphase transmission lines, $[U]$ and $[\lambda]$ ($[\tau]$) are both constant value. In theory, $[Q]$ is constant, and the inverse matrix $[\tau]$ of potential coefficient $[\lambda]$ is capacitance. Checked by actual measure, the capacitance between copper conductor in high pressure casing and shielding net is 8P~30P. Because the acquisition terminal of the high voltage charged indicator signal is connected to the shield, it can get a voltage signal in high voltage charged indicator signal, and the value of voltage is only a few volts to dozens of volts.

4 ACQUISITION PRINCIPLE OF PHASE DETECTOR BASED ON PHASE DIFFERENCE

During the phase detection of high voltage switchgear, two electrodes are respectively inserted into the corresponding holes of potential indicator on the switchgear panel, and another one connects to the ground. Terminal s of phase detector contact with the capacitance sensor output of the measured high voltage wire, and the phase detector holes are coupling the voltage signal with phase relation frequency

as power frequency. It takes out the voltage wave signal with phase relation through the limiter circuit and delivers it to the comparator for shaping, as signal A and signal B in the [figure 5](#), and then sends the phase difference (impulse signal) to the microprocessor for acquisition. The pulse width can be calculated through gathering phase difference.

As there is no uniform standard defining what is in phase and what is out phase during phase detection of circuits at present, it's usually considered that when the phase difference value is less than 10° , it's in phase; when the phase difference value is more than 30° , it's out phase. And this phase detector for high voltage circuits based on phase difference is based on this standard.

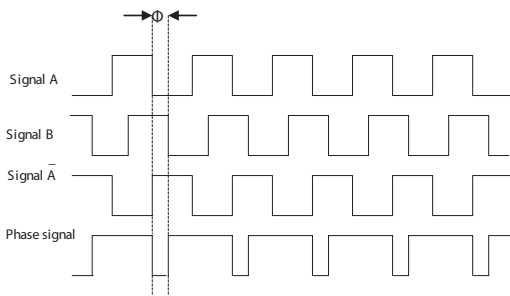


Figure 5. Acquisition principle of phase difference of high voltage circuit.

Signal 1 and signal 2 are the two cofrequency signals of phase difference of Φ . Make phase reversion of signal 1 and operate with another signal to obtain the phase difference signal, which is delivered to the comparator through integrating circuit to output the result of phase detection, as is shown in [figure 6](#).

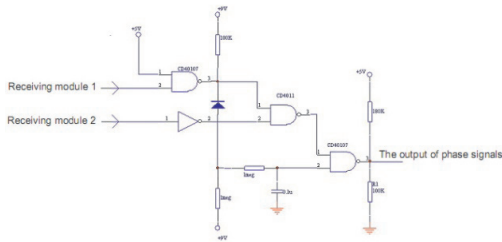


Figure 6. Phase detection for gate circuit.

5 HARDWARE PRINCIPLE OF PHASE DETECTOR

Phase detector includes: signal treatment circuit, microprocessor, power source, switching circuit, power light, 2 circuit lights, on phase light, and out phase light. Refer to [Figure 5](#) in the functional block diagram.

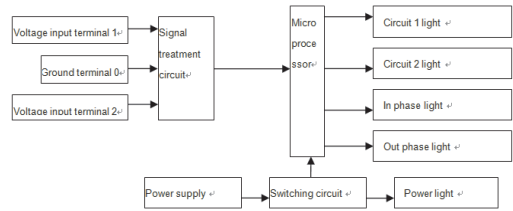


Figure 7. Functional block diagram of phase detector.

6 CONCLUSION

By adopting the above-mentioned principle, this paper designs a portable phase detector with power source based on the phase difference. This phase detector has the features of accurate judgment, portable and convenient operation. During the phase detection, two electrodes are respectively inserted into the corresponding holes of potential indicator on the switchgear panel, and another one connects to the ground, so we can accurately determine what is in phase or out phase. When the phase difference value is less than 10° , it's in phase; when the phase difference value is more than 30° , it's out phase. The phase detector has been put into use in many power companies, and it plays an important part in ensuring the safety of electric power.

REFERENCES

- [1] ZHANG Jinbo etc. 2005a. Design of wireless phase detector for high voltage power transmission line [J], *Electric power automation equipmen* 25(10):65–67.
- [2] ZHANG Jinbo. etc. 2005b. Design of practical wireless phase detector for high voltage transmission lines [J], *Computer measurement and control* (1):45–47.
- [3] ZHANG Jinbo. etc. 2006a. Study and realization of wireless phase measurement for high voltage electric power line [J], *Chinese Journal of Scientific Instrument* 27(11):1505–1507.
- [4] ZHANG Jinbo. etc. 2006b. Design of wireless phase detector for high voltage transmission lines [J], *High voltage technology* 32(8):108–110.
- [5] LI Guoding. 2000. *Fundamental Theory of Electromagnetic Field* [M], *Tsinghua University Press*: 8
- [6] OUYANG Qing. 2002. Phase-detect experimental procedures in Transmission Project Operation Field [J], *Electric Safety Technology* 13(2):21–23.
- [7] CHEN Benli. 1999. Phase sequence verification for 6KV bus bar of auxiliary power system [J], *Hubei Electric Power* 26(4):56–57.
- [8] TANG Mingdong. 1983. Generalized bridge method 110-220KV phase detection based on resistance [J], *East China Electricity* (11):72–84.
- [9] SU Wenbo. 2003. Study on phase-check method during start-up of power generation unit [J], *Electricity Construction* 24(8):51–54.
- [10] CHEN Qizhan, HUANG Wei. 2000 Apply voltage amplitude comparison method to determine the result of phase detection [J], *Rural Electrification* (05):36–37.

Three-level full-bridge bi-directional DC-DC converter

C. Yin

NR Electric, Nanjing, China

Y. Rao & K. Jin

Nanjing University of Aeronautics and Astronautics, Nanjing, China

ABSTRACT: A bi-directional DC-DC converter for high voltage application is employed in this article. To decrease the switcher voltage stresses, three-level full bridge architecture with PWM control in both power flowing directions is used. The operational principles of the proposed converter are discussed in details and tested in a 2kW prototype. The experiment waveforms are given and analyzed at the end.

1 INTRODUCTION

With the fast development of power electronics, bi-directional DC-DC converters are widely used in UPS, aircraft electric, electric car, renewable energy, etc. Bi-directional DC-DC converters are divided into two types: isolated ones and non-isolated ones. A Buck/Boost non-isolated bi-directional converter [1] could operate in Buck mode or Boost mode, depending on power flows. In [2], an isolated full-bridge (FB) bi-directional DC-DC converter, which is simple in architecture and control, is adopted in high power level application. However, in high voltage application, such as wind power generation, electric motor, the transistors sustain high voltage stress, bringing about rough optimizing in converter designation.

Three level (TL) converters [3], [4],[5] are preferred in high input/output voltage applications because of lower switch voltage stress, which is only half of the input voltage. In [6], non-isolated TL bi-directional DC-DC converter for high voltage and high power level application is proposed. The non-isolated converter hardly matches the high input-to-output voltage ratio.

This article adopts TL in full-bridge isolated bi-directional DC-DC converter and analyzes the operational principles in detail. A 2kW prototype is made and tested.

2 OPERATION PRINCIPLE

2.1 Control mode

PWM control is employed in the proposed bi-directional converter. In figure 1(b), TL bridge leg in primary or high voltage side has 4 transistors $Q_{H1} \sim Q_{H4}$. FB leg in secondary or low voltage side also has 4 transistors $Q_{L1} \sim Q_{L4}$. The switch transistors Q_{H1} (Q_{H3}) and

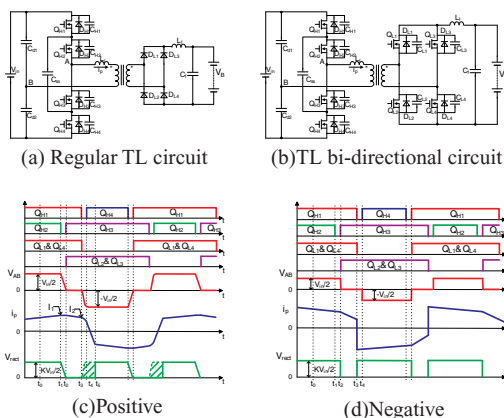


Figure 1. Main circuit and waveforms.

Q_{H4} (Q_{H2}) conduct complementary. Q_{H1} (Q_{H2}) and Q_{H3} (Q_{H4}) conduct with a phase delay of 180° . In this article, power flowing from high voltage to low voltage is defined as positive. Correspondingly, power flowing from high voltage to low voltage is defined as negative.

2.2 Operation principle

Figure 1(a) shows a unidirection TL Full-bridge DC-DC converter. The secondary side is generally full-bridge diode rectifier. To realize bi-directional power flows, diodes need to be replaced by switch transistors. The TL bi-directional full bridge DC-DC converter is shown in figure 1(b). TL bridge leg in primary side has 4 switch transistors $Q_{H1} \sim Q_{H4}$ (including body diodes $D_{H1} \sim D_{H4}$ and capacitances $C_{H1} \sim C_{H4}$). FB bridge leg also has 4 transistors $Q_{L1} \sim Q_{L4}$ (including body diodes $D_{L1} \sim D_{L4}$ and capacitances $C_{L1} \sim C_{L4}$). C_{d1} and C_{d2} are large enough

and divides equally the input voltage, $V_{cd1}=V_{cd2}=V_{in}/2$. Flying capacitor C_{ss} is large enough too, and its voltage keeps $V_{in}/2$. L_r is resonant inductor, it helps $Q_{H1}\sim Q_{H4}$ realize zero voltage switching. Before analyzing the converter, a few assumptions are made: (1) All switch transistors, diodes, inductor, capacitor and transformer are ideal; (2) C_{d1}, C_{d2}, C_{ss} and L_r are large enough. The voltage of C_{d1}, C_{d2}, C_{ss} could be seen as a voltage source, while the current of L_r as a current source.

2.2.1 Positive direction principle

Fig.1(c) shows the main operation waveforms of the TL bi-directional FB DC-DC converter work in positive power flowing direction.

Detailed analyses are as follows:

- $[t_0, t_1]$. In figure 2(a), Q_{H1} and Q_{H2} conduct, Q_{L1} and Q_{L4} conduct.
- $[t_1, t_2]$. In figure 2(b), Q_{H2} switches off at t_1 . Primary current i_p charges C_{H2} and discharges C_{H3} . The current i_p is approximate constant, because resonant inductor L_r and large filtering inductor L_f are in series. The voltage of C_{H2} starts a linear increase while C_{H3} a linear decrease. Meanwhile the junction capacitors of secondary side transistors Q_{L1} and Q_{L4} discharge until reaching 0.
- $[t_2, t_3]$. In figure 2(c), at t_2 , Q_{H3} , Q_{L2} and Q_{L3} turn on with zero voltage switching. Primary current i_p flows through a flying capacitor. The secondary side four transistors all conduct, which clamps the transformer voltage at zero. So it is a freewheeling process.

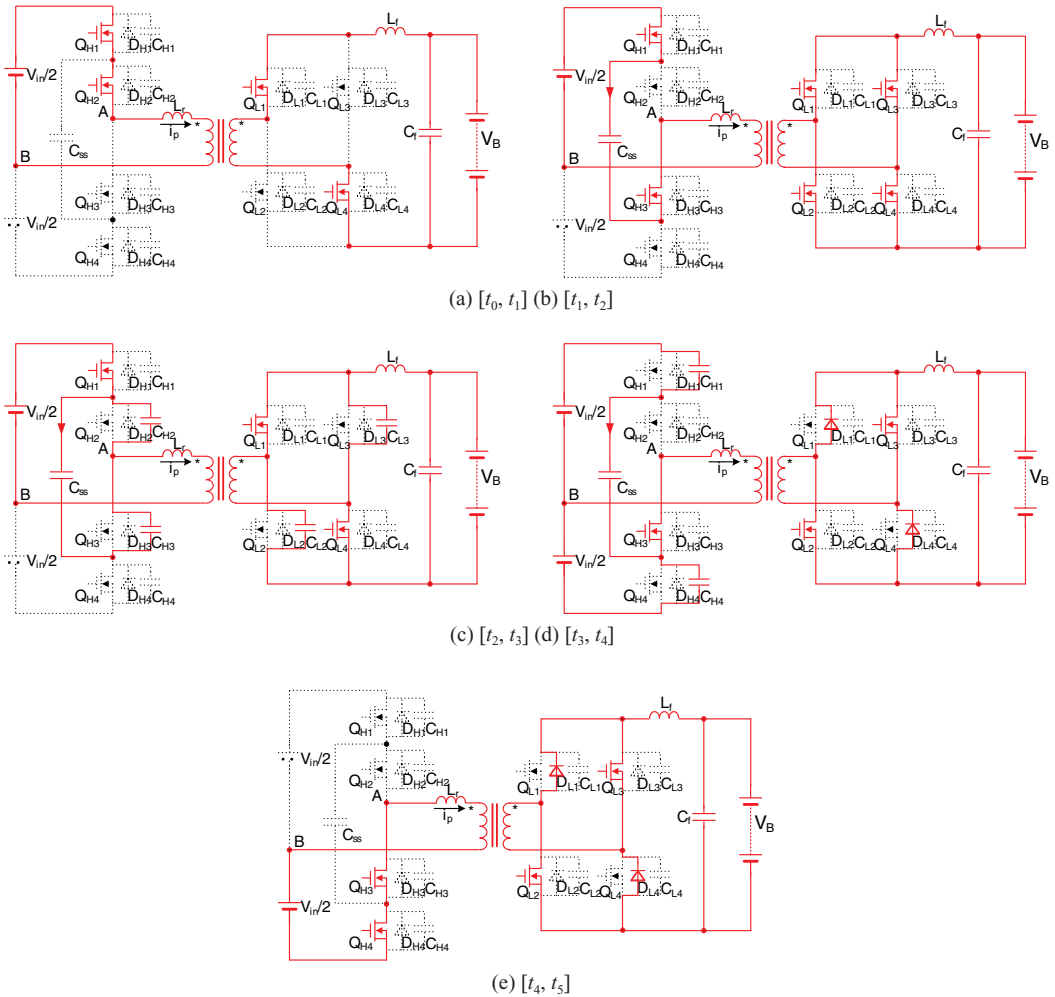


Figure 2. Positive direction switch modes circuits.

- $[t_3, t_4]$. In figure 2(d), at t_3 , Q_{H1} , Q_{L1} and Q_{L4} turns off. Primary current i_p charges C_{H1} and discharges C_{H4} . It decreases due to small resonant inductor L_r , resulting in inadequate current need of secondary side. Thus, in secondary side, transistors Q_{L2} , Q_{L3} and body diodes D_{L1} , D_{L4} conduct. This switch mode continues until the voltage of C_{H1} reaches $V_{in}/2$ and the voltage of C_{H4} reaches 0.
- $[t_4, t_5]$. In Figure 2(e), at t_4 , Q_{H4} turns on with zero voltage switching. Primary current i_p reduces to zero, and then increases inversely. The body diodes naturally turn off when i_p reaches load equal.

After t_5 , the converter starts the other half period and works likewise.

2.2.2 Negative direction principle

Figure 1(d) shows the main operation waveforms of the TL bi-directional FB DC-DC converter working in a negative power flowing direction.

Detailed analyses are as follows:

- $[t_0, t_1]$. In Figure 3(a), Q_{H1} and Q_{H2} conduct, Q_{L1} and Q_{L4} conduct. Power flows from secondary to primary side. The power inductor L_f releases energy, and its current decreases linearly.
- $[t_1, t_2]$. In Figure 3(b), Q_{H2} switches off at t_1 . L_f continues releasing energy. The existence of resonant inductor L_r keeps i_p from changing. And current i_p flows through Q_{H1} and D_{H2} .
- $[t_2, t_3]$. In Figure 3(c), Q_{H3} , Q_{L2} and Q_{L3} turn on at t_2 . For the voltage of flying capacitor is $V_{in}/2$, the body diode suffers inverse voltage and cuts off. The primary current flows through Q_{H3} and decreases linearly. Simultaneously the secondary side four transistors all conduct. The power inductor L_f stores energy.
- $[t_3, t_4]$. In figure 3(d), at t_3 , Q_{H1} , Q_{L1} and Q_{L4} turns off. The secondary side transformer current changes direction while the primary current doesn't. Considering that L_f is larger than L_r , primary

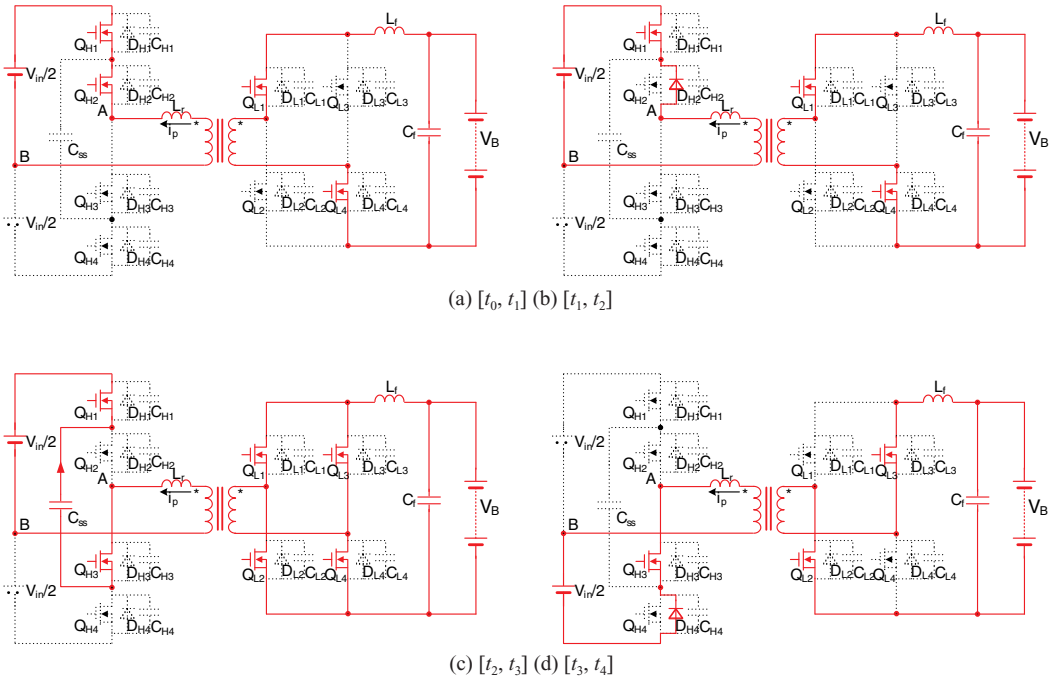


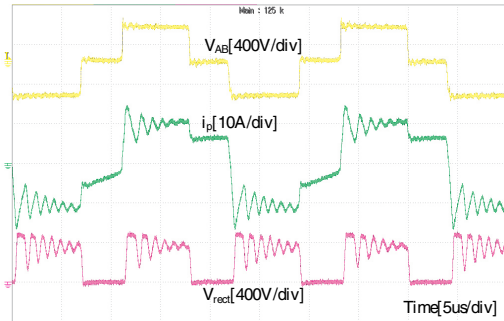
Figure 3. Negative direction switch modes circuits.

current is forced inverted, which leads to voltage spikes. The spikes could be reduced by RCD snubber circuit or active clamping circuit. Current i_p flows through Q_{H3} and D_{H4} .

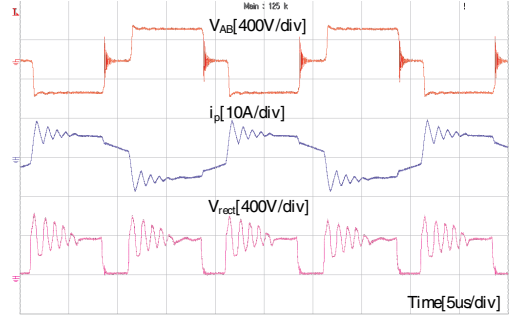
After t_4 , the converter starts the other half period and works likewise.

3 EXPERIMENT RESULTS

In order to verify the operation principle of this TL FB bi-directional DC-DC converter, a 2kW prototype is designed and tested. The parameters are as follows: Input voltage $V_{in}=690V$; Output voltage



(a)Positive



(b)Negative

Figure 4. Positive and negative experiment waveforms.

$V_b=216\sim 300\text{V}$; Power rating: $P_N=2\text{kW}$; Transformer ratio: $n_p=19:20$; Resonant inductor: $L_r=6.5\mu\text{H}$; Filter inductor: $L_f=515\mu\text{H}$; Flying capacitor: $C_{ss}=4.7\mu\text{F}\times 4$; Dividing capacitor: $C_{d1}=C_{d2}=470\mu\text{F}$; Transistors: IPW60R199CP; Switching frequency: $f_s=50\text{kHz}$. Figure.4(a), (b) shows positive and negative test waveforms, respectively. The primary side voltage of transformer V_{AB} changes among three levels: $+V_{in}/2$, 0, and $-V_{in}/2$. Through comparison between Figure.4(a), (b) and Figure.4(c), (d), the operating principle is convinced.

4 CONCLUSION

In this article TL is employed in bi-directional DC-DC converter to reduce the voltage stress of transistors and fit the high voltage and high power application. The proposed converter realizes soft switching within a certain load range when it is operated in the positive power flowing direction. Based on the analysis of converter operation principles, a prototype is made and the experimental results are discussed.

REFERENCES

- [1] Sable D M, Lee FC. A zero-voltage-switching bidirectional battery charger/discharger for the NASA EOS satellite. Proc. IEEE APEC, 1992. 614–621.
- [2] Yujin Song, Prasad N. Enjeti, A New Soft Switching Technique for Bi-directional Power Flow, Full-Bridge DC-DC Converter. IEEE IAS, 2002, pp.2314–2319.
- [3] Pinheiro J R, Barbi I. Wide load range three level ZVS-PWM dc-to-dc conversion. Proc. IEEE PESC, 1993. 171–177.
- [4] Pinheiro J R, Barbi I. Three level ZVS-PWM dc-dc converters-A comparison. Proc. IEEE PESC, 1995. 914–919.
- [5] Xinbo Ruan, Linqun Zhou, Yangguang Yan. Soft-switching PWM three-level converters. IEEE Transactions on Power Electronics, 2001, 16(5):612–622.
- [6] Yu Du, Xiaohu Zhou. Review of Non-isolated Bi-directional DC-DC Converters for Plug-in Hybrid Electric Vehicle Charge Station Application at Municipal Parking Decks. IEEE APEC, 2010, pp.1145–1151.

Fabrication of *n*-DLC/*i*-SiC/*p*-Si photovoltaic cells by pulsed laser deposition at the electrical field

S. Kurumi & K. Suzuki

College of Science and Technology, Nihon University, Surugadai, Kanda, Chiyoda-ku, Tokyo, Japan

S. Yonekawa

Graduate Student, Electrical Engineering, College of Science and Technology, Nihon University, Surugadai, Kanda, Chiyoda-ku, Tokyo, Japan

ABSTRACT: We report on the fabrication of *n*-DLC/*i*-SiC/*p*-Si photovoltaic cell by pulsed laser deposition at electrical field. Raman spectra showed that the $I(D)/I(G)$ ratio of *n*-DLC increased with increasing temperature, as well as open voltage and current density. The SiC layers were analyzed under irradiation by electron spectroscopy for chemical analysis. Our method could synthesize *i*-SiC layers between *n*-DLC and *p*-Si.

1 INTRODUCTION

Photovoltaic cells are attractive from the energy viewpoint, and they are now a required new material. Whereas general materials such as photovoltaic cells made of single-crystal Si are investigated, which have a theoretical efficiency of approximately 26%. [1] On the other hand, carbon materials are developed into better efficiency photovoltaic cells [2-11]. In a previous study, we had synthesized the hetero-junction of P-doped *n*-DLC and *psi* as a photovoltaic cell by pulsed laser deposition (PLD). [12] This paper shows that the efficiency of photovoltaic cells is increased by infrared light heating. In order to increase more, we attempted to synthesize SiC layer between DLC-layer as a tandem photovoltaic cell. SiC layer was grown by PLD in an electrical field. Because carbon plumes, which ablated by laser irradiation, are accelerated in this field. Therefore, carbon and the surface of a Si substrate combine with SiC layers. In this study, we report on the growing technique of the *n*-DLC/*i*-SiC/*p*-Si layers by PLD at the electric field.

2 EXPERIMENTAL

Figure 1 shows experimental system. We chose the PLD method for depositing phosphorus (P) -doped DLC. *n*-DLC:P films were prepared on *p*-Si substrates by YAG laser (LOTIS, LS2147, wave length: 1064 nm, pulse width: 20 ns, energy: 100 mJ/pulse, and repetition rate: 10 Hz), which was focused on the target bulk at an incident angle of 45 degrees. The target bulks were prepared by mixing the graphite powder with of 1.0 wt%

P.[8] The substrates were cleaned by ultrasonic cleaning using ethanol and wet etching with ammonium fluoride (NH_4F) at 343 K in an argon atmosphere of 30 mTorr. Anode and cathode electrodes were attached back of the targets and Si substrate, respectively. Direct voltage was applied to the electrodes. The laser irradiated the target bulks for 15 min. The substrates and ablation plumes were irradiated by infrared (IR) light and the substrates were heated with an IR light to 373, 473, 573, 673, and 773K. *p*-DLC:P film structure was analyzed by Raman spectroscopy (Renishaw System-1000) measured with a spectrometer (StellarNet, EPP2000C-50). The chemical composition of the *n*-DLC/*i*-SiC/*p*-Si layers were analyzed by electron spectroscopy for chemical analysis (ESCA Shimadzu ESCA-850). Current-voltage (*I*-*V*) characteristics of the DLC films were measured by a two-probe method under an illuminated Xe lamp. Ion sputtering was used for forming an Au ohmic contact between the anode and the cathode of the cell.

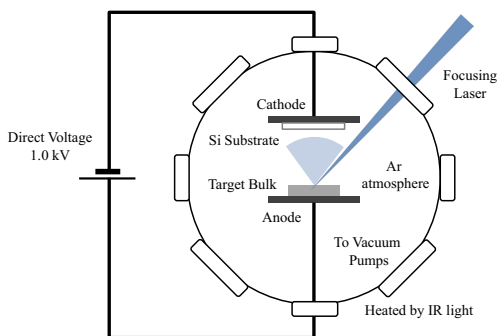


Figure 1. Experimental system of PLD at electrical field.

3 RESULTS AND DISCUSSION

Typical Raman spectra of DLC is presented in Figure 2. The Raman spectra of DLC shows two sharp modes, the G peak around 1580–1600 cm^{-1} and the disorder peak (D peak) around 1350 cm^{-1} [13]. The graphite peak (G-peak) occurs because of the bond stretching of all pairs of sp^2 atom sites in both rings and chains. The D peak occurs because of the breathing mode of sp^2 atoms in the rings [14]. To examine the crystallization of DLC films, the intensity ratio of the peaks is denoted as $I(D)/I(G)$. This ratio increases from 0.32 to 0.79 with increasing temperature and saturates above 473 K. This means that heating over 473 K can deform the DLC microstructure and affect the DLC film structure. Depending on temperature and open voltage, current density is shown in Figure 3. The open voltage of films increases in

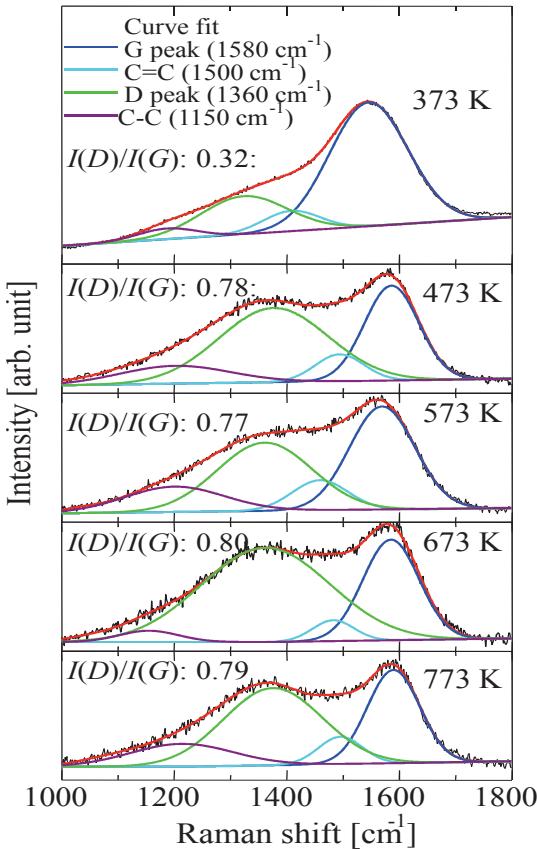


Figure 2. Raman spectra of DLC films deposited by PLD at electrical field.

substrate temperature increasingly. The open voltage of 773 K films was 68.2 mV, while that of 373 K films was 2.4 mV. What's more, those of current density are increased 0.03 mA/cm^2 and 325 mA/cm^2 , respectively. Increasing the substrate temperature shows growth films have a positive influence on the solar cell properties. Volt-ampere characteristics of films are as shown in?? In order to examine relationships between crystallization of growth films and solar cell properties, we measured the cross-section structure of those films by ESCA as shown in Figures. 4. The binding energies of C-C, Si-Si, and C-Si are 284.3, 99.3, and 100.4 eV, respectively [15-17]. The growth films were etched for 240 min (45 cycles) by Ar-ion sputtering. All films had C-C bond, Si-C bond and Si-Si bond. This means the growth films were consisted of DLC/SiC/Si layer. C-C bond of 373 K films was observed from 1 cycle to around 10 cycles, while that of 573 K and 773 K films were observed from 1 cycle to around 25 cycles. This result shows that the etching rate of 373 K films was higher than that of 573 K and 773 K films, because of $I(D)/I(G)$ ratio was different between 373 K films and over 473 K films. Si-C bond of 773 K films was observed from 19 cycles to 39 cycles, while that of 573 K films was from 18 cycles to 29 cycles. It should be noted that SiC layer of 773 K films was longer than that of 573 K. This result indicates that increasing the substrate temperature can control the thickness of SiC layer. This means that the amount of ions which can reach the substrate are increasing by applying to the electrical field, and that electrical field has a positive effect on synthesizing the SiC layer by attracting the carbon ions to Si substrate.

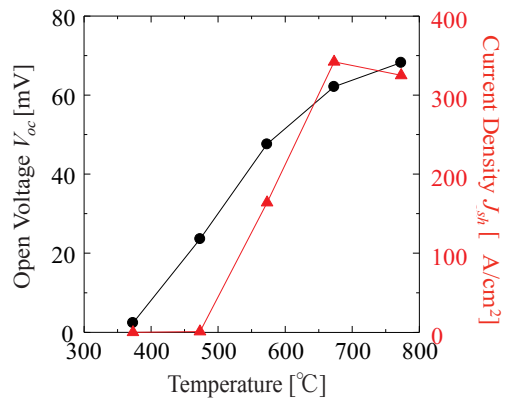
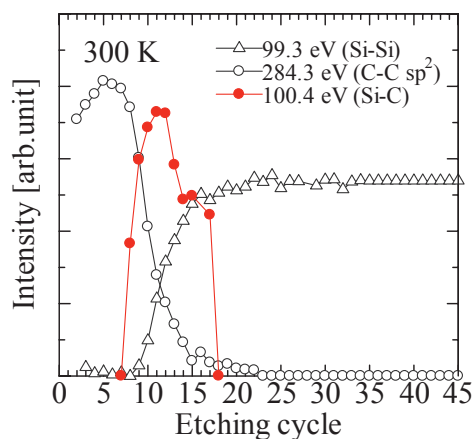
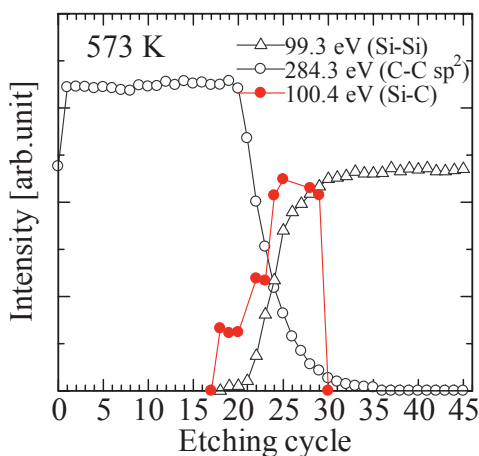


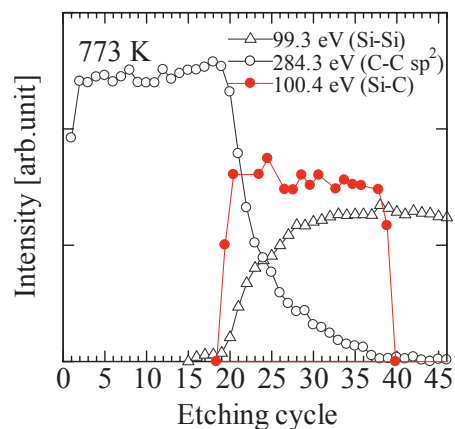
Figure 3. Relationships between temperature and open voltage, current density of films deposited by PLD at electrical field.



(a)



(b)



(c)

Figures 4. Cross-Section structure of DLC films by ESCA.

4 CONCLUSION

We report on the fabrication of n-DLC/i-SiC/p-Si photovoltaic cell by pulsed laser deposition at electrical field. Raman spectra showed that the $I(D)/I(G)$ ratio of n-DLC increased from 0.32 to 0.79 with increasing temperature. As well as, open voltage and current density were increased from 2.4 mV, 0.03 mA/cm² to 68.2 mV 325 mA/cm² with increasing temperature, respectively. The SiC layers were analyzed under irradiation by ESCA. Our PLD method could synthesize i-SiC layers between n-DLC and p-Si.

ACKNOWLEDGMENT

We would like to thank Takashi Kaito and Ken-ichi Matsuda for their helpful discussions and we also want to thank Yoshiaki Aoyagi and Ken Takayama for assisting this experiment. A part of this study was financially supported by the Ministry of Education, Science, Sports and Culture, a Grant-in-Aid for Scientific Research (C) 2008–2010, 20510098, and Nihon University Strategic Projects for Academic Research “Nihon University, Nanotechnology Excellence-Nanomaterial based Photonic and Quantum Technologies.”

REFERENCES

- [1] W. Shocley and H. J. Queisser, *J. Appl. Phys.*, Vol. 32 p. 510 (1961).
- [2] Y. Tawada, J. Takada, N. Fukada, M. Yamaguchi, H. Yamagishi, K. Nishimura, M. Kondo, Y. Hosokawa, K. Tsuge, T. Nakayama and I. Hatano, *Appl. Phys. Lett.*, Vol. 48, p. 584 (1986).
- [3] B. H. Lung, M. J. Chiang and M. H. Hon, *Chem. Phys.*, Vol. 72 p. 163 (2001).
- [4] N. Martin, L. Sanchez, M. A. Herranz, B. Illescas and D. M. Guldi, *Acc. Chem. Res.*, Vol. 40, p. ≈1015 (2007).
- [5] Y. Liu, A. Erdemir and E. I. Meletis, *Surf. Coat. Technol.*, Vol. 94–95, p. 463 (1997).
- [6] Ch. Schwarz, J. Heeg, M. Rosenberg and M. Wienecke, *Diamond Relat. Mater.*, Vol. 17, p. 1685 (2008).
- [7] W. Y. Yeh, J. Hwang, A. P. Lee, C. S. Kou and H. Chang, *Appl. Phys. Lett.*, Vol. 79, p. 3609 (2001).
- [8] Q. Zhang, S. F. Yoon, S. Zhgoon, V. Ligatchev, J. Ahn, Rusli and Z. Sun, *Thin Solid Films*, Vol. 397, p. 276 (2001).
- [9] A. Kongkanand, R. M. Dominguez and P. V. Kamat, *Nano Letter*, Vol. 7, p. 676 (2007).
- [10] E. Kymakis and G. A. J. Amaratunga, *Appl. Phys. Lett.*, Vol. 80, p. 112, (2002).
- [11] Y. Tawada, K. Tsuge, M. Kondo, H. Okamoto and Y. Hamakawa, *J. Appl. Phys.*, Vol. 53, p. 5273 (1982).

- [12] K. Takayama, S. Soma, T. Inoue, H. Kakinuma, T. Haraguchi and K. Suzuki, *Appl. Phys. A*, Vol. 101, p. 727 (2010).
- [13] A. C. Ferrari and J. Robertson, *Phys. Rev. B*, Vol. 61, p. 14095 (2000).
- [14] A. C. Ferrari and J. Robertson, *Phys. Rev. B*, Vol. 64, p. 075414 (2001).
- [15] J. C. Lee, et. al. *Appl. Phys. Lett.*, Vol. 84, p. 1305 (2004).
- [16] R. Fu, H. Zeng, Y. Lu, S. Y. Lai, W. H. Chang and C. F. Ng, *Carbon*, Vol. 33, p. 657 (1995).
- [17] Z.-W. Deng and R. Souda, *Diamond. Relat. Mater.*, Vol. 11, p. 1676 (2002).

Applying the technique of data analysis to estimate the electrical consumption for household in Taiwan

F.F. Chen

Taiwan Power Company Research Institute, Taipei, Taiwan

T.K. Lu

Department of Electrical Engineering, National Taiwan Ocean University, Keelung, Taiwan

ABSTRACT: The purpose of this research is to estimate the electrical consumption of household electrical appliances in Taiwan, in order to improve the efficiency of demand-side management (DSM). This research draws upon expertise from areas such as power utilities, field research, questionnaire surveys, and data quantitative analysis, to establish a base of estimating electrical consumption for decision support in DSM in electricity.

1 INTRODUCTION

1.1 The research background

In the past, residential and commercial customers in Taiwan utilized electricity, mostly for lighting purposes. However, due to the booming economy, the increase of national income, the improvement of living standards, and most importantly, the prevalence and a variety of household electrical appliances, there has been a significant change in the original electrical structure, and customers no longer use electricity merely for the purpose of lighting. To understand the impact of such changes on power consumption, this research carried out a survey on the prevalence and usage of household electrical appliances in Taiwan in order to further understand the prevalence of household electrical appliances and their usage rates in Taiwan.

1.2 The research objective

In this paper, the authors propose an energy informatics framework with quantitative analysis techniques, to deal with the huge raw data from survey efficiently and provide multi-dimensional inquiry of the energy consumption information.

The data quantitative analysis techniques are used to extract the profiles of the energy consumption from the questionnaire survey and to establish the bases for multi-dimensional inquiry of the energy consumption information. Then, the profiles of each household electrical applicant are analyzed and the analysis results are reported. The objective of the analysis is to provide the decision-makers with sufficient

information to develop the future power supply and support the demand-side management in electricity.

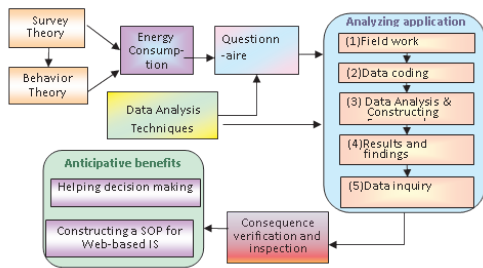


Figure 1. Research framework.

2 THE EMPIRICAL SOCIAL SURVEY MODEL

2.1 The research model

Based on the theory and the application's characteristics, the questionnaire survey and data analysis processes can be implemented effectively, as shown in Figure 1.

The field survey covers all provinces of Taiwan. The target population is shown as Table 1.

2.2 The sampling method

The sampling method is shown as the following. We totally have 6,070 samples in this survey, and the d value (the error level) is less than 1.5%.

Table 1. The Target Population of Survey unit: house.

| Type | Household | | | | | |
|------------------|-----------|---------|---------|-------------|-------------|-----------|
| Power (Kwh) Area | -249 | 250-599 | 600-999 | 1,000-1,499 | 1,500 above | Sub-total |
| Area1 | 116569 | 146532 | 36885 | 8164 | 1355 | 309505 |
| Area2 | 109544 | 217000 | 116056 | 42427 | 8236 | 493263 |
| Area3 | 183210 | 380094 | 140485 | 34723 | 5319 | 743831 |
| Area4 | 93487 | 187382 | 83745 | 26671 | 4925 | 396210 |
| Area5 | 153727 | 304337 | 117737 | 32287 | 5394 | 613482 |
| Area6 | 272811 | 370800 | 104496 | 26435 | 4946 | 779488 |
| Area7 | 121884 | 168051 | 53575 | 14796 | 2888 | 361194 |
| Area8 | 113117 | 111020 | 26079 | 6296 | 1069 | 257581 |
| Area9 | 443536 | 487727 | 124639 | 33184 | 6778 | 1095864 |
| Area10 | 148133 | 126143 | 25796 | 6432 | 1283 | 307787 |
| Area11 | 317817 | 247576 | 61941 | 15512 | 2928 | 645774 |
| Area12 | 259498 | 143366 | 29840 | 7387 | 1515 | 441606 |
| Area13 | 211613 | 165400 | 36095 | 8253 | 1519 | 422880 |
| Area14 | 104636 | 76892 | 15599 | 3699 | 775 | 201601 |
| Area15 | 310503 | 330436 | 83355 | 19610 | 3612 | 747516 |
| Area16 | 278510 | 345931 | 101439 | 23474 | 4050 | 753404 |
| Area17 | 207336 | 256256 | 73026 | 15232 | 2384 | 554234 |
| Area18 | 205171 | 185397 | 48926 | 10858 | 1933 | 452285 |
| Area19 | 46209 | 37169 | 9606 | 2515 | 438 | 95937 |
| Area20 | 60849 | 58957 | 16701 | 4575 | 918 | 142000 |
| Area21 | 97579 | 89852 | 18146 | 4103 | 778 | 210458 |
| Area22 | 13936 | 14880 | 4702 | 1075 | 160 | 34753 |
| Area23 | 5548 | 5719 | 2334 | 736 | 137 | 14474 |
| Area24 | 1262 | 956 | 359 | 105 | 19 | 2701 |
| Total | 3849019 | 4467877 | 1343977 | 352871 | 64087 | 10077828 |

$$d = Z_{\frac{\delta}{2}} \times \sqrt{\sum_{i=1}^6 w_{i^2} \left(\frac{N_{i^2} - n_i}{N_i - 1} \right) \frac{1}{4n_i}}$$

2.3 The equation of household electrical appliances popularization rate

i : 1=residential, 2=commercial

j : 1,2,.....,24, areas

k : 1,2,.....,5, levels

3 RESULTS

Some energy consumption survey studies from different countries indicate that residential energy consumption accounts for approximately 20 ~ 40 percent of total energy consumption. In Taiwan, it accounts for approximately 20%. In Taiwan the majority of residential energy consumption is power-related; moreover, the key factor that influences residential energy consumption is the use of household electrical appliances.

According to the quantitative investigation of “2011 Household Appliances Availability Survey”, after quantifying the results of our survey as shown in Table 2, we found that electric fans and refrigerators have the highest rate of availability, followed by televisions, mobile phone chargers, air conditioners, rice cookers/electric cookers, washers, range hoods,

computers and fluorescent lamps. In regard to the electricity consumption in summertime, air conditioners rank as highest, followed by lighting facilities, refrigerators and computers. Concerning the electricity consumption in non-summertime, lighting facilities come first, followed by refrigerators, computers and water heaters. Hence it is obvious that air conditioners and lighting facilities are both major sources of electricity consumption.

In Taiwan, which is situated at lower latitudes, the low usage rate of room heaters is different from that found in Europe and the USA.

In addition, plasma/LCD television sets have started to replace traditional TV sets. The increased

or decreased popularization of certain electrical appliances can have a significant effect. It can be observed that once a revolutionary product emerges, the replacement of older generation electrical products will follow. This phenomenon could serve as a worthwhile foundation for promoting energy-saving electrical appliances in the future.

In respect of lighting facilities, the availability of fluorescent lamps has seen a decrease, whereas compact fluorescent lamps and LED items have become more popular in recent years. As for the willingness to purchase among consumers the willingness to buy LED lamps is increasing surprisingly, so are frequency-convertible products such as refrigerators and air conditioners, which indicate energy-saving household appliances are getting momentum among local consumers.

Table 2. Household Electrical Appliances Popularization Rate 2011.

| Appliance | Popularization Rate (%) |
|------------------------|-------------------------|
| Electric fan | 95.9 |
| Refrigerator | 91.8 |
| Television | 89.4 |
| Mobile Phone Charger | 87.6 |
| Air conditioner | 84.5 |
| Electric pot | 80.1 |
| Washing machine | 78.5 |
| Range hood | 78.4 |
| Computer | 76.5 |
| Fluorescent light | 72.4 |
| Energy saving lamp | 58.5 |
| Traditional TV sets | 57.4 |
| Incandescent lamp | 51.9 |
| VCD/DVD player | 49.5 |
| Drinking fountain | 49.4 |
| Plasma/LCDTVsets | 45.6 |
| Blender/Juicer | 43.6 |
| Electric thermos | 43.4 |
| Audio devices | 43.1 |
| Microwave oven | 41.3 |
| Plug-in Phone | 40.2 |
| Electric oven | 38.9 |
| Vacuum cleaner | 32.8 |
| Induction oven | 32.8 |
| Pump | 26.5 |
| Printer | 25.3 |
| Electric boiler | 21.6 |
| Power-driven door | 21.5 |
| Electric radiator | 20.4 |
| Dehumidifier | 20.3 |
| Mosquito zapper | 17.5 |
| Dish Dryer | 14.8 |
| Air purifier | 10.1 |
| All-in-one printer | 8 |
| LED light | 6.8 |
| Security Monitor | 5.5 |
| Electric massage chair | 4.3 |
| PL Fluorescent light | 3.6 |
| Halogen light | 2.9 |
| Treadmill | 2.2 |
| HID light | 0.5 |

4 CONCLUSIONS

Over the past years, due to the socioeconomic and climate changes and the increase of public consciousness on energy-saving, apart from the rapidly developing trend of newly invented appliances, how the consumption of electricity will likely change is an issue worth attention and concerns. Here with the objective of the research lies in trying to comprehend the availability of all kinds of major household appliances and accurately comprehend the electricity consumption of domestic residential sectors, which may serve as a reference for policy planning in future.

REFERENCES

- [1] Al-Grandeur, A, et al. 2008, "Residential past and future energy consumption: Potential savings and environmental impact," *Renewable and Sustainable Energy Reviews*.
- [2] Chun-Lang Hsu, Sheng-Yuan Yang b, Wei-bin Wua, "3C intelligent home appliance control system – Example with refrigerator." *Expert Systems with Applications* 37 (2010) 4337–4349.
- [3] Marsh, Rob; Larsen, Vibeke Grupe; Kragh, Mikkel; "Housing and energy in Denmark: past, present, and future challenges." *Building Research & Information*, Jan/Feb 2010, Vol. 38 Issue 1, p92–106.
- [4] MV Shuma-Iwisi, GJ Gibbon, "Appliance Standby Power and Energy Consumption in South African Households." *The CPWT Active Web (CAW) Unit, Domestic Use of Energy(DUE) annual international conference, 2009.*
- [5] Paula Fonseca, Anibal de Almeida, Nicolai Feilberg, George Markogiannakis, Casper Kofod, "Characterization of the household electricity consumption in the EU, potential energy savings and specific policy recommendations." *European Council for an Energy Efficiency Economy, 2009.*

- [6] Willington O. Onuh, Alice T. Valerio, Ananias Permalino, Jr., 2011, "Residential Demand for Electricity Dasmariñas, Cavite, Philippines," *Journal of Global Business & Economics*, Jan2011, Vol. 2 Issue 1, p1–22.
- [7] Z.X. Cheong, N.N. Barsoum, "An Investigation on the Energy Saving Potential of Eelctromagnetic Ballast Fluorescent Lamps." *AIP Conference Proceedings*, Volume 1159, pp.155–160 (2009).

Design of electric vehicle charging battery swap station power quality data transmission system based on lab view

D.S. Bai, G. Chen, X.Q. Li, J. Liu & W.F. Li

Beijing PuriHigh-voltage Transmission Technology Co., LTD of State Grid, Changing, Beijing, China

H. Xing

China Electric Power Research Institute, Haidian, Beijing, China

ABSTRACT: In this paper, computer technology, virtual instrument technology and network communication technology, on the basis of a unified monitoring platform based on Ethernet networking and hierarchies through integrated data monitoring, communications, servers and other application subsystems in one, integrated monitor, communication server and the central station server, to design a filling station network of power quality data transmission system is suitable for electric vehicles.

1 INTRODUCTION

Currently, the shortage of fossil energy supply, deteriorating ecological environment of the world economy, has become a significant threat [1]. Development of energy saving electric cars is an effective way for implementation of energy strategy and the sustainable use of low-carbon economy [2–3].

One of the core technology of electric vehicles is the energy reserves and supply system, electricity use technology to become the key for the development of the electric car industry [2,4]. Electric vehicle charging facilities required grid power quality. Meanwhile, a large number of electric vehicle charging for electric facilities run will have a great impact on power quality, while the electric car charge for electricity with randomness and disorder [4], resulting in electric vehicle charging for electricity facilities presented volatility, further affecting the diversity of electric power quality [3,5].

There are a variety of power quality monitoring methods, such as a power quality monitoring system based on Internet-based virtual instrument power quality monitoring system, application PcAnywhere power quality monitoring systems, DSP-based power quality monitoring system [7-8], with three-tier C / S structure to achieve power quality database management systems [3,9], the network power quality management systems, distributed power quality monitoring system. However, these indicator monitoring system can monitor less common poor, lower accuracy, low degree of automation, standardization,

serialization, modular, multi-functional, integrated, intelligent level is not high [7–8,10].

In this paper, computer technology, virtual instrument technology and communication technology can be used, based on a unified monitoring platform, Via Ethernet networking and hierarchies, integrated data monitoring, communications, servers and other application subsystems in one, integrated monitors, and central station server communication server, to design a suitable electric vehicle filling station network of power quality data transmission system.

2 THE PRINCIPLE AND STRUCTURE OF POWER QUALITY DATA TRANSMISSION SYSTEM

2.1 *National power quality standard system indicators*

The IEEE Standardization Coordinating Committee gives the corresponding power quality technical meanings: passing power quality concept refers to the sensitive equipment to provide power and a grounding system is suitable for the device to work normally (in many cases, the grounding system of the power quality has a great influence).

Power quality national standard indicators are: utility grid harmonics, voltage flicker and voltage fluctuation allowed, three-phase voltage unbalance factor, supply voltage tolerance, the power system frequency tolerance.

2.2 Level of power quality data transmission system

Hierarchical distributed power quality data transmission system can be divided into monitoring layer, communication layer, the management structure three levels. Monitoring layer is the underlying data collection terminals, containing one or more independent PQM series power quality monitoring devices, collect, measure and analyze the relevant line of electrical quantities. Communication layer is a data transmission channel for data transmission based on Internet, PSTN, ISDN, ADSL, optical fiber transmission. Energy management is a central quality of the data transmission system for communication and management of power quality monitoring equipment via fiber optic network or cable communication networks and substations, collect and store data for each monitoring point, and support multiple users simultaneously. In addition, management can also send a message through a variety of alarms or internal systems.

2.3 Structure of power quality data transmission system

Power quality data transmission system based on computer technology, the virtual instrument technology and network communication technology, connects to the grid of each monitoring site, in order to achieve power quality monitoring of networking. Most substations have access to a local area network, and LAN traffic can guarantee real-time data transmission, reliability, power quality monitoring system can therefore take advantage of the existing LAN to set up a data transmission system.

The data transmission system of the power quality system components can be divided into the data monitoring subsystem, communication subsystem, server subsystem three parts, the system is shown in Figure 1.

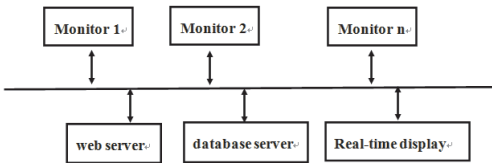


Figure 1. Power quality data transmission system diagram.

The core part of the data monitoring subsystem is power quality monitoring instrument which is based on virtual instrument technology, computer analysis, processing, display, store data, and equipped conditioning circuits and monitoring software to obtain the

measurement signal. The communication subsystem supports Ethernet interface, modem interface and serial port communication, responsible for uploading real-time data, commands issued information management center. The server subsystem includes a Web server and database server, providing real-time database and historical database to provide an object-oriented interface to provide query access to the historical data platform that provides information maintenance functions, provide user management and permissions management tools, and provide report viewing and records check.

3 ETHERNET NETWORK DESIGN POWER QUALITY DATA TRANSMISSION SYSTEM

3.1 Power quality data transmitted by the transmission system

There are three data in power quality data transmission system: real-time data, order data, monitoring data. Real-time data include real-time display data and alarm data, which is collected by monitor power quality analysis software acquisition server and sent to the central station. Upon receipt of the start control commands to transfer data line monitor data acquisition and control software, according to the reference count to decide whether to forward the data forwarded to the central station server via a communications server. Command data from the central station server via the communication server forwards to monitor data acquisition and control software and perform. The data monitoring power quality monitor data that is collected, in the specified time interval to the data acquisition and acquisition and control software, and then forwarded to the central station server via a communications server, after processing database statistical analysis program stored for future queries, display.

3.2 System architecture of power quality data transmission

Ethernet-based networking process-oriented design, the communication subsystem structure consists of three parts, monitors, central stations and the communication server, the communication protocol using TCP / IP protocol. The structure shown in Figure 2.



Figure 2. Ethernet networking power quality data transmission system architecture.

Power quality monitor, including data collection, data processing section, data communications

transmission part, can monitor, analyze and specify line voltage and current signals, the data types are integer and character. The monitor is installed in the power quality monitoring and analysis software for high-speed acquisition of three-phase voltage and current signals, analog to digital conversion, and analyze information and to obtain a series of steady-state transient information. The communication server receives power quality monitor real-time data upload and monitoring data, monitoring data stored in the database, real-time data will be forwarded to the central station server, and forwards the command information. Division manner using different types of data transmission channels to communicate the flow of data. The central station server is mainly used to store and manage the various monitoring points, upload monitoring data, database statistical analysis of data processed. Central station server database into real-time database and real-time database, system data can be simultaneously written to two or more databases for easy expansion. From the central station server functionality for database servers and Web servers, which offers a central database, data statistical analysis software monitoring data after processing, historical database for long-term storage of data processed.

3.3 Power quality data transmission system communication model based on Ethernet networking

The way inter-process communication are the mail slot, news pipes, window messages, dynamic data exchange, messaging pipeline, Windows Sockets, Internet communications, serial and parallel communication, we mainly use the CAayncSocket communication technology, mail slot technology and message pipes technology, the communication model is shown in Fig.3.

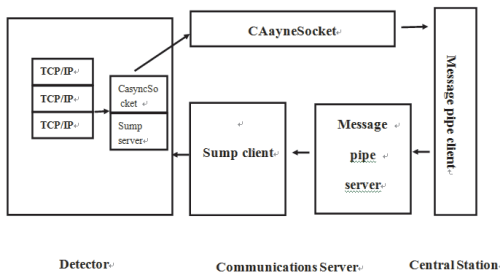


Figure 3. Ethernet networking power quality data transmission communication model diagram.

Where in the communication to communicate with each CAayncSocket communication technology allows two or more applications, or between the same system between applications, where the software

for data acquisition and control and communication server to transmit data. In the communication server uses server-side program, receiving a plurality of data sent by the client and in the end use of the monitor client program to send data. Mail slot technology for the client to the server from a one-way transfer of data, can facilitate the achievement of many of the small amount of data communications. Message pipes shared memory for inter-process communication, which is used for one-way or two-way communication between a pipe server and pipe client communication modes are byte mode and message mode.

3.4 Advantages of Ethernet networking power quality data transmission system

The system has to provide a monitoring, interactive interface, real-time data acquisition and storage, operation monitoring and alarm, running records, operation control, screen display and tabulation printing, communications, self-diagnosis, self-healing capabilities, user-friendly, full-featured, expansion convenient, highly intelligent, standardized, comprehensive, remote, real-time, intuitive and effective monitoring of power quality.

The biggest advantage of mail slot is that you can send broadcast messages to one or more servers for transmitting and receiving short messages in the LAN environment. Named pipes have a good flexibility of use, and its main advantages are: both for the local network can also be used; may be applied by name; supports multiple client connections; supports bidirectional communications.

4 CONCLUSION

Global energy and environmental problems have become increasingly severe, it is very necessary to develop electric vehicles. With the gradual promotion of electric vehicles, which charge for electricity facilities run impact on power quality cannot be ignored.

In this paper, the use of computer technology, virtual instrument technology and communication technology, based on a unified monitoring platform, via Ethernet networking and hierarchies, integrated data monitoring, communications, servers and other application subsystems in one, integrated monitors, and central station server communication server, to design a network of power applied to the electric vehicle charging station quality of the data transmission system. The system realizes the same power system, power quality monitoring in different locations, and even centralized monitoring of multiple different power supply systems. Meanwhile, in function, more emphasis on intelligence, in addition to a calculation, display, and other functions, but it also has some judgment, analysis, decision-making

and other functions, such as to conduct the event prediction, fault identification, interference source identification and real-time control, preliminary having the utility automatically advanced computational intelligence assessment function. The system is conducive to the large-scale promotion of electric vehicles with significant economic and social benefits.

ACKNOWLEDGMENTS

This work was financially supported by The National High Technology Research and Development of China 863 Program (2011AA05A109), National Energy Application Technology Research and Engineering demonstration project (NY20110703-1), SGCC project “Research on Test and Measurement Technology of charging and swapping infrastructure for Electric vehicles”.

REFERENCES

- [1] Zhang Wenliang, Wu Bin, Li Wufeng, to the well-off and explore the development direction of China's energy supply mode of pure electric vehicles [J] Power System Technology, 2009,33 (3): 1–5.
- [2] Chen Liangliang, Zhang Hao, Ni Feng and other electric vehicle energy supply facilities construction and development status quo [J] Automation of Electric Power Systems, 2011,35 (14): 11–17.
- [3] Tzoo-Hseng S. Li, Shih-Jie Chang, Yi-Xiang Chen, etc. Implementation of human-like driving skills by autonomous fuzzy behavior control on an FPGA-based car-like mobile robot using Infrared Sensors [C]. IEEE Transactions on Industrial Electronics, 2003: 145–164.
- [4] Xietian Hui, Li Wufeng, Bode Sheng, Cui Yu, Stereotype .GIS in electric vehicle charging facilities for electric running applications [J] Chinese CSEE, 2012,32: 89–92.
- [5] Sortomme E, El-Sharkawi M A. Optimal Charging Strategies for Unidirectional Vehicle-to-Grid [J] Smart Grid, IEEE Transactions on, 2011,2 (1):.131–138.
- [6] Guimarin DC, Janik W M. Battery exchange system for electric vehicles: US Patent 5,612,606 [P] 1997: 3–18.
- [7] Hanshao Fu Du tree new design and realization of power quality monitoring system [J] Electric Power Automation Equipment83.-, 2006, 26 (4):. 80.
- [8] Xiangning Xu Yonghai power quality analysis grid technology, 2001,3 (7): 66–69.
- [9] MOTTO K, LI Y, HUANG A Q. Comparison of the State-of-the-art High Power IGBTs, GCTs and ETOs [A]. New Orleans (LA, USA), 2000.1129–1136.
- [10] Wang Bin synchronous processing and plant design power quality monitoring data, power system automation, 2002,26 (11): 45–49.

Research on intelligent integrated charging equipment for electric vehicle

X.Q. Li

State Grid Electric Power Research Institute of Beijing SG-EPRI UHV Transmission Technology Company, Ltd, Beijing, China

H.L. Wang

China Electric Power Research Institute. Beijing, China

S.Y. Li & K.X. Li

State Grid Electric Power Research Institute of Beijing SG-EPRI UHV Transmission Technology Company, Ltd, Beijing, China

ABSTRACT: The power supply mode of electric vehicles depends on the electric vehicle charging scenarios and charging requirements. The research of combing multiple power supplying mode charging device has a wide range of adaptability. According to the features of conventional slow charging and DC fast charging for power supply, we researched the combination of AC and DC charging equipment integrated electrical topology. Considering the safe operation of equipment and environmental adaptability, we proposed an intelligent integration of charging equipment for security program and environmental adaptation program. Considering the convenience of current payment, it is based on the multiple payment modes payment scheme, such as payment card, non-contact card and online payment. We have tested the device output characteristics, safety features, environmental characteristics, and the payment process, equipment program can work as AC, DC simultaneously and meet the needs of the equipment for long-term operation.

1 INTRODUCTION

The development of electric vehicles is an important way to solve the energy shortage and environmental crises. And it depends on the construction and development of the charging facilities. The power supply mode of electric vehicle includes three ways, which are AC charging, DC charging and battery swap. [Table 1] compares the advantages and disadvantages of the three current power supply modes. By contrast, we can find that in the table, AC charging or DC charging has its own advantages and applications. Achieving AC and DC power supply complementary mode has aroused wide concern. The research and development of an AC-DC combined charging device has both conventional charging and emergency charging modes. The applications of it become increasingly prominent.

The research on the intelligent integrated device of electric vehicle includes the design of the hardware system, security protection, software system and human-computer interaction system. The design of hardware system includes an AC charging system, DC charging system, the information gathering system and security protection hardware circuit. The design of human-computer interaction

system includes a display interface, the charging indicator system and the card reader system. The design of safety protection includes the structure of equipment, threshold limit alarm. The design of software system includes card reader process, charging process and equipment operation process. This article introduces each subsystem design process and scheme in detail.

Table 1. Three kinds of power supply mode contrast.

| Supplement patterns | Mode advantages | Mode disadvantages |
|---------------------|--|--|
| AC charging | Small current charge, extending battery life, lowering installation costs. | Long charge time, the charge can not meet the emergency needs |
| DC charging | Big charging current, short charging time | High installation costs, centralized charging has a big impact on the grid |
| Battery swap | High efficiency power supplement, unified charging management | Large area, charging interface standard is hard to uniform |

2 MAIN CONTENTS

2.1 Hardware system design

The hardware system of the integrated intelligent charging device includes AC system, DC system, metering and billing system, card reader system, integrated monitoring and control system and the main control system. The core of hardware system is master controller; it can be connected to the other hardware devices through different data communication. The block diagram of overall equipment, hardware is shown in Figure 1.

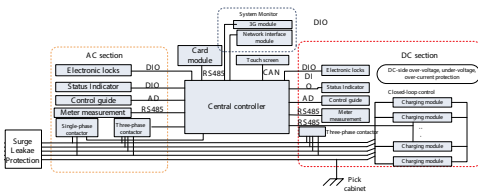


Figure 1. The block diagram of overall equipment hardware.

The selection of hardware system design mainly involves three main parts: the main circuit device, the central controller, integrated monitoring systems. Other parts of the AC system are mainly based on the state control of central controller and data exchange. The DC charging system is mainly based on the paralleled charging module and the measurement and control of the input and output information.

2.1.1 Selection of the main circuit device

The main circuit of intelligent integration charging equipment includes breaker, electric meter, surge protector, leakage protector. The selection of equipment considers different operating modes such as working status, fault status, which is to ensure the safety of equipment and personal safety.

The selection of AC breaker considers a rated operating voltage, rated operating current, maximum braking capacity. The main selection accords to formula, (1), (2).

$$I_d \geq I_{KN} \geq k_i I_N \quad (1)$$

$$U_{KN} \geq k_u U_N \quad (2)$$

Where, U_{KN} is the rated operating voltage of the switch and I_{KN} is the rated operating current of the switch. U_N and I_N are the rated operating voltage and current of the line. I_d is the minimum short circuit current of the line. k_i is the margin coefficient of the current,

and generally select 1.2~1.5. k_u is the margin coefficient of the voltage, and generally select 1.05~1.15.

Selection of meter, mainly considers rated capacity, rated voltage, rated current, the main selection accords to formula, (3), and (4).

$$\frac{I_n}{k_{b \max}} \leq I_b \leq \frac{I_n}{k_{b \min}} \quad (3)$$

$$P_{NMAX} = \sqrt{3} U_N I_b \quad (4)$$

Where, I_n is the rated current of load, and I_b is the rated maximum current of the meter. P_{NMAX} is the maximum operating power of the lines. U_N is rated operating voltage of the meter. According to the calculation of meter operating current, we can determine the installation of the meter. If direct access does not meet the measurement requirements, we need to use a transformer.

The surge protector mainly protects equipment from lightning damage, considering an impulse withstand voltage of the device, grounding system type and the maximum operating voltage of the grid. And its operating voltage of the surge protectors should operate between them.

According to the national standard, the leakage protector can choose the 30mA acting threshold leakage protector.

2.1.2 Central controller design

A central controller is the core unit of the integrated charging device, which is responsible for collecting the switch variables, sub-module states and data. The hardware of the controller includes a CAN communication unit, RS485 communication unit, RS232 communication unit, AD acquisition unit, I/O interfaces, PWM communication interface and so on, through which we can achieve the unified control and management of the entire device.

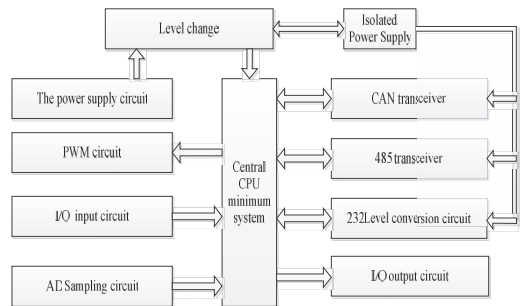


Figure 2. The block diagram of central controller design system.

2.2 Software system design

The software system is the most important part of the integrated charging device, through which we can control the overall operation of the system. The software system has two main parts that are the main flowchart of equipment operation and system operation. The equipment operation main flowchart is shown in Figure 3.

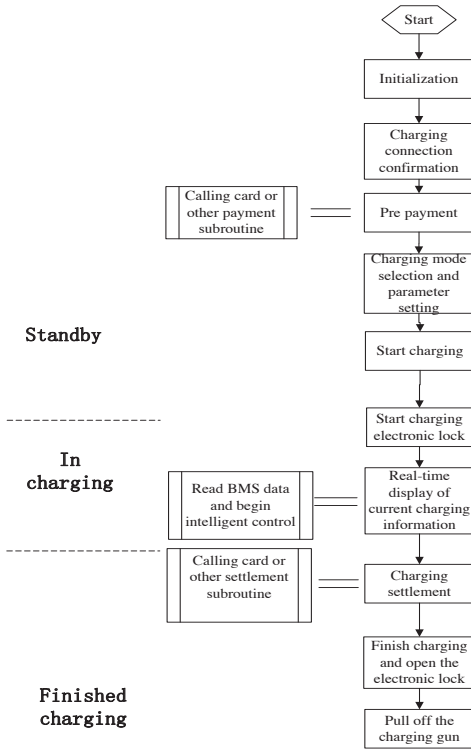


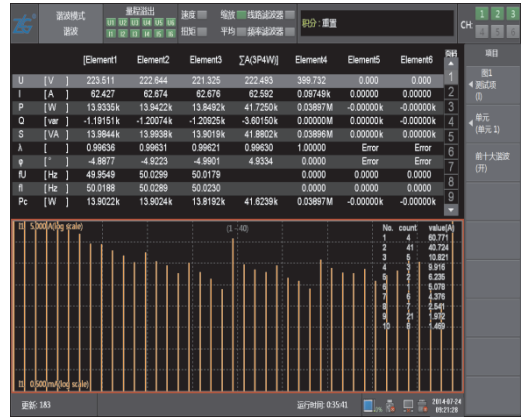
Figure 3. Main flow chart of the intelligent integration equipment.

2.3 Experimental tests

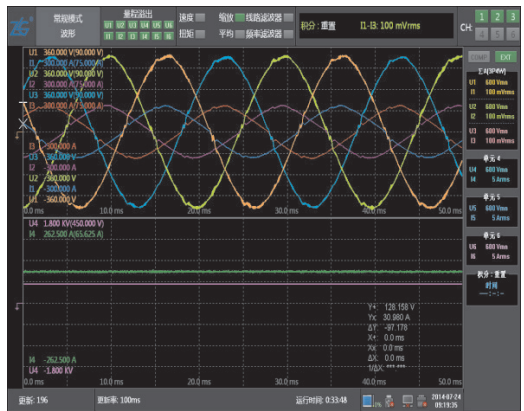
Based on the above design scheme, we developed an AC output 380V/32A, DC output 400V/150A integrated device and conducted some performance tests such as AC input and output, DC output. The testing results are shown in Figure 4.

3 CONCLUSION

For the purpose of the electric vehicles, power supplement demanding, we developed the design scheme from hardware, software, structural design and other aspects. Then we made a prototype whose indicators



(a) Numerical indicators



(b) DC output and AC input waveform

Figure 4. Device function detection.

meet the practical application needs, taking the coordinated charge control into account, which has a very significant impact for the future promotion of electric vehicle charging facilities.

ACKNOWLEDGMENTS

This work was financially supported by The National High Technology Research and Development of China (863 Program) (2011AA05A109) and National Energy Application Technology Research and Engineering demonstration project (NY20110705-1).

REFERENCES

- [1] Murat Yimaz, (2012). Review of Charging Power Levels and Infrastructure for Plug-In Electric and Hybrid Vehicle Conference (IEVC), IEEE International.

- [2] Melike Erol-kantaric, Janhangir H. sarker. Hussein T. Mouftan , (2012), Quality of Service in Plug-In Electric Vehicle Charging Infrastructure[C].Electric Vehicle Conference (IEVC), IEEE International.
- [3] Xueliang Ren, Juntian He,Zhizhou Sun. Design and Implementation of the Integration of Electric Vehicle Charger [J]. Information Technology and Information 2013,2(2):88–94.
- [4] Shuai Han,Chunmei Lv,Miao An,Zhenjun Zhu. Research and Design of Electric Vehicle Charging Terminal Communication 2011,8(1):18–26.
- [5] Tao Wang, Donghua Zhang, Research and Design of Control System of Electric Vehicle Charging Pile [J]. HuBei Electric Power,2011,(1):31–37.
- [6] Xiaoyan Yan. Application of REIP Prepaid Payment System Based on Charging Station [J] Electric Power Information: 2012,8(7):10–13.
- [7] Bofei Gu,Junjie Zhao,Kaihua Wu, Design of Single-chip Microcomputer Touch Screen Based on Modbus Protocol Communication [J]. Mechanical and Electrical Engineering, 2012,29(1):104–107.
- [8] Guo Wei, Huijie Gao,Chunbo Zhu, (2012) ,Pile Charging System Design for AC Motor [J].Electronic Design Engineering.
- [9] Xiaoyan Fan,Libo Ding, (2013) ,Pile Design of Charging Electric Vehicle Communication Based on Embedded System [J].Electronic Design Engineering.
- [10] SAE J1772, (2010), Electric Vehicle and Plug in Hybrid Electric Vehicle Conductive Charge Coupler[S]. American: Society of Automobile Engineers.
- [11] Liangliang Chen. Status and Development of Energy Supply Facilities Building Electric Cars [J], 2011, 35(14):11–17.
- [12] Robert Bruninga ,(2010).Overlooking L1Charging At-Work in the Rush for Public Charging Speed[C]. Electric Vehicle Conference (IEVC), IEEE International.
- [13] IEC 62196-2, (2011), Plugs, Socket-outlets, Vehicle Couplers and Vehicle Inlets-Conductive Charging of Electric Vehicles, Part 2: Dimensional Interchangeability Requirements for Pin and Contact-tube Accessories [S].Switzerland International Electric technical Commission.
- [14] Steven L, (2007). Plug-in Hybrid Electric Vehicles and the Vermont Grid: a Scoping Analysis [R]. University of Vermont Transportation Center.
- [15] Parks K, Denholm P, Markel T, (2007). Costs and Emissions Associated with Plug-in Hybrid Electric Vehicle Charging in the Xcel Energy Colorado Service Territory [R], National Renewable Energy Laboratory.

Design and development of driver protection circuit of self turn off devices of electric vehicle DC charger

D.S. Bai, X.Q. Li, G. Chen, J. Liu & W.F. Li

Beijing Puri High-voltage Transmission Technology Co., LTD of State Grid, Changping, Beijing, China

H.L. Wang

China Electric Power Research Institute. Beijing, China

ABSTRACT: In this paper, we design and develop a driver protection circuit of self turn off devices of electric vehicle DC charger based on the chip UAA4002. Take power transistor, for example, we study and design the topological structures of the Optocoupler isolation circuit, control circuit, anti-saturation circuit, composite buffer circuit. Then we integrate these circuits and obtain the formation of self turn off device drivers, isolation and protection circuit. The circuit has a simple circuit structure, perfect function, high reliability, low switching losses and fast response.

1 INTRODUCTION

Currently, the social and economic development of the world is facing a shortage of fossil energy supply, environmental deterioration and other challenges, and energy and environmental issues have become national major strategic issues needed to be resolved [1]. Electric vehicles have the advantages of energy saving and environmental protection [2–3]. Therefore, many countries vigorously develop electric vehicles and efforts to achieve a transformation of transport energy.

With the strategic development of electric vehicles and electric vehicles making up the current electricity demand, electric vehicles with fast charging will become the main way of energy supply [4]. Therefore, the development of charging machines in electric vehicle with fast charging mode and other infrastructure is a key route for optimizing EV charging for electric resource allocation, and it is the basic link to promote the development of the electric car industry.

For changes in electric vehicle charging mode, the charger used for power transistors and other self-turn-off device reliability, higher requirements of response speed [3, 5], as well as enhancement of the efficiency of the charging machine, you need to reduce the self-turn-off switching losses of the device and simplify the structure of their driver protection circuit as far as possible.

The role of high-power transistors and other self shut down the device driver protection circuit is the control circuit outputs a control signal current amplification enough to ensure reliable self-turn-off

devices on and off. [1] Since the device is turned off the base drive that can directly affect their working conditions, certain characteristics can be improved or hurt parameters and driver protection circuit should be matched with the main circuit to ensure that self-turn-off device is turned on at any time during the quasi-saturation state. [2]. Plus off quickly when the reverse bias voltage is large enough, and have enough protection.

Since the turn-off devices are mostly used baseband reverse bias tape drive circuit or overload, short-circuit protection of the base drive circuit topology [4]. Baseband reverse bias gate drive circuit and potential compensation diode clamp diodes can self-turn-off device is always turned on in the critical saturation [3, 6]. Zener diode reverse biased and self-turn-off devices can cut off acceleration. However, Zener voltage is more difficult to determine, and low reverse bias effect is not obvious. Simultaneously, high power components will be damaged, and may generate high-frequency parasitic oscillation [4, 7]. Since the shutdown of the heat capacity of the device is generally small. And over current capacity is very low. When it is running on the inverter system, because of its longer off-time, it easily leads to short-circuit fault arm [8].

In this paper, based on chip UAA4002, it designed a suitable mode for electric vehicle with fast charge of self-protection circuit that can shut down the device driver, with a simple circuit structure, function, high reliability, low switching losses, fast response and so on.

2 UAA4002 WORKS

2.1 UAA4002 Introduction

UAA4002 chip is a power switching transistor that can achieve the optimum base drive and decentralized situ conservation standard 16-pin plastic DIP LSI.

2.2 UAA4002 Features

2.2.1 The use of integrated high-speed internal logic processor

Internal use was integrated with high-speed logic processor protection switching transistor. The processor monitors the collector current, collector saturation voltage emitter saturation voltage, and positive and negative power supply voltage of the chip temperature and manifold work.

2.2.2 Set by the user on-time extremes

Set by the user on-time extremes, fault information is stored until it is turned over, to avoid repeated switching.

2.2.3 Base current with adaptive link

Base current, which change from logic signal, includes adaptive link, [3] to ensure the switching transistor operating in the critical saturation.

2.2.4 The output current can be adjusted

The output current can be adjusted for maximum output current chip 0.5A, and can be adjusted by adding external transistors.

2.2.5 Turn-off may be applied to the base of the reverse base current

The shutdown can be applied to the base of the reverse bias current, to increase the turn-off of the effectiveness and real-time and to reduce the switching losses.

3 DRIVER PROTECTION CIRCUIT DESIGN

3.1 Circuit overall structural design

In this paper, we designed a high-power transistors, including the driver, isolation, protection circuit. [4] As shown in Figure 1, driver protection circuit includes positive and negative power supply, opto-isolation circuit, control circuit, anti-saturation circuits, power transistors, and composite buffer circuit.

After the positive input control signal by the RC filter circuit optocoupler original light emitted diode, [5] light emitting diodes emit negative ground via a resistor, and transistor optocoupler deputy pole circuit is directly connected to the control input of the chip UAA4002 and via a resistor and a switch, it is

connected to the negative supply and the positive supply via a switch connected phototransistor collector. The positive and negative power supply are connected to the positive and negative supply pin UAA4002 chip, and are connected to two follow-up and down on the tube collector, and down the road for two tubes were treated with RC filter and then through a resistor connected to the anti-saturation circuit current limiting, and anti-saturation circuit is in parallel between the base and collector of the power transistor, and the control circuit chip UAA4002 were passed down resistor connected and subsequent RC filter two of the base pipe. The setting of its internal end and the external circuit via the transmitting power transistors are commonly grounded, and the negative power source is connected through the switch and the power transistor and the emitter resistor, and a composite buffer circuit with high-power is directly in parallel with the collector and emitter of the transistor.[6]

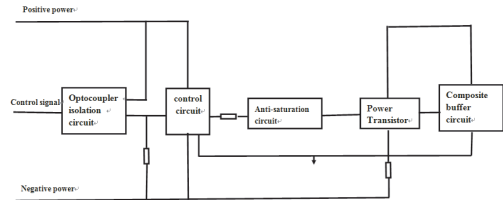


Figure 1. Schematic view of the overall structure of the circuit.

3.2 Optocoupler isolation circuit design

The opto-isolator circuit includes RC filter links, optocoupler, resistor. RC filter using resistors and capacitors connected in parallel, and then connected to the positive of light-emitting diodes, which is optocoupler original parties. The negative electrode of the light emitting diode connected to ground through a resistor, the collector of the phototransistor, which Optocoupler deputy party, is connected to the positive supply through the switch. Negative power connected with emitter via a resistor and a switch, and connected Follow-up control circuit directly.

3.3 Control circuit design

The control circuit includes chip UAA4002 its peripheral circuits, and down on the tube and two additional circuits.

3.4 Anti-saturation circuit design

Anti-saturation circuit Baker clamp circuit topology, directly connected in parallel between the base and collector of the power transistor, is connected through

a current limiting resistor in the control circuit which comprises three branches. The first branch comprises a clamping diode, the diode with fast recovery diodes, connected limiting resistor and power transistor collector; the second branch comprises two series of potential compensation diode, which are used in fast diode connected limiting resistor and large base power transistor; third branch includes an anti-parallel diode and a direct parallel with the second leg.

3.5 Composite buffer circuit design

The composite buffer circuit includes both the opening and off snubber circuit, directly paralleled between the collector and emitter of the power transistor, which includes an inductor, a resistor, capacitor, diode.

Series inductance power transistor collector, followed by a branch connected in parallel, the branch in which resistor and capacitor in series. A diode between the inductor branch and resistor branch, the diodes and inductors, resistors make an opened absorption circuit, while it makes a constitute off snubber circuit with a resistor and capacitor. [7] [8]

3.6 Schematic circuit implementation

In summary, the design of Opto-isolation circuit, control circuit, anti-saturation circuit, a composite buffer circuit that has great significance for the EV fast charge mode of self-turn-off device driver protection circuit. Schematically shown in Figure 2.

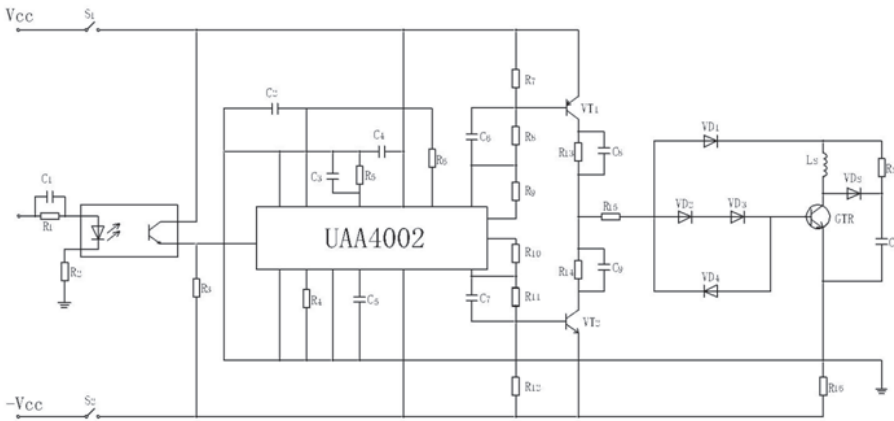
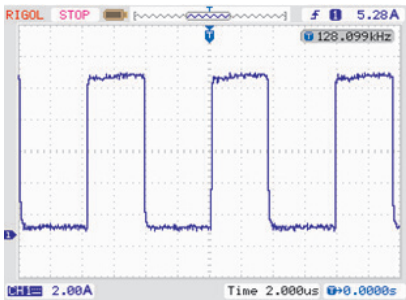


Figure 2. Achieve schematic.

4 EXPERIMENTAL WAVEFORMS RECORDED

4.1 GTR base drive current waveform

After power, GTR base drive current I_b waveform:

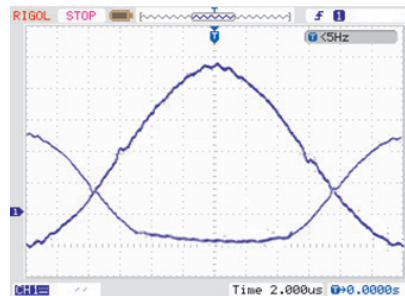


When GTR opened, the drive current frontier steep, and there will be overshoot current, and the opening time and the opening loss can be reduced. When the GTR opened, the drive current is small, controlled

GTR in quasi-saturated conduction state, reducing the driving power, reducing the storage time. When GTR shut down, the reverse base current is large, it can reduce the turn-off time and turn-off loss.

4.2 Composite buffer circuit voltage and current waveforms

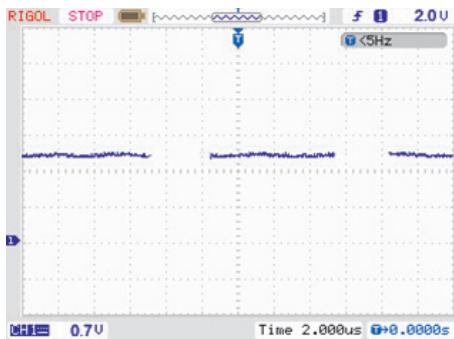
Composite buffer circuit collector current I_c , collector-emitter voltage U_{ce} waveforms:



The composite buffer circuit can reduce switching losses, reducing the voltage and current overshoot. The inductor in series with the GTR may suspend rate rise of the collector current I_c when it opens. Resistor connected in parallel with the GTR current can flow off the subsequent rapid decay, the diode can be isolated in the GTR conduction resistance of the inductor shunt effect.

4.3 Anti-saturation circuit terminal voltage waveform

U_{cb} anti-saturation circuit terminal voltage waveform:

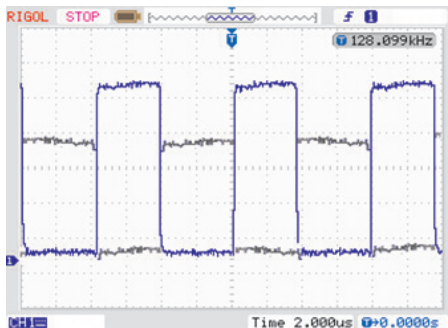


Anti-saturation circuitry can prevent GTR into the enlarged area and depth of the saturated zone, which will improve the operating speed of GTR.

When GTR is in conduction, $U_{cb} = 0.7V$. When the GTR is off, U_{cb} is off.

4.4 Control circuit VT1, VT2 control waveform

Control circuit VT1, VT2 terminal voltage waveform:



The control circuit contains discrete components are generally complex and less stable. The large-scale integrated control circuit can be greatly simplified, and have an automatic protection function, which

VT1, VT2 terminal voltage waveform alternating square wave.

4.5 Optocoupler isolation circuit, transistor voltage waveform

Optocoupler isolation circuit, transistor voltage waveform:



The Optocoupler isolation circuit can solve the electrical isolation between the control circuit and main circuit, and improve the overall safety performance and anti-jamming capability of the circuit. When GTR is opened, the voltage of the Optocoupler isolation circuit the transistor is 0, when the GTR is closed, the voltage of the Optocoupler isolation circuit the transistor is positive or negative V_{cc} . [10]

5 CONCLUSION

In this paper, based on UAA4002 chip it designs a suitable mode for electric vehicle with fast charge of self-turn-off device driver protection circuitry, including circuit overall structure and Optocoupler isolation circuit, control circuit, anti-saturation circuit design composite buffer circuit, and then turn off the device to protect the circuit. [10]

The circuit has a very good performance in the drive, isolation, protection, With simple structure, high reliability, ideal driving effect, high system efficiency, fast response, flexible regulation, characteristics, it must have a wide application.

ACKNOWLEDGMENTS

This work was financially supported by The National High Technology Research and Development of China 863 Program (2011AA05A109), National Energy Application Technology Research and Engineering demonstration project (NY20110701-1)

REFERENCES

- [1] Zhang Wenliang, Wu Bin, Li Wufeng, to the well-off and explore the development direction of China's energy supply mode of pure electric vehicles [J] Power System Technology, 2009, 33 (3): 1–5.
- [2] Tzuo-Hseng S. Li, Shih-Jie Chang, Yi-Xiang Chen, etc. Implementation of human-like driving skills by autonomous fuzzy behavior control on an FPGA-based car-like mobile robot using Infrared Sensors[C]. IEEE Transactions on Industrial Electronics, 2003:145–164.
- [3] Hu Chunyu, Li Wufeng, Yan Hui. Development of a new electric vehicle smart charging machines [J]. China CSEE, 2012, 32: 99–104.
- [4] MOTTO K, LI Y, HUANG A Q. Comparison of the State-of-the-art High Power IGBTs, GCTs and ETOs[A].New Orleans (LA, USA), 2000.1129–1136.
- [5] charges the people, Zhang Yanli PROCEEDINGS. Current limiter based on a new solid-state bridge short-circuit fault IGCT of [J]. Power system automation, 2006, 30 (7): 60–64.
- [6] Samuel V. Araújo, Benjamin Sahan, Peter Zacharias. Comparative Evaluation of SiC-JFETs applied to Power Converters in Renewable Energy Systems[C]. Proceedings of the 9th International PCIM China Conference, 2010:215–221.
- [7] Xu Guohua, Yu Kun, Shen Xiong. A Depth-setting TUV control system based on RS-485 network[C]//2nd International Conference on Manufacturing Science and Engineering.Clausthal-Zellerfeld: Trans Tech Publication, 2011:2295–2299.
- [8] Chen Gang, Jiang Tao burning, Cai Yonghua, et al. Novel solid converter fault has shunt inductor limits. China Electrical Engineering, 2004, 24 (7): 200–205.

4. Chemical, biological and agricultural engineering

Drying of agricultural products using thin layer principle furnace

J. Spišák, A. Olijár, M. Reisová & E. Oravcová

Development and realization workplace of raw materials extracting and treatment, Faculty of mining, ecology, process control and geotechnology, Technical University of Kosice, Nemcovej, Kosice, Slovak Republic

ABSTRACT: For the agricultural postharvest products, moisture content can achieve 50%. Products with high moisture content are unfit for storage in silos. Drying of agricultural materials is the most demanding operations of postharvest processing. The drying process is affected not only by the size of grains, but also by their general properties such as stability, germination, nutritional properties, etc. The process of drying is technically, technologically and economically difficult. Proposed drying device represents a new technology for drying of agricultural materials. The main advantage of the current devices is the high degree of modifiability covering large scale of agricultural products for drying.

1 INTRODUCTION

The grain moisture content can achieve 25%, corn moisture content 30–36% during harvest. It is obvious that moistures are not homogeneous among the grains. With an average moisture content of the grains 22%, there are approximately 10% of grains with the moisture content below 17% and more than 20% of the grains with the moisture content about >25%. During the wet season, the moisture content can reach 20–22%, extremely 29%. For storage of leguminous plants, limit of the moisture content is about 17%, and for cereals it is about 15%. The moisture content of long-term stored seeds (more than one year) should reach 11–12%, and the moisture content of long-term stored grains is 13–14%. During short storage, the grains need to reach the moisture content of 15,5–18%. If the moisture content of long-term stored grains is high, the drying has to be implemented in post-harvest processing. The drying device has to be chosen with respect to the required performance, regime of use, positional conditions and preferred drying medium. For the heating the oil fuel, natural gas, methane or propane/butane are being used. The dryers considering operational condition dependency can be divided as follows:

- Continual (working on one pass),
- Batch.

The drying devices can be direct or indirect (pure air is heated by flue gas in heat exchanger). And quality of dried grains is influenced by three key factors:

- Heating temperature of grains,
- Grain input moisture content,
- Heating time (exposure).

2 A DEVICE THAT WORKS ON THE THIN LAYER PRINCIPLE USED FOR GRAINS DRYING

An Integrated thermal aggregate (ITA) is a thermal aggregate used for the granular material heat treatment. (Figure 1a). The aggregate is based on the principle of a thin layer.

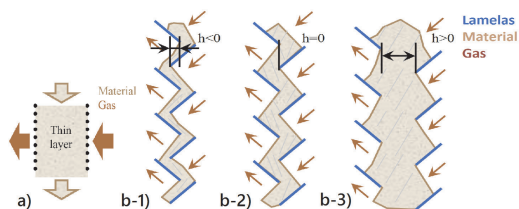


Figure 1. Thin layer a) principle; b) The shape of the layer thickness 1) $h < 0$; 2) $h = 0$; 3) $h > 0$.

A thin layer consists of a processed material with a thickness of a few cm and it does usually not exceed 25 cm. There is the principle based on a cross material moving. A material layer is formed by a system of the lamellas. The lamellas' slope exceeds the angle of repose of the material (Figure 1b). The gap between the lamellas can be more, equal or less to zero. Thus we can obtain the specific configuration of the layer.

The layer thickness depends on the size of the maximum grain and rheological properties of the material. The minimum thickness should be less than triple of a maximum grain size.

In terms of rheological properties, the material powder flow ability is deciding. It is associated with the formation of the arches and the stop of the

material movement. In case of the dry materials, this parameter is not limiting. In case of the wet material, a layer thickness depends on the granularity and powder flow ability decreases in reducing the average grain. At lower speeds maximum 0.5 m/min, the layer movement is a piston. At larger speeds, mutual mixing of grains is obtained.

The main advantage of the ITA technology is an intensive heat transfer due to cross flow of heat medium through a material layer by the size and number of zones in accordance with the technological requirements.

2.1 Modelling of the drying process

The modelling of the drying process in the ITA equipment has been performed in order to simulate the drying process. The elementary balances method has been used for the modelling. The subjects of the modelling are the rheological processes (material motion), hydromechanical (flow) and thermodynamic processes (heating, drying process).

Experimental device proposal.

A proposed experimental device (Fig. 4.) consists of two parallel layers of material with a thickness of 90 mm. The device is vertically divided into 3 parts (zones - heating, drying, cooling). A heat source is placed in a separate combustion chamber. A natural gas and preheated air from the cooling element has been used for burning. A required temperature of the flue gas is ensured by the ambient air addition. Material input is carried out via a hopper located on the top of the device. Material output is carried out via a conveyor belt located under a cooling zone of the thermal equipment (Figure 2).

Conditions of the experiment are shown in Table 1.

The experiment consisted of a start-up phase and the operational phase (Figure 3). Beginning of the

Table 1. Conditions of the experiment.

| | | | | |
|-------------------|----------------------|--------|-------------------|-------|
| Dried material | 15,51 | kg/h | | |
| Material moisture | 4,00 | % | | |
| Amount of product | dm ³ /min | l/hod. | m ³ /h | kg/h |
| Average | 0,304 | 18,24 | 0,0182 | 15,14 |
| Product moisture | 1,58 | % | | |
| Moisture loss | 2,42 | % | | |

experiment was preceded by filling the device with wheat with a moisture content of 4% and a temperature of 22.3°C. The experiment began with the drying medium temperature of 100°C which had fallen to 80°C during a start-up phase of the experiment. In the initial stage, a heating of the material reached a temperature of 80°C which gradually decreased due to decline of input at 60°C. An operational phase of the experiment was initiated by step change of the input to achieve a constant temperature of the drying medium 120°C. An output of the device 15,14 kg/h has achieved a maximum temperature of the material in the drying zone 100°C. The material at the outlet of the cooling zone had a temperature of 43–53°C. A wheat moisture loss 1.58% representing a decrease of 2.42% was secured by drying process in the integrated thermal aggregate.

2.2 The mathematical model calibration

The mathematical model calibration extrapolated to the required extent of the parameters used for the pilot plant equipment design has been performed based on the results of the experiment. Calibration results are shown in Table 2.

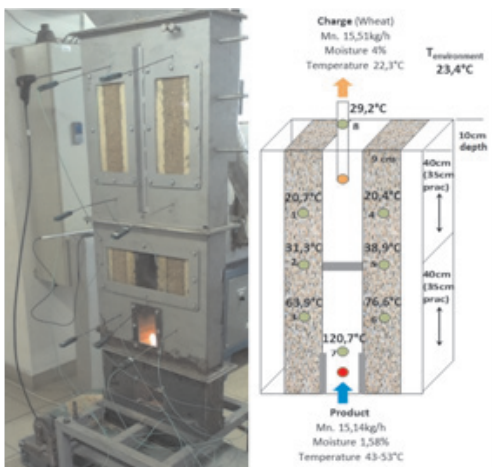


Figure 2. ITA a) physical model; b) description of the experiment.

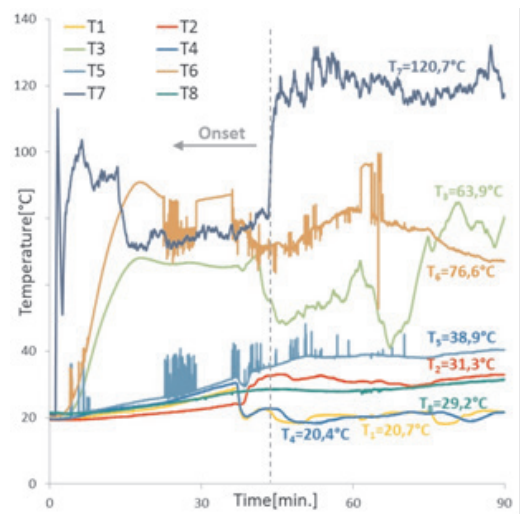


Figure 3. Course of temperatures during experiment.

Table 2. Comparison of the parameters of the experiment and mathematical model.

| Parameters | Experiment | Mat. model | |
|--|------------|------------|--------|
| Charge | 15,51 | 15,51 | [kg/h] |
| Input moisture | 4 | 4 | [%] |
| Product | 15,14 | 15,13 | [kg/h] |
| Output moisture | 1,58 | 1,54 | [%] |
| $T_{3,6 \text{ str}} = T_{\text{mat}}$ | 70,25 | 72 | [°C] |
| $T_8 = T_{\text{waste gases Out}}$ | 29,2 | 32 | [°C] |

The mathematical model highly represents the work of the experimental device (Tab 2). A calibrated mathematical model is a key element of the innovation chain and digital plant.

2.3 The proposal of the pilot plant equipment for grains drying

The device with the parameters shown in Table 3 has been designed for the further research of the device working on the thin layer principle used for agricultural materials drying.

Table 3. Dimensions of the device.

| | | |
|--------------------|---------|------|
| Number of zones | 4 | [-] |
| Height of zone | 800+200 | [mm] |
| Working height | 700 | [mm] |
| Number of layers | 2 | [-] |
| Thickness of layer | 100 | [mm] |
| Depth of layer | 400 | [mm] |

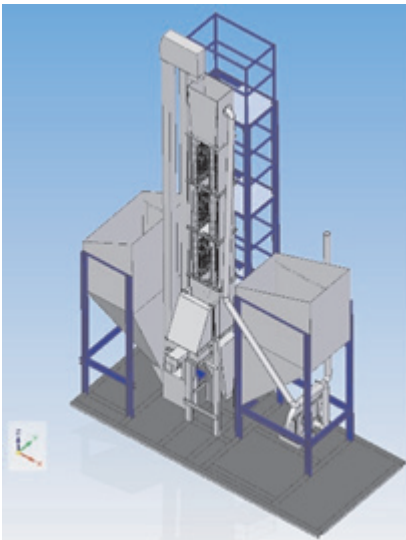


Figure 4. 3D Model of pilot furnace.

2.4 Drying simulations

Based on the requirements on wheat storage in silos presented as introductory part, a series of simulations for wheat drying with different performance of the furnace and different temperature of drying medium are provided. The simulation results are shown in Figure 6-11.

First are simulations for the maximum flue gas temperature of 80°C (Figure 5-7), then it is a simulation for the input moisture content of wheat 18, 21 and 29% (Figure 8-10).

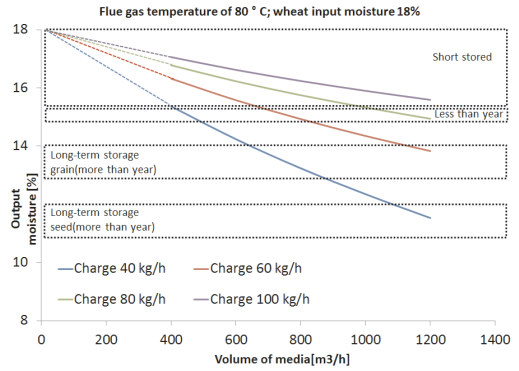


Figure 5. Flue gas temperature of 80°C; wheat input moisture 18%.

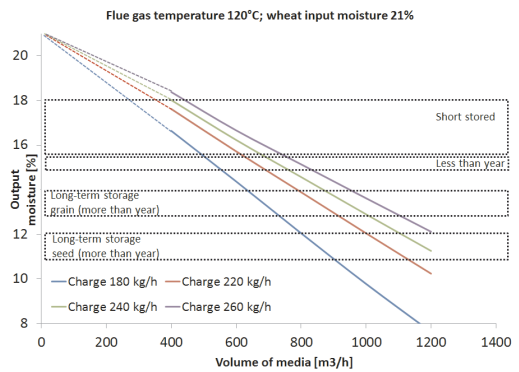


Figure 6. Flue gas temperature of 80°C; wheat input moisture 21%.

3 CONCLUSIONS

In the post-harvest processing of agricultural materials, drying belongs to the necessary and one of the most complex technological operations. Device operating on the principle of thin layer is a revolutionary solution in the field drying of agricultural materials. The main advantage of integrated thermal apparatus lays in a high degree of modifiability which is

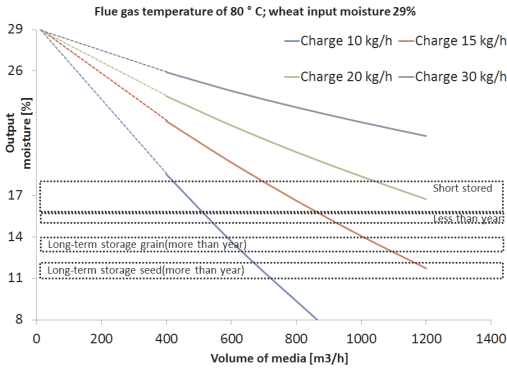


Figure 7. Flue gas temperature of 80 °C; wheat input moisture 29%.

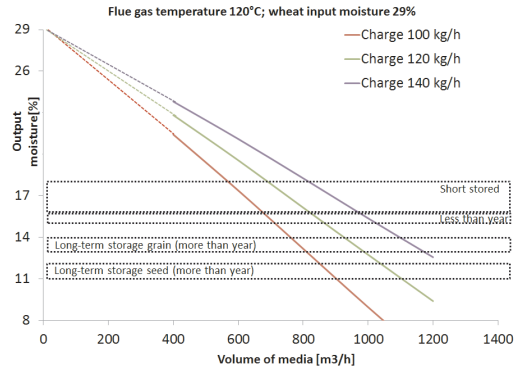


Figure 10. Flue gas temperature of 120 °C; wheat input moisture 29%.

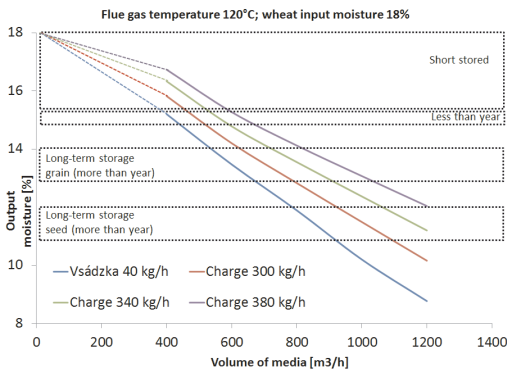


Figure 8. Flue gas temperature of 120 °C; wheat input moisture 18%.

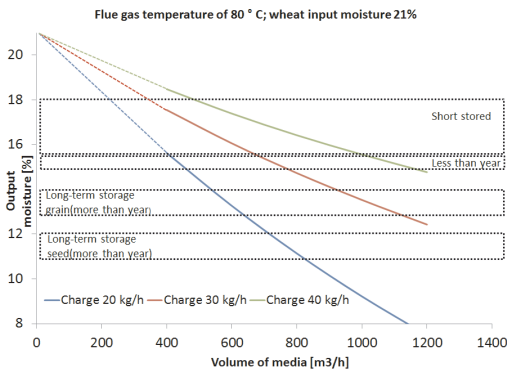


Figure 9. Flue gas temperature of 120 °C; wheat input moisture 21%.

ensured use of the furnace as a drying device for a large scale of agricultural materials (drying of grains, biomass). The uniqueness of the device represents its construction solution which significantly facilitates frequent problems with moving technological devices. The equipment for drying of agricultural products operating on the principle ITA technology equipment has been suggested. The device is applicable for a wide range of agricultural products. The mathematical model has been created for the device design. Model parameters have been determined on a physical model based on the principle of equivalence. The prototype used to design pilot plant equipment has been created on the basis of the calibrated mathematical model.

ACKNOWLEDGMENT

This contribution/publication is the result of the project implementation Advanced technology for the mining company of 21st century supported by the Research & Development Operational Program funded by the ERDF⁺. (ITMS: 26220220131)

REFERENCES

- [1] Mikula J. et al. : Mathematical modelling of lumpy and granular material thermal treatment. In: Acta Metallurgica Slovaca, ISSN 1335-1532, vol. 15, no. 1 (2009), pp. 197–204.
- [2] Mikula J. et al. : Generation of mathematical hierarchical models for virtual reality environment. In: ICCV'2009. Zakopane, Poland, May 24–27, 2009, Krakow : AGH - University of science and Technology, 2009. ISBN 8389772-51-5, pp. 419–422.
- [3] Ertekin, C. and Yaldiz, O., Drying of Egg-plant and Selection of a Suitable Thin layer Drying

- Model, *Journal of Food Engineering*, Vol.63, pp.349–359, 2004.
- [4] Zhang jinjiang, Liu Qinghua, Wu Yaosen, ect. Research on New Techniques and Equipments to Dry Agricultural Products [J], *Modem Agricultural Equipments*, 2009, 12, P42–44.
- [5] Cui Lu, Wang Xiangying, Dou Zhihao, ect., Situation and expectation on domestic and international solar drying agricultural products [J], *Agriculture Engineering Technology (Agricultural Product Processing Industry)*, 2010, 5, P37–39.
- [6] Brooker, D.B., Bakker-Arkema, F.W., & Hall, C.W. (1992). *Drying and Storage of Grains and Oilseeds*. New York: Van Nostrand Reinhold.
- [7] Maier, D.E., & Bakker-Arkema, F.W. (2002). Grain Drying Systems. Facility Design Conference of the Grain Elevator & Processing Society, St. Charles, Illinois, U.S.A., 28–31 July 2002. *Proceedings*, 1–53.
- [8] Nellist, M.E., & Bruce, D.M. (1995). Heated-air grain drying. In: Jayas, D.S., White, N.D.G., & Muir, W.E. (Eds.) *Stored-Grain Ecosystems* (pp.609–659). New York: Marcel Dekker Inc.

Effect of gamma irradiation on saccharification of *Undaria* sp. biomass

J. Choi

Department of Biotechnology and Bioengineering, Chonnam National University, Gwangju, South Korea

ABSTRACT: The application of irradiation technique to the saccharification process of seaweed was investigated. The brown seaweed *Undaria* biomass was irradiated at doses of 0, 10, 50, 100, 200, and 500 kGy, and then hydrolyzed using sulfuric acid. The effect of gamma irradiation was measured by the concentration of the reducing sugar of hydrolysates. The concentration of reducing sugar of hydrolyzates significantly increased as a result of gamma irradiation, with or without acid hydrolysis. Microscopic images show that gamma irradiation causes structure breakage of the *Undaria* cell wall. These results indicate that the combined method of gamma irradiation with acid hydrolysis can significantly improve the saccharification process for bioenergy production from marine algae biomass.

1 INTRODUCTION

Consumption of fossil fuels increased rapidly during the age of global industrialization. High energy consumption caused environmental problems, such as global warming, and it was also causing the exhaustion of fossil fuel. A potential solution to these worldwide problems is the development of alternative energy from renewable biomass.

Bioenergy is a renewable alternative fuel that could replace petroleum-based fuel like gasoline and diesel. Although land plant biomass is easily fermented for bioenergy, bioenergy production from land plants causes rising of food prices through competition with food sources (Adams et al., 2010).

Marine algae is an abundant and year-round source of carbohydrates. Approximately 20–30% of marine algae biomass is made up of cellulose or starch, and marine algae also contains various mono-saccharides, such as glucose, xylose, and galactose. Brown algae, also known as marine macroalgae, contains up to 67% carbohydrates by dry weight, including materials such as alginate, laminarin, and mannitol (Horn et al., 2000). Moreover, algal cells can be harvested within a short span of time as compared to other land feedstocks, and hence they can meet the increasing demand of feedstocks for bioenergy production (Harun et al., 2010).

These marine algal sugars can be extracted using saccharification method, and these extracts can be converted into bioethanol and biodiesel (Saha et al., 2005). However, the saccharification yield from algae is not sufficient for economical production of bioethanol.

Several pretreatment methods of improving saccharification on cellulosic raw materials have been

investigated (Esteghlalian et al., 1997; Silverstein et al., 2007).

Ionizing irradiation can modify and disrupt the structure of lignocellulose and can be an effective method of pretreatment of lignocellulosic biomass for sugar production. Radiation degradation for increasing sugar yield has been reported in regards to various lignocellulosic materials, such as bagasse (Han et al., 1981), wheat straw (Yang et al., 2008), chaff, and sawdust (Kumakura and Kaetsu, 1984).

Therefore, in this work, the effect of gamma irradiation on the saccharification of seaweed was investigated.

2 MATERIALS AND METHODS

2.1 Seaweed preparation

The brown seaweed *Undaria* sp. was harvested in the region near Wando Island, South Korea. Before the experiment, samples were washed with distilled water and then stored at 4°C until they were used.

2.2 Gamma irradiation

The wet samples were irradiated in a cobalt-60 gamma-irradiator (IR-221, Nordion International Ltd., Ontario, Canada) equipped with 11.1 PBq strength at $22 \pm 0.5^\circ\text{C}$ and operated at a dose rate 10 kGy/hr. The applied dose levels were 10, 50, 100, 200, and 500 kGy. Dosimetry was performed in 5 mm-diameter alanine dosimeters (Bruker Instruments, Rheinstetten, Germany), and after irradiation the sample was stored at 4°C for further experiments. Non-irradiated samples were also prepared for control studies.

2.3 Acid hydrolysis

The samples were re-suspended in 1% sulfuric acid (H₂SO₄) and incubated at 121°C for 0 to 240 min. The hydrolyzates were cooled down to room temperature and then neutralized with calcium carbonate (CaCO₃). After neutralization, calcium sulfate (CaSO₄), a product of the chemical reaction, was removed using a centrifuge operated at 8,000 rpm for 15 min.

2.4 Concentration of reducing sugar

The concentrations of reducing sugar after the saccharification process were determined by the 3, 5 dinitrosalicylic acid (DNSA) method (Miller, 1959). Specifically, 1 mL pretreated samples were transferred into 15 mL glass tubes, and 2 mL of the modified DNSA reagent (0.5 g dinitrosalicylic acid, 8 g sodium hydrate, and 150 g Rochelle salt in 500 mL distilled water) was added to the tube. The mixture was vortexed for 5 sec, boiled at 90°C for 10 min, and then cooled in ice. The absorbance of the samples was measured at 550 nm using a spectrophotometer (UV-1601 PC, Shimadzu Co.), and the OD values were converted to concentration of reducing sugar using a standard curve prepared with different concentrations (0.625, 1.25, 2.5, 5, and 10 mg/mL) of glucose.

3 RESULTS AND DISCUSSION

3.1 Concentration of reducing sugar after acid hydrolysis

The brown seaweed *Undaria* biomass was saccharified at 121°C with 1% sulfuric acid concentrations for different residence times (0 to 240 min). **Table 1** shows the changes in the concentration of reducing sugar resulting from acid hydrolysis. The concentration of reducing sugar increased with increasing

Table 1. Concentration of reducing sugar on algal biomass after acid (1% H₂SO₄) hydrolysis.

| Time for acid hydrolysis (min) | Reducing sugar concentration (g/L) |
|--------------------------------|------------------------------------|
| 0 | 0.0169 |
| 15 | 0.0186 |
| 30 | 0.0193 |
| 60 | 0.0232 |
| 120 | 0.0298 |
| 180 | 0.0383 |
| 240 | 0.0380 |

residence time. In sulfuric saccharification, a low acid concentration (e.g., 0.1 to 1% sulfuric acid) at high temperature is commonly used because a high acid concentration is extremely corrosive and dangerous (Taherzadeh and Karimi, 2008).

The positive effects of conversion of cellulosic biomass into fermentable sugars for bioethanol production through the use of low acid concentration have been reported (Cara et al., 2007); however, inhibitors that form at the low pH usually inhibits the microbial growth and fermentation, and this results in lower ethanol yield. Moreover, this process requires acid recovery for economic reasons (Taherzadeh and Karimi, 2008). Therefore, the pretreatments at low pH should be selected properly in order to reduce the formation of these inhibitors, and economic efficiency should be taken into consideration.

3.2 Concentration of reducing sugar after gamma irradiation

The marine algae samples were irradiated at doses of 0, 10, 50, 100, 200 and 500 kGy. **Table 2** shows the changes in the concentration of reducing sugar in marine algae, resulting from the gamma irradiation. The concentration of reducing sugar increased with increasing gamma radiation dose. The reducing sugar concentration of the non-irradiated sample was about 0.017 g/L, whereas the concentration in the sample irradiated at a dose of 500 kGy increased to about 0.048 g/L. In general, irradiation leads to the degradation of polysaccharides such as starch and cellulose because it breaks the glycosidic bonds in the presence of water (Sokhey and Hanna, 1993). Saccharification of cellulosic biomass from plants for bioethanol production has been enhanced by the use of high-energy gamma irradiation (Youssef and Aziz, 1999). The effects of gamma irradiation on depolymerization were consistent with those reported in other studies in regards to algal polysaccharides such as starch, agar, alginate, and carrageenan (Aliste et al., 2000).

Table 2. Concentration of reducing sugar on algal biomass after gamma irradiation.

| Absorbed dose (kGy) | Reducing sugar concentration (g/L) |
|---------------------|------------------------------------|
| 0 | 0.0167 |
| 10 | 0.0181 |
| 50 | 0.0219 |
| 100 | 0.0273 |
| 200 | 0.0380 |
| 500 | 0.0493 |

3.3 Concentration of reducing sugar after combined acid hydrolysis with gamma irradiation

Treatment involving a combination of irradiation and other methods such as acid treatment and mechanical crushing can further accelerate the saccharification process (Mamar et al., 1990). Irradiation can enhance cellulosic degradation of cellulose into glucose. In this study, the effects of irradiation used for pretreatment of marine algae prior to acid hydrolysis (1% sulfuric acid at 121°C for 180 min) were studied (Table 3). The reducing sugar concentration of gamma-irradiated marine algae at the dose of 500 kGy increased from about 0.017 g/L to about 0.048 g/L, while the reducing sugar concentration significantly increased to about 0.235 g/L through combined treatment involving gamma irradiation and acid hydrolysis.

Table 3. Concentration of reducing sugar on algal biomass after acid hydrolysis with gamma irradiation.

| Absorbed dose (kGy) | Reducing sugar concentration (g/L) |
|---------------------|------------------------------------|
| 0 | 0.036 |
| 10 | 0.042 |
| 50 | 0.068 |
| 100 | 0.087 |
| 200 | 0.136 |
| 500 | 0.235 |

3.4 Change in cell wall construction from gamma irradiation

The cell walls of brown algae generally contain three component polysaccharides, namely, cellulose, alginic acid, and sulfated polysaccharides. Among these polysaccharides, cellulose is derived from D- glucose units, which condense through beta (1-4)- glycosidic bonds and provide structural support to the algal cell wall.

The radiation-induced reactions in the cellulose materials are known to be initiated through rapid localization of the absorbed energy within the molecules; this produces high-activity radicals that cause the secondary degradation of materials through chemical reactions such as chain scission (Khan et al., 2006). High-dose gamma irradiation of ligno-cellulosic materials causes cell wall degradation by weakening the cell wall constituents, or it depolymerizes and delignifies the fiber (Al-Masri and Zarkawi, 1994).

To investigate the pretreatment effect of gamma irradiation, irradiated marine algae was observed

using an optical microscope. The cell walls of the marine algae were destroyed by high-dose gamma irradiation. In contrast to those of the non-irradiated sample, the cell walls and the membranes of the irradiated sample were broken, and thus algal cells were released. This result indicates that when irradiating marine algae for pretreatment, gamma irradiation can increase not only the production of mono- saccharides but also the efficiency of sugar extraction.

4 CONCLUSION

Pretreatment involving acidic and enzymatic digestion is commonly used for producing bioethanol from polysaccharides, but these processes are characterized by high cost, low yield, long processing time, and waste (Taherzadeh and Karimi, 2008). Gamma irradiation, on the other hand, is simpler and more eco-friendly than chemical or enzymatic digestion. The results of this study indicate that the concentration of reducing sugar in algal biomass significantly increases as a result of gamma irradiation, with or without acid hydrolysis.

In this study, it is shown that irradiation increases the concentration of reducing sugar and increases the saccharification yield. Thus, gamma irradiation could be an effective method of converting algal polysaccharides to fermentable sugar for bio- ethanol production.

ACKNOWLEDGMENT

This study was financially supported by the Nuclear R&D program of the Ministry of Science, ICT & Future Planning (MSIP), by Golden Seed Project, Ministry of Agriculture, Food and Rural Affairs (MAFRA), Ministry of Oceans and Fisheries (MOF), Rural Development Administration (RDA) and Korea Forest Service (KFS), by the research supporting program by Chonnam National University 2013.

REFERENCES

- [1] Adams, J.M., Ross, A.B., Anastasakis, K., Hodgson, E.M., Gallagher, J.A., Jones, J.M. & Donnison, I.S. 2010. Seasonal variation in the chemical composition of the bioenergy feedstock *Laminaria digitata* for thermochemical conversion. *Bioresour. Technol.* 102:226–234.
- [2] Aliste, A.J., Vieira, F.F. & del Mastro, N.L. 2000. Radiation effects on agar, alginates and carrageenan to be used as food additives. *Radiat. Phys. Chem.* 57:305–308.
- [3] Al-Masri, M.R. & Zarkawi, M. 1994. Effect of gamma irradiation on cell wall constituents of some agricultural residues. *Radiat. Phys. Chem.* 44:661–663.

- [4] Cara, C., Ruiz, E., Oliva, J.M., Saez, F. & Castro, E. 2007. Conversion of olive tree biomass into fermentable sugadilute acid pretreatment and enzymatic saccharification. *Bioresour. Technol.* 99:1869–1876.
- [5] Esteghlalian, A., Hashimoto, A.G., Fenske, J.J. & Penner, M.H., 1997. Modeling and optimization of the dilute-sulfuric-acid pretreatment of corn stover, poplar and switchgrass. *Bioresour. Technol.* 21:129–136.
- [6] Han, Y.W., Timpa, J., Clegler, A., Courtney, J., Curry, W.F. & Lambremont, E.N. 1981. γ -Ray-induced degradation of lignocellulosic materials. *Biotechnol. Bioeng.* 23:2525–2535.
- [7] Harun, R., Danquah, M.K. & Forde, G.M. 2010. Microalgal biomass as a fermentation feedstock for bioethanol production. *J. Chem. Technol. Biotechnol.* 85:199–203.
- [8] Horn, S.J., Aasen, I.M. & Ostgaard, K., 2000. Ethanol production from seaweed extract. *J. Ind. Microbiol. Biotechnol.* 25:249–254.
- [9] Khan, F., Ahmad, S.R. & Kronfli, E. 2006. Gamma-radiation induced changes in the physical and chemical properties of lignocellulose. *Biomacromolecules.* 7:2303–2309.
- [10] Kumakura, M. & Kaetsu, I. 1984. Heat enhancement effects in radiation pretreatment of cellulosic wastes. *Industrial & Engineering Chemistry, Product Research and Development* 23:88–91.
- [11] Mamar, S.A.S. & Hadjadj 1990. A. Radiation pretreatments of cellulose materials for the enhancement of enzymatic hydrolysis. *Radiat. Phys. Chem.* 35:451–455.
- [12] Miller, G.L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Biochem.* 31:426–428. Saha, B. C., Iten, L. B., Cotta, M. A. & Victor Wu, Y. 2005. Dilute acid pretreatment, enzymatic saccharification and fermentation of wheat straw to ethanol. *Process Biochem.*, 40:3693–3700.
- [13] Silverstein, R.A., Chen, Y., Sharma-Shivappa, R.R., Boyette, M.D. & Osborne, J., 2007. A comparison of chemical pretreatment methods for improving saccharification of cotton stalks. *Bioresour. Technol.* 98:3000–3011.
- [14] Sokhey, A.S. & Hanna, M.A., 1993. Properties of irradiated starches. *Food Struct.* 12:397–410.
- [15] Taherzadeh, M. J. & Karimi, K. 2008. Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: A Review. *Int. J. Mol. Sci.* 9:1621–1651.
- [16] Yang, C., Shen, Z., Yu, G. & Wang, J. 2008. Effect and aftereffect of γ radiation pretreatment on enzymatic hydrolysis of wheat straw. *Bioresour. Technol.* 99:6240–6245.
- [17] Youssef, B.M. & Aziz, N.H., 1999. Influence of gamma-irradiation on the bioconversion of rice straw by *Trichoderma viride* into single cell protein. *Cytobios.* 97:171–183.

Pythium oligandrum colonization enhances wheat growth and resists disease caused by *Pythium arrhenomanes*

Y. Zhu

Institute of Biology, Gansu Academy of Sciences, Lanzhou, China
School of Life Sciences, Lanzhou University, Lanzhou, China

Z.Y. Wang, H.S. Qi, H. Yang & J.P. Zhou

Institute of Biology, Gansu Academy of Sciences, Lanzhou, China

Z.B. Wang

School of Information Science and Engineering, Lanzhou University, Lanzhou, China

ABSTRACT: A kind of *Pythium* species is isolated from wheat roots and identified as *Pythium arrhenomanes*. *Pythium arrhenomanes* is virulent to wheat plants, but not to seeds sprouting. However, another kind of *Pythium* specie named *Pythium oligandrum* has the potential for use against *Pythium arrhenomanes* and induces plant disease resistance. The disease incidence of plants inoculated by *Pythium arrhenomanes* and *Pythium oligandrum* is significantly ($P=0.05$) decreased, and the plant height and biomass production are visible increased compared with only *Pythium arrhenomanes* treatment plants. The pH of xylem sap commonly changes under adverse situation, which is an important component of root to shoot communication. The pH of the stem is also a little higher during treatment with fungus than control. Catalase (CAT) and malondialdehyde (MDA) are not significantly changed, suggesting that they are not the main factors for disease resistance to *Pythium arrhenomanes*. Superoxide dismutase (SOD) is sharply augmented in dual fungus treatments, while it is a little reduced in *Pythium arrhenomanes* treatments compared with control, which may play the key role to activate the disease resistance. *Pythium oligandrum* could boost plants tiller, aggrandize plant height and enhance biomass production in our experiments. So, *Pythium oligandrum* also has the ability to improve plant growth.

KEYWORDS: *Pythium Oligandrum*, wheat, *Pythium Arrhenomanes*, protection, antioxidant enzymes.

1 INTRODUCTION

Pythium oligandrum (P.O) is ubiquitous in the field or greenhouse and firstly isolated from pea root disease and wrongly considered the fungus blight pathogen [1]. P.O is a common soil fungus with worldwide distribution such as: Central Africa, Pakistan, Germany, Hawaii, South Australia and China. Along with further studies, the new functions have been found that P.O has an ability to defense soil borne plant diseases caused by a number of fungal and bacterial pathogens [2]. It has been studied that gray mold caused by *Fusarium oxysporum* and *Trichoderma harzianum*, bacterial wilt disease caused by *Ralstonia solanacearum* [3], and root rot caused by *Fusarium crown* [4] in tomato are suppressed by P.O. Some researches show that P.O not only is used for biological control agent, but also improves rice growth and production [5]. Molecular mechanism has been proved that genus *Pythium* includes more than 80 species, but only several species have these functions; and it has been demonstrated that cell wall protein fractions (CWPs), cellulases, and crab shell Chitosan of P.O play a key role in disease resistance [6].

Pythium arrhenomanes (P.A) is the fungal pathogen that heavily infects wheat [7], maize, barley, sugarcane, Ornamental plants and soybean. P.A infection is common to cause damping-off diseases of crops and impact all stages of plant development, which can bring in economic losses to farmers. P.A has been isolated and identified by our group from wheat roots in this study. P.A could cause wheat blight and could lead to the death rate about 80%. We want to know that whether P.O could repress this pathomycete P.A. In the present study, P.O was tested for the biocontrol to P.A and could suppress the pathogen and enhance wheat growth.

2 MATERIALS AND METHODS

2.1 Isolation and identification of the fungus from wheat root.

Wheat plants showing a little of ill-being are collected from pot tests in Agriculture Drought Ecological Experiment Station, located at YuZhong Lanzhou University Campus in June, 2011. Wheat roots are

washed thoroughly in running tap water, distilled water and are surface disinfected in 0.1% mercuric chloride for 15s, then washed with distilled water for 3 times, finally blotted dry with sterile filter paper. Prepared roots samples are chopped in 0.2 mm length, and inoculated on potato medium (PDA) agar plates, then cultured incubator for 3 days at 25°C. After 3 days, there are mycelia growing in some root segments. The mycelia are picked with inoculating needle and then planted on PDA agar plates for isolation and purification. The purified fungus is conserved on strains cant at -80°C for further use. The fungus is identified for *Pythium arrhenomanes* with rDNA ITS sequence analysis [8].

To enrich P.A, we culture mycelia in the PDA liquid medium. 500 ml flask filled with 200ml PDA medium is inoculated with 5 inoculating loops from P.A strains cant, and then cultured in the concussion incubator with the speed of 280 r/min for 7 days at 25°C. Mycelium and fermentation broth are fragmented with a sterilized blender as previously described, and then mixed with sterile soil at the ratio of 1:1000 by dry weight. Each of the pots is filled with 2.5 kg P.A inoculated soil.

2.2 *P.O cultures and inoculum*

The strain of P.O is isolated and purified from one kind of biocontrol agents, which is kindly provided by the Institute of Biology, Gansu Academy of Sciences. The fungus is identified as *Pythium oligandrum* with rDNA ITS sequence analysis. We use the same way with P.A to obtain P.O and P.O-inoculated soil.

2.3 *Wheat growth*

Wheat seeds (*Triticum aestivum* L.) are sterilized by 0.1% mercuric chloride for 1min and thoroughly rinsed three times with sterile-distilled water, and then put on sterile filter paper to dry. The seeds are planted in pots about 5cm depth and 15 seeds per pot. They are randomly arranged on a greenhouse under 400W mercury halide lamps. The temperature is controlled at 25°C by air conditioning and the light application time is 12 hours.

Seedling emergence is relatively complete after 10 days. Only ten seeds in similar height are remained. Then, all plants are watered with nutrient solution mixed with NH_4NO_3 (0.64g/kg soil) and KH_2PO_4 (0.28g/kg soil) in one time. In order to keep soil water content between 60% and 80%. Sterile water is added twice a week.

2.4 *Parameters measured*

2.4.1 *Plant height*

When plants are 3-week-old, three plants height are randomly measured with ruler in each pot and was

analyzed by Statistical Analysis Software 17.0 and Origin 8.5.1.

2.4.2 *Disease assessment*

Disease incidence (DI) is assessed for each treatment when plants are 3-week-old. The DI is estimated by dividing the number of dead plants by 10 (the total keeping plants in each pot) $\times 100$.

2.4.3 *Tiller rate*

When plants are 6-week-old, the productive tiller number is counted. The tiller rate is calculated by dividing the number of tiller plants by total number of keeping strains in each pot $\times 100$.

2.4.4 *Biomass production*

After 6-week-old, the root system is separated from shoots. The ground biomass part and underground part biomass are collected. The shoot tissues are taken off separately and immediately frozen in liquid nitrogen and kept at -20°C until use for pH and antioxidant enzyme analysis.

2.4.5 *Antioxidant enzymes*

Sample preparation: about 1.0 g fresh leaves are pulverized with liquid nitrogen in mortar and homogenized with 10mL 0.5M phosphate buffer (pH 7.5) in a centrifuge tube. The supernatant obtained after centrifugation (15000 rpm at 4°C) is used for enzyme assay.

Superoxide dismutase (SOD) activity is assayed with spectrophotometry. 50 μL supernatant is added into 3mL reaction mixture, which is consisted of Na_2CO_3 , 200mM methionine, 2.25mM NBT, 3mM EDTA, 60 μM riboflavin and 0.1M phosphate buffer (pH 7.8). Matching test is done by using 50 μL buffer (0.5M phosphate buffer, pH 7.5) instead of supernatant. Absorbance is read at 560 nm.

Lipid peroxidation is determined by malondialdehyde (MDA). MDA is also assayed with spectrophotometry. 2mL supernatant is added into 2mL 0.6% thiobarbituric acid (TBA), heated in boiling water for 15min, and centrifuged after cooling. Supernatant is obtained for colorimetric analysis at 600 nm, 532 nm and 450nm respectively. 2mL supernatant replaced by 2mL distilled water serves as blank. The MDA content is calculated according to the formula: $(6.45 \times (\text{OD}_{532} - \text{OD}_{600}) - 0.56 \text{OD}_{450}) \times 0.03 / W$ (W stands for sample weight).

Catalase (CAT) activity is assayed by measuring the initial rate of disappearance of H_2O_2 . 0.1mL supernatant is added into 1mL 0.3% H_2O_2 and 2mL H_2O , measuring the reduce rate of OD value at 240nm. The CAT activity meaning the amount necessary to decompose 1 μL of H_2O_2 per minute at 25°C , is calculated by using the formula: $\text{mA}_{240} \times V = V_a \div W$ (mA_{240} stands for the change of OD value in per minute; V stands for the total

volume; V_a stands for sample volume; W stands for sample weight).

2.4.6 The pH of stem

Stems of wheat are juiced, dipped the juice with glass bar on precision pH test paper and read the pH value of the standard calorimetric plate.

3 RESULTS

3.1 Plant height is enhanced by the colonization of *Pythium oligandrum*

P.O-colonized plants showed significantly higher growth performance compared to P.A-colonized

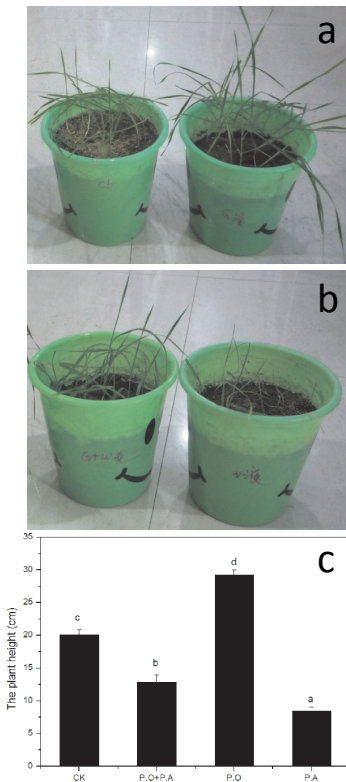


Figure 1. Plant height of wheat inoculated or not with *Pythium oligandrum* and *Pythium arrhenomanes*. a: The wheat growth comparing with control and P.O treated. b: The wheat growth comparing with P.A-colonized and dual fungi-colonized. c: Data correspond to the means of twenty-seven experiments (3 strains in each pot \times 9 pots). The bars with the same letter do not differ significantly at $P \geq 0.05$ determined by LSD test. CK: control; P.A: *Pythium arrhenomanes* inoculated plants; P.O: *Pythium oligandrum* inoculated plants; and P.O+P.A: dual inoculated with *Pythium oligandrum* and *Pythium arrhenomanes*.

plants, dual fungi-colonized plants and non-colonized plants (Figure 1). This most significant growth is always associated with an increase of the roots (Figure 2). The plants colonized by P.O and P.A are much taller than those of only P.A-colonized plants, but a little shorter compared with the control plants (Fig. 1). The plants colonized by P.A are lower in all experiments (Fig. 1 b and c).

3.2 Disease assessment can be reduced by P.O.

In inoculating P.A treatments, disease incidence (DI) is twice times higher than those of dual-culture treatments and significantly higher than control treatments. In spite of the fact that the treatment of P.O lower DI to zero (Table 1), treatment with both P.O and P.A significantly reduces DI after 3 weeks. In this connection, significantly decrease in DI is achieved with dual-culture treatments, while treatment of P.O alone could enhance live incidence compared to control treatment.

3.3 Tiller rate is boosted by P.O.

The tiller rate is significantly varied among all treatments (Table 1). However, the plant demonstrates the same tiller rate between the single (P.A) and the dual (P.O+P.A) fungi treatments, as P.A is virulent to wheat plants and effected tiller rate. The plants colonized by P.O have the tiller rate of 52.5%, with is almost twelve-fold higher than control (4.4%) (Table 1), implying that the P.O could promote tiller.

Table 1. The Effect of P.O and P.A Inoculated on Disease Incidence, Tiller Degree and the Ratio of Root and Shoot.

| Treatment | 3 Weeks after planting | 6 Weeks after planting | The ratio of root and Shoot (%) | |
|-----------|------------------------|------------------------|---------------------------------|---------|
| | Disease incidence (%) | Tiller rate (%) | The pH of stem | |
| CK | 3 a | 4.4b | 5.88 a | 11.00 a |
| P.O | 0 b | 52.5c | 6.11 a | 17.67 a |
| P.A | 79 c | 0 a | 5.96 a | 19.00 a |
| P.O+P.A | 34 d | 0 a | 5.97 a | 16.67 a |

CK: control; P.A: *Pythium arrhenomanes* inoculated plants; P.O: *Pythium oligandrum* inoculated plants; and P.O+P.A: dual inoculated with *Pythium oligandrum* and *Pythium arrhenomanes*. Values of each column followed by the same letters are not significantly different according to Duncan's multiple range test ($p = 0.05$), each value represents the mean of nine replicates.

3.4 Biomass production is improved by *P.O*

The dry weights of P.O-colonized root systems are significantly heavier than other experiments (Fig. 2). The root systems colonized by P.O and P.A has more weight than those of only P.A-colonized, but a little lighter compared with the control (Fig. 2 A). The root systems colonized by P.A are lighter in all experiments, but not significant contrast to control plants and dual fungi-colonized tests (Fig. 2 A).

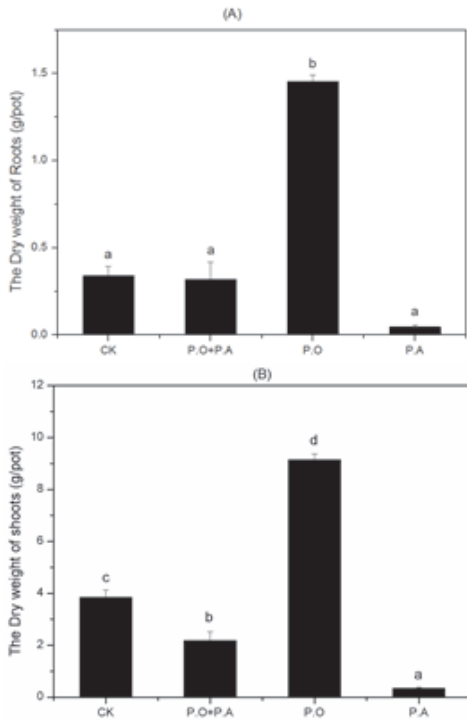


Figure 2. Biomass production of wheat inoculated or not with *Pythium oligandrum* and *Pythium arrhenomanes*. Data correspond to the means of nine experiments. The bars with the same letter do not differ significantly at $P \geq 0.05$ determined by LSD test. CK: control; P.A: *Pythium arrhenomanes* inoculated plants; P.O: *Pythium oligandrum* inoculated plants; and P.O+P.A: dual inoculated with *Pythium oligandrum* and *Pythium arrhenomanes*.

The shoots of plants with dual fungi-colonized develop significantly compared with the plants inoculated by P.A (Fig. 2 B). Inoculation with P.O has a good effect on shoot weight, almost triple of controls (Fig. 2 B). Although the weights of root systems are not significant in each other (Fig. 2 A), the shoot weights are remarkable in all tests (Fig. 2 B). However, the ratio of root and shoot is not significant in plants, whether colonized with single and dual fungi or control (Table 1).

3.5 The changes of antioxidant enzymes and the pH of stem

MDA is slightly increased in plants colonized with microorganism suggesting induction of oxidative stress, but not very significant compared to control (Fig. 3 c). The SOD activity of P.A-colonized is low, while inoculation with P.O and P.A has the highest

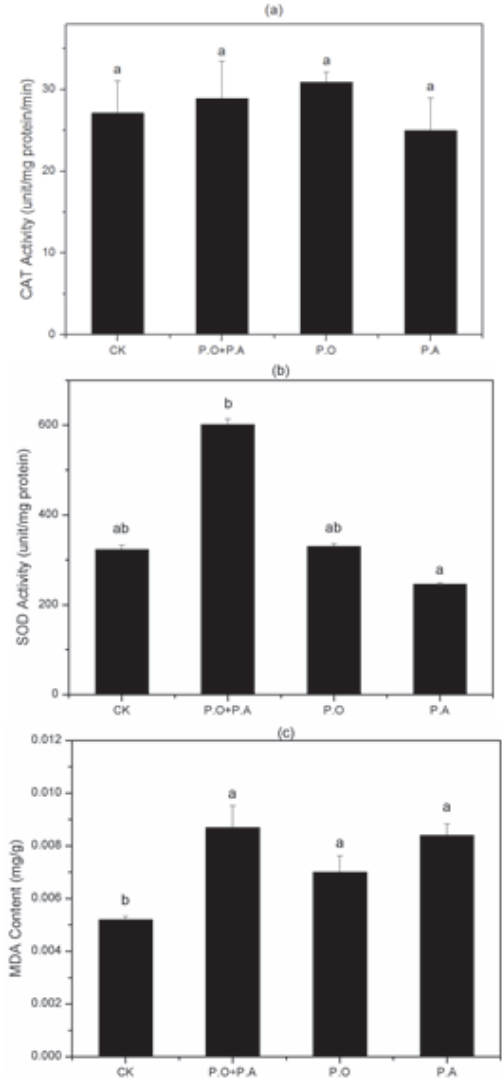


Figure 3. MDA content and the activity of SOD and CAT in wheat inoculated or not with *Pythium oligandrum* and *Pythium arrhenomanes*. Data correspond to the means of three experiments. The bars with the same letter do not differ significantly at $P \geq 0.05$ determined by LSD test. CK: control; P.A: *Pythium arrhenomanes* inoculated plants; P.O: *Pythium oligandrum* inoculated plants; and P.O+P.A: dual inoculated with *Pythium oligandrum* and *Pythium arrhenomanes*.

SOD activity (Fig. 3 b). However, the SOD activity of only P.O.-colonized is not significant change compared with control. The CAT activity has a little alteration in every treatment, but not very significant (Fig. 3 a). With fungus colonized, the pH of the stem has obviously increased corresponding to root to shoot communication (Table 1), but is not very significant.

4 DISCUSSION

4.1 Disease resistance

Suppression of blight disease caused by P.A is observed in wheat cultivars treated with dual inoculated by P.A and P.O. The mechanism of P.O putting down disease is better studied as follows: Akira Masunaka and Kazuhiro Nakaho suggest that P.O represses bacterial wilt disease by inducing the plant defense, not directly competing for infection sites [3], while Floch et al demonstrates that P.O restrains fungal disease caused by *Pythium* genus pathogens by not only competing for infection sites, but also inducing the plant defense [6]. However, some studies mention that P.O parasitizes *Pythium* genus pathogens and eventually leads the host destruction [4]. In this experiment, all treatments seeds germinate entirely, so P.A cannot influence wheat seeds sprouting. But P.A is virulent to wheat seedlings and results in almost 80% plants death (Table 1). When treated with P.A and P.O, plant death rate is reduced to 34% indicating that P.O can suppress P.A effectively.

Physiological changes may produce in response to infection by the pathogen. In order to understand the disease resistance, some physiological properties including SOD, CAT, MDA and stem liquid pH are analyzed. The data indicate that physiological processes are impacted when fungus is introduced. CAT and MDA have an equable change (Fig. 3 a and b), while the activity of SOD in plants inoculated with P.O and P.A is doubled compared with that of P.A treatment (Fig. 3 c). These results manifest that SOD plays the key role in response to disease caused by P.A. The result can confirm the oxidative burst phenomenon when treated with CWP of P.O [6]. The pH of stem is very important for root to shoot communication. The pH of stem liquid is slightly higher treated with fungus than control and might have certain relationships with disease resistance.

4.2 Promoting growth

In this study we also examine the ability of P.O to promote wheat growth. First, DI of wheat treated with P.O is zero, while DI of control is 3% (Table 1).

The data indicates that P.O can reduce plant death rate. Productive tiller number (PTN) is a very important component of grain yield in wheat [7] and rice. The data verifies that P.O treatment can remarkably enhance tiller rate (Table 1) suggesting the potential ability to increase wheat yield. The plants inoculated by P.O have the tallest plant height and significant biomass production. These results indicate that P.O colonization rhizosphere can promote wheat growth. P.O increasing plant growth is mediated by the fungus-produced auxin compound, tryptamine [9]. So, P.O is a promising biocontrol strain in agro-ecological.

ACKNOWLEDGMENT

This research was supported by International S&T Cooperation Program of China (2013DFA30950), National Science and Technology Support Program of China (2012BAD14B10), National Spark Program of China under Grant (2010GA860003), National Natural Science Foundation of China Youth Science Foundation (61201421), Livelihood Fund Projects of Gansu Province (1209FCMA003), Youth Science and Technology Fund Projects of Gansu Province (1208RJYA058), Youth Science and Technology Fund Projects of Gansu Academy of Sciences Fund (2014QN-03) and Development & Application of Gansu Academy of Sciences Fund (2012JK-03).

REFERENCES

- [1] Drechsler, C. (1936). New Conidial Phycomycetes Destructive to Terricolous Amoebae. *Mycological Society of America* Vol. 28, No. 4 363–389.
- [2] Martin, F. N., Hancock, J.G. (1987). The use of *Pythium oligandrum* for biological control of preemergence damping-off caused by *P. ultimum*. *Phytopathology* 77: 1013–1020.
- [3] Akira Masunaka, K. N., Masao Sakai, Hideki Takahashi, Shigehito Takenaka (2009). Visualization of *Ralstonia solanacearum* cells during biocontrol of bacterial wilt disease in tomato with *Pythium oligandrum*. *J Gen Plant Pathol* 75:281–287.
- [4] Le Floch, G., N. Benhamou, et al. (2006). *Pythium oligandrum* associated with two antagonistic fungi: Rhizosphere compatibility and tomato grey mould biocontrol. *Phytopathology* 96(6):S65–S65.
- [5] Takenaka, S. and H. Tamagake (2009). Foliar spray of a cell wall protein fraction from the biocontrol agent *Pythium oligandrum* induces defence-related genes and increases resistance against *Cercospora* leaf spot in sugar beet. *Journal of General Plant Pathology* 75(5): 340–348.
- [6] Le Floch, G., N. Benhamou, et al. (2005). Characterisation of the early events in atypical tomato root colonisation by

- a biocontrol agent, *Pythium oligandrum*. *Plant Physiol Biochem* 43(1):1–11.
- [7] Borrás-Gelónch, G., G. J. Rebetzke, et al. (2012). Genetic control of duration of pre-anthesis phases in wheat (*Triticum aestivum* L.) and relationships to leaf appearance, tillering, and dry matter accumulation. *Journal of Experimental Botany* 63(1):69–89.
- [8] Schroeder KL, O. P., Tambong JT & Paulitz TC (2006). Identification and quantification of pathogenic *Pythium* spp. from soils in eastern Washington using real-time polymerase chain reaction. *Phytopathology* 96:637–647.
- [9] Gaétan Le Floch, P. R., Emile Benizri, Nicole Benhamou, Yves Tirilly (2003). Impact of auxin-compounds produced by the antagonistic fungus *Pythium oligandrum* or the minor pathogen *Pythium* group on plant growth. *Plant and Soil* 257:459–470.

Hyperspectrum detection for nitrogen content of apple leaves

J. An, G.Q. Yao & H.L. Deng

Institute of Information Engineering, China University of Geosciences, Beijing, China

ABSTRACT: The nitrogen content of leaves is an important indicator for estimating growth status of apple tree. Using ASD FieldSpec 3 spectrometer to measure the spectral reflectance for apple leaves. We constructed multiple regression analysis of correlation between nitrogen content of apple tree leaves and their transformation forms. Next, we selected four more sensitive wavebands and constructed the retrieval model for nitrogen content of apple leaves using BP artificial neural network algorithm. Finally, we optimized and tested the model. The results showed that it is an effective way to improve the ability of predicting apple tree nitrogen content based on BP artificial neural network algorithm.

1 INTRODUCTION

Quantitative information on plant nutrient concentration is necessary to manage nutrient applications in real time. Nitrogen is the main component which makes up organic compound such as protein, chlorophyll and enzymes. It not only is a component of the plant, but is also involved in the multiple biological processes. It affects the metabolic activity and growth of fruit trees in many ways directly or indirectly^[1], therefore, it is the essential nutrients for growth of fruit tree, fruit production and fruit quality^[2]. The nitrogen content of leaves is an important indicator for estimating growth status of apple tree. Although the traditional chemical analysis has higher accuracy in monitoring the nutrient content of apple tree, it is limited by high consumption and complexity. Therefore, it is not able to diagnose the fruit tree nutrition, rapidly in large range. Hyperspectral as a real time, fast, non-destructive and environmental technology, plays an important role in the plant nutrient element acquisition, etc. Detection of nitrogen content is an important application of remote sensing. Due to the fact that the hyperspectrum is extremely sensitive to nitrogen content and chlorophyll content in the plant, the change of nitrogen content in the leaves affects the spectral reflectance inevitably^[3]. As a result, using the spectral information of leaves, we can estimate the content of nitrogen in it.

As early as 1972, through the study Thomas^[4] found that N content in bell pepper leaf was highly correlated with spectral reflectance in 550 ~ 675 nm band of the leaf. The deviation between the measured value and predicted value of the leaves is less than 7 percentage. The study showed that spectral analysis can provide a rapid, easy or accurate method to detect

the N content of plants. The research by Yan zhu etc^[5] showed that the reflectance at 610 nm and 680 nm waveband of rice canopy had a good correlation with nitrogen content in rice leaf. They raised that the single regression coefficient can improve the precision of estimation of nitrogen content in rice leaf.

The influence caused by environment background and instrument makes the target spectrum contain noise, and this research adopted various methods to eliminate those errors. In order to improve the accuracy of spectral information extraction, we need to preprocess spectrum data mainly with better classing. In this research, the spectrum and nitrogen content of apple tree leaves were detected by ASD for multiple regression analysis and BP artificial neural network, realizing higher precision of detection of nitrogen content in apple tree leaves. This research aims at studying and exploring the models to predict apple leaf nitrogen content in different physiological and phonological phases with high accuracy which can provide a theory basis for quick apple nutrition diagnosis and growing monitoring.

2 METHODS AND MATERIALS

2.1 Experiment

Samples were collected from the six apple orchards of Gaodu and Menyin Town in Shandong province.

Menyin county is located in Shandong province, China's eastern coastal city, which is one of the primary apple bases. This region accounts for about one-fourth of the total apple-barley producing areas in China. Menyin has the monsoon climate of warm temperate zone, with abundant sunshine and rainfall and four clearly distinct seasons. Its annual temperate

averages 12.8°C and the annual average precipitation is 820 mm.

The apple breed is called red Fuji, and the soil type is brown earth. The samples were collected on September 23, 2009. The sampling points were selected on the basis of land utilization map and orchard landscape. 86 apple trees were selected randomly from different orchards containing different growth situations. 1–2 leaves were selected from the canopy periphery in four directions, and these were for all the subsequent experiments. We putted the apple tree leaves into some fresh-keeping bags and ensured proper labeling, coding and sealing, and then brought it back to the lab rapidly.

2.2 Spectrum determination

The spectral reflectance was detected by the ASD FieldSpec 3 spectrometer manufactured in the USA. Its spectral range is 350–2500 nm, and its spectral resolution is 1.4 nm in the 350–1000 nm range and 2nm in the range of 1000–2500 nm. The spectra were measured under controlled light condition in a dark-room. When spectra was measured, the instrument was fitted with a 25 field of view lens and held in a nadir orientation 0.1 m above canopies. Before data collection, we need to make corrections by using the standard whiteboard (reflectivity of whiteboard is 1). Each sample was measured for 10 times, then the mean value was calculated as the relative reflectance.

2.3 Nitrogen content detection

Firstly, they were put into the oven at 80°C to pre-dry for 15–30 minutes, then the oven temperature was turned down to 60°C, and the samples would be dried for 24 hours. Secondly, the dried samples were ground into powder by using a mortar, subsequently concentrated $H_2SO_4-H_2O_2$ was added to take nitration. Finally, we detected Nitrogen content in apple tree leaves by Flame photometry.

2.4 Methods

The artificial neural network uses the methods available in physics or the computer to imitate the same structure and function of the neural network in the biology, and apply it to the engineering field. Because of its characteristics of self-organizing, self-learning, strong fault tolerance, distributed storage and parallel processing, which is more similar to the people, the neural network technology is widely applied in remote sensing image recognition [6, 7, 8, 9]. The topological structure of neural network model includes input layer, hidden layer and output layer, the quality of the network depends on

the network structure, transfer function, number of hidden layers, number of neurons and the training function [10].

The BP neural network is regarded as the nonlinear mapping from input variables to outputs [11], that is

$$F: R^n \sim R^m, f(x) = Y \quad (1)$$

Therefore, this study used BP artificial neural network algorithm to establish a retrieval model of apple leaves.

R^2 (correlation coefficient) and RMSE(root-mean-square error) are two parameters for measuring quality of models, using the following Eq.(2)

$$R^2 = \frac{\left[\sum_{i=1}^n (y_{ai} - \bar{y}_a)(y_{mi} - \bar{y}_m) \right]^2}{\sum_{i=1}^n (y_{ai} - \bar{y}_a)^2 \cdot \sum_{i=1}^n (y_{mi} - \bar{y}_m)^2} \quad (2)$$

Where R^2 stands for correlation between the measured value and predictive value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{ai} - y_{mi})^2}{N - 1}} \quad (3)$$

Where y_{ai} , y_a , y_{mi} , \bar{y}_m and N are measured, mean measured, predictive, mean predictive value respectively. While RMSE stands for the deviation between predictive value and measured value.

2.5 Spectral data processing (Input variables)

We used spectrum processing software ViewSpec Pro 5.0 to deal with the measured spectral reflectance of apple leaf and applied EXCEL and SPSS software for further analysis. In order to reduce the various influences, we have transformed original spectral reflectance R into various forms, which is as shown in Figure 1.

In the figure 1, $1/R$, $\ln R$, (R) , $(\ln R)'$ and $(1/R)'$ benefit R's reciprocal, logarithm, derivative, derivative of R's logarithm, derivative of R's reciprocal respectively.

We selected four more sensitive wave bands' derivative from these transformation forms above for SPSS's multiple regression analysis to establish characteristic spectral parameters, which is as showed in

$$y = 2.328 + 51.141X_1 + 0.978X_2 - 90.456X_3 - 273.488X_4 \quad (4)$$

x_1, x_2, x_3, x_4 benefit the derivative of 364nm, 373nm, 392nm, 998nm waveband respectively, while y is the

nitrogen content. The unit is $\text{mg}\cdot\text{kg}^{-1}$. The derivative of the four wavebands can predict nitrogen content accurately, which is as shown in Figure 2. Then BP neural network model was established for nitrogen content with the derivatives of the four wavebands as input.

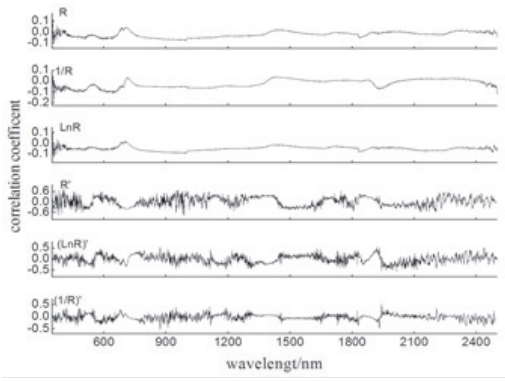


Figure 1. Correlation between reflectivity(R) or its transformation and nitrogen content of apple leaves.

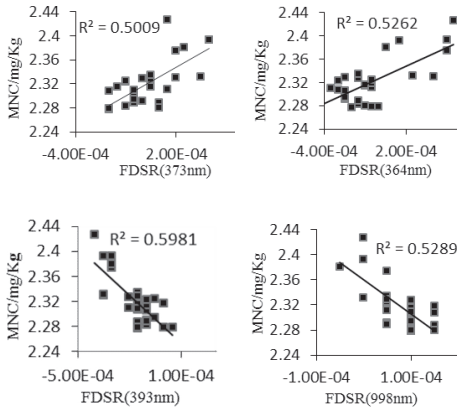


Figure 2. Correlation between first-order derivatives of four wavebands and nitrogen content of apple leaves.

Where MNC is measured nitrogen content and FDSR is first-order derivatives of spectral reflectance.

2.6 Constructing model

Leaf samples were divided into two groups, 19 samples were training group and the remaining 9 samples were validation group. The number of hidden layers of neural network model neurons was set to 10. The structure of the BP model was 4-10-1, which got an ideal result, as is shown in Figure 3(a). Nine sets of

data were used for testing training results of the BP neural network. The derivatives of four bands in the training group were regarded as the inputs for predicting the measured values of nitrogen content, as shown in Figure 3(b). The training R^2 of the BP neural network model reached 0.8878, and the RMSE was 0.0142. The validation R^2 reached 0.9001, and the RMSE was 0.0150.

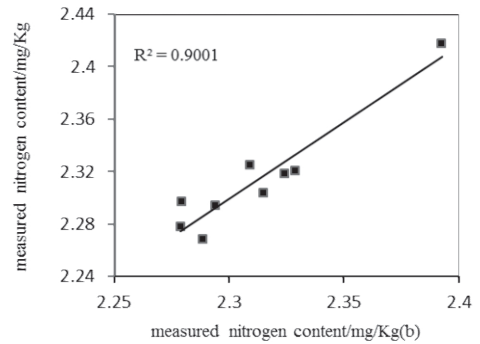
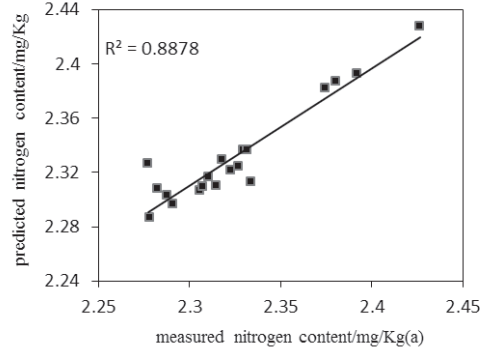


Figure 3. Nitrogen prediction models with BP neural network model.

2.7 Discussion

Stepwise polynomial regression was used for predicting nitrogen content with 28 samples, as it is shown in Figure 4. The R^2 of the model reached 0.7163, and the RMSE was 0.0222. The four wavebands were used as inputs for constructing the model.

Contrasting BP neural network model with multiple regression model, the former one which is established got the higher modeling and predicting accuracy than the latter one. When the training R^2 of the BP neural network model reached 0.8878, the RMSE was 0.0142. And when the validation R^2 reached 0.9001, the RMSE was 0.0150. In contrast, when the R^2 of the multiple regression model only reached 0.7163, the RMSE was 0.0222. It indicated that a BP neural network model had the highest accuracy in predicting

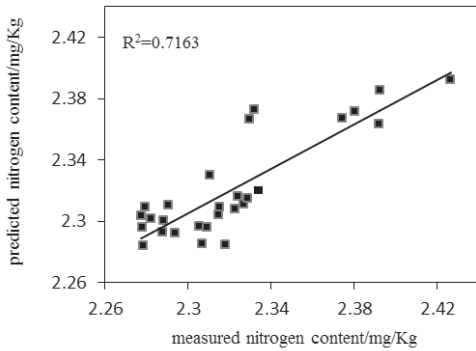


Figure 4. Nitrogen prediction models using stepwise polynomial regression.

apple leaf nitrogen content and could be regarded as the best estimation model in this area.

3 CONCLUSION

This research aimed at exploring high accuracy models to forecast nitrogen content in the apple tree leaves using spectra. The BP neural network model was established with derivatives of original spectrum, which got the highest precision and lowest deviation. The optimal structure of BP neural network model is 4-10-1. When the training R^2 of the BP neural network model reached 0.8878, the RMSE was 0.0142, and when the validation R^2 reached 0.9001, the RMSE was 0.0150. Compared with predecessors, the methods adopted in this research have a great improvement. And the accuracy of the best model illustrated that it was possible to improve the spectra prediction ability of apple tree leaf nitrogen content using a BP neural network algorithm.

REFERENCES

[1] Broge N H, Mortensen J V. Deriving green crop area index and canopy chlorophyll density of winter wheat

from spectral reflectance data[J]. *Remote Sensing Environment*, 2002, 81(1):45–57.

[2] Z.J. Qiu, H.Y. Song, Y. He, et al. Variation rules of the nitrogen content of the oil seed rape at growth stage using SPAD and visible-N IR [J]. *Transactions of the CSAE*, 2007, 23(7):150–154. (in Chinese with English abstract).

[3] Y. Wang, J.F. Huang, F.M. Wang, et al. Predicting nitrogen concentrations from hyperspectral reflectance at leaf and canopy for rape. *Spectroscopy and Spectral Analysis*, 2008, 28(2):273–277 (in Chinese).

[4] Thomas J R, Oerther G F. Estimating nitrogen content of sweet pepper leaves by reflectance measurements[J]. *Agronomy Journal*, 1972, 64:11–13.

[5] Y. Zhu, Y.X. Li, D.Q. Zhou, Y.C. Tian, X. Yao, W.X. Cao et al. Quantitative relationship between leaf nitrogen concentration and canopy reflectance spectra in rice and wheat[J]. *Acta Ecologica Sinica*, 2006, 26(10): 3463–3469 (in Chinese with English abstract).

[6] K.S. Song, B. Zhang, Z.M. Wang, et al. Inverse model for estimating soy bean chlorophyll concentration using insitu collected canopy hyperspectral data [J]. *Transactions of the CSAE*, 2006, 22(8):16–21. (in Chinese with English abstract).

[7] X.G. Tang, D.W. Liu, K.S. Song, B. Zhang, G.J. Jiang, F. Yang, J.P. Xu. A Study for Estimating the Main Tree Species Leaf Area Index in Northeast Based on Hyperspectral Data[J]. *Remote Sensing Technology and Application*, 2010 (3):334–341 (in Chinese with English abstract).

[8] F.Q. Yao, Z.H. Zhang, R.Y. Yang, Y.C. He, Y. Zhang, S.F. Cui, H.J. Wang. Application of mobile collocation method in the interpolation of height abnormal[J]. *Science of Surveying and Mapping*, 2010 (1):109–112 (in Chinese with English abstract).

[9] Lubac B, Loisel H. Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea[J]. *Remote Sensing of Environment*, 2007, 110(1):45–58.

[10] G.G. Shen, R.C. Wang. Study on Bi-directional Reflectance Model of Rice Using a Artificial Neural Network[J]. *Journal of Remote Sensing*, 2002, 6(4): 252–258.

[11] P. Wang, X.G. Liu, F. Huang. Retrieval Model for Subtle Variation of Contamination Stressd Maize Chlorophyll Using Hyperspectral Data[J]. *Spectroscopy and Spectral Analysis*, 2010 (1):197–201.

Effect of heavy metals on growth and metal accumulation of an energy crop, *Pennisetum purpureum* schum

C. Tian, X.F. Zhang, Y.H. Su, P.H. Tan, H.F. Zhang & G.Z. Zhang

Guangxi Scientific Experiment Center of Mining, Metallurgy and Environment, Guilin University of Technology, Guilin, PR China

College of Environmental Science and Technology, Guangxi Key Laboratory of Environmental Pollution Control Theory and Technology, Guilin University of Technology, Guilin, PR China

ABSTRACT: An experiment was conducted to evaluate the effect of Cu, Cd, Ni, As, Zn, Pb on growth, and metal accumulation of an energy crop, *Pennisetum purpureum* Schum. The dry weight was significantly decreased under the heavy metal pollution, while the shoot length was not suppressed significantly in the treatments of Cu, As, Pb. The concentration of Cu, Cd, Ni, As, Zn, Pb in plant shoot were 4.7~10.43, 5.37~11.07, 59.53, 0.08~9.33, 20.73~310.62, 37.77~194.94 mg/kg, respectively. The BCF of *P. purpureum* in all treatments were all lower than 1. *P. purpureum* could produce 15.35-24.03 t/ha of dry biomass under Cu, As, Pb pollution each year. The study results indicated that *P. purpureum* could be cultivated as an energy crop in Cu, As, Pb -contaminated soils and might be useful for the phytostabilization of this pollution soil.

KEYWORDS: Heavy metals, *Pennisetum purpureum* schum, energy crop, phytoremediation.

1 INTRODUCTION

Recently, energy crisis and heavy metal pollution are becoming increasingly severe worldwide [1, 2]. Cultivating energy crops in metal-contaminated lands which are not suitable for agriculture can achieve a win-win situation for biomass production and soil remediation [3]. *Pennisetum purpureum* Schum, belonging to the gramineous family, is a potential energy crops in many reports [4, 5]. It is a perennial forage grass that is widely cultivated in tropical and subtropical areas in the world. It is a fast growing and high yielding grass with the 4 m shoot length. Obviously, if *P. purpureum* could be cultivated in heavy metal-contaminated soils, it would save a large amount of clean land for agricultural use. However, its performance in tolerance and accumulation characteristics of heavy metals in polluted soils was few to be documented. Therefore, the objectives of the present study are to evaluate the effect of metal pollution on the growth and metal accumulation of *P. purpureum*, and assess its biomass production.

2 MATERIALS AND METHODS

The pot culture experiment was conducted in the greenhouse at Guilin University of Technology of China. Thirteen treatments (0, 95mg/kg Cu, 190mg/kg Cu, 6mg/kg Cd, 12mg/kg Cd, 105mg/kg Ni, 210 mg/kg Ni,

27.5mg/kg As, 55 mg/kg As, 360mg/kg Zn, 720 mg/kg Zn, 265mg/kg Pb, 530mg/kg Pb) with 4 replicates were set up and designated as CK, Cu95, Cu190, Cd6, Cd8, Ni105, Ni201, As27.5, As55, Zn360, Zn720, Pb265, Pb 530 respectively. The soil metal concentration was set up according to the soil environmental quality standard of Holland. Top soil (0–20 cm) was collected from a local abandoned garden. Dried soil (5 kg) sieved through 1 cm mesh was put into each plastic bucket (25 cm in diameter and 20 cm in height), and then saturated with a heavy metal solution containing the required amount of CdCl₂·2.5H₂O, CuSO₄·5H₂O, NiSO₄·H₂O, K₂Cr₂O₇, Na₂HAsO₄·7H₂O, ZnSO₄·7H₂O, and Pb(NO₃)₂, respectively. One month later, 3 uniform-sized plants were planted into each plot. The plants were obtained by cutting stem nodes of *P. purpureum* from a plantation in Pinghui Animal husbandry Company in Fujian, China, and then were rooted in soil. Ten days later, when the cuttings had grown to 10–12 cm height and were used for transplanting. All pots were watered each day. After 90 days of cultivation, the shoot length was determined, and then cut and washed with tap water, followed by deionized water. Then all the samples were dried at 75°C until completely dry, weighed, ground to <0.5 mm, and sealed into plastic bags for analyses.

Bioaccumulation factor (BCF) was used to evaluate phytoremediation efficiency. It is defined as ratio of plant concentrations of heavy metals in different parts

to the soil metal concentrations. Statistical analysis was performed using SPSS 16.0 statistical software. Data were analyzed by one-way ANOVAs using the Least Significant Difference (LSD) to compare significant differences between treatments ($P < 0.05$).

3 RESULTS AND DISCUSSION

Generally, shoot length and biomass are often used to evaluate the tolerance of plants [6, 7]. In this study, though the dry weight were significantly decreased in the heavy metal pollution, the shoot length was not suppressed significantly in the treatments of Cu95, Cu190, As27.5, As55, Zn360, Pb265, Pb530 (Figure 1 and 2). *P. purpureum* had a rapid growth rate with the dry weight of 25.1~39.3g/pot and the shoot length of 109~194cm in the treatments of Cu, As, Pb pollution. The results indicated that *P. purpureum* had high tolerance to Cu, As and Pb pollution, while was severe toxic by the Zn, Cd and Ni pollution. Cadmium concentrations in different plant tissues are shown in Table 1. The concentrations of Cu, Cd, Ni, As, Zn and Pb in *P. purpureum* increased with increasing metal concentrations in soil, which were 4.7~10.43, 5.37~11.07, 59.53, 0.08~9.33, 20.73~310.62, 37.77~194.94 mg/kg in stems and leaves, respectively. Meanwhile, the BCF of *P. purpureum* were all lower than 1 which were 0.003~0.923 (Table 2).

The ideal plant used for phytostabilization should have a high biomass and low bioaccumulation factor [8]. In this study, *P. purpureum* had high biomass and low BCF, which indicated that it might be useful for the phytostabilization of soil Cu, As, Pb pollution.

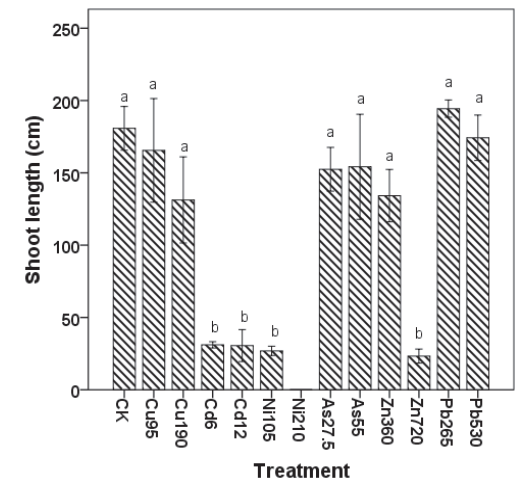
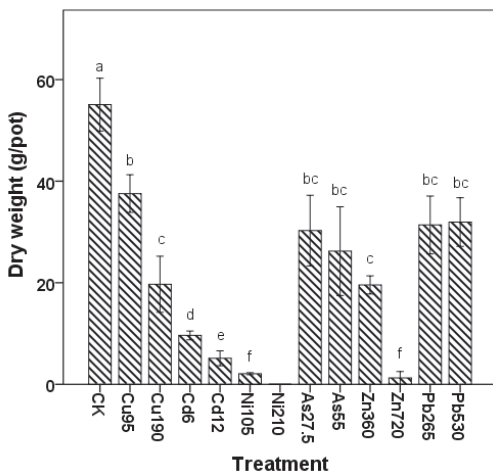


Figure 2. Cadmium-induced changes in the shoot length of *P. purpureum*.

Table 1. Heavy metal concentration in *P. purpureum* under pot culture conditions.

| Metals | Treatments | Stem (mg/kg) | Leaf (mg/kg) |
|--------|------------|----------------|----------------|
| Cu | CK | 2.84±0.81* b | 6.57±1.39 b |
| | Cu95 | 6.66±1.83 a | 9.02±2.66 ab |
| | Cu190 | 4.70±1.32 ab | 10.43±2.58 a |
| Cd | CK | 0.18±0.05 | 0.04±0.02 c |
| | Cd6 | - | 5.37±0.24 b |
| | Cd12 | - | 11.07±0.78 a |
| Ni | CK | 0.53±0.23 | 1.81±0.36 b |
| | Ni105 | - | 59.53±37.81 a |
| | Ni210 | - | - |
| As | CK | 0.002±0.0008 b | 0c |
| | As27.5 | 0.08±0.08 a | 2.06±0.09 b |
| | As50 | 0.47±0.15 a | 9.33±3.31 a |
| Zn | CK | 14.96±3.96 b | 20.73±13.73 b |
| | Zn360 | 310.62±72.25a | 124.64±30.89 a |
| | Zn720 | - | - |
| Pb | CK | 0 b | 0.78±0.78 b |
| | Pb265 | 37.77±11.52 a | 194.94±49.34 a |
| | Pb530 | 49.91±6.06 a | 184.37±50.61 a |

Figure 1. Cadmium-induced changes in the dry weight of *P. purpureum*.

* Mean ± S.E.; “-” means no datas; data followed by the same letters in the same column for the same metal indicate no significant difference at $P = 0.05$ level by LSD test.

Table 2. Cadmium concentration and BCF in *P. purpureum* under pot culture conditions (“-” means no datas).

| Treatments | BCF of stem | BCF of leaf |
|------------|-------------|-------------|
| Cu95 | 0.070 | 0.095 |
| Cu190 | 0.025 | 0.055 |
| Cd6 | - | 0.895 |
| Cd12 | - | 0.923 |
| Ni105 | - | 0.567 |
| Ni210 | - | - |
| As27.5 | 0.003 | 0.075 |
| As50 | 0.009 | 0.187 |
| Zn360 | 0.863 | 0.346 |
| Zn720 | - | - |
| Pb265 | 0.143 | 0.736 |
| Pb530 | 0.094 | 0.348 |

Table 3. Biomass productivity of *P. purpureum* (It is assumed that the *P. purpureum* may be mowed 3 times each year).

| Treatments | Shoot dry biomass (t ha ⁻¹ for each year) |
|------------|---|
| CK | 33.69 |
| Cu95 | 22.99 |
| Cu190 | 15.35 |
| Cd6 | 5.87 |
| Cd12 | 4.16 |
| Ni105 | 1.35 |
| Ni210 | - |
| As27.5 | 18.53 |
| As50 | 24.03 |
| Zn360 | 11.98 |
| Zn720 | 1.53 |
| Pb265 | 22.32 |
| Pb530 | 22.32 |

In fact, based on the data from the present experiment (Table 3), on the basis of each year, *P. purpureum* could produce 15.35-24.03 t/ha of dry biomass under

Cu, As, Pb pollution. The study results indicated that *P. purpureum* could be cultivated as an energy crop in Cu, As, Pb -contaminated soils.

ACKNOWLEDGMENTS

The project was supported by the Natural Science Foundation of Guangxi (2013GXNSFB019026), the Special Funding for Guangxi ‘BaGui scholars’ Construction Projects, and the Guangxi Talent Highland for Hazardous Waste Disposal Industrialization.

REFERENCES

- [1] Weyens N, van der Lelie D, Taghavi S, Newman L, Vangronsveld J. Exploiting plant-microbe partnerships to improve biomass production and remediation. *Trends in microbiology*, 2009, 27(10):591-8.
- [2] Kumar N, Baudhh K, Kumar S, Dwivedi N, Singh DP, Barman SC. Accumulation of metals in weed species grown on the soil contaminated with industrial waste and their phytoremediation potential. *Ecological Engineering*, 2013, 61:491-495.
- [3] Zhang XF, Zhang XH, Gao B, Xia HP, Li ZA, Li Haifang, Li Jian. Effect of cadmium on growth, photosynthesis, mineral nutrition and metal accumulation of an energy crop, king grass (*Pennisetum americanum* × *P. purpureum*). *Biomass & Bioenergy*, 2014, 67:179-187.
- [4] Guo XH, Yu DS, Wang J, Tang DY, Liu XM. Comparative Study of the Lignocellulose Content in Six Plants. *Journal of Hunan University (Natural science)*, 2008, 35(9):76-781. (In Chinese).
- [5] Li P, Sun XL, Han JG, Liu TM. A New Source of Bio-energy—Energy Grass. *Chinese Journal of Grassland*, 2010, (5):97-100. (In Chinese).
- [6] Gonzaga MIS, Santos JAG, Ma LQ. Phytoextraction by arsenic hyperaccumulator *Pteris vittata* L. from six arsenic contaminated soils: repeated harvests and arsenic redistribution. *Environmental Pollution* 2008, 154(2):212-8.
- [7] Zhang XF, Gao B, Xia HP. Effect of cadmium on growth, photosynthesis, mineral nutrition and metal accumulation of bana grass and vetiver grass. *Ecotoxicology and Environmental Safety*, 2014, 106:102-108.
- [8] Marques APGC, Rangel AOSS, Castro PML. Remediation of heavy metal contaminated soils: phytoremediation as a potentially promising clean-up technology. *Critical Reviews in Environmental Science and Technology*, 2009, 39:622-654.

Research in disproportionation reaction mechanism of methyltrichlorosilane and trimethylchlorosilane catalyzed by modified catalyst $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$

W.Y. Xu, J.X. Zhou, Y.P. Liu & H.H. Wan

College of Science, East China Jiaotong University, Nanchang, China

S.G. Hong

College of Science, University of Nan Chang, Nanchang, China

ABSTRACT: Disproportionation reaction mechanism of methyltrichlorosilane and trimethylchlorosilane catalyzed by modified catalyst $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ was calculated by DFT(B3LYP/LANL2DZ) method. The conclusions are following: the reaction process which modified catalyst $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ participated in mainly was divided into main and side reactions. The rate-determining steps of main reaction and side reaction were respectively the first and the third reactions. The active energies of each step were $E_{a1}=272.3956\text{kJ/mol}$, $E_{a2}=207.7569\text{kJ/mol}$, $E_{a3}=330.1417\text{kJ/mol}$, $E_{a4}=184.0111\text{kJ/mol}$ respectively. The reaction energy barrier of the first step reaction was high, and main reaction occurred difficultly in the catalytic effect of modified catalyst. The reason is that the Lewis acid of the Al position on catalyst decreased, resulting in the catalytic activity becoming small, which is the disadvantage of the disproportionation reaction.

1 INTRODUCTION

ZnCl_2 , as Lewis acid, has gotten more and more attention in recent years^[1]. JF Wu^[2] et al. calculated the reaction mechanism of activation and conversion of methane catalyzed by Zn modified H-ZSM-5 catalyst, the result of which was that the Lewis acid can strengthen after introducing Zn^{2+} on zeolite and the reaction can carry on at room temperature. Some researches report^[3] the Bronsted acid amount of ZSM-5 zeolite decreased at the same time the Lewis acid amount increased because of Zn modified. The cracking activity of Zn/ZSM-5 catalyst decreased, but the dehydrogenation ability strengthened, which led to the result that the amount of aromatic hydrocarbon liquid phase increased obviously, promoting hydrocarbon aromatization. Xi-ping Luo^[4] et al found that the amorphous substance ZnCl_2 and the supporter formed by synergistic reaction can improve the catalytic activation of catalyst. The reference^[5] indicated that ZnCl_2 had a better catalytic effect of the disproportionation reaction of methyltrichlorosilane and trimethylchlorosilane. Our previous experiments^[6] have calculated the mechanism of the disproportionation reaction prepared dimethyldichlorosilane that ZnCl_2 participated in and we had done the experiment research. We can get the conclusion that ZnCl_2 has some catalytic effect of the disproportionation reaction by contrasting the activate energies and the yields of dimethyldichlorosilane got through experiments

that $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ and AlCl_3 catalysts participated in the reaction but the catalytic effect of $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ catalyst was weaker than AlCl_3 . Thus we can imagine that if we modify the ZSM-5 catalyst that has supported $-\text{AlCl}_2$ group by ZnCl_2 , the catalytic effect of catalyst may increase. So the paper reports the disproportionation reaction mechanism that ZnCl_2 modified catalyst participates in the reaction. From the reaction activation energy, molecular configuration et al facets, we explore the catalytic effect of catalyst.

2 COMPUTATIONAL DETAILS

The geometries for all the reactants, products, intermediates and transition states were fully optimized by density functional theory at B3LYP/LANL2DZ levels, of which energies were corrected by the secondary perturbation MP2/6-31++G** method. And the geometries for transition states were fully optimized by the eigen vector tracking method AM1 firstly, and then further optimized by density functional theory(B3LYP/LANL2DZ), certificated through vibration analysis and calculated through IRC, finally the energies were corrected by the secondary perturbation MP2/6-31++G** method again. All the calculations have been done by the GAUSSIAN 09 program (Rev B01) in computing workstation of East China Jiaotong University.

3 RESULTS AND DISCUSSIONS

The reaction process of disproportionation reaction of methyltrichlorosilane and trimethylchlorosilane that $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ catalyst participated in were mainly divided into five-step reactions. The first step reaction is that trimethylchlorosilane reacted to $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ catalyst firstly, generating intermediate I ($\text{CH}_3\text{AlCl}_2/\text{ZnCl}_2\text{-ZSM-5}$) and dimethyldichlorosilane. The second step reaction is that methyltrichlorosilane reacted to intermediate I through transition state TS2, reducing to $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ catalyst and producing the main product dimethyldichlorosilane. The third step reaction is the rate-determining of the main reaction that the activation energy is 272.3956kJ/mol; the fourth step reaction is that methyltrichlorosilane reacted to $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ catalyst by functional group interconversion firstly, yielding the intermediate I ($\text{CH}_3\text{AlCl}_2/\text{ZnCl}_2\text{-ZSM-5}$) and chemical byproduct silicon tetrachloride, then trimethylchlorosilane reacted to I ($\text{CH}_3\text{AlCl}_2/\text{ZnCl}_2\text{-ZSM-5}$) reducing to $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ catalyst and obtaining byproduct tetramethyl silane. The fifth step reaction is the rate-determining of side reaction whose activate energy is 330.1417kJ/mol. The intrinsic reaction coordinate of transition state (TS1~TS4), reaction enthalpies and reaction activate the energy of each step reaction are as shown in Fig 1. Fig 2 lists that the spacing trend of key atoms along IRC that $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ catalyst participates in the reaction process, which confirms the structures of the transition states and the reaction pathway respectively.

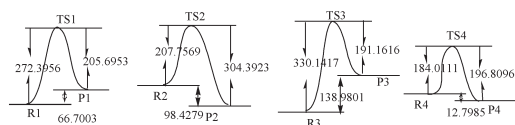


Figure 1. The intrinsic reaction curve of $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ catalyst participated in reaction (kJ/mol).

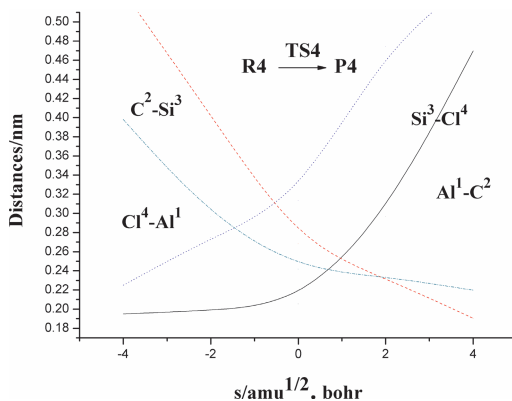
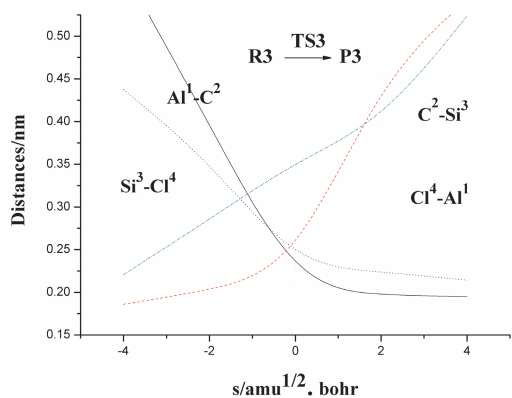
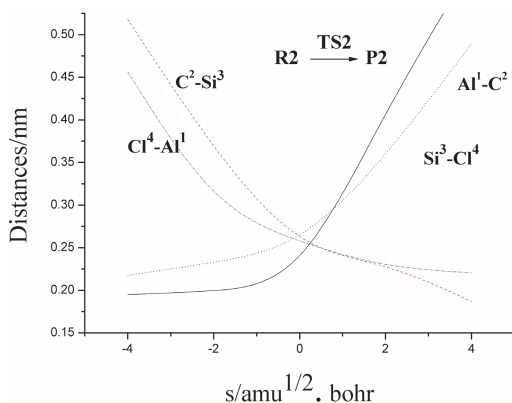
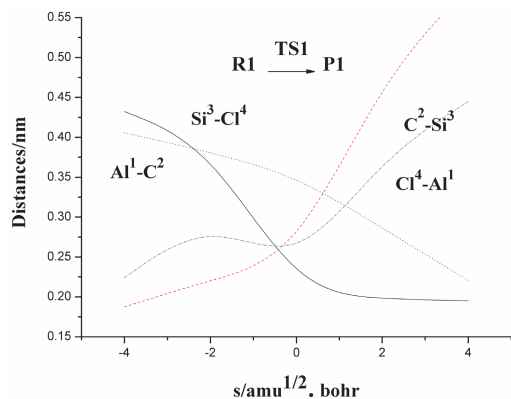


Figure 2. The spacing trend of key atoms along IRC.

4 CONCLUSION

Disproportionation reaction mechanism of Methyltrichlorosilane and Trimethylchlorosilane catalyzed by modified catalyst $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ was calculated with DFT(B3LYP/LANL2DZ) method. The conclusions are as follows: the reaction

process of modified catalyst $\text{AlCl}_3/\text{ZnCl}_2\text{-ZSM-5}$ participated in was mainly divided into main and side reactions. The rate-determining steps of main reaction and side reaction were respectively the first and the third reactions. The active energies and reaction enthalpies of each step were $\Delta H_1 = 66.7003$ kJ/mol, $\Delta H_2 = -98.4279$ kJ/mol, $\Delta H_3 = 138.9801$ kJ/mol, $\Delta H_4 = -12.7985$ kJ/mol $E_{a1} = 272.3956$ kJ/mol, $E_{a2} = 207.7569$ kJ/mol, $E_{a3} = 330.1417$ kJ/mol, $E_{a4} = 184.0111$ kJ/mol respectively. The reaction energy barrier of the first step reaction was high, and main reaction occurred difficultly, in the catalytic effect of modified catalyst. The reason is that the Lewis acid of the Al position on catalyst decreased, resulting in the catalytic activity becoming smaller, which is the disadvantage of the disproportionation reaction.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 21163005 and 31160187) and the National Natural science foundation of Jiangxi Province (No. 20132BAB203013).

REFERENCES

- [1] Topic, D.; Aschwanden, P.; Falssler, R.; Carreira, E. 2005. *M.Org. Lett.* Washington:ACS, 7, 5329.
- [2] Wu J F, Yu S M, Wang W D, 2013. Mechanistic Insight into the Formation of Acetic Acid from the Direct Conversion of Methane and Carbon Dioxide on Zinc-Modified H-ZSM-5 Zeolite *Journal of the American Chemical Society*, New York :ACS, 135(36): 13567–13573.
- [3] Xu.J.; Zheng. A.M, Wang X.M, 2012 Room temperature activation of methane over Zn modified H-ZSM-5 zeolites: Insight from solid-state NMR and theoretical calculations.*J. Chem. Sci.*,3, 2932–2940.
- [4] Biscardi.J. A.; Meitzner. G.D.; Iglesia E. 1998 Structure and density of active Zn species in Zn/H-ZSM5 propane aromatization catalysts. *J. Catal. California: Elsevier* .179, 192–202.
- [5] Zhou. CH.H, Luo X.P, Ge Z.H, LI X.N, Cai.Y, Guo HongCiang. 2003. Preparation of acid-treated clay-supported ZnCl_2 and its catalytic Properties for benzylation of benzene bybenzyl chloride. *Chinese Journal of Catalysis, Dalian :Elsevier.* 24(8): 579–584.
- [6] Xu W.Y, He Z.Y, Liu X.L, Hao W, LI F.Y, Hong S.G. 2008. Study on the Preparation of DichlorodimethylSilane by Redistribution—The Catalysis of ZnCl_2 , *Journal of Jiangxi normal university (Natural Science)*, Nanchang: (4): 462–466.

The correlation studies about histone modification and H2A. Z nucleosome positioning

Z.H. Shan, J.H. Feng, P.F. Chen, H.H. Wei & H. Hu

School of Electrical and Information Technology, Yunnan University of Nationalities, Kunming, China

ABSTRACT: In eukaryotes chromatin, H2A. Z is highly conservative histone variants and closely associated with transcriptional regulation and the stability of the genome. In order to explore the link between histone modification data, and H2A.Z nucleosome positioning data, we discussed the linear or nonlinear relationships between those data around the transcription start site in the yeast genome-wide by MIC algorithm. The results showed there are two kinds of histone modification data (H4ac and GCN4) existed in an undiscovered nonlinear relationship with nucleosome position and discussed the reasons from biological sense. **Keywords:** Cr(VI), sodium sulfite, humic acid, *L. hexandra*.

1 INTRODUCTION

In the eukaryotes, DNA and histone combine together to form a chromosome, and nucleosome is the basic structural unit of chromosomes, and it contains a core group of eight protein polymers, the structure is composed by histones H2A, H2B, H3 and H4, each protein are formed from two molecules and about 200 bp of DNA molecules wind outside of the core protein with eight polymers to form a nucleosome unit [1–2].

Histone modification is a covalent modification that under the action of related enzymes and happened in the end of the histone N. The covalent modifications include methylation, acetylation, phosphorylation and so on, and different histone modification plays a different role in the expression of genes[4]. Nucleosome precisely positioned on gene DNA molecules is called nucleosome location, it has been confirmed in the process of transcription regulation, DNA replication, and so on, which plays an important role[6].

According to the above, based on the experiment of yeast histone variants H2A. Z histone modification data, and nucleosome positioning data, we studied the association between histone modification and nucleosome position data. Different from previous studies, in addition to the linear correlation between research data, we also discussed the nonlinear correlation with the help of MINE algorithm.

2 EXPERIMENT METHODS

2.1 MIC algorithm description

MIC is used to measure the dependence of two variables, it is able to capture a wide range of connection

between two variables, including the function or not (including the linear and non-linear relationship). For the functional relation, it can get an approximate value which is equal to the sample determination coefficient. It belongs to a category which based on the nonparametric test statistics and maximizes the information. Its general idea is: if there is a relationship between two variables, first of all, draw a grid on the two variables scattering plot to part the data and encapsulate the correlation. Then, explore all of the grid to its biggest grid resolution and according to the sample size to calculate every integer (x, y) maximum mutual information. After that, normalize the mutual information value. Finally, make each group of the normalized maximum mutual information value into a matrix M (characteristic matrix), and the MIC value is the maximum of the characteristic matrix.

The Main Formula Is As Follows:

For a limited definition set with h the integer x, y, we define: $D \subset R^2$

$$I^*(D, x, y) = \max I(D | G) \quad (1)$$

X represents column, y on behalf of the line, $I(D|G)$ represents the mutual information of (D|G)

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) \quad (2)$$

Characteristic matrix $M(D)$:

$$M(D)_{x,y} = \frac{I^*(D, x, y)}{\log \min\{x, y\}} \quad (3)$$

with $xy < B(n)$:

$$MIC(D) = \max \{M(D)_{x,y} \} \tag{4}$$

N represents sample size, $B(n)$ on behalf of the network size and usually $B = n^{0.6}$, $MIC(D)$ represents the maximum information coefficient.

2.2 The data source

Our primary research data, including several parts: the nucleosome histone methylation and acetylation data of yeast, which come from Julia Zeitlinger et al [4], and the second is a histone variant H2A.Z nucleosome positioning data from Luc Gaudreau et al [17]. Due to the formation and precision of the experimental data is not unified, we carried out the necessary dispose of the original data.

2.3 Data preprocessing

In order to unite all kinds of data for 1 bp precision, we first interpolated all groups of the original data. After the comprehensive comparison of several common interpolation methods and using the atlas after the interpolation (aligned) compared with the results of thesis [11], we found a gaussian interpolation effect is good.

2.4 Data alignment

According to the position of gene data, around the TSS site of each gene, we cut the length of the upstream and downstream with the 1200 bp, respectively (The length is through trial and error and the effect is the best), then aligned the cut nucleosome positioning and histone modification data respectively, after that did the overlay, average and normalization process. Finally, we got the whole genome in the vicinity of the TSS after normalized of nucleosome positioning (Fig.1) and histone modifications map (Fig.2).

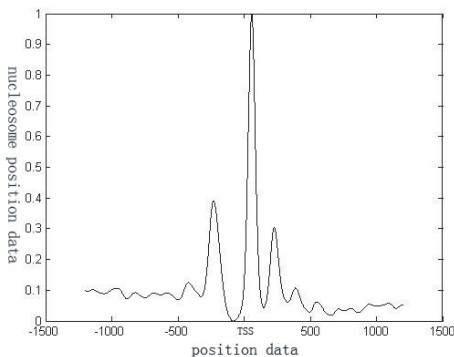


Figure 1. Near the TSS after normalized of nucleosome positioning data chart.

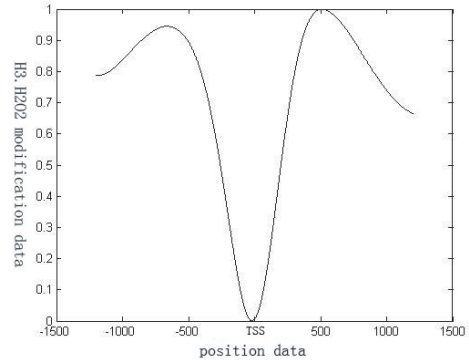


Figure 2. Near the TSS after normalized of histone modification data chart.

We got the map (Figure.1 and Figure.2) compared with the experimental data measured by Yuan GC[1] found the distribution and the trendence of data is the same, so the result of genome-wide data we got is correct. Histone modification of data (methylation and acetylation) has 28 units, because of the limited length of the article, we only lists the H3. H2O2 modification chart here.

3 RESULTS

In order to study the association of protein modification with nucleosome positioning data, we first calculated the correlation between two groups data based on the MIC algorithm and got the following bar charts (Fig.3).

From the chart (Fig.3), we found the data have a strong correlation, but the link is linear or nonlinear? for this, we got the Pearson correlation coefficient chart of the nucleosome positioning, respectively, with 27 kinds of modification data (Fig.4)

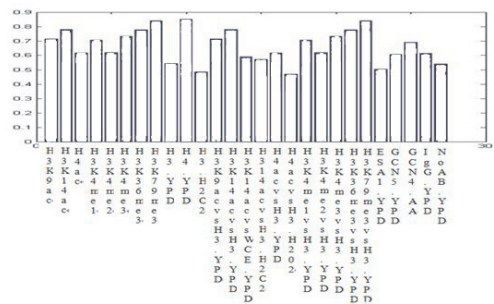


Figure 3. The MIC value bar chart of the modification and nucleosome positioning data.

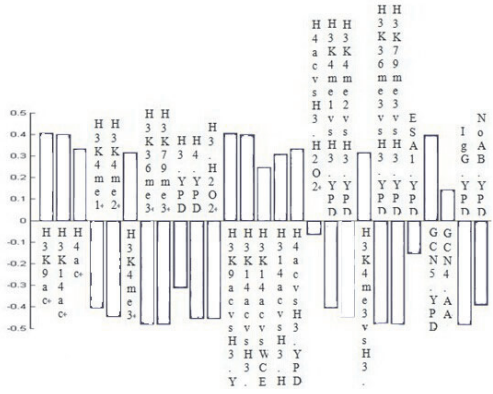


Figure 4. The Pearson correlation coefficient value bar chart of the modification and nucleosome positioning data.

We carefully compared the two different methods and found most data conform to the actuality. Such as high MIC value with high Pearson correlation coefficient (for example the first modification H3K9ac), when the two MIC value are close to, their Pearson correlation coefficients are also the same (the first modification H3K9ac and the fourth modification H3K4me1), and the low MIC value also with low Pearson correlation coefficient (the histone modification ESA1.YPD). But we found two groups of data distinction are obvious (the seventeen modification H4ac and the twenty-five modification GCN4. AA), they have high MIC value, but the Pearson correlation coefficient is very low, so we guessed there may exist an unknown nonlinear relationship between the two kinds of modifications.

Based on this, we under the whole genome and around the TSS, drew two-dimension and three-dimensional figure between the two histone modification data and nucleosome position data.

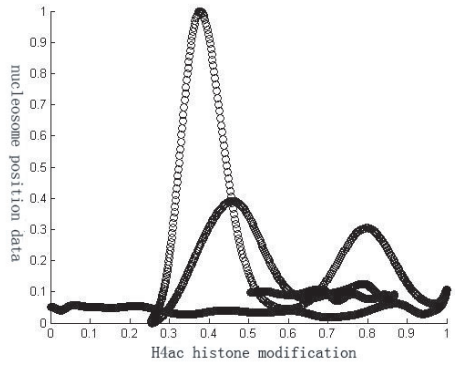


Figure 5. The two-dimension figure of H4ac modification and nucleosome positioning data.

From the above two-dimension and three-dimension figure, we found the projection of the two-dimension graphic was a positively cosine function (the local is more apparent), and its extreme value are all

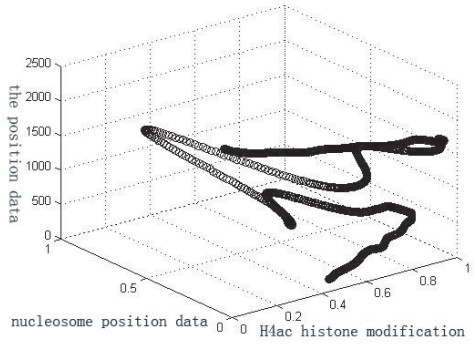


Figure 6. The three-dimension figure of H4ac modification and nucleosome positioning data.

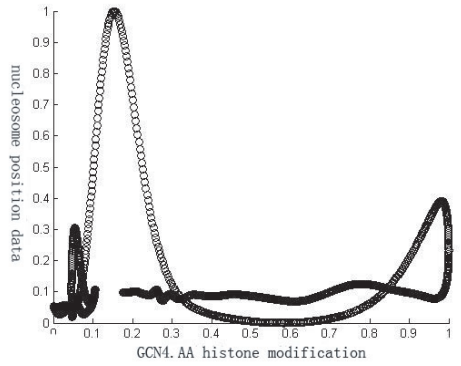


Figure 7. The two-dimension figure of histone GCN4.AA modification and nucleosome positioning data.

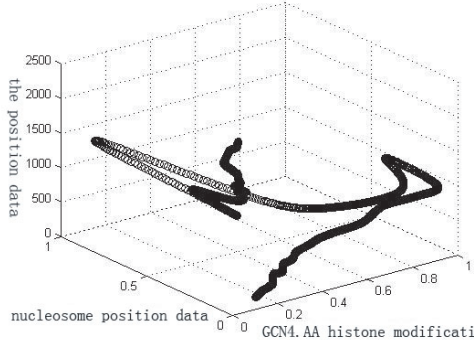


Figure 8. The three-dimension figure of GCN4.AA modification and nucleosome positioning data.

around TSS. We referred to the thesis [15] found in the yeast organisms, acetylation of histone H4 occurred at the end of several different histone lysine, its acetylation peak was over than active gene starting position and also associated with the transcription rate, and the transcription positivity. And it also can't rule out at the N end of histone H4.

4 CONCLUSION

After got the nucleosome positioning data, histone modifications data interpolated, aligned and normalized, we first calculated the correlation between the data by the MINE algorithm (including linear and nonlinear relation). And in order to define which properties of the data sets, we calculated the Pearson correlation coefficient. By comparing and combining the overall trendence of data from two-dimensional and three-dimensional chart, We found there are two kinds of histone modifications (H4ac and GCN4) located undiscovered nonlinear relationship and also discussed the reasons from biological sense.

ACKNOWLEDGMENTS

First and foremost, I would like to show my deepest gratitude to my supervisor, J.H. FENG, who has provided me with valuable guidance in every stage of the writing of this thesis. I'd also like to thank all my friends, especially my roommates, for his

encouragement and support, and I also want to thank the fund project ,the National Natural Science fund project(31160234) and the Applied Basic Research fund project(2011FB082).

REFERENCES

- [1] LIU, H. ZHUANG, Z.H. GUAN, J.H. ZHOU, S.G. 2012. Transcriptional Regulation Functions of Nucleosome Positioning: a Survey*. *Progress in Biochemistry and Biophysics* 39:843–852.
- [2] Vincent, M. Cedric, V. 2008. DNA physical properties determine nucleosome occupancy from yeast to fly. *Nucleic Acids Research* 36(11):3746–3756.
- [3] Itay, T. Judith, B. Naama, B. 2007. The pattern and evolution of yeast promoter bendability. *TRENDS in Genetics* 23(7):318–321.
- [4] Yuan, G.C. Liu, Y.J. Dion, M.F. Slack, M.D. Wu, L.F. Altschuler, S.J. Rando, O.J. 2005. Genome-scale identification of nucleosome positions in *S.cerevisiae*. *Science* 309(5734):626–630.
- [5] XING, Y.Q. LIU, G.Q. CAI, L*. 2013. An analysis and prediction of nucleosome positioning based on information content. *Chromosome Research* 21:63–74.
- [6] David, N. 2011. Detecting Novel Associations in Large Data Sets. *Science* 334:1518–1524.
- [7] Pham, T. Tran, D. 2005. Qualitatively Predicting Acetylation and Methylation Areas in DNA Sequences. *Genome Informatics* 16(2):3–11.
- [8] Peckham, H.E. 2007. Nucleosome positioning signals in genomic DNA. *Genome Res* 17(8):1170–1177.
- [9] Terry, S. 2011. A Correlation for the 21st Century. *Science* 334:1502–1503.

Molecular cloning, overexpression, purification, sequence analysis of matrix Gla protein from the Giant Panda (*Ailuropoda melanoleuca*)

P.P. Wang, X. Ding, Y.L. Hou, Y.X. Zhu, H.Q. Zhu, L. Fu & N. Zhang

College of Life Sciences, China West Normal University, China

ABSTRACT: Purposes: We have cloned the cDNA and genomic sequence of *MGP* from the Giant Panda by RT-PCR. Procedures and Methods: We extract DNA and RNA of the *MGP* from the Giant Panda. We construct a recombinant plasmid that highly expresses the Giant Panda matrix Gla protein in *Escherichia coli* and induce the bacteria to express proteins. And then we have purified the protein with the N-terminally His-tagged. Results: The result shows that the *MGP* cDNA cloned from the Giant Panda is 358 bp in size containing one open reading frame of 312 bp encoding 103 amino acids. The matrix Gla protein's estimated molecular weight is 12.18 KDa and the pI of the protein is 9.58. The genomic sequence is found to be 3075 bp long with four exons and three introns. Topology prediction shows that there are one Gla domain profile, one vitamin K-dependent carboxylation and one protein kinase C phosphory site in the matrix Gla protein of the Giant Panda (*Ailuropoda meanoleuca*). Conclusions: Alignment analysis indicates that the nucleotide sequence and the deduced amino acid sequence are highly conserved to *Canis familiaris*. The homologies for nucleotide sequences and amino acid sequences of the Giant Panda *MGP* to that of these species are 93.91 % and 96.12 %, respectively. This study has made contribution to the study of the Giant Panda and to protect this species.

1 INTRODUCTION

Matrix Gla protein, which exists in bone, cartilage, heart, kidney and lungs and other organs is a vitamin K-dependent protein. Matrix Gla protein (MGP) is an extracellular mineral-binding protein expressed in several tissues, but it only accumulates in bone and calcified cartilage under physiological conditions. Available evidence indicates that it acts as a physiological inhibitor of mineralization. MGP has inhibited the abnormal deposition of calcium ion in cartilage, blood vessels and other soft tissues. Matrix Gla protein (MGP) has been recognized as a potent calcification inhibitor and a regulator for some proteins such as bone morphogenetic protein-2 (Ran L. Et al. 2012, Sweatt A. et al. 2003).

In the past years, scientists have concentrated on classification, feeding, morphology, disease and physiology and biochemistry for the Giant Panda (*Ailuropophilic melanoleuca*). Recently, the genetics of the Giant Panda researches have developed from the cellular level to the molecular level and the researches focus on genetic diversity, parentage and physiology.

Recently, primers are designed according to the homogenous sequences of *MGP genes* from some mammals. We use an RT - PCR technique to amplify the CDNA of *MGP* gene from the total RNA of the skeleton muscle of the Giant Panda. At the same time

we analyze the characteristics of the protein sequence encoded by the cDNA and compare it with those of human and other animals.

2 MATERIALS AND METHODS

2.1 Material

Skeletal muscle is collected from a dead Giant Panda at the Wolong Conservation Center for the Giant Panda (Sichuan, China).

2.2 DNA and RNA isolation

DNA is isolated by Kits. The DNA obtained is then dissolved in TE buffer and kept at -20°C . Total RNAs are isolated from about 400 mg muscle tissue by using the Total Tissue / Cell RNA Extraction Kits (Waton Inc., Shanghai, China). The total RNAs extracted are dissolved in DEPC water and kept at -70°C .

2.3 Primer design

The PCR primers are designed by Premier 5.0, based on the mRNA sequence of *MGP* from *Homo sapiens* (NG_023331), *Bos Taurus* (AC_000162), *Canis familiaris* (NC_006609), *Mus musculus* (NC_000072), *Rattus norvegicus* (NC_005103) and

Sus Scrofa (NC_010447). The specific primers of cDNA sequence are as follows:

MGP-F: CCGAATTCATGAAGAGCCTT

MGP-R: TCAAGCTTTCATTGGCCCC

2.4 Cloning the genomic sequence of MGP

The PCR primers are the same as the *MGP-F* and *MGP-R* mentioned in the front. The genomic sequence of the *MGP* gene is amplified by using touchdown-PCR.

2.5 Construction of the expression vector and overexpression of recombinant MGP

We have designed restriction sites *EcoR* I (GAATTC) and *Hind* III (AAGCTT) in the *MGP-F* sequence and *MGP-R* sequence, respectively. The amplified PCR product is cut and ligated into the corresponding site of pET28a vector. The resulting construct is transformed into *E.coli* BL21 (DE3) strain and used for the induction by adding IPTG, by using the empty vector transformed BL21 (DE3) as a control.

3 RESULTS AND ANALYSIS

3.1 The cDNA of MGP and genomic DNA from the giant panda

The amplification products by RT-PCR are appeared at about 350 bp. The genomic sequence of the *MGP* gene has been submitted to GenBank (accession number:KJ504105). The length of the cDNA fragment amplified from the Giant Panda cloned is 358 bp (Fig.1). The *MGP* sequence has been submitted to Genbank (accession number: KJ504106). The cDNA sequence contains a 312 bp ORF and the ORF encodes 103 amino acids (Fig.2). The *MGP* gene of the Giant Panda is found high homology with *MGP* genes of other mammals by Blast at NCBI. The result of Gene scan indicates that *MGP* gene contains four exons and three introns. Alignment analysis of *MGP* of the Giant Panda and those of *Homo sapiens*, *Bos Taurus*, *Canis familiaris*, *Mus musculus*, *Rattus norvegicus* and *Sus Scrofa* indicates that both the nucleotide sequence and the deduced amino acid sequence are highly conserved with *Canis familiaris* of 93.91% and 96.12%, respectively.

3.2 Prediction of protein function sites in the MGP protein of the Giant Panda

Primary structure analysis reveals that the molecular weight of the putative MGP protein of the giant panda is 12.18 kDa with a theoretical pI = 9.58. Topology

```

1 GAGAGACC ATG AAG AGC CTT CTC CTT CTC TCC ATT CTG
1 M K S L L L S L L
39 GCT GCC TTG GCG GTG GCA GCT CTG TGC TAC GAA TGT
11 A A L A V A A L C Y E S
78 CAT GAA AGC ATG GAA TCC TAT GAA ATT AAT CCC TTC
23 H E S M E S E I N P F
111 ATT AAC CGG AGA AAT GCG AAT ACT TTT ATA TCC CGG
35 I N R R N A N T F I S P
147 CAA CAG AGA TGG AGA GCG AAA GCC CAA GAG AGG ATC
47 Q Q R W R A K A Q E R I
183 AGA GAA AGC ACC AAG CCC GCC TAT GAG CTC AAT CGG
59 R E S T R P A Y E L N R
219 GAA GCC TGT GAT GAC TTC AGG GTT TCG GAA CGC TAT
71 E A C D D F R L C E R Y
258 GCC ATG GTT TAT GGA TAC AAT GCT GCC TAC AAT CGC
83 A M V Y G Y A A Y N R
291 TAT TTT CGG AAG CGC CGA GCG GCC AAA TGAGATTGGAG
95 Y F R K R R G A K *
329 AAAGTCTCTTTCTTCAGACGGCTGGCC

```

Figure 1. Nucleotide sequence of cDNA encoding the Giant Panda MGP and the amino acid sequence deduced from the ORF. Nucleotides are numbered in the 5'- to -3'direction. The predicted amino acid sequence of the gene is shown under the unclenotide. The asterisk (*) represents termination codon.

prediction shows that there is one Protein kinase C phosphorylation site, one Vitamin K-dependent carboxylation domain and one Gla domain profile in the matrix Gla protein of the Giant Panda (*A. melanoleuca*) (Fig. 2). Further analysis detects a few polymorphic sites in the amino acid sequences of other six species compared. Also, most base transitions of the gene coding sequence in these mammals are synonymous mutations.

| | | |
|-------------------------------|---|-----|
| <i>Atiuropada melanoleuca</i> | MKSLLLSI LAALVAALCYESHESMESYEI N PFI NRRNA | 40 |
| <i>Home sapiens</i> | MKSLLLAI LAALAVTL CYESHESMESYELN PFI NRRNA | 40 |
| <i>Bos Taurus</i> | MKSLLLSI LAALVAALCYESHESLESYEI N PFI NRRNA | 40 |
| <i>Canis familiaris</i> | MKSLLLSI LAALVAALCYESHESMESYEI N PFI NRRNA | 40 |
| <i>Mus musculus</i> | MKSLPLAI LAALAVTL CYESHESMESYEI S PFI NRRNA | 40 |
| <i>Rattus norvegicus</i> | MKSLPLAI LAALVAALCYESHESMESYEVSPFI NRRNA | 40 |
| <i>Sus Scrofa</i> | MKSLLLSV LAALVAALCYESHES LESYEI N PFI NRRNA | 40 |
| <i>Atiuropada melanoleuca</i> | NTFI SPQQRWRAKQERIR RESTKPAVELNREACDDFKLCE | 80 |
| <i>Home sapiens</i> | NTFI SPQQRWRAKQERIR BERSKPVELNREACDDVRLCE | 80 |
| <i>Bos Taurus</i> | NSFI SPQQRWRAKQERIR BELNKPQVELNREACDDFKLCE | 80 |
| <i>Canis familiaris</i> | NTFI SPQQRWRAKQERIR REITKPAVELNREACDDFKLCE | 80 |
| <i>Mus musculus</i> | NTFMSPPQQRWRAKQKRVQRNKPQVELNREACDDVRLCE | 80 |
| <i>Rattus norvegicus</i> | NTFI SPQQRWRAKQERIR BELNKPQVELNREACDDVRLCE | 80 |
| <i>Sus Scrofa</i> | NTFI SPQQRWRAKQERIR BELNKPVELNREACDDVRLCE | 80 |
| <i>Atiuropada melanoleuca</i> | RYAMLYGYNAAVNYFERRRGAK | 103 |
| <i>Home sapiens</i> | RYAMLYGYNAAVNYFERRRGAK | 103 |
| <i>Bos Taurus</i> | RYAMLYGYNAAVDRYFERRRGAK | 103 |
| <i>Canis familiaris</i> | RYAMLYGYNAAVNYFERRRGAK | 103 |
| <i>Mus musculus</i> | RYAMLYGYNAAVNYFERORRGAKY | 104 |
| <i>Rattus norvegicus</i> | KYAL I YGYNAAVNYFERRRGAK | 103 |
| <i>Sus Scrofa</i> | RYAMLYGYNAAVNYFERRRGAK | 103 |

Figure 2. Comparison of the MGP amino acid sequences among the different species.

- :Gla domain profile
- ⋯ :Vitamin K-dependent carboxylation domain
- :Protein kinase C phosphorylation site
- :Microbodies C-terminal targeting signal
- :cAMP and cGMP-dependent protein kinase phosphorylation site

3.3 Overexpression and purification of the MGP gene in E.coli

The results indicate that the protein MGP fusion with the N-terminus His-tagged form gives rise to the accumulation of an expected 16 kDa polypeptide that

forms inclusion bodies by SDS-PAGE. Apparently, the recombinant protein is expressed after half an hour of induction and after 2 hours reached the highest level (Fig.3). These results suggest that the protein is active.

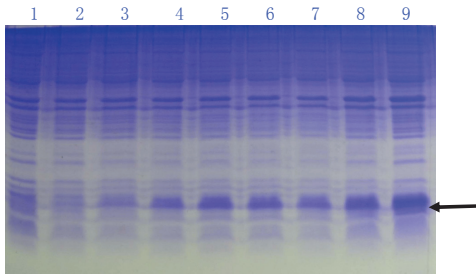


Figure 3. Protein extracted from recombinant *E. coli* BL21 strains was analyzed by SDS-PAGE gel stained with Commassie blue R250. Numbers on right shows the molecular weight and the arrow indicates the recombinant protein bands induced by IPTG with 0h, 0.5h, 1h, 1.5h, 2h, 2.5h, 3h and 4h (lane 2–9), respectively. The lane 1 represents the products of the *E. coli* strains with the empty vectors.

3.4 Prediction of tetrameric structure of MGP protein of the Giant Panda

Prediction of tertiary structure of recombinant protein MGP is simulated by the SWISS-MODEL software (<http://swiss-model.expasy.org/>). The results show that there are many alpha helix and some coil in the secondary structure of MGP protein of the Giant Panda (Fig. 4a). What's more, it also shows that helix is the main form. The secondary structure of MGP of the Giant Panda is similar with other six species and the little differences have no effect on the tertiary structure of MGP (Fig. 4b).



(a)



(b)

Figure 4. The secondary (a) and tertiary (b) structure of recombinant protein MGP from the Giant Panda.

3.5 The construction of the phylogenetic tree

Based on the data of MGP of the main mammalian, which have been published on NCBI, we have constructed the phylogenetic relationship of the Giant Panda (Fig.5).

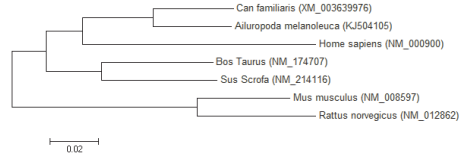


Figure 5. The phylogenetic relationship of the Giant Panda.

4 DISCUSSION

Matrix Gla protein, which exists in bone, cartilage, heart, kidney and lungs and other organs is a vitamin K-dependent protein. And MGP belongs to extracellular mineral-binding protein family. MGP can inhibit the abnormal deposition in calcium ions in the organization. Tanaka (2001) thinks that osteoporosis is closely related to the hardening of the arteries and both of them are affected by MGP.

As we all know that there are two forms of the MGP protein: one is γ -Glutamate carboxylation state and the other is Serine phosphorylation state. The former is activatory which can keep down or inhibit the calcium ion precipitated in the vein.

Lei Guo (2010) and his partners have shown that the MGP protein is overexpressed in tumor tissues. In their study, it indicates that the overexpressed MGP protein plays a role in gastric epithelial cell proliferation. We all know calcium ion is one of the second messengers in the signaling pathways which can activate some enzymes which can inhibit cell multiplication. The calcium ion can be integrated with MGP proteins when the proteins are activated so that the cells can multiply. The conclusion is consistent with our results in figure 2. On the basis of figure 2, we can find that the Gla domain profile locates in 67–97 amino acids and the domain carries negative charges so that it can have a high affinity with the calcium ions and can inhibit the abnormal deposition in calcium ions in bone, cartilage and other organs.

According to figure 4, we can discover that the C-terminal of the protein has plenty of helix and coil. The helix may be the receptor part for the activity factors or other donors. And one of the most important effects is to offer function sits of the protein. The activation depends on its carboxylation Gla residues and the process needs vitamin K. The vitamin K-dependent carboxylation/ gamma-carboxylutamic

region locates in the C-terminal because the C-terminal has conservative sequence.

We can find that there are a high homology between *Ailuropoda melanoleuca* and *Canis familiaris* in both encoding sequence and protein sequence. And from evolutionary tree based on *MGP* gene, we also can discover that they have close genetic relationship. The conclusion is in accord with the conclusion of Shenzhen Huada Gene Research Institute that has reported that the genome of the closest to *Ailuropoda melanoleuca* with *Canis familiaris*.

It should be noted that this study has examined only technology and preliminary results. We have to put out that we do not go into its function because of any limitations of this study. Notwithstanding its limitation, this study does lay a foundation for further research.

5 ACKNOWLEDGMENTS

This project This project was supported by the National Natural Science Foundation of China (31400016 and 31200012), the Application Foundation Project of Sichuan Province (2013JY0094), the Science and Technology Support Project of Sichuan Province (2014SZ0020 and 2014FZ0024), the Cultivate Major Projects of Sichuan Province (14CZ0016), the Open Foundation of Microbial Resources and Drug Development of Key Laboratory and of Guizhou Province (GZMRD-2014-002) and the Doctor Startup Foundation Project of China West Normal University (11B019 and 11B020).

REFERENCES

- [1] Beazley K.E. et al 2013. Two sides of MGP null arterial disease: chondrogenic lesions dependent on transglutaminase 2 and elastin fragmentation associated which induction of adipsin. *The Journal of Biological Chemistry* 288(43): 314–318.
- [2] Cindy F. et al 2009. Dual transcriptional regulation by runx2 of matrix Gla protein in *Xenopus laevis*. *Gene* 450(1): 94–102.
- [3] Cranenburg E.C.M. et al 2010. Characterisation and potential diagnostic value of circulating matrix Gla protein (MGP) species. *Thrombosis and Haemostasis* 104(4): 811–822.
- [4] Fumio H. et al 2001. Expression and location of MGP in rat tooth cementum. *Archives of Oral Biology* 46(7): 585–592.
- [5] Guo L. et al 2010. Discovery and verification of matrix gla protein, a TNM staging and prognosis-related biomarker for gastric cancer. *Chinese Journal of Pathology* 39(7): 436–441.
- [6] Julien M. et al 2009. Phosphate-dependent regulation of MGP in osteoblasts: role of ERK1/2 and Fra-1. *Journal of Bone and Mineral Research* 24(11): 1856–1868.
- [7] Liao M.J. et al 2003. cDNA cloning of growth hormone from giant panda (*Ailuropoda melanoleuca*) and its expression in *Escherichia coli*. *Comp Biochem Phys B* 135: 109–116.
- [8] Merteck S. et al 2009. Matrix Gla protein : an overexpressed and migration-promoting mesenchymal component in glioblastoma. *BMC Cancer* 9(1): 302.
- [9] Noriyuki M. 2005. Expression of matrix Gla protein (MGP) in the spinal ligaments. *Yamaguchi Medical Journal* 53(1): 29–38.
- [10] Ran L. et al 2012. Quantitative determination of matrix Gla protein(MGP) and BMP-2 during the osteogenic differentiation of human periodontal ligament cells. *Archives of Oral Biology* 57(10): 1408–1417.
- [11] Simes D.C. et al 2003. Purification of matrix Gla protein from a marine teleost fish, *Argyrosomus regius*: calcified cartilage and not bone as the primary site of MGP accumulation in fish. *Journal of Bone and Mineral Research* 18(2): 244–259.

The kinetic and thermodynamic study on the OH-initiated degradation mechanism of polychlorinated dibenzo-*p*-dioxins

X. Zhang, C.X. Zhang, L.Y. Kang, C. Gong & X.M. Sun

Environment Research Institute, Shandong University, Jinan, P. R. China

J. Yu

Affiliated school of Shandong 83 carbon factory, China

ABSTRACT: Quantum chemical method is adopted to elucidate the OH-initiated degradation mechanism of polychlorinated dibenzo-*p*-dioxins. Among the entire possible product channels, the main reaction pathways are discussed and show that the concerted isomerization reaction of druidical adduct is more favored than the step-by-step decomposition of monoradical adduct. The kinetic calculation is performed and the rate constants are calculated over a temperature range of 200~1200 K. The forward and reverse rate constants and the equilibrium constants are obtained. The four-parameter formulas of rate constants with the temperature are fitted.

1 INTRODUCTION

The polychlorinated dibenzo-*p*-dioxins (PCDDs) is a kind of important persistent organic pollutants (POPs). They act as not only a potent carcinogen but also an extremely persistent endocrine disrupter that may have serious effects on human and animals (Schechter, 1994). Furthermore, there are many sources including the combustion of waste materials, industrial chemicals, unintended by-products and so on (You et al.1994; Piao et al.1999). Since the PCDDs possess some distinct properties: persistent in the environment, bioaccumulates in fatty tissues, and long-range transport, they can strongly impact the environment and human health (Stalling et al.1985). Therefore, the rapid and simple degradation methods of PCDDs is now required urgently. Generally speaking, the degradation methods include photo degradation, thermal dechlorination, biodegradation, chemical dechlorination and heterogeneously catalytic hydrodechlorination (Konstantinov et al.2000; Choi et al.2004; Hilmi et al.1998; Sako et al.2000).

There are many experimental and theoretical studies on PCDD. In 2004, Ciesla's report indicated that a combination of H₂O₂ and UV irradiation with Fe(II) (photo-Fenton process), can significantly enhance decomposition of many refractory organic compounds (Ciesla et al.2004). Katsumata (Katsumata et al.,2004) established the degradation method using this system for various environmental pollutants including 2,3,7,8-TeCDD, *penta*, *hexa*, *hepta* and *octa* CDDs (PeCDD, HxCDD, HpCDD and OCDD).

So far, the theoretical studies focus on the molecular structure, vibrational frequencies, formation pathways, and decomposition process in homogeneous phase (Lee,2004). However, the degradation mechanism in heterogeneous phase is investigated seldomly.

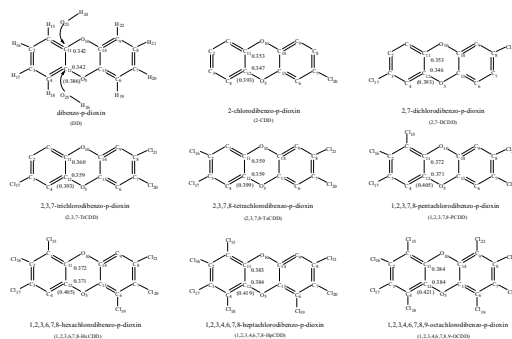


Figure 1. The structure of PCDD (n = 0–8).

In this paper, the OH-initiated degradation mechanism of polychlorinated dibenzo-*p*-dioxins (PCDD) is calculated. The information of the stationary points including the reactants, intermediates, transition states and products is calculated. The kinetic calculation is performed and the rate constants are calculated over a temperature range of 200~1200 K. The forward and reverse rate constants and the equilibrium constants are obtained. The four-parameter formulas of rate constants with the temperature are fitted.

2 COMPUTATIONAL METHODS

2.1 Geometry Optimization

High-level molecular orbital theory calculations have been carried out for the reactions of 2,3,7,8-TeCDD with OH radical. The geometries have been optimized at the B3LYP level with the 6-31G(d) basis sets. All the work is performed using the Gaussian 03 programs (Frisch et al. 2003) and SGI workstation.

2.2 Kinetic calculation

The canonical variational transition-state (CVT) theory is used to calculate the rate constants coupled with the small curvature tunneling (SCT) method to compute the tunneling properties (Zhang et al. 2011). The kinetic calculations have been carried out using POLYRATE 9.7 program (Corchado et al. 2007). The CVT rate constant for temperature T is given by:

$$k^{CVT}(T, s) = \min_s k^{GT}(T, s) \quad (1)$$

$$k^{GT}(T, s) = \frac{\sigma k_B T}{h} \frac{Q^{GT}(T, s)}{\Phi^R(T)} e^{-V_{MEP}(s)/(k_B T)} \quad (2)$$

3 RESULTS AND DISCUSSION

The structure of PCDD (n=0-8) is illustrated in Scheme 1.

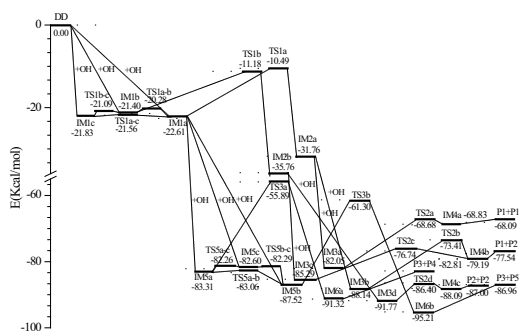


Figure 2. The potential energy surface profile of the OH-initiated reaction pathways of DD.

3.1 Degradation mechanism

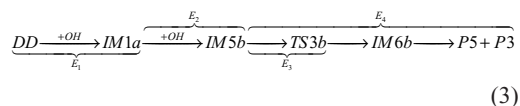
There are two perspectives that should be pointed out. On one hand, each C atom adjacent to the O atoms in DD may be attacked by O atom in OH radical to form the adduct DD-OH (IM1) which results in the rupture of the dioxin ring to obtain the phenol, benzoquinone and phenolic ketone. On the other hand, the rupture

of the dioxin ring is not affected by the number and orientation of Cl atoms in the PCDDs. The adducting and ring-opening mechanism initiated by OH radical is chosen to study the degradation of DD.

Because of the orientation of OH group, the adduct DD-OH (IM1) has three isomers labeled as IM1a, IM1b and IM1c, respectively. The energies of IM1a, IM1b and IM1c are 22.61, 21.40, and 21.83 kcal/mol lower than those of the total energy of the original reactants. The barriers and the reaction heats can be seen in Figure 1. The barrier is very slow which implies that the conformational transformation occurs easily. Then the adduct IM1a is the most stable isomer among the three stable geometries.

In the ring-opening and adducting mechanism, the adduct DD-OH crosses the potential barrier about 10 kcal/mol, (IM1a, 12.12 and IM1b, 10.22 kcal/mol) to rupture a C-O bond in the dioxin ring. While in the adducting and ring-opening mechanism, the association of the adduct DD-OH with OH radical will release heat above 60 kcal/mol. Obviously, the second mechanism takes the major when OH radical is abundant. The second OH radical can be added from the same side or the opposite side to several positions, such as C12, C13, and C14. Another three isomers labeled as IM5a, IM5b and IM5c are obtained with equilibrium distributions of 0.001:0.998: 0.001. Obviously the IM5b is the preferential conformation. The isomerization of IM5b to IM6b will open up the two O-C bonds via the transition state TS3b. The energy of TS3b is higher by 26.23 kcal/mol than that of IM5b. The subsequent unimolecular decomposition of IM6b produces pyrocatechol (P5) and o-benzoquinone (P3).

While the isomerization of IM5b has a barrier with 26 kcal/mol. Then the isomerization of IM5b is the rate-determining step in the degradation mechanism. The whole degradation path is as follows and the energy changes of the PCDD degradation path (n=0-8) are labeled and listed in Table 1. The adducts of PCDD and OH are labeled as PCDD-OH.



Where E1 and E2 are the heat of the barrierless reaction, E3 and E4 are the potential barrier of the forward and reverse isomerization rupture the dioxin ring, and E5 is the dissociation energy. From Table 1, it can be seen that the energy varies non-monotonically with the number of Cl atom. The forward potential barrier E3 varies from 25 to 29 kcal/mol, while the E4 lies in the range of 32~36 kcal/mol. The number and orientation of Cl atoms in PCDD affect the charge distribution of the C atoms adjacent to the O atom in

Table 1. The energy changes of the PCDD degradation path (n=0-8) (unit: kcal/mol).

| Energy | DD | CDD | DCDD | TrCDD | TeCDD | PCDD | HxCDD | HpCDD | OCDD |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| E1 | -17.37 | -18.34 | -18.63 | -18.69 | -18.59 | -21.33 | -21.48 | -21.41 | -22.06 |
| E2 | -59.68 | -59.19 | -59.40 | -59.71 | -59.73 | -60.86 | -61.21 | -62.97 | -63.21 |
| E3 | 26.23 | 26.60 | 26.07 | 25.85 | 26.15 | 27.86 | 26.30 | 28.40 | 27.76 |
| E4 | -33.91 | -35.18 | -34.06 | -32.60 | -32.57 | -34.08 | -32.99 | -34.88 | -34.38 |
| E5 | 8.25 | 9.20 | 8.72 | 8.29 | 7.33 | 7.31 | 6.80 | 7.31 | 6.59 |

Table 2. The four-parameter formulas of forward and reverse rate constants.

| | Forward | | | | Reverse | | | |
|----------|----------|-------|----------|----------------|----------|--------|----------|----------------|
| | A | n | Ea | R ² | A | n | Ea | R ² |
| DD-OH | 3.99E+11 | 0.35 | 1.13E+05 | 0.999 | 2.97E-23 | 1.66 | 1.14E+05 | 0.999 |
| CDD-OH | 1.55E+18 | 0.28 | 1.03E+05 | 0.999 | 9.18E-34 | 1.65 | 1.10E+05 | 0.999 |
| DCDD-OH | 6.68E+12 | 0.82 | 1.01E+05 | 0.999 | 2.37E-37 | 1.68 | 1.17E+05 | 0.999 |
| TrCDD-OH | 4.41E+08 | 1.58 | 9.95E+04 | 0.999 | 4.07E-21 | 1.65 | 1.08E+05 | 0.999 |
| TeCDD-OH | 6.76E+09 | 1.08 | 1.04E+05 | 0.999 | 1.11E-22 | 1.67 | 1.13E+05 | 0.999 |
| PCDD-OH | 8.95E+15 | -0.95 | 9.66E+04 | 0.997 | 5.19E-17 | 0.0093 | 9.77E+04 | 0.998 |
| HxCDD-OH | 4.46E+16 | -0.52 | 8.39E+04 | 0.999 | 7.26E-04 | -3.75 | 1.36E+05 | 0.999 |
| HpCDD-OH | 3.24E+05 | 2.39 | 9.48E+04 | 0.998 | 9.54E-22 | 1.70 | 1.15E+05 | 0.998 |
| DCDD-OH | 1.36E+10 | 1.26 | 1.12E+05 | 0.999 | 3.22E-21 | 1.75 | 1.24E+05 | 0.999 |

the dioxin ring since Cl atom belongs to the orthopara orientation group and electron-withdrawing group. The charge distribution of the C atoms adjacent to the O atom affects the addition of OH radical. The higher the positive electricity is, the easier the addition of OH radical is.

3.2 Rate constants

Taking the temperature into consideration in the troposphere, the flue gas, the waste incineration, and the wastewater treatment, a wider range of 200~1200 K is chosen to study the relationship between the temperature and the rate constants.

3.2.1 The forward rate constants

Both the rate constants of TST and CVT vary bigger from 200 to 1200 K. The rate constant of the TST method is about 15 times of those of the CVT method at the 200K and about 25 times at 1200K, which suggests that the variational effect is not to be neglected in this reaction, and then higher the temperature is, the larger the variational effect is. The tunneling effects are considered to compute the rate constants. The CVT rate constants are larger than the CVT/ZCT and CVT/SCT ones over the temperature range of 200~1200 K. For example, at 250 K, the CVT rate constant is 1.07×10^{-11} sec⁻¹, while the CVT/ZCT and CVT/SCT rate constants are 8.71×10^{-12} and 8.72×10^{-12} sec⁻¹. The former is 1.2 times of the latter. The higher

the temperature is, the less the ratio is, which means that only in the lower temperature range the tunneling effects play an important role in the calculation of the rate constants. Another phenomenon is that the CVT/ZCT rate constants are very similar to the rate constants of CVT/SCT in the higher temperature range. The CVT/SCT rate constants are more accurate in a large temperature range. It is obvious that the calculated rate constants exhibit typical non-Arrhenius behavior, i.e., $k(T) = a \times 10^b T^c \exp(d/T)$. The rate constants are fitted by four-parameter formulas over the temperature range of 200~1200 K and given in Table 2. The correlative coefficient R² is above 0.99. It can be observed that the relativity between actual values and predicted values is excellent. The formulas of the rate constants with temperature can provide helpful information for the atmospheric model study on the formation and degradation of dioxin and the management of hazardous materials.

3.2.2 Atmospheric lifetimes

According to the rate constants of elementary reaction, the atmospheric lifetime of the reactants can be calculated and analyzed. For the first order reaction, the lifetime τ is the reciprocal of the rate constants k . The lifetime at the room temperature is taken as an example to discuss.

For the monomolecular isomerization reaction, the CDD-OH and HxCDD-OH belong to the short lifetime species, and the lifetimes are 4.6 and 25.1 seconds,

respectively. The DCDD-OH (~2.88 hours) and TrCDD-OH (~15.49 days) belongs to medium life-time species. Other species belong to long life-time ones, such as, TeCDD-OH (~2.13 months), OCDD-OH (~10.34 months), PCDD-OH (~11.57 months), DD-OH (~1.77 years) and HpCDD-OH (~2.31 years). Generally speaking, the higher the temperature is, the shorter the lifetime is.

4 CONCLUSIONS

The OH-initiated degradation mechanism of polychlorinated dibenzo-p-dioxins is studied at the B3LYP/6-31G* level and the kinetic calculation is performed and the rate constants are calculated using CVT/ZCT method (1) The isomerization of PCDD-OH adduct is the rate-determining step in the degradation mechanism. (2) The four-parameter formulas of rate constants with the temperature are fitted and the lifetimes of the reaction species are estimated according to the rate constants. (3) The large equilibrium constants imply that the forward reactions are superior to the reverse ones.

ACKNOWLEDGMENT

This work is supported by Marie Curie International Research Staff Exchange Scheme Fellowship within the 7th European Community Framework Program (No. 295132), National Natural Science Foundation of China (No. 21277082, 21337001), Promotive Research Fund for Excellent Young and Middle-aged Scientists of Shandong Province (No. BS2012HZ009) and Program for New Century Excellent Talents in University (NCET-13-0349).

REFERENCES

[1] Choi, J.; Choi, W.; Mhin, B.J. 2004. Solvent-Specific Photolytic Behavior of Octachlorodibenzo-p-dioxin. *Environ. Sci. Technol.*, 38:2082–2088.

[2] Ciesla, P.; Kocot, P.; Mytych, P.; Stasicka, Z. 2004. Homogeneous photocatalysis by transition metal complexes in the environment. *J. Mol. Catal. A Chem.* 224:17–33.

[3] Corchado, J.C.; Chuang, Y.Y.; Fast, P.L.; Villa, J.; Hu, W.P.; Liu, Y.P.; Lynch, G.C.; Nguyen, K.A.; Jackels, C.F.; Melissas, V.S.; Lynch, B.J.; Rossi, I.; Coitino, E.L.; Ramos, A.F.; Pu, J.; Albu, T.V.; Garrett, R. B.C.; Truhlar, D.G. 2007. POLYRATE Version 9.7. University of Minnesota, Minneapolis.

[4] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr. T.; Kudin, K. N.;

Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. 2003. Gaussian, Inc., Pittsburgh PA.

[5] Hilmi, A.; Luong, J.H.T.; Nguyen A.-L. 1998. Applicability of micellar electrokinetic chromatography to kinetic studies of photocatalytic oxidation of dibenzo-p-dioxin. *Chemosphere*, 36:3113–3117.

[6] Katsumata, H.; S. Kawabe, S. Kaneco, T. Suzuki, K. Ohta. 2004. Degradation of bisphenol A in water by the photo-Fenton reaction. *J. Photochem. Photobiol. A Chem.* 162:297–305.

[7] Konstantinov, A.D.; Johnston, A.M.; Cox, B.J.; Petrulis, J.R.; Orzechowski, M.T.; Bunce, N.J.; Tashiro, C.H.M.; Chittim, B.G. 2000. Photolytic Method for destruction of dioxins in liquid laboratory waste and identification of the photoproducts from 2,3,7,8-TCDD. *Environ. Sci. Technol.* 34:143–148.

[8] Lee, J.E.; Choi, W.; Mhin, B.J.; Balasubramanian, K. 2004. Theoretical Study on the Reaction of OH Radicals with Polychlorinated Dibenzo-p-dioxins. *J. Phys. Chem. A* 108:607–614.

[9] Piao, M.; Chu, S.; Zheng, M.; Xu, X. 1999. Characterization of the combustion products of polyethylene. *Chemosphere*, 39:1497–1512.

[10] Sako, T.; Sugeta, T. 2000. Decomposition of environmentally toxic substances with supercritical water oxidation process. *Materia* 39:314–319 (in Japanese).

[11] Schecter, A. 1994. *Dioxins and Health*. Plenum, New York.

[12] Stalling, D.L.; Norstrom, R.J.; Smith, L.M.; Simon, M. 1985. Patterns of PCDD, PCDF, and PCB contamination in Great Lakes fish and birds and their characterization by principal components analysis. *Chemosphere*, 14:627–643.

[13] Sun, X.M.; Zhang, Q.Z.; Wang, W.X. 2008. Degradation mechanism of PCDDs by photo-fenton process: Quantum chemistry and quantitative structure-activity relationship. *Sci. Total Environ.* 402:123–129.

[14] You, J.H.; Chiang P.C.; Chang K.T.; Chang S.C. J. 1994. Polycyclic aromatic hydrocarbons (PAHs) and mutagenicity of soot particulates in air emissions from two-stage incineration of polystyrene. *Hazard. Mater.* 46:1–17.

[15] Zhang, C.X.; Sun, T.L.; Sun, X.M. 2011. Mechanism for OH-initiated degradation of 2,3,7,8-tetrachlorinated dibenzo-p-dioxins in the presence of O₂ and NO/H₂O. *Environ. Sci. Technol.* 45:4756–4762.

Thermal, electrical and mechanical properties of Lead-free Tin solder alloys containing Bi and In

S. Attia

Department of Biomedical science, Al-Farabi College of Dental and Nursing, Kingdom of Saudi Arabia, Jeddah.

ABSTRACT: Low melting temperature lead-free soldering alloys with tin base metal 90 Wt % , indium ≤ 10 wt % and bismuth ≤ 10 Wt. % namely Sn₉₀In₁₀, Sn₉₀In₅Bi₅ and Sn₉₀Bi₁₀ alloys have been prepared. The samples are subjected to thermal, electrical and mechanical studies to reveal their physical properties. Among these three samples the current study verifies that the low soldering as Bi free alloy namely (Sn₉₀In₁₀), shows the best ductile characters so it responds to the creep and microstructure mechanical tests. The addition of little bismuth to the prepared samples abruptly increases in their brittle character. The addition of Bi to Sn-In alloys causes decreasing in their amorphous character and improvement in their crystalline behavior. This study also establishes the Sn₉₀Bi₁₀ alloy offer the most convenient properties for low temperature solder applications.

KEYWORDS: Tin alloys, resistivity, young's modulus, microhardness, wetting and creep.

1 INTRODUCTION

Eutectic Sn37Pb solder constitutes a common alloy that finds wide applications for interconnecting and packaging electronic components. However, Pb and Pb – containing compounds are considered toxic. Due to environmental concerns and health hazards associated with Pb, Numerous researches have been focused on Pb – free solders and soldering techniques aiming at replacing Sn – Pb solders with Pb – free alternatives [1–3]. Low temperature soldering is one of the key technologies before the accomplishment of total lead-free conversion in electronics industries and in the assembly process. When the devices soldered are susceptible to thermal damage, low-temperature soldering is necessary [4,5]. It reduces the damage due to thermal cycling caused by thermal expansion mismatch among the various materials in electronic packaging.

The melting temperature is a matter of concern when developing a new lead-free composite solder because a wide alteration of temperature will largely affect the solder application techniques when new equipment and systems need to be developed [6]. However, if small changes occur in the melting temperature by the reinforcement, material can greatly improve the service temperature capability of the monolithic base – matrix solder [7].

Bismuth–Tin and In-Sn are classified as low-temperature special solders. They are used in applications where a low melting temperature point alloy is

required, such as step soldering or when components in an assembly are sensitive to temperature. In addition, the solder joints of temperature-sensitive zones and the outer layers of classification packaging, can reduce the influence of soldering temperature on the inner layer of classification packaging [8, 9].

The present work aims at measuring the physical properties of lead free soldering for the Sn90In₁₀, Sn₉₀In₅Bi₅ and Sn₉₀Bi₁₀ alloys, as well as to investigate some thermal and mechanical properties, determines some useful thermal and mechanical parameter for the alloys. The establishment of these properties is expected to enable us to understand the behavior of these three alloys and to reveal whether they are favored to be good in soldering usage or not.

2 EXPERIMENT

The samples are prepared from pure elements (99.5%) of tin, bismuth and indium. The samples are weighed using a highly sensitive microbalance with an accuracy of 10⁻⁴ grams. The weighed samples are placed in a very clean Pyrex tube under a natural fluxing agent “colophony” of the chemical constituent (Abietic acid) by heating, and it changes Laevopimaric acid to prevent the sample oxidation in air and to improve mixing and melting. The syntheses of the samples are carried out on a benzene fire for more than half an hour while the tube is shaken to ensure homogenization of the melt. After solidification, the tube is

broken to obtain a bulky alloy. Excess flux is removed by soaking the ingot in pure carbon tetra chloride (CCl₄) for a sufficient time. The Densities for Sn alloys are shown in table (1).

Table 1. Densities for soldering alloys.

| Solder alloys | Density calculated | Density Experimental |
|--|--------------------|----------------------|
| Sn ₉₀ In ₁₀ | 7.30 | 7.31 |
| Sn ₉₀ In ₅ Bi ₅ | 7.61 | 7.55 |
| Sn ₉₀ Bi ₁₀ | 7.93 | 7.83 |

3 RESULTS AND DISCUSSION

3.1 X-ray analysis

The X-ray diffraction patterns of (Sn₉₀In₁₀, Sn₉₀In₅Bi₅ and Sn₉₀Bi₁₀) alloys are Figure (1) which shows the crystalline and amorphous phase for Sn90In10 alloy. This is concordant with the work of Allen and Jesser [10]; as they mentioned, both amorphous and crystalline phases for SnIn alloys with different concentration are present. The diffraction patterns of these amorphous particles show diffusion of the characteristic of amorphous phases. Compositional analysis of the crystalline and amorphous particles reveals trends in the occurrence of these phases. The X-ray

diffraction patterns, Figure (1) shows that the Sn₉₀In₁₀ alloy contains the amorphous nature with the crystalline structure tetragonal tin phase, a tetragonal indium phase and a hexagonal InSn4 phase. The bismuth addition to SnIn of the soldering causes the increasing crystalline character and revealing the sharp lines as it occurs in Sn₉₀In₅Bi₅ alloy.

The X-ray diffraction analysis of SnIn solder alloys carries out the presence of both amorphous and crystalline structures which considers a composite material alloys. Moelans, et al. [11] states that, in In-Sn system, there are two intermediate phases with large homogeneity range. These are designated as β-InSn and γ-InSn, where the β- InSn phase have a tetragonal structure. The Bi-In system has four intermediate phases.

3.2 Differential thermal analysis (DTA) results

Figure (2) shows the DTA profiles of solder alloys. The Sn₉₀In₁₀ alloy exhibits a sharp endothermic peak at about 391.3 K that corresponds to the solidus temperature. The second endothermic peak for the alloy is due to the complete melting until the liquidus point reaches 468.1 K which is as shown in table (2). The alloy Sn₉₀Bi₁₀ the sharp endothermic reaction occurs at about 411 K, which corresponds to the solidus temperature and 416 K corresponds to liquidus temperature.

The result of this alloy is concordant with the result of Suganuma [12], who states that Sn-Bi binary alloy exhibits a sharp endothermic reaction at about 413 K, which corresponds to the eutectic temperature i.e., 138 °C [13]. The temperature of the endothermic reaction beyond 411 K that is observed in the hypoeutectic alloy corresponds to the liquidus line. The Sn90In5Bi5 alloy has one endothermic peak with solidus and liquidus temperature 457.7 and 478.4 K respectively.

From the table, the melting temperature range for the SnIn alloys are suitable compared with SnPb solder alloy. For any alloy to be worthwhile solder in the electronics industry, it must have a liquidus temperature (above which it is molten) that is sufficiently low so that components and boards are not damaged during soldering. In practice, this means that it must be usable at 533 K, which is the maximum temperature exposure limit for the majority of electronic components. By the same argument, it must have a solidus temperature (below which it is completely solid) that is sufficiently high so that during service, and the solder joints point of SnIn alloys provides a useful compromise between these two criteria. It is much more preferable to use an eutectic solder or a solder with a narrow pasty range. From table (2) the narrow pasty range is shown for the eutectic alloy Sn₉₀Bi₁₀, which offers the suitable property for solder application.

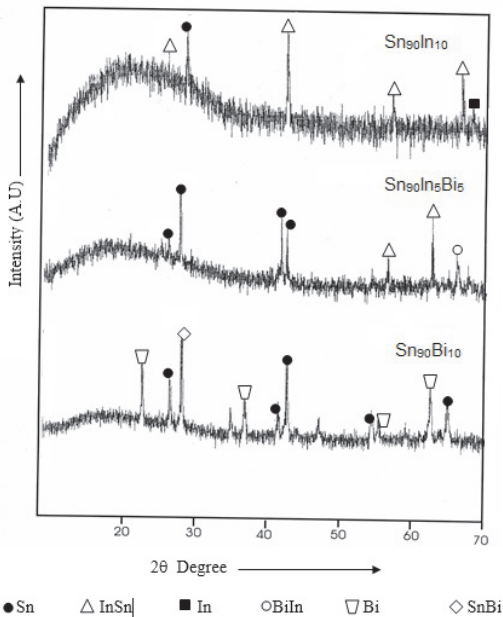


Figure 1. X-ray diffraction pattern of SnIn solder alloys.

Table 2. Solidus and liquidus temperature in (K) for the SnIn alloys.

| Solder alloy | Solidus Temperature (K) | Liquidus Temperature (K) | pasty range (K) |
|--|-------------------------|--------------------------|-----------------|
| Sn ₉₀ In ₁₀ | 391.9 | 468.1 | 76.2 |
| Sn ₉₀ In ₅ Bi ₅ | 457.7 | 478.4 | 20.7 |
| Sn ₉₀ Bi ₁₀ | 411 | 416 | 5.0 |

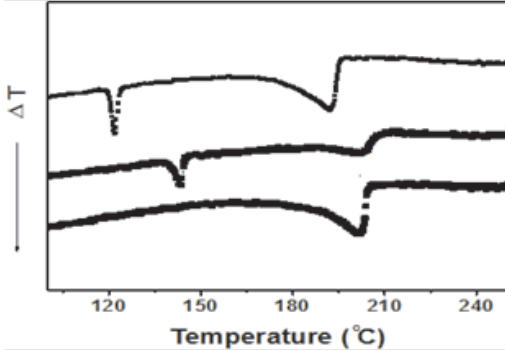


Figure 2. The DTA thermogram of the SnIn solder alloys.

3.3 Wettability

Wettability is the precursor of solderability, which describes the solder's ability to form an actual joint on a circuit board. Wettability is quantitatively assessed by the contact angle formed at the solder-substrate's flux triple point. The contact angles of SnIn solder alloys at 523 K after 40 s elapsed time are shown in table (3). It is obvious that these alloys have adequate contact angles, leading to their spreading on the substrate. The contact angle decreases by adding bismuth to tin-indium alloys.

3.4 Resistivity measurements

Electrical properties of SnIn samples depend sensitively on the structural state of the material as well as on the alloy composition. The measurements of the temperature dependence of the electrical resistivity

Table 3. Contact angles of the alloys on different substrates after 40 sec elapsed time for soldering alloys.

| Solder | Contact angles (present) at temperature 573 K | |
|--|---|------------------------------|
| | Al substrate | CuZn ₃₀ substrate |
| Sn ₉₀ In ₁₀ | 117 | 35 |
| Sn ₉₀ In ₅ Bi ₅ | 115 | 29 |
| Sn ₉₀ Bi ₁₀ | 113 | 21 |

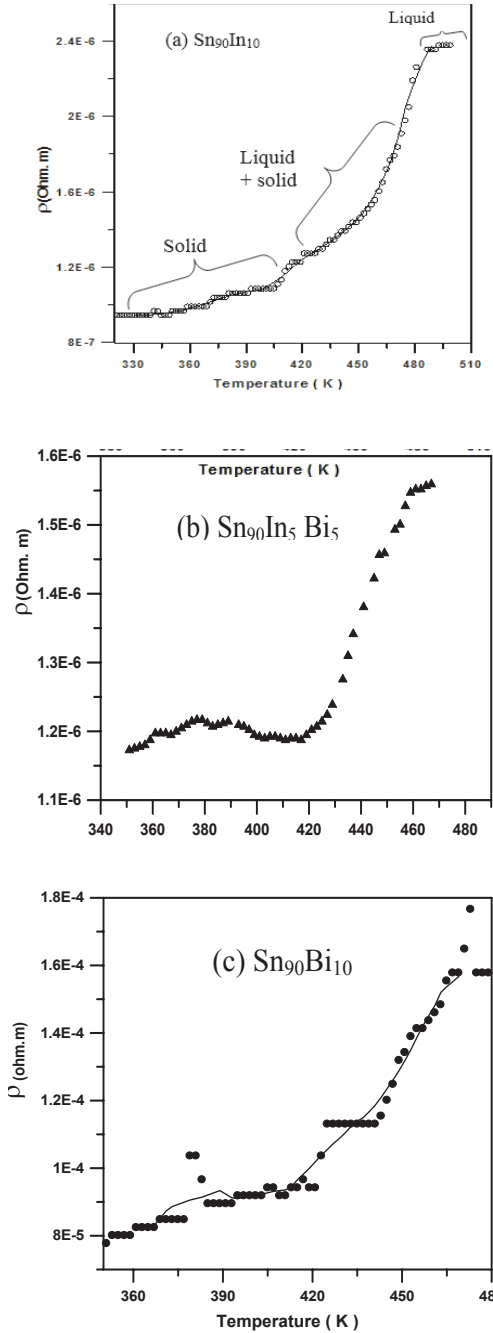


Figure 3. Resistivity measurements of the SnIn solder alloys.

might thus be a useful tool to study the process of any phase transformation of the material. Figures (3a, b and c) show the temperature dependence of resistivity

for alloys. It shows the linear dependence of resistivity on temperature, and no phase transformation is observed for the three alloys. For metals above Debye temperature, with no phase transformation, the resistivity can be expressed as;

$$\rho = \rho_{rt} [1 + \alpha (T - T_{rt})] \quad (1)$$

where ρ_{rt} is the resistivity at room temperature T_{rt} and α are the temperature coefficient of resistivity. And α is given by

$$\alpha = \text{slope} / \rho_{rt} \quad (2)$$

At low temperature below the melting point, the resistivity of the solid is governed by residual scattering of the conduction electrons by crystal imperfections, dislocation and impurity concentrations. Little increasing temperature, electron phonon scattering processes attribute to the thermal lattice vibrations that in turn contribute to resistivity increase. All these contributions represent “disorder scattering” which would be expected to increase on melting and indeed the resistivity generally double value as it is shown in the figures.

3.5 Mechanical properties

3.5.1 Young’s modulus

Both the young’s modulus E and the internal friction or the damping capacity Q^{-1} have been measured by the dynamic resonance method for all as-quenched SnIn solder alloys. The values of Young’s modulus, E , bulk modulus, B , and Shear modulus, μ of SnIn solder alloys are shown in table (4). It is obvious that adding Bi decreases the elastic modulus values of Sn-In, because Bi atoms dissolved in the Sn-In matrix cause shallow defects in the matrix by decreasing the interatomic bonds between the matrix atoms, thus lowering the elastic modulus. In addition, it is interesting to note that Ledbetter’s theoretical value μ/E is in good agreement with the experimental data [14, 15]. From table (4) the Ledbetter’s values of μ/E are about (0.36).

Table 4. Young’s modulus (E), Shear modulus (μ), Bulk modulus (B), Hardness (H_V), Internal friction (Q^{-1}) and Shear stress (τ_{max}) of SnIn solder alloys.

| Solder alloy | E | B | μ | H_V (MPa) | $Q^{-1} \times 10^{-2}$ | μ/E | τ_{max} (MPa) |
|--|-----|------|-------|-------------|-------------------------|---------|--------------------|
| Sn ₉₀ In ₁₀ | 47 | 1.26 | 16.8 | 9.87 | 0.391 | 0.358 | 2.961 |
| Sn ₉₀ In ₅ Bi ₅ | 43 | 23.9 | 15.4 | 21.3 | 0.593 | 0.358 | 6.391 |
| Sn ₉₀ Bi ₁₀ | 42 | 23.3 | 15.1 | 21.64 | 0.313 | 0.357 | 6.492 |

3.5.2 Internal friction

The damping capacity or the ability of metals to damp out vibrations is important in reducing the danger of failure in machines and structure subject to mechanical vibration or repeated shock loading. The internal friction Q^{-1} of SnIn solder alloys is decreasing by adding Bi as it is seen in table (4). That is because the presence of dissolved Bi atoms increases the internal atomic mobility, and therefore the resonance frequency of this material is enhanced.

3.5.3 Microhardness measurements

The hardness of a material is often equated with its resistance to abrasion or wear and this characteristic is of practical interest since it may determine the durability of a material during usage period and it may also decide the suitability of the material for special applications. The results obtained for the microhardness values of the SnIn solder alloys are represented in table (4).

From the table, we can see the microhardness increases with increasing of Bi concentration of SnIn solder alloys. Generally, there is an excellent correlation between hardness and wear resistance, as it is well known [16]. In wear of a metal or alloy, small particles are torn from its surface by materials moving against it. High hardness tends to prevent both penetration of the surface and the deformation necessary to permit relatively large contact areas. Where in crystalline solids the deformation is believed to be through crystallographic slipping.

3.5.4 Maximum shear stress

Table (4) shows the maximum shear stress values τ_{max} . It increases with adding Bi to Sn-In alloys to maximum value (6.492 MPa). Hence, the variation in hardness is attributed to the structural changes of the matrix phase.

3.6 Creep results

The mechanical creep of a soldering material occurs more readily when elevated temperatures one involves. At elevated temperature, the mechanical properties of metals are a result of simultaneous processes of strain hardening, due to plastic deformation, and the softening effect of recovery and recrystallization [17]. Sn metals have a melting temperature of 505 K and its lead-free alloys have almost similar melting temperatures. For most electronics product, the service temperature ranges from room temperature is up to 373 K and from 313 to 423 K for vehicles engine. Within this temperature range Sn alloys creep as well as plastically deformed, makes the understanding of Sn’s mechanical behavior important. When the test specimen is subjected to a constant load, the initial

instantaneous response includes elastic and time independent plastic flow. Creep then precedes in the three successive stages namely, primary, secondary and tertiary creep [18].

The creep specimens of SnIn solder alloys are swaged in the form of wires (diameter 0.66 mm with cold work of approximately 94.5%. The wire samples are annealed for one hour at 383 ± 2 K. Following this heat treatment, the samples are allowed to cool slowly to room temperature with a rate of ≈ 0.04 ks^{-1} . On carrying out the creep experiment, all the samples are cracked except the $\text{Sn}_{90}\text{In}_{10}$ alloy. This can be attributed to the presence of bismuth in the above mentioned samples. The addition of Bi increases the brittle character of its alloy, the sample alloy that contains high Bi cannot be responsible for the cold work while the low Bi content is responsible for the cold work however all of them are not crept.

Creep experiment is carried out at five different constant applied stresses (2.8, 3.3, 3.6, 4.2 and 5.6 MPa) and at test temperatures (383, 388, 393, 398 and 403 K) respectively. Figure (4) is a representative example of the isothermal creep curves for $\text{Sn}_{90}\text{In}_{10}$ alloy. The creep behavior observed is of the normal type, i.e. the instantaneous and three stages are observed. By increasing the test temperature, the creep time is decreased. The maximum ductility of $\text{Sn}_{90}\text{In}_{10}$ alloy decreases (form 18 %) with increasing the test temperature 398 K.

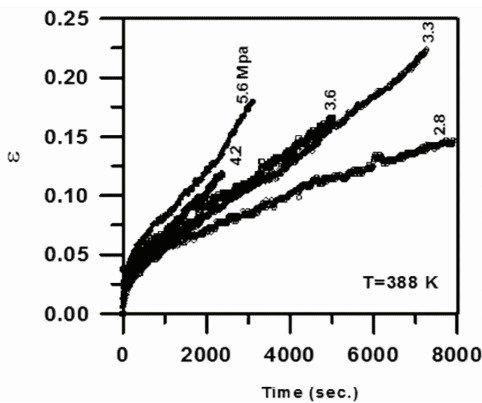


Figure 4. Isothermal creep curves for $\text{Sn}_{90}\text{In}_{10}$ alloy under different stresses and 383 ± 2 K annealing temperature for (1 hr).

3.6.1 Stress and temperature dependence of creep, on the (n) exponent

The double-log plot of minimum creep rate vs. stress at different test temperatures for $\text{Sn}_{90}\text{In}_{10}$ alloy is shown in Figure (5). The values increase with the increasing of the applied stress and/or test temperature

for $\text{Sn}_{90}\text{In}_{10}$ alloy. The increase of ϵ . With increasing test temperature could be explained in view of an easier motion of dislocation already existing, where the thermal energy introduced facilitates the dislocation motion by helping it to overcome the precipitates which act as barrier [18]. The stress exponent n for $\text{Sn}_{90}\text{In}_{10}$ are shown in table (5) and governed by (Nabarro-subgrain mechanism).

Table 5. The stress exponent n at different test temperature.

| Annealing temperature | Test temperature | Exponent n |
|-----------------------|------------------|--------------------------------|
| | | $\text{Sn}_{90}\text{In}_{10}$ |
| 373 ± 2 K | 383 | 1.13 |
| | 388 | 0.82 |
| | 393 | 0.49 |
| | 398 | 0.40 |
| | 403 | 0.53 |

As the temperature is increased marked variations in the compositions of the two phases and subsequently associating diffusion activity occurs. This process gives rise to a solution process at one of the interphase boundaries and precipitation process at the other interphase boundary. These changes in compositions are connected to the movement of the dislocations at the interphase boundaries [12].

3.6.2 Activation energy of the steady state creep

the activation energy, Q of the steady state creep region of the $\text{Sn}_{90}\text{In}_{10}$ alloy is calculated from the slopes of the relation between \ln and $1/T$ shown in

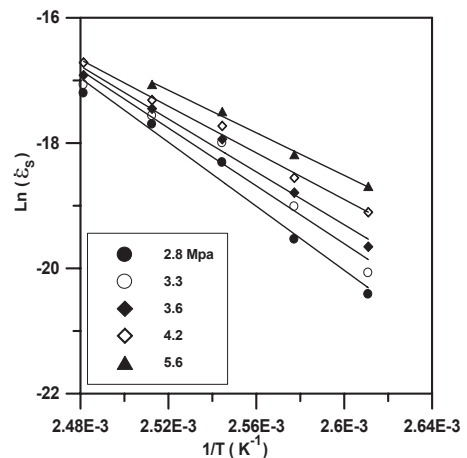


Figure 5. Double-log plot of minimum creep rate versus stress at different test temperature for $\text{Sn}_{90}\text{In}_{10}$ solder alloy.

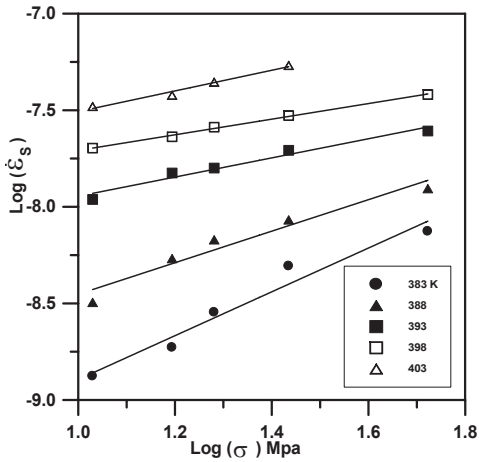


Figure 6. Relation between $\ln(\dot{\epsilon}_s)$ and $1/T$ (K⁻¹) curves for Sn90In10 solder alloy.

Figure (6) shows linear curves for values of the activation energy of SnIn solder alloys are given in table (6). Frenkel [19] studies the creep of some pure metals and find that the activation energies for polycrystalline tin and indium is 0.91 and 0.72 eV

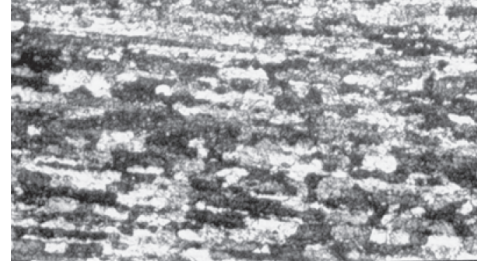
Table (6). The activation energy from the steady state creep at different stresses of solder alloy.

| Annealing temperature | Stress (MPa) | Activation energy (eV) |
|-----------------------|--------------|-----------------------------------|
| | | Sn ₉₀ In ₁₀ |
| 373 ± 2 K | 2.8 | 2.21 |
| | 3.3 | 1.99 |
| | 3.6 | 1.82 |
| | 4.2 | 1.60 |
| | 5.6 | 1.46 |
| | | average 1.8 |

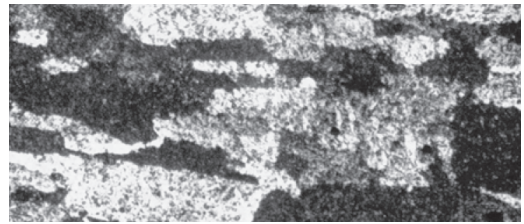
3.6.3 Microstructure test

The microstructure study of a material can provide the information regarding the morphology and distribution of constituent phases as well as the nature and pattern of certain crystal imperfections [19]. The metallographic micrographs of Sn₉₀In₁₀ alloy before creep test (with cold work 94.5%) at annealing temperature 383 ± 2 K for one hour with magnification (500x). The plastic deformation of the alloys caused by cold work are shown in Figure (7a). The microstructure of the Sn₉₀In₁₀ alloy under (3.6 Mpa) and test temperature 393 K are shown in Figure (7b) with magnification (500x). The grain size decrease observed might be attributed to the formation of the dislocation pile-up formed at grains and phase boundaries where

the system remains containing two phase system since the creep temperature is in two phase fields a- phase (Sn-rich phase) and b-phase (In-rich phase). It will disturb the continuity of boundaries and might result in grain cleavage.



(a)



(b)

Figure 7. Microstructure of Sn₉₀In₁₀ in wire form at annealing temperature 383 ± 2 K with magnification 200 x (a) before creep test (b) after creep test under 3.6 MPa at test temperature 393 K.

4 CONCLUSION

The different physical properties of lead free soldering, for the Sn₉₀In₁₀, Sn₉₀In₅Bi₅ and Sn₉₀Bi₁₀ alloys are measured as follows:

- 1 The X-ray diffractions show crystalline as well as amorphous pattern. It is also found that, the increase in the Bi content to Sn-In alloys results in decreasing the amorphous nature.
- 2 The DTA profiles of SnIn solder alloys, show a small sharp endothermic peak with a narrow pasty range between 5 and 17.9 K for the Sn₉₀Bi₁₀ sample alloy.
- 3 The contact angles of SnIn solder alloys at 523 K after 40 s elapsed time are found favorable for soldering uses.
- 4 The values of TCR of SnIn solder alloys ranges between 2.4×10^{-3} and 2.4×10^{-1} (K⁻¹).
- 5 Both the Young's modulus and the internal friction or the damping capacities have been measured for all as-quenched soldering alloys. It is obvious that, the addition of Bi causes the decrease of the elastic

modulus values of the Sn-In alloys. The maximum shear stress t_{max} . increases with adding Bi to Sn-In alloys to the maximum value 6.49 (MPa) of SnIn solder alloys. The internal friction of SnIn solder alloys is decreasing by adding Bi to Sn-In alloys. The theoretical values of μ/E are in good agreement with the experimental results.

- 6 The stress exponent n for $Sn_{90}In_{10}$ alloy ranges between 0.40 and 1.13. The activation energy of the steady state creeps at different stresses for $Sn_{90}In_{10}$ alloy is 1.82 EV.

From the above-mentioned descriptions for SnIn solder alloys, it can be concluded that the alloys which offer the most suitable properties for a low temperature solders application $Sn_{90}Bi_{10}$. This is because they have the lowest electrical resistivity, low pasty range and good wettability with Cu and suitable mechanical properties.

REFERENCES

- [1] S. K. Das, A. Sharief, Y.C. Chan, N. B. Wong. 2009. *J. Alloys Comp.* 481:67–172.
- [2] Chi – Yang Yu Chen, Jeng – Gong Duh. 2012. *Intermetallics.* 26:11–17.
- [3] Jenn-Ming Song, Bo –Ron Huang, Cheng- Yi Liu, Yi – Shao Lai, Ying –Ta Chiu, Tzu-Wen
- [4] Huang. 2012. *Mater. Sci.Eng.* A534:53–59.
- [5] J.F. Li, S.H. Mannan, M.P. Clode, D.C. Whalley, and D.A.Hutt. 2006. *Acta Mater.* 54:2907.
- [6] J.F. Lia, S.H. Mannan, M.P. Clode, K. Chen, D.C. Whalley, C. Liu, and D.A. Hutt. 2007. *Acta Mater.* 55:737.
- [7] J. Shen, Y.C. Chan. 2009. *Microelectron. Reliab.* 49:223–234.
- [8] F.Guo. 2006. *J. Mater. Sci.: Mater. Electron.* 18:129–145.
- [9] A. Elmahfoudi, S. Furtauer, A. Sabbar and H. Flandorfer. 2012. *Thermochimica Acta* 534:33–40.
- [9] D.M. Zhang, G.F. Ding, H. Wang, Zh. Jiang, and J.Y. Yao, *J. Funct. Mater. Dev.* 12:211. [10] G.L. Allen & W.A Jesser. 1984. *J. Cryst. Growth.* 70:546.
- [10] N. Moelans, K.C. Hari Kumar & P. Wollants. 2003. *J. Alloys & Comp.* 360:98.
- [11] K. Suganuma. 1998). *Scripta Mater.* 38:1333.
- [12] M. Hansen. 1958. "Constitution of Binary Alloys" McGraw-Hill Book Co. Inc. New York.
- [13] H. M. Ledbetter. 1977. *J. Mater. Sci. Eng.* 27:133.
- [14] C. Swkker. 1954. "Physical properties of solid materials". Pergamon. New York.
- [15] G. Albert Guy & J. John Hren. 1974 "Elements of Physical Metallurgy". Addison-Wesley Publishing Company.
- [16] B. Arzamasov. 1989. "Material Science". Mir Publisher Moscow.
- [17] S. Mohammed. 2014. *Inter. J. of Innov. Res. in Sci., Eng. and Tech.* 3:9439.
- [18] E. R. Frenkel, M.Sc. 1992. Thesis. University of California. Berkeley.

*5. Computer algorithm, simulation and
information technologies*

A deletion cost based compression algorithm for linear vector data

Q.X. Chen & Y. Hou

Institute of Urban Planning Engineering and Information Technology, Zhejiang University, China

N. Wu

College of Civil Engineering and Architecture, Quzhou University, China

ABSTRACT: A new deletion-cost based compression algorithm (DCA) for linear vector data is proposed. To each curve—the basic element of linear vector data, all the deletion costs of its middle nodes are calculated, and the minimum deletion cost is compared with the pre-defined threshold. If the former is greater than or equal to the latter, all remaining nodes are reserved and the curve's compression process is finished. Otherwise, the node with the minimal deletion cost is deleted, its two neighbors' deletion costs are updated, and the same loop of the compressed curve repeats till the termination. Experiment results show that DCA outperforms Douglas-Peucker algorithm (DPA) in compression accuracy and execution efficiency as well.

1 INTRODUCTION

Douglas-Peucker algorithm (Douglas & Peucker, 1973) are widely used to compress the linear vector data in many fields. It is easy to implement. However, there still exist the following shortcomings of this algorithm: (1) it will delete lots of important nodes by mistake if the distance from the middle node to the line segment (with the kept nodes as end points) is utilized as the criteria to judge whether it should be kept or not, which will make the simplified curves deviate from the original one and affect the subsequent analyses (Yang et al., 2002; Ebisch, 2002; Wu & Marquez, 2003); (2) the time complexities of DPA is $O(n \log n)$, which needs to be improved due to the demands for processing massive spatial data (e.g. vector data retrieved from remote sensing images) and real-time network transmission.

Several solutions for the above problems are as follows: the time efficiency is improved at different levels (Hershberger & Snoeyink, 1992); the method of radial distance constraints is proposed to avoid removing the nodes which had great influence on area deviation (Yang et al., 2002); Taking the total distance between the middle node and the end nodes as the basis of evaluation was recommended so as to avoid deleting important nodes by mistake (Ebisch, 2002). Unfortunately, however, the above improved methods are all based on the classical DPA with the limitation on efficiency that in each compression step, it is inevitable to recalculate distance for all middle nodes. Other recently developed methodologies for

linear features simplification (Balboa & Lopez, 2009; Park & Yu, 2011) are somehow much more complicated and the corresponding algorithm's efficiency should be improved further. In consideration of the above reasons, DCA for linear vector data is proposed in this paper.

2 INTRODUCTION OF DCA

Before the introduction of DCA, it is necessary to introduce a new feature node [Feature nodes are those that have great effect on the contour of certain curve. They are always intended to be reserved after compression.] evaluation index—"Deletion Cost".

Assume that $P = (P_1, \dots, P_n)$ indicates a two dimensional linear curve, n is the total amount of nodes in it, p_{i-1}, p_{i+1} ($1 < i \leq n-1$) are two adjacent or neighbor nodes of P_i and $d(p_x, p_y)$ denotes the distance between two nodes p_x, p_y , then the definition of deletion cost of node P_i is as follows:

$$del\ cost(p_i) = d(p_i, p_{i+1}) + d(p_{i-1}, p_i) - d(p_{i-1}, p_{i+1}) \quad (1)$$

It is easy to see that the smaller $del\ cost(p_i)$ is, the closer to the line composed with P_{i-1} and P_{i+1} the node P_i is, thus the impact on the curve shape from this node is smaller and this node will be more likely to be removed. On the contrary, the impact from this node is larger and it will be more likely to be kept.

The basic steps of DCA are as follows:

- 1 Calculate deletion costs for all the middle nodes on curve $p = (p_1, \dots, p_n)$, and obtain deletion-cost set $C = \{\text{cost}(p_2), \text{cost}(p_3), \dots, \text{cost}(p_j), \dots, \text{cost}(p_{n-1})\}$;
- 2 Find out the minimal deletion cost $\text{cost}(p_x) = \min_{1 < j < n} \text{cost}(p_j)$, and compare it with the predefined threshold T :

If $\text{cost}(p_x) \geq T$, stop the compression process, and all the remaining nodes will be reserved.

If $\text{cost}(p_x) < T$, remove a node, and update its neighbors' deletion costs. Repeat step 2.

Fig. 1 shows the entire compression process of DCA on one curve. Although both DCA and DPA implement as a recursive loop and their termination depends on the comparison with thresholds, the basic idea of these two algorithms is just the opposite: DPA assumes that all middle nodes need to be removed at the very beginning, and then gradually adds feature nodes through comparison between maximal distance and the threshold; while DCA assumes that all middle nodes need to be retained at first and then progressively removes redundant nodes by comparison between minimal deletion cost and the threshold.

DPA is a simplified version of DCA. By taking only perpendicular distance into consideration, DPA actually ignores the influence of the position of the foot of the perpendicular drawn from the middle node

on line composed with two end nodes. That is the reason why DCA outperforms DPA in terms of shape representation or compression accuracy.

DCA never splits curves. The deletion costs of all middle nodes need to be calculated only for once at the very beginning, and in each step only the costs of the two nodes adjacent to that removed node need to be updated. Hence, compared with DPA, DCA is likely more efficiency in implementing, which provides a favorable condition for massive spatial data processing. The more objective and convincing comparison of time consumption between DCA and DPA is proposed in section 3.

3 EXPERIMENTS

Due to spatial resolution limitation, noise interference and errors generated by discretization of raster data, the curves of vector spatial data (especially those transformed from remote sensing images) are usually in an irregular jagged shape. For anti-aliasing and higher compression precision, the Multiple Iterated Gaussian Filtering method (Sarfraz and Khan, 2004) is recommended for the pre-process of smoothing vector graphs (this step can be ignored if it is non-aliasing). Since this pre-process exerts no effect on the analytic results, all experiments and analyses presented below were based on the data smoothed by three times iterated Gaussian filtering in advance.

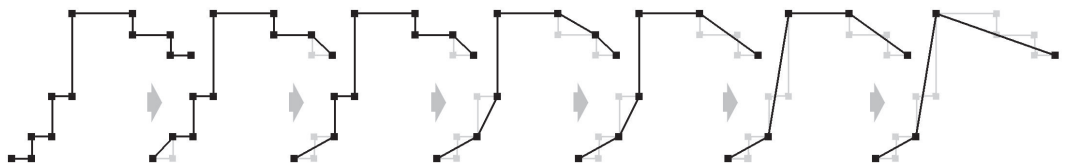


Figure 1. Diagram of DCA compression process.

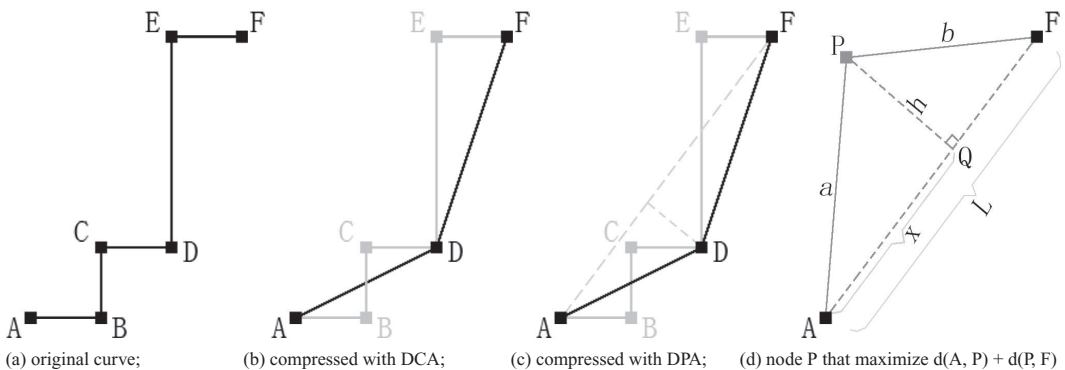


Figure 2. Comparison of algorithms' compression results.

3.1 Comparison under equal processing ratio condition

Processing ratio is the ratio of the number of processed nodes (processing quantity) to the total amount of nodes in original graph. The comparison under equal processing ratio turns out to be the time consumption comparison with the same quantity of processed nodes, which can reflect the per-node processing efficiency and the practical processing ability of algorithms. As DCA is the one that continuously detects and deletes redundant nodes until the minimum deletion cost is greater than the predefined threshold and the remaining nodes are reserved as feature nodes, its processing quantity equals the total amount of redundant nodes

that have been removed in the whole graph. On the contrary, DPA is the one that continuously extracts feature nodes until the perpendicular distance from the node to the line of two end nodes is less than predefined threshold and the remaining nodes are removed as redundant ones. So the processing quantity of DPA is the total amount of remaining nodes in the whole graph after compression. Here in this paper comparative experiments are conducted on the following three vector graphs (shown in Fig.3), obtained from remote sensing images of different spatial resolutions and sizes after segmentation and vectorization) under various threshold (e) conditions. The results under equal processing ratio condition are recorded in Table [1].

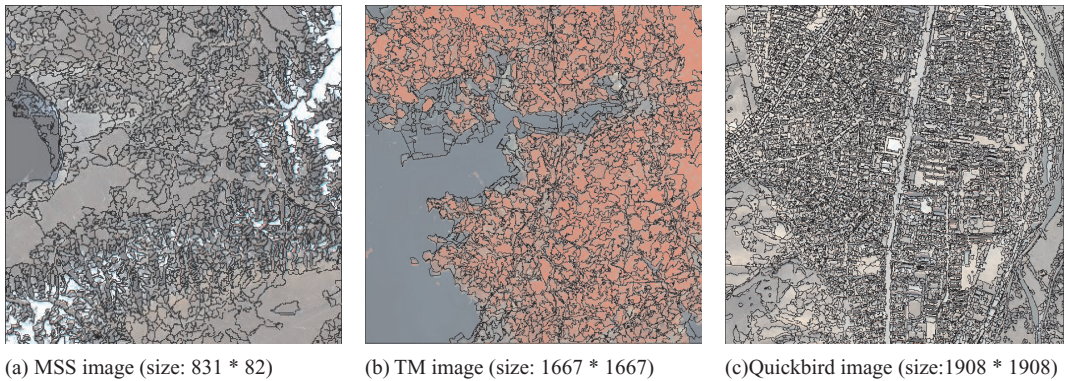


Figure 3. Remote sensing images and their vector graphs.

Table 1. Time consumption comparison under equal processing ratio condition.

| Vector graph | Nodes Number | DCA | | | | DPA | | | | Ratio of time (DCA/DPA,%) |
|--------------|--------------|-----------|---------------|-----------------|----------|-----------|------------|-----------------|----------|---------------------------|
| | | Threshold | Removed nodes | proc. ratio (%) | Time (s) | Threshold | Nodes left | Proc. ratio (%) | Time (s) | |
| Fig. 3-a | 87361 | 5E-04 | 9719 | 11.13 | 0.406 | 2.5 | 9681 | 11.08 | 0.641 | 63.34 |
| | | 9E-04 | 12821 | 14.68 | 0.437 | 1.5 | 12324 | 14.11 | 0.750 | 58.27 |
| | | 0.005 | 26996 | 30.90 | 0.532 | 0.3 | 26888 | 30.78 | 1.047 | 50.81 |
| | | 0.01 | 36019 | 41.23 | 0.593 | 0.175 | 37582 | 43.02 | 1.141 | 51.97 |
| | | 0.02 | 44886 | 51.38 | 0.656 | 0.125 | 43360 | 49.63 | 1.172 | 55.97 |
| | | 0.05 | 59304 | 67.88 | 0.734 | 0.08 | 58545 | 67.02 | 1.219 | 60.21 |
| | | 1.0 | 78302 | 89.63 | 0.844 | 0.02 | 78730 | 90.12 | 1.266 | 66.67 |
| Fig. 3-b | 187611 | 5E-05 | 8412 | 4.48 | 0.734 | 8.00 | 8657 | 4.61 | 1.5 | 48.93 |
| | | 0.005 | 58741 | 31.31 | 1.188 | 0.40 | 51099 | 27.24 | 3.093 | 38.41 |
| | | 0.02 | 94665 | 50.46 | 1.485 | 0.15 | 94338 | 50.28 | 3.391 | 43.79 |
| | | 0.05 | 124241 | 66.22 | 1.688 | 0.08 | 126726 | 67.55 | 3.500 | 48.23 |
| | | 0.1 | 140093 | 74.67 | 1.813 | 0.05 | 147311 | 78.52 | 3.563 | 50.88 |
| | | 0.5 | 163768 | 87.29 | 1.985 | 0.03 | 160561 | 85.58 | 3.578 | 55.48 |
| | | 5.0 | 177675 | 94.70 | 2.094 | 0.01 | 177136 | 94.42 | 3.609 | 58.02 |
| Fig. 3-c | 443168 | 0.001 | 76989 | 17.37 | 2.282 | 1.30 | 73484 | 16.58 | 3.61 | 63.21 |
| | | 0.005 | 143254 | 32.32 | 2.75 | 0.275 | 145869 | 32.92 | 4.984 | 55.18 |
| | | 0.01 | 187474 | 42.30 | 3.062 | 0.18 | 187107 | 42.22 | 5.266 | 58.15 |
| | | 0.02 | 232825 | 52.54 | 3.281 | 0.115 | 230160 | 51.94 | 5.516 | 59.48 |
| | | 0.04 | 288728 | 65.15 | 3.578 | 0.86 | 289035 | 65.22 | 5.718 | 62.57 |
| | | 0.1 | 331810 | 74.87 | 3.844 | 0.055 | 334138 | 75.40 | 5.890 | 65.26 |
| | | 5.00 | 395045 | 89.14 | 4.079 | 0.02 | 393770 | 88.85 | 6.062 | 67.29 |

* if the error of two processing quantities is within 4% of the total amount of nodes in original graph, they can be taken as the same. A laptop with an Intel T23 processor with 1.8 GHz and 4 GB memory was used, the developing tool was Visual .net C++ 2008 in Win XP (similarly hereinafter).

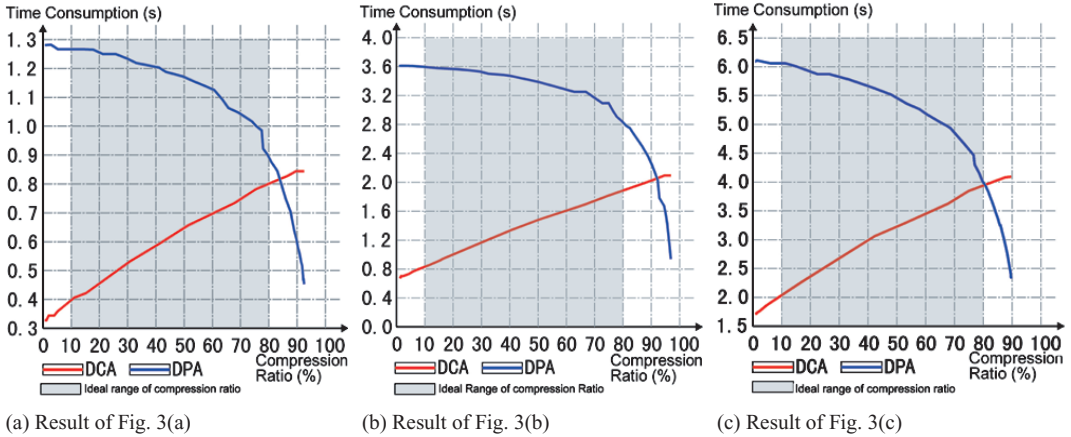


Figure 4. The variation pattern of time consumption and compression ratio.

It can be seen from Table [1] that for vector graphs obtained from vectorization of various remote sensing images, when the processing ratio is within the broad range of 4%~94% (this means the feasibility within more possible cases), the time consumption ratio of DCA and DPA remains at a relatively lower level of 38%~67%, and the optimum can be reached when the processing ratio is about 30%. This indicates that the per-node processing ability of DCA, compared with that of DPA, has a significant advantage.

3.2 Comparison under equal compressing ratio condition

From the above results, some rule can be learned: (1) the time consumption of DCA increases almost linearly as the compressing ratio grows, while that of DPA is inversely proportional to the compressing ratio, and as the latter grows, the rate of the decline gradually increases; (2) when the compressing ratio is close to 0%, the value of T_{DCA}/T_{DPA} gets the minimum, about 0.2~0.3. As the compressing ratio increases, the value of T_{DCA}/T_{DPA} also gradually grows and goes up to 1 when the compressing ratio stays at a certain point between 80%~100%, and remains greater than 1 after that point. Although the value of T_{DCA}/T_{DPA} may reach 1 or even become greater than 1, not all compressing ratios can guarantee a desirable accuracy in actual applications. It is known through large numbers of experiments that when the compression ratio is greater than 80%, larger deviation exists between the curves of vector graph and boundaries of corresponding real objects, and there always turns out to be an over-compressed phenomenon. Similarly, when the compressing ratio is less than

10%, the deviation between curves and real boundaries is so little that an under-compressed phenomenon appears. As a result, 10%~80% is reckoned as a relatively pleasant range of compressing ratio, which is called "Rational Range of Compressing Ratio", and the best range is general 40%~75%.

It can be seen that DCA has an overwhelming advantage over DPA in terms of the performance speed. The integrative reduction of time consumption in all kinds of cases is within the range of 15%~70%, and its average reduction is about 55%, which means there is a 120% increase of average execution efficiency.

4 CONCLUSION

In consideration of the deficiencies of DPA, DCA has been proposed in this paper. Comparative experiments have been conducted under equal processing ratio and equal compressing ratio conditions, and show that the proposed algorithm outperforms DPA in accuracy and efficiency. For a further fulfillment of improvement of efficiency and its application in massive spatial data processing and real-time network transmission, parallelization of DCA is worth to be studied further.

ACKNOWLEDGMENT

This study was supported by the International S&T Cooperation Program of China (No: 2010DFA 92720-27).

REFERENCES

- [1] Balboa, J.L.G. & Lopez, F.J.A. 2009. Sinuosity pattern recognition of road features for segmentation purposes in cartographic generalization. *Pattern Recognition*, 42:2150–2159.
- [2] Douglas, D.H. & Peucker, T.K. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *Cartographica: The International Journal for Geographic Information and Geovisualization* 10(2):112–122.
- [3] Ebisch, K. 2002. A correction to the Douglas-Peucker line generalization algorithm, *Computers & Geosciences*, 28(8):995–997.
- [4] Hershberger, J. & Snoeyink, J. 1992. Speeding up the Douglas-Peucker line-simplification algorithm, *In: Proc. of the 5th International Symposium on Spatial Data Handling*, 34–143.
- [5] Park, W. & Yu, K., 2011. Hybrid line simplification for cartographic generalization, *Pattern Recognition Letters*, 32(9):1267–1273.
- [6] Sarfraz, M. & Khan, M.A. 2004. An automatic algorithm for approximating boundary of bitmap characters, *Computer Graphics and Geometric Modeling*, 20(8):1327–1336.
- [7] Wu, S.T. & Marquez, M.R.G. 2003. A non-self-intersection Douglas-Peucker algorithm, *in: Proc. of the XVI Brazilian Symposium on Computer Graphics and Image Processing*, 60–66.
- [8] Yang, D.Z. & Wang, J.C., Li, G.N. 2002. Study of realization method and improvement of Douglas-Peucker algorithm of vector data compression, *Bulletin of Surveying and Mapping (in Chinese)*, 10(2):112–122.

A design of highly-available and cost-effective data center infrastructure

Y.J. Chen

Department of Electronic Engineering, Ming Chi University of Technology, Taiwan

W.J. Hsu

M.A. Program, Department of Electronic Engineering, Ming Chi University of Technology, Taiwan

ABSTRACT: The motivation is to provide a design of cost-effective and highly-available data center infrastructure for small/medium-scale businesses. The proposed design has the features of full-level HA, system monitoring, recovering, and controlling, and event handling. It controls the routing among network-level HA clusters to solve the “PPPOE connection racing” and “winding path” problems. Being implemented with free, open source software, it is a cost-effective solution, as well as the design of generating events only on state changes. This saves the space and time to store and de-duplicate events, caused by the same-reason sources. This paper describes testing scenarios and exhibits their experiment results for verifying and elaborating of the approaches. The future work is to analyze and strengthen the performance on the monitoring, recovering, and controlling of full-level HA.

1 INTRODUCTION

With the development of Information Technology (IT) in recent years, the concepts of Cloud (Kanso, A. & Lemieux, Y. 2013) have been utilized in the data centers of enterprises. Cloud concepts emphasizing on a High Availability (HA) (Araujo, J.A. et al. 2013, Liu, Z. et al. 2010) of a system, are usually realized by virtualization and clustering technologies (Wei, Y. et al. 2012). For small/medium-scale companies, where the number of employees is below 200, normally in Taiwan, they cannot provide the resources as large-scale ones can. A cost-effective and highly available design of a data center for small/medium-scale business is very essential (Katiyar, E. et al. 2013).

The basic concept of the proposed design is to establish an infrastructure consisting of primary and backup sides through the clustering technology. A clustering technology is designed to fail over the devices in a cluster group. However, there are several clusters in a data center and the routing between them is not considered in the current clustering technology. The primary and backup devices in a cluster determine whether the other one is alive by issuing “Heartbeat packets” (Tong, R. & Zhu, X. 2010) to each other. The devices also need to determine that each connected network segment works properly. A simple approach to detect the states of a device in its associated network segments is to check the states of the device’s interfaces connected to the network segments.

To solve the problems mentioned above and consider a full-level of High Availability (HA) on a

data center, the proposed design applies clustering technology to introduce HA in the network and server levels individually and Load Balancing (LB) (Kanso, A. & Lemieux, Y. 2013) in the application level. In addition, it utilizes a monitor server to monitor the network-level devices for controlling the routing among clusters. The monitor server recovers the objects in each level by issuing reboot or restart commands to the objects. The monitor server also acts as an event center and uses Syslog protocol to record the events of state changes or transitions. In order to achieve management-level HA, dual monitor servers are designed and deployed to support the resilience of event center functionalities.

The proposed design can achieve a highly-available and cost-effective infrastructure since the HA functions and monitor mechanisms are designed in these four levels as mentioned previously and implemented by free, open source software, such as Linux Virtual Server (LVS) (LVS 2014), Vyatta Router and Firewall Server (Vyatta 2014), and NMap Scanner (NMap 2014). The “cost-effective” is also included in the design that events are only generated on state changes. This saves the space and time to store and de-duplicate events. The rest of the paper is organized as follows: Section 2 presents the background knowledge and related work of the proposed design. Section 3 illustrates and presents the details of the data center infrastructure design and its implementation. Section 4 shows the experiment results for verifying and elaborating of the approaches. The final section concludes the paper and describes the future work.

2 BACKGROUND AND RELATED WORK

2.1 Background knowledge

High Availability is a commonly seen cluster mechanism. Application servers in corporate environments implement this mechanism by setting two servers as a cluster. If one server is damaged due to uncontrollable reasons, the other can automatically takeover in a short period of time. Load Balancing is another cluster mechanism allowing for a large number of service requests. The Front-end application server in a corporate environment often plays a role by allocating multiple application servers to simultaneously service a lot of requests from the client side. This mechanism works by accepting user's requests in the load balancer, then using load-balancing rules (Manimala, R.x& Suresh, P. 2013) to allocate requests to application servers for processing. The structure is shown in Fig. 1, where the real server refers to the application server. Failover technology is achieved by using "IP address takeover"; therefore, the active node of a cluster will have a virtual IP address to provide client-side services. When problems occur in the active node, Heartbeat functions will automatically toggle the service to standby node and give it the virtual IP address.

Vyatta provides software-based virtual router, virtual firewall and VPN products for Internet Protocol

networks (IPv4 and IPv6). A free download of Vyatta has been available since March 2006. The system is a specialized Debian-based Linux distribution with networking applications such as Quagga, OpenVPN, and many others.

2.2 Related work

In order to deploy high availabilities of a data center to protect its mission-critical services, the intelligent techniques need to be introduced or monitor in the three main resource pools, i.e., computer, storage, and network, involving multi-layered approaches. The layered approach is the basic foundation of a data center design that seeks to improve scalability, performance, flexibility, resiliency and maintenance (HP 2014, Cisco 2014). For example, in a Cisco approach, the layers of a data center design are the core, aggregation, and access layers. The core layer provides connectivity to multiple aggregation modules and a resilient Layer 3 (i.e., network layer) routing. The designs, at the aggregation and access layers, include "passive" redundancy built into data centers to overcome power or internet provider failures, as well as "active" redundancy that leverages sophisticated monitoring to detect issues and initiate failover procedures (Cisco 2014). Furthermore, additional log shipping can be adapted to achieve data storage protection (HP 2014, Microsoft 2014).

However, using high availability under a virtual environment does not guarantee the vital services will be available at all times (Techtarget 2014). If a service goes down while the virtual server keeps running, the high availability offered by the virtualization software or hardware will not detect anything. Therefore, high availability at the application level is an important consideration under this circumstance. Apparently, to achieve high availability, several different approaches are feasible as illustrated previously, but the common elements are: layer architecture, load balancing, reliable monitoring system, redundancy, logging capability, and more.

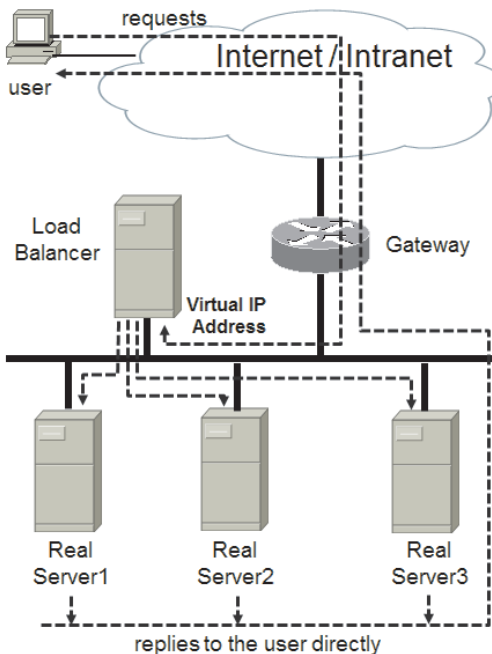


Figure 1. Network and system architecture of Load Balancing.

3 ECONOMIC DESIGN OF FULL-LEVEL HA

The proposed design for a data center infrastructure has the features: full-level HA, system monitoring recovering, and controlling, and event handling, which are presented in the following subsections.

3.1 HA model of data center infrastructure

The proposed HA design is based on a general model of data center infrastructure, as is shown in Fig. 2, suitable for a small/medium-scale company (SMSC). The design uses two tiers of firewalls to create the Public and Private DMZ network segments for

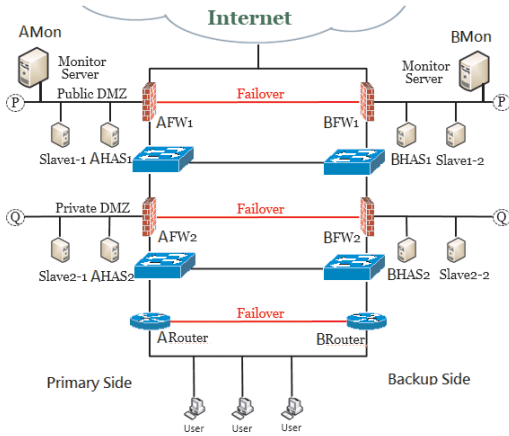


Figure 2. HA model of data center infrastructure for SMSC.

accommodate the Internet servers and intranet servers, respectively. The Internet servers are utilized to settle the public-domain service applications such as official Web, Email, FTP, etc. while the intranet servers settle the private-domain ones such as HR, Financial, Asset systems, etc. In addition, one router tier is designed to create one or multiple network segments for the internal departments of company. The HA designs in network, server and application levels based on this model are described in the following subsections.

3.1.1 Network-level HA

The network system in this model consists of the network devices of two-tier firewalls and one-tier routers. For network-level HA, each tier is grouped as a network cluster and divided into two sides: Primary and Backup. By default, the Primary-side firewalls and router are active in their network-level clusters while the Backup-side ones are standby. The internal user can issue a connection to the Internet through the active router (ARouter), then the active firewall 2 (AFW2), and finally the active firewall (AFW1). On the other side, the firewall BFW1 and BFW2 and the router BRouter are standby for backup. There is a failover network cable, between the active and standby devices of each tier, used as a passage for mutual monitoring. Once the active device fails and is unable to provide service, the virtual IP (VIP) address representing the cluster will be taken over by the backup device to realize uninterrupted network services. The network devices are implemented with general host computers running Vyatta system image. A Vyatta machine can act as a firewall or router and support HA functions.

There are two routing problems among the network-level clusters. First, many small/medium-scale companies use PPPOE connections in ADSL links to connect to the Internet. In the proposed model (Fig. 2), when firewalls AFW1 and BFW1, implemented as Vyatta machines, are in powered-on state and if their PPPOE connections are not controlled, both firewalls will fight for the PPPOE connections through ADSL and stall the network. In order to avoid this situation, a PPPOE connection coordination program must be set in AFW1 and BFW1 individually. The program in the active firewall sets PPPOE commands to issue the PPPOE connection while that in the standby firewall removes all PPPOE commands to close the PPPOE connection. It thus avoids simultaneous PPPOE connections through ADSL. The PPPOE setting is never saved in configuration file. Thus, whenever the firewalls reboot, no PPPOE connection will be issued.

The second problem is that a winding path occurs while the routing among clusters is not controlled. Consider the scenario that AFW2 in a non-preempt HA reboots due to version upgrade. The BFW2 becomes active while AFW2 becomes standby. A winding path for an internal user's connection to the Internet occurs from ARouter, then through BFW2, and finally to AFW1. As is shown in Fig. 2, there are four switches in this path. If AFW2 finishes the reboot phase and works well, the better path is from ARouter, then through AFW2, and finally to AFW1 or from BRouter, then through BFW2, and finally to BFW1. There are only two switches in the individual paths. This routing can be adjusted by controlling the active/standby state of a network device. The monitor and control processes are designed in the monitor servers and describe in Section 3.2.

3.1.2 Server-level and application-level HA

In the server level, there are 8 Ubuntu servers created in the public and private DMZ network segments, named AHAS-1~2, BHAS-1~2, Slave1-1~2, and Slave2-1~2 (Fig. 2). In the public DMZ segment, AHAS-1 and BHAS-1 are grouped in a server cluster and placed into the Primary and Backup sides, respectively. For server-level HA, AHAS-1 is active by default while BHAS-1 is standby for backup. It is implemented with Linux LVS and Keepalived utilities, which provide the Heartbeat mechanism to determine if the other side server is alive. Once the active server fails and is unable to provide service, the virtual IP (VIP) address representing the cluster will be taken over by the backup server. The similar arrangement is done in the private DMZ segment.

For application-level HA, an application service is distributed into the servers of DMZ segment, which dynamically share the service requests and thus achieve the HA of application service with the way of service Load Balancing (LB). In the

public DMZ segment, by default, AHAS-1 acts as an active Master server while BHAS-1 acts as a standby Master, but they do not offer any application service. The active Master plays a role of load balancer to dispatch service requests to the Slave servers actually hosting the application service. Slave1-1 and Slave1-2 are the Slave servers in the public DMZ segment. The design of the LB adopts the Direct Routing (Fig. 1) architecture for reducing the traffic load of the Master servers. It is implemented with Linux LVS and Keepalived utilities, which are installed in the Master servers to provide the health checks of application services in the Slave servers and thus know how to dispatch service requests. For the private DMZ segment, the LB operation is implemented with the same way as that in the public DMZ.

3.2 Monitoring, recovering, and controlling HA

A monitor cluster is created in the Public DMZ and consists of two Monitor servers, in the Primary and Backup sides, named AMon and BMon, respectively. By default, AMon is active to monitor the objects in network, server, and application levels and then recover them, if failures (also including HA failures) are detected, through reboot or restart commands in the Vyatta or Ubuntu systems. Simultaneously, BMon is standby for backup of AMon to achieve the management-level HA. The active Monitor server uses the Ping utility to monitor the objects in network and server levels while using the NMap utility to monitor the services in application level.

After the monitoring and recovering phases, the active Monitor server starts the HA control process. First, it checks the route to the Internet to prevent AFW1 and BFW1 from issuing the PPPOE connections simultaneously. The control flow is shown in Fig. 3. Initially, it notifies the two firewalls to clear their PPPOE setting and then checks which firewall is active in HA. It notifies the active firewall to run the PPPOE program to issue the PPPOE connection. Then, the control process Pings the IP address to determine if the route to the Internet is successful. If the Ping detection fails, the process tries to reboot the firewall failing on PPPOE.

Secondly, the active Monitor server controls the routing among the network-level clusters to avoid a winding path through the clusters. The way is to control all active network devices in one side, either the Primary or Backup side (Fig. 2), and all standby ones in the other side. The control process is called One-side-active process, as is shown in Fig. 4. It tries to control the firewall successfully issuing

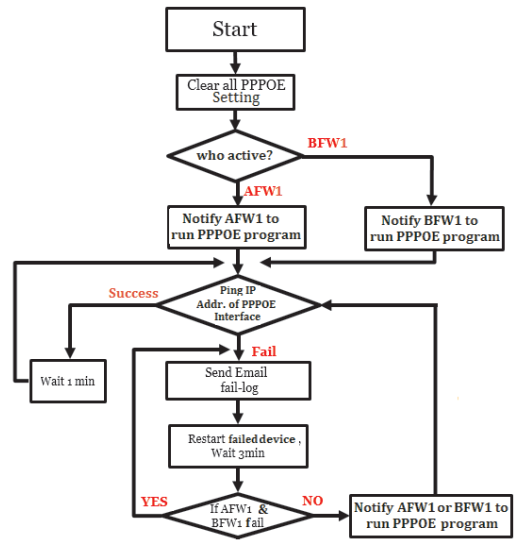


Figure 3. The control of PPPOE by Monitor server.

PPPOE connection, either AFW1 or BFW1, together with other active network devices in the same side. Consider the following scenario. Assume that AFW1 has successfully established a PPPOE connection to the Internet. Obviously, AFW1 is active in its cluster. If AFW2 is standby in its cluster, the One-side-active process will reboot BFW2 to make AFW2 active and BFW2 thus standby.

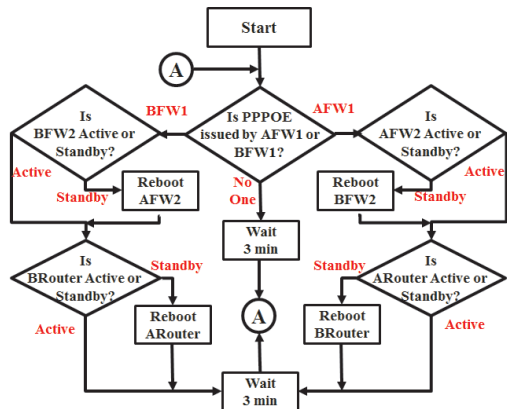


Figure 4. The control of One-side-active by Monitor server.

3.3 Event handling

In the HA model of data center infrastructure, in order to coordinate the operations among the HA clusters of all levels, the concept of Event Center

(EC) is introduced into and implemented in the Monitor server. As an EC, the Monitor server monitors and controls the whole data center and records and analyzes events to alert the administrator while a critical event occurs. Its system architecture is shown in Fig. 5. The Monitor & Controller module undertakes the monitoring, recovering, and controlling of the objects in all levels for achieving full HA. The operation states of each object can be simply defined as Normal and Abnormal. An event from an object is defined as the state change of the object.

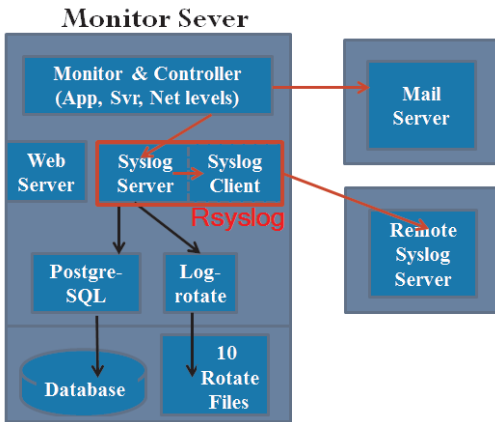


Figure 5. System architecture of Monitor Server.

The monitor server is implemented with Linux Ubuntu system and utilizes the mechanisms of Syslog, Database, and Rotated Files to systematically record events. Whenever the monitor and control module detects a state change of an object, it creates an event to describe the state change. The event is sent to the local Syslog server with the Linux Logger utility or encapsulated as an email and then sent to a mail server. The design uses state change to trigger the creation of an event and thus reduces the quantity of events. This is a cost-effective design, compared to an event design which continuously generates the same-reason events from periodical detection of an Abnormal object until the object becomes Normal.

The states of objects in all level are exhibited with the signs drawn in a web page. The Normal and Abnormal states are presented as Green and Red signs, respectively. The logs record the stage changes of objects in different levels are exhibited in individual web pages. These pages are formed by a web server, which is implemented with the Tomcat utility.

4 EXPERIMENT RESULTS

In order to verify the proposed design works well, the data center infrastructure is implemented mainly with the Linux Ubuntu and Vyatta system software. The cluster HA functions in the two operating systems are used to provide individual-level cluster HA. The Monitor & Controller module designed in the Monitor server solves the routing problems of “PPPOE connection racing” and “winding path” among the network-level clusters. The experiment results exhibited in the following subsections explain how to verify the approaches.

4.1 PPPOE connection test

The testing scenario is as follows: The active side is the Primary side. Close AFW1, the active firewall in the first tier. The test is to check if the Monitor server can detect the failure of AFW1, then recover it, and next notify BFW1 to establish the PPPOE connection since BFW1 has become active.

The time is reset to zero when AFW1 is closed. As is shown in Fig. 6, the PPPOE connection is failed at 2.243 sec. However, the Monitor server detects the failure of AFW1 after 1 min 21.520 sec (Fig. 7).

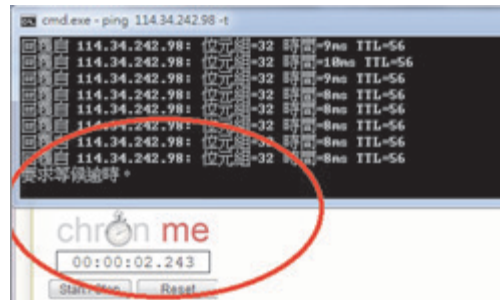


Figure 6. The route to the Internet is failed.

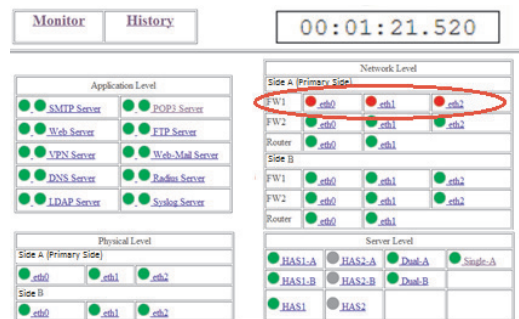


Figure 7. AFW1 is detected to be failed.

It then recovers AFW1 through rebooting. AFW1 is detected to be successful after 2 min 26.020 sec (Fig. 8). Finally, the PPPOE connection is recovered at 3 min 4.009 sec (Fig. 9).

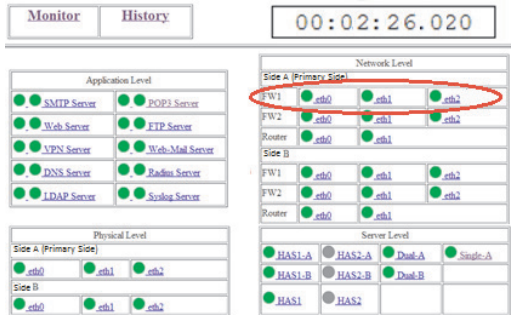


Figure 8. AFW1 is recovered and detected to be successful.

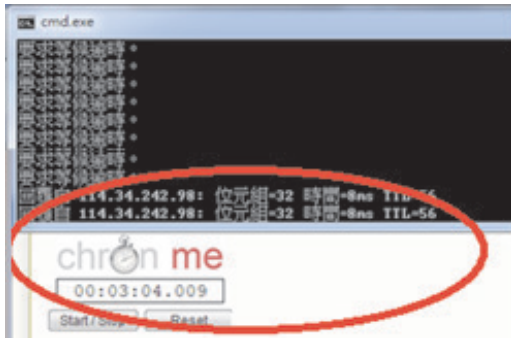


Figure 9. The route to the Internet is recovered.

4.2 One-side-active process test

This test scenario continues the PPPOE connection test. When the PPPOE connection is recovered as is shown in Fig. 9, One-side-active process will enforce BFW2 and BRouter to be active since BFW1 is active (Fig. 10), and issues the PPPOE connection.

By following the timing of PPPOE connection test, in which the time is reset to zero when AFW1 is closed, the scenario shows AFW2 and ARouter are rebooted and their failures are detected at 4 min 47.361 sec (Fig. 11). Then, they are detected to be successful at 5 min 58.760 sec (Fig. 12).

5 CONCLUSIONS

The motivation is to provide a design of cost-effective and highly-available data center infrastructure for small/medium-scale businesses. The proposed design

```
itex@BFW1:~$ show cluster status
=== Status report on secondary node BFW1 ===

Primary AFW1: Active (standby)
Secondary BFW1 (this node): Active
Resources [192.168.99.254/24/eth1 192.168.254.2/28/eth2]:
Active on secondary BFW1 (this node)
```

Figure 10. BFW1 is active while AFW1 is standby.

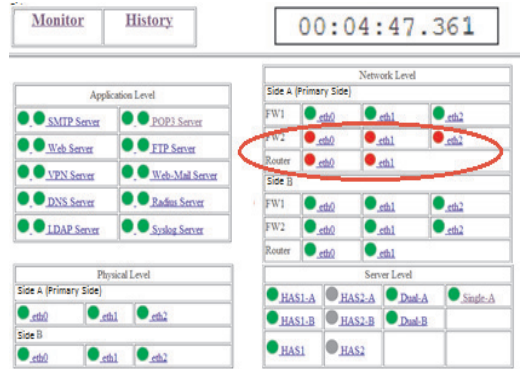


Figure 11. One-side-active process reboots AFW2, ARouter.

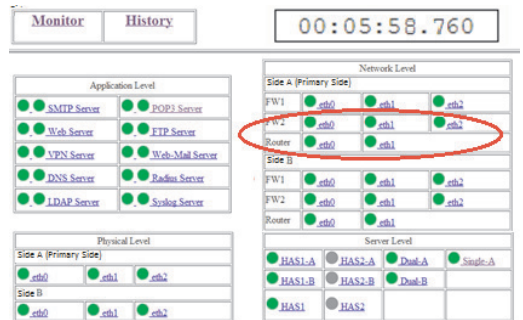


Figure 12. AFW2 and ARouter reboot successfully.

has the features of 1) providing full-level HA on network, server, application, and management levels, 2) controlling the routing among network-level clusters to solve the “PPPOE connection racing” and “winding path” problems, 3) monitoring and recovering the objects in each level with an economic and effective way, and 4) handling events resulting from object state changes with an event center. The design is cost-effective since it can be done with freeware. In addition, the event center generates events only on state changes. This saves the space and time to store

and de-duplicate events, caused by the same-reason source. The experiment results show the Monitor & Controller module in Monitor server can avoid PPPOE connection racing and winding path among HA clusters automatically.

The future work is to analyze and enhance the performance of the Monitor & Control module. This includes: 1) what is the proper lower and upper bounds of the monitoring interval to the objects of all levels, 2) how to shorten the time for silencing the rebooted or restarted objects for recovery, 3) to find a better method of recovery instead of reboot.

ACKNOWLEDGMENT

This work was supported in part by R.O.C. (Taiwan) Ministry of Science and Technology under Grant MOST103-2622-E-131-008-CC3.

REFERENCES

- [1] Araujo, J.A., Lazaro, J., Astarloa, A., Zuloaga, A. & Garcia, A. 2013. High availability automation networks: PRP and HSR ring implementations. *ISIE2012; Proc.*
- [2] Cisco 2014. *Cisco Data Center Infrastructure 2.5 Design Guide*. Cisco Validated Design.
- [3] HP 2014. Designing Disaster Tolerant High Availability Clusters. Manufacturing Part Number: B7660-90006. *HP Document*.
- [4] Katihar, E., Khendek, F. & Toeroe, M. 2013. Operating system upgrade in high availability environment. *SoftCOM2013; Proc.*
- [5] Kansa, A. & Lemieux, Y. 2013. Achieving High Availability at the Application Level in the Cloud. *CLOUD2013; Proc.*
- [6] Liu, Z., Yang, X., Sha, J. & Wan, Y. 2010. Fault detection for high availability RAID system. *NCM2010; Proc.*
- [7] LVS 2014. <http://www.linuxvirtualserver.org/> as of 2014.
- [8] Manimala, R. & Suresh, P. 2013. Load balanced job scheduling approach for grid environment. *ICICES2013; Proc.*
- [9] Microsoft 2014. HA & DR. [http://technet.microsoft.com/en-us/library/hh393576\(v=sql.110\).aspx](http://technet.microsoft.com/en-us/library/hh393576(v=sql.110).aspx) as of Aug. 12, 2014.
- [10] NMap 2014. <http://nmap.org/> as of Aug. 15, 2014.
- [11] Techtarger 2014. Ensuring high availability in the data center. <http://searchdatacenter.techtarget.com/tip/Ensuring-high-availability-in-the-data-center> as of Aug. 12, 2014.
- [12] Tong, R. & Zhu, X. 2010. A Load Balancing Strategy Based on the Combination of Static and Dynamic. *DBTA2010; Proc.*
- [13] Vyatta 2014. Vyatta Community Documentation. <http://www.vyatta.org/> as of Aug. 15, 2014.
- [14] Wei, Y., Dong, Y., Ma, R., Liang, A. & Guan, H. 2012. High Availability through Output Continuity. *IEEE Cluster Computing Workshops*.

A method to detect memory leak of C++ class field member based on static analysis

S.W. Liu & D.H. Jin

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China

ABSTRACT: The defect of memory leak in C++ language is so common but hard to be detected. Although many methods have been proposed to detect the common kinds of memory leak so far. While, this article shows you a kind of memory leak from another angle which is caused by the class inheritance. The article gives you the reason why this kind of defect produces and comes out a method to detect the defect. Based on static analysis this paper using the technology of abstract syntax tree and XPATH to locate the target node. Then the defect could be found out by comparing the allocation operation for the same pointer.

KEYWORDS: Memory leak, static analysis, software testing, class field member memory leak.

1 INTRODUCTION

In recent years, the rapid development of the Internet comes into being all kinds of large-scale softwares which bring convenience to people's lives. People may be subjected to a huge threat caused by the reliability of software along with enjoying the convenience. In this era of highly dependent on the Internet, the software reliability problems have been concerned on by more and more researchers.

Software reliability directly determines the normal operation, robustness and anti-attack capability of the system. Speaking from the perspective of program code, in addition to ensuring the syntax and semantics of the code, the fatal problems, such as the null pointer, memory leaks, resource leaks and the designing of interface should also take your attention. Therefore, designing and implementing a framework according to specification is the first step to ensure your work security. Finding out the bugs^[1] hidden in the code during your coding process is an effective method to improve the quality.

Software Testing is an essential link to insure the reliability and discovers the implied defects during the development phase. Statistics indicate that a well-designed strategy of software testing will save more than 40 percent of the cost used in the development and more than 60 percent of the cost used in the maintenance of the system. Usually, there are two ways to detect defects which are static testing and dynamic testing. This article is using static testing^[2] to detect the memory leak which is caused by class members' memory mismanagement.

2 MEMORY LEAK CLASSIFIED

A memory leak is a phenomenon that memory space possessed by some object instance which can't be recycled because of the error or incomplete code designing. Unlike the JVM, the C++ developers need to control their own applications and release memory manually. Especially memory leak has no obvious feature in a project, then a coder could be forgotten to release the applied memory after using it. Then it produces a defect. As the accumulating of memory leak, the system will run slower or crash down at worst. There are three kinds of memory leaks as follows.

```
1. void MyFunction(int nSize)
2. {
3.   char* p= new char[nSize];
4.   //doSomething...
5.   if(condition)
6.   {
7.     free(p);
8.   }
9. }
```

Figure 1. Example of memory leak.

2.1 Common type

After the program allocating memory using 'malloc' or 'new', remember to free the memory using the corresponding function 'free' or 'delete'. Fig. 1

illustrates a common type of memory leak. After allocating memory for pointer 'p', there is no release operation for it until the end of the function.

2.2 Condition type

Fig.2 presents a memory leak of condition type. In this code, if we get true from 'GetStringFrom()', the function returns without releasing the memory, then a memory leak defect appears. Otherwise, the function works normally.

```

1. void MyFunction(int nSize)
2. {
3. char* p= new char[nSize];
4. if( !GetStringFrom( p, nSize ) )
5. {
6.     MessageBox("Error");
7.     return;
8. }
9. ...//using the string pointed by p;
10. delete p;
11. }

```

Figure 2. Example of memory leak.

2.3 Inherited type

This kind of memory leak plays the main role of this paper. Fig.3 gives the code segment, the

```

1. class A
2. {
3.     public:
4.         int* p;
5.         A()
6. {
7.         p = new int[10];
8.     }
9. };
10. class B:public A
11. {
12.     public:
13.         B():A()
14.         {
15.             p = new int[11];
16.         }
17. };

```

Figure 3. Example of memory leak.

allocation operation of memory for variable 'p' which is inherited from its parent class will be executed twice when the constructor B() executes. Then memory leak happens.

The other kind of memory leak in this type is caused by destructor function. As Fig.4(a) and (b) give an example that when the instance of class A named 'a' is released, the destructor function defined in class A will be called rather than that in class B, so the pointer 'q' defined in class B can't be released which causes a defect.

```

1. class A
2. {
3.     public:
4.         int* p;
5.         A()
6.         {
7.             p = new int[10];
8.         }
9.         ~A()
10.        {
11.            delete[] p;
12.        }
13. };

```

(a)

```

1. class B:public A
2. {
3.     public:
4.         int* q;
5.         B():A()
6.         {
7.             q = new int[11];
8.         }
9.         ~B()
10.        {
11.            delete[] q;
12.        }
13. };
14. int main(){
15.     A *a = new B();
16.     delete[] a;
17.     return 0;
18. }

```

(b)

Figure 4. Example of memory leak.

3 SOLUTION OF DETECT DEFECTS

The process of detecting defects using static analysis consists of the following steps: program pre-analysis, producing AST (abstract syntax tree), generating CFG (control flow graph), calculating the available path and designing the defect mode state machine.

3.1 Using of AST and XPATH

The syntax tree of a code segment is a data structure of the tree and is produced according to the syntax rules of the language. It is used to store the structural information of source programs and as a basis for the latter part of static analysis.

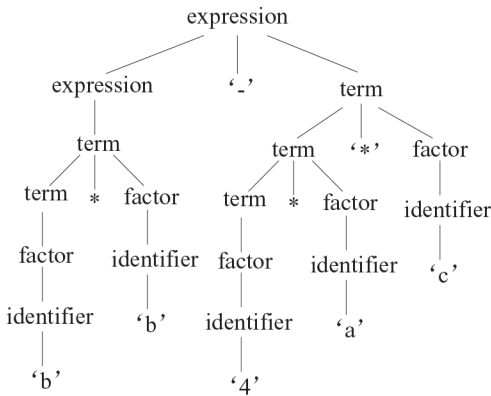


Figure 4. The analysis tree of express $b*b-4*a*c$.

3.2 Generate CFG

CFG^[3] is an essential structure in the process of program analysis and testing. CFG could be obtained from an abstract syntax tree and it is a directed graph which reflects the logic control process of program. Usually a control flow graph can be expressed as a CFG (V, E). Wherein V represents the set of nodes which contains the information about the variable's value and E represents a set of directed edges. E has the structure of $E = \langle V_{begin}, V_{end} \rangle$ which is the reflection of the relationship between statements. Fig. 5 shows a control flow graph

3.3 Design the state machine

Defect mode^[4] is an algorithm of detecting defect which is in the form of a state machine. A state machine describes a series of states and transition conditions. With this system, all the state machines start from the 'START' state, and then the state converts to another according to the calculation of the conversion

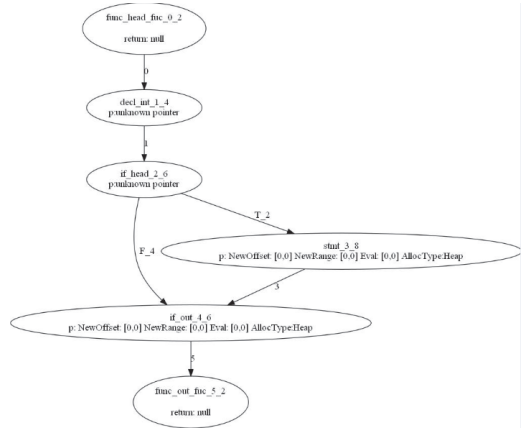


Figure 5. Control flow graph.

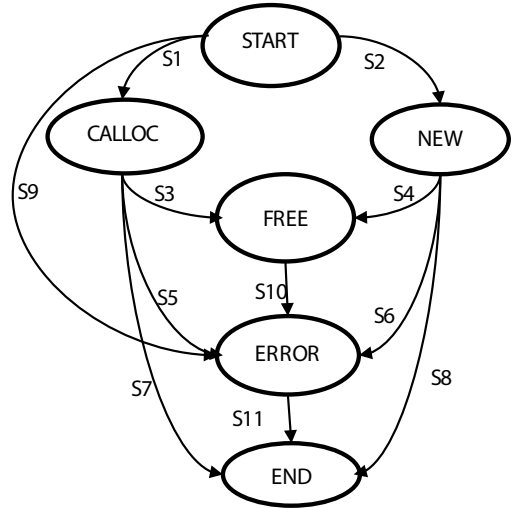


Figure 6. Converting of the defect state machine.

condition. Once the 'ERROR' state occurs, a defect will be reported. Expansibility is the advance of using defect detection based on a defect mode machine. The memory leak state machine is shown as Fig.6.

3.4 Algorithm of detecting class field memory leak

The algorithm can be divided into two parts. One is to find out whether the constructor has an allocation operation for the pointer variable inherited from the parent class which has been allocated memory in the parent class constructor. The other is to check out whether the parent class has a virtual destructor if the child class's destructor has the memory release operation. As the algorithm is to detect inherit type defect, so the target


```

1. FIELDMEMORYLEAKDETECT (Source src)
2. begin
3.   for (each Constructor in src)
4.     do generate the CFG
5.     do collect vars malloced in parent
       constructor(var_parent)
6.     do collect vars malloced in this
       Constructor(vars)
7.     for(each var in vars)
8.       if (var in var_parent)
9.         then report this var node
10.    for (each Destructor in src)
11.      do collect free operation (FREE())
12.      if ( FREE() is not empty)
13.        if (parent's destructor is not virtual)
14.          then report this destructor node
15. end

```

Figure 7. Algorithm of detecting defect.

class must be inherited from a parent class. The process of the algorithm is described as Fig.7.

4 RESULT

The method of detecting defects mentioned in this paper is implemented and tested on DTS^[5] platform. To evaluate our approach, we implemented the algorithm and conducted the experiments using five open-source projects. Table 1 shows the experiment results.

Table. 1 DTS testing result.

| Project | File | Lines | IP | Defect | FP |
|-----------|------|--------|-----|--------|-----|
| Avisynth | 86 | 55798 | 215 | 182 | 15% |
| BWChess | 17 | 6738 | 18 | 11 | 17% |
| Dgvideo | 66 | 23676 | 94 | 65 | 30% |
| Playa | 32 | 18711 | 53 | 40 | 24% |
| SpGateWay | 25 | 6995 | 8 | 5 | 37% |
| Total | 226 | 111918 | 388 | 303 | 21% |

In this table, 'IP' refers to the checkpoint, which is the total number of DTS actually reported. 'Defect' refers to the number of certain defects after judging by manual. 'FP' is the rate of fault reported which is calculated by $(IP - Defect) / IP * 100\%$.

5 SUMMARY AND FUTURE WORK

The class field memory leak of C++ language mentioned in this paper is an extension of the case appeared in the function scope. Therefore, it's hard to be detected by a simple single file analysis. Although the method that combines recording the structural information about inheritance and ordinary memory leak defect detection method can achieve the purpose to some extent. While, due to the complexity of C++ syntax, the proposed method cannot effectively detect this type of memory leak in the case of the target has a complex or multi-layered inherited structure. The gist of this article is to present memory leak from another angle and provide a simple detection method for the defect. From this article, we also hope to prompt people to analysis memory leak from multiaspects and be able to find a better algorithm to detect class field memory leak after a deeper research.

ACKNOWLEDGMENTS

This work was supported by the National Grand Fundamental Research 863 Program of China (No. 2012AA011201), the National Natural Science Foundation of China (No. 61202080), the Major Program of the National Natural Science Foundation of China (No. 91318301), and the Open Funding of State Key Laboratory of Computer Architecture (No. CARCH201201).

REFERENCES

- [1] I. Dillig, T. Dillig, and A. Aiken. Static error detection using semantic inconsistency inference. In Proc. of the 2007 ACM SIGPLAN Conf. on Prog. Lang. Design and Impl. , pages 435–445, June 2007.
- [2] Y.Z.Gong,. *Software Testing*. Beijing: China Machine Press. 2008:65–78.
- [3] YANG Zhao-hong, GONG Yun-zhan, XIAO Qing, WANG Ya-wen. A Defect Based Model Testing System[J].JOURNAL OF BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS, 1980, 31(5): 1–4.
- [4] Wang YW, Chen JL. Research on software testing technology based on defect pattern [Ph.D. Thesis]. Beijing: Beijing University of Posts and Telecommunications, 2009
- [5] Y. W. Wang, Y. Z. Gong, et al., A method of variable range analysis based on abstract interpretation and its applications, Dianzi Xuebao(Acta Electronica Sinica), vol. 39(2), pp. 296–303, Februray. 2011.

SSD lifespan-oriented cache optimization in SSD/HDD hybrid storage system

X. Wang, Q.X. Zhang & Y.A. Tan

Beijing Engineering Research Center of Massive Language Information Processing and Cloud Computing Application, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China

J.Q. Qu

Shanghai Radio Equipment Research Institute, Shanghai, China

ABSTRACT: Layered SSD/HDD hybrid storage system has been widely used. However, most traditional designs are performance oriented, and not optimized for SSD lifespan. In this paper, SSD lifespan oriented layered model and cache algorithm are proposed. To decrease SSD wear rate, HDD-RAM switch tunnel is appended to the traditional layered hybrid structure. Cache window size is considered in cache algorithm to filter data, avoiding unnecessary cache cost. Results show optimized designs significantly reduced erase times on SSD, and slightly influenced original performance.

1 INTRODUCTION

As semiconductor technology develops, computer storage design has been enormously changed. Solid state drive (SSD) is gradually displacing traditional magnetic rotation drive (HDD). Compared with HDD, SSD has higher performance and better reliability, while it consumes lower power and creates fewer fragments. [1] SSD greatly improves nonvolatile I/O, which eliminates bottleneck between RAM and HDD.

Although SSD has such advantages, its cost is very high. The cost of traditional HDD is only 0.06\$/GB, while the cost of SSD is generally 0.5\$/GB. [2] Considering limited budget, it is unrealistic for traditional storage systems to replace all HDDs with SSDs. [3] In order to balance cost and performance, a hybrid storage system needs to be deployed, which contains both high-capacity HDD and high-performance SSD. [4]

In most hybrid storage system, highly accessed data are cached in SSDs, and cold data are stored on HDDs. [5] There are two typical deployments: One is layered storage structure, which deploys SSD as upper cache and HDD as lower switch storage; the other is parallel storage structure, which deploys SSD and HDD in the same layer. The layered storage structure is easier to realize, generally just appending SSD to original systems. While parallel storage design needs to modify original structures and access controls, which reduces system reliability and portability. [6]

But recent researches show layered storage design doesn't agree with HDD-SSD-Hybrid storage system. Compared with HDD, SSD is more delicate to write wear. Traditional cache designs increase too many write and erase operations on SSD, which greatly reduces SSD's lifespan. [7] Above all, most data especially movie data cached in SSD are barely accessed again, which results in unnecessary cache costs. [8] These data need to be accessed directly between HDD and RAM, avoiding SSD cache. [9]

In this paper, a SSD cache algorithm in the layered hybrid storage system is optimized. Based on PriorHDD model and cache window size, two strategies are appended to the traditional cache algorithm: frequently accessed small data are cached in SSD; infrequently accessed data and big data are switched directly between HDD and RAM. Compared with traditional cache algorithm, optimized design effectively reduces unnecessary write operations on SSD, which improves SSD's lifespan. Experiments show erase times on SSD were significantly reduced, while system original performance was maintained.

2 REFERENCE

Boukhobza has studied SSD's characters, and proposed SSD is easily worn out with many erase operations. [10] He designed a flash-specific cache system to maximize lifetime by absorbing some erase operations. The system was flexible as it offers many

configuration possibilities and can be adapted according to the I/O workload. Since SSD was less utilized, the system gave less than 15% performance in most cases.

Puyuan Yang proposed a practical way to integrate SSD and magnetic disk and then to obtain a better tradeoff between two storage medium. [11] He investigated the issues of integrating SSD and disk in the storage layer of a database management system. Since big data cached on magnetic disk, writes on SSD were significantly reduced. The results showed his design ensures most read operations were performed on SSD and most write operations were focused on disk, while the system has less runtime than the single-disk-based mechanism.

Liu Yang designed a hybrid storage architecture combining both of SSD based disk cache and disk drive subsystem.[12] This architecture greatly extended lifetime of the SSD, while improving system performance through providing priority to cache random access data. Based on data size, random access was split. Big data were serviced by hard disks directly to even the full workload between SSD and disk storage, while small data were cached in SSD. The results showed that such design can significantly reduce flash wear and improve performance, compared with the state-of-art flash cache architecture.

3 DESIGN AND IMPLEMENT

Traditional layered hybrid storage design directly adopts Cache-Switch access model for HDD-SSD-RAM structure, which greatly increases SSD wear rate. In this section, novel access model and cache algorithm are proposed to reduce unnecessary write operations on SSD.

3.1 Prior HDD model

3.1.1 Model structure

Traditional hybrid designs generally divide HDD-SSD-RAM structure into HDD-SSD and SSD-RAM parts. In each part, Cache-Switch access model is simply utilized. All data are written to and read from SSD, which greatly reduces SSD's lifespan. The traditional model is shown in Fig.1 a.

To decrease the I/O throughput on SSD, another data tunnel between HDD and RAM is appended, and all data are prior stored in HDD, instead of caching in SSD. The prior HDD model is shown in Fig.1b. In this model, the HDD has higher priority to be accessed. Only frequently accessed data are cached in SSD. Since most data, especially movie data are barely accessed, this model reduces enormous unnecessary data cache on SSD. HDD-RAM tunnel also provides a parallel I/O, which improves storage system throughput.

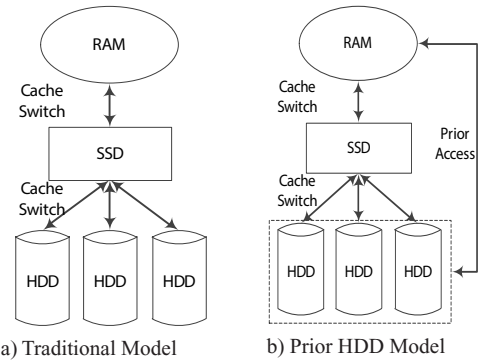


Figure 1. Data access models.

3.1.2 Priority factor

Since SSD performs much better than HDD, HDD-RAM tunnel only benefit time insensitive data. To maintain high performance, time sensitive I/O in HDD-RAM tunnel must be limited. Priority factor α is defined as the percentage of time sensitive data switched in the HDD-RAM tunnel. α is presented as

$$\alpha = \frac{m}{N} \quad (1)$$

N is total time sensitive data switched between RAM and HDD. m is time sensitive data switched in RAM-HDD tunnel.

The hybrid system performance P influenced by the prior HDD model is calculated as

$$P = P_{SSD} - P_{HDD} / P_{SSD} \cdot \alpha \cdot \beta \quad (2)$$

$(P_{SSD} - P_{HDD}) / P_{SSD}$ is sequential performance disparity between SSD and HDD. β is time sensitive data percentage in total data. Since random access is generally achieved by the cache, in HDD-RAM tunnel only sequential performance disparity is considered.

According to reference [13], $(P_{SSD} - P_{HDD}) / P_{SSD}$ is generally 40% to 60%, and β is 15% to 40%. To maintain original high performance, decreased performance P needs to be less than 20%. [14] So α is calculated to be less than 83%, which is easily achieved in general systems.

3.2 Cache algorithm

3.2.1 Cache window size

Traditional cache algorithms generally apply small cache window size, ex.4KB, to improve cache performance and hit rate. But such design also increases

write frequency and data fragments, which has detrimental effects on SSD. To decrease SSD wear rate, bigger cache window size needs to be utilized.

Based on reference [15], 50000 common files were collected and analyzed to get proper window size. The file data size and access frequency distributions are shown in Fig.2 and Fig.3. Since big cache window significantly reduces system random IOPS, experiment analysis size was limited less than 1024KB.

Results show 64KB and 256KB are points of data inflection, which have higher profits. 64KB window size covers 35% file data and 47% access, while 256KB window size covers 65% file data and 79% access. Since small cache window benefits random

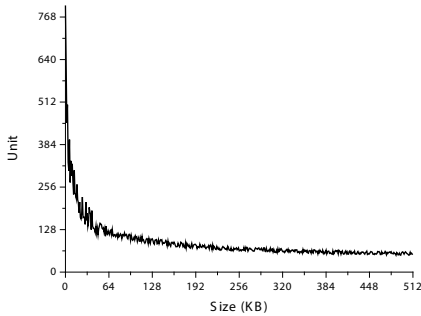


Figure 2. File data size distribution.vv

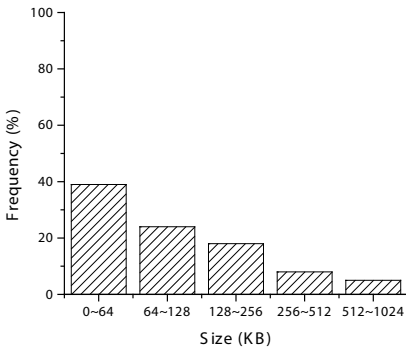


Figure 3. Access frequency distribution on data size.

IOPS, although 64KB window size covers less data and access, it provides higher random throughput. So 64KB is better employed in random busy environments, while 256KB is better in sequential busy environments.

3.2.2 Cache strategy

Based on cache window size, optimized cache strategy is proposed in Fig.4.

As traditional cache strategies, data frequency is previously checked. There are many frequency calculation algorithms, and ARC algorithm [16] is generally applied. Infrequently accessed data are switched in the HDD-RAM tunnel, and frequently accessed data is further checked. Only small data which size meet cache window size are cached by SSD. For sequential access, max data size is 256KB; for random access, max data size is 64KB. Big data are still switched in the HDD-RAM tunnel.

4 EXPERIMENT

In this section, assimilation experiment and a standard trace experiment are conducted to check SSD erase times and system performance.

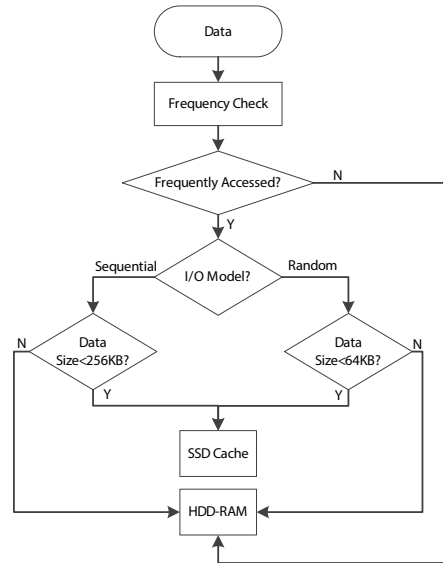


Figure 4. Optimized cache algorithm.

4.1 Environment

4.1.1 Hardware

We used a server which has Intel Xeon E5-2620 processors running at 2.00GHz, 64GB 1333MHz DDR3 RAM, Samsung 840pro 128GB SSDs and Seagate Constellation 7200rpm 2TB HDDs. The operating system is Ubuntu server 12.04 LTS with 3.2.14 Linux kernel. We used standard C coding language and GCC 4.4.5 compiler.

4.1.2 Software

We used FlashSim [17] to create simulation environments. This software is able to efficiently simulate I/O operations and record I/O information. Standard traces such as Financial1, Financial2 and WebSearch1 were utilized to provide real workloads.

4.2 Results

We first conducted a simulation experiment to measure SSD erase times. The result is shown in Fig.5.

Compared with traditional one, SSD erases times in an optimized hybrid system is significantly decreased by 10% to 40%, which proves optimized designs effectually decreasing SSD wear rate. The results also show that optimized designs perform better in busy write environments.

We then conducted a standard trace experiment to measure storage system performance. The result is shown in Fig.6.

Compared with un-optimized one, performance jitter in optimized system is more significant. This is because HDD is also directly accessed by RAM in optimized system, which results in performance decline. But since the average performance remains similar, it is reasonable to efficiently improve SSD lifespan at such slight jitter cost.

| | Financial1 | Financial2 | WebSearch1 |
|------------|------------|------------|------------|
| Read Ratio | 23.16% | 82.34% | 99.98% |

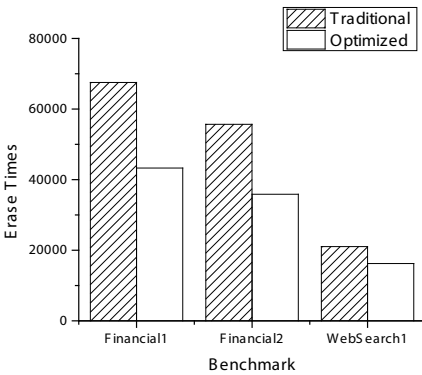
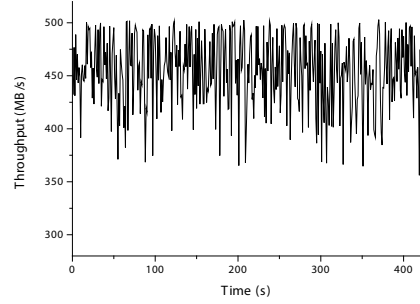
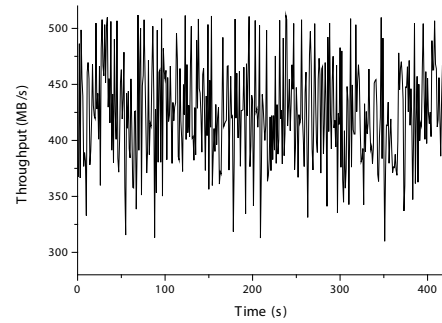


Figure 5. SSD erases times of traditional and optimized hybrid system.



a) Performance of Un-optimized System



b) Performance of Optimized System

Figure 6. System performance comparison.

5 SUMMARY

In this paper, a novel layered model and an optimized cache algorithm are proposed to improve SSD lifespan in layered SSD/HDD hybrid systems. The optimized designs significantly reduced erase times on SSD, and slightly influenced original performance. In the future, the model and algorithm will be further optimized to decrease the performance jitter influence.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61272511& 61370063), 863 Program (No. 2013AA01A212), students' Innovative Plan of BIT (No. 201410007045), and Shanghai Aerospace Science and Technology Fund(SAST201341).

REFERENCES

- [1] JinsunSuk,Jaechun No, "Performance Analysis of NAND Flash-based SSD for designing a Hybrid Filesystem", HPC, 2009.
- [2] Jung-WookPark,Seung-Ho Park,Charles, "A hybrid flash translation layer design for SLC-MLC flash memory based multibank solid state disk", Microprocessors and microsystems, 2011.1.
- [3] SangWonLee,BongkiMoon,Chanik Park,"A case for flash memory ssd in enterprise database applications",SIGMOD, 2008.
- [4] Xiao Li, Yuan Tan, Zhizhuo Sun, "Semi-RAID: A Reliable Energy-Aware RAID Data Layout for Sequential Data Access" MSST, 2011.
- [5] No J., "NAND flash memory-based hybrid file system for high I/O performance", Journal of Parallel and Distributed Computing, 2012.12.
- [6] Ruixuan Li, Chengzhou Xiao, "An Efficient SSD-based Hybrid Storage Architecture for Large-Scale Search Engines", International Conference on Parallel Processing, 2014.
- [7] Sun, Feng-Bin Parkhomovsky, "Physics-based life distribution and reliability modeling of SSD", ARMS, 2013.
- [8] Zhizhuo Sun, Yuan Tan, Yuanzhang Li, "An energy-efficient storage for video surveillance.", Multimedia Tools and Applications, 2014.1.
- [9] GuanYingWu,XuBin He, "An Adaptive Write Buffer Management Scheme for Flash-Based SSDs", ACM Transactions on Storage, 2012.1.
- [10] Boukhobza, Jalil, "CACH-FTL: A Cache-Aware Configurable Hybrid Flash Translation Layer", Euromicro International Conference on Parallel, 2011.
- [11] Puyuan Yang, "Hybrid Storage with Disk Based Write Cache", DASFAA , 2011.1.
- [12] Liu Yang, "RAF: A Random Access First Cache Management to Improve SSD-Based Disk Cache", IEEE International Conference on Networking, 2010.1.
- [13] SanamShahlaRizvi, "Flash SSD vs HDD: High Performance Oriented Modern Embedded and Multimedia Storage Systems", ICCET, 2010.1.
- [14] SoojunIm, "Flash-Aware RAID Techniques for Dependable and High-Performance Flash Memory SSD", IEEE Transactions on Computers, 2011.1.
- [15] Brent Welch , "Optimizing a hybrid SSD/HDD HPC storage system based on file size distributions", MSST, 2013.
- [16] WEI Wenguo, ZHAO Huimin , "An Improved Clock Adaptive Cache Replacement Algorithm", ActaScientiarumNaturaliumUniversitatisSunyatseni, 2015.51(6).
- [17] Youngjae Kim, "FlashSim: A Simulator for NAND Flash-Based Solid-State Drives", International Conference on Advances in System Simulation, 2009.1.

A collaborative filtering algorithm based on the improved similarity measure method and rating matrix pre-filling

D.W. Cai, Y. Liu & T. Cao

Institute of Intelligent Vision and Image Information, China Three Gorges University, Yichang, Hubei province, China

ABSTRACT: When the magnitudes of users and commodities grow rapidly, the rating matrix becomes extremely sparse. In this condition, algorithms based on the traditional similarity measure method have poor performance. In order to overcome data sparse and cold start issue, a new similarity measure method was proposed in this paper. The method combined the rating behavior of users and the attributes of items, and then filled the sparse matrix based on the similarity measure method proposed. Experimental results show that the algorithm is more accurate compared with other traditional algorithms.

1 INTRODUCTION

With the rapid development of information technology, E-commerce platform has brought great convenience to people's life. A recommendation system has become the core function of e-commerce platform and it is an important means to promote sales [1]. Nowadays, there are many recommendation systems among which collaborative filtering recommendation is the most successful. The basic idea of collaborative filtering algorithm is searching the nearest neighbors of users or items based on rating matrix and using the nearest neighbors to do recommend [2].

With the enlargement of e-commerce, the number of users and goods increase exponentially, but traditional collaborative filtering technology performance is getting worse. The main questions are: 1. Data sparse and precision issue. In many recommendation systems, user-item rating matrix is greatly sparse, which makes it hard to find the similar neighbors according to the traditional calculation method. 2. Cold start issue. When a new user is added, since the absence of evaluation over any item, you cannot get their interest preferences and know its nearest neighbor for further recommendation [3]. Similarly, when a new item is added, because there is no evaluation made by its user, so it cannot do filter recommendation.

For enhancing recommendation quality and overcoming data sparse and cold start issue, some scholars also do the work related. Based on the analysis of traditional similarity measure method and rating matrix pre-filling algorithm, this paper proposes a new similarity measure method which combines the rating behavior of users and the attributes of the item, and fills the sparse matrix based on the similarity measure method proposed. Experiments show that the algorithm this paper proposed can partly improve the accuracy of recommendation.

2 THE COLLABORATIVE FILTERING ALGORITHM

2.1 *Traditional similarity measure methods*

Similarity measure is the most important step in collaborative filtering recommendation algorithm. Here we present three methods such as correlation-based similarity, cosine based similarity and adjusted-cosine similarity [4].

1) Cosine-based similarity

In this case, two users are thought of as two vectors in the n dimensional item space. The similarity between them is measured by computing the cosine of the angle between these two vectors. Formally, similarity between user i and j , denoted by $\text{sim}(x,y)$ is given by

$$\begin{aligned} \text{sim}(x, y) &= \cos(X, Y) = \frac{X \cdot Y}{X \times Y} \\ &= \frac{\sum_{c=1}^n R_{x,c} R_{y,c}}{\sqrt{\sum_{c=1}^n R_{x,c}^2} \sqrt{\sum_{c=1}^n R_{y,c}^2}} \end{aligned} \quad (1)$$

Here $R_{x,c}, R_{y,c}$ denotes user x and y rated on item c .

2) Correlation-based similarity

In this case, the similarity between two users x and y is measured by computing the Pearson-r correlation. To make the correlation computation accurate, we must first isolate the co-rated cases. Let the set of items denoted by I_{xy} which were rated by both x and y then the correlation similarity between user x and y denoted by $\text{sim}(x, y)$ is given by

$$\begin{aligned} \text{sim}(x, y) &= \\ &= \frac{\sum_{c \in I_{xy}} (R_{x,c} - \bar{R}_x)(R_{y,c} - \bar{R}_y)}{\sqrt{\sum_{c \in I_{xy}} (R_{x,c} - \bar{R}_x)^2} \sqrt{\sum_{c \in I_{xy}} (R_{y,c} - \bar{R}_y)^2}} \end{aligned} \quad (2)$$

Here \bar{R}_x, \bar{R}_y denotes the rating of user x and y on item c , \bar{R}_x, \bar{R}_y denotes the average rating of user x and user y rated.

3) Adjusted-cosine similarity

Computing similarity using basic cosine measure in item-based case has one important drawback-the differences in rating scale between different users are not taken into account. The adjusted cosine similarity offsets this drawback by subtracting the corresponding user average from each co-rated pair. Formally, the similarity between user x and y using this scheme is given by

$$\begin{aligned} \text{sim}(x, y) &= \\ &= \frac{\sum_{c \in I_{xy}} (R_{x,c} - \bar{R}_x)(R_{y,c} - \bar{R}_y)}{\sqrt{\sum_{c \in I_x} (R_{x,c} - \bar{R}_x)^2} \sqrt{\sum_{c \in I_y} (R_{y,c} - \bar{R}_y)^2}} \end{aligned} \quad (3)$$

Here \bar{R}_x, \bar{R}_y denotes the average rating of user x and user y rated.

2.2 The problems existing in the traditional methods

The traditional similarity measure method based on two user's common rating vector and the disadvantages of this kind of measure method are mainly sparse, extensibility and a cold start problem. In the

extremely sparse rating matrix, the number of items rated by two users is fewer which will make the traditional recommendation precision not high.

For improving recommendation quality and overcoming sparse and cold start issue, some scholars put forward the algorithm that fills the sparse rating matrix in advance. Deng A L, etc.^[5] has proposed A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction. Feng Z J, etc.^[6] has proposed an optimized collaborative filtering recommendation algorithm by using a new revised conditional probability expression. Peng S, etc.^[7] has proposed a collaborative filtering algorithm based on rating matrix pre-filling. YU Z, etc.^[8] has proposed a data filling algorithm based on Cloud model. ZHANG Yu-fang, etc.^[9] has proposed a Collaborative filtering algorithm based on two-step filling for alleviating data sparse. Sun J G, etc.^[10] has proposed a collaborative filtering recommendation algorithm based on item attribute and cloud model filling.

All these algorithms improved the recommendation accuracy effectively, but the similarity measure methods in these algorithms only considered the rating values, but not the relationship between the user's behavior of rating and the attributes of items. There is still room for improvement. According to the above analysis, this paper proposes a new similarity measure method which combines the routing behavior of users and the attributes of items, and then fills the sparse matrix based on the similarity measure method proposed.

3 THE IMPROVED SIMILARITY MEASURE METHOD

In the actual behavior of user ratings, the user's rating behavior can show the user's similarity of interest. The greater number of users who were both rated on item i and j , the more similar the items i and j was. In order to measure the similarity between the two items, this paper considers the two elements of user ratings: the number of users who both rated item i and j and the value is very close, the number of users who only rated on an item i or j . In this paper, $(U_i \cap U_j)$ denotes the set of users who both rated on item i and j and the two values differ by less than two. The formula definition is given by

$$\begin{aligned} (U_i \cap U_j) &= \\ &= \{u_m \mid (R_{m,i} \neq 0) \cap (R_{m,j} \neq 0) \cap \mid R_{m,i} - R_{m,j} \mid \leq 2\} \end{aligned} \quad (4)$$

$n = \mid(U_i \cap U_j)\mid$ denotes the numbers of users defined above. m denotes the number of users who rated on only i or j . The improved similarity measure method

formula which synthesized the rating behavior of users is given by

$$\text{sim}_k(i, j) = \frac{\min(n, r)}{r} \text{sim}(i, j) \times \left(1 - \frac{\alpha \times m}{m+n}\right) \quad (5)$$

Here r is a parameter that takes a value increasing with the number of the users both rated. a is the factor of influence which denotes the degree of impact by the number of users who only rated one item.

The similarity measure method of the traditional collaborative filtering algorithm is using the rating information table, and the attributes of items can also provide important auxiliary factors of similarity measure. The attribute information of items can play an important role in the similarity measure especially when rating information is insufficient to similarity measure. Item attributes information generally uses vectors to express as shown in table 1. Take movie recommendation as examples, In the matrix of item attributes, Attr_1 denotes Action, Attr_2 denotes adventure, Attr_3 denotes animation, Attr_4 denotes western, etc.

Table 1. The attributes of items.

| Item | Attr ₁ | Attr ₂ | Attr ₃ | ... | Attr _t |
|-------------------|-------------------|-------------------|-------------------|-----|-------------------|
| Item ₁ | 1 | 0 | 1 | ... | 1 |
| Item ₂ | 1 | 1 | 0 | ... | 1 |
| ... | ... | ... | ... | ... | ... |
| Item _n | 1 | 1 | 1 | 0 | 0 |

In this paper, C_i denotes the set of attributes of item i , C_j denotes the set of attributes of item j . $|C_i \cap C_j|$ denotes the number of common attributes between item i and j . $|C_i \cup C_j|$ denotes the number of attributes to which neither item i nor j belong. The similarity of item i and j can be measured as follows:

$$\text{sim}_c(i, j) = \frac{|C_i \cap C_j|}{|C_i \cup C_j|} \quad (6)$$

In order to combine the rating behavior of users and the attributes of items and improve the accuracy and stability of the similarity measure method, this paper imports two weighting factors φ_k and φ_c , their range is $[0, 1]$. φ_k is the dynamic influence degree about the similarity of rating behavior and φ_c is the dynamic influence degree about the similarity of item attributes. The final formula definition is given by

$$\text{sim}'(i, j) = \varphi_k \text{sim}_k(i, j) + \varphi_c \text{sim}_c(i, j)$$

$$\text{Here } \varphi_k = \frac{\text{sim}_k^2(i, j)}{\text{sim}_k^2(i, j) + \text{sim}_c^2(i, j)} \text{ and}$$

$$\varphi_c = \frac{\text{sim}_c^2(i, j)}{\text{sim}_k^2(i, j) + \text{sim}_c^2(i, j)}.$$

4 THE IMPROVED COLLABORATIVE FILTERING ALGORITHM

The improved collaborative filtering algorithm of this paper is pre-filling the sparse rating matrix based on the improved similarity measure method which combines the rating behavior of users and the attributes of items and then does the recommendation on the filled rating matrix.

The main process of the Algorithm is as follows:

Input: the target user U_x , user-item rating matrix, item attribute matrix.

Output: the top-N recommended set.

Method:

(1) Traverse the set of items.

(2) For each item i , compute the similarity between i and others, according to the formula (7) and create the set of similar items.

(3) Based on the similarity set of item i , fill the rating matrix that no user rated on item i according to the formula (8).

$$P_{\text{item}}(i, U_x) = \frac{\overline{R}_i + \sum_{j \in \text{Nei}(i)} \text{sim}'(i, j) \times (R_{x,j} - \overline{R}_j)}{\sum_{j \in \text{Nei}(i)} \text{sim}'(i, j)} \quad (8)$$

Here, $\text{Nei}(i)$ denotes the neighbors of item i , \overline{R}_i denotes the average rating of target item i , \overline{R}_j denotes the average rating of the neighbors of item i .

(4) Based on the pre-filled rating matrix, compute the similarity between target user U_x and other users according to the formula (3) and create the set of similar users.

(5) Based on the set of similar users, compute the prediction of non-rated items according to the formula (9) and then create the set of recommendation.

$$P_{\text{user}}(U_x, i) = \frac{\overline{R}_x + \sum_{k \in \text{Nei}(U_x)} \text{sim}(k, x) (R_{k,i} - \overline{R}_k)}{\sum_{k \in \text{Nei}(U_x)} \text{sim}(x, k)} \quad (9)$$

Here, $\text{Nei}(U_x)$ denotes the neighbors of target user U_x , $R_{k,i}$ denotes the value that user k rated on item i . \overline{R}_x and \overline{R}_k denotes the average rating of user x and k .

5 EXPERIMENT RESULTS AND ANALYSIS

5.1 Data set

The data set in the experiment is the famous movie rating data provided by the research group GroupLens which includes 943 users' 100,000 rating records on 1682 movies, every user gives their rates on 20 movies at least. The rating value is integer from 1 to 5, higher value means higher preference. The whole data set are divided in training and test, respectively, by 80% and 20%.

5.2 Evaluation criteria

MAE(mean absolute error) is used as evaluation criteria, and it is the common evaluation criteria for the recommendation algorithm.

$$MAE = \frac{\sum_{u \in T} |R_{u,j} - R'_{u,j}|}{|T|} \quad (10)$$

Here, $|T|$ denotes the number of rating items in the test set. $R_{u,j}$ denotes the rating user u rated on item j in the test set, $R'_{u,j}$ denotes the rating prediction based on an algorithm.

5.3 Experimental results and analysis

In this paper, the experiment takes respectively 5 to 50 nearest neighbors to predict rating. This paper compares tradition item-based collaborative filtering algorithm (IBCF)with the algorithm proposed in paper^[9](CFBTSF $\alpha = 60$, $\beta = 10$, $\Upsilon = 0.7$), the algorithm proposed in paper^[10](Attribute and Cloud Filling) and the algorithm(IMCF) proposed in this paper in the same data environment. The result of the experiment is shown in Figure 1.

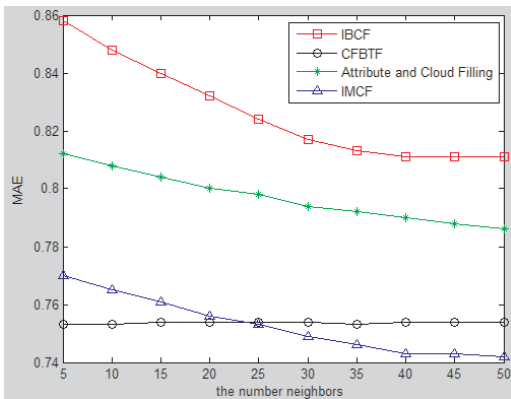


Figure 1. Comparison of prediction quality of different algorithms.

The experiments show that the MAE values basically tend to reduce with the increase of neighbor numbers except CFBTF. When the number of neighbors is from 5 to 25, the MAE value of CFBTF is lower than the IMCF, but when the number of neighbors is more than 25, the result is opposite. When the number of neighbors is nearly 40, the MAE value of IMCF tends to be stable. Compared with the MAE value of four algorithms, the obvious conclusion is that the algorithms proposed in this paper is better than other three algorithms.

6 CONCLUSIONS

This paper analyzed the advantages and disadvantage of the traditional similarity measurements in order to deal with the problem that user rating matrix is extremely sparse, and then proposed a new similarity measure method which combines the rating behavior of users and the attributes of items. The result of the experiment shows that the proposed algorithm is feasible and the recommendation result is more precise. In this paper, we mainly focus on the attributes of items and user behavior, but don't consider the other factors such as the attributes of users. In the future work, we will combine other factors to calculate the similarity and do recommendation.

REFERENCES

- [1] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering[J]. Internet Computing, IEEE, 2003, 7(1):76–80.
- [2] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems (TOIS), 2004, 22(1):5–53.
- [3] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[J]. Knowledge and Data Engineering, IEEE Transactions on, 2005, 17(6):734–749.
- [4] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the 10th international conference on World Wide Web. ACM, 2001:285–295.
- [5] Deng A L, Zhu Y Y, Shi B L. A collaborative filtering recommendation algorithm based on item rating prediction[J]. Journal of Software, 2003, 14(9):1621–1628.
- [6] Feng Z J, Xian T, Feng G U O J. An optimized collaborative filtering recommendation algorithm[J]. Journal of Computer Research and Development, 2004, 10:34.
- [7] Peng S, Zhou Z B, Wang G J. Collaborative Filtering Algorithm Based on Rating Matrix Pre-filling[J]. Computer Engineering, 2013, 39(1).

- [8] YU Z, QI Y. A Data Filling Algorithm Based on Cloud Model[J]. Computer Technology and Development, 2010, 12:010.
- [9] ZHANG Yu-fang, DAI Jin-long. Collaborative filtering algorithm based on two-step filling for alleviating data sparsity[J]. Application Research of Computers, 2013, 09:010.
- [10] Sun J G, Ai L R. Collaborative filtering recommendation algorithm based on item attribute and cloud model filing[J]. Journal of Computer Applications, 2012, 32(2).

Testing the radio proximity fuse with hardware-in-loop simulation

Y.J. Teng

Changchun University of Science and Technology, Changchun City, China

H.Y. Liu

Beihang University, Beijing, China

ABSTRACT: The performance of a radio proximity fuse is very crucial for ammunition attacking effect on targets. With hardware-in-loop simulation testing, a radio proximity fuse can effectively and reliably achieve the response when the fuse approaching target. In this paper, theory analysis is performed focusing on how to test the performance of a radio proximity fuse, and the transformation relation between the far-field RCS and the near-field RCS in a microwave dark room is derived. At the same time, the technical construction way of the radio proximity simulation testing system is presented.

1 INTRODUCTION

The radio proximity fuse can significantly improve the operational effectiveness of ammunition, so its dynamic performance should be tested. Due to the very high and still increasing target speed, such as aircraft and missile, the intersection condition of the fuse and the target changes drastically. Therefore, it is very difficult to control the intersection process[1–3]. Meanwhile, some kinds of high performance weapons are very expensive, it is costly in testing a fuse with real ammunition consumption. Hence, it is meaningful for testing fuse with simulation method [4]. But the first problem should be solved is how to simulate the dynamic scattering field from the target when the fuse approaching targets.

2 THE ARCHITECTURE OF THE SIMULATION-TESTING SYSTEM FOR RADIO PROXIMITY FUSE

Simulation the approaching process of the radio proximity fuse must be performed in microwave anechoic for avoiding noise. There are many kinds of targets and their electromagnetic scattering characters are various, accompanying the room of the microwave anechoic chamber is very limited, so the dynamic back scattering field should be generated with antenna and signal source. In the simulation system, the explosive part of the radio fuse is removed for safety. The output signal for electro-explosive is measured and recorded. The dynamic performance and the detection ability of the proximity fuse under test can be judged by analyzing these signals. The technical construction way of

the radio proximity fuse simulation testing system is shown as Figure 1.

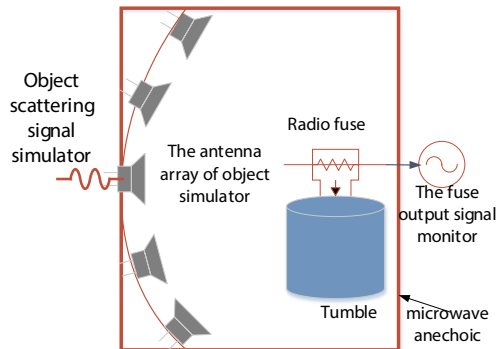


Figure 1. The technical construction way of the radio proximity simulation testing system.

3 THE SIMULATION METHOD OF THE SCATTERING FIELD RECEIVED BY RADIO FUSE ANTENNA

3.1 Achieving the target far field electromagnetic scattering characteristic

The electromagnetic scattering characteristic of the target is the basis of searching and detection for radio fuse. According to electromagnetic scattering characteristic, targets can be divided into two kinds, one is metal targets and the other is background targets. The metal targets are those with fixed shape, and made of metal, such as aircraft and missile, etc. The background targets usually have huge area, for example,

the ground and the sea, only a small portion of an area can be illuminated by an electromagnetic field from radio fuse. If an attacked target is near the ground or sea surface, the radio fuse will receive the scattering field from both the attacked target and the background. The confidence level of simulation testing result is directly decided by whether the simulated signal is as same as the target electromagnetic scattering characteristic. Because the electromagnetic scattering characteristics are various greatly for different targets, each kind of radio fuse has its specific targets. The electromagnetic scattering characteristics of these targets, and the spectrum and energy variation during projectile target crossing, will directly affect the fuse operating distance, the action volume shape, the Doppler signal bandwidth and the fuse circuit parameters. Therefore, in order to infield simulate the echo signal back to fuse with reality, the electromagnetic scattering characteristics of targets should be acquired in advance. There are several methods for acquisition target electromagnetic scattering characteristics, including strictly electromagnetic analyzing, numerical calculation, high frequency approximation calculation, measurement the target and its model [5].

3.2 Infield simulation the near-field scattering characteristics of targets

After Having achieved the far field scattering characteristics of a target, it should be researched how to simulate the scattering signal when the fuse signal incident to the target with an antenna or antenna array. It is not correct that taking the far field electromagnetic scattering characteristics as the near field electromagnetic scattering characteristics, in other words, taking the target far field RCS as near field RCS is wrong when the target and the radio fuse is very close during the intersection of the fuse and the target. In order to simulate the target near field scattering characteristics with antenna array, the far field RCS should be converted into near field RCS, and then use the nearest field RCS data control the antenna array.

3.3 The conversion method of far field to near field RCS

The conversion relationship between near and far field RCS can be derived strictly with scattering reaction integral equation [6-7]. In order to get useful results for practical engineering, also for easy, the multiple coupling between the antenna and the target and the target coupling with the environment is ignored. As shown in Figure 2, the voltage on antenna B induced by target scattering field can be expressed with the spheric angle function (SAF) of the receiving and transmitting antenna, and the target scattering SAF integrals.

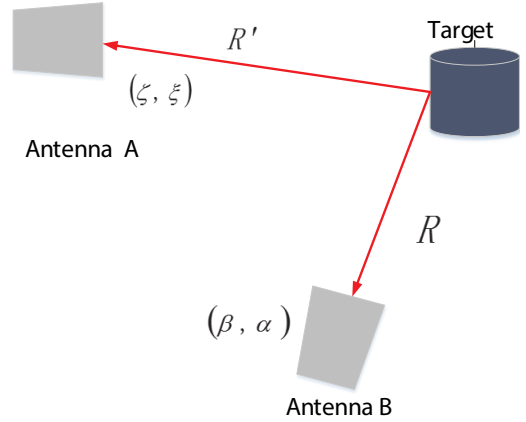


Figure 2. Target near-field double station scattering.

Set the SAF of the transmitting antenna to be

$$F_t = re^{jkr} E_{ff}(\theta, \varphi) \quad (1)$$

In (1), $E_{ff}(\theta, \varphi)$ is the far field electric field, its spherical wave factor e^{-jkr}/r just being counteracted by the factor re^{jkr} in (1). So the spheric angle function F_t does not contain a spherical wave factor, only is the function of spheric angle (θ, φ) . The SAF of the receiving antenna is

$$F_r = \frac{2}{i_0 \eta_0} F_t^* \quad (2)$$

Where, i_0 is the input unit current, $\eta_0 = 377\Omega$ is the free space wave impedance, the superscript '*' stands for adjoint. Set the SAF of the target scattering to be

$$F^s = re^{jkr} E_{ff}^s(\theta, \phi) \quad (3)$$

Where $E_{ff}^s(\theta, \phi)$ is far field electric field. Assuming that the receiving antenna can receive two orthogonal linear polarization signal, then the voltage to be measured is

$$V(R | R') = V_\theta(R | R')\hat{\theta} + V_\varphi(R | R')\hat{\varphi} \quad (4)$$

Where R and R' are the position vector directing to receiving antenna and transmitting antenna, respectively. V_θ stands for vertical polarization signal channel, V_φ stands for horizontal polarization signal channel. The SAF of two receiving channels can be expressed with dyadic vector, that is

$$\underline{\underline{F_r^b}} = \begin{bmatrix} \hat{\theta}\hat{\theta}F_{r\theta\theta}^b & \hat{\theta}\hat{\varphi}F_{r\theta\varphi}^b \\ \hat{\varphi}\hat{\theta}F_{r\varphi\theta}^b & \hat{\varphi}\hat{\varphi}F_{r\varphi\varphi}^b \end{bmatrix} \quad (5)$$

Where, $F_{r\theta\theta}^b$ is SAF when θ direction polarization transmitting and vertical receiving, the rest of the meaning of each component and so on.

The object scattering plane wave dyadic function can be expressed by spherical polarization components, that is

$$\overline{\overline{S}}(\theta, \varphi, \theta', \varphi') = \begin{bmatrix} \hat{\theta} \hat{\theta} S_{\theta\theta} & \hat{\theta} \hat{\varphi} S_{\theta\varphi} \\ \hat{\varphi} \hat{\theta} S_{\varphi\theta} & \hat{\varphi} \hat{\varphi} S_{\varphi\varphi} \end{bmatrix} \quad (6)$$

If the antenna aperture $D^a > 20\lambda$ and the distance between the object and antenna is greater than πD^a , the receiving antenna inductive voltage is

$$V(R|R') = \iint_{\Omega} \overline{\overline{F}}_r^b(\theta - \beta, \varphi - \alpha) \bullet F^s(\theta, \varphi | R', \zeta, \xi) \bullet \exp[-jk(\theta, \varphi) \bullet R] d\Omega \quad (7)$$

Where, $\overline{\overline{F}}_r^b(\theta - \beta, \varphi - \alpha)$ is the rotation dyadic vector of the receiving antenna, $F^s(\theta, \varphi | R', \zeta, \xi)$ is the scattering SAF when the transmitting antenna at R' and direct along (ζ, ξ) . The wave vector can be written as

$$k(\theta, \varphi) = k_0 [\sin \theta \cos \varphi \hat{x} + \sin \theta \sin \varphi \hat{y} + \cos \theta \hat{z}], \quad \text{and} \\ d\Omega = \sin \theta d\theta d\varphi.$$

The object scattering vector F^s can be calculated by integration of SAF, that is

$$F^s(\theta, \varphi | R', \zeta, \xi) = \frac{1}{\sqrt{4\pi}} \int_{\Omega'} \overline{\overline{S}}(\theta, \varphi | \theta', \varphi') \bullet \quad (8)$$

$$F_t^\alpha(\theta' - \zeta, \varphi' - \xi) \bullet \exp[+jk(\theta', \varphi') \bullet R] d\Omega'$$

Where, $F_t^\alpha(\theta' - \zeta, \varphi' - \xi)$ is the rotation SAF vector of the transmitting antenna. The wave vector can be written as

$$k(\theta', \varphi') = k_0 [\sin \theta' \cos \varphi' \hat{x} + \sin \theta' \sin \varphi' \hat{y} + \cos \theta' \hat{z}], \\ \text{and} \quad d\Omega' = \sin \theta' d\theta' d\varphi'$$

The angle (β, α) and (ζ, ξ) stand for the pitching angle and azimuth angle of the antenna A and antenna B, respectively.

Take (8) into (7), the receiving voltage (near field scattering reaction) equation

$$V(R|R') = \frac{1}{\sqrt{4\pi}} \iint_{\Omega} \iint_{\Omega'} \overline{\overline{F}}_r^b(\theta - \beta, \varphi - \alpha) \bullet \overline{\overline{S}}(\theta, \varphi | \theta', \varphi') \bullet F_t^\alpha(\theta' - \zeta, \varphi' - \xi) \bullet \exp[-jk(\theta, \varphi) \bullet R] \bullet \exp[+jk(\theta', \varphi') \bullet R'] d\Omega' d\Omega \quad (9)$$

The equation (9) associates the target near field scattering response $V(R|R')$ with the far field scattering response $\overline{\overline{S}}(\theta, \varphi | \theta', \varphi')$. Equation (9) is a convolution calculation relation. Therefore, when the distance between the fuse and the target doesn't satisfy the far field condition, the signal backing to receiving antenna reflected by the target should be calculated according to equation (9), instead of according to the target far field RCS. The analysis above is based on double station RCS. For the fuse, the receiving antenna and transmitting antenna are the same antenna, only set the pitching angle equal to the azimuth angle, equation (9) is still valid.

4 CONCLUSION

This paper presents a method for testing the performance of a radio proximity fuse with hardware-in-loop infield simulation. The method is based on the basic principle of electromagnetic scattering and good for sparing, fast testing radio fuse. But in this paper the method only can be proved feasible according to electromagnetic theory, therefore, further simulation and testing should be engaged for comparing and verifying.

REFERENCES

- [1] GJB49.35-41. The conventional weapon test methods, aviation ammunition fuse. Defense science, technology and industry committee, 1990
- [2] GJB573A-98. Fuse environment and performance test methods. Defense science, technology and industry committee, 1991
- [3] MIL-HDBK-757, 1994
- [4] Wang ShiYu. Modern fuse design problems and solutions. Nanjing China: Nanjing university of science and technology press, 2000
- [5] Huang PeiKang. The signal characteristics of radar target. Beijing China: Aerospace press
- [6] Fang DaGang. The spectrum domain method in electromagnetic theory. HeFei China: Anhui education Press, 1995
- [7] COWN B J. Near-field scattering measurements for determining complex target RCS. IEEE-AP 1989, 37(5)

An improved moving object tracking algorithm based on Camshift for dynamic background

G.W. Yuan, H. Zhou & D. Xu

College of Information Science and Engineering, Yunnan University, Kunming, China

Y. Zhao

Department of Animation, Yunnan Normal University, Kunming, China

Q.S. Tao

Department of Science and Technology, Yunnan University, Kunming, China

ABSTRACT: Tracking moving objects from video frames is one of the key techniques in computer vision. An improved Camshift algorithm for dynamic background is presented in this paper. The algorithm uses Harris corner matching to estimate background movement parameters, adopts two-frame difference to detect moving objects, combines merged hue histogram with edge direction histogram to describe moving objects and employs Camshift to track moving objects. The experiments show that the improved algorithm can achieve better result in the accuracy and robustness of detecting and tracking moving objects for dynamic background.

1 INTRODUCTION

Moving object tracking is an important step in computer vision. A moving object tracking method detects moving objects from video frames, matches the objects in the adjacent frames and computes motion parameters [Liu, Z. et al. 2012]. In the recent years, many methods have been presented, among which particle filter [Katja, N. et al. 2003, Rui, T. et al. 2013], Camshift [Dorin, C. et al. 2003, Zhang, H.Z. et al. 2006] and Kalman filter [Ristic, B. et al. 2004, Moussakhani, B. et al. 2014] are the most important methods.

The algorithm based on Camshift has been widely adopted for fixed camera scenes. But Camshift still has some disadvantages in complex applications. On the one hand, Camshift has a worse robustness when moving objects' color is similar to the background color, moving objects are occluded or moving objects' scale is variable. On the other hand, it could not get high accuracy in dynamic background.

An improved moving object tracking algorithm based on Camshift for dynamic background is presented in this paper. The algorithm can overcome the two demerits of the traditional Camshift, and it achieves better effect in the experiments.

2 MOVING OBJECTS DETECTION FOR DYNAMIC BACKGROUND

Moving object detection is the foundation of tracking. According to the motion of the camera or not,

it is classified into two kinds: fixed background and dynamic background. According to the principle of the algorithm, it is classified into three kinds: frame difference, background subtraction and optical flow. It is more difficult to detect moving objects from dynamic backgrounds than from fixed background. Optical flow can relatively detect moving objects in dynamic background, but it has high algorithm complexity. The traditional algorithms based on frame difference and background subtraction cannot get an accurate moving objects' region with dynamic background. Frame difference is very simple, and it can achieve good effect as long as the background is corrected. Harris corner matching is used to estimate background movement parameters, and this method is described as follows.

2.1 To estimate background movement parameters

Assuming that the background's motion includes only translation and rotation, the movement can be estimated using the following affine coordinate transformation model with 6 parameters [Zhang, L.Y. & Zhang G.L. 2005].

$$\begin{cases} x' = a_1x + b_1y + c_1 \\ y' = a_2x + b_2y + c_2 \end{cases} \quad (1)$$

In the formula 1, (x', y') is the corresponding pixel's coordinate in the current frame when (x, y)

is a pixel coordinate in the previous frame. The 6 parameters $a_1, a_2, b_1, b_2, c_1, c_2$ can be calculated by substituting three pairs of corresponding pixels' coordinate into the formula 1 [Yan, X.L. et al. 2008].

Now, the key issue is how to select the three pairs of corresponding pixels. Because Harris corner is a kind of representative feature point, it was selected as the matching points. The three pairs of Harris corners are selected as follows.

- Each frame is divided averagely into 16 blocks;
- Harris corners are extracted in each frame, and the corners between adjacent frames are matched;
- In order to ensure the matched corners out of the moving objects' region and be not near to each other, 3 non-adjacent blocks, which contain Harris Corners and do not contain moving objects, are selected randomly;
- One point is chosen randomly from the each block selected above Step c. Then the 3 pairs of corresponding pixel coordinates are substituted into the formula 1 to estimate background movement parameters.

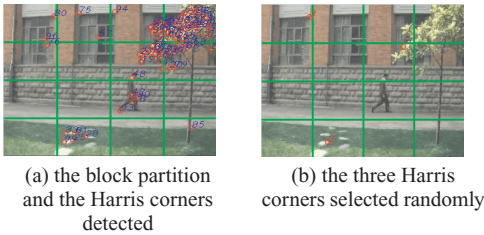


Figure 1. Select Harris corners to estimate background movement parameters.

2.2 To correct the current frame's background according to the background movement parameters

The movement distance of background can be estimated by the formula 2 according to the background movement parameters from the above.

$$\begin{cases} \mu = x' - x \\ \nu = y' - y \end{cases} \quad (2)$$

According to the movement distance of background (μ, ν) , each pixel in the current frame can be calculated back to the previous frame's position. If there are no corresponding pixels in the corrected frame, the pixels' gray-scale value is filled with zero.

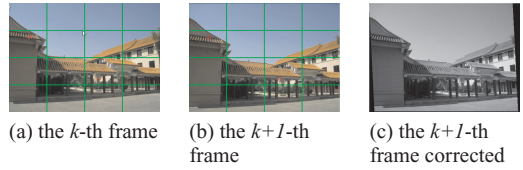


Figure 2. Correct background according to the background movement parameters.

2.3 To detect moving objects using frame difference

After correcting background, moving objects can be detected using frame difference. Figure 3a is the result without background correction, and then Figure 3b is the result after background correction. It is shown that the moving object detection with background correction is more accurate than one without background correction.

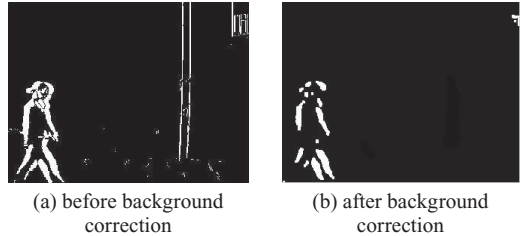


Figure 3. Result of moving objects detection.

3 AN IMPROVED MOVING OBJECTS TRACKING ALGORITHM BASED ON CAMSHIFT FOR DYNAMIC BACKGROUND

3.1 Traditional Camshift algorithm

Camshift is used widely in tracking moving objects, developed by Mean Shift algorithm [Dorin, C. et al. 2003, Zhang, H.Z. et al. 2006]. It is based on color histograms. The Camshift algorithm can be described by three parts: back projection, Mean Shift algorithm and Camshift algorithm. The detailed steps are as follows.

- 1 Set the whole image as searching area;
- 2 Initialize the tracking window's position and size;
- 3 Calculate the color's probability distribution of tracking window;
- 4 Run the Meanshift to obtain the new position and size of tracking window;
- 5 Set current tracking window's position and size with the result in Step 4, jump to Step 4.

3.2 Feature selection for moving object tracking based on Camshift

The traditional Camshift algorithm uses color histogram to describe the matching template. When the background's color is similar to the moving objects or the moving objects are obstructed, the tracked objects are easy to lose using color histogram. At the same time, the traditional color histogram is a global color histogram with complex calculation, and it is not conducive to real-time tracking. In addition, an edge direction histogram is robust to the occlusion and the similarity between background and moving objects. Therefore, our improved algorithm uses a combination of merged hue histogram and edge direction histogram to describe moving objects in order to track moving objects accurately and robustly in a complex environment. The multiple feature fusion method can be described as follows.

3.2.1 To calculate the edge direction histogram

First, the Sobel operators are used to filter the grayscale image I to generate a horizontal edge image G_x and a vertical edge image G_y respectively. Then the edge magnitude $Mag(i, j)$ and the edge direction $Dir(i, j)$ can be calculated by the formula 3 and 4.

$$Mag(i, j) = \sqrt{G_x(i, j)^2 + G_y(i, j)^2} \quad (3)$$

$$Dir(i, j) = \arctan \left| \frac{G_y(i, j)}{G_x(i, j)} \right| \quad (4)$$

Second, $Dir(i, j)$ is quantized to N types.

$$\theta(i, j) = k,$$

$$\text{if } Dir(i, j) \in \left[-\frac{(N/2+1)\pi}{N} + \frac{\pi}{N}k, -\frac{\pi}{2} + \frac{\pi}{N}k \right) \quad (5)$$

where $Dir(i, j) \in \left[-\frac{\pi}{2}, \frac{\pi}{2} \right]$ and $k = 1, \dots, N$.

Third, edge magnitude histogram of each type direction $\theta(i, j)$ is summed:

$$EdgeOrient(k) = \sum_{i,j} Mag(i, j) \delta[\theta(i, j) - k],$$

$$\text{if } \delta[x - k] = \begin{cases} 1, & x = k \\ 0, & x \neq k \end{cases} \quad (6)$$

where $k = 1, \dots, N$.

Fourth, the probability of $EdgeOrient(k)$ is calculated by formula 7:

$$EdgeHist(k) = \frac{EdgeOrient(k)}{\sum_{r=1}^N EdgeOrient(r)} \quad (7)$$

3.2.2 To calculate the merged hue histogram

First, the moving object area is transformed from RGB color space to HSV color space, and the hue values are obtained and normalized to 0.0~360.0 by formula 8 [Gonzalez, R.C. & Woods, R.E. 2002]:

$$Hue = \begin{cases} (G - B) \times 60 / HW, & \text{if } \max(R, G, B) = R \\ 120 + (B - R) \times 60 / HW, & \text{if } \max(R, G, B) = G \\ 240 + (R - G) \times 60 / HW, & \text{if } \max(R, G, B) = B \end{cases}$$

$$HW = \max(R, G, B) - \min(R, G, B) \quad (8)$$

Second, the hue is merged to M types by formula 9:

$$HueDiv(i, j) = l, \quad \text{if } Hue(i, j) \in \left[\frac{360}{M}(l-1), \frac{360}{M}l \right) \quad (9)$$

where $l = 1, \dots, M$.

Third, the probability of $HueDiv(i, j)$ is calculated by formula 10:

$$HueHist(l) = \frac{n_l}{n_{total}} \quad (10)$$

where n_{total} is the number of the moving object pixels and n_l is the number of the l -th hue type pixels.

3.2.3 To compare between the feature vectors

The moving objects are extracted using the above algorithm of moving object detection for dynamic background. According to the characteristics of edge direction and merged hue defined above, the merged hue histogram and edge direction histogram are calculated as the feature template of moving objects. The proximity measure $D(p, q)$ between the template p and q can be estimated by the formula 11.

$$D(p, q) = (1 - \lambda)D_{Edge}(p, q) + \lambda D_{Hue}(p, q) \quad (11)$$

where $\lambda \in [0.0, 1.0]$, and $D_{Edge}(p, q)$ and $D_{Hue}(p, q)$ are defined using the histogram intersection by the formula 12 and 13.

$$D_{Edge}(p, q) = \sum_{k=1}^M \min(EdgeHist_p(k), EdgeHist_q(k)). \quad (12)$$

$$D_{Hue}(p, q) = \sum_{l=1}^M \min(HueHist_p(l), HueHist_q(l)). \quad (13)$$

In the formula 11, $1-\lambda$ is the proportion coefficient of edge orientation feature and λ is that of the merged hue feature. The coefficient λ can be set according to the specific environment. The coefficient λ will be increased when the objects' color is not similar to the background. λ will be decreased when the moving objects are obstructed or the scene's illumination is not enough. The mechanism ensures the robustness of the tracking algorithm for various environments.

3.3 An algorithm flowchart of moving object tracking based on Camshift for dynamic background

The algorithm flowchart of moving object detection for dynamic background is shown in Figure 4, and the flowchart of moving object tracking based on multiple features fusion is shown in Figure 5.

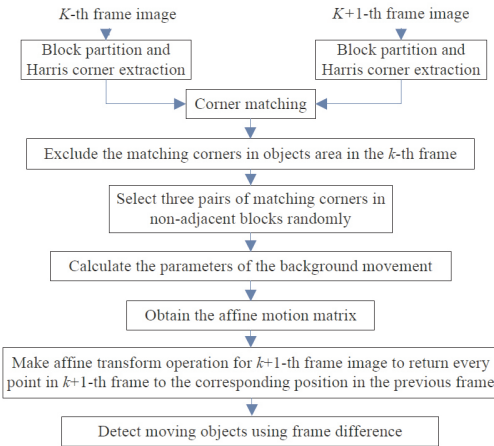


Figure 4. Algorithm flowchart of moving objects detection for dynamic background.

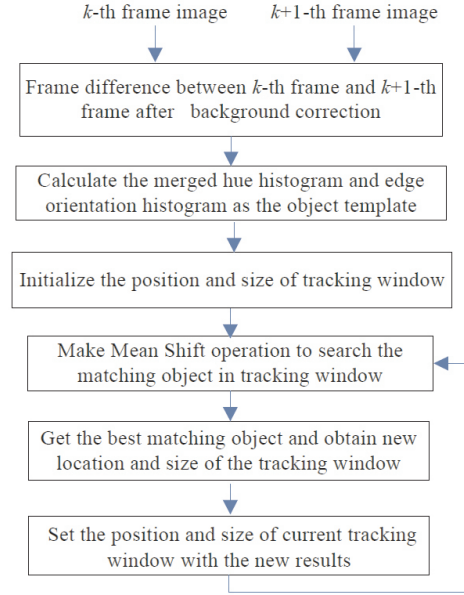


Figure 5. Algorithm flowchart of moving objects tracking based on multiple features fusion.

4 EXPERIMENTS AND ANALYSIS

We used a personal computer with an Intel Core i5-3470 processor and a 4GB memory in the experiments, in which Visual C++ and OpenCV are installed. The improved algorithm is applied to some outdoor videos including dynamic background.

4.1 Compare our method using background correction with a traditional frame difference for moving object detection

The result in Figure 6(a) is gotten by direct frame difference, and Figure 6(b) is gotten by our method with the background. The moving objects' contour detected in Figure 6(a) is bigger than the moving objects because of the absence of background correction. In Figure 6(b), the moving objects' contour is more accurate.

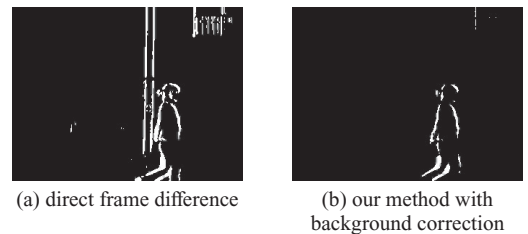


Figure 6. Comparison results of traditional frame difference and our method for dynamic background.

4.2 *Compare our method using background correction with the algorithm in the reference [Yan, X.L. et al. 2008]*

Figure 7a is the result of the algorithm in the reference 9, and Figure 7b is our improved algorithm's result. Our improved algorithm eliminates the error of estimating background motion parameters by avoiding the condition in which the matching Harris corners are selected in the moving objects. Therefore, the background correction using our improved method is more accurate, and the moving objects' contour is more accurate than the algorithm in the reference [Yan, X.L. et al. 2008].

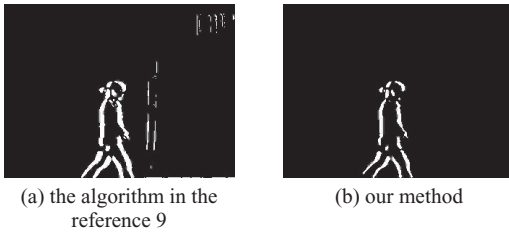


Figure 7. Comparison results of our method and the algorithm in the reference [Yan, X.L. et al. 2008].

4.3 *Compare traditional Camshift with our improved Camshift using multiple features fusion for dynamic background*

Figure 8a is gotten by the traditional Camshift in dynamic background, which uses color characteristics. Figure 8b is gotten by our improved Camshift algorithm combining the merged hue histogram with the edge histogram. In our experiment, the parameter N was assigned to 16 in the formula 5, 6, 7 and 12, and the parameter M was assigned to 24 in the

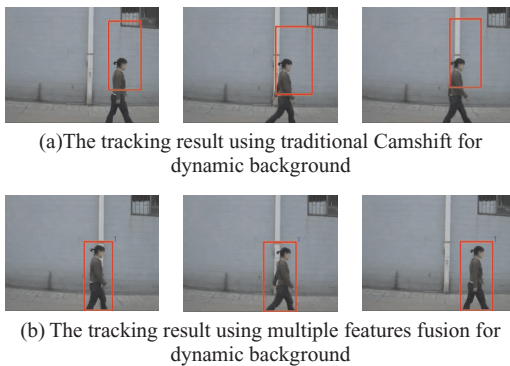


Figure 8. Comparison results of traditional Camshift and our improved Camshift using multiple features fusion for dynamic background.

formula 9 and 13. In the formula 11, and the fusion parameter λ was assigned to 0.4.

From Figure 8a, although traditional Camshift can track moving objects in dynamic background, there are some tracking errors. It can be seen in Figure 8b that our improved method can accurately track moving objects.

4.4 *Continuous tracking results using our improved Camshift in dynamic background*

Figure 9 is the continuous tracking results using our improved Camshift with multiple features fusion in dynamic background. It can be seen in Figure 9 that our improved algorithm can achieve good effect in the continuous change of background.

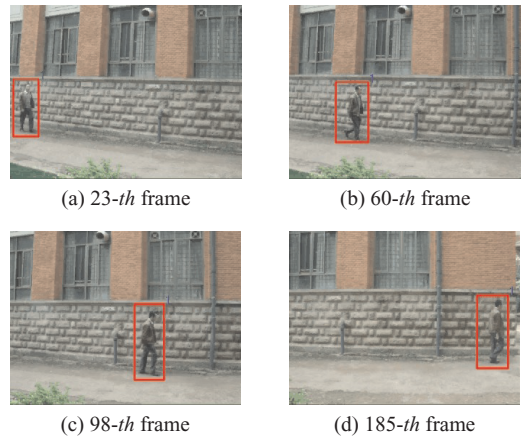


Figure 9. Continuous tracking results using our improved Camshift in dynamic background.

5 CONCLUSIONS

In this paper, an improved moving object tracking algorithm based on Camshift for dynamic background is presented in this paper. The algorithm corrects background by estimating motion parameters, employs a combination of merged hue histogram and edge direction histogram as matching feature and uses Camshift to track moving objects. The experiments show that the improved algorithm can achieve better result in detecting and tracking moving objects for dynamic background.

ACKNOWLEDGMENTS

This work is supported by the Natural Science Foundation of China (No.61163024), the Application and Foundation Project of Yunnan Province

(No. 2011FB019), the Science Research Foundation of Education Department of Yunnan Province (No.2012Z057) and the Young Teachers Training Program of Yunnan University.

REFERENCES

- [1] Dorin, C., Visvanathan, R. & Peter, M. 2003. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(5):564–577.
- [2] Gonzalez, R.C. & Woods, R.E. 2002. *Digital image processing*. Prentice Hall, 2nd Edition.
- [3] Katja, N., Esther, K.M. & Luc, V.G. 2003. Object Tracking with an Adaptive Color-Based Particle Filter. *Lecture Notes in Computer Science* 2449:353–360.
- [4] Liu, Z., Shen, H., Feng, G.Y. & Hua D.W. 2012. Tracking objects using shape context matching. *Neurocomputing* 83(15):47–55.
- [5] Moussakhani, B., Flam, J.T., Ramstad, T.A. & Balasingham, I. 2014. On change detection in a Kalman filter based tracking problem. *Signal Processing* 105:268–276.
- [6] Ristic, B., Arulampalam, S. & Gordon, N. 2004. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, London: Artech House.
- [7] Rui, T., Zhang Q., Zhou Y. & Xing, J.C. 2013. Object tracking using particle filter in the wavelet subspace. *Neurocomputing* 119(7):125–130.
- [8] Yan, X.L., Liang, B. & Zeng, G.H. 2008. Object tracking method based on block motion estimation. *Journal of Image and Graphics* 13(10):1869–1872.
- [9] Zhang, H.Z., Zhang, J.H., Yue, H. & Huang, S.L. 2006. Object tracking algorithm based on CamShift. *Computer Engineering and Design* 27(11):108–110.
- [10] Zhang, L.Y. & Zhang, G.L. 2005. Background motion model parameters estimation based on extended Kanade-Lucas tracker. *Computer Applications* 25(8):1946–1947.

Study on sustainability and life cycle assessment of RFID antenna

R.K. Kanth, P. Liljeberg & H. Tenhunen

University of Turku, Turku, Finland

H. Tenhunen & L.R. Zheng

Royal Institute of Technology, Kista, Stockholm, Sweden

Y. Amin

Telecommunication Engineering Department, University of Engineering and Technology Taxila, Pakistan

H. Kumar

Rama University, Geeta Nagar, Kanpur, Uttar Pradesh, India

ABSTRACT: This work gives a unique insight into evaluating sustainability, life cycle assessment and environmental emissions of the specified RFID tag antenna that validates the potential for ultra-low cost mass production for green electronics. We have designed a printed RFID antenna that employs commercially available paper substrates, and has capability to endure itself at varying environmental effects. The geometry and dimension of the antenna has been set as per requirement of read range of RFID tag. The environmental emissions profiles global warming, acidification potential, ozone layer depletion and human toxicity potentials are illustrated in this paper.

1 INTRODUCTION

Information and Communication Technology (ICT) plays an important role in the global economy. People around the world are voicing about the state of the physical environment, global warming and growing adverse environmental effects. Market analyses and research reports estimate that printed and stretchable electronics will become a huge, several hundred billion dollar business in the near future [1]. Motivated by the evidence of increasing demand of printed electronics in the near future, our aim is to develop an understanding of the life cycle assessment of printed electronic products and to analyze the environmental impacts in terms of global warming, acidification potential, ozone depletion potential and human toxicity potentials. The soaring growth of printed, organic and flexible electronics not only put the conventional devices aside but also make manufacturers have an opportunity to replace with cheaper and high performance alternatives. The proposed research method will utilize the general life cycle assessment approach which involves the collection and evaluation of quantitative data on the inputs and outputs of materials and energy associated with printed RFID antenna. The expected outcomes from this research work will contribute to characterize the environmental and health impact indexes such as global warming, ozone

depletion, acid rain and human toxicity potential, GHG computation and so on. These environmental analysis results obtained from printed RFID antenna will help in implementing life cycle thinking in policy or decision making nationally and internationally. The proposed research is highly relevant to state-of-art environmental research and this will be a continuation of my previous research works [2]–[5] on life cycle assessment of printed antennas.

2 THE GEOMETRY AND DIMENSION OF RFID ANTENNA

Fig.1 explicitly shows the dimension of the antenna which leads to determine the area required to accommodate antenna on the substrate that is 10cm × 4.3cm. Inkjet printing is considered for tracing the pattern on the substrate.

The proposed antenna is optimized in an organized approach for maximizing the read range while minimizing the ink consumption and effects of environmental harsh conditions on RFID tag. One of the critical observations made while field testing (in supply chain industry) of antennas having thinner traces used for matching network, is that the effect of even minor scratches on those tracks can cause loss of RFID tags during transportation. Therefore,

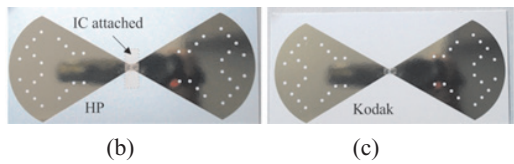
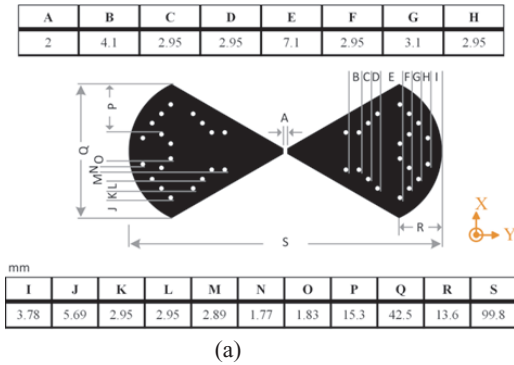


Figure 1. (a) The geometry & dimensions of RFID antenna; inkjet printed RFID tag on: (b) HP photo paper, (c) Kodak photo paper.

it led towards the realization of a structure which can eliminate such risk factors while exhibiting better performance. The main purpose of iterations and parametric enhancements is to maximize the feeding power to the load (RFID IC) and reduce the antenna area, in order to demonstrate highly efficient antenna structure on paper substrate for eco-friendly tags. The NXP UCODE G2XM (SOT1040AB2 (A1 strap), for instance the goal IC impedance at 915MHz is 13.3 - j122Ω) is preferred. The detailed explanation for structure of the antennas is depicted in papers [6]–[8].

3 PARAMETRIC MODEL

The environmental impacts model is created as shown in Figure 2 by using life cycle assessment software GaBi.

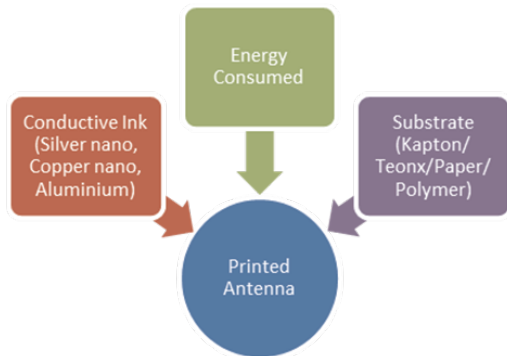


Figure 2. Parametric model required for sustainability analysis.

Inkjet printing is accomplished by using Fujifilm Diamatrix printer which is operated at 230V AC, 50Hz with maximum power consumed at any moment is 375Watt. The estimated time required for printing a pattern is approximately 15 minutes. If the printer operates at its maximum power rating, the energy required to print 1 million patterns is 337500 Mega Joules i.e. $(375 \times 15 \times 60 \times 10^6)$. The substrate under test is high gloss inkjet photo paper with standard weight of 230g/m² and thus for 1 million tags the required weight of paper substrate is 975.545kg $(230 \times 99.8 \times 42.5 \times 10^{-3})$. The capability of the printer to produce the antenna trace on paper substrate is 4000 structures with 1000ml (1kg) of conductive ink. Therefore, the estimated amount of conductive ink consumed for production of 1 million tags is 250kg. The mass of each component for manufacturing 1 million proposed RFID antennas is summarized in Table 1.

Table 1. Characterized/Evaluated parameters.

| S.N. | Material | Mass |
|------|---|----------|
| 1 | Substrate (paper/Kapton/ Teonex) | 976kg |
| 2 | Conductive Ink (Silver nano/ copper nano/aluminium) | 250kg |
| 3 | Energy Consumed | 337500MJ |

4 EMISSIONS PROFILES

Based on the modeling parameters, possible environmental impacts are explored in terms of global warming potential (KGCO2 equivalent), acidification potential (KGSO2 equivalent), ozone layer depletion potential and human toxicity potentials.

The Institute for Environment and Sustainability, European Commission Joint Research Centre, has published the research report [26] that recommends a preferred methodological approach which aims to calculate the environmental footprints of the products and organizations. For product environmental footprint, “ISO 14044: Environmental management- Life Cycle Assessment” is a recommended method. This method is analysis of the proposed printed antenna. Life cycle approach takes into consideration the spectrum of resources flows and environmental interventions associated with the product, services, or organization from a supply chain perspective, including all phases from raw material acquisitions through processing, distribution, use and end of life processes.

There are different methods such as TRACI (Tools for reduction and Assessment of Chemical and other Environmental Impacts) and CML (Centre for Environmental Studies, University of Leiden)

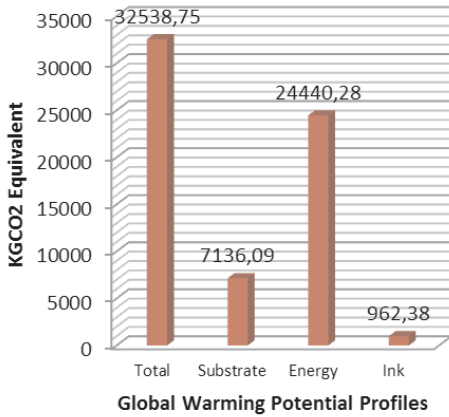


Figure 3. Global warming potential profiles.

that are employed to perform life cycle impact assessment. In this study, LCIA-CML methodology is considered to achieve the projected research analysis.

Figure 3 to Figure 6 show the potential environmental impacts in the manufacturing process of the proposed printed antenna. Four main environmental impacts are demonstrated and in each emission profile, there are four columns that correspond to the quantity of emissions caused by a specific component such as substrate, conductive ink or the associated energy consumptions in manufacturing 1 million printed antennas. KGCO₂ and KGSO₂ are carbon dioxide and sulphur dioxide equivalent emissions while ozone layer depletion and human toxicity potentials are computed by the release of

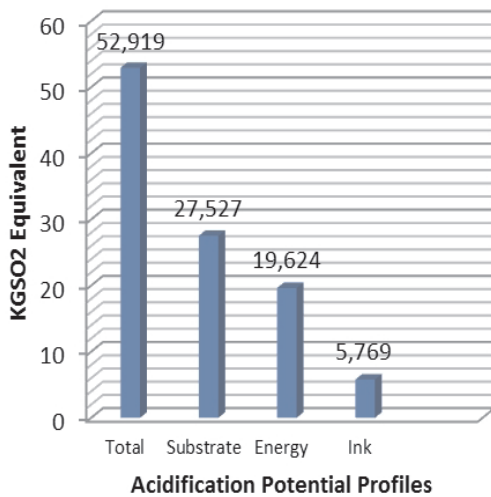


Figure 4. Acidification potential profiles.

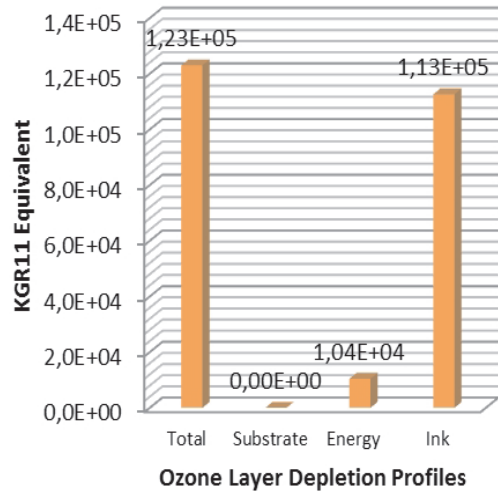


Figure 5. Ozone Layer Depletion potential profiles.

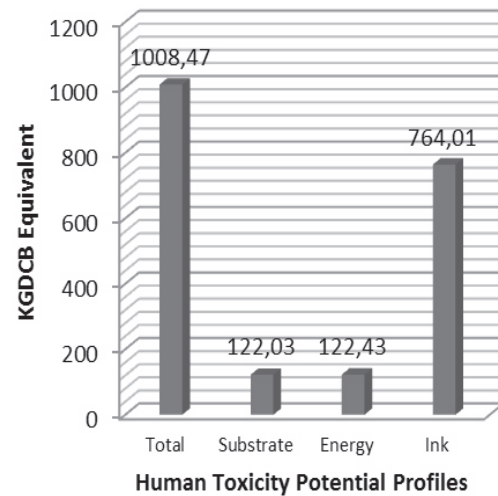


Figure 6. Human Toxicity potential profiles.

trichlorofluoromethane (CFC-11, also known as R11) and dichlorobenzene (DCB), respectively.

5 CONCLUSION

The sustainability and environmental impacts reckoning for the printed antennas can be applied as a manufacturing yield benchmark for future productions. In this work, we have analyzed environmental impacts in the form of global warming potentials, acidification potentials, ozone layer depletion, and human toxicity potentials in the production process of 1 million RFID antennas.

ACKNOWLEDGMENTS

This work was financially supported by Vinnova (The Swedish Governmental Agency for Innovation Systems) through the iPack Vinn Excellence centers, Royal Institute of Technology (KTH) program.

REFERENCES

- [1] Raghu Das, Dr Peter Harrop, *Organic & Printed Electronics Forecasts, Players & Opportunities 2007–2027*, ID TechEx
- [2] Rajeev Kumar Kanth, Qiansu Wan, Harish Kuamr, Pasi Liljeberg, Qiang Chen, Lirong Zheng, Hannu Tenhunen, "Evaluating Sustainability, Environment Assessment and Toxic Emissions in Life Cycle Stages of Printed Antenna", Elsevier, *Procedia Engineering*, Volume: 30, Issue:1, PP: 508–513, 2012.
- [3] Rajeev Kumar Kanth, Pasi Liljeberg, Hannu Tenhunen, Harish Kumar, Lirong Zheng, Nanotechnology, Life Cycle Assessment and Green ICT, Published in Proceedings of International Conference on Sustainability, Technology and Education (STE2013), Pages: 1–5, ISBN: 978-989-8704-00-9
- [4] Rajeev Kumar Kanth, Pasi Liljeberg, Yasar Amin, Qiang Chen, Lirong Zheng, Hannu Tenhunen, "Comparative End-of-Life Study of Polymer and Paper Based Radio Frequency Devices", *International Journal of Environmental Protection*, Volume: 2, Issue: (8), Pages: 23–27, Published year: 2012. Online ISSS: 2224–7777, Print ISSN:2226–6437
- [5] Rajeev Kumar Kanth, Qiansu Wan, Harish Kumar, Pasi Liljeberg, Lirong Zheng, Hannu Tenhunen, Life Cycle Assessment of Printed Antenna: Comparative Analysis and Environmental Impacts Evaluation. In: Matthew Eckelman (Ed.), *Conference Proceedings of IEEE International Symposium on Sustainable Systems and Technology (ISSST)*, 1, 1, IEEE, 2011.
- [6] Amin, Y.; Qiang Chen; Botao Shao; Hallstedt, J.; Tenhunen, H.; Li-Rong Zheng; *Design and analysis of efficient and compact antenna for paper based UHF RFID tags*. Antennas, Propagation and EM Theory, 2008, pp 62–65.
- [7] Botao Shao; Weerasekera, R.; Woldegiorgis, A.T.; Li-Rong Zheng; Ran Liu; Zapka, W.; *High frequency characterization and modelling of inkjet interconnects on flexible substrate for low-cost RFID applications*. Electronics System, 2008, pp 695–700.
- [8] Amin, Y.; Botao Shao; Hallstedt, J.; Prokkola, S.; Tenhunen, H.; Zheng, L.-R.; *Design and characterization of efficient flexible UHF RFID tag antennas*, Antennas and Propagation, 2009, pp 2784–2786.

Probabilistic data model based on privacy tipping points

M.V. Rallapalli

IBM, Financial Services Sector (FSS), Atlanta, USA

ABSTRACT: Many companies routinely alter privacy policies without taking any web user inputs into consideration. In 2012, Facebook decided to approve changes to its data use policy and statement of rights and responsibilities without any user input. This resulted in a huge backlash in the social media against Facebook policies by web users and privacy advocates. Web users and privacy advocates have taken matters into their own hands, posting enough comments on the note in the Social media, forcing Facebook to put them to a vote. What this shows is user community collectively can allow or disallow privacy changes made by the big social media companies. This paper addresses a probabilistic data model that captures the privacy thresholds to give a better understanding of acceptable privacy changes to a published privacy agreement. The computations are based on Random Walk theory formulae applied to privacy data sets collected in a real life survey.

KEYWORDS: Big data, privacy, framework, privacy negotiation, framework approach, tipping points.

1 INTRODUCTION

In the aftermath of the uproar over changes to Instagram’s privacy policy and terms of service earlier in December, 2012, the company did an about-face. In a blog post on the company’s site, Kevin Systrom, Instagram’s co-founder, said that where advertising was concerned, the company would revert to its previous terms of service, which have been in effect since October 2010 [1]. Users had been particularly concerned by a clause in Instagram’s policy introduced on December 17th that suggested Instagram would share users’ data — like their favorite places, bands, restaurants and hobbies — with Facebook and its advertisers to better target ads. Web users also took issue with an update to the company’s terms of service that suggested users’ photos could be used in advertisements, without compensation and even without users’ knowledge. The terms of that user agreement said, “*You agree that a business or other entity may pay us to display your user name, likeness, photos (along with any associated metadata) and/or actions you take, in connection with paid or sponsored content or promotions, without any compensation to you*”. Following a reaction that included customers defecting to other services, Instagram issued a statement that the new policy had been misinterpreted. “*It is our mistake that this language is confusing*,” and promised an updated agreement. This paper addresses one way to design a probabilistic data model to understand tipping points in privacy policies absorption (acceptance or rejection). Though not exact, related research is done by the entities as outlined in [table 1](#). This paper assumes

that the six research studies identified in [Table 1](#) in the area of web services privacy technologies is a sufficient number to validate the current state of privacy frameworks and in understanding the shortcomings, gaps and the focus areas of each of these research studies

Table 1. Research studies in the area of privacy technologies.

| # | Research Study | Affiliation |
|---|---|--|
| 1 | Standards for XML and Web Services Security | Martin Naedele, ABB Corporate Research [2] |
| 2 | Towards Standardized Web Services Privacy Technologies | University of Ontario Institute of Technology & University of Insubria at Como [3] |
| 3 | Privacy Policy Compliance for Web Services | National Research Council Canada [4] |
| 4 | Exploring Privacy Issues in Web Services Discovery Agencies | University of Ontario Institute of Technology & University of Insubria at Como [5] |
| 5 | A Framework for Building Privacy-Conscious Composite Web Services | Stony Brook University & University of Illinois at Chicago [6] |
| 6 | A Privacy Agreements Model for Web Services | University of Lyon & University New South Wales [7] |

2 PRIVACY TIPPING POINTS

A Tipping Point is a point of state and is defined as “the moment of critical mass, the threshold, the boiling point” [8]. The reverse is also true, if a few trend setters begin to reject a particular style or model, the product rejection tends to accelerate leading to a withdrawal from the market. Applying same logic to privacy policies, web user comments either favoring or opposing privacy policies may force organizations and their Socio-Technical Systems (STS) to either implement or withdraw their privacy policy changes. To be able to measure these privacy tipping points by using a probabilistic model is covered later in this paper.

In the instance of Facebook unilaterally changing its privacy policies, web users and privacy advocates have taken matters into their own hands, posting enough comments on the note in the Social media, forcing Facebook to put them to a vote. Since the social media postings has gained a substantial amount of comments (more than 19,000 from greater than 7,000 users), Facebook decided to open up a seven-day voting period, allowing users to make their voices heard. However, it will take 30 percent of Facebook’s roughly 1 billion-strong population to stop the changes. In this case, an arbitrary tipping point of 30% is being used for privacy policies approval. Twenty-four hours after Facebook opened the polls, more than 100,000 users have cast their votes 10 to 1 against Facebook’s proposed changes to its policies. That includes a proposal that would do away with Facebook users’ right to vote on future changes [9]. The negative publicity prompted Facebook to begin letting users vote on major changes to how it handles their personal information.

The Instagram and Facebook incidents reflect the fact that, the web user reaction to privacy policies can make or break the policy updates.

3 TIPPING POINTS INSTANTIATION

This section describes various parameters used in computations, use cases, research assumptions, and threshold values leading to the probabilistic data model. This research assumes privacy tipping points to take place at two threshold values:

‘a’ is an upper tipping point threshold, when the first time of hitting equals this value, P is acceptable to the web user group, and,

‘-b’ is a lower tipping point threshold, when the first time of hitting equals this value, P is NOT acceptable to the web user group.

Where,

P = Privacy policy,

N = Instance of the user (e.g. N= 1, means it’s the first user)...N is always be $\leq M$, and,

M = Subset number of users (for e.g. a million users).

When a P is presented to M users,

Some agree with the privacy terms (p), and,

Some disagree with the privacy terms (q).

Let’s take a random variable R, and apply the following values:

For users $i = 1 \rightarrow N$ (N indicates a large number)

(p) = $R_i = 1 \rightarrow$ Accept

(q) = $R_i = -1 \rightarrow$ Reject

Summation Series:

Summation $S_N = R_1 + R_2 \dots\dots R_N$ (N = running parameter 1 thru ∞)

$S_0 = 0$ (At N = 1, no disposition is available)

$S_{0+1} = R_1$

$S_N = S_{(N-1)} + R_N \rightarrow$ Summation at any given N equals previous summation + New disposition (a 2-barrier Random Walk)

Following assumptions are taken into consideration for calculations:

Users, once they make their choice, do not change their disposition and

Each user makes a choice independent of other users (simplified assumption).

3.1 Use case scenarios

Summation, S_N , can take one of the three possible values as shown below:

- 1 When $S_N = a$, Privacy terms are acceptable to the web user group
- 2 When $S_N = (-b)$, the web user groups rejects the privacy terms provided by the service provider
- 3 Privacy negotiation takes place when (S_N) is in between ‘a’ & ‘-b’:

Service provider’s goal is to push web users to higher threshold (a).

If the service provider does NOT budge on negotiation, then there is a risk of losing a potential group of clients (web users)

The following use cases to describe the above scenarios with six sample users and arbitrary threshold values.

Use Case 1: First hitting time of upper threshold tipping point ‘a’ (with 6 sample users)

First hitting time of tipping points (a or -b), the process of group voting stops

a = 4 (-b) = -2 (Arbitrarily selected tipping points)

$S_6 = R_1 + R_2 + R_3 + R_4 + R_5 + R_6$

$= (+1) + (+1) + (+1) + (-1) + (+1) + (+1) = 4$ (5 users accept, 1 reject)

$S_6 = 4$ (First hitting time of upper threshold Tipping point reached), terms are acceptable to the web user group

Use Case 2: First hitting time of lower threshold Tipping point '-b' (with 6 sample users)

$$S_6 = R1 + R2 + R3 + R4 + R5 + R6$$

$$= (+1) + (-1) + (+1) + (-1) + (-1) + (-1) = -2$$

(3 users accept, 4 reject)

$S_6 = -2$, terms are NOT acceptable to the web user group

Use Case 3: Summation is in between the thresholds (with 6 sample users)

$$S_6 = R1 + R2 + R3 + R4 + R5 + R6$$

$$= (+1) + (-1) + (+1) + (-1) + (+1) + (+1) = 2$$

(4 users accept, 2 reject)

S_N is between 'a' and '-b', the user group voting continues

3.2 Research computations

This section documents various random walk formulae used for model computations.

Computation of probability of absorption (a) using random walk:

- 1 Computation of probability of absorption (b) $p(a) = pa * (pb - qb)$

$$\frac{(p^{(a+b)} - q^{(a+b)})}{p - q}$$

$$= b / (a + b) \text{ (for } p = q)$$

Where a= upper threshold

b = lower threshold

$p(a)$ = probability of absorption (accept)

$q(-b)$ = probability of absorption (reject)

- 2 Computation of probability of absorption (b) $q(-b) = 1 - p(a)$

- 3 Summation series

Summation $SN = R1 + R2 \dots \dots RN$ (N = running parameter 1 thru ∞)

$SN = a$ (probability of first hitting time of upper threshold), implies user group 'accepts' the privacy policy

$S0 = 0$ (At N = 1, no disposition is available),

$S0 + 1 = R1$, and,

$SN = S(N-1) + RN \implies$ Summation at any given N equals previous summation + New disposition (a 2-barrier Random Walk, R= random variable)

$SN = (-b)$ (probability of first hitting time of lower threshold) implies user group 'rejects' the privacy policy, $-b < SN < a$, implies voting continues

- 4 $E(N)$ = average number of users that SP should negotiate with, to get privacy policy acceptance

$$E(N) = a / (1 - 2p) \text{ (for } p < q)$$

$$E(N) = a / (2p - 1) \text{ (for } p > q)$$

$$E(N) = a / (2p - 1)$$

Actual computations on a sample data are provided in Appendix C.

3.3 Model verification

To verify that data computations by using Random Walk, this research used 3 data sets collected independently using a formal but anonymous survey process. The methodology involves doing the computations first by using Random Walk formulae as is shown above and we compare it with the actual data to determine the deviation. As part of this research, a survey was done by using 3 datasets and computations are applied on the survey data to determine the comparison gaps and observations. Table 2 displays the survey conducted on three different independent sources, collecting three different datasets.

Table 2. Survey data to verify against Random Walk computations.

| Surveys at a glance!!!! | |
|-----------------------------|-------------------------|
| Total number of subjects | 1486 |
| Valid responses | 436 |
| Average time for completion | 4 days |
| Tools used | TOLOR.COM , SurveyGizmo |
| Survey method | Web anonymous |

Providing actual survey data is beyond the scope of this paper, but following are the summaries from each of the datasets:

Dataset 1:

When a privacy term on a Credit card value is altered, the upper threshold reached in four days as shown in Fig. X below:

Upper threshold (a) = 20

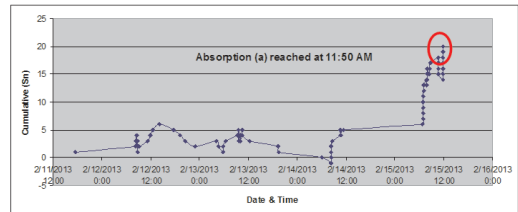


Figure 1. Probability of absorption (a) for CC as a privacy term.

Dataset 2:

In this data set, web users accepted the privacy terms after negotiation for a lower value privacy term for personal telephone information.

Upper threshold (a) = 18

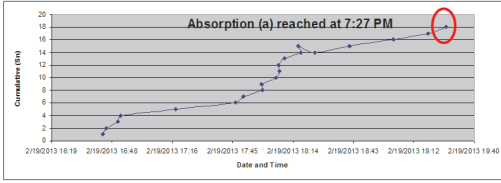


Figure 2. Probability of absorption (a) for telephone information.

4 CONCLUSIONS

The outcome of this research work is a model that will enable service providers to consider privacy as a vital aspect of every interaction with the web user. This would not only provide peace of mind on privacy protection, it will also gain the confidence of the web users, resulting in a market share gain. The major beneficiaries of this research will be the B2B and B2C e-commerce websites who are continuously faced with the task of gaining more customers and ensuring that their privacy data are protected as per the privacy agreement.

REFERENCES

- [1] Nicole Perloth and Jenna Wortham, “Instagram Does an About-Face”, 12/20/2012, <http://bits.blogs.nytimes.com/2012/12/20/instagram-does-about-face-reverts-to-previous-policy/?hpw>.
- [2] Martin Naedele, ABB Corporate Research, “Standard for XML and Web Services Security”, pages 96–98,

- April 2003WikiBon blog “A Comprehensive List of Big Data Statistics“, August 1, 2012, <http://wikibon.org/blog/big-data-statistics/>.
- [3] Barbara Carminati, (University of Insubria at Como), Elena Ferrari, (University of Insubria at Como), Patrick C.K. Hung, University of Ontario Institute of Technology, “Exploring Privacy Issues in Web Services Discovery Agencies”, IEEE Security and Privacy, September/October 2005, pp. 14–21; <http://doi.ieeecomputersociety.org/10.1109/MSP.2005.121>.
- [4] George Yee and Larry Korba, National Research Council Canada, “Privacy Policy Compliance for Web Services”, 2004 IEEE Int’l Conf. Web Service (ICWS 2004), 2004, pp 158-165; DOI bookmark: <http://doi.ieeecomputersociety.org/10.1109/ICWS.2004.1314735>.
- [5] Barbara Carminati, Elena Ferrari, Patrick C.K. Hung, “Exploring Privacy Issues in Web Services Discovery Agencies”, pages 14–21, September/October 2005.
- [6] Wei Xu, R. Sekar, I.V. Ramakrishna (Stony Brook University), V.N. Venkatakrishnan (University of Illinois at Chicago), “A Framework for Building Privacy-Conscious Composite Web Services”, 2006 IEEE Int’l Conf. Web Service (ICWS 2006), 2006, pp 655–662; DOI bookmark: <http://doi.ieeecomputersociety.org/10.1109/ICWS.2006.4>.
- [7] Salima Benbernou, Hassina Meziane, Mohand-Said Hacid, Yin Hua Li, “A Privacy Agreement Model for Web Services”, pages 196–203, 2007.
- [8] Malcolm Gladwell, “The Tipping Point: How Little Things Can Make a Big Difference (ISBN 978-0-316-34662-7)”, 2000 (page 12).
- [9] Jessica Guynn, “Facebook opens polls, users vote against privacy-policy changes“, Los Angeles Time, <http://www.latimes.com/business/technology/la-fi-tn-facebook-opens-polls-users-vote-against-privacy-policy-changes-20121204,0,3550668.story>, December 4th, 2012.

Research on hybrid parallel computing model and key technology

J.P. Dai, L.P. Zhu, B. Hu & A.H. Zhou

Information & Communication Department, China Electric Power Research Institute, Nanjing, China

ABSTRACT: This paper discussed MPI and OpenMP programming paradigm and different features. Then a hybrid computing model is designed, including architecture, hierarchical structure and implementation. The characteristic of hybrid computing model shows that it has better scalability and performance.

1 INTRODUCTION

High performance computer (HPC) can be divided into two categories according to the memory structure, i.e. shared memory structure and distributed memory structure. SMP (symmetric multi-processors) cluster system is a hybrid of the two kinds of memory structures. It is built by a multi-level system structure composed of multiple processors-SMP nodes and rapid connection network among each node. Currently, the computer with multi-level parallel architecture of SMP cluster, as shown in figure 1, is developing rapidly. It has become the mainstream of the current high performance computer [1].

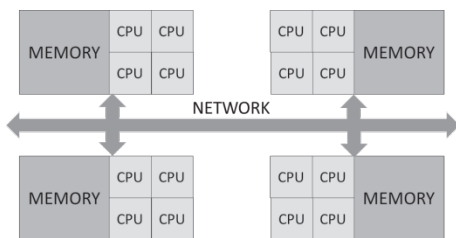


Figure 1. Multi-level parallel architecture of SMP cluster.

This architecture provides two levels of parallelism intra and inter nodes. Therefore, parallel computing must take account of the characteristics of the two kinds of system structures. In distributed memory systems, using the MPI message passing has been considered as the ideal model of program design. In shared memory systems, OpenMP is the standard of compiler-based parallelism. On the basis of an SMP system, there are several parallel computing modes, such as pure MPI mode using single message passing, Open MP mode running on the SMP node, and hybrid mode combined with MPI and OpenMP. Due

to the two level-memory architectures, the parallel mode, which uses multiple threads and message passing among nodes can better match the multi-level architecture of SMP cluster, so as to achieve better performance.

2 MPI AND OPENMP MODEL

2.1 MPI computing model

MPI (Message Passing Interface) [2] is an implementation standard on the message passing computing model developed by academia, government and industry association. It is the mainstream computing model of the current distributed memory system. It is not an independent programming language, but a library, providing Fortran and C/C++ language binding. MPI is suitable for both shared memory and distributed memory computing environment. Any program developed by MPI can run on SMP cluster directly.

During the development of a single program multiple data (SPMD) parallel model with MPI, each process can only read or write data in local memory. The access to remote data is accomplished by explicit message passing (library function calls) among the processes. This kind of explicit parallelism usually provides better performance and portability.

However, development with MPI has some disadvantages. It is difficult to decompose, develop and debug, requiring a great deal of changes to the code in most of the cases. Also, the communication overhead may be quite high. It often needs large code granularity to minimize the communication delay.

2.2 OpenMP computing model

OpenMP[3] is an industry standard for shared memory system programming. It is not an independent language, but an extension to basic language (such

as Fortran, C/C++). The goal is to provide a portable and extensible development interface for SMP systems. OpenMP specifies a series of compilation guidance, runtime library and environmental variables to explain the parallel mechanism of shared memory structure. Compilation guidance is the extension to programming language, to further support parallel region, job sharing, synchronous structure, making data sharing and privatization easier. Users can adjust the execution environment of parallel program by the runtime library and environmental variables. Also, OpenMP implements the thread-level parallelism. The inter thread communication is realized through reading and writing shared variables.

With the development of the emerging multicore systems and standardization, OpenMP has recently become more and more popular. OpenMP programming is relatively simple, making full use of the characteristics of the shared memory architecture. It avoids the overhead of message passing, and provides both fine-grained and coarse-grained parallel mechanism. Another feature of OpenMP is the simplicity of the code transformation. It is not necessary to make a big change to the codes when transforming serial programs to parallel programs. Simply adding some guidance statement is enough. However, the weakness of OpenMP is its poor scalability. OpenMP is based on a shared memory machine, so it is subject to the limitation of the available computing unit in the system. When the problem size increases, it is difficult to add new resources to accelerate the computation.

3 HYBRID COMPUTING MODEL

3.1 Hybrid model architecture

Given that MPI is the represented standard of message passing parallel computing model, it can easily support the distributed memory and shared memory structure. On the other hand, as an application programming interface designed for shared memory parallel programming environment, OpenMP currently is the industrial standard supporting shared memory parallel programming. In the SMP cluster, hybrid computing model has some successful application cases. But the study of hybrid computing model for multi-core PC cluster has just started. In the multi-core PC cluster, the combination of MPI and OpenMP technology can bring more advantages, and achieve better performance with smaller developing overhead. The architecture of hybrid model with MPI and OpenMP is shown as [figure 2](#).

According to the method of message passing between process in MPI, the hybrid model can be classified into two types as follows.

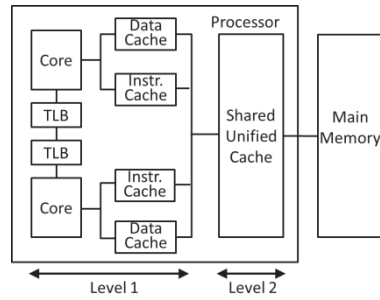


Figure 2. Architecture of MPI and OpenMP hybrid model.

- 1 Single-level hybrid model (hybrid master-only). MPI calls occur outside the multi-thread parallel region of applications. The inter-process communication is executed by the main thread. This kind of hybrid computing model is easy to be realized. Just add the loop statement `#pragma omp parallel` to the key computation part of program based on MPI model.
- 2 Multi-level hybrid model (hybrid multiple). The MPI calls may occur inside the multi-thread parallel region of applications. The inter-process communication can be completed by any one thread or several threads within the region of application programs. In this model, when some threads are executing communication, other non-communication thread can execute computation simultaneously. This mechanism makes the concurrent execution of communication and computation possible, and also optimizes solution to the inter-process communication congestion problem.

3.2 Hierarchical structure of hybrid mode

In order to make full use of SMP cluster architecture, the MPI and OpenMP hybrid computing model provides two-level parallelism inside and outside nodes. It can take advantages of the shared memory computing model and the message passing computing model, mapping cluster system structure of multi-processor computers perfectly.

The hierarchical structure of hybrid computing model is shown as [figure 3](#). The MPI level is located on the top, representing the parallelism outside nodes. The OpenMP level is located on the bottom, representing the parallelism inside nodes. At the very beginning, the problem is decomposed by MPI. Each MPI process is assigned to a SMP node. Inter process communication is conducted through message. Next, a set of threads are created with OpenMP guidance command in each MPI process, and assigned

to different processors of SMP node to be executed concurrently[4].

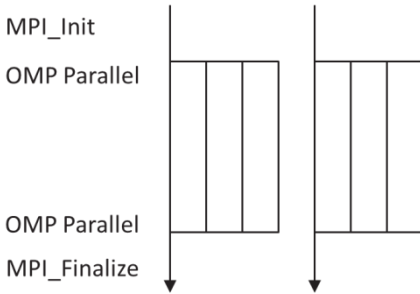


Figure 3. Hierarchical structure of MPI/OpenMP hybrid model.

Hybrid computing model makes full use of the advantages of both programming modes. Using MPI may solve the problem of coarse-grained communication among multiple processors. Meanwhile, the lightweight threads provided by OpenMP may deal with the interaction between processors within multi-processor computers well. The MPI and OpenMP hybrid model use, fast shared memory access instead of message passing within nodes, reducing the communication overhead. Furthermore, when the number of the MPI process increases, the parallel performance usually gets worse. But the hybrid model may overcome this problem. Because the number of MPI processes in hybrid model is equal to the number of nodes rather than the number of processors in MPI model. Thus, the number of processes may be dramatically reduced while the number of processors does not change. Therefore the hybrid model has the potential advantage of scalability.

3.3 MPI/OpenMP hybrid parallel model

In OpenMP, inter-thread parallelism may be realized through fine-grained or coarse-grained methods. When combined with MPI code, we can get two kinds of hybrid parallel model.

3.3.1 MPI and fine-grained OpenMP hybrid parallelism

Fine-grained parallelism of OpenMP refers to using OpenMP to perform computation on the loop part concurrently, also known as the loop level parallelism. Therefore, a fine-grained hybrid parallel programming method is the simplest way. It uses the multi-thread parallelism of OpenMP on shared memory structures within an MPI process. Its process runs as follows:

- a. There is only one MPI process on each node, and the process is initialized.

- b. The MPI process on each node can perform some local computation independently. The inter-node communication can also be carried out when necessary.
- c. The main part of computation within an MPI process (usually loop part) is performed by multi-thread parallelism of OpenMP.
- d. The MPI process may run local computation, communication and synchronization, when the computation with OpenMP has ended.
- e. The multi-thread computation with OpenMP and the local computation and communication with MPI may run in turn. After all the computation tasks finish, the processes of MPI end.

3.3.2 MPI and coarse-grained OpenMP hybrid parallelism

In the coarse-grained hybrid parallelism of OpenMP, the SPMD computing model is combined with MPI. Its process runs as follows:

- a. There is only one MPI process on each node, and the process is initialized.
- b. The MPI process on each node can perform some local computation independently. The inter-node communication can also be carried out when necessary.
- c. The multiple threads of the OpenMP start after the MPI initialization.
- d. The whole computation is performed through multiple threads. If the MPI communication is required between nodes, the messages are passed within single or master region.
- e. After all the computation tasks finish, the multiple threads of OpenMP end firstly, then the processes of MPI end.

OpenMP first generates multiple threads in the main program. Each thread is similar to a process of SPMD computing model. Therefore, these threads may be used as the processes of the SPMD computing model. They perform the same code on their own different data, respectively.

3.4 Implementation of hybrid programming

3.4.1 MPI support for OpenMP

During the realization of multi-thread parallelism of OpenMP inside the process of MPI, for the purpose of guaranteeing the safety of threads, the program must be initialized by MPI_INIT_THREAD function instead of MPI_INIT, as follows.

MPI_Init_thread (int * argc, char * ((*argv) []), int required, int * provided)

The parameter required indicates the required supporting degree of multiple threads. The parameter's possible values are as follows:

- 1 MPI_THREAD_SINGLE. Only one thread is running in the process. The MPI of current parallel machine does not support multiple threads in the process.
- 2 MPI_THREAD_FUNNELED. Multiple threads are running in the process. But only the main thread can execute the MPI function.
- 3 MPI_THREAD_SERIALIZED. Multiple threads are running in the processes. Each thread can execute the MPI function. But only one thread is allowed to do that one time.
- 4 MPI_THREAD_MULTIPLE. Multiple threads can execute the MPI function.

The multi thread level parameters for required request support for initialization, pro-vided returns the MPI environment practical support multi threading level. When the program runs, we can also use the MPI_Query_thread (int * provided) this function to judge the MPI environment supports thread level, and use MPI_Is_thread_main (int * flag) this function to determine whether the main thread running thread.

3.4.2 *Implementation of hybrid master-only model*
 The initialization and termination of MPI is realized by MPI_Init_thread () and MPI_Finalize (). For a better code portability, the number of threads in each MPI process is determined by omp_set_num_threads (n). This is one kind of simplest hybrid computing model, in which MPI is at the top layer, while OpenMP is located in the lower layer. MPI communication thread exists safely outside the parallel domain of OpenMP.

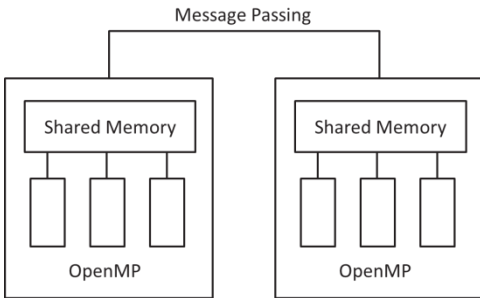


Figure 4. Hybrid master-only model.

3.4.3 *Implementation of hybrid multiple model*
 When conducting MPI inter-process communication in the OpenMP parallel region, MPI function call should be in the corresponding region of Critical, Master or Singe. OpenMP guidance can be used in the outermost layer of inter-process of communication, so that communication between processes can occur outside the parallel region, and also can occur inside the parallel region. In the parallel region, a

thread is created after the MPI process generation and initialization. The MPI inter-node communication needs thread synchronization setup, increasing the cost of the program itself and reducing the performance. In the premise of guaranteeing the thread safe of MPI system message passing library function, multiple threads can also call the message passing library functions simultaneously without mutual interference. In this computing model, inter-process communication can be executed by any one thread or some threads in the multi-thread parallel region. At the same time of inter-process communication, the other threads in the parallel region can perform computation operation, so as to realize the overlap of communication and computation.

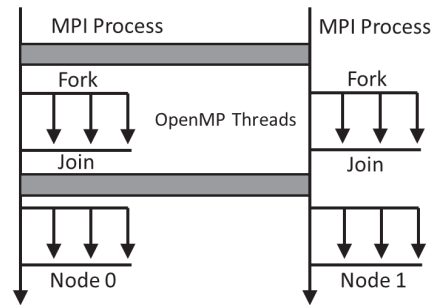


Figure 5. Hybrid multiple model.

4 CONCLUSION

Compared with the MPI message passing model, the hybrid computing model with MPI and OpenMP can make full use of the architectural features of multi-processor computer clusters better. When using MPI and OpenMP combination of programming, user usually achieves better performance of parallel programs. And the program will have good portability. Definitely, the hybrid computing model with MPI and OpenMP will be used more widely in parallel program design.

REFERENCES

- [1] HUANG K. Advanced computer architecture: parallelism, scalability, programmability[M]. Beijing: Machinery Industry Press, 1999.
- [2] MPI: A message-passing interface standard [EB/OL]. <http://www.mpiforum.org/>.
- [3] OpenMP Application program interface version 2.0 [EB/OL]. <http://www.openmp.org/>.
- [4] Quinn M J, Parallel Programming in C with MPI and OpenMP [M]. Mc Graw Hill, 2004.

Scheme study of construction of TSMDB

H.K. Tang, M.J. Liu, Y.L. Fan & C.W. Li

University of Jinan, Jinan, Shandong, China

ABSTRACT: This article analyzes the necessity of construction of a teaching source material database (TSMDB), designs the retrieve framework based on distributed database and designs the database layer of management, development and application. It also analyzes and designs the application layer from the interface of teacher construction, teacher quote and student inquiry, discusses the application style about the material database. Finally, the construction and application of the teaching material database were summarized.

1 OVERVIEW

From a historical perspective of the development of education, teaching in the form of evolution has gone through three processes. The first is one to one teaching approach, which was popular in the early history of China. This way of teaching in the early stages of development received good results. With this method, students can learn teacher's knowledge and skills more comprehensively. Students' practical ability can get plenty of exercise, but the knowledge and skills can't get quickly spread. The second stage is one teacher teaching many students. The advantage of this system is the rapid dissemination of knowledge and the optimization of teacher source. But its disadvantage is that it is not conducive to student's individualized education, and emphasize too much on book knowledge, the link between theory and practice is not effective. The last stage is combined teaching and field teaching mode, which is better for linking between theory and practice and enhancing students' ability to solve practical problems. But the implementation of teaching is more complex and, requirements to teacher are higher.

The development of modern teaching has undergone great changes, especially nowadays in the rapid development of information technology. On the basis of the comprehensive advantages of the various forms of teaching, there comes the flexible teaching mode. There are research-based teaching mode, project-driven teaching mode, and tutorial and so on. Recently, on the basis of study of teaching experience abroad reversal classroom, discussion teaching, etc. are proposed. Distance learning, self-teaching like MOOC, small course and so on are developed with the advanced teaching equipment. The Material Library is an important contemporary teaching source for students' self-study, individualized development, teaching efficiency improvement, linking between

theory and practice, etc. Current Material Library is mainly for primary and secondary education; it is not enough for higher education. Higher education also needs a material library; it is critical for students to acquire knowledge and skills and improve their practical ability.

There is a variety of modern teaching sources, including text, images and other static sources. It also includes dynamic source such as sound, image, video, animation and so on. How to build abundant teaching sources and how to use them to organize the teaching is the main problem that needs to be solved.

In this paper, the necessity of TSMDB construction, the Library's structural framework, how to use the Library to organize teaching, and how students can use the database were discussed.

2 CONSTRUCTION OF TSMDB

The construction of a teaching source materials database is a complex project. It is a project with rich content and strong usability and requires not only a lot of energy, experience and strong sense of responsibility of teachers, but also requires advanced support, follow-up database maintenance. Its management also requires the joint efforts of the teachers and students.

2.1 *The contents of TSMDB*

Content of TSMDB is organized according to profession and courses. According to the manifestations of material, it can be divided into a tiny courseware material database, multimedia material database, solutions materials database and exercise materials database.

The manifestations of the tiny courseware materials database are basic concepts, nature, theory, grammar, etc., such as the definition of Huffman tree in

data structure, referential integrity in database theory and so on. The multimedia material database includes lecturing video, audio, algorithm, animation and so on. Different from traditional instructional videos, video in material database is a micro-video; it can be an explanation of a specific problem or algorithm. It may be only a few minutes long. Solutions materials database will provide a system for process development, system design steps, etc. Exercises materials database provides exercises, questions and consolidate what they have learned and conduct unit tests, midterm tests and so on.

2.2 Distributed database technology

The traditional construction of teaching sources generally confines to school, college or department. It is not only easy to create a single and one sided teaching source and not conducive to mutual exchange and learning. So the material database construction needs joint efforts of different areas and teachers from different professional courses. Distributed database technology and cloud computing technology become necessary technical means. The following discussion of TSMDB construction is from analyzing distributed database technology.

Distributed database using the network channel connects different databases. Remote users can access remote database sources through the network.

In the database, local user access material database via LDBMS, and global user access through GDBMS.

2.3 Structural design of material database

Types of materials include images, video, animation, sounds, text and other information, the size and

complexity of their data are significantly different from traditional database, so the construction of materials database is more complex than traditional databases, the key technologies include multimedia data model, data compression and decompression, data storage management and storage, good user interaction tools, distributed technology.

According to the structure of TSMDB, it divides into the management layer, development layer, application layer, the overall structure as shown in Figure 1.

The Management Layer manages the database through MDBMS, including the structure definition of the source material database, manipulating of the source material database, the operational management of the database and the establishment and maintenance of the database.

The Development Layer develops the application program. The application program uses B / S mode structure, its benefit is that the users do not need to install the client software. The user can use and query the source material database through the browser.

The Application layer enables the user access the database through the application programming interface. The users include the teachers and students, in two categories. The teachers are not only the builders of the database, but also the users. The students are mainly the users. So the application program of the development layer mainly includes three interfaces, teacher building database, the teacher using databases and the students' inquiry the database. The interface of the teacher building the database is used for teachers' construction the database, making the materials, the upload, classification and management etc. of the database.

The teacher using database is mainly used for teachers to use the materials to build courseware and download the materials. The student inquiry

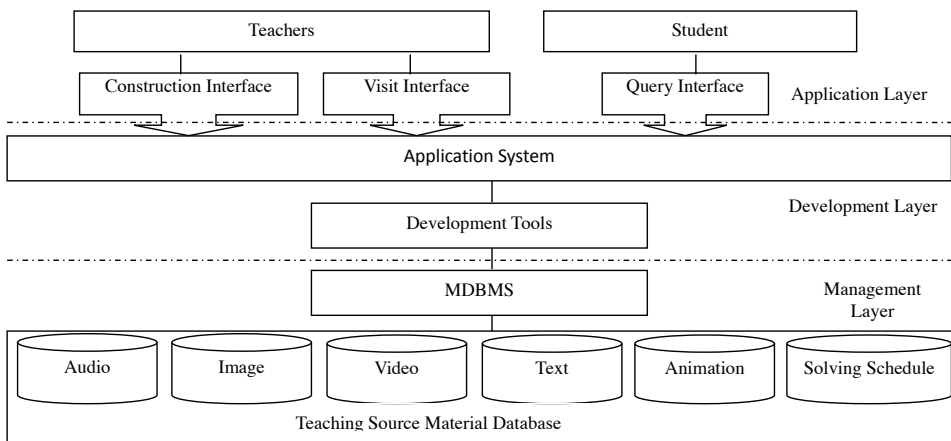


Figure 1. Structure of TSMDB.

interface mainly includes a preview of the material and download of the materials etc.

3 TEACHERS USE THE MATERIAL DATABASE ON TEACHING

The teachers are the organizer and leader of the teaching activities, will be the personnel to construct, control and building the material database. The teachers use the database in the following aspects.

3.1 *Construction material library*

Each teacher user can provide materials to the database, and also can use the materials to make the courseware in their teaching activities. Each teacher has very rich experience during the long time teaching activities. The direct result of these experiences can be described as recordings, images, video, animation, solutions and text, etc. The teachers upload these materials to the database, benefit for other teachers reference and study, at the same time, all students can inquiry it from the network. These materials can be very big (e.g., classroom video), or small fragments of few minutes (such as explanation of a certain knowledge point, even the perception and interpretation of one issue), or even the whole knowledge system of a mixture of varying materials.

3.2 *Make the teaching materials with TSMDB*

The material database is a learning platform for the students. And the material database can also provide the materials to the teachers for classroom teaching. The teachers can use the materials make the courseware to help students preview and review. The teachers can embed objects and inside links, etc. to help the students learning and review. Also for some higher learning ability students, it can guide them to solve a real problem according to the solution project, enrich after-school communications between the students and teachers, provides the effective support for the students' independent learning after school.

3.3 *Teachers can use the Library for teaching exchange*

In the source material database, the recording, images, video, animation and solutions, etc., are the experiences of the teachers for many years teaching, which focus the teacher's wisdom, experience and effort. The other teachers can use it as a reference and learning for better teaching activities, so that teachers

fully understand and master the teaching activities, the knowledge point distribution and understanding of the other school. At the same time it enriches the knowledge and skills of the teachers, provide a valuable reference for subsequent teaching and guidance. It not only solves the difficult problems of students, but also enrich the students' knowledge, train the students open thinking mode, and also avoid the limitations of the traditional teaching for the teaching discussion and exchange.

4 STUDENTS' STUDY THROUGH TSMDB

Students are another important user of TSMDB, also are the essences of TSMDB construction, and the final service object, students can use TSMDB through the following way.

4.1 *Long distance learning through TSMDB*

Classroom and laboratory is not the only way to study at the moment, students can learn whenever and wherever through a variety of ways and platforms. TSMDB not only provide support for the classroom teaching, students can also achieve distance learning and communication through the network client, such as mobile terminals, APP.

4.2 *Student querying from TSMDB*

TSMDB provides rich teaching sources, through query students can get the further explanation of certain knowledge and more detailed analysis. The database owns richer source material in the database, the source type is more complex, not only including traditional courseware, text messages, but also audio, video, animation, multimedia data, so this kind of multimedia material database can not only provide basic data information query, also provide retrieval based on the characteristics of the database, so that the students can learn and ask in anytime and anywhere. Break through the time and place limitations of traditional data query.

4.3 *Carry out the second classroom*

Through the second classroom, students can consolidate knowledge, enrich knowledge and application of knowledge in an effective means of solving practical problems, students can not only learn the knowledge from the TSMDB, and also can use a lot of solution of problems to solve the practical problems in study and life, to achieve the purpose of using their learn to the practice.

5 SUMMARY

The construction of TSMDB is a systemic engineering, its construction process needs many teachers pay a lot of time and energy, especially the teacher of cooperation and exchange from different professionals. Currently TSMDB mainly facing the students from primary and secondary schools, higher education TSMDB construction is still in an early stage. But with the abundance of teaching mode, the development of network and mobile communication technology, it provides the technical support of the construction of TSMDB. Universal education and lifelong learning concept are deeply rooted in the hearts of the people, put forward requirements of the construction of TSMDB.

REFERENCES

- [1] Wen Hengfu. Discussion on Transformation of Chinese Teaching Mode. *Journal of Chinese Society of education*, 2002, 12. 42–45.
- [2] Chu Hongqi. On the Transformation of Education Development Mode. *Education Research*. 2011(11), 3–10.
- [3] Zhang Huai-nan, Yang Cheng. A Research on Construction of Regional Universities Digital Instructional Research Based on Cloud Computing. *Modern Education Technology*, 2012(6), 45–50.
- [4] Fang Bing, Yang Cheng. Study on Universities' Information Construction of Instructional Sources in I-era: Exemplified with Open University. *JOURNAL OF DISTANCE EDUCATION*, 2013(6), 88–94
- [5] Cai Yuezhong. Discuss on the Construction of Teaching-Video Database and Routine Management. *Journal of Guangdong Polytechnic Normal University*, 2007(3), 33–36.
- [6] Guan Lili. The Frame System Study of the Material Library of Online Teaching Sources. *Journal of Shanghai Finance University*. 2005(2), 70–73.
- [7] Zhang Zheng, He Xiuling, Feng Gang, Liang Miaoyuan. Model Design of Web-Based Teaching Material House. *Computer Engineering and Applications*, 2001(11), 125–126.
- [8] Chao Jing. Study on Contents-based Retrieval. *Journal of Wuhan Engineering Institute*, 2005(12), 51–54.

Damage diagnosis of derrick steel structures based on vibration parameters and improved genetic algorithm

D.Y. Han & S.M. Wei

College of Vehicles and Energy Yanshan University, Qinhuangdao, Hebei Province, China

N. Zhang & P.M. Shi

Key Laboratory of Measurement Technology and Instrument Yanshan University, Qinhuangdao, Hebei Province, China

ABSTRACT: The damage diagnosis is one of the key techniques for safe service of derrick steel structures. In this article, the damage diagnosis of derrick steel structures is divided into two steps. According to the mechanical properties of load-bearing steel structure, the element strain modal difference is calculated by the ANSYS finite element software before and after damage, which can locate the damage location. Then, according to the frequency change of steel structure, damage level is diagnosed by genetic algorithm. It shows that the damage diagnosis results of derrick steel structures are accurate, namely, this method can successfully detect the damage location and damage level for single damage and multiple damage of large load-bearing steel structure.

1 INTRODUCTION

Derrick steel structures are widely used in many project fields, such as petroleum, mining, geological prospecting, construction, etc. (Han & Shi 2011). Because of the bearing design load and cyclic loading, the corrosive and aging of materials under the natural environment and all kinds of unexpected external forces, the derrick steel structures will inevitably lead to different damage levels in the process of manufacturing, installing, moving and decades of service.

When the local damage of structure accumulates, the capacity and durability of structure bearing large load will decline, which causes damage or even collapse for derrick steel structures. In recent years, the survey shows that most of derrick steel structures collapse accidents are upper body or the whole body collapsed, which leads to incalculable economic damage and casualties, and the fundamental reason is lacking of periodic monitoring for the key parts of the device and failing to timely discover and eliminate hidden trouble (Zou & Xie 2008). The damage diagnosis research can effectively reduce the probability of accidents.

Damage diagnosis is prerequisite for safety assessment and remaining life estimation of derrick steel structures. The research of damage location is increasingly developing and diagnosing damage level is becoming the current focus, especially for large

load-bearing steel structures. In this paper, it uses element strain modal difference to locate the damage position and then applies the natural frequency changes of structure to diagnose the damage level of derrick steel structures.

2 VIBRATION PARAMETER IDENTIFICATION THEORY

2.1 *The recognition factor of damage location*

When the structure is damaged, the local rigidity decreases and the internal stress redistributes. According to Saint Venant's principle, the change of strain far away from the damage location decays rapidly, while the strain near by changes obviously, making the local variable-strain sensitive to the damage location, which can locate damage location (Dong et al. 1999).

The dynamics equation of multi-degree of freedom linear system can be expressed as:

$$[M]\{\ddot{x}\} + [C]\{\dot{x}\} + [K]\{x\} = \{f(t)\} \quad (1)$$

Where $[M]$, $[C]$ and $[K]$ respectively are mass matrix, damping matrix and stiffness matrix of system; $\{\ddot{x}\}$, $\{\dot{x}\}$ and $\{x\}$ respectively are column vector of acceleration, velocity and displacement for all points in the physical coordinate; $\{f(t)\}$ is external load of system.

Each strain modal vector is corresponding to a displacement modal vector, and both modal coordinates are equal.

Make $\{f(t)\} = \{F\}e^{j\omega t}$, $\{x\} = \{X\}e^{j\omega t}$, through decoupling coordinate transformation, and we can get:

$$X = \Phi q = \sum_{r=1}^n q_r \Phi_r \quad (2)$$

Where Φ is regularization main modal matrix; q is generalized coordinate. It shows that the generalized coordinate component q_r defines the relative weights of components corresponding to displacement $\{X\}$ in the r -order main modal Φ_r .

Change (1) into the frequency domain equation:

$$(-\omega^2 [m_i] + [k_i] + j\omega [c_i])q = \Phi^T F \quad (3)$$

Substitute (3) into (2):

$$X = \Phi Y_i \Phi^T F \quad (4)$$

$$\text{Where } Y_i = (-\omega^2 [m_i] + [k_i] + j\omega [c_i])^{-1}$$

In the three-dimensional space, the displacement $\{X\}$ can be expressed as:

$$X = \begin{Bmatrix} U \\ V \\ W \end{Bmatrix} = \begin{bmatrix} \Phi_u \\ \Phi_v \\ \Phi_w \end{bmatrix} \begin{bmatrix} \ddots & & \\ & Y_i & \\ & & \ddots \end{bmatrix} \begin{bmatrix} \Phi_u^T & \Phi_v^T & \Phi_w^T \end{bmatrix} \begin{Bmatrix} F_x \\ F_y \\ F_z \end{Bmatrix} \quad (5)$$

According to the relationship between strain and displacement in elastic mechanics, we can get the positive strain components of the structure:

$$\{\varepsilon\} = \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{Bmatrix} = \begin{bmatrix} \frac{\partial U}{\partial x} \\ \frac{\partial V}{\partial x} \\ \frac{\partial W}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{\partial \Phi_u}{\partial x} \\ \frac{\partial \Phi_v}{\partial x} \\ \frac{\partial \Phi_w}{\partial x} \end{bmatrix} \begin{bmatrix} \ddots & & \\ & Y_i & \\ & & \ddots \end{bmatrix} \begin{bmatrix} \Phi_u^T & \Phi_v^T & \Phi_w^T \end{bmatrix} \begin{Bmatrix} F_x \\ F_y \\ F_z \end{Bmatrix} \quad (6)$$

Where $\left\{ \frac{\partial \Phi_u}{\partial x}, \frac{\partial \Phi_v}{\partial x}, \frac{\partial \Phi_w}{\partial x} \right\} = \{\psi_x, \psi_y, \psi_z\}$ is positive strain modal,

(6) can also be written as:

$$\{\varepsilon\} = \Psi Y_i \Phi^T F \quad (7)$$

Then the first-order differential of $\{\varepsilon\}$ is:

$$\{\Delta\varepsilon\} = [\Delta\Psi Y_i \Phi^T + \Psi \Delta Y_i \Phi^T + \Psi Y_i \Delta\Phi^T] F \quad (8)$$

From (8), we know that the change of strain modal of structure is mainly caused by $\Delta\Psi$, ΔY_i and $\Delta\Phi$. It means that all of them can be used as recognition factors for damage location. According to Saint-Venant's principle, the sensitivity of the three recognition factors for damage location decreases successively, and $\Delta\Psi$ and $\{\Delta\varepsilon\}$ are one-to-one corresponding relationship on the location coordinates (Huo 2010). So this paper uses element strain modal of structure as the damage location factor.

2.2 Natural Frequency

The natural frequency of the structure is easy to obtain with high test accuracy. The frequency change caused by structural damage is determined by the location and extent of the damage. If the structural damage locations are determined, the greater the damage level, the greater the frequency change (Han & Shi 2011). Though the peak of strain modal difference increases with damage level, it cannot reflect the damage level directly. However the global variable-frequency can show the damage level, by combining with genetic algorithm, we can use the frequency change of derrick steel structures before and after damage as the objective function, and then optimize calculation for damaged elements located, which can diagnose the damage level accurately.

3 THE IMPROVED GENETIC ALGORITHM

Genetic Algorithm (GA) is a global optimization algorithm that follows the process of genetic and evolution of creatures in the natural selection. It starts with generating randomly initial feasible solution which follows the principle of survival of the fittest and continues loop execution, and gradually approaches the global optimal solution finally (Chou & Ghaboussi 2001).

The GA can transform the damage diagnosis problem into variable optimization. In the case of less test information, it also can simulate the damage level. The GA is more used in the simple structural damage diagnosis, because the research on the complex structures has just started. Affected by the number of variable of the objective function, the GA has the features of large-scale and long time-consuming calculation.

In this article, damage position is located at first, and then the located damaged elements are used as variables to reduce the calculation of GA and enhance the timeliness of diagnosis of damage level. The objective function of genetic algorithm optimization is defined as the difference of frequency changes before and after damage by derrick steel structures:

$$F_\omega = \sum_{i=1}^m \left(\frac{f_i^{test} - f_i^{cal}}{f_i^{test}} \right)^2 \quad (9)$$

Where f_i^{test} is the measured i -order natural frequency and f_i^{cal} is the calculated i -order natural frequency.

The article utilizes the strategies of real valued coding and best individual preservation simultaneous, and sets initial value of 1. By using several located elements as program variables, we can make the main program of damage diagnosis of derrick

steel structures in MATLAB, and then repeat calling the APDL program of ANSYS to realize the diagnosis of damage level by deducing the extent of structural damage.

4 CALCULATION EXAMPLE AND ANALYSIS

Through above analysis, we can apply a two-step method based on strain and frequency parameters to diagnose the damage of JJ220/42-K type derrick steel structure model. In the finite element analysis model, the main structure of derrick can be simplified as a three-dimensional elastic beam element (BEAM188). The K-type derrick steel structure model has four segments, each segment has five elements, and it altogether has eleven kinds of sectional properties. The mass density of Q235 type steel which is applied to K-type derrick steel structure model is 7830 kg/m^3 , and the Poisson's ratio is 0.3. The derrick model is 3.5 meter high and is divided into 290 nodes and 197 elements. Derrick bottom nodes limit the moving degree of freedom in the x, y, z directions and the rotational degree of freedom in the y, z directions. The junction that connects the front leg of the derrick and the bottom beam limits the degree of freedom in the y direction. The 5100N load is separately applied to the derrick top four nodes to simulate the load of derrick in actual working conditions, including the maximum hook load, rope and crown block self-weight etc. Derrick Steel structure diagram is shown in Figure 1.

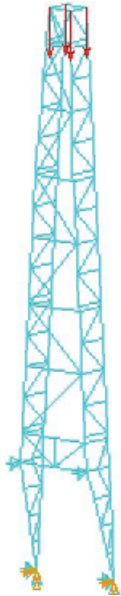


Figure 1. JJ220/42-K type derrick steel structures model.

This paper only uses the left rear leg of derrick steel structures as the research subject which is divided into 22 elements from bottom to top. This article sets up several damage working conditions in order to simulate actual damage of derrick steel structures. Damage working conditions of derrick steel structures are shown in Table 1:

Table 1. Damage Working Conditions of Derrick Steel Structures

| Damage condition | Damage location | Damage level |
|------------------------------|-----------------|--------------|
| The first working condition | 13th element | 5% |
| The second working condition | 13th element | 30% |
| The third working condition | 7th element | 20% |
| | 13th element | 50% |
| | 18th element | 10% |

4.1 Damage Location Identification

This article uses finite element software for modal analysis of derrick steel structures, and separately extracts the first-order element strain modal of derrick steel structures left rear leg before and after the damage to calculate the differences. The results show that the strain modal of damaged elements has changed remarkably when structures are damaged, regardless of damage level, while strain modal of other undamaged elements has changed slightly mainly. The element strain modal difference of derrick steel structures in three working conditions can be expressed as bar graphs (Figs 2-4).

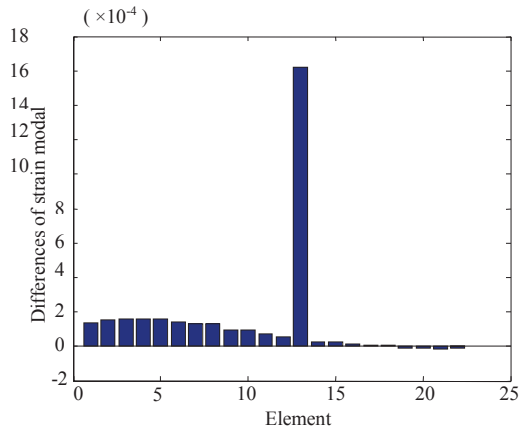


Figure 2. Element strain modal and damage location diagram of the first working condition.

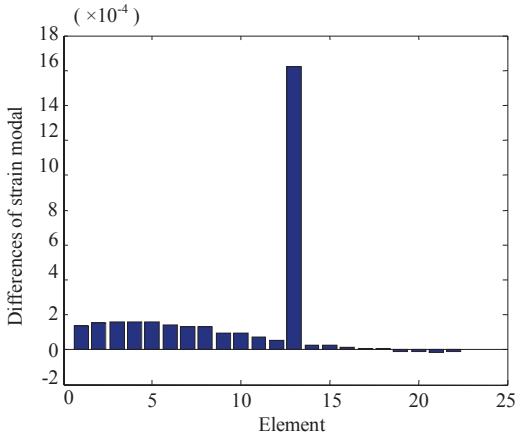


Figure 3. Element strain modal and damage location diagram of the second working condition.

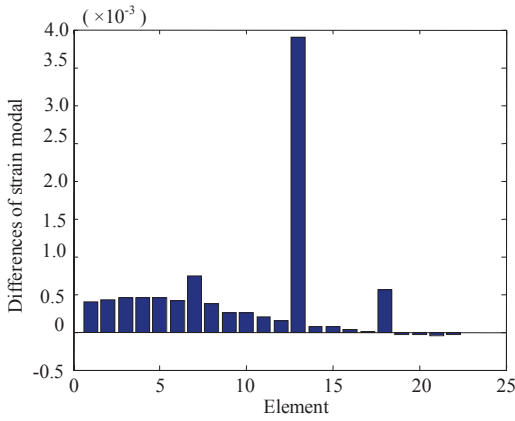


Figure 4. Element strain modal and damage location diagram of the third working condition.

4.2 Damage Level Identification

After locating the damage position, this article uses genetic algorithms to establish the objective function whose variables are damaged elements to calculate damage level. In this paper, real value coding is used in genetic algorithm, namely, each element corresponds to one variable. In the process of calculation, two variables are randomly selected from cross function to cross for increasing the range of function optimization. For the single damage scenario, the one element located by the method of damage diagnosis and the other two elements randomly selected from cross function are used as variables to calculate damage level. The experimental results show that it does not affect the computational results, which enhances the versatility of the main

function of genetic algorithm. The computational results are expressed as Figures 5–7:

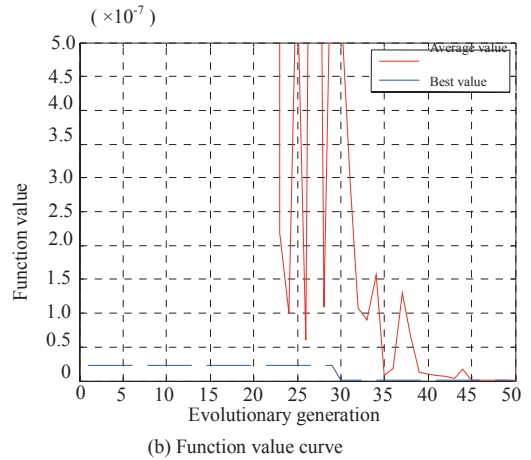
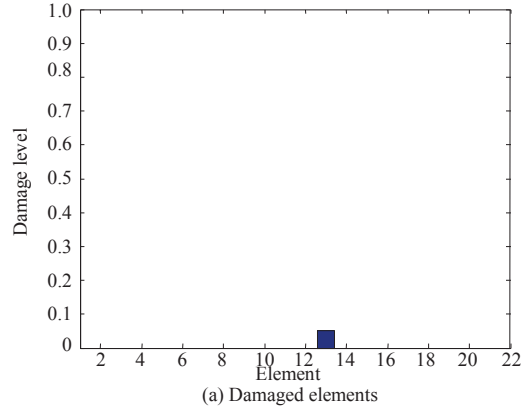
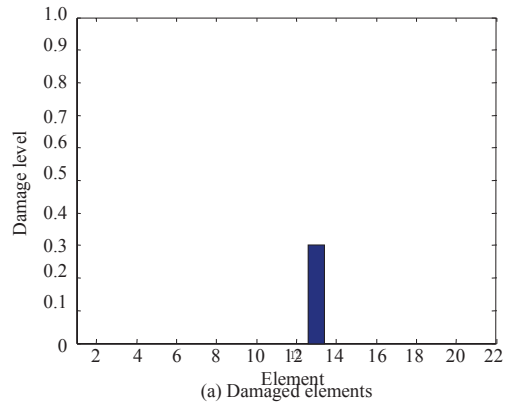


Figure 5. Damage level identification of the first working condition.



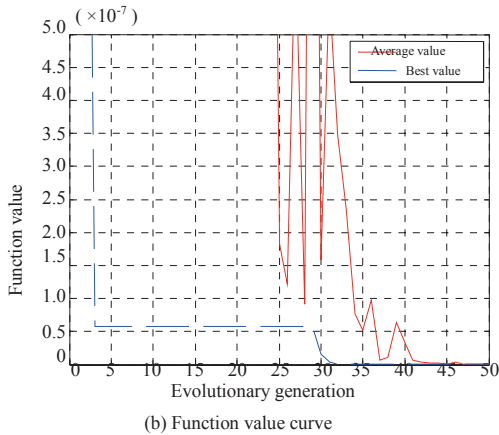


Figure 6. Damage level identification of the second working condition

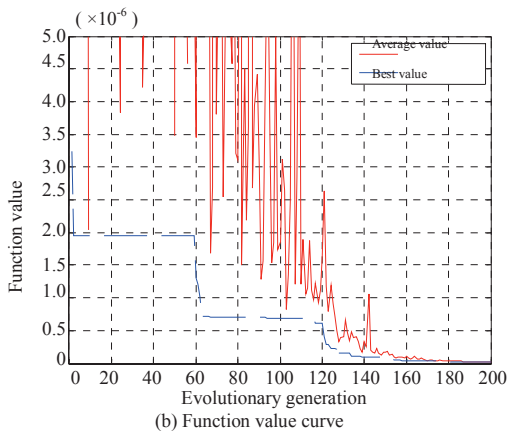
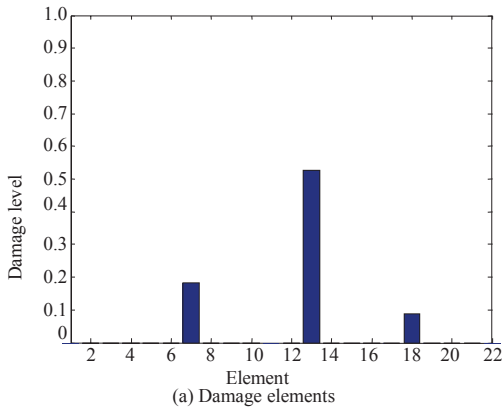


Figure 7. Damage level identification of the third working condition

In conclusion, after the damage position are located successfully, the damage level of derrick steel structures can be diagnosed accurately by only using natural frequency in genetic algorithms. It needs to be pointed that, in the case of multiple damage, there are few errors in the damage level detection result of genetic algorithm which only uses natural frequency as varied. Meanwhile, it needs longer time for calculation compared to single damage scenario.

5 CONCLUSIONS

- 1 Damage location can be located accurately by using the element strain modal difference of derrick steel structures before and after the damage.
- 2 Basing on the result of the first step, by only using natural frequency as variable, the damage level of derrick steel structures can be diagnosed by genetic algorithm.
- 3 In the case of multiple damage, there are few errors in the detection result of genetic algorithm which only uses natural frequency as varied. Meanwhile, it needs longer time for calculation compared to single damage scenario.

ACKNOWLEDGMENT

Project 51104129 and 51475407 supported by the National Natural Science Foundation of China.

REFERENCES

- [1] Chou J.H. & Ghaboussi J. 2001. Genetic algorithm in structural damage detection. *Computers and Structures* 79(14): 1335–1353.
- [2] Dong C. Ding H. & Gao S. 1999. Basic principle and method of structural damage identification and location. *China Railway Science* 20(3): 89–94.
- [3] Huo X. 2010. Research on the structural damage identification based on strain mode technology. Beijing: Beijing University of Post and Telecommunications.
- [4] Han D.Y. & Shi P.M. 2011. Damage identification of derrick steel structures based on frequency and equivalent damage coefficient. *Engineering Mechanics* 28(9): 109–114.
- [5] Zou L.Q & Xie C.Q. 2008. Damage identification of oil derrick based on normalized change ratio of frequency and neural network. *China Safety Science Journal* 18(10): 30–33.

A resource allocation economic model for Diffserv network

B. Zeng, L. Yao & W. Hu

Department of Management, Naval University of Engineering, Wuhan, Hubei, China

ABSTRACT: A network system can be said to be efficient if the overall user satisfaction of the network usage in this system is maximized, which results in the concept of differentiated services (DiffServ) network. An economic approach facing different service classes has been shown, under the assumption of a competitive system with fixed pricing for service classes and monotone user utility function, users will choose the service class that can provide them with the highest value. The value depends both on the price and on the system quality. Because the differentiated services model is formulated quite generally, the result here not only explains well the behavior of bandwidth usage, but also gives important managerial implications for differentiated service provider.

1 INTRODUCTION

To address the problem of Internet network performance degradation, many researchers have recently resorted to the economic approach. The network bandwidth provision can be thought of as a problem of resource allocation. Thus, by applying proper resource allocation mechanism, the network can get substantial improvements in its performance.

The goal of network design should be meeting users' needs of network usage. That is, users should be able to get their desired services of class at their chosen cost. Thus, we think that a differentiated services network with the pricing mechanism will be the prevalent and proper infrastructure of the future Internet. The DiffServ architecture of the IETF is one of the main proposals that aim to provide multiple and differentiated service classes in the network systems (Hosek 2013).

Several differentiated network services pricing schemes have been proposed in the recent years. The Paris Metro Pricing (PMP) proposal aims to provide a rather simple and scalable minimalist differentiated services solution. The PMP mechanism divides the network into several separate channels with each being charged differently. The number of designed channels should be small to fit the users' preference for simplicity (Chi-Kin 2010, 2014). The key feature of PMP is that it relies solely on pricing to provide differentiated services (Dongmyung 2012). However, the PMP mechanism does not give a detailed and economic model to analyze user behavior under the PMP system.

However, many other studies focus more on the priority mechanism to provide differentiated services. Hui et al (2011) developed a pricing scheme with

priorities. Huan et al (2012) also presented a priority pricing model and showed that pricing with priority services could result in an optimal allocation of resources that maximized the total welfare of all users. In their model, the network changes dynamically the rental prices for the corresponding priorities. Users collect the current pricing information and predict waiting times from the network and then decide whether it is worthy of initiating their transmissions.

In our research, while the above proposals and research works do give a valuable background to our studies, our work focuses on analyzing a general differentiated services pricing model. We try to construe and model user behavior in a differentiated services network regardless of how the service classes are differentiated, i.e. the PMP or priority-based approaches. Also, by gaining more information about user behavior, we then try to analyze the system equilibrium for the various differentiated pricing designs. We have given a much simpler method for analyzing user behavior in the differentiated services system. Most importantly, our analysis can provide a deep and valuable economic description of user behavior, which may aid in judging the design and implementation of the differentiated services.

2 NETWORK MODEL

2.1 *The network model*

In this chapter, we study and model the differentiated service provision problem with an economic approach. To simplify the analysis, in this research, we consider only the single link case. The differentiated services model is highly abstracted as a

link to a set of service classes $\{C_1, C_2, \dots, C_N\}$. We assume all the service classes to be different or differentiated only in the associated transmission probabilities $\{p_1, p_2, \dots, p_N\}$. That is, we assume that the only difference to distinguish these different service classes is reflected on their corresponding transmission probabilities. Also, we assume the bandwidth to be indefinitely divisible; that is, the link is regarded as a fluid model. Let u_i be the unit price of the unit bandwidth usage (or per packet) with service class C_i . Considering the complexity of dynamic price updating mechanisms as those in (Perambur 2011), we adopt their suggestions that the users prefer a predictable or rather flat setting of pricing rates. Thus, we further assume that the unit price u_i is held fixed and will not change with the link congestion level¹.

Assume that there are J users using this link, let $x_j(i)$ be a user j 's allocation in the service class C_i . Note that user j allocates $x_j(i)$ but is not guaranteed to have successful transmissions up to $x_j(i)$ in C_i because C_i is associated with the transmission probability, p_i . On the other hand, user j can only anticipate an average transmission of $x_j(i) * p_i$ for his or her allocation in C_i .

2.2 Modeling user behavior

In our research, similar to most network pricing studies, we adopt the utility functions to model and analyze the user behavior. It is rather difficult to find an exact utility function to describe precisely users' preferences for bandwidth usage; however, careful observations on the network applications can provide some proper assumptions and simplifications on the utility function forms that can be used to model the user behavior very well. In short, the network usage can be categorized as data and real-time applications. Data traffic (or sometimes being referred to as elastic traffic) can be much more tolerant of occasional network performance degradation. Current network traffic is dominated by data or elastic applications such as E-mail, Web surfing, terminal communications, and file transfers. However, real-time traffic has stringent requirements of bandwidth and cannot stand network jitters or performance variations.

For our analysis, we assume user j 's utility or satisfaction as a function of the average throughput of his or her allocations in all the service classes.

$$X_j = \left(\sum_{i=1}^{i=N} p_i * x_j(i) \right) \quad (1)$$

Modeling users' perceived performance with anticipated measures is common in the literature. Though such representations are appropriate for elastic applications, they cannot catch the properties

of real-time applications very well because of the hard requirements of real-time quality of service. We focus on the elastic users for two main reasons. First, by observing elastic applications such as E-mail and WWW reveals that current Internet usage is undeniably dominated by elastic traffic. So it is of great interest to analyze such user behavior under the deployment of a differentiated services pricing architecture. Second, the characteristics of real-time traffic are quite different from those of elastic traffic. To model real-time traffic properly one should at least put performance variations such as loss variations or jitters into consideration. Thus, to make the analysis more tractable, in this chapter, we focus on modeling the behavior of elastic users only.

Assume user j 's utility function, $U_j : R^+ \rightarrow R^+$ to be a strictly increasing function in the sense that $U_j(x) > U_j(y)$ if and only if $x > y$. Note that the assumption does not require any continuity or differentiability of U_j . Furthermore, the subscript of j allows heterogeneous preferences for users as long as each U_j obeys the strictly increasing property. Thus, compared with other model formulations, our approach presents a more general analysis of this problem.

3 DIFFERENTIATED SERVICES ECONOMIC MODEL

3.1 The non-cooperative game model

The phenomenon of congestion externality implies that a good formulation such as the non-cooperative game model that can catch human selfishness well is particularly appropriate for analyzing this problem. Thus, in our model we assume that all users are maximizing their own net benefit of bandwidth usage simultaneously. Our modeling is in essence a non-cooperative game. To further simplify the problem, we formulate the network as a competitive market in the sense that no single user can affect the transmission probabilities. If we regard the transmission probabilities as the offered quality, then the system is competitive because every user is a quality taker. Note that, as pointed out in Monsef (2014), this assumption is reasonable when there are so many small users on the link that individual influence on the transmission probabilities is negligible or when it is impossible or too expensive for a user to determine how the choice of an allocation may affect the transmission probabilities. Thus, the formulations here can also be used in the case of a big Internet service provider (ISP) with numerous small users; however, it does not fit well for some corporate Intranets or campus networks where congestion is often ascribed to some particular departments or agents.

In addition, we assume that the unit price u_i and the transmission probability p_i of each service class

C_i can be accessible to every user. As the price setting is kept fixed, the price information can be posted in advance to all users. The probabilities, though changing with the system load, can be made available to all users through some feedback mechanism. In our analysis, we assume that users can obtain enough information for deciding their packet allocations.

Let $B_j = U_j(\sum_{i=1}^N p_i * x_j(i))$ be a user j 's benefit function. Then in our differentiated services model we analyze user j 's behavior as an optimization problem in which user j determines an ideal allocation $\{x_j(1), x_j(2), \dots, x_j(N)\}$ to maximize his or her benefit function according to the announced prices $\{u_1, u_2, \dots, u_N\}$ and the transmission probabilities $\{p_1, p_2, \dots, p_N\}$ in the network system. If M_j denotes user j 's budget to access the net, then the optimization model (I) of user behavior can be defined as follows

$$\text{Max} B_j = U_j(\sum_{i=1}^{i=N} p_i * x_j(i)) \quad (2)$$

$$\text{subject to } \sum_{i=1}^{i=N} u_i * x_j(i) \leq M_j$$

The assumption of competitiveness that any single user cannot affect the transmission probabilities means $\frac{\partial p_i}{\partial x_j(i)} = 0$. That is, for the optimization problem of every user, the transmission probabilities $\{p_1, p_2, \dots, p_N\}$ can be taken as exterior constants. With the assumptions on competitiveness and the utility function, we then can have the following theorem.

(Theorem I) : If $U_j : R^+ \rightarrow R^+$ is a strictly increasing function and $\frac{\partial p_i}{\partial x_j(i)} = 0$, then the optimal allocation

$\{x_j(1), x_j(2), \dots, x_j(N)\}$ in Optimization Model (I) is the same as that in Optimization Model (II) defined as follows.

$$\text{Max} B_j = \sum_{i=1}^{i=N} p_i * x_j(i) \quad (3)$$

(Proof)

$U_j(X_j)$ is a strictly increasing function in X_j and because $\frac{\partial p_i}{\partial x_j(i)} = 0$, $X_j = (\sum_{i=1}^{i=N} p_i * x_j(i) \geq 0)$ is clearly also increasing in $x_j(i)$. Then by the composition of increasing functions, the allocation of $\{x_j(1), x_j(2), \dots, x_j(N)\}$ that maximize $X_j = \sum_{i=1}^{i=N} p_i * x_j(i)$ under the

linear constraint of $\sum_{i=1}^{i=N} u_i * x_j(i) \leq M_j$ should also act as the optimal solution for Optimization Model (I).

Theorem I states that though U_j is allowed to be heterogeneous, if the system is competitive and U_j is strictly increasing, we can solve Optimization Model (I) by Optimization Model (II). Theorem I also reveals some interesting and meaningful economic implications. If we further combine the assumption of differentiability on U_j , combined with

the competitiveness assumption of $\frac{\partial p_i}{\partial x_j(i)} = 0$, then $U_j(X_j) = U_j(\sum_{i=1}^{i=N} p_i * x_j(i))$ denotes in essence that the

user has the preference of "perfect substitute" for different service classes. In other words, the economic marginal rate of substitution (MRS) turns out to be a constant ratio in this case. The MRS is an economic terminology that describes a subjective exchange ratio between goods. Take the system with only two service classes for example. Such differentiated services system is commonly discussed in the literature and the two service classes are often called the premium and the best-effort classes. Let transmission probabilities of the premium and the best-effort classes be p_p and p_b , respectively. Then $X_j = p_p * x_j(p) + p_b * x_j(b)$, so the MRS of user j for the premium and the best effort classes is denoted as:

$$\begin{aligned} \text{MRS}_{pb}^j &= \frac{MU_{x_j(p)}}{MU_{x_j(b)}} = \frac{\frac{\partial U_j(X_j)}{\partial X_j} * \frac{\partial X_j}{\partial x_j(p)}}{\frac{\partial U_j(X_j)}{\partial X_j} * \frac{\partial X_j}{\partial x_j(b)}} \quad (4) \\ &= \frac{\frac{\partial U_j(X_j)}{\partial X_j} * p_p}{\frac{\partial U_j(X_j)}{\partial X_j} * p_b} = \frac{p_p}{p_b} \end{aligned}$$

Which is a constant ratio for user j under the assumption of competitiveness. Thus, the subjective exchange ratio between the premium and the best-effort classes for user j is $\frac{p_p}{p_b}$. In other words, for user j , one packet of the premium class is considered of the same satisfaction as $\frac{p_p}{p_b}$ packets of the best-effort class. If user j is asked to cut down his or her transmissions by one premium class packet, for he or she to maintain the same satisfaction level of network usage, user j should be compensated with $\frac{p_p}{p_b}$ packets of the best-effort class.

According to the above analysis, if U_j has the property of perfect substitute, then the MRS is indeed a constant. That is, the utility function U_j can be reduced to a linear function that can represent the same preference of perfect substitution for user j . Under our differentiated services model where the service quality is assumed to be differentiated only in transmission probabilities, the corresponding quality of each service class can be defined accordingly by the transmission probability. Therefore, the subjective exchange ratio (the MRS), which equals the ratio of transmission probabilities between different service classes, reflects essentially the ratio of perceived quality between different service classes. Thus, in the example of the system with two service classes, the quality of one premium class packet is perceived the same as $\frac{P_p}{P_b}$ packets of the best-effort class.

3.3 Economic interpretation of the user behavior

For convenience, we also illustrate our analysis with the system of only two service classes. As seen in the proof of Theorem 1 given above, it is clearly seen that user j judges whether the premium or the best-effort class should be used by comparing the

$MRS_{pb}^j (= \frac{P_p}{P_b})$ with the price ratio ($= \frac{u_p}{u_b}$). The MRS_{pb}^j represents user j 's subjective comparison between the two service classes. That is, one premium packet and $\frac{P_p}{P_b}$ best-effort packets are considered of the same quality in user j 's mind. In other words, it does not matter subjectively to user j (that is, without gains and without loss) in abandoning one premium packet but getting $\frac{P_p}{P_b}$ best-effort packets in return.

However, the price ratio is an objective exchange ratio in the market. Once announced in the system, it does not change with the user preference. In fact, it should be fair, open, and identical (under the fixed price setting) to all users. Thus the price ratio of $\frac{u_p}{u_b}$ means that the user can save *objectively* the expense of one premium packet for getting $\frac{u_p}{u_b}$ best-effort packets. Therefore, if the subjective exchange ratio ($\frac{P_p}{P_b}$) is less than the objective one ($\frac{u_p}{u_b}$), then the behavior of exchanging premium packets with best-effort packets will be beneficial to the user. If the system is

competitive such that the MRS for the user turns out to be a constant, then the inequality between the two ratios continues and so does the exchange behavior. So finally all the allocation of the user will be put in the best-effort class. On the other hand, with similar analysis, we can assert that the user will transfer all their traffic from the best-effort class to the premium class under the inequality of $\frac{P_p}{P_b} > \frac{u_p}{u_b}$. However,

if the subjective and objective ratios are equal, then it makes no difference to the user in any allocations between the two service classes and therefore both classes would be in use concurrently.

4 CONCLUSION

In this paper, we have presented through reviewing the mainstream research works on the network service architecture that the differentiated bandwidth services mechanism is getting necessary for network service provision. From the views of both network management on the technical side and service provision management on the business side, we also find that pricing schemes, especially differentiated pricing, go very well with the proposed differentiated services architecture. Therefore, we try to analyze theoretically the differentiated services pricing system by abstracting highly the differentiated services model. Our analysis aims to discover the interaction and behavior of network users toward the differentiated services pricing architecture.

REFERENCES

- [1] Chi-Kin, C., Qian, W. & Dah-Ming, C. 2010. On the Viability of Paris Metro Pricing for Communication and Service Networks. *2010 Proceedings IEEE INFOCOM, San Diego, CA, 14–19 March 2010*. USA: IEEE.
- [2] Chi-Kin, C., Qian, W. & Dah-Ming, C. 2014. Economic Viability of Paris Metro Pricing for Digital Services. *ACM Transactions on Internet Technology* 14(2): 16–28.
- [3] Dongmyung, L., Taehyun, K., Jeonghoon, M., et al. 2012. Paris Metro Pricing for Internet Service Differentiation. *Game Theory for Networks* 75(2): 628–641.
- [4] Hui, C., Jun, M., Huaying, S. 2011. Priority Pricing Based Network Resources Allocation Scheme for Mobile Data Communication. *7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*. Wuhan, China. 23–25 Sept. 2011. USA: IEEE.
- [5] Huan, H., Ke, X., Ying, L. 2012. Internet resource pricing models, mechanisms, and methods. *Networking Science* 1(4): 48–66.

- [6] Hosek, J., Molnar, k., Rucka, L., et al. 2013. SNMP-based acquisition system for DiffServ parameters. *Telecommunication Systems* 52(3): 1595–1604.
- [7] Monsef, E., Anjali, T., Kapoor, S. 2014. Price of Anarchy in network routing with class based capacity *2014 Proceedings IEEE INFOCOM, Toronto, ON. April 27 2014-May 2 2014*. USA: IEEE.
- [8] Perambur, S., Renata, C. 2011. Consumer benefit versus price elasticity of demand: a nonlinear complex system model of pricing internet services on QoS-centric architecture. *NETNOMICS: Economic Research and Electronic Networking* 12(1): 31–60.

A study of active steering system for safety of a railway vehicle

K.S. Sim, T.W. Park, I.K. Hwang & M.J. Kim

Graduate School of Mechanical Engineering, Ajou University, Suwon, Korea

ABSTRACT: Analysis for reducing wear of wheels is implemented using a railway vehicle. In order to reduce wear number of wheels, an active-steering system is investigated. An active steering bogie system of a railway vehicle is designed to decrease wear of wheel/rail interaction. Results of simulation, which are wear number of wheels, show more improvement than the conventional railway vehicle. The proposed model is more effective on curved track.

1 INTRODUCTION

The conventional railway vehicle has a first suspension, which has very stiffness characteristics of springs or bushings because of running stability. So the performance of steering of the railway vehicles is limited on curved track. Also, as a result of wheel/rail contact interaction, the lateral forces of these railway vehicles are significantly increased. These phenomena cause the wear of wheel/rail, noise and vibration, etc.

In order to improve curving performance of a railway vehicle, concepts of steering system are presented for a long time. One commercialized steering bogie is forced by steering bogie which is only composed of links and joints. This concept of mechanism is the basic and early stage of development systems for steering. Wheel-sets can be steered by the connection element when the links between a car-body and a bogie frame are activated on the curving negotiation. The performance of this concept, however, is limited according to circumstances because of the passive element.

In this study, an active steering system is implemented when a railway vehicle passes a curved track. Additionally, the practical possibility of an active steering system that adopts the direct method is discussed by virtual analysis of an active steering system compared with a passive system.

2 DYNAMIC MODEL OF RAILWAY VEHICLE

2.1 Model of full-scaled vehicle

In this paper, a conventional train using numerical model is shown in [Figure. 1](#). This model consists of single car-body and two bogies. Dynamic characteristics of a conventional train in service are applied to MBS (multi-body system) simulation tool.

The suspension is divided into a primary suspension system and a secondary system. A primary suspension consists of axle springs which are connected to a wheel-set and a bogie frame. A secondary suspension consists of air springs and dampers. These suspension elements connect a bogie frame with a car body. Also, a center-pivot connects them and provides additional degrees of freedom with each other. This system is implemented as a kinematic and compliance structure. The bogie model is shown in [Figure. 2](#).

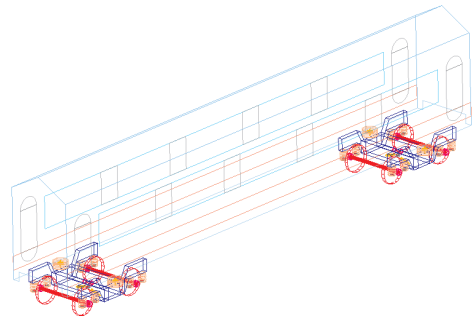


Figure 1. Full scaled model of railway vehicle.

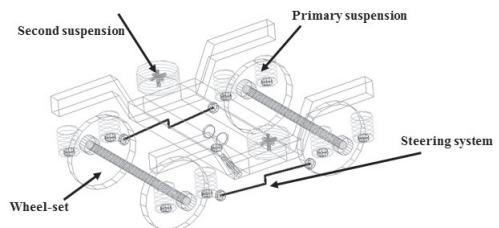


Figure 2. Bogie model of a railway vehicle.

2.2 Concept of active steering system

The concept of an active steering system is a method of reducing the angle of attack in the radial direction of the wheel-set on a curved track in order to locate the wheel-set in the radial direction. In this study, the active steering bogie is implemented by a simple control law: that is PID control. PID control is implemented by using MATLAB/ Simulink using dynamic analysis.

2.3 Condition of analysis

In this study, a dynamic model for active steering is implemented by using commercial software VI-Rail. This software is based on multi-body dynamics and is an exclusive analysis program for railway vehicles.

In order to analyze the wheel and rail contact interactions, Kalker's FASTSIM is applied to an MBD model. Also, in order to make a comparison of a conventional railway vehicle and an active steering vehicle, a small curved track is modeled. The radius of curved track is 300m, the length of transition curve is 200m and the cant is 152mm. The running speed of a simulation model is the constant speed of 70km/h between a straight and a curved track.

2.4 Results of analysis

The results of the simulation between the passive and active steering vehicles are shown in Figures 3-4. First, this study is proposed for a steering method in the radial position of the wheel-sets.

The wear of leading wheel-sets in the passive bogie is significantly larger than those of trailing wheel-sets. Thus, the wear of leading wheel-sets should be reduced using the radial steering method for smooth and stable passing curves. Wear numbers are significantly reduced in the active steering bogie.

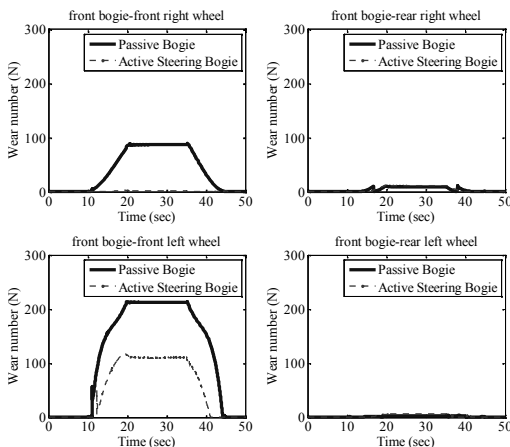


Figure 3. Wear number of a front bogie.

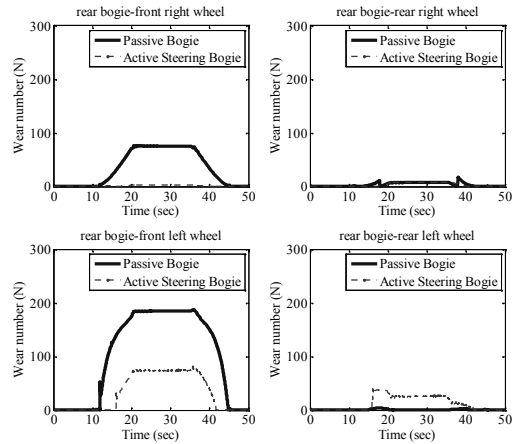


Figure 4. Wear number of a rear bogie.

3 CONCLUSIONS

The aim of this study is to assess the ability of an active steering bogie to run effectively on a curved track. A simple control logic is applied to an MBD model of an active steering system using VI-Rail and MATLAB/Simulink. Additionally, the status of the wheel and rail contact interaction of the proposed steering model is compared to that of a passive vehicle that runs on a curved track. The proposed method, which reduces wear number in the radial direction of the wheel-set on the curved track, is applied to an active steering bogie.

For further study, an active steering bogie is being developed and will be tested on a test-bed and a real track.

ACKNOWLEDGEMENT

This research was supported by a grant (14RTRP-B067983-02) from Railroad Technology Research Program funded by the Ministry of Land, Infrastructure and Transport of Korea government.

REFERENCES

- [1] Gear, C.W. 1971. Simultaneous numerical solution of differential algebraic equations, *Trans. of IEEE circuit theory* 18(1): 89–95.
- [2] Hur, H.M., You, W.H., Park, J.H., Kim, M.S. & Kim N.P. 2010. Active steering control strategy of railway rolling-stock using the estimated curvature of curve section. *Proc. of spring of Korea Society of Mechanical Engineers*:326–327.
- [3] Mei, T.X. & Goodall, R.M. 2000. Wheelset control strategies for a two axle railway vehicle. *J. Vehicle system dynamics* 33:653–664.

- [4] Pearson, J.T., Goodall, R.M., Mei, T.X. & Himmelstein, G. 2004. Active stability control strategies for a high speed bogie. *Control Engineering Practice, Elsevier* 12:1381–1391.
- [5] Sim, K.S., Park, T.W., Lee, J.H. & Kim, N.P. 2014. Optimization of characteristics of longitudinal creepage for running stability on sharp curved track, *Journal of the Korean Society for Railway* 17(1):19–27.
- [6] Swenson, C.A & Scott, R.T. 1996. The effect of locomotive radial steering bogies on wheel and rail wear, *ASME/IEEE joint railroad conference*: 91–100.
- [7] Thompson, D. 2009. *Railway noise and vibration, Amsterdam, The Netherlands.*

A development of the wheel-rail dynamic analysis program model with two-point contact algorithm

I.K. Hwang, T.W. Park, K.S. Sim & M.J. Kim

Graduate School of Mechanical Engineering, Ajou University, Suwon, Korea

ABSTRACT: As time goes by, the driving speed of railway vehicles has been increased. So the dynamic performance and stability of the vehicle have also become an important issue these days. The analysis of dynamic performance, therefore, is one of the main considerations before manufacturing and experiment. In this study, a module of dynamic analysis has been developed for railway vehicles. And an analysis of the dynamic behavior of a railway vehicle's wheel-set is carried out.

1 INTRODUCTION

The dynamic behavior of the vehicle is considerably depending on the contact between wheel and rail. The contact between them is able to be defined by the geometry of wheel-rail and the driving conditions. The dynamic behavior of wheel-set and the penetration phenomena or force between the wheel-rail can be calculated from the contact. Items to analyze the dynamic attitude such as creep and normal force are also computed by the contact. In this study, a dynamic analysis module for railway vehicle has been developed. Wheel-rail geometry and driving condition have been considered, which are attributed to the multi-body dynamics of a railroad vehicle. A contact area between wheel and rail assumed as elliptical surface based on the Hertzian theory, and the creep forces are computed by the creepages that Kalker's linear and polach nonlinear theory provided.

2 FORM OF MODULE

Two-point contact between wheel and rail which may be occurred during driving on the track is also applied in the code. The analysis with driving condition is more reasonable to consider the two-point contact. The lateral force on the wheel-set has also been assigned to the driving conditions. As shown in Fig. 1, all procedure of computation is not only considered two-point contact but also operated based on force calculation. There is no force constraint to search the contact point in the procedure, so the contact point does not have to be existed. Unlike others, therefore, Derailment can be detected in this module. The more reasonable phenomenon can be expressed by the calculation with regard to considering the reaction force.

3 DYNAMIC ANALYSIS

At first, the properties of material and profile of wheel/rail that are received are used for making generalized coordinate system. Also, it would be used to define surface parameters. As you can see in Figs. 2 and 3, there are four surface parameters which are to facilitate programming in dynamic analysis.

Therefore, the vector between the coordinate system to the point of contact can be expressed by using an Euler transformation matrix (1) and the local coordinate (2) that contains the rotation information

$$A^W = \begin{bmatrix} \cos \psi \cos \theta - \sin \psi \sin \phi \sin \theta & -\sin \psi \sin \phi & \cos \psi \sin \theta + \sin \psi \sin \phi \cos \theta \\ \sin \psi \cos \theta + \cos \psi \sin \phi \sin \theta & \cos \psi \sin \phi & \sin \psi \cos \theta - \cos \psi \sin \phi \cos \theta \\ -\cos \phi \sin \theta & \sin \phi & \cos \phi \cos \theta \end{bmatrix} \quad (1)$$

$$\bar{u}^W = \begin{bmatrix} g(s_1^W) \sin(s_2^W) \\ \mp 0.5L_W \pm s_1^W \\ g(s_1^W) \cos(s_2^W) \end{bmatrix}, \bar{u}^r = \begin{bmatrix} s_1^r \\ \mp 0.5L_R \pm s_1^r \\ f(s_2^r) \end{bmatrix} \quad (2)$$

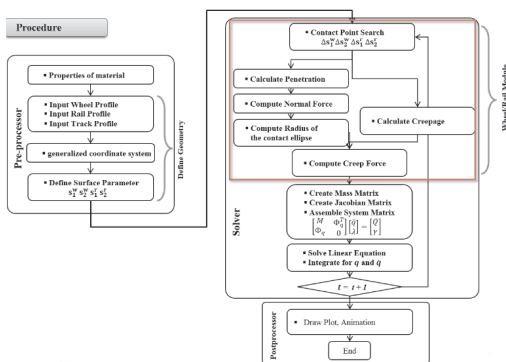


Figure 1. Process of the analysis procedure.

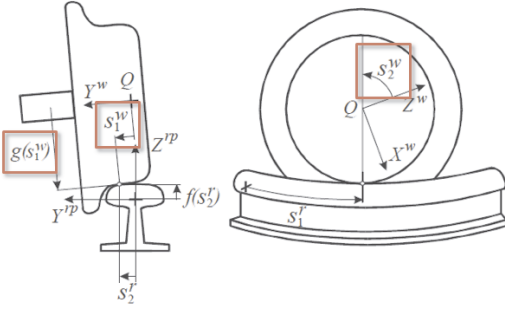


Figure 2. Wheel surface parameters.

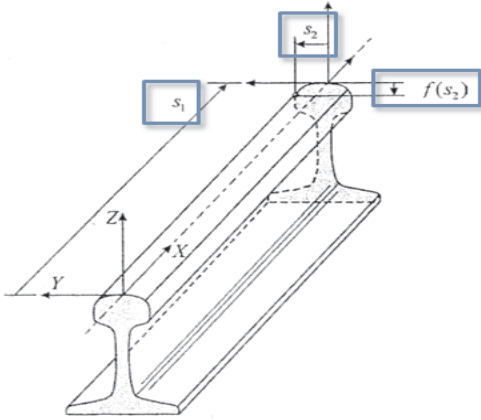


Figure 3. Rail surface parameters.

$$\bar{r}^w = \bar{R}^w + A^w \bar{u}^w, \bar{r}^r = \bar{R}^r + A^r \bar{u}^r \quad (3)$$

At equation (1), respectively express yaw, roll, and spin. and also indicate the center of left-right wheel.

The second process would be at the heart in this module. Because it looks for the point of contact between wheel and rail, by using the preceding information. And then calculates the penetration. In this process, Elastic Contact Formulation-Algebraic (ECF-A) is used.

$$\begin{bmatrix} t_1^w \cdot t_1^w & t_1^w \cdot t_2^w & \frac{\partial t_1^w}{\partial s_1^w} \cdot r^{wr} - t_1^w \cdot t_1^r & \frac{\partial t_1^w}{\partial s_2^w} \cdot r^{wr} - t_1^w \cdot t_2^r \\ t_2^w \cdot t_1^w & t_2^w \cdot t_2^w & \frac{\partial t_2^w}{\partial s_1^w} \cdot r^{wr} - t_2^w \cdot t_1^r & \frac{\partial t_2^w}{\partial s_2^w} \cdot r^{wr} - t_2^w \cdot t_2^r \\ \frac{\partial t_1^w}{\partial s_1^w} \cdot n^r & \frac{\partial t_1^w}{\partial s_2^w} \cdot n^r & \frac{\partial n^r}{\partial s_1^r} \cdot t_1^w & \frac{\partial n^r}{\partial s_2^r} \cdot t_1^w \\ \frac{\partial t_2^w}{\partial s_1^w} \cdot n^r & \frac{\partial t_2^w}{\partial s_2^w} \cdot n^r & \frac{\partial n^r}{\partial s_1^r} \cdot t_2^w & \frac{\partial n^r}{\partial s_2^r} \cdot t_2^w \end{bmatrix} \begin{bmatrix} \Delta s_1^w \\ \Delta s_2^w \\ \Delta s_1^r \\ \Delta s_2^r \end{bmatrix} = - \begin{bmatrix} t_1^w \cdot r^{wr} \\ t_2^w \cdot r^{wr} \\ t_1^w \cdot n^r \\ t_2^w \cdot n^r \end{bmatrix} \quad (4)$$

This formula (4) has an advantage that can distinguish contact from non-contact.

In the end, the normal force is calculated using the penetration. Also, position and velocity are required to calculate creepages and creepforces. Like

a wheel-rail, between two bodies in contact with moving occurs Rolling or Sliding, which are occurring when it have a different linear velocity and the angular velocity of two objects.

$$\begin{aligned} \xi_x &= \frac{(v^w - v^r)^T \cdot t_1^r}{V} \\ \xi_y &= \frac{(v^w - v^r)^T \cdot t_2^r}{V} \\ \varphi &= \frac{(w^w - w^r)^T \cdot n^r}{V} \end{aligned} \quad (5)$$

As shown at (5), and respectively express longitudinal creepage, lateral creepage and spin creepage.

The creep force is caused by creepage.

In this paper, as mentioned earlier, to calculate the creep force using polach nonlinear theory (6) was much faster to compute than the linear theory of kalker and it can be non-linear analysis.

$$\begin{aligned} F_x &= F_{px} \frac{\xi_x}{v_c} \\ F_y &= F_{px} \frac{\xi_y}{v_c} + F_{py} \frac{\varphi}{v_c} \end{aligned} \quad (6)$$

4 ANALYSIS OF DRIVING CHARACTERISTIC

Table 1. Condition of analysis.

| | |
|---------------------|------------|
| Mass of Wheel | 1800kg |
| Mass of body | 40000kg |
| Mass of | 4000kg |
| Curvature of rail | 0.002(1/m) |
| Velocity of vehicle | 18km/h |

In this paper, using the Table 1, an experiment was performed. Radius of track is 500m and wheel-set's velocity is 5 m/s. As a result of the experiment, the output is similar to the Adams-wheel/rail which is a commercial program.

5 CONCLUSION

As you can see from the following graphs, a simple analysis of the dynamic behavior of a railway vehicle wheel-set is also carried out. The red or dot line indicates left wheel. And the blue line indicates right wheel. The data from the wheel-rail contact model

provides the initial values to predict the dynamic behavior of a railway vehicle model. A bogie and car body was simply designed as linearized models based on dynamics. The dynamic performance of the simple car model was estimated by numerical analysis. Regarding the results of the model, the movement of the car body which was produced by the contact between the wheel and rail was appeared as a result. The multi-contact between the wheel and rail was also considered, so the effect of the contact was able to be performed more similarly as the actual features.

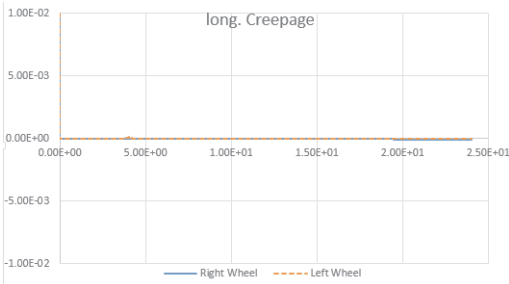


Figure 4. Creepage X.



Figure 5. Creepage Y.

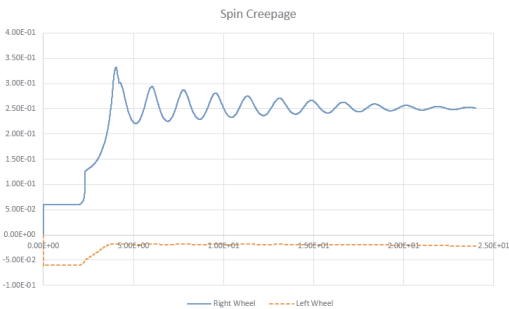


Figure 6. Creepage spin.

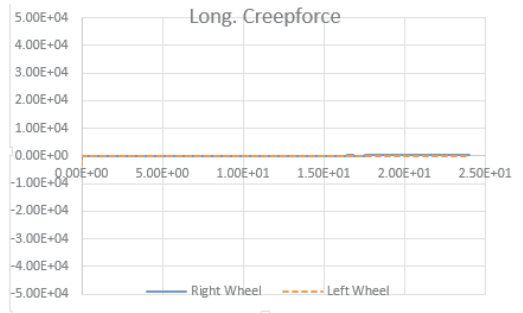


Figure 7. Creep force X.

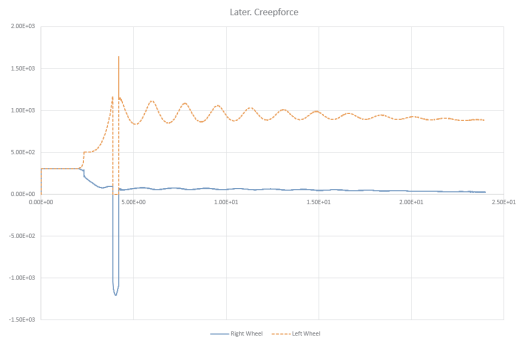


Figure 8. Creep force Y.

REFERENCES

- [1] J. Kalker, 1990. *Three Dimensional Elastic Bodies in Rolling Contact*.
- [2] O. Polach, Journal of Nonlinear Dynamics, vol. 45, "A Fast Wheel-Rail Forces Calculation Computer Code," pp. 109–130, 1999.
- [3] A. Shabana, 2007. *Railroad vehicle dynamics*, CRC Press.
- [4] J. Pombo, J. Ambrosio & M. Silva, 2007. Vehicle System Dynamics, *A new wheel-rail contact model for railway dynamics* 45(2):165–189.
- [5] E. Meli, M. Malvezzi, et al., December.2008. Vehicle Systems Dynamics, *A railway vehicle multibody model for real-time applications*, 46(12):1083–1105.
- [6] A. A. Shabana, K. E. Zaazaa, J. L. Escalona and J. R. Sany, Journal of Sound and Vibration, *Development of Elastic Force Model for Wheel/Rail Contact Problems*, vol.269, no. 1–2, pp. 295–325, 2004.

Tight frames and iterative shrinkage based measurement matrix optimization for compressive sensing

E. Oey

Department of Computer Engineering, Bina Nusantara University, Jakarta, Indonesia

ABSTRACT: The famous technique to sense the signal with a fewer number of samples compared to the traditional Shannon-Nyquist sampling theory is known as compressive sensing (CS). CS provides the joint sensing and compression by exploiting sparsity of the signal in the appropriate sparsifying dictionary. Measurement matrix is used to project the signal and provide a dimensional reduction of the signal. The performance of the CS is determined by the properties of equivalent dictionary that is the multiplication between measurement matrix and specifying dictionary. One of important properties is the averaged mutual coherence which is defined as the average of correlation between the column of the equivalent dictionary. The smaller averaged mutual coherence will provide a better performance of CS, thus optimizing the measurement matrix is how to minimize the averaged mutual coherence. This paper proposes an algorithm to optimize measurement matrix by reducing the averaged mutual coherence based on tight frame design. The results of numerical simulations show that our proposed method improves the performance of the CS system in terms of signal reconstruction accuracy and outperforms the previous algorithms for measurement matrix optimization.

KEYWORDS: Compressive sensing, measurement matrix, averaged mutual coherence, tight frame.

1 INTRODUCTION

Most signals are sparse, which means that it can be represented by a few coefficients by using an appropriate specifying dictionary. The sparsity of the signal gives the possibility to acquire a signal so that the obtained signal is in the already compressed form. Compressive sensing (CS) provides this possibility by projecting the signal into a measurement matrix and results dimensional reduction of the signal. Donoho is the one who introduces the term of Compressed Sensing [1], and alternative names are Compressive Sensing or Compressive Measurements. The important works of Donoho, Candès, Romberg and Tao provide the theoretical foundation of CS [1]–[4]. In the recent decade, CS has attracted a lot of attention from signal and image processing communities and some excellent reviews of CS and its applications have been published [5]–[6].

The measurement matrix in CS is also known as projection matrix or sensing matrix must be incoherent with the sparsifying dictionary in order that the signal can be reconstructed accurately [7]. Random matrices such as Gaussian, Bernoulli or random partial Fourier matrix satisfies this requirement because with high probability they are incoherent with many sparsifying dictionaries. However, the measurement matrix can further be optimized by minimizing the averaged mutual coherence of the equivalent dictionary. Several

methods have been proposed to reduce the averaged mutual coherence by reducing the off-diagonal's absolute values of the Gram matrix of equivalent dictionary such as shrinkage method [8], and alternating projection into the sets of matrix, which is well known, has incoherent properties [9] and make the Gram matrix close to identity matrix [10]–[11]. In this paper, based on a tight frame design we propose the improvement of measurement matrix optimization. The numerical experiments show that by using the optimized measurement matrix from our method, the signal reconstruction accuracy of CS system can be improved compared to the previous methods.

The paper is organized as follows. In the Section 2, we provide some preliminaries on CS and the measurement matrix optimization problems. In Section 3, we present our proposed algorithm in the measurement matrix optimization. Section 4 presents the results of numerical experiments to show that our proposed algorithm outperforms the previous methods. Finally, we conclude our paper in section 5.

2 RELATED WORKS

Consider that a signal vector $x \in \mathfrak{R}^{N \times 1}$ is linear projected into the measurement matrix $\Phi \in \mathfrak{R}^{M \times N}$:

$$y = \Phi x \quad (1)$$

resulted a measurement vector $y \in \mathfrak{R}^{M \times 1}$ with $M \ll N$. How to reconstruct accurately x from y is the task of CS. By using an *overcomplete dictionary* $\Psi \triangleq [\psi_1 \ \psi_2 \ \dots \ \psi_K] \in \mathfrak{R}^{M \times K}$ where $K > N$, we can represent x as a linear combination of columns ψ :

$$x = \sum_{i=1}^K \theta_i \psi_i \triangleq \Psi \theta \quad (2)$$

where $\theta \in \mathfrak{R}^{K \times 1}$ is a vector coefficients. CS requires that the signal x is sparse, which means $\|\theta\|_0 \ll N$, where $\|\theta\|_0$ denotes the number of non-zero elements in θ . The signal x is said S -sparse in ψ if $\|\theta\|_0 = S$. We can rewrite (1) by substituting (2) into (1) as :

$$y = \Phi \Psi \theta \triangleq D \theta \quad (3)$$

where the matrix $D = \Phi \Psi \in \mathfrak{R}^{M \times K}$ is called an *equivalent dictionary* in CS. The reconstructed signal \hat{x} can be recovered from y by solving the following constrained problem :

$$\min \|\hat{\theta}\|_0 \quad s.t. \quad y = D \hat{\theta} \text{ and } \hat{x} = \Psi \hat{\theta} \quad (4)$$

The *orthogonal matching pursuit* (OMP) [12] that is known as greedy pursuit algorithm can be used to obtain the approximated solution of (4). The *Mutual coherence* is a property of measurement matrix that is easy to evaluate and can be easily manipulated to improve the performance of CS systems [8],[13]-[15]. Let $D = [d_1 \ d_2 \ \dots \ d_K] \in \mathfrak{R}^{M \times K}$ and denote $g_{ij} \triangleq d_i^T d_j$ as the (i, j) th elements of the Gram matrix of D and $Z_c \triangleq \text{diag}(g_{11}^{-1/2} \ \dots \ g_{ii}^{-1/2} \ \dots \ g_{KK}^{-1/2})$. The normalized equivalent dictionary $\bar{D} \triangleq D Z_c$ and the normalized Gram matrix of D is $\bar{G} = \bar{D}^T \bar{D}$ such that $\bar{g}_{ii} = 1, \ \forall i$. The mutual coherence of D is defined :

$$\mu(D) = \max_{i \neq j} |\bar{g}_{ij}| \quad (5)$$

Any S -sparse signal x can be exactly recovered from compressive measurement $y = D \theta$ by using (4) [13] if:

$$S < \frac{1}{2} \left(1 + \frac{1}{\mu(D)} \right) \quad (6)$$

and for a matrix D of dimension $M \times K$, $\mu(D)$ is bounded with $\mu_B \leq \mu(D) \leq 1$ where :

$$\mu_B \triangleq \sqrt{\frac{K-M}{M(K-1)}} \quad (7)$$

is Welch bound [16]. However, (7) is just a worst-case bound and does not do justice to the actual behavior of the average signal recovery performance. A better measure was introduced by Elad [8] when optimizing the measurement matrix that is *t-averaged mutual coherence* $\mu_t(D)$:

$$\mu_t(D) := \frac{\sum_{i \neq j, 1 \leq i, j \leq K} \left(|\bar{g}_{ij}| \geq t \right) \cdot |\bar{g}_{ij}|}{\sum_{i \neq j, 1 \leq i, j \leq K} \left(|\bar{g}_{ij}| \geq t \right)} \quad (8)$$

where $0 \leq t < 1$ and indicator function $\left(|\bar{g}_{ij}| \geq t \right) = 1$ if the condition is true and otherwise is zero.

A finite frame $F \in \mathfrak{R}^{M \times K}$ in M -dimension Hilbert space $\in \mathfrak{R}^{M \times 1}$ is a sequence of $K > M$ vectors $\{f_k\}_{k=1}^K$, where $f_k \in \mathfrak{R}^{M \times 1}$ that satisfies a generalized Parseval condition [17] :

$$\alpha \|v\|_2^2 \leq \sum_{n=1}^K |\langle v, f_n \rangle| \leq \beta \|v\|_2^2, \text{ for all } v \in \mathfrak{R}^{M \times 1} \quad (9)$$

where α and β are two positive constants as the *lower bound* and *upper frame bound*. The frame F is said α -tight if in (9) $\alpha = \beta$. The α -tight frame must satisfy:

$$F_{\alpha-ENTF} F_{\alpha-ENTF}^T = \alpha I_M \quad (10)$$

The Frame F is called α -equal norm tight frame (ENTF) if it has equal norm column vectors $c_k \triangleq \|f_k\|_2 = e, \forall k$ and satisfies :

$$\alpha M = K e^2 \quad (11)$$

ENTF is called α -unit norm tight frame (UNTF) if $e = 1$ and $\alpha = \frac{K}{M}$ is redundancy of the frame.

Equiangular tight frame (ETF) is an α -UNTF that fulfills $|\langle f_k, f_l \rangle| = c, \ \forall k \neq l$. The optimal *Grassmanian frame* is an ETF that achieves $c = \mu_B$ [16]. Because ETF has a very nice averaged mutual coherence behavior, it becomes important in the measurement matrix optimization [9]. The averaged mutual coherence of the equivalent dictionary can be minimized by making the Gram matrix $G = D^T D = \Psi^T \Phi^T \Phi \Psi$ as close as possible to an identity matrix :

$$\min_{\Phi \in \mathfrak{R}^{M \times N}} \left\| I_K - \Psi^T \Phi^T \Phi \Psi \right\|_F^2 \quad (12)$$

where $\|\cdot\|_F$ denotes the *Frobenius* norm and I_K is an identity matrix of dimension K . The small averaged mutual coherence of the equivalent dictionary will improve the performance of the CS system to reconstruct the signal from compressive measurement. Duarte et al. [10] proposed the method to solve (12) and rewritten as :

$$\min_{\Phi \in \mathbb{R}^{M \times N}} \|\Lambda - \Lambda \Gamma^T \Gamma \Lambda\|_2^F \quad (13)$$

where $V\Lambda V^T$ is Eigen-decomposition of $\Psi\Psi^T$, and $\Gamma = \Phi V$. Abolghasemi et al. [11] solved (12) by using the gradient descent method.

3 PROPOSED MEASUREMENT MATRIX OPTIMIZATION

In this section, we propose an algorithm to make equivalent dictionary and achieve a smaller averaged mutual coherence than the previous methods. Denote $\bar{G}^{(k)} = \bar{D}^{T(k)}\bar{D}^{(k)}$ as the normalized Gram matrix of equivalent dictionary $\bar{D}^{(k)}$, and we first reduce the off-diagonal entries of $\bar{G}^{(k)}$ by using the shrinkage function [18] :

$$\bar{H}^{(k)}(i, j) = \begin{cases} 1 & i = j \\ \frac{4}{\pi} \gamma \tan^{-1}(\bar{G}^{(k)}(i, j)) & i \neq j \end{cases} \quad (14)$$

Then, we find the equivalent dictionary $\hat{D}^{(k)}$ of $\bar{H}^{(k)}$ by using SVD such that $\bar{H}^{(k)} = \hat{D}^{T(k)}\hat{D}^{(k)}$ and compute $\tilde{G}^{(k)}$ by using :

$$\tilde{G}^{(k)} = \bar{H}^{(k)}\hat{D}^{+(k)}\hat{D}^{(k)} \quad (15)$$

where M^+ denotes a pseudo inverse of matrix M . By using SVD, we find the equivalent dictionary $\tilde{D}^{(k)}$ of $\tilde{G}^{(k)}$ so that $\tilde{G}^{(k)} = \tilde{D}^{T(k)}\tilde{D}^{(k)}$, then we find a nearest α -tight frame $\tilde{D}^{(k)}$ to $\tilde{D}^{(k)}$ with respect to the Frobenius norm [19] :

$$\tilde{D}^{(k)} = \frac{K}{M} (\tilde{D}^{(k)}\tilde{D}^{T(k)})^{-1/2} \tilde{D}^{(k)} \quad (16)$$

Compute the normalized Gram matrix $\bar{\bar{G}}^{(k)} = \bar{\bar{D}}^{T(k)}\bar{\bar{D}}^{(k)}$, and we find ETF Gram $G_{ETF}^{(k)}$ that close to $\bar{\bar{G}}^{(k)}$ by using :

$$G_{ETF}^{(k)}(i, j) = \begin{cases} 1, & i = j \\ \bar{\bar{G}}^{(k)}(i, j), & |\bar{\bar{G}}^{(k)}(i, j)| \leq \xi \\ \text{sign}(\bar{\bar{G}}^{(k)}(i, j)) \xi & \text{otherwise} \end{cases} \quad (17)$$

We need to find a point between the $G_{ETF}^{(k)}$ and $\bar{\bar{G}}^{(k)}$ to get $G_{opt}^{(k)}$ by using :

$$G_{opt}^{(k)} = \beta G_{ETF}^{(k)} + (1 - \beta) \bar{\bar{G}}^{(k)} \quad (18)$$

By using SVD again, we find the equivalent dictionary $D_{opt}^{(k)}$ of $G_{opt}^{(k)}$ so that $G_{opt}^{(k)} = D_{opt}^{T(k)}D_{opt}^{(k)}$, and finally we can obtain an optimum measurement matrix :

$$\Phi_{opt}^{(k)} = D_{opt}^{(k)}\Psi^+ \quad (19)$$

Based on the discussion above, we summarize our proposed method for finding the optimized measurement matrix in an iterative algorithm below :

Initialization: Given sparsifying dictionary Ψ , initial measurement matrix $\Phi^{(0)}$ and iteration number R_{iter} .

- *Step I :* While $1 \leq k \leq R_{iter}$, compute $\bar{G}^{(k)}$.
- *Step II :* Find $G_{opt}^{(k)}$ by using (14) until (18).
- *Step III :* Find $\Phi_{opt}^{(k)}$ by using (19).
- *Step IV :* End while if $k > R_{iter}$.

4 EXPERIMENTAL RESULTS

In this section, we present some experiments to compare the performance of our proposed measurement matrix optimization in the CS systems of the previous algorithms. We generate a $N \times K$ sparsifying dictionary Ψ with zero mean and unit variance Gaussian distributed entries. We generate a measurement matrix $\Phi^{(0)}$ with zero mean and variance $= \frac{1}{\sqrt{M}}$ Gaussian distributed entries as random measurement matrix. This measurement matrix Φ_0 is also used as initial conditions for measurement matrix optimization. For convenience, we denote the measurement matrix optimization algorithms with *Elad* [8], *XPC* [9], *DS* [10], *AFMS* [11], and *EN* for our proposed algorithm. We use $t = 25\%$ in *Elad*, which means it targets the top 25% of the absolute of diagonal entries in the normalized Gram matrix, and $\eta = 0.01$ is used in *AFMS* while $\beta = 0.20$ is used in *XPC* and *EN*. $R_{iter} = 1000$ is used in *Elad*, *XPC*, *AFMS*, and *EN*, and $\gamma = 0.90$ is used in *Elad* and *EN*. $\xi = \mu_B$ is used in *XPC*, while in *EN*, we use time varying ξ , starting with $\xi = 1.2\mu_B$ and ending with $\xi = 0.8\mu_B$ so we have averaged $\xi = \mu_B$. Averaged mutual coherence in (8) is computed with $t = \mu_B$ for all methods except *EN*, and in *EN* we used time varying ξ for t .

For generating a $N \times K$ in sparsifying dictionary Ψ , we generate a set of 1000 S -sparse $K \times 1$ vectors $\{\theta_l\}$ ($l = 1, 2, \dots, 1000$) in which each non-zero elements of θ_l is randomly positioned with an Gaussian distribution of *iid.* zero-mean and unit variance. The 1000 test signal vectors $\{x_l\}$ is created by using (2) : $x_l = \Psi\theta_l, \forall l$. The 1000 measurements $\{y_l\}$ are obtained by using (1) : $y_l = \Phi x_l, \forall l$, where Φ is an $M \times N$ matrix taken from six different measurement matrices yielded in the previous discussion above. The reconstructed signal \hat{x}_l is obtained by using $\hat{x}_l = \Psi\hat{\theta}_l$, where $\hat{\theta}_l$ is the solution of (4) by using OMP. The *relative error figure*, denoted as e_f was used to measure the signal reconstruction performance of a CS system :

$$e_f \triangleq \frac{1}{N_s} \sum_{l=1}^{N_s} \frac{\|x_l - \Psi\hat{\theta}_l\|_2^2}{\|x_l\|_2^2} \triangleq \frac{1}{N_s} \sum_{l=1}^{N_s} e_l \quad (20)$$

with N_s the total number of testing signals where $N_s = 1000$ in our experiments.

For $N = 80, K = 120, S = 4$: **Figure 1** shows the signal reconstruction accuracy of the CS systems by using different measurement matrices for the measurement dimension M varying from 12 to 36. **Figure 2** shows the behavior of $\mu_l(D)$ by the six different CS systems for M varying from 12 to 36.

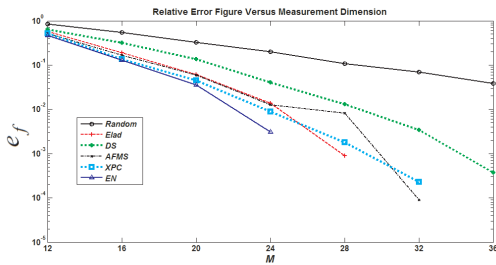


Figure 1. Relative error figure e_f versus measurement dimension M , for $N=80, K = 120, S = 4$.

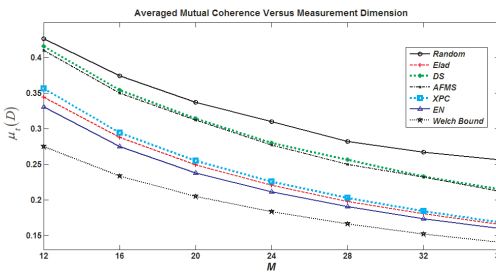


Figure 2. The behavior of the averaged mutual coherence $\mu_l(D)$ versus measurement dimension M , for $N = 80, K = 120, S = 4$.

For $N=80, K = 120, M = 25$: **Figure 3** shows the signal reconstruction accuracy of the CS systems by using different measurement matrices for signal sparsity S varying from 1 to 7. **Figure 4** shows the behavior of $\mu_l(D)$ by the six different CS systems for S varying from 1 to 7.

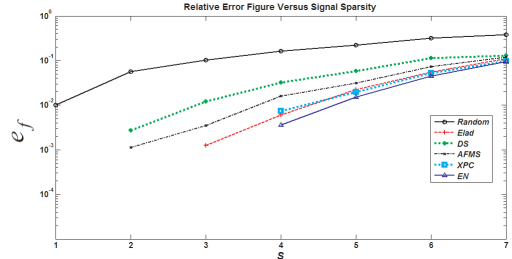


Figure 3. Relative error figure e_f versus signal sparsity S , for $N=80, K = 120, M = 25$.

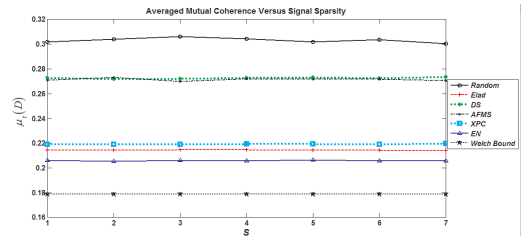


Figure 4. The behaviour of the averaged mutual coherence $\mu_l(D)$ versus versus signal sparsity S , for $N=80, K = 120, M = 25$.

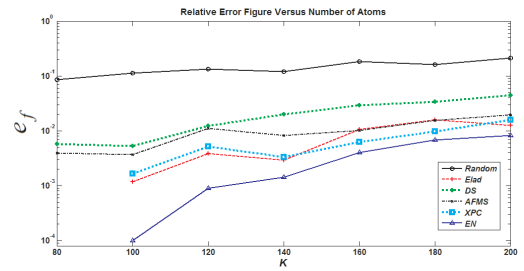


Figure 5. Relative error figure e_f versus number of columns (atoms) K , for $N = 80, K = 30, S = 5$.

For $N=80, K = 30, M = 5$: **Figure 5** shows the signal reconstruction accuracy of the CS systems by using different measurement matrices for the number of columns (atoms) K varying from 80 to 200. **Figure 6** shows the behavior of $\mu_l(D)$ by the six different CS systems for K varying from 80 to 200.

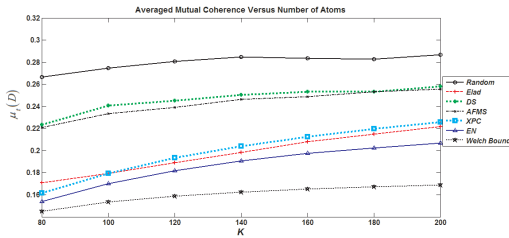


Figure 6. The behaviour of the averaged mutual coherence $\mu_1(D)$ versus number of columns (atoms) K , for $N = 80$, $K = 30$, $M = 5$.

As is seen from the results, all of the measurement matrix optimization algorithms provide a smaller averaged mutual coherence than random matrix and improve the signal reconstruction performance. Our proposed algorithm (*EN*) provides a smaller averaged mutual coherence compare to all of the previous algorithms and outperforms the others in terms of signal reconstruction accuracy.

5 CONCLUSIONS

In this paper, we have presented the algorithm for measurement matrix optimization based on tight frame design. The experiment results show that our proposed algorithm outperforms the previous algorithms in terms of signal reconstruction accuracy. Further improvements can be attempted in future work such as employing the optimized measurement matrix for natural image rather than just synthesizing signal, the improvement in ETF construction for measurement matrix optimization and joint optimization of sparsifying dictionary and measurement matrix.

REFERENCES

- [1] David L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] Emmanuel J. Candès, Justin Romberg, and Terence Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [3] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.
- [4] Emmanuel J. Candès and Terence Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

- [5] T. Strohmer, "Measure what should be measured: Progress and challenges in compressive sensing," *IEEE Signal Process. Lett.*, 2012.
- [6] Saad Qaisar, Rana Muhammad Bilal, Wafa Iqbal, Muqaddas Naureen and Sungyoung Lee, "Compressive sensing: From theory to applications, a survey," *Journal of Communications and Networks.*, vol. 15, no. 5, pp. 443–456, Oct.2013.
- [7] E. Candès and J. Romberg, Sparsity and incoherence in compressive sampling, *Inverse Prob.*, vol. 23, no. 3, pp. 969–985, 2007.
- [8] M. Elad, "Optimized projections for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
- [9] Jianping Xu, Yiming Pi and Zongjie Cao, "Optimized Projection Matrix for Compressive Sensing," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 560349, 8 pages, doi:10.1155/2010/560349, 2010.
- [10] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395–1408, July 2009.
- [11] Vahid Abolghasemi, Saideh Ferdowsi, Bahador Makkiabadi and Saeid Sanei, "On Optimization of The Measurement Matrix For Compressive Sensing," in *Proc. 18th European Signal Processing Conference, EUSIPCO 2010*, August 23–27, 2010.
- [12] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [13] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthonormal) dictionaries via minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
- [14] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [15] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.
- [16] T. Strohmer and R.W. Heath, "Grassmannian frames with applications to coding and communication," *Appl. Comp. Harmon. Anal.*, vol. 14, no. 3, pp. 257–275, May 2003.
- [17] Kovacevic, J. and Chebira, A., "Life Beyond Bases: The Advent of Frames (Part I)," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 86–104, July 2007.
- [18] Evaggelia V. Tsiliagianni, Lisimachos P. Kondi and Aggelos K. Katsaggelos, "Construction of Incoherent Unit Norm Tight Frames With Application to Compressed Sensing," *IEEE Trans. Inf. Theory.*, vol. 60, no. 4, pp. 2319–2330, April 2014.
- [19] J. Tropp, I. S. Dhillon, R. W. Heath, Jr., and T. Strohmer, "Designing structured tight frame via alternating projection," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, 2005.

Amplitude control of the flow tube based on PID control

J. Yu & Z.B. Gui

Beijing University of Posts and Telecommunication, China

ABSTRACT: Based on the measurement requirements of fast, stable, zero steady-state error performance in the metering filling machine process, we study on Coriolis flowmeters, analyze the formation of the flow tube self-excited vibrations, and propose a simple, practical PID control algorithm. Compared with traditional complex algorithms, PID control algorithm improves the accuracy of the flow tube amplitude control, and is simpler and more applicable. It also provides a theoretical basis for the design of other types of flow meters, and a guarantee to the CNG filling machine accurate measurement.

1 FLOW METERS USED IN FILLING MACHINE

1.1 The basic principle of the Coriolis flow meter

Coriolis flow meters are flow measurement instruments which are designed based on the principle of the Coriolis force. Allowing fluid via a rotating or vibrating measuring tube, the fluid flowing in the pipe which is equivalent to the linear motion, rotation or oscillation of the measuring tube will produce an angular rate, since rotation or vibration is driven by the electromagnetic fields, it has a fixed frequency, and thus the fluid in the pipeline by the Coriolis force is only related to its mass and velocity, and therefore by measuring the Coriolis force, mass flow rate can be measured.

In CNG filling machine development process, we are using the Micro Motion CNG050. When the fluid passes through, U-shaped tube Importers subject a pair of couple equal in magnitude and opposite in direction, the angular twists affected by the Coriolis force, the twist angle is α . Coriolis force has the following mathematical relationship between the flow tube mass flow:

Let the size of the Coriolis force F , the fluid within the flow tube mass is m , the measuring tube vibrating angular velocity ω , the measuring tube to the fluid flow rate V .

According to Newton's second law:

$$F = ma, \quad a = 2\omega V \quad (1)$$

Measuring tube is suffered by the torque:

$$M = F_1 \times R_1 + F_2 \times R_2 \quad (2)$$

According to [equation \(1\) \(2\)](#) :

$$2M = 2F \times R = 4M\omega r \quad (3)$$

According to the flow tube's parameters we can know that, set the flow tube length L , when the time is t , the flow through the tube at a flow rate $V = L/t$, the mass flow flowing through the flow tube is $V_m = m/t$. So the [equation \(3\)](#) is equal to:

$$M = 4\omega r v_m L \quad (4)$$

Due to the impact of the flow tube's metal material, it is deformed due to the rigidity, when the twist angle is α and stiffness is K_s , torque formed by deformation of the flow tube is:

$$T = K_s \alpha \quad (5)$$

According to [equation \(4\)](#) and [\(5\)](#), we can know that:

$$4\omega r v_m L = K_s \alpha, \quad \text{so } v_m = K_s \alpha / 4\omega r L \quad (6)$$

Assuming the speed of the midpoint of the tube vibration is vt , $vt = \omega L$, so

$$\sin \alpha = \omega L \Delta t / 2r \quad (7)$$

When the value of α is small, $\alpha = \sin \alpha$, according to [equation \(6\) \(7\)](#), we can know

$$v_m = K_s \Delta t / 8r^2 \quad (8)$$

1.2 Coriolis flow meters basic structure

Coriolis flow meter is composed of two parts: sensors and transmitters. The sensor portion is used for detecting a flow rate signal, it is mainly composed by the shunt, the measuring tube, the drive, the drive coil, the detection and the detection magnet. The transmitter is mainly used for the conversion, operation, flow display and signal output of sensor-driven and flow tube detection signal, and the transmitter is mainly composed of power supply, drives, detect, display and other parts of the circuit components.

2 STUDY ON FLOW TUBE SELF-EXCITED VIBRATION SYSTEM

2.1 Flow tube self-excited vibration system

Coriolis flow meter consists of two parts: sensors and transmitter, transmitters include the signal processing circuit and driver circuit. In the Coriolis flow meter, driving portion of the sensor portion and the transmitter portion form a self-excited oscillation circuit system. And this self-excited system is the core to control the magnitude of the flow tube.

2.2 Establishment mathematical model of the sensor section

In the research process of self-excited system, a mathematical model of the sensor part of the drive system is a prerequisite for the study.

Coriolis flow meter sensor section includes flow tube, electromagnetic exciter and speed sensors. We can use a simplified one-dimensional damped and forced vibration system to describe its vibration system, the vibration differential equation is:

$$m\ddot{x}(t) + c\dot{x}(t) + k_s x(t) = f(t) \quad (9)$$

We use m , c , k , $x(t)$, $f(t)$ to indicate in the flow tube mass, damping, stiffness, and the driving force of the displacement of the flow tubes.

The size of the attraction generated by the electromagnetic exciter is:

$$f(t) = B_1 l_1 i(t) \quad (10)$$

Wherein B_1 , l_1 , $i(t)$ are the magnetic induction, the coil length and the input current.

The output voltage magnitude generated by the speed sensor is:

$$v(t) = B_2 l_2 \dot{x}(t) / t = B_2 l_2 \dot{x}(t) \quad (11)$$

Wherein B_2 , l_2 , $x(t)$ are the strength of the magnetic induction, the coil length and the vibration displacement.

According to (4) (5) (6), we can get the relationship between the speed sensor output voltage $v(t)$ and current $i(t)$, and according to Laplace transformation:

$$\frac{v(s)}{i(s)} = \frac{Ks}{s^2 + 2\zeta_s \omega_n s + \omega_n^2} \quad (12)$$

Wherein $K=B_1 l_1 B_2 l_2 / m$, $\omega_n = \sqrt{k_s / m}$, $\zeta_s = c / (2 \cdot \sqrt{k_s \cdot m})$, is damping ratio, ω_n is natural angular frequency.

2.3 Analysis on flow tube self-excited vibration

Combined with the equations above, we can obtain the differential equations describe the flow tube vibration:

$$\ddot{x}(t) + 2\zeta_s \omega_n \dot{x}(t) + \omega_n^2 x(t) = \frac{f(t)}{m} \quad (13)$$

The transfer function of the flow tube:

$$H(s) = \frac{X(s)}{F(s)/m} = \frac{1}{s^2 + 2\zeta_s \omega_n s + \omega_n^2} \quad (14)$$

Based on the actual situation, filling machine is mainly to transport gas and flow, the flow tube's damping coefficient is small and essentially unchanged.

Meanwhile, the differences between the different inherent angular frequency meter are obvious. In order to simplify research, we take $\zeta_s = 0.00004$, $\omega_n = 140 \text{ rad/s}$, so the equation (14) is changed to:

$$H(s) = \frac{1}{s^2 + 0.03519s + 193444.24} \quad (15)$$

According to the formula, we can make its frequency response bode diagram:

As it can be seen from the upper part of Fig 1, the correlation between the excitation frequency and the natural frequency, so as to constitute a self-excited system, it needs to satisfy $|H(j\omega)F(j\omega)|=1$ and $\gamma_n + \gamma_f = n\pi$ conditions.

From the phase-frequency characteristics of the flow tube, when the excitation frequency is close to the natural frequency, that is when the phase displacement of the flow tube lags reaches a stable vibration, since the displacement of the flow tube lags, in the system configuration, it needs the

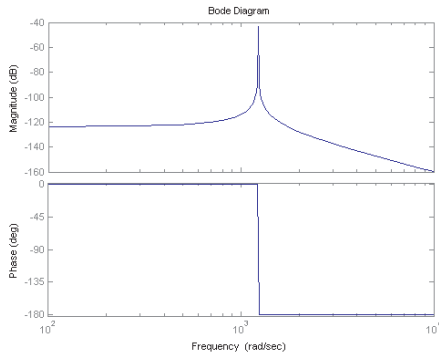


Figure 1. Flow tube's frequency characteristics.

feedback link with -90° phase shift, which makes the feedback signal in phase with the driving signal. Therefore, the feedback signal uses the speed signal, so that it reaches the phase equilibrium condition. In order to minimize the energy consumption, the excitation frequency of the driving force should be closer to the natural frequency of the flow tube. Therefore, in the control system design process, we can consider the speed signal as a feedback signal, so that the feedback signal and the drive signal in the same phase, thus the phase equilibrium conditions in self-excitation system could be reached.

3 STUDY ON AMPLITUDE CONTROL ALGORITHM IN FLOW TUBE

The above analyses use the speed signal as a feedback signal, so that the feedback signal and the drive signal are in the same phase, which forms a relatively stable self-excited vibration system. In industrial control, the most widely used regulator is proportional, integral, differential controller, also known as PID controller. Since its simple structure, good stability, reliable, easy to adjust, PID controller has become one of the main industrial control technologies. In general, this paper optimizes control of the flow tube vibration system based on a PID control algorithm, thereby to enhance the flow tube vibration start-up speed.

3.1 The basic principle of the amplitude control algorithm

To start the electromagnetic vibration exciter, we need to provide a drive signal, which comes from the output signal of the speed sensor. In order to achieve precise control of the amplitude in the control process, when the system achieves the steady-state, we have to accelerate the speed control.

Through the analyses above, the flow tube self-excitation system structure is shown in Figure 2:

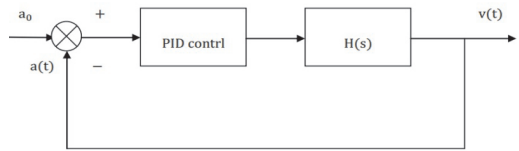


Figure 2. Amplitude control structure of the flow tube.

3.2 Study on amplitude control algorithm

According to the above analyses, we can obtain the transfer function of the flow tube:

$$H(s) = \frac{1}{s^2 + 0.03519s + 1934444.24}$$

Through the schematic diagram of the flow tube self-excited system, we use simulation software to get its step response curves.

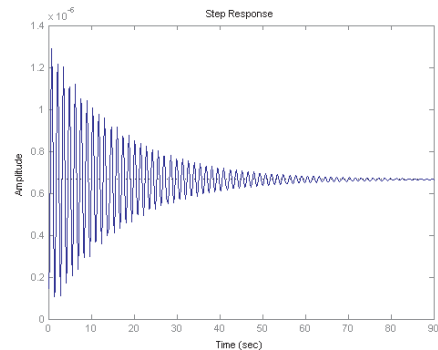


Figure 3. Diagram of flow control step response.

On the analysis of Figure 3, the system can reach a final steady state, but the time to reach steady-state response is too long, and the error is in a large number. For high accuracy flow meter, a device for fast response, simple feedback obviously cannot reach precise measurement requirements. In industrial automation equipment, we often use group of electrical or pneumatic units combined correction device, which can use proportional (P) unit K_p , differential (D) unit $K_d s$ and integral (I) unit $(T_i s)^{-1}$ to make PD, PI and PID three kinds of corrector.

3.3 Controller selection and error analysis

The effect of the modulating law is: deviation is able to be adjusted in time, fast response, and timely control. The integral regulating rule is: as long as deviation exists, the regulator will exist until the deviation is 0, so it can eliminate static bias, but if the integral action is too strong, the regulation will also be too strong, causing parameters overshoot, or even oscillation. The differential regulation, rule is: adjusting according to changes in the speed deviation, greatly reducing the amount of deviation and dynamic regulation time within a certain range of the system. Therefore, according to the requirement above, we can select the PI controller that both control timely and can eliminate the residual error or PID controller that can improve the start-up speed.

Via analyzing Figure 8 we know that, when we add PI control structure, the system needs to be able to reach our final system requirements for zero steady-state error, and eliminate the overshoot, because the effect of the proportional and integral regulation. But according to the actual measurement requirements, aerated measure needs rapid response and can reach a steady state. If the response time is too long, not only energy will be wasted, but also the gas stations cannot afford the loss. Therefore, according to the PID control theory, we can choose right differential time, eliminate overshoot, and reduce adjustment time.

We use simulation software to get the step response that without PI control structure, step response that add PI control structure, and step response that add PID control structure.

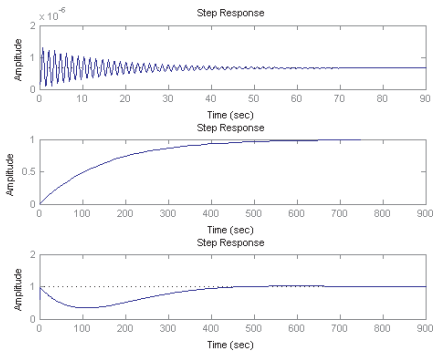


Figure 4. Simulation diagram with the same PI control coefficient.

We analyze step response of PI controller and a PID controller which is shown in figure 4, and without changing the ratio coefficient and the integral coefficient, we could add differential control aspects. Thus we can maintain a zero steady-state error

brought by PI controller, eliminate the overshoot, and reduce the system response time. But according to figure 4, we can find that, although add differential part could optimize the system performance, due to restrictions on all aspects of parameter selection, the response time is negligible even we do not maintain zero steady-state error and eliminate overshoot.

3.4 PID controller parameter tuning

Through the above analysis, we know the transfer function, and through comparing the simulation software results, we find that PID control is a relatively effective and simple control method. Therefore, we use trial and error method in tuning all coefficients. According to the above steps, we tune the coefficients of the PID controller in figure 4: firstly, we maintain the proportion coefficients and differentiation coefficients, and next we reduce the integration coefficients, and then we increase the different coefficients, finally we can get two different results:

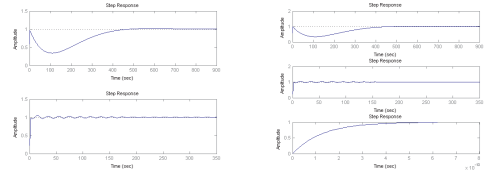


Figure 5. Decrease integration coefficients and increase differential coefficients.

Via analysis on the left part of figure 5, we find that after reducing the integration coefficients, the system response time reduces obviously, and achieves the required response time. But in the time up to about 190 seconds, the system still has a certain level of oscillation. And then we increase the differentiation coefficients, the result are shown in the right part of Figure 5. Both the oscillation and the response time is reduced. The system achieves stability in a very short time, the overshoot is eliminated and zero steady state error can meet the system requirements, through tuning PID controller's parameters reasonably, we can make the flow meter to meet the accurate measurement requirements.

4 CONCLUSION

The Coriolis mass flow meter is widely used in measurement due to its own advantages. On the basis of relevant research studies, we model the flow tube controller, and use Matlab to simulate the step response of the PI and PID controller, and finally we tune the parameters

to optimize the controller's performance. And get the PID controller that meets the system requirement. This paper gives a control method for simulation in the flow tube amplitude, and it is also a common, simple, practical control method. This will provide follow-up study a theoretical basis for the semi and total digital control methods in the tube amplitude control.

REFERENCES

- [1] Shi Er, Yi Xiaoqing, Chen Jinhong. The Application of Coriolis Flowmeter in Gas Fueling Station for Metering Compressed Natural Gas. *GuangzhouChemical industry*,2009,37(2):36–39
- [2] Dong Hongchao, Chen Liang. Coriolis Mass Flowmeter and its Application.*Chemical Defence On Ships*,2008,4:44–47
- [3] M.P.Henry, Michael Tombs, Mihaela Duta,etc..Two-phase flow metering of heavy oil using a Coriolis mass flow meter:A case study[J].*Flow Measurement and Instrumentation*.2005, 17:399–413
- [4] Li Xianggang.Researches on Modeling of Primary Meter and Amplitude Control Method of Flowtube for Coriolis Mass Flowmeter[D].HeFei:HeFei University of Technology,2009.4
- [5] Xu Kejun,Zhang Han.A study on the Digital Signal Processing and Driving Method for Corioils Mass Flowmeter.*Acta Metrologica Sinica*.2004,25(4):339–343,379
- [6] Mayela Zamora and Manus P.Henry.An FPGA Implementation of a Digital Coriolis Mass Flow Metering Drive System.[J].*IEEE Trans.Ind. Electron.*,vol.55,no.7,pp.2820–2831,July.2008.
- [7] N.Mole,G.Bobovnik,J.Kutin,B.Stok,I.Bajsic. An improved three dimensional coupled fluid-structure model for Coriolis flowmeters[J].*Journal of Fluids and Structures*,2008,24:559–575.
- [8] Xu Kejun,Yu Xincui,Su Jianhui etc. Development of excitation circuit for Coriolis mass flowmeters.Journal of Hefei University of Technology[J].2002,23(1):37–40[2]
- [9] Tian Shuhang, Jiang Lijuan. Research on the fuzzy PID controller with parameters self-tuning. *ELECTRIC DRIVE AUTOMATION*.Vol.25, No.6, 2003, 25(6):28–30
- [10] Han Banghua. The Research of PID controller parameter tuning methods and applid:[D].Hangzhou: Zhejiang University,2009:1–3

The noise reduction by the application of the vibration isolators

A.A. Igolkin, L.V. Rodionov & F.S. Pomatilov

Samara state aerospace university, Samara, Russia

ABSTRACT: In the living quarters, located above the bakery, the noise levels have been measured and compared to the sanitary standards. The bakery equipment has been the source of the noise. The main types of the noise sources and paths of propagation in the building have been considered. To reduce the noise in the apartments some arrangements have been offered. Vibrator isolators, which provide installation permissible sound pressure levels, have been designed and manufactured.

1 INTRODUCTION

1.1 History

In Russia (earlier in the USSR), unlike the Western countries, in the construction of residential, industrial and office buildings, soundproofing issues have never been neglected. Although the noise is a harmful factor of the human's life, the costs of acoustic comfort have been seemed excessive.

The overall noise level in apartments significantly increases every year. The norm of the sound insulation between flats' walls and ceilings established by the current national standard SNiP- $R_w = 50$ dB becomes more and more inadequate.

Moreover, the poor quality of construction does not always meet the regulatory indicators. In Figure 1 the results of the measurement and calculation of the index of airborne sound insulation in one of the new buildings are presented. The figure shows that the soundproofing partitions are considerably non-compliance with the building norms.

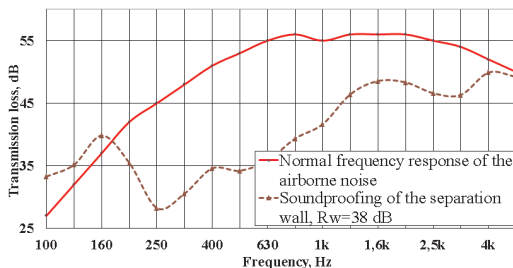


Figure 1. The airborne sound insulation frequency response for separating wall.

The index requirements increase to $R_w = 60$ dB for the airborne sound insulation between the rooms on the floors of the apartment and the rooms located underneath (for example, restaurants, cafes, catering, etc.), if under the living quarters there are sources of the increased noise (process equipment, refrigerators and ventilation system) (Igolkin, A. et al. 2012).

1.2 The problem

The measurements whose results revealed the excess of the existing standards were held in the flat above the bakery (Fig. 2). Because of the process equipment the daily norm was reduced by 5 dB below the permissible sound pressure level for the living quarters

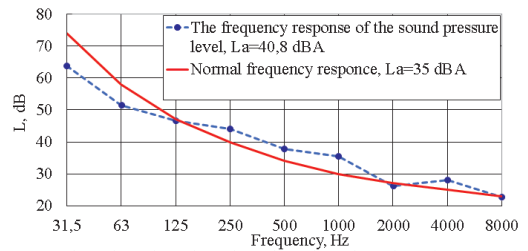


Figure 2. The frequency response of the sound pressure level in the residential area.

The process equipment was the source of the noise (kneading machine and cream machine). However, the preliminary research showing the main source was the kneading machine (Fig. 3).



Figure 3. Kneading machine.

2 THEORY

2.1 Types of the noise

Consider the main types of the noise and how it spreads in the building (Figure. 4). If the source of the noise is not connected to the constructions (for example, a loudspeaker: the transmission of the sound energy is a result of fluctuations in the structure separating two rooms) this noise is called the air-born noise (Fig. 4, pos. 1). The noise is called the structure-born noise when the “impacts” on the overlap is taking place (walking, dancing) and the energy transfer occurs via wall fluctuations (Fig. 4, pos. 2).

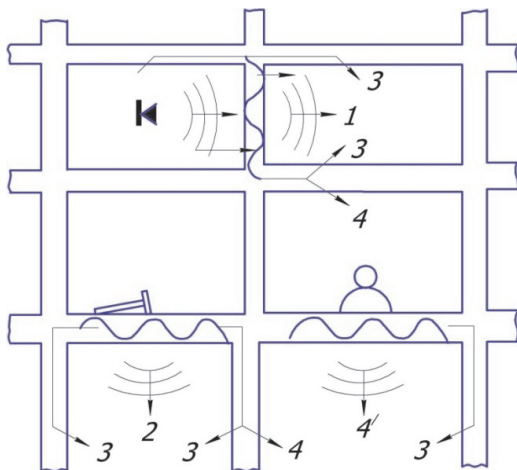


Figure 4. Diagram of noise propagation in buildings: 1 air-borne noise; 2- impact noise; 3,4 - structure-borne noise.

2.2 The noise transmission ways

The transmission ways of the noise in the room can be direct (1 and 2) and indirect (roundabout) (3 and 4).

This transfer is possible because the vibrations caused by the airborne or the structure-born noise, spread through the entire building's walls. Vibrating structures produce the noise on the premises located at a considerable distance from the source. This noise is often called the structure-born noise. The noise emitted by the construction tightly connected with any vibrating machinery (pump, ventilation or elevator) (4') is also the structure-borne noise. In this case, the path (3) will be indirect with respect to the direct (4').

In the propagation of the noise through the building indirect channel intensity is decreasing due to absorption of the material structure's vibration energy, the energy loss in the joints and energy distribution on the larger area walling. There is a more intense transmission of the noise through indirect ways in modern buildings by reducing the weight of walling, increasing the rigidity of interfaces in joints and reducing their numbers. As a result, the noise propagates with low long distance attenuation from the source, this leads to uncomfortable conditions in a number of premises, which is typical for many modern buildings made of large-scale structures.

2.3 The theory of the noise propagation

Fluctuations of the various technical devices in residential and industrial buildings are spreading through the buildings and structures in the form of structure-born noise and vibrations. Structure-born noise distribution depends on the physical properties of construction materials, structural parameters and connecting nodes. The transfer structure-born noise from the one building element to another may be characterized by the factor (Vér, I.L. & Beranek, L.L., 2006):

$$H(f) = \frac{4z_1z_2}{2\sqrt{l} \cdot z_1z_2 + z_1^2 + z_2^2}, \quad (1)$$

Where z_1/z_2 is the ratio of the mechanical impedances of the first and the second construction elements.

The structure-born noise is distributed constructions of the residential and industrial buildings. The magnitude of the noise level depends on the intensity of the noise sources and on the mechanical impedances of the structural elements of buildings and structures.

From the equation (1) follows the greater difference between the values of the mechanical impedances of the structures, then less the transfer of sound energy from the one component to another. The attenuation of the sound transmission structure can be calculated by the equation:

$$\Delta L = 10 \lg \frac{l}{H(f)}, \quad (2)$$

Where “ $H(f)$ ” is a transfer coefficient.

A 3 dB reduction of sound transmission occurs when the angular connection of the construction elements is made from the same material and has the same cross section, despite the impedances difference. If the second element is twice the thickness of the first, the reduction of the sound transmission reaches a 6 dB value. Five times the cross-sectional changes linearly arranged construction elements gives a structurally sound attenuation of the transmitted only by 3 dB.

3 THE NOISE REDUCING

3.1 The techniques of the noise reducing

For the noise reducing in industrial and residential areas engineers use various methods and tools. These methods are selected in accordance with the origin and nature of the noise requirements for acceptable noise levels, technical information and cost means to reduce the noise and the other factors (Igolkin, A.A. et al. 2008).

One of them is associated with the implementation of measures to reduce noise at its own sources. The method of the noise reducing at the source of its occurrence is to improve the structure of the source and the process changing. The most effective using of this method is in the development phase of the new equipment. When the noise source is the speaker it is not necessary to reduce their acoustic quality because their intended purpose is to convert the electrical energy into the sound.

The methods of the noise decreasing on its transmission ways received the greatest development. These methods include soundproofing (sound insulation).

- There are the following methods of sound insulation:
- Thickening of the external walls of the buildings;
- Double glazing windows application;
- Hollow glass balconies, double doors;
- Windows, doors and aperture seal;
- Walls with multilayer acoustic panels.

Nowadays, the multi-layer panel (“sandwich”) are often used as the acoustic panels. In the multilayer modular acoustic panels sound absorbing materials are between two sheets of soundproofing. For rigid elements used steel, plasterboard, gypsum, lead or duralumin profiled sheets, chipboard and other materials. As the sound absorbing materials can be applied in foams, pressed straw, porous polymeric materials, mineral wool and fiberglass. As damping material is used in bitumen compounds, as sealing - rubber putty, plastic, elastomeric materials. Domestic construction material losses increase its soundproofing.

The elementary layer changing of the parameters and their order gives the opportunity to adjust the character of the frequency dependence of the laminated construction sound insulation. Thus, to reduce the noise emitted by the process equipment in a residential apartment offers the following activities:

- 1 Vibration isolation of the noise sources by the installation of technological equipment on the vibration isolators (hereinafter VI).
- 2 Separation of the floor and walls and installation of “the floating floors”.
- 3 Increased the sound insulation of the walls.
- 4 Increased the ceiling insulation.

3.2 The vibration isolator

The proposed activities should be carried out consistently with the collection of measurements after each item. The technological equipment installation on the VI was the first step selected because of the low cost of their production and the convenience implementation of this action.

Suggested VI has the following structure (Fig. 5) (Igolkin, A.A. et al. 2011).

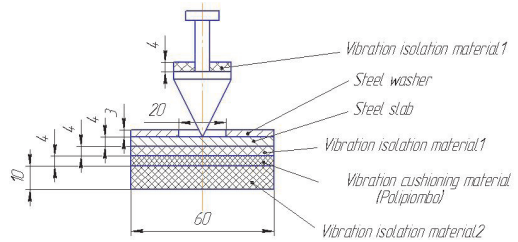


Figure 5. Schematic of the vibrator isolator.

3.3 Results

After the introduction of the first action item (installation of process equipment on the VI (Fig. 6)) held interim measure for evaluating the effectiveness of the action item (Fig. 7).

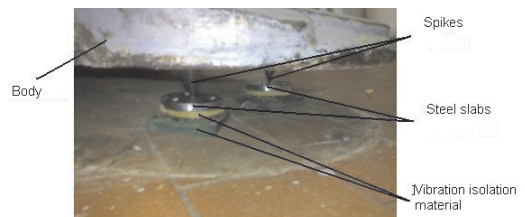


Figure 6. Technological equipment on vibration isolators.

The graph shows that the vibration on the floor and the wall around the unit after applying VI decreased, and on the unit increased. It happens because the

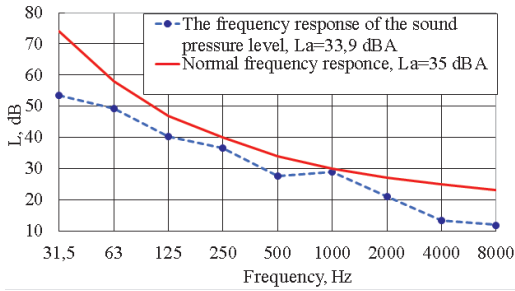


Figure 7. Frequency response of the sound pressure level in the residential area after the installation of technological equipment on vibration isolators.

amount of energy that was transferred from the operating unit to the floor and walls decreased due to the reflection and absorption of vibrations.

The graph shows that the application of the VI helps to achieve the norms for the sound pressure level in a residential area. In addition to the measurement of the sound pressure level, vibration measurements were made before and after the vibration application of the VI (Figs. 8–10), which confirmed the effectiveness of the proposed means of the vibration isolation.

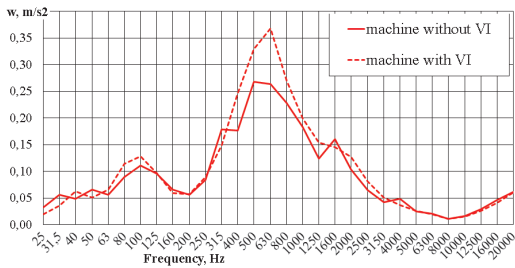


Figure 8. The vibration level of the machine before and after the installation of vibration isolators.

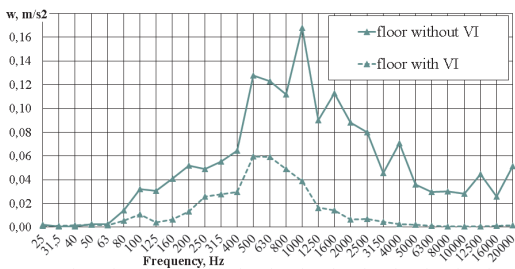


Figure 9. The vibration level of the floor before and after the installation of vibration isolators.

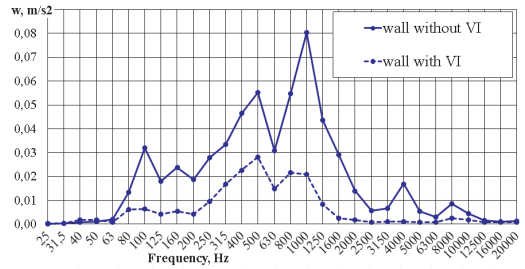


Figure 10. The vibration level of the wall before and after the installation of vibration isolators.

3.4 Conclusions

Thus, the installation of the vibration isolators designed for the technological equipment provided the permissible sound pressure levels and sound levels in residential and public buildings. To further reduce noise in the living room (to achieve night norms), it is necessary to implement all of the proposed set of the action items.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of the Russian Federation.

REFERENCES

- [1] Igolkin, A.A., Izzheurov, E.A., Hongyuan, J.b, Shakhmatov, E.V. 2011. Acoustic performances of metal rubber, *Proceedings of the 18th International Congress on Sound and Vibration*, Rio de Janeiro, Brazil, 10–14 July.
- [2] Igolkin, A.A., Rodionov L.V., Shakhmatov, E.V. 2008. Noise Decrease in Premises at the Expense of Application Dampers, *All-Russian Scientific, Methodical and Information Magazine "Safety in Technosphere"*, 4, 40–43.
- [3] Igolkin, A.A., Safin, A.I., Makaryants, G.M., Kruchkov, A.N., Shakhmatov, E.V. 2013 Non-contact registration and analysis of the product machine vibration with a three-component laser scanner, *Applied Physics*, Issue 4, Pages 49–53.
- [4] Igolkin, A., Koh, A., Kryuchkov, A., Safin, A., Shakhmatov, E. 2012. Pressure reducing valve noise reduction *19th International Congress on Sound and Vibration 2012, ICSV 2012, Volume 3, 2012, Pages 2458–2464, 19th International Congress on Sound and Vibration 2012, ICSV 2012, Vilnius.*
- [5] Lazutkin, G.V., Ulanov, A.M., Fedorova, I.V. 1990 Mathematical model of the deformation of structural-damping systems and its software implementation. *Strength of Materials*. Volume 22, Issue 8, August, Pages 1144–1150. Norwell.

- [6] Lazutkin, G.V., Ulanov, A.M.1989. Mathematical description of the deformation of complex structural damping systems under arbitrary loading, *Soviet machine science*. Issue 4, Pages 34–38. New York.
- [7] Ulanov, A.M., Lazutkin, G.V. 1997. Description of an arbitrary multi-axial loading process for non-linear vibration isolators, *Journal of Sound and Vibration* Volume 203, Issue 5, Pages 903–907.
- [8] Vér, I.L., Beranek, L.L. Noise and Vibration *Control Engineering: Principles and Applications: Second Edition*. Hoboken.

Intelligent management platform of forestry based on LBS cloud services

F. Li, Y. Ma & X. Zhang

Institute of Forest Resource Information Technique, Chinese Academy of Forestry, Beijing, China

M.B. Zhang

Research Institute of Forestry Policy and Information, Chinese Academy of Forestry, Beijing, China

X.W. Yu & P.F. Feng

Institute of Forest Resource Information Technique, Chinese Academy of Forestry, Beijing, China

ABSTRACT: Based on the analysis requirements of the cloud computing technology, location service technology and forestry business management, the paper proposes to build an integrated the basic geographic information data, professional software and high performance processing algorithm of forestry intelligent LBS platform technology solutions and methods of the management which is based on cloud computing technology.

1 INTRODUCTION

LBS (Location Based Services) refers to the spatial location of the mobile information service which uses the related technologies to cross dissolve such as GPS, GPRS, GIS, Internet and database, and provides users with many based services, technology, including location query, location search, vehicle tracking, urgent call by getting the location of the mobile user information.

Forestry demand for location services is mainly manifested in the following several aspects:

- 1 The demand of the forestry resources survey, from the survey of forest resources to wetland, desertification of land survey, forest resources field investigation has become a systematic and mass work. To complete a large number of field investigation work, using GPS positioning and integrated application of 3S technology to accomplish planning and design investigation has become the development direction of forest survey. Quickly and accurately positioning, area measurement, fieldwork two-way communication for field sample can achieve planning, designing, downloading survey tasks, automatically reporting in the data collection process, realizing image pass back; and the function of personnel, vehicles and aircraft navigation, choosing the best path, real-time positioning, behavior management, intelligent monitoring.
- 2 In the forestry project management, the ranger management is one of the important work of national key public welfare forest, local public welfare forest and natural forest resource protection, and

is to realize the necessary foundation of forestry management modernization and maintain ecological security. In order to further strengthen the ranger management work, it improves the ranger at the village work sense of responsibility and prevents forest fires, pests, attack and deforestation and no checking or inspection does not reach the designated position phenomenon happens, need information for acquisition and transmission to the position of the rangers, in order to monitor the management of a forest area effectively.

- 3 Biodiversity protection, technical requirements for location service is mainly to realize quickly, accurately positioning, area measurement, fieldwork two-way communication for biodiversity monitoring, achieve the field sample, object of protection and complete the task of download, automatically report in data collection process, image pass back, realize personnel, manage the daily patrol and protect of vehicles, combine with sensor devices and monitor the position of wild animals and plants, behavior trace(animals), environment.

2 LBS CLOUD SERVICE TECHNOLOGY

Cloud computing is a technology revolution of the IT industry, which has become the direction for the future development of IT industry. Cloud computing as a new network application mode, through the Internet to get the autonomy of heterogeneous computing resources together; can be able to handle huge amounts of equipment.

Forestry information development lags behind that of other industries, and many forestry businesses do not have the basic hardware condition and IT maintenance personnel. Through the LBS technology combined with cloud computing technology to form LBS of cloud services, LBS cloud service makes the location data storage, spatial retrieval, map showing tasks, host in the cloud, both to reduce development cost and improve development efficiency. With the LBS services in the cloud, the forestry business units do not need to purchase a large database software and geographic information software to manage the spatial data and the corresponding hardware environment; it is a big savings in operating costs.

The platform adopts the development interface of Baidu LPS cloud services to achieve its goals, we can design and storage management for their own data of the forestry business department by using the video LBS cloud storage API, Using visual data editing tools and map labeling or the way of batch data importation can manage their own data of the forestry business department easily. In the cloud, utilizing the open Baidu map engine operation ability index of the stored data in real time, use Baidu LBS cloud retrieval API returns the various location-based data retrieval request. Simultaneously, it can realize map showing by using Baidu map API. Using Baidu cloud services development, LBS application system process is shown in figure 1.

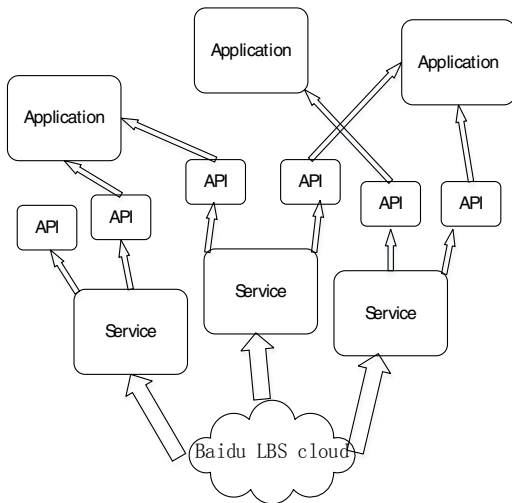


Figure 1. Develop an application system using a Baidu LBS cloud service.

3 SYSTEM FUNCTION DESIGN

Intelligent management platform of forestry is an information service platform which adopts GPS, GPRS, GIS, Internet, database and related technologies to cross and combine. In communications network coverage within the scope of any location, any time or any environment, through reserve patrol's smartphones, which installs intelligent inspection APP, implements the two-way communication between management station and the patrol. The platform is divided into patrol platform (contains mobile patrol client management), patrol data collection platform、platform for data analysis and decision and system management platform. Intelligent management platform of the forestry function architecture diagram is shown in figure 2.

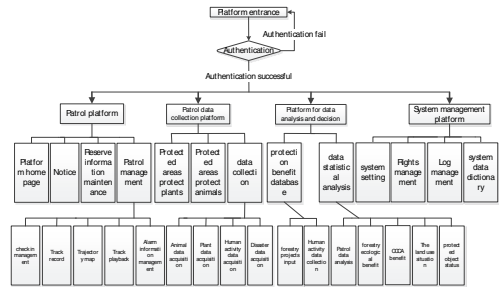


Figure 2. Intelligent management platform of forestry function architecture diagram.

- 1 In patrol management, realizing the acquisition and transmission of the position information of the patrol, historical trajectory playback, can save all the patrol trajectory of the patrols and is effective to monitor the patrols' management area;
- 2 In terms of biological information acquisition, the platform realizes two-way communication for field operations, tasks downloaded, field data collection and reported, image back, wild animals and plants positioning, behavior trajectory record (animals), and environmental monitoring, etc.
- 3 In terms of emergency response, through the way of one key alarm, timely feedback of the information such as thieves cutting the unlawful, forest fire, forest diseases and insect pests to the platform through a mobile terminal, and taking photos, videos, the management station can take manner to respond to these alarm information timely.

4 THE UNDERLYING SUPPORT ENVIRONMENT DESIGN BASED ON LBS CLOUD SERVICES

The operation of the platform is divided into data acquisition terminal, server terminal and terminal operation. The design of the data flow model is: firstly call the LBS location service for mobile client's position, and passes back to the server through GPRS, and then the server calls baidu LBS cloud service interface to transmit data to the cloud, save the coordinate data in the cloud storage, and the server by calling the baidu LBS cloud map displays interface, and through the manipulation of the platform in a web page to display patrol and protect monitoring data , at the same time the server can sent various commands to the mobile phone client.

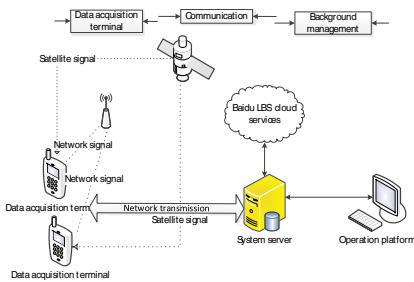


Figure 3. Intelligent management platform of forestry data process.

Data acquisition terminal generally uses the current mainstream Android smartphone, collection of information including:

- 1 Text information: records of all kinds of information, with SQLite files .db stored in local mobile phone;
- 2 Data information: all kinds of data, generally SQLite files (or with text information records);
- 3 Image information: all kinds of pictures, generally .JPG and .tif files, etc.

According to the characteristics of the forest area, using radio communication network and GPS to get position information; network transmission, which uses the GPRS, information transfer mainly include two types:

- 1 Real-time information transfer: when there are GPRS network signals, location information feedbacks to the remote server in time, to facilitate instant response;
- 2 Non-real time information transfer: if there is no GPRS network signals or when the situation of

the poor, temporarily cache location information feedbacks to the terminal device, and then back to the remote server.

Android smartphones as acquisition terminal equipment encapsulates the Location service API. Baidu location of SDK is a set of LBS location-based services interface for Android mobile terminal applications which is simple and easy to use. After Baidu location of the SDK integration in the application, it can easily through the location of the SDK interface to Baidu LBS request location information, with the help of GPS, base station, wi-fi and sensor information, to realize high precision hybrid location services.

Mobile phones, cloud and system server using Web Service technology to communicate. Web service is a platform independently, loosely coupled, self-contained, based on the programmable Web application, and you can use open source XML (a subset of standard generalized markup language) standard to describe, publish, discover, coordinate and configure these applications, for the development of interoperability distributed application. All the communication methods are encapsulated into the class "conn. WebserviceHelper". The web service is released in the system, server, the Android client can invoke the Web Service through a third party ksoap2 - Android libraries. The Web service is responsible for storing the location information and the collection of data in the server database, and platform can get the patrol location information and the detailed information via real-time access to the system server database.

The server system is the bridge between the mobile terminal and Baidu LBS cloud services. All collected data from mobile phones come together on the server system, and upload a copy of the location information in the Baidu LBS cloud for storage and management, cloud storage interface functions: support for a single gleotable billion level data and the size of the hundreds of G storage, support millions of magnitude read and write every day, and support per square midleweight read and write.

5 CONCLUSION

Through the LBS cloud services to build intelligent management platform of forestry, reduce the threshold of the forestry management department by using GIS technology, it is more convenient and faster to build the application of forestry management business. At the same time as the data and a large number of operations in a public cloud, it can solve the data security, uniform access interface, information isolated island, software and hardware repeated investment problems.

ACKNOWLEDGMENT

This paper was supported in part by National Major Special Projects and Development Program of China and Science Foundation of The Chinese Academy of Forestry, under contract number 21-Y30B05-9001-13/15, and IFRIT201104.

REFERENCES

- [1] Z.L. Yu, Y.X. Chen, M.J. Zheng etc. Android SDK development model greatly entire [M]. Beijing: Posts & Telecom Press, 2009. (In Chinese).
- [2] Y. Chen, G. Li, J. Xu. Mobile GIS development research based on ArcGIS Mobile [J]. Modern commercial and trade industry, 2009(23). (In Chinese).
- [3] S. Wang, H.Q. Liu, S.S. Chen Research and application of LBS in mobile guide system [J]. Microcomputer Information (control integration), 2010, Vol.26, 7–3. (In Chinese).
- [4] P.X. Tian, Z.Y. Zhu. Based on J2EE LBS server platform design III[J]. Microcomputer Applications, 2005, (2). (In Chinese).
- [5] G.S. Zhang, B. Gao, H.M. Yao. The mobile location services and digital map [J]. Journal of Geomatics Science and Technology, 2007, (3). (In Chinese).
- [6] Y. Song, Q.Q. Li, N.B. Zheng, Based on LBS distributional vehicles monitoring information service platform [J] Computer engineering, 2007, (6). (In Chinese).
- [7] C.M. Wang, Z.M. Wang, L.Z. Zhu. Problems and countermeasures existing in the LBS [J] Geomatics & Spatial Information Technology, 2009, (3). (In Chinese).
- [8] Q. Li, X. Zheng, Cloud computing research overview [J]. Computer Science, 2011, 38 (4): 31–37. (In Chinese).
- [9] D.P. Yang, W.M. Lang, H.S. Li, Cloud computing service model research [J]. Telecommunications Information, 2011 (11). (In Chinese).
- [10] B.Y. Fang, Y.Y. Zhang, Q.J. Chen. Unified SaaS platform in cloud computing environment[J]. Telecommunications Network Technology, 2011 (05). (In Chinese).
- [11] S.G. Zhu. Study on the development trend of satellite navigation systems and industry [J]. Digital communication world 2010 (S1): 34–36. (In Chinese).
- [12] Saha AK, Gupta RP, Arora MK. GIS-based data integration for landslide hazard zonation in the Garhwali Himalayas[C]//Anon.Map India Conference. [S.l.]: [s.n.], 2003.
- [13] L. Wu, Y. Liu, J. Zhang, et al. Geographic information system principle, method and application [M] . Beijing: Science Press, 2001 (In Chinese).

Focus+context visualization with multi-resolution meshes

Y.P. Zha, Z.J. Li, Y.M. Liu & F.X. Li

Beijing Institute Of Technology, Beijing, China

K. Yang

China Aerospace Science & Industry Corp, Beijing, China

ABSTRACT: Focus+Context Visualization has been widely used in many fields and confirmed to be a good method to help us to see focus regions in detail and keep the context. But most of the algorithms didn't consider the characteristics of the models. In this paper, we divide the 2D-model to multi-resolution mesh depend on the characteristics of the models like distribution of roads on a road network, and use energy optimization model to get minimization distortion while magnify an area of interest. In particular, we show how to forbid edge fold, which generates after the deforming. And we demonstrate the efficacy and less distortion of our method.

1 INTRODUCTION

Geographic information system, which includes many map styles like a road network map, road traffic map, tube map, terrain map, etc., has been widely used in many fields. We can get many important information from these maps, like we can get a comprehensive observation of a city's road construction from a whole looking on the road network in the city, and study of a specified region of the city if we are interested in some place by magnifying the interested area, then we can get a microscopical condition of our interest regimes.

When we want to observe a region we are interested in more detail, a simple solution is, we find a new place to put an enlarged version of the region, nonetheless, this straightforward sub-window technique requires the user to make an accurate map of the sub region to the original version. Another solution is, to magnify the specified region in the map, but its violent side—this solution would cut off the other regions which are far away from this region, thus we could not get a full view of the model.

When we display a complex model, the Focus+Context technique can magnify the area of interest without cutting off the other parts, even the parts which are far away from the specified region. Thus we can observe the model's detail without losing the overall view of the model's shape and topology, all of which just likes to use a magnifying lens to enlarge the focal region that the user are interested.

To observe the maps like road network, tube map and terrain map which are generally very complex, we propose a framework to interactively deform a multi-resolution-mesh model to achieve Focus+Context

visualization with distortion minimization. In consideration of the internal characteristics of the map, for example, the nodes in a road and the lines in a tube map network are non-homogeneously distributed, the surface features are not familiar at all the land. We construct a given model to multi-resolution meshes using the mind of a quadtree, split the sparse region to lower-resolution meshes and the dense region to higher-resolution meshes. We can reduce the other regions to keep the entire model within the global bounding space. We map every node of the model to one mesh, thus we can deform the model by computing each model node as a linear combination of its respective set of mesh vertices in the deformed mesh space.

The maps are sensitive to the aspect ratio of all its sub-regions. Since our goal is to constrain the model to be within a bounding space, when the focal region is expanded, some regions must be reduced. But using different scale on the same map implies distortions—if we want to measure distances, such a distorted map is useless. So we aim to keep it similar to its original appearance, avoiding squeezing or squashing which would result in distortion. We use the optimization procedure which is proposed by Yu-Shuen Wang, allows empty grids to be stretched to absorb the resulting distortion. However, the transformations of connected grids are not identical, the deformation inevitably causes distortion. But we should minimize the distortion, so we use a set of energy terms to form an optimization system and solve for the grid vertex positions of the deformed space in a least-squares sense.

In Yu-Shuen's article^{vi}, he partitions a 3D model using a n^3 uniform grid space. We need to achieve

the Focus+Context visualization on 2D model, so we split our model to n^2 uniform meshes. And we use the energy optimization model the same to Yu-Shuen's. We use the road network of Beijing as the background and divide the model to different resolution to compare the effects, and we use two criteria, deformation quantity and the operating efficiency to judge how good the algorithm is. Obviously, an interaction can be responded in three seconds, which can be tolerated by users. According to our repeated experiments, we find that higher resolution of the meshes of the model we divide, less deformation quantity would get after we deform the model. But we find a general problem while dividing the model with a very high resolution to get as less deformation quantity as we can—the operating efficiency, higher resolution we divide, more running time needs. On figure 1, we represent the deformation with variable transparency, higher deformation represent by darker red.

Our paper is structured as follows. We firstly introduce the Focus+Context visualization in Sect.1. Then we give an overview of related work in Sect.2 and we present our method in detail in Sect.3. We discuss our experiments in Sect.4. Finally, we conclude the paper and give an outlook on future research in Sect.5.

2 RELATED WORK

Focus+Context has been applied in the information visualization field widely. As the growing of scientific technology, more and more approaches to study the data in the fields of medical science, earthquake, biology, etc. have been proposed. To explore the data efficiently, Focus+Context have been highly used in the field of science visualization.

Various Focus+Context frameworks have been used widely. In 1973, William Farrand [1] first raised the idea of “An efficient data visualization method, which should provide the user the sense in global context when user wants to see detailed information of local area”. Researchers raise Focus Context technology to accomplish the target that both show local detail of the user interests and the global context in a limited resolution screen. In 1980, Robert Spence and Mark Apperley [2] raised Bifocal display, which was used Focus+Context visualization.

In the development process of information visualization, the Focus+Context visualization has evolved to several classes. In 1986, Furnas [3] learnt from the fisheye view and raised fisheye algorithm which had a huge influence on Focus+Context visualization. The fisheye algorithm is applied in the rendering of urban map, because in the real world, the road network and buildings are non-uniform distributed, in some regions, there are dense buildings, and in some others, the buildings are very loose. Feng, Kun-Chuan,

etc. [4] published an article on IEEE TVCG which did research on Focus+Context of complex networks. In their paper, they proposed according to analysis the importance of a node and analysis of clusters, and they found the correlations of main nodes directly. In medical diagnosis field, the diagnosis of disease is always depending on the size, shape, appearance and etc. of the diseased organ, so if we highlight the lesion area while present the whole organ, we can give strong support to the doctors to give an accurate diagnosis. Wei-Hsien Hsu [5], and etc. gave a rendering framework for multiscale views of 3D Models. The basis of their approach is a set of non-linearly bent camera rays that smoothly cast through multiple scales. Yu-Shuen, W, and etc. [6] use the uniform cube to realize the Focus+Context visualization of surface of 3D models. In this algorithm, it can keep distortion minimization while magnify the focal zone, but this algorithm can't keep the overall shape.

Huamin Qu [7], et al. used Focus+Context visualization on 3D Urban Environments. Through the creative utilization of the empty space in an urban environment, their technique can informatively reveal the focus region and minimize distortions to the context buildings. In 2011, Yu-Shuen, W, et al.[8] realized the Focus+Context visualization in metro maps. Users can highlight the specified line, and the technique has been used in mobile, pad and so on. And in the same year, Jan-Henrk, et al. [9] realized Focus+Context visualization for the road network. In the algorithm, Jan-Henrk regarded the road network as a map, and each road is an edge of the map, he used CQP (convex quadratic program) to solve the objective function to minimize the distortion of the edges, and used constraint to avoid cross of two edges.

3 FOCUS+CONTEXT VISUALIZATION OF ROAD NETWORK

Every road network model is consisted of nodes and edges. One node represents a switch of roads, and one edge represents one road. We represent a black dotted circle as the magnifying lens, the user can move the circle to the interest region to magnify the focal region. The user can set parameter α as the magnification factor of the focal region. We set several meshes in a yellow frame, which stand for the meshes covered in the focal region. As the circle moved to a region, we deforms the meshes by the magnification parameter z , the vertexes of the meshes would be changed automatically. Our method solves the mesh vertex positions by minimizing a set of energy functions retaining the aspect ratios of the meshes covering the given model as well as constraining the mesh vertices to be within the bounding space. We computes the position of every node in each model by a linear combination of its respective

set of mesh vertices in the deformed mesh space, and we redraw the edges connected by two nodes which are neighbored to reconstruct the model. As the mesh vertices are all constrained to be within the bounding space, the nodes of the model would stay within the bounding space, which makes us accomplish the aim of Focus+Context Visualization.

3.1 Merge the model to multi-resolution meshes

We can easily get that higher resolution of the meshes of the model we divide, less deformation quantity would get after we deform the model. But we find a general problem while dividing the model with a very high resolution to get as less deformation quantity as we can—the operating efficiency, higher resolution we divided, more running time we need. So we achieve the focus+context visualization with multi-resolution mesh to get less deformation quantity while guarantee the operating efficiency. In the areas in which nodes are very dense, we divide them into higher resolution to make sure that we get less deformation quantity, and in the areas in which nodes are very sparse, we divide them to lower the resolution to guarantee the operating efficiency.

We follow the steps to do the focus+context visualization:

step 1: Divide the model to the highest resolution based on LOD, the size of the highest resolution should be a power of 4.

step 2: Set LOD levels of the model.

step 3: Map every node of the model to the highest resolution meshes.

step 4: Initial index to all the nodes of every mesh unit in each LOD level.

step 5: Get the amount of nodes in every mesh unit of the highest resolution.

step 6: Set the max amount of nodes as the threshold.

step 7: Combine the meshes cyclic, from the first mesh of the second highest resolution, if the sum of the nodes in the mesh is smaller than the

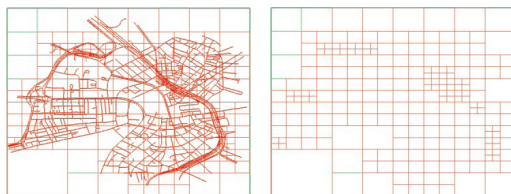


Figure 2. Multi-resolution meshes of road model of Boston we divided while the initial resolution is 32*32, the left graph is grids with model, the right is without model.

threshold, then combine its four children, or turn to the next mesh until this mesh is the last mesh. Then start from the first mesh cyclic until it arrives to the lowest resolution. Finally, we get the multi-resolution mesh.

In figure 2, we show the multi-resolution mesh we divide of the road model of Boston.

3.2 Eliminate the edge fold drawback

After we get the multi-resolution mesh, we solve the non-linear constrained objection which proposed by Yu-Shuen Wang^v to minimization the distortion. However, there is a drawback in the edge between different resolution mesh that we call it edge-fold problem. When the resolution of the focal regions is different from the resolution of the adjoining areas, after focus+context visualization, three vertexes are collineation and won't be collinearity.

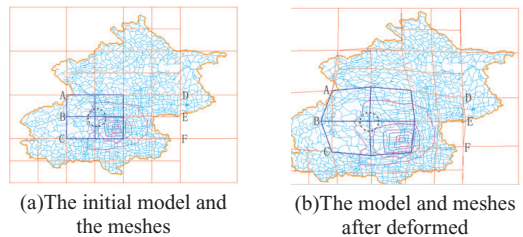


Figure 3. The edge fold problem in multi-resolution.

The plan to solve the edge-fold problem is: we get the LOD levels of two neighboring meshes through the labels of each mesh we set before, so we can get edges which have the fold problem. Then we add constraint condition to the edges.

The algorithm is as follows:

Step 1. Get the edges which are between two meshes that have different resolution:

If the edge between two meshes which have the same resolution won't have the edge-fold problem, we will consider the other situation, and two neighboring meshes have different resolution, and the vertexes of the higher resolution mesh on the edge would come to the problem.

In the process of dividing the model to multi-resolution mesh, we label every mesh to three state: subdivide, render and reject. The reject state means the mesh don't cover the model or call the Trivial Cell. We analyze every mesh as follows: compare the current mesh to all its neighbors. If its neighbor's state is render, it means the two meshes are on the same LOD level, and it won't have the problem; if its neighbor's state is subdivided, it means its LOD level is lower than its neighbor's, and it won't have the problem, too; if its

neighbor's state is rejected, it means this mesh is on the border, and it also won't have the problem; if its neighbor hasn't labeled a state, it means its neighbor is rendered or rejected on a lower LOD level, then we should get the neighbor's parent's state cyclic, until its state is rejected or rendered, if its state is rejected, it won't have the problem, but if its state is rendered, we need to add constraints to the edge between the two meshes.

Step2. Add constraints to the edges which have the edge-fold problem:

The main idea of add constraints to the edges is to make the three points on the edge of the higher resolution mesh to be collinear.

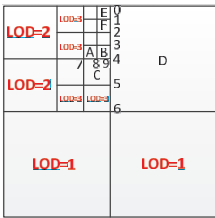


Figure 4. The reason of edge fold problem.

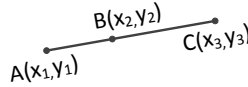


Figure 5. Three points collinear.

As it is shown on Figure 4, there has multiple edges between two meshes which have different resolution that would have the edge-fold problem, we should add constraints on them. If the resolution level differs by one between two neighbored meshes, we just add constraints to make points(7,8,9) to be collinear. But when the resolution level differs by $n(n > 1)$, we should add constraints to the points whose index interval is 2^{i-1} ($i = 1, 2 \dots n$), the resolution level differs by 3 between E and D, we should add constraints to make points(0,1,2), points(0,2,4), points(0,2,6) collinear.

In Figure 5, points A, B, C are collinear, and the coordinates are (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , and B is in the middle of A and C, so we get $\vec{AB} = \lambda \vec{AC}$ ($0 < \lambda < 1$), we can get the equation which is as below:

$$(x_2 - x_1, y_2 - y_1) = \lambda(x_3 - x_1, y_3 - y_1) \Rightarrow \begin{cases} x_2 - x_1 - \lambda x_3 + (1-\lambda)x_1 = 0 \\ y_2 - y_1 - \lambda y_3 + (1-\lambda)y_1 = 0 \end{cases} \quad (0 < \lambda < 1) \quad (1)$$

Set a tetrad $e_i = (V_{iA}, V_{iB}, V_{iC}, \lambda_i)$ stands for the i th constraints and should be added. λ_i stands for the length ratio of line $V_{iA}V_{iB}$ and line $V_{iB}V_{iC}$, and ($0 < \lambda_i < 1$)
 $\circ E = (e_1, e_2 \dots e_m)$ stand for all the constraints and also should be added for collinear.

The i th constraints is $l_i V_p = 0$, in formula (1), l_i which is defined as below:

$$l_{i,u} = \begin{cases} -(1-\lambda)\sigma & \text{if } u=V_{iA} \\ \sigma & \text{if } u=V_{iB} \\ -\lambda & \text{if } u=V_{iC} \end{cases} \quad (2)$$

σ is a constant, which stands for the punish factor when the three points against collinear, in our experiment, we set $\sigma = 2$.

All the constraints can use $l_i V_p = 0$ to denote, $L = [l_1, l_2, \dots, l_m]$. We add all the constraints to the objective function, then we can solve the problem.

4 EXPERIMENTAL RESULTS

We implement our method in Windows 7, using C++, OpenGL and Taucs library of Intel Core i5 750 2.67HZ, Nvidia Quadro FX 3400/4400 512MB. We get the data of road map in Boston from Jan-Henrik Haurert.

We calculate the distortion between the original model and the deformed model by calculating the residuals of each node in the model. And the residual between two nodes which are neighbored is calculated by formula (6) which is raised by Haurert^{viii}:

$$\begin{cases} \sigma_{x_{uv}} = s_u(X_v - X_u) - (x_v - x_u) \\ \sigma_{y_{uv}} = s_u(Y_v - Y_u) - (y_v - y_u) \end{cases} \quad (3)$$

In formula (6), u represents any node of the model, v represents one neighbor of u (two nodes are neighbors means a line connected by the two nodes), $\sigma_{x_{uv}}$ and $\sigma_{y_{uv}}$ stand for the residual on x axis and y axis between u and v , s_u represents the scale of the mesh which lies in node u . And we calculate the weighted square sum of residual of every node in the model by formula (7).

$$\sum_{v \in Adj(u)} ((w(u,v) \cdot \sigma_{x_{uv}})^2 + (w(u,v) \cdot \sigma_{y_{uv}})^2) \quad (4)$$

with $w(u, v) = 1 / \sqrt{(X_v - X_u)^2 + (Y_v - Y_u)^2}$.

We set the magnification parameter $z = 3$ which is the same as Haurert and magnify the region in the black circle on Figure 6, and the results are showed in Figure 7.

From Figure 7, we can find the distortion of our algorithm is far less. And the final effects comparisons are shown in Figure 8. We list the times needed on deforming while divide the model to different resolution in table 1.

5 CONCLUSION AND FUTURE WORK

As the distribution of a road network is non-uniformed, so we take the idea of multi-resolution into Focus+Context visualization to solve the problem between deformation quantity and operating efficiency. But due to the fact that we divide the model to multi-resolution mesh, it comes out a drawback-edge fold. We use Yu-Shuen's method to minimize the resulting distortion to keep the original mesh appear similar to its original counterpart. And our algorithm inherits the advantages of Yu-Shuen's method and solves the edge fold problem in the result of the visualization.

Our method solves instances of about 5000 road segments in a few seconds. We are concerned about the resolution rather than the amount of the model, so we can solve a very compressive model in a very fast time and, in the mean time, we can make less distortion. Furthermore, we plan to improve the optimization model. And in future we will consider multi-focus-context visualization.



Figure 6. The original road network of Boston.

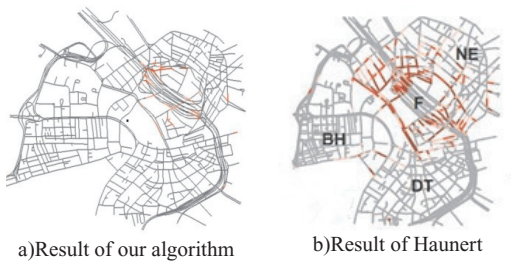


Figure 7. The comparison of distortion.



Figure 8. The effect comparison of Hauner's and our.

Table 1. The times needed on deforming while divided the model to different resolution.

| Model | Edges | Nodes | Resolution | Running Time(s) |
|----------------------|-------|-------|------------|-----------------|
| Beijing road network | 5195 | 65365 | 32*32 | 16.589 |
| | | | 16*16 | 1.888 |
| | | | multi | 3.46 |
| Boston road network | 2852 | 3379 | 32*32 | 6.331 |
| | | | 16*16 | 3.017 |
| | | | multi | 4.073 |

REFERENCES

- [1] Farrand, William A. (1973). Information display in interactive design, Doctoral Thesis. University of California at Los Angeles
- [2] Spence, Robert and Apperley, Mark (1982): Data Base Navigation: An Office Environment for the Professional. In Behaviour and Information Technology, 1 (1) pp. 43–54
- [3] Furnas, George W. (1986): Generalized Fisheye Views. In: Mantei, Marilyn and Orbeton, Peter (eds.) Proceedings of the ACM CHI 86 Human Factors in Computing Systems Conference April 13-17, 1986, Boston, Massachusetts. pp. 16–23
- [4] Feng, K.-C., Wang, C., Shen, H.-W. & Lee, T.-Y. Coherent Time-Varying Graph Drawing with Multifocus+Context Interaction. Visualization and Computer Graphics, IEEE Transactions on 18, 1330–1342, doi:10.1109/tvcg.2011.128 (2012).
- [5] Hsu W H, Ma K L, Correa C D. A rendering framework for multiscale views of 3D models[J]. ACM Trans. Graph., 2011, 30(6): 131.
- [6] Yu-Shuen, W., Tong-Yee, L. & Chiew-Lan, T. Focus+Context Visualization with Distortion Minimization. Visualization and Computer Graphics, IEEE Transactions on 14, 1731–1738, doi:10.1109/tvcg.2008.132 (2008).
- [7] Qu H, Wang H, Cui W, et al. Focus+ context route zooming and information overlay in 3D urban environments[J]. Visualization and Computer Graphics, IEEE Transactions on, 2009, 15(6): 1547–1554.
- [8] Yu-Shuen, W. & Ming-Te, C. Focus+Context Metro Maps. Visualization and Computer Graphics, IEEE Transactions on 17, 2528–2535, doi:10.1109/tvcg.2011.205 (2011).
- [9] Haunert, J. H. & Sering, L. Drawing Road Networks with Focus Regions. Visualization and Computer Graphics, IEEE Transactions on 17, 2555–2562, doi:10.1109/tvcg.2011.191 (2011).

Modeling channel variations at 433MHz for wireless sensor networks in indoor scenarios

M.W. Khan & T. Zahid

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China
Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, Guangdong, China

Y.M. Zhou & G.Q. Xu

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China
Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, Guangdong, China
The Chinese University of Hong Kong, Hong Kong, China

ABSTRACT: The indoor channel model for wireless sensor networks for the office environment is presented in this paper. The diversity of an indoor wireless channel due to time-variant nature of channel can add some great impact on the accomplishment of the wireless communication systems. A channel model plays a vital role to classify the consumption and understanding of the communication system. In this paper we have discussed (to the best of our knowledge) the characteristics of the received power along with path loss, power delay profile, impulse response for multipath signals and scattering for the indoor channel model at 433MHz by using FSK modulation.

1 INTRODUCTION

Wireless sensor networks have attracted so many researchers in the past decade. In wireless communication. It is a well known fact that the underlying communication channel characteristics heavily change the performance of the overall communication system. A proper channel helps us in the simulation for the better performance of the estimation.

The said model can be very useful for finding the data of paths in a very huge number of territories. At any time when such a territory is established, with the known properties of geometric and materials, it's very much likely to interchange statistical ability of the paths for accurate observation of unique paths utilizing the simulators for different type of environments [1]. In indoor channels; WSN has many major applications inclusive of the management system and body sensor networks and security surveillance for both terrestrial and underwater [2]. Radios effects for these kinds of networks are highly constrained due to the sharing of bandwidth with other large number of nodes along with the limited transmit power [3]. Implementing and running an experiment to study the real time presence of different types of impairments of the channel and performance of the network need a great deal of work [4, 5, 6]. In [4,5,6] they did the work by using experimental data with a much used

Mica2 Motes [7] and all of them agreed, that due to the time variation the real wireless links show a random and uncertain behavior.

Performance of WSN, e.g., consuming of energy, reliability, accuracy and capacity mainly depend on the properties of the channels between sensor nodes. [2]. The experiments and simulations for a channel model should be proper enough without incrementing the complexity and cost, therefore the main objective of this paper is to get the more proper and low complex channel model to provide more realistic WSN simulations.

This paper presents the indoor channel model at 433MHz for an indoor office environment. The measurements are carried out in an office environment, including building floor, cubicles, halls and tea rooms.

The remainder of the paper is as follows: section II explains different terminologies used in propagation channels. Section III summarizes about the experiments and results Section IV concludes the paper.

2 INDOOR PROPAGATION MODEL

Usually indoor channel models are difficult to implement because of the channel variation with the environment. The factors which affect the indoor channel

include building architecture, material used in the construction and layout of different rooms and offices. For a better understanding of the presented model, there are few basic terminologies to be explained to the scholars along with different parameters of the problem.

2.1 Path loss

The main indoor channel is the one with free space line of sight with no hurdles between the transmitter (Tx) and receiver (Rx) or in the surrounding paths between them. In this case, the received power for line of sight (LOS) is given by

$$P_r = P_t \frac{\sqrt{G_t G_r}}{4\pi d^2} \quad (1)$$

where P_t is the power transmitted, G_t is the result of the Tx and Rx antenna radiation field, wavelength is the λ , and d is the given distance. The received signal suffers from attenuation while transmitting is as defined by [8]

$$PL \text{ dB} = 10 \log \frac{P_t}{P_r} \quad (2)$$

Ground contributes toward the reflection of the waves and reaches to the transmitter, causing the waves to make a 180 degree shift. It may also reduce the total power received at the receiver side. A duplex estimation for path loss is defined by equation 3.

$$P_r = P_t \frac{G_t G_r h_t^2 h_r^2}{d^4} \quad (3)$$

where h_t , h_r are the heights of the transmitter and receiver, gains of Tx and Rx antenna are G_t , G_r respectively. The above equation is not valid for $d=0$, because of the distance factor in the denominator. Therefore, a lot of the propagation channels use different methods for a close distance, d_0 , known as the power received reference point [9]. In real time radio channels, a free space is not a recommended medium to choose, a more regular path loss model uses α to define the interconnection between received power and the separation distance. The path loss equation is given by

$$PL(d) = PL(d_0) + 10 \alpha \log(d/d_0) + X_\sigma \quad (4)$$

where $\alpha = 2$ for free space. However the value of α is usually higher for wireless communication channels. $PL(d)$ is the path loss with distance d . The d_0 , is the reference distance while X_σ represents a zero-mean Gaussian random variable with standard deviation σ .

The path loss is the fundamental contributor for propagating a channel and is correlated for the coverage of the area of radio system

2.2 Power delay profile

Random and complex radio channels can be represented by the channel impulse response. The impulse response $h(t)$ is represented by the linear filter of the channel for each point in 3D environment. The impulse response contains detailed information about the channel and all necessary information to simulate different types of radio transmission through given channel [13]. By taking the average of $|h(t)|^2$ over a local area, the statistical power distribution can be calculated by power delay profile which is useful for deriving many multipath channel parameters. Power-delay profiles are presented in the form of plots of the P_r as a function of an additional or excess delay with respect to a fixed time delay reference [13]. Due to the finite velocity of propagation of the EMT signals there is a delay between the time of transmission and at the time the signal is received. A density function of the power delay profile is given by equation

$$P \tau = \frac{|h(t)|^2}{\int_{-\infty}^{\infty} |h(t)|^2 dt} \quad (5)$$

2.3 Multipath

Usually the entities over the wireless signals reflect the particular signal and most of the time, the receiver also receives these reflected signals. As different paths are taken by reflecting signals due to which they have different amplitudes and phases, these multipath signals may have more or less power at the receiver. The channel measurements can lead us to the channel model by adding all the multipath signals. If the signal input is a unit, $\delta(t)$, the output of the channel will be the channel impulse response, which can be written as [11]

$$t = \sum_{n=1}^N A_n \delta(t - \tau_n) \exp -j\phi_n \quad (6)$$

where A_n is attenuation, n the arrival time of delay, and ϕ_n is the phase to parallel path n .

Multipath is the main cause of the severe degradation of the communication systems for both indoor and outdoor.

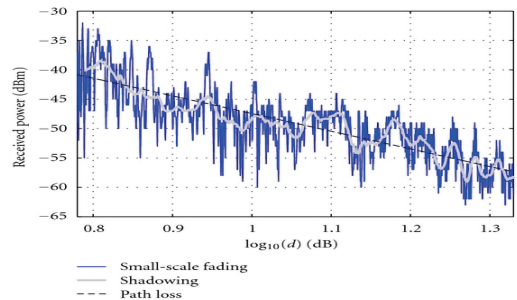


Figure 1. Comparison between small scale fading, shadowing and path loss.

We can only have a good performance if the pre-existing knowledge of the propagation is known and whether it has effects on the system. Scattering, diffraction, and reflection are the three main contributors which affect the radio communication systems. Figure 1 [12] shows the shadowing, small scale fading and path loss.

2.4 Scattering

When the receiving signal at the receiver hits any object whose size is equal to or less than the wavelength of the signal, the scattering occurs. Usually scattered waves are generated by the rough and sharp surfaces included in the channel. In general case the scattering can be omitted [14], while in our model we have included it. The complex field received from the various incident paths is given by [14]

$$E_{\text{receive}} = \sum_i E_i \tag{7}$$

$$E_i = E_o \varphi_{ti} \varphi_{ri} L_i d \prod_j \Gamma \varphi_{ji} \prod_p T \varphi_{pi} e^{-jkd} \tag{8}$$

E_o is the reference incident amplitude (V/m); where φ_{ti} and φ_{ri} are the Tx and Rx-antenna field radiation patterns the direction of interest for the i th multipath component; $L_i(d)$ and $\Gamma \varphi_{ji}$ are the i th multipath component for the path loss and the reflection coefficient for the j th reflection respectively; where $T \varphi_{ji}$ is the transmission coefficient for the p th transmission of the i th multipath component; e^{-jkd} is the propagation phase factor due to the path length $d(k = 2\pi/\lambda$, with λ representing the wavelength); d is the path length (m); and E_i is the field strength of the i th multipath component [14]. Figure 2 presents the scattering function for indoor environments.

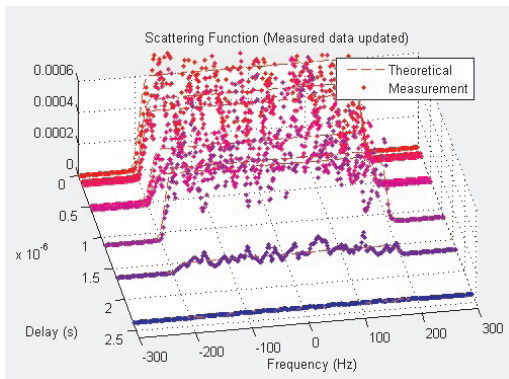


Figure 2. Scattering theoretical and measured for indoor.

3 EXPERIMENTS AND RESULTS

The experiment is conducted in an indoor office environment. The building has 12 floors and 3 wings where office area is occupied by different shape and size of the cubicles made from the combination of thin fiberglass and wood, and office rooms for workers.



Figure 3. Office Environment.

The average area covered by the cubicle is 1.5 m x 2.0 m of office space and covered by 1.3 m tall partition. We conducted our experiment in the B wing of our office and most of the area is covered by conference rooms and professors' offices separated by concrete walls. The schematic of our office area can be seen in figure 3.

On the right side there are offices and conference rooms, while on the left side there are cubicles. We assume that there is no hostile movement, e.g., no one is walking around and the office doors are not being open and closed all the time (we performs the experiment at night to ensure about the movements caused by the above) no movement means static nodes and no dynamic object around it.

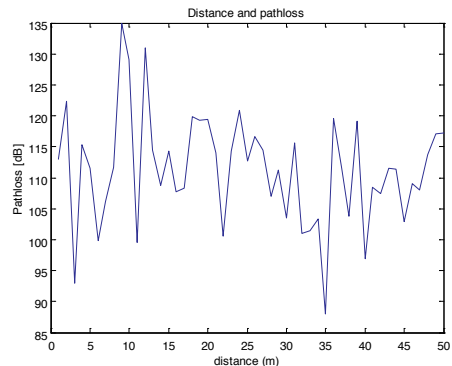


Figure 4. Path loss vs Distance(m).

We also have the blueprints of our office building and the exact location of the transmitter and receiver are known, with all the above information we have developed our model for path loss as a function of distance. Some of the parameters σ and α are obtained from [10].

The figure 4 presents the result of path loss in the scenario explained above. It can be seen that the path loss is directly proportional to distance where the distance is expressed in meters and path loss in dB. It can be also seen that the path loss also increases with the increase of the frequency. Figure 4 presents the power delay profile for the method explained above. As the PDF gives the intensity of a signal received through multipath as a function of time delay which is depicted in figure 5 and figure 6 shows the impulse response for the multipath signals.

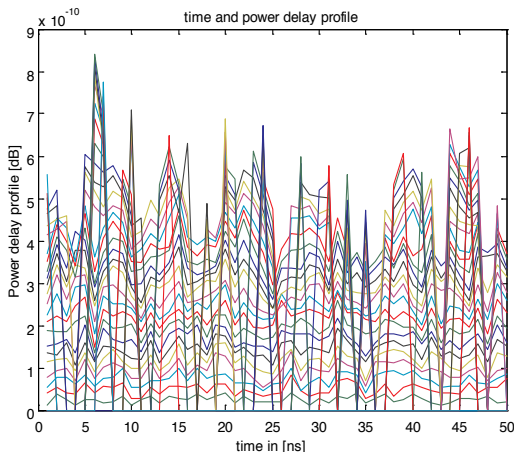


Figure 5. Indoor channel Power delay profile.

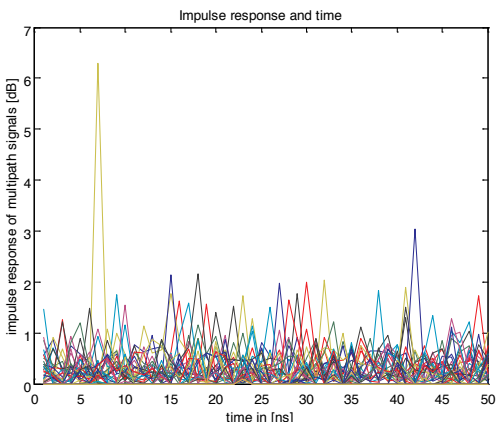


Figure 6. Impulse response.

4 CONCLUSION

Many research papers on the indoor channel modeling have presented simulation result for WSN. Mostly, these papers are based on theory and assumptions of channel model that can lead to the miscalculated results.

In this paper, we have presented different results based on our own experiments in an indoor office environment in order to get more accurate results based on real time data and low complex channel models to feed the real time simulation. The results of the, measured data characterize the studied indoor environment through path loss, power delay profile and impulse response of the system. For both LOS and NLOS the effects are different in indoor environments. Also the path loss is on average higher in frequencies for indoor.

REFERENCES

- [1] Kaya, Aliye Özge. "Channel modeling approaches to wireless system design and analysis." PhD diss., Rutgers University-Graduate School-New Brunswick, 2010.
- [2] Akyildiz, Ian F., Tommaso Melodia, and Kaushik R. Chowdhury. "A survey on wireless multimedia sensor networks." *Computer networks* 51, no. 4 (2007): 921–960.
- [3] Wyne, Shurjeel, Amit P. Singh, Fredrik Tufvesson, and Andreas F. Molisch. "A statistical model for indoor office wireless sensor channels." *Wireless Communications, IEEE Transactions on* 8, no. 8 (2009): 4154–4164.
- [4] Szewczyk, Robert, Alan Mainwaring, Joseph Polastre, John Anderson, and David Culler. "An analysis of a large scale habitat monitoring application." In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pp. 214–226. ACM, 2004.
- [5] Ganesan, Deepak, Bhaskar Krishnamachari, Alec Woo, David Culler, Deborah Estrin, and Stephen Wicker. *Complex behavior at scale: An experimental study of low-power wireless sensor networks*. Vol. 13. Technical Report UCLA/CSD-TR 02, 2002.
- [6] Zhou, Gang, Tian He, Sudha Krishnamurthy, and John A. Stankovic. "Impact of radio irregularity on wireless sensor networks." In *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pp. 125–138. ACM, 2004.
- [7] Mangold, Stefan, Matthias Lott, Dave Evans, and Rob Fifield. "Indoor radio channel modeling-bridging from propagation details to simulation." In *Proc. of PIMRC'98*, pp. 625–629. 1998.
- [8] Bertoni, Henry L. *Radio propagation for modern wireless systems*. Pearson Education, 1999. pp. 90–92.
- [9] Andersen, Jorgen Bach, Theodore S. Rappaport, and Susumu Yoshida. "Propagation measurements and models for wireless communications

- channels.” *Communications Magazine*, IEEE 33, no. 1 (1995): 42–49.
- [10] Jain, Raj. “Channel Models A Tutorial1.” (2007).
- [11] Seidel, Scott Y., and Theodore S. Rappaport. “Site-specific propagation prediction for wireless in-building personal communication system design.” *Vehicular Technology*, IEEE Transactions on 43, no. 4 (1994): 879–891.
- [12] Gonzalez-Ruiz, Alejandro, Alireza Ghaffarkhah, and Yasamin Mostofi. “A comprehensive overview and characterization of wireless channels for networked robotic and control systems.” *Journal of Robotics* 2011 (2012).

A key management scheme for wireless sensor networks based on bridge and ECC

J. Zheng, F.F. Zhao, D.X. Wang & S.C. Xu

School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China

ABSTRACT: Providing a key suitable management protocol is a challenging task in wireless sensor network. Based on Elliptic Curve Cryptography and Bridge, a key management scheme for wireless sensor network is proposed in the paper. It adopts the bridge and hash chain technologies to achieve session keys establishment and fast authentication between nodes. This protocol also supports new nodes join and invalid nodes elimination, so it has good dynamics and extendibility. It is proved that the protocol is secure under the Elliptical Curve Discrete Logarithm Problem. The analysis result indicates that the scheme has advantages of good security, a small amount of calculation and less storage consumption.

KEYWORDS: Wireless sensor networks, bridge, one-way hash chain, session key establishment.

1 INTRODUCTION

Wireless sensor network (WSN) consists of a large number of sensors which are capable of probing the environment and reporting the collected data to the sink node. Each sensor node is highly vulnerable to many kinds of attacks, both physical and digital attacks. It makes the task of incorporating security in wireless sensor network (WSN) a challenging problem^[1]. In order to ensure data security, data must be encrypted to transmit. Thus, key management scheme becomes the technical core component in dealing with the security settling^[2].

In this paper, we propose an effective key management scheme. The paper is organized as follows: section II describes related works. Section III presents the preliminary knowledge of our scheme. Section IV describes our scheme in detail. Section V describes correctness, security and performance analysis of our scheme. Finally, we summarize our results in Section VI.

2 RELATED WORKS

Recently, many scholars have proposed a variety of key management schemes. Eschenauer and Gligor proposed the basic Random Key Pre-distribution (RKP) scheme in the literature [3]. In the scheme, each sensor node is pre-distributed a set of random keys from a key pool before deployment, and the random pre-distribution process makes any two nodes share a common key on a certain probability. Based on the scheme, Chan et al. developed two key pre-distribution

schemes in the literature^[4]: Q-composite scheme and a session key pre-distribution scheme. Both of these schemes enhance the resilience of the RKP scheme. However, in the former scheme, as the number of compromised nodes increases, the fraction of affected session keys increases quickly^[5], and in the latter scheme, the number of the stored keys linearly increases depending on the network size^[6]. Du et al.^[7] proposed a key pre-distribution scheme with a definite node compromise threshold λ . The scheme exhibits a nice threshold property: when the number of compromised nodes is less than the threshold λ , the probability that any other nodes are affected is near to zero. Blundo et al.^[8] proposed a polynomial-based key pre-distribution scheme, which distributes a polynomial share to each sensor node. In this scheme, every pair of nodes can generate a link key. Symmetric polynomial $f(x, y)$ of degree λ is used in literature [8]. Thus, the solution is λ -secure, which means that a coalition of less than $\lambda + 1$ sensor nodes knows nothing about others' session keys.

Existing schemes which addressed the network resiliency either introduce an important storage and communication overhead or degrade the secure connectivity coverage. In contrast to these solutions, our goal is to enhance the resiliency of WSN key management schemes without introducing any new storage or communication overheads while maintaining high secure connectivity coverage.

In this paper, we combine elliptic curve cryptography with the bridge to form an effective key management scheme, which makes sensor nodes faster and more secure to transmit information. Meanwhile, the application of the time slice mechanism makes

sensor nodes more efficiently to prevent the cipher text analysis. Lastly, we design a dynamic key update technique, which significantly reduces the communication overhead in the phase of key establishment and updating.

3 RELATED KNOWLEDGE

3.1 Related mathematical problems

Our scheme uses the following mathematical problems. Let G_1 be a cyclic additive group with prime order q .

1 Discrete Logarithm Problem(DLP)

Given two elements $P, Q \in G_1$, to find an integer $n \in \mathbb{Z}_q^*$, such that $Q = nP$.

2 Elliptical Curve Discrete Logarithm Problem (ECDLP)

An elliptic curve, defined modulo a prime p is a set solution for (x, y) to an equation of the form $y^2 \equiv x^3 + ax + b \pmod{p}$ where a and b are integers. The elliptic curve discrete logarithm problem is as follows:

Determining the base point P is on the elliptic curve. Let kP represent the point P added to it k -times, where k is an integer. Suppose another point Q is a multiple of P , such as $Q = kP$ for some integer $k > 0$. Then the elliptic curve discrete logarithm problem is to determine k given P and Q ^[9].

3.2 One-way hash chain

One-way hash chain's format as it is shown in formula 1, where v_0 is the seed variable, $H(*)$ is a one-way hash function, and the hash chain's length is long enough, e.g.100. We guarantee each variable in the hash chain is different from each other.

$$v_n = H(v_{n-1}) = H^2(v_{n-2}) \cdots = H^{n-1}(v_1) = H^n(v_0) \quad (1)$$

In our scheme, we utilize time slice mechanism as follows: We divide the time into a series of time intervals. After sensor nodes deployment, the time interval is S_0 . In the time interval S_0 , we use v_0 as an input variable, and in the time interval S_i , we use v_i as the input variable. When the time interval exceeds the hash chain's length, we use v_0 as the input again. Thus, we can ensure the key session in each communication link is different from time to time during every time interval^[10].

3.3 Bridge

All the variables are sorted as what Fig.1 shows. The format is called a bridge. Each column (e.g. $a_i, b_i, c_i, \dots, h_i (1 \leq i \leq n)$) is a road. Each row

(e.g. $l_1, l_2, \dots, l_n (l \in \{a, b, \dots, h\})$) is a segment. From left to right the serial numbers of the roads are the first, the second, etc. (e.g. $road_1, road_2, \dots, road_n$). From the top to down the serial numbers of the segments are the first, the second, etc. (e.g. $segment_1, segment_2, \dots, segment_n$). Each variable surrounded by roads and segments is called the bridge element (e.g. q_1), which connects different pieces of the roads (e.g. $road_1, road_2$) and divides roads into several segments (e.g. $segment_1, segment_2$). The basic bridge is a bridge of two roads and two segments. Its format is showed in Fig.2.

The calculation rule of the bridge is to compute each roads variables multiplication (e.g. $a_1 b_1, a_2 b_2$); multiplication of the bridge elements (e.g. q_1) and variables (e.g. a_1, b_2) on different roads and segments. The bridge polynomial is the sum of those multiplication results^[11], e.g. the basic bridge polynomial expansion is:

$$\begin{vmatrix} a_1 & a_2 \\ p \\ b_1 & b_2 \end{vmatrix} = a_1 b_1 + a_1 p b_2 + a_2 p b_1 + a_2 b_2 \quad (2)$$

$$\begin{vmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ q_1 & q_2 & \cdots & q_{n-1} & \\ b_1 & b_2 & b_3 & \cdots & b_n \\ r_1 & r_2 & \cdots & r_{n-1} & \\ c_1 & c_2 & c_3 & \cdots & c_n \\ & & \cdots & & \\ v_1 & v_2 & \cdots & v_{n-1} & \\ h_1 & h_2 & h_3 & \cdots & h_n \end{vmatrix}$$

Figure 1. Bridge.

$$\begin{vmatrix} a_1 & a_2 \\ p \\ b_1 & b_2 \end{vmatrix}$$

Figure 2. The basic bridge.

4 PROPOSED KEY MANAGEMENT SCHEME

Before introducing our scheme, we must make some assumptions:

Assumptions 1: Each sensor node is capable of determining its location by using GPS.

Assumptions 2: Each sensor node updates session keys at each random time interval and the time interval is determined by the base station (BS).

Assumptions 3: Before capturing a sensor node, or cracking its key, we can update the key in time.

There are some basic notations we will use in our scheme. We list them in the form of table as below:

Table 1. Notations.

| Symbol | Specification |
|---|--|
| k_{init} | The system initial key |
| $H_1 : \{0,1\}^* \rightarrow Z_q^*$ | Mapping from arbitrary-length strings to a value in Z_q^* , be used in generating derivative keys |
| $H_2 : \{0,1\}^* \rightarrow \{0,1\}^n$ | Mapping from arbitrary-length strings to n-length strings, be used in generating the Message Authentication Code (MAC) |
| G | The base point on the elliptic curve |
| ID_u | The sensor node u 's identity |
| SK_u | A sensor node u 's private key |
| $PK_u = SK_u \cdot G$ | A sensor node u 's public key |
| (x_u, y_u) | The coordinate of a node u |
| S_{Max} | A sensor node's maximum survival time interval |
| $E_x(M)$ | Encrypt the message M with the key K |
| $D_x(M)$ | Decrypt the message M with the key K |

4.1 Network initialization

Before the network deployment, each sensor node is initially preloaded with a random and unique number as its identity, two one-way hash functions H_1 and H_2 , a distinct private key and a system initial key K . All the sensor nodes have the same system initial key K . The BS selects a random symmetric polynomial ^[12] $f(x, y)$ of degree t with coefficients over a finite field F_q . Each CH node (the head node of the cluster) will get its personal secret (e.g. a CH node u will get its personal secret $f(ID_u, y)$) and the group key. Each CH node shares a group key with the BS. We refer the literature [13] to calculate the group key.

4.2 Cluster formation phase

We adopt cluster head election algorithm proposed in the literature [14] to choose cluster heads. In the cluster formation phase, each CH node records the other

CH nodes' identities and transfers its identity to the BS. The BS sends the corresponding personal secret and the group key to the CH node.

Each CH node obtains its location and then broadcasts a 'HELLO' message to the nearby sensor nodes. Each sensor node joins the nearest cluster. If there is a sensor node which doesn't receive any 'HELLO' message during the phase, it will actively look for a neighboring CH node and join it. After the phase, each CH node records the identity of its members [15].

4.2.1 Neighborhood discovery phase

After the cluster formation phase, the phase of neighborhood discovery begins. The detailed description is as follows:

- 1 Firstly, a sensor node u initializes a timer, then computes and broadcasts the message BM_u :

$$BM_u = E_{k_{init}} \left(\underbrace{ID_u}_{M_u} \parallel \underbrace{H_2(k_{init} \parallel ID_u \parallel k_{init})}_{M'_u} \right)$$

- 2 After receiving the message BM_u , a neighboring node v responds with an acknowledgment (ACK) message $AckM_v$:

$$AckM_v = E_{k_{init}} \left(\underbrace{ID_v \parallel (x_v, y_v)}_{M_v} \parallel \underbrace{H_2(k_{init} \parallel ID_u \parallel (x_v, y_v) \parallel k_{init})}_{M'_v} \right)$$

- 3 The node u 's message authentication process is as follows:

Step1: The node u decrypts the message $AckM_v$ using the system key k_{init} , and then gets two messages M_v and M'_v .

Step2: The node u computes $M_v'' = H_2(M_v)$, then checks if the equation $M_v'' = M'_v$ holds. If the answer is positive, the authentication succeeds. Otherwise, it fails.

- 4 If the node u 's message authentication succeeds, it calculates an angle θ among the BS, the node v and itself. Fig.3 describes the calculation diagram, where the node u is the origin of the coordinates.

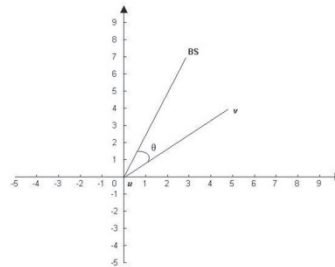


Figure 3. Angle of calculating diagram.

5 The node u checks whether $\theta \leq 45^\circ$. If so, it records the node v 's identity ID_v . It aims to reduce memory consumption and make sensor nodes transfer information faster. When the timer runs out, if there is still a sensor node which doesn't have any neighborhood, we can appropriately lower the limitation of the angle.

4.3 Key establishment

4.3.1 Intra-cluster session key establishment

After the network deployment, the time interval is S_0 . In the time interval, each sensor node can establish session keys with its neighbors. During the phase, the BS broadcasts each cluster members' public keys and the base point G . The description of session key establishment is as follows:

1 A sensor node u firstly computes the equation:

$$K_0^u = H_1(k_{init} \parallel SK_u)$$

, then computes and broadcasts the message M_u^v :

$$M_u^v = E_{k_{init}} \left(\underbrace{ID_u \parallel K_0^u}_{\tilde{r}_u} \parallel \underbrace{H_2(k_{init} \parallel ID_u \parallel K_0^u \parallel k_{init})}_{\tilde{r}_u} \right)$$

2 After received the message M_u^v , then a neighboring node computes the equation:

$$K_0^v = H_1(k_{init} \parallel SK_v)$$

, and then responses with an ACK message $AckM_v^u$:

$$AckM_v^u = E_{k_{init}} \left(\underbrace{ID_v \parallel K_0^v}_{\tilde{r}_v} \parallel \underbrace{H_2(k_{init} \parallel ID_v \parallel K_0^v \parallel k_{init})}_{\tilde{r}_v} \right)$$

3 If the message authentication succeeds, the two nodes calculate the session key shared between them.

The message authentication process is similar to the neighborhood discovery process. So we omit it. The detailed calculation steps are as follows:

Step1: The node u calculates the key k_w .

$$\begin{aligned} k_w &= \begin{vmatrix} H_1(ID_u)G & PK_u \\ K_0^u K_0^v & \\ H_1(ID_v) & SK_v \end{vmatrix} \\ &= H_1(ID_u)H_1(ID_v)G + H_1(ID_u)K_0^u K_0^v SK_v G + H_1(ID_u)K_0^u K_0^v PK_v + SK_u PK_v \\ &= H_1(ID_u)H_1(ID_v)G + H_1(ID_u)K_0^u K_0^v SK_v G + H_1(ID_u)K_0^u K_0^v PK_v + SK_u SK_v G \end{aligned}$$

Step2: The node v calculates the key k_w .

$$\begin{aligned} k_w &= \begin{vmatrix} H_1(ID_u)G & PK_u \\ K_0^u K_0^v & \\ H_1(ID_v) & SK_v \end{vmatrix} \\ &= H_1(ID_v)H_1(ID_u)G + H_1(ID_v)K_0^u K_0^v SK_u G + H_1(ID_v)K_0^u K_0^v PK_u + SK_v PK_u \\ &= H_1(ID_v)H_1(ID_u)G + H_1(ID_v)K_0^u K_0^v SK_u G + H_1(ID_v)K_0^u K_0^v PK_u + SK_v SK_u G \end{aligned}$$

4 After established session keys, then each sensor node computes the equation:

$$k_{init}^0 = H_2(k_{init})$$

, and then erases the old value k_{init} . In the next time interval, each sensor takes the key k_{init}^0 as the system key. Session keys establishment between CH node and sensor nodes is similar to the intra-cluster, so we omit it.

4.3.2 Inter-cluster session key establishment

After the network deployment, the time interval is S_0 . In the time interval, each CH node can establish session keys with the other CH nodes. During the phase, the BS broadcasts the base point G . The description of session key establishment is as follows:

1 A CH node u randomly selects an ephemeral value $t_u \in Z_q^*$, then computes and broadcasts the message CH_u :

$$CH_u = E_{k_{init}} \left(\underbrace{ID_u \parallel t_u G}_{M_u} \parallel \underbrace{H_2(k_{init} \parallel ID_u \parallel t_u G \parallel k_{init})}_{M_u} \right)$$

2 After receiving the message CH_u , another CH node v also randomly selects an ephemeral value $t_v \in Z_q^*$, and then responses with an ACK message $ACKCH_v^u$:

$$ACKCH_v^u = E_{k_{init}} \left(\underbrace{ID_v \parallel t_v G}_{M_v} \parallel \underbrace{H_2(k_{init} \parallel ID_v \parallel t_v G \parallel k_{init})}_{M_v} \right)$$

3 After the message authentication succeeds, the two nodes calculate the session key shared between them. The detailed calculation steps are as follows:

Step1: The node u calculates the key k_{ch_u, ch_v} :

$$k_{ch_u, ch_v} = f(ID_u, ID_v) t_u t_v G$$

Step2: The node v calculates the key k_{ch_u, ch_v} :

$$k_{ch_u, ch_v} = f(ID_v, ID_u) t_v t_u G$$

- 4 After established session keys, then each CH node computes the equation:

$$k_{init}^0 = H_2(k_{init})$$

, and then erases the old value k_{init} . In the next time interval, each CH node takes the key k_{init}^0 as the system key.

4.4 Session key updating

If a network doesn't refresh keys periodically, it will be compromised by an attacker. Furthermore, to continue with normal network operations, the failure nodes need to be revoked, and new nodes need to be added into the network to replace failure nodes.

4.4.1 Key refreshment

In this scheme, the time of the network is split into many time intervals. On the time interval S_i , each node (e.g. a sensor node u) updates the keys K_{i-1}^u and k_{init}^{i-1} , and then erases the old value. The detailed calculation steps are as follows:

Step 1: compute $K_i^u = H_1(k_{init}^{i-1} \| SK_u)$

Step 2: compute $k_{init}^i = H_2(k_{init}^{i-1})$, and then take the key k_{init}^i as the system key.

4.4.2 Node deletion

The BS can detect failure nodes in time with the help of intrusion detection systems. The node deletion mechanism is illustrated as follows:

Step 1: The BS removes the failure nodes' keys and broadcasts a failure nodes list.

Step 2: After receiving the list, all the CH nodes redirect it to their members.

Step 3: After receiving the list, all the sensor nodes check whether the failure nodes are its neighbors. If so, they remove the nodes. Otherwise, they ignore the information.

4.4.3 Node addition

On the time interval S_i , we need to add a new node into the network (e.g. a sensor u). The node addition mechanism is illustrated as follows:

- 1 The node is preloaded with some essential data, such as the system initial key k_{init}^i , a unique identity, a private key and the other materials.
- 2 After the node development, it joins the nearest cluster and discovers its neighbors.
- 3 On the time interval S_{i+1} , the node can establish session keys with its neighbors (e.g. another

sensor node v). The detailed calculation steps are as follows:

Step1: The sensor node u computes the equation:

$$K_0^u = H_1(k_{init}^i \| SK_u)$$

, then calculates and broadcasts the message M_u^v :

$$M_u^v = E_{k_{init}^i} \left(\underbrace{ID_u \| K_0^u}_{T_u} \| \underbrace{H_2(k_{init}^i \| ID_u \| K_0^u \| k_{init}^i)}_{T_u} \right)$$

Step2: After receiving the message M_u^v , then a neighboring node v computes the equation:

$$K_0^v = H_1(k_{init}^i \| SK_v)$$

, and then responses with an ACK message $AckM_v^u$:

$$AckM_v^u = E_{k_{init}^i} \left(\underbrace{ID_v \| K_0^v}_{T_v} \| \underbrace{H_2(k_{init}^i \| ID_v \| K_0^v \| k_{init}^i)}_{T_v} \right)$$

Step 3: After the message authentication succeeds, the two nodes calculate the session key shared between them. The detailed calculation process is similar to the intra-cluster, so we omit it.

- 4 After establishing session keys, each sensor node computes the equation:

$$k_{init}^{i+1} = H_2(k_{init}^i)$$

, then erases the old value k_{init}^i . In the next time interval, each sensor node takes the key k_{init}^{i+1} as the system key.

5 ANALYSIS

5.1 Correctness analysis

The correctness proof of this paper is in the following theorems. If the calculations of the legal key negotiators are correct, it can negotiate a consistent session key.

Theorem1. The key agreement protocol in Intra-cluster satisfies the correctness. Each pair of sensor nodes (e.g. two sensor nodes u and v) can negotiate a consistent session key in this protocol, which means the following equation holds ($0 \leq i \leq S_{Max}$):

$$\left| \begin{array}{cc} H_1(ID_v)G & PK_v \\ K_i^u K_i^v & \\ H_1(ID_u) & SK_u \end{array} \right| = \left| \begin{array}{cc} H_1(ID_u)G & PK_u \\ K_i^u K_i^v & \\ H_1(ID_v) & SK_v \end{array} \right|$$

Proof: Since $PK_u = SK_u \cdot G$, $PK_v = SK_v \cdot G$ and the properties of the bridge and elliptic curve, we have:

Left:

$$\begin{aligned} & \begin{bmatrix} H_1(ID_v)G \quad PK_v \\ K_v^u K_v^v \\ H_1(ID_u) \quad SK_u \end{bmatrix} \\ &= H_1(ID_v)H_1(ID_u)G + H_1(ID_v)K_v^u K_v^v SK_u G + H_1(ID_u)K_v^u K_v^v PK_v + SK_u PK_v \\ &= H_1(ID_v)H_1(ID_u)G + H_1(ID_v)K_v^u K_v^v PK_u + H_1(ID_u)K_v^u K_v^v PK_v + SK_v SK_u G \\ &= k_{uv} \end{aligned}$$

Right:

$$\begin{aligned} & \begin{bmatrix} H_1(ID_u)G \quad PK_u \\ K_u^v K_u^u \\ H_1(ID_v) \quad SK_v \end{bmatrix} \\ &= H_1(ID_u)H_1(ID_v)G + H_1(ID_u)K_u^v K_u^u SK_v G + H_1(ID_v)K_u^v K_u^u PK_u + SK_v PK_u \\ &= H_1(ID_u)H_1(ID_v)G + H_1(ID_u)K_u^v K_u^u PK_v + H_1(ID_v)K_u^v K_u^u PK_u + SK_u SK_v G \\ &= k_{vu} \end{aligned}$$

Conclusion: $k_{uv} = k_{vu}$

By observing the above equations, we know that the nodes share the session key with each other.

Theorem2. The key agreement protocol in Inter-cluster satisfies the correctness. Each pair of CH nodes (e.g. two CH nodes u and v) can negotiate a consistent session key in this protocol, which means that the following equation holds:

$$f(ID_u, ID_v)t_u t_v G = f(ID_v, ID_u)t_v t_u G.$$

Proof: Since the properties of the symmetric bivariate polynomial and elliptic curve, we have:

$$\begin{aligned} & f(ID_u, ID_v) = f(ID_v, ID_u) \\ & \Downarrow \\ & f(ID_u, ID_v)t_u t_v G = f(ID_v, ID_u)t_v t_u G \\ & \Downarrow \\ & k_{yv} = k_{vu} \end{aligned}$$

By observing the above equations, we know that the nodes share the session key with each other.

5.2 Security analysis

We define typical attacks on WSN and demonstrate how our proposed scheme can prevent them. The typical attacks can be classified as follows: guessing attack, replaying message attack, eavesdropping attack, node capture attack and denial of service attack.

5.2.1 Guessing attack

Assuming that an attacker can get data transmitted between nodes and based on the information he maybe guess the key k_{mit}^i ($0 \leq i \leq S_{Max}$). Nevertheless, the network periodically updates the key k_{mit}^i in each time interval. Furthermore, all the messages are encrypted by session key. So, the guessing attack doesn't have any effect on the scheme.

5.2.2 Replaying message attack

If an attacker keeps sending old message, the operation will threaten message freshness. The attack doesn't have any effect on the scheme in that all the session keys are different in each time interval. Since the system key is periodically updated to generate the session keys in each time interval.

5.2.3 Eavesdropping attack

In this attack, an adversary creates connections with sensor nodes and takes over the handling of message between any two sensor nodes. The attack makes the sensor nodes believe that they directly communicate with each other. Since each node frequently updates session keys in each random time interval, the adversary can't modify the content of the message. So, the eavesdropping attack is useless for the scheme.

5.2.4 Node capture attack

After network development, each attacker maybe want to get sensor nodes' (e.g. a sensor u) keys (e.g. session keys, the system key k_{mit}^{i-1} , private key SK_u and the key K_i^u ($0 \leq i \leq S_{Max}$, $k_{mit}^{-1} = k_{mit}$). In the scheme, after establishing session keys, each sensor node erases the obsolete element k_{mit}^{i-1} . When the attacker captures a sensor node, he can't construct session keys by using the key k_{mit}^i . What is more, the hash function H_1 is one-way hash function, and there is scarce polynomial time algorithm in solving the elliptical curve discrete logarithm problem (ECDLP). So, if an attacker captures one sensor, the attack still can't infect the others.

5.2.5 Denial-of-service attack

If an attacker keeps sending illegal data to a valid node in order to make energy consumption and cause network paralysis, in the scheme, we utilize one-way function to generate message authentication code (MAC) to detect illegal packet.

5.3 Performance analysis

We mainly discuss cost in message authentication and session key agreement phase, including

communication, computation and storage. The performance comparison among the protocols of DUBKM^[16], HKMS^[17] and the proposed scheme is as follows:

5.3.1 Communication cost

In the scheme, each node needs to exchange two messages to establish a session key. One is a broadcast message; the other is an ACK message. While, in the scheme HKMS^[17] each node must exchange three messages. In the scheme DUBKM^[16] each node only needs to exchange two messages, but its broadcast message contains the node's identity, the node's location and a neighbor list. In the scheme, the broadcast message only consists of a node's identity (e.g. a node u 's identity ID_u) and its key K_i^u ($0 \leq i \leq S_{Max}$). Compared with the scheme DUBKM^[16], our proposed scheme uses a smaller message.

In the key refreshment phase, the proposed scheme doesn't need to transmit any information to update session keys. While, the other schemes we mentioned above need to transmit information. Therefore, our scheme better reduces communication cost. The communication cost in key establishment and refreshment phase is showed in Fig.4.

5.3.2 Storage cost

Before the network development, each node is preloaded with the system initial key and its private key. We assume that the numbers of CH nodes and sensor nodes are respectively N_{ch} and N_s . So the total number of preloaded keys is $2N_{ch} + 2N_s$. In the scheme DUBKM^[16], each node is also preloaded with the initial key and its private key.

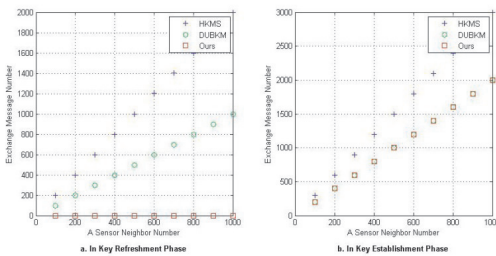


Figure 4. Communication cost.

5.3.3 Computation cost

Main calculation operations for key maintenance are scalar multiplications on an elliptic curve, multiplication, addition and hash operation. Table 2 shows the computation cost of session key establishment phase. In the scheme, each sensor node needs to calculate six multiplications, two hash operations, three

numbers of scalar additions, and four numbers of scalar multiplications.

Observing the table 2, the scheme HKMS^[17] introduces a high computation cost of multiplication and addition operations. Meanwhile, the table shows that our scheme introduces more computation cost compared with the scheme DUBKM^[16]. The cost is reasonable as our scheme provides a high-level security and a more comprehensive key management scheme including key refreshment, key deletion and new node addition phase.

Table 2. Computation cost comparison.

| | DUBKM | HKMS | Ours |
|-----------------------|-------|--------------------|------|
| Multiplication | 0 | $3t^3 + 6t^2 + 3t$ | 6 |
| Addition | 0 | $t^3 + 3t^2 + 3t$ | 0 |
| Hash Operation | 4 | 1 | 2 |
| Scalar Addition | 0 | 0 | 3 |
| Scalar Multiplication | 0 | 0 | 4 |
| Key Refreshment | Yes | Yes | Yes |
| Key Deletion | No | Yes | Yes |
| New Node Addition | Yes | No | Yes |

t : a random symmetric bivariate polynomial is t-degree

6 CONCLUSIONS

In this paper, we have proposed a novel key management mechanism, which utilizes elliptic curve cryptography and the bridge. We combine one-way hash chain with time slice mechanism to design a mechanism. The mechanism significantly reduces communication overhead in key establishment and refreshment phase. The analysis section shows that our scheme has some advantages in communication, storage and computation cost. Furthermore, our scheme provides perfect scalability and resilience against node capture.

ACKNOWLEDGMENTS

This paper is supported by the National Natural Science Foundation of China (No.61272511).

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

REFERENCES

- [1] Deng Jing, Han Richard, Mishra Shivakant. INSENS: Intrusion-tolerant routing for wireless sensor

- networks[J]. *Computer Communications*, 2006, 29(2): 216–230.
- [2] Ma X, Dong Z, Li J. A Novel Key Management Scheme for Wireless Sensor Networks[C]. *Internet Computing for Science and Engineering (ICICSE)*, 2012 Sixth International Conference on: IEEE, 2012: 198–203.
- [3] Eschenauer L, Gligor V D. A key-management scheme for distributed sensor networks[C]. *Proceedings of the 9th ACM conference on Computer and communications security*. ACM, 2002: 41–47.
- [4] Chan H, Perrig A, Song D. Random key predistribution schemes for sensor networks[C]. *Security and Privacy, 2003. Proceedings. 2003 Symposium on*. IEEE, 2003: 197–213.
- [5] Zhao Huawei, Liu Ruixia. Hierarchy Cluster Model and Key Management in Wireless Sensor Network[C]. *Wireless Communications, Networking and Mobile Computing, 2009. WiCom'09. 5th International Conference on*. IEEE, 2009: 1–4.
- [6] Bechkit W, Challal Y, Bouabdallah A, et al. A highly scalable key pre-distribution scheme for wireless sensor networks[J]. 2013.
- [7] Du W, Deng J, Han Y S, et al. A pairwise key predistribution scheme for wireless sensor networks[J]. *ACM Transactions on Information and System Security (TISSEC)*, 2005, 8(2):228–258.
- [8] Blundo C, De Santis A, Herzberg A, et al. Perfectly-secure key distribution for dynamic conferences[C]. *Advances in cryptology CRYPTO92*. Springer Berlin Heidelberg, 1993:471–486.
- [9] Udin M N, Halim S A, Jayes M I, et al. Application of message embedding technique in ElGamal Elliptic Curve Cryptosystem[C]. *Statistics in Science, Business, and Engineering (ICSSBE)*, 2012 International Conference on. IEEE, 2012: 1–6.
- [10] Zhou R, Yang H. A hybrid key management scheme for Heterogeneous wireless sensor networks based on ECC and trivariate symmetric polynomial[C]. *Uncertainty Reasoning and Knowledge Engineering (URKE)*, 2011 International Conference on: IEEE, 2011, 1: 251–255.
- [11] Shen Xiaofeng, Yu Lan, Shen Yu. *Logic algebra* [M]. Beijing: Science press, 2008: 109–123. (in Chinese)
- [12] Wang L, Jiang T, Zhu X. Updatable key management scheme with intrusion tolerance for unattended wireless sensor network[C]. *Global Telecommunications Conference (GLOBECOM 2011)*, 2011 IEEE: IEEE, 2011:1–5.
- [13] Tian B, Han S, Parvin S, et al. A key management protocol for multiphase hierarchical wireless sensor networks[C]. *Embedded and Ubiquitous Computing (EUC)*, 2010 IEEE/IFIP 8th International Conference on: IEEE, 2010: 617–623.
- [14] Kim K B, Zhang Y Y, Yang W C, et al. An authentication protocol for hierarchy-based wireless sensor networks[C]. *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*: IEEE, 2008:1–6.
- [15] Boujelben M, Cheikhrouhou O, Abid M, et al. A pairing identity based key management protocol for heterogeneous wireless sensor networks[C]. *Network and Service Security, 2009. N2S'09. International Conference on*: IEEE, 2009: 1–5.
- [16] Rao Y, Wang J, Tian R, et al. Dynamic updating based key management algorithm for wireless sensor networks[C]. *Wireless Communications and Signal Processing (WCSP)*, 2011 International Conference on: IEEE, 2011: 1–5.
- [17] Zhou R, Yang H. A hybrid key management scheme for Heterogeneous wireless sensor networks based on ECC and trivariate symmetric polynomial[C]. *Uncertainty Reasoning and Knowledge Engineering (URKE)*, 2011 International Conference on: IEEE. 2011, 1: 251–255.

Proposal of a universal model for managing research and development center

J. Spišák, M. Reisová, E. Oravcová & S. Pořáková

Development and realization workplace of raw materials extracting and treatment, Faculty of mining, ecology, process control and geotechnology, Technical University of Kosice, Kosice, Slovak Republic

ABSTRACT: Economic growth and progress currently requires innovation and implementation support focused on obtaining financial resources for their implementation and effective use of the so-called good practice from innovative leaders. Many businesses are now to concentrate them only on cost reduction strategy, which is by itself not prospective. Competitiveness must provide properly selected combination of innovation and methods leading to rationalize the organization of work, which will lead to increased productivity and efficiency. Therefore study submitted includes a set of useful knowledge not only about the innovation process, but also about the field of project management, logistics, innovative organizational forms and process approach of the research organization to ensure its sustainable development.

1 INTRODUCTION

The need proposes an effective management system of research and development department, which fulfills its objectives in the field of raw material extraction and treatment through the scientific solutions, infrastructural and implementation projects and which has a number of conflicting specifics as compared to standard production workplaces, is highly current one. Such a workplace with a wide area of interest has its own unique characteristics that distinguish it from other organizations and it is necessary to be taken into account in the process of preparation and implementation of innovative projects and in the formation of innovations and new approaches at all. Therefore, for the management of research and development department whose activities are carried out through the disposable VVI projects, it is necessary to project management tools extended by the new application of the logistics principles, which enables parallel ensuring of the efficient operation of the workplace and is carefully applying to management system in the specific environment of the development department. The proposed management system must therefore use not only the tools and methods of project management, but also the logistics tools, knowledge from the field of innovations, work organization and algorithmization.

1.1 Methodology

The management of department implements innovative projects, which are not tools and techniques of

project management sufficient since it does not take into account the environment in which the workplace operates. Workplace never works in isolation, its operation is always the result of operation of environment and internal structures that are interconnected by linkages [1]. The management of the workplace must rest in the coordination of projects, in the ability of variations and constant evaluation and updating of the model situation, in parallel with the changing real conditions. Therefore, the workplace that seeks to project preparation in the highest possible quality in the shortest time with the maximum use of available capacities must manage tools necessarily extended to a logistics tools and draw on the number of advantages that brings the knowledge of micro and macro environment in which the project is implemented [2].

1.2 Proposal of a universal model for managing research and development center

For streamlining the management of innovation projects can be project-oriented organizations based on these recommendations [3]:

- 1 The formation of the process model - based on the knowledge of regularities of processes that relate not only to the preparation and implementation of the project, but run across the workplace, it is possible to create a process model that must be implemented as an open, allowing his immediate adjustment, adaptation to new conditions and numerous modifications. Creating a process model of the workplace should be based on five known postulates of logistics (system approach,

coordination, planning, algorithmic thinking, global optimization), taking into account regularities (micro, macro) of the workplace environment;

- 2 The learning organization - each organization should have a built-in strategy of continuous learning: evaluation of the results and analysis of good and faulty decisions. Subsequently, it is necessary to transform the results of this analysis into improved processes, newly defined qualities and so on;
- 3 Stand - by organization accepted standards simplifies not only routine operations, but properly created a standard procedure also creates from the complex activities a routine. Standards can make transparent all defined and de-scribed processes, accelerate and overall streamline the communication inside the organization, but also outwardly;
- 4 Shortening of the innovation cycle - organizations should look the ways how to shorten the stages of the innovation process, to find the ways how to create such an organizational ensuring, which will lead to the maximum overlapping of the individual phases of the innovation process in time. Organizations, which can create such support, may offer innovations in the market for the position of the first seller. It can determine the sectoral standards and acquire the possibility to occupy the most preferred and strategically most important markets;
- 5 Virtual structure - not to be afraid for key innovative projects to use the option of virtual structures, temporary network of organizations respectively their components, for example the common research centers concluded on the basis of cooperation contracts;
- 6 Organizational structure - organizational structure adapts to the project and not conversely. It must be dynamic, must support project-oriented environment, which has a high degree of flexibility, to allow to solve complex innovation projects with the maximum use of available capacity, at minimum cost and time, which is optimized for the organization;
- 7 Exploiting the potential of innovation project for the development of the entire organization - projects suffer the so called tunnel vision, because organizations often focus only on the achievement of its limited objectives, taking not into account the comprehensive benefit for the entire organization. The added value of the project can be created in the planning stage. Conceived the project so that its implementation has generated another opportunity for a new project, but contributes to the development of the workplace.

2 DEVELOPMENT – REALIZATION WORKPLACE OF RAW MATERIALS EXTRACTION AND TREATMENT

In the workplace in terms of the method of preparation and implementation are two types of projects. The projects of applied research to characterize the formation of solution from top to bottom, which is also typical for the framework program projects. For this type it provides at first strategy and concrete solutions are brought up with project implementation. In terms of the risk prevention, workplace can be based on a rich database of experience in the implementation of similar projects.

The second, a rapidly growing group of projects implemented in the workplace are the projects for practice. In these projects, the planning takes place in the form of bottom-up, i.e. is finding a solution of the specific problem. For this type risks and deficiencies may not appear immediately. Some of them will be identified only during the project solution.

The current structure of cooperation with project partners is centralized and is characterized by direct ties to the management workplace (VRP), which provides their interconnection and simultaneously coordinates their activity [4]. This model can be currently considered obsolete and the workplace passes to the new structure by creating a network of participating workplaces (see Fig. 1).

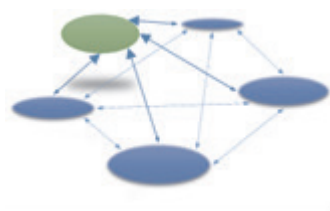


Figure 1. Organizational structure – network.

Irrespective of the way of the project solutions (from top to bottom or bottom-up) at the workplace identified three main factors that most influence the success of the implemented projects:

- 1 Infrastructure - suggests, how is the organization/workplace structured, but also organized,
- 2 System - what system the organization/workplace used to support innovative projects, processes and processes for ensuring the operation and development of the whole workplace,
- 3 Culture - which type of culture predominates in the organization, including communication, informal relations inside the workplace and also outwards.

From the perspective of the preparation and management process of innovative projects it is essential that the workplace promotes the creation of an environment that will support innovation activities.

2.3 Structure and parameters of the workplace management system

For planning and management of all activities of the workplace was created by its own management system, which is based on a thorough knowledge of micro and macro model, takes into account the specific position of the workplace and respecting the legislative conditions in which the workplace operates. A model for the creation of the system was the process model of the workplace.

The management system of the workplace, in visual form represented by four-side pyramid (see Fig. 2), has a simple structure. Top of the pyramid forms the vision of the workplace. Its fulfillment is divided into four sides of the pyramid, i.e. into four basic fields of the workplace operation:

- research, development and innovation,
- transfer of the VVI results into practice,
- transfer of the VVI results in education,
- infrastructure of the VRP.

In each area of operation are identified activities and processes carried out when it created algorithms for their planning, implementation and control with a clear attribution of responsibility for their implementation. From the perspective of the pyramid, it is possible for each pillar to pursue three levels:

- 1 Strategic, which includes vision and strategic planning (shown in the 3D model by purple color),
- 2 Tactical, which includes the area of tactical planning, annual plans and their evaluation (in the 3D model displayed by 4 colors - according to the four subject areas,
- 3 Operational, where the area is represented by the monthly and weekly plans of each department workplace.

2.4 Methodology for the preparation and management of innovative projects in the workplace

Innovative project management methodology is based on the management system of work. One of the reasons, why it was within an established management system necessary to create a methodology to streamline the process of preparation and implementation of innovative projects, is the complexity of these processes. Although the solution is usually divided into partial goals, represented mini projects with their

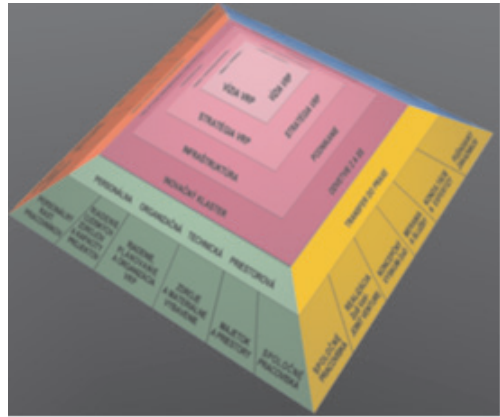


Figure 2. Hierarchical process model of the system structure.

own procedures and goals too high number of solvers, different culture in understanding and solving problems at length and content targeting create space for a possible disorganization and inconsistency. It appropriately chose organizational support which can clarify this whole process and lead the project team effectively without wasting time, finances and other resources.

Fig. 3 shows the sequence of steps for an innovative project designed by the Structural funds, the agency ASFEU, for the Operational Program Research and Development. The proposed methodology, based on a process approach, is specific in that it does not distribute the project processes into key and ancillary, does not give more weight to real fulfillment as a formal site. It understands these two areas as part of one whole. Follow the structure of forming by the element that has been identified in the preparation of projects and solutions based on a thorough knowledge of the processes of innovation projects.

3 CONCLUSIONS

Developed management system of the workplace is simple - the execution of processes corresponds to the structure, workloads are clearly assigned, it is the transparent and very easily controlled, substantially eliminates bureaucracy and is implemented as an open - if needed it will prompt innovation precisely those workers who need to upgrade. Use a number range of support files, number of documents, models and experiences. This will create ideal conditions for enhancing the quality and productivity of work will perform also in the future. Introducing a process approach in the workplace, resulting in the creation

of a process model of the workplace, was these improvements:

- 1 The processes, identification of the workplace,
- 2 The precise definition of links between processes,
- 3 The allocating process to owners and clarifications to the responsibilities and competences,
- 4 Create a comprehensive overview of the activities carried out by the department.

The elimination of duplication of activities carried out,

- 1 The coverage of unsecured actions,
- 2 The creation of substitutability,
- 3 To ensure duplicate checks,
- 4 Clear definition of the job description,
- 5 Streamline administrative work in the workplace.

By clear attribution of responsibilities and defined sequence of steps resulting from the interactive workload of all employees, there were the following positive aspects the shortening of the activities performed, the shortening of communication and information channels, the efficiency of the action was maximized, the maximum capacity utilization of workplace. The results achieved by workplace represent [Table 1](#).

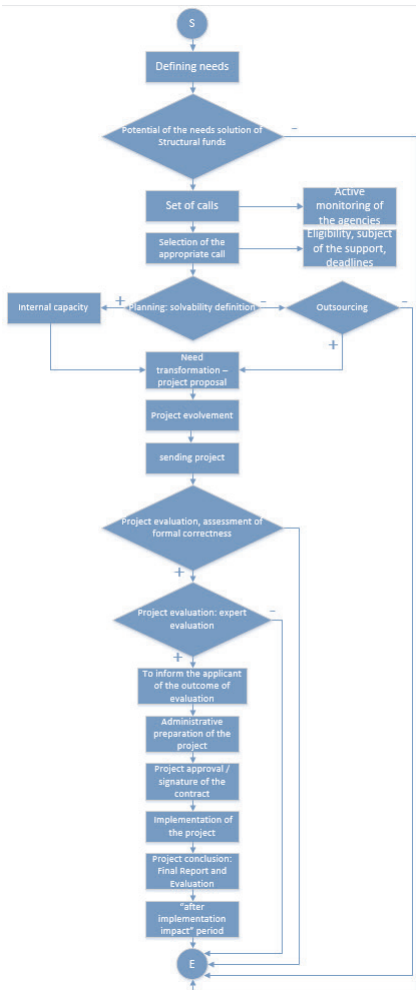


Figure 3. Sequence of steps for an innovative project designed by the Structural funds, the agency ASFEU, for the Operational Program Research and Development.

Clarifications on labor organization

As a result of the introduction of a process model in the workplace, it has been observed changed in the capacity of which was the following improvements:

Table 1. The workplace success of innovative projects.

| Project type | submitted | approved |
|--|-----------|-----------|
| International projects | 3 | 3 |
| Structural projects | 8 | 8 |
| APVV projects | 6 | 6 |
| Projects of the Ministry of Education Basic Research | 4 | 4 |
| Projects of the Ministry of Education Applied Research | 6 | 6 |
| Projects of the Ministry of Economy | 1 | 1 |
| Projects for practice | 29 | 29 |
| Total | 57 | 57 |

VRP's vision to become a national research and innovation cluster in the area of raw materials extraction and treatment to effectively provide a comprehensive and efficient research and development and innovative services, both within the Slovak and European market.

ACKNOWLEDGMENT

This contribution/publication is the result of the project implementation Advanced technology for the mining company of 21st century supported by the Research & Development Operational Program funded by the ERDF. (ITMS: 26220220131)

REFERENCES

- [1] Zelko, M., Husarova, M., 2009: Principles of successful project management concepts digital enterprise to the thermal treatment of mineral re-sources. *Transport and Logistics*, 6, pp. 272–276.
- [2] Husarova, M., Kostial, I., Lisuch, J., 2011. In-formation, Dissemination and Knowledge support for Centre of excellent research project results exploitation digital fabric: In: *Advanced technology in the acquisition and processing of raw materials*, Technical university of Kosice, pp. 35–39.
- [3] Lavrin, A., Husarová, M., Lisuch, J., 2011. Integrated system of information, knowledge and dissemination services. In: *Advanced technology in the acquisition and processing of raw materials*, Technical university of Kosice, pp. 142–146.
- [4] Husarova, M., Spišák, J., 2013. Management system for streamlining the innovation process of re-search and development workplace. In: *European Scientific Journal*, 9, pp. 38–46.

Research on human resources complex standardization system of large state-owned enterprises

L. Zhao

China Institute of Water Resources and Hydropower Research, China

ABSTRACT: In this paper, the influence of complex standardization method on technical standards is used to study human resources management standardization, analyze the status quo and existing problems of HR management of domestic large state-owned enterprises, put forward the principles on structuring HR complex standardization system and propose the solution to possible questions in the course of establishing complex standardization system with the aim to materialize the HR complex standardization system that meets the requirement of the enterprise's HR strategy through structuring it.

1 STATUS QUO AND EXISTING PROBLEMS OF HR MANAGEMENT OF LARGE STATE-OWNED ENTERPRISES

1.1 *Status quo of HR Management of large state-owned enterprises*

As a key pillar of the national economy, large state-owned enterprises shoulder the important mission of supporting and guaranteeing national economy and social development. Along with the rapid economic growth and increasing tense competition in recent years, large state-owned enterprises have improved HR management through deepening reform, transformation and innovation, greatly promoted economic benefits and competitiveness and made great contribution to the fast economic growth.

However, as a result of the historic planned economic system, the gap between large state-owned enterprises and foreign leading enterprises in the industry kept emerging. While dealing with external environmental changes, the improvement of HR management not matching with the enterprise's scale expansion poses a large obstacle for the enterprise's development and has already become a bottleneck for the enterprise to be good, strong, scientific and sound. Thus, it is imperative to tackle the task of making HR work better serve the enterprise development strategy.

1.2 *Existing Problems*

1.2.1 *Difficulty in HR policy-making of group enterprises*

As large state-owned enterprises expand business areas, scale and management scope, the difficulty in stipulation of HR policy, standards and system increases. Overemphasis on implementation rigidity might cause affiliated companies to lose drive and

flexibility while overemphasis on differences of affiliated companies might cause the chaos in policy making from multiple departments at the headquarters and that HR management might lose control. In the course of making HR policy and setting criteria, quantity, individual, technology tend to be focused on while quality, overall, management and coordination tend to be ignored. Different levels and business models summarize individual topics to define plans on personnel and salary and make separate policy and criteria. As a result, difficulties in meeting requirements arise while solving comprehensive problems across levels and business models. The headquarters can hardly conduct effective HR arrangement or realize the effective combination of resources and scale effect.

1.2.2 *Redundant basic management data & lack of standardization on process control*

Large state-owned enterprises usually have a large span of management, being supervised by SASAC and instructing business and operation of affiliated enterprises. Due to lack of interpretation and management requirement of all types of basic data and standardized index of basic process, the headquarters are unable to acquire actual HR data of all branch companies and may easily misunderstand, while accessing data provided by subsidiary companies as it is hard to judge the accuracy of data resources and make statistic analysis on the data mentioned above.

Meanwhile, due to lack of complete, standard and unified approval management process, HR management cannot adapt to rapid external environment changes or fast expansion requirement. HR staff at all levels may read the policy differently, thus the HR management process at different levels may lack standardization. The same business at different enterprises may differ a lot and refined management process of the enterprise may be hindered.

1.2.3 *Inadequate coordination among internal business models*

As a result of the influence of the original planned economy, large state-owned enterprises usually divide HR management work into labor organization, salary management, cadre management and education and training. Management within each business model is relatively mature. However, coordination among different business models is inadequate and HR staffs at all levels tend to focus on their own work and requirement. Inner conflict may therefore frequently arise. The HR standard without overall coordination cannot become an organic whole and play the role of being the largest fundamental support.

1.2.4 *Low management sensitivity and slow response*

One notable feature of the traditional HR management policy, standard and systems is that every policy and standard is formulated separately. All standards are established under different backgrounds with different purposes and functional requirements. Thus the entire system is in a still and separate state and can only touch upon partial production businesses in a passive manner.

1.2.5 *Lack of support from effective hr management information system*

For the moment, most large state-owned enterprises start to develop an HR management information system and some have already started to develop ERP system for the overall business processes due to the pressure caused by refined management. But due to the disunity in overall standards, large span of management, too many basic HR management, data and low internal business coordination, the system is in a state of low efficiency or of no effect. Unable to be simultaneously integrated, the basic data can only be processed through a labor statement in an ineffective way which cannot effectively support the strategic growth of the enterprise.

2 IMPLEMENTATION OF HR COMPLEX STANDARDIZATION SYSTEM IN LARGE STATE-OWNED ENTERPRISES

2.1 *The concept of complex standardization*

The complex standardization was proposed by standardization experts from the former Soviet Union in the 1960s. It is a scientific methodology and systematic theory guiding modern modernization. The concept of complex standardization given by China National Standard GB/T 12366-2009 Guideline for the work of complex standardization goes like this: to establish

standard complex and put into implementation with a systematic analytic method for a definite goal.

Judging from the definition, complex standardization is a systematic project of standardization. HR work happens to possess the application requirements of complex standardization: with a definite goal that conforms to the enterprise's strategy, application of systematic analytic methods, establishment of standards complex and putting into implementation in an all-round manner. Thus we can explore the application of complex standardization in HR management.

2.2 *How to structure hr complex standardization system*

2.2.1 *To unify the understanding of structuring HR complex standardization system*

To structure HR complex standardization system starts from people's correct understanding of the system in the first place. The goal of complex standardization is to solve specific problems and the ultimate result is embodied in the extent of problem solving. Complex standardization is a system management tool that organizes the factors of people, material, technology, management and provides guidance. It tells you who do it first and who follows, to what extent you do it, with whom you do it and how you collaborate. Theoretically speaking, the working process and every section in each process of HR management system structured based on complex standardization is science and logic. While meeting the management and analysis and prediction requirement on HR business from the executives, it can satisfy the requirement of management and actual business operation from subsidiary companies. But this structure is not a denial or rejection of traditional HR management business, but transcendence or a supplement.

2.2.2 *To integrate the variety of policies and standards & to regulate basic business processes*

In the course of structuring HR complex standardization and regulating business processes, the improvement of quality is fundamental goal and the core is to define the rights and responsibility of enterprises at all levels and the complex standardization only serves as a means to realize the goal. The business functional requirement of the above-mentioned aspects has to be taken into consideration while undertaking the design so that all policies and standards as well as business process can meet the requirement of rational orientation of HR management and control at all levels from the headquarters so that the maximum effect of the company's HR can be utilized to organically integrate flexibility and unity.

The headquarters mainly charges the responsibility of formulating the policies and standards on the entire company's HR strategic goal and all affiliated enterprises are responsible for examination and coordination of HR work. As a key link that connects the preceding and the following, the branch companies or subsidiary companies are responsible for implementing the standards established by the headquarters, conducting specific business operation and providing information support for the decision-making at the grassroots enterprises. The grassroots enterprises are the specific implementation bodies of the headquarters' HR management with a similar operational mode to non-group (individual) enterprises with business as the core. Different from the individual enterprises, the grassroots enterprises are subjected to the guidance and restriction from superior level enterprises with whom necessary information exchange is required.

2.2.3 *To improve internal coordination, systematic sensitivity and response speed*

It is necessary to strengthen complex standardization's response and service to routine HR work, give prominence to compulsory policies such as standards on resources planning and arrangement, overall efficiency and equity and performance management, enhance the work on formulating and implementing the standards on turnover staff, promotion and salary changes, solve problems of HR standards lagging behind development and inadequate coverage, ensure internal coordination between relevant standards and improve synergy of HR businesses within the entire enterprise.

Based on the HR complex standardization structuring principle mentioned above, formulating large state-owned enterprise HR standards system form can help to realize the completeness of standards within certain range. HR management standards index at different levels at grassroots enterprises should be defined and fine control of standards through the index can be realized. Thus standards particles in the HR complex standardization system can be adjusted. Along with the improvement of standards from the headquarter level to grassroots enterprises, the object of standards will become more specific and the particles smaller. Different levels should be used to show the similarity and coverage of HR management work and to improve sensitivity and response speed to specific business.

2.2.4 *To enhance application scope and depth of information system*

To further meet the objective requirement of refined management after establishment of standardization

system, it needs to adopt information means to effectively improve management efficiency, make HR information become centralized management instead of original separate management, set overall HR strategy and management regulations based on enterprise operation strategy, establish sharing and coordination mechanism to realize the sharing and unified arrangement of HR, training and performance management resources, and to exercise the enterprise's overall strength.

To cater for features such as complex organization, large number of personnel category, different authority for roles and positions in group enterprises, the system must support different authority along with different people or team, positions and roles. The functional authority needs to be assigned to operational action level and data authority needs to be assigned to field level. Meanwhile it has to materialize authority assignment at different angles such as meeting the requirement of user management, process configuration and affiliated organization management. To realize an in-depth management at multiple latitudes must realize HR development and management in the entire enterprise at different latitudes such as organizations, objects and functions so that overall optimal configuration can be carried out.

3 EFFECTS EVALUATION OF HR COMPLEX STANDARDIZATION SYSTEM

As management standards touch upon both technical part in technical standards and production relations with human factors in production management courses, the effects evaluation always remains a difficulty. The effects of standards penetrate from the beginning when effects produced to the end thus the contribution of standards is a vague factor. The overall effect is a result of combined action of multiple factors. The financial index that can be easily quantified is not direct embodiment of standards effect but an indirect one. The impact of HR management standards on the enterprise management in particular is in a dynamic changing process. Therefore, there are lots of difficulties to tackle in order to evaluate whether a system is the optimum.

Evaluation of standards effect is a subjective understanding course on the objective. To ensure the scientificity and accuracy of basic research and evaluation, the structuring must start from a systematic point based on full analysis of responsibility, rights and interests of three level responsibility entities. In accordance with the principles of combing qualitative with quantitative, long-term with short-term, internal evaluation with external evaluation, economic benefits evaluation and social benefits evaluation, the effective HR management evaluation system to

establish can evaluate the adaptability, performability and effectiveness in the three dimensions of coordination, efficiency and effects. The evaluation of coordination orients towards the internal and external coordination of the enterprise's HR management system, i.e. adaptability evaluation. The evaluation of efficiency addresses the implementation of each HR management obligations, reflecting the efficiency of HR management activities, i.e. performability evaluation. The evaluation of effects addresses the contribution of HR management to the goal realizing of the enterprises, employees' demand satisfying and goal supporting, i.e. effectiveness evaluation.

The key of standardization lies in implementation. The realizing of defined complex standardization goal depends on the scientificity and rationality of standards complex and implementation even more. Systematic and effective evaluation of HR management can not only help to understand the HR status and management, but also get to know the business operations at all levels of different affiliated enterprises. Through further analysis and diagnosis of existing problems in HR management, overall operations, continuous improvement and optimizing, efficient HR management can be maintained and competitive strength of HR can be created to support the realizing of enterprise's strategy.

4 VISION OF HR COMPLEX STANDARDIZATION OF LARGE STATE- OWNED ENTERPRISES

As a scientific method to comprehensively solve problems within a certain space and time, the complex standardization can be universally applied. To judge the complicated standards complex of large state-owned enterprises with a systematic view, the complex standardization possesses all essentials of a

system. The methodology can be effectively applied to detailed management work such as HR management and can effectively integrate relevant management policies and standards. Through sorting out HR standards of large state-owned enterprises and internal HR business processes, HR complex standard system model is designed to form an HR complex standardization system to meet the objective requirement of fast-changing internal and external environment on HR management and improve HR management standard system from individual level to a systematic level and thus realize the qualitative leap on HR management level.

Confronted with ever-growing intense, competitive internal and external market environment, large state-owned enterprises have become aware of the fundamental, overall and guiding role of HR as the most important assets with most vitality and creativity to the enterprise's growth. The future development tendency of the HR management standard system is to start from the most front-end of specific businesses, center on defined specific goal, structure a set of HR management complex standardization system with comprehensiveness, good coordination and high sensitivity that can realize dynamic adjustment according to the internal and external environment, improve internal operation efficiency and work quality and thus enhance the enterprise's competitive edge.

REFERENCES

- [1] ISO. Assessing Economic Benefits of Consensus-Based Standards-The ISO Methodology [R]. 2010
- [2] Beuth Verlag. Economic benefits of standardization: summary of results. DIN German Institute for Standardization E.V., 2000
- [3] Paul Temple, Robert Witt, Chris Spencer, Knut Blind, Andre Jungmittag, G.M.Peter Swann. The Empirical Economic of Standards[R]. London: DTI, 2005.

How to cope with the challenges of internet-based finance: New finance mode of commercial bank in China

P.H. Wang

Glorious Sun School of Business and Management, Donghua University, China

ABSTRACT: The Internet-based finance is a new model of finance featuring the combination of the Internet and finance. With the help of the Internet and big data technology, Internet companies can reduce the cost in financial transactions, alleviate information asymmetry, and extend the scope and duration of financial transactions. This paper deals with the impact of Internet-based finance on traditional financing, especially the banking industry, and by analyzing their respective advantages and disadvantages, proposes three business models for the banking industry in response to the challenges from the Internet-based financing

KEYWORDS: E-business finance, internet-based finance, financial disintermediation, transaction cost.

In recent years, with the quick development of Internet technology, new internet finance modes have been emerging continuously in China. The internet finance generated from the combination of internet and finance, is a new and convenient financial service mode which realizes funds allocation, payment and information intermediary depending on internet technology and mobile communication technology. Therefore, in a sense, Internet finance is changing the existing operation mode, profiting mode and surviving mode of commercial banks.

terms of product innovation, business development, customer service, risk management and many other aspects. This paper discusses how traditional commercial banks should meet the challenges from Internet-based finance. By analyzing the respective advantages and disadvantages of Internet companies and traditional financial institutions, the paper proposes three business models for commercial banks. Hopefully, those innovative modes of financing can prove to be an effective means by which traditional banks can achieve a sustained healthy development.

1 INTRODUCTION

In recent years, with the rapid development of the Internet technology, Internet-based finance has been emerging and thriving in China. Different from the indirect financing among commercial banks and the direct financing on stock and bond markets, Internet-based finance, resulting from the combination of the Internet and finance, is a new and convenient mode of financial service with the function of funds allocation, payment and information intermediation based on the use of the Internet and telecommunications. In a sense, Internet-based finance is changing the existing mode of survival, operation and profit-making in commercial banks.

Some dot-com companies have gradually made their way into the financing industry. Some business platforms for Internet-based financing have been built by Internet companies like Alibaba and JD.COM, which have exerted a profound impact on traditional finance service providers (Yang,2007). Commercial banks, an important sector in the modern economy, have to face the severe challenges from Internet-based finance in

2 THE IMPACT OF INTERNET COMPANIES ON THE TRADITIONAL FINANCE

The existence of traditional financial institutions is based on the theory of financial intermediation, whose importance and necessity have been clearly manifested through a series of financial theories, ranging from the notions of transaction cost and asymmetric information in the early days to the modern concepts of financial intermediary, risk management, participation cost and value creation (Xie & Zou, 2012). The impact of Internet-based finance on traditional finance mainly lies in the fact that the widespread use of the Internet has greatly reduced the transaction cost, facilitated information access and processing, radically altered the situation of information asymmetry, and bred some Internet companies that have begun to assume certain functions of traditional financial institutions (Xiao,2013). Internet-based finance might evolve into a new mode of third-party financing different from both the indirect financing of

commercial banks and the direct financing on capital markets.

2.1 *Business scope of Internet-based financing*

Internet-based finance has insinuated its way into the domain of traditional financial institutions by starting from payment, then to loan and financing, and finally to deposit. Many dot-com companies, with Alibaba as a typical representative, have begun to offer savings deposits and do some marginal banking business such as supply chain finance and factoring finance, as well as some non-bank financing like Internet-based securities and online insurance.

2.2 *Typical cases of Internet-based financing*

The first case is of Alibaba starting Internet-based finance from e-commerce. AliPay was originally established by Alibaba to solve the online payment problem for Taobao, an e-business under Alibaba, but it has now developed into one of the most important platform of third-party micropayment. Aliloan, an online credit loaner under Alibaba, granted an annual loan of over 100 billion to more than 700,000 clients in 2013. Yuebao, the fund management platform under Alipay of Alibaba Group and Tian Hong Asset Management, had attracted about RMB 500 billion from its users by March, 2014. More than 30 fund companies have started to sell funds through Taobao.com. (Huang-Hailong, 2013). Alibaba has been shortlisted as one of the pioneers of a pilot program to set up private banks in China.

The second case is about Tencent starting Internet-based financing as a platform of social media. In addition to selling insurance policies online, Tencent is making a series of moves toward the expansion of Internet-based financing, which involves online payment, financing, issuing Internet credit cards in cooperation with CITIC bank. It has also received the green light to set up a private bank.

Suning is a different case. Suning, originally an offline chain store group, has begun a strategic shift toward e-commerce since 2010. It has made the foray from traditional business into e-business and then into Internet-based finance. It started from setting up small loan companies, insurance companies, third-party payment companies, and has applied for a bank license.

Cases of setting foot in the finance sector from other businesses include Baidu, a search engine, and Sina.com, a news portal, and so on.

Internet companies of various kinds are all vying for entry into the Internet-based financing by meeting the financial needs of customers through their core Internet applications.

2.3 *More IT companies attracted to P2P and crowd funding*

Now P2P (Peer to Peer) and crowd funding have become a hot spot in the field of Internet-based finance. P2P has been legalized in the United States since the Jumpstart Our Business Startups Act (JOBS Act) was signed into law in March 2012. There had been more than 700 crowdfunding websites in the United States by the end of 2012.

China is also witnessing the blooming of P2P and crowdfunding platforms, the number of which is around 1000 at present. However, with no official regulations and supervision and no accurate statistical data (Huang&Jun, 2013), the potential financial risk related with it grows with the increase in the number of customers, which should be noted with some concern.

2.4 *Yuebao and its impact on bank deposits*

Recently much public attention has been drawn toward online fund management platforms like Yuebao, whose business includes deposit, the core business area of banks. Instead of collecting deposits in explicit forms, funds like Yuebao draw money from clients by indirect means, for example, by tying the clients' online payment accounts with their bank accounts and then yielding returns for their clients via some monetary funds. In this way, a considerable amount of deposits in banks has been diverted away to Yuebao. Statistics from the Central Bank of China in January show that Renminbi deposits in banks has been reduced by RMB 940.2 billion, a large portion of which has been moved into Yuebao in the form of inter-bank deposits.

Yuebao in essence is just an integrated front end of sales, and it is not much different from banks in terms of assets, for its assets mainly take the form of inter-bank deposits via monetary funds. The money travels in a cycle back to banks again, but two changes have taken place in the process---the nature of the money has changed from individual demand deposits to interbank deposits, and the bank deposit interest rates has changed from 0.385% to the much higher interest rates on interbank deposits. Banks are actually the ultimate source of yields for the clients.

Internet companies have developed a new mode of financing different from traditional financing by obtaining licenses for small loans, guarantees and third-party payment, and by serving as an unlicensed financial intermediary via the Internet (Wang, Zhang&Li,2013). Internet-based financing has greatly blurred the division of labor and weakened the specialization in the finance sector, since much of the work in finance now can now be done with the use of the Internet and software. The financial markets, instead of being confined to an inner circle

of professionals, are now open to the general public, who can also have a share in the high yields from the Internet-based financing, a mode of financing that is more democratic and inclusive rather than being controlled by elite professionals.

3 THE RELATIONSHIP OF INTERNET COMPANIES AND FINANCIAL INSTITUTIONS AND AN ANALYSIS OF THEIR MERITS AND DEMERITS

Internet companies' forage into the finance industry has exerted a profound impact on and posed a severe challenge to traditional financial institutions. Faced with an inevitable diversion of clients and business, banks are forced to transform its mode of service and profit making. However, Internet companies have their own weaknesses and limitations, which make it impossible for them to subvert or replace traditional banks. The relationship between dot-com companies and banks and their respective advantages and disadvantages should be viewed from an objective and dialectic perspective.

3.1 *Advantages of Internet companies*

- a. Absence of official regulations is right now the most important advantage of Internet companies, which therefore can have greater room for financial innovation and business development without government supervision over capital, assets and liabilities, business access, security, customer information protection and many other aspects.
- b. Driving force for innovation
Most dot-com companies are private enterprises, which have a stronger sense of crisis, greater creativity, and better adaptability to the changing world. To survive and succeed in the fierce competition, Internet enterprises have to keep up with the rapid development of Internet technology through constant innovation.
- c. Customer-centered service.

Internet companies stress customer experience over safety, and even give it top priority. Customer experience is valued most because if it proves to be unsatisfactory, Internet users will not be attracted or attached to the products or platforms. Without a network of bank outlets, Internet companies rely on websites and mobile apps to as the most important entry for clients, and therefore might put emphasis on user experience, sometimes even at the cost of security.

- d. No historical burden.

Internet companies occupy a moral high ground over banks, for on the one hand they have no income from financial business, and on the other hand they

pass on to clients most of the yields earned through investing the money collected from them, and sometimes even offer subsidies to clients with income from other sources. While banks' only sources of income are banker's spread and service fees, therefore it is very hard for banks to redistribute profits.

3.2 *The disadvantage of Internet companies*

- a. There are higher risks involved and no reliable mechanism of protection against clients' loss for lack of financial licenses and official regulations.
- b. The yields earned for clients are not completely controllable and rely heavily on traditional financial institutions.
- c. The technological and capital security is relatively lower, which is not for technological reasons, but because user experience is put before everything else, even to the sacrifice of security to a certain extent.
- d. Lack of liquidity management experience.

3.3 *The advantages and disadvantages of financial institutions*

Finance institutions are just the opposite of Internet companies in terms of their merits and demerits in financial business, with the merits of financial institutions being the disadvantages of Internet enterprises, and vice versa (Zeng, 2013). Financial institutions have financial licenses and therefore are legally protected in banking business. They have strict risk management, especially liquidity management, higher level of security, better reputation, greater reliability, a complete line of products, especially asset business that is highly dependent on risk management. Risk management has always been the core of banks and other financial intermediaries, and is stressed even more today in light of financial innovation (Zhang, 2001). Financial institutions, through risk management, can provide customers with assets that generate cash flow, while Internet enterprises have to rely on financial institutions for interest-bearing assets.

Where financial institutions are weak is just where Internet companies are strong. The disadvantages of financial institutions include strict government regulations, inflexible mechanism, slow response to market changes, lack of strong motivation for innovation, inadequate attention to customer experience, historical burdens and hindrance from vested interests.

3.4 *Limitations of Internet-based finance*

Internet is powerful tool of communication and interaction, and may bring about revolutionary changes regarding sales, customer service

channels and interaction with customers. By reducing intermediation, it lowers transaction costs, improves service efficiency, optimizes customer experience, and extends the time and scope of service to achieve financial inclusion, which is a very powerful aspect of the Internet.

However, it should also be noted that Internet-based finance, despite its inexorable development, has some limitations. Upon closer examination, it is merely an application of Internet technology to financing, which might cause some changes to the traditional financing but will not replace or subvert it. The future trend is not only about Internet enterprises assuming the role of financial institutions, but also about financial institutions shifting toward the Internet and fully employing the Internet to the advantage of its business development.

Prof. Chen (2014) recently published an article entitled "How new is Internet-based finance?", which discusses the limitations of Internet-based finance and points out that "Internet-based finance is not a new form of finance, but an innovation with respect to the channels of financial sales and financial access. In other words, Internet-based finance challenges traditional banking and capital markets only in channels but is not much different from banks, insurance companies or capital markets regarding product structure and product design. It has not changed the essence of finance, which is the inter-temporal exchange of value and credit among the parties concerned. The emergence of the Internet has changed financial transaction in terms of its scope, participants, volume and environment, but not in nature. The use of the Internet hasn't changed the essence of financial products---the inter-temporal exchange of certain and uncertain cash flows. Neither has it changed the liquidity, profitability and security balance of banks.

3.5 *The relationship between Internet companies and financial institutions*

Financial institutions and Internet companies are not exclusive of each other. Instead, the relationship between them involves cooperation as well as competition.

Banks have a comprehensive range of business covering assets, liabilities and intermediation, while Internet companies are better at payment service and absorption of liabilities, but weaker in assets. The Internet has strong advantages in retailing, while traditional banks still dominate B2B markets, for they still have an advantage in wholesale business where huge capital, greater security and credibility are required. There is ample room for cooperation between banks and Internet companies. For example, Internet companies could serve as the front end and sales channels for financial institutions to draw

idle social funds and handle micropayments, while banks can focus on assets business, especially large and medium-sized corporate financing at the back-end. Dot-com companies and financial institutions should take full advantage of their respective strong points and cooperate where their clients and businesses overlap, and thus build a harmonious financial ecosystem.

4 THREE MODELS OF FINANCE FOR BANKS TO COPE WITH CHALLENGES

Faced with the four major trends of capital regulation, financial disintermediation, marketization of interest rates and use of Internet technology, banks must take a capital-light and low-cost approach to meet the changes in the external business environment.

Transaction costs are considered the reason for the existence of financial intermediation. The major means of reducing transaction costs is economy of scale and scope, as costs will decrease with the increase in trading volumes. Another reason for the existence of financial intermediaries is asymmetric information, which might cause adverse selection and moral hazard (Benston and Smith, 1976).

Due to the widespread use of information technology and Internet, people now have an easier access to more information, which helps to reduce adverse selection and moral hazard. So transaction costs and asymmetric information as the root reasons for financial intermediation has been brought under question. (Merton&Bodie, 1993) What matters most is not the structure but the function of a financial system.

Changes in financial markets and the advancement of information technology have been furthering the disintermediation of financial intermediaries in two forms --- financial disintermediation and technological disintermediation. Financial disintermediation means that big clients rely on direct financing via capital markets and non-bank financial institutions, instead of turning to indirect financial intermediaries (Song & Zhong, 2010). Technology disintermediation means that micro enterprises and individual customers directly carry out financial activities on the Internet, which can solve the problem of information asymmetry with an easier access to information, and which in turn can reduce transaction costs and increase trading opportunities.

Banks will have to promote Internet-based financing in response to financial disintermediation. There are three major models now in practice in China: (1) Online sales of financial products, which shows a development path from networked banks, to e-banks, then to direct banks, and finally to future banks; (2) Running self-owned e-businesses alongside financing (3) Financing in cooperation with e-businesses

4.1 *The model of setting up e-banks*

Online banks in China have been developing for more than ten years, from providing merely online information in the first generation, to offering online products in the second generation, and then to today's more comprehensive online services which includes sales as well as customer service. Many banks have launched third-generation online banking products, such as "I finance" under China Merchants Bank, a virtual financial mall under CITIC Bank, and the Sunshine Business Hall under China Everbright Bank.

The third-generation e-banking means that banks go online, provide user-friendly and customer-centered service, and carry out effective target marketing via data-mining of the behavior and transaction records of customers. In addition, customer managers and financial managers of banks also work online, which enables not only the two-way customer-bank interaction, but also the online handling of various businesses such as pre-sales consulting, sales, post-sale service and complaints. It is different from the earlier e-banks which offered only very limited self-service banking. Of course, technological factors play an important role in the emergence of the third-generation online banks, such as speed of access to the Internet, the handling capacity of the terminal, and the use of smart phones and 4G networks.

Alongside the use of physical outlets, banks have begun to encourage their customers to use online banking. This is O2O pattern. Recently some banks have opened some community banks in residential areas, with the purpose to reach more offline clients and then convert them into online clients.

4.2 *The model of running self-owned e-businesses*

The second pattern is that banks run their own e-businesses by setting up their own B2C or B2B platforms for e-commerce, providing online trading services, and then extending the business to payment and financing. Recently, many banks have been engaged in e-commerce, which has become a new trend in the domestic banking sector. Lots of e-businesses have been founded, like "Good Financial mall" under CCB, "Jiaobohui" under Bank of Communications and "Financial E-Purchase" under ICBC, all of which are run by a particular department of the banks. The e-business under Minsheng Bank is run and controlled by its shareholders. Maybe there will appear e-businesses run by a subsidiary of the bank in the future.

It is disputable as to whether banks should establish and run e-businesses by themselves. Some are supportive of this, arguing that banks have the ability to do so, and whereby can obtain authentic transaction

data of customers, based on which banks can provide customers with value-added services such as online payment and financing and improve bank-customer attachment. But opponents point out that it is against the principle of social division of labor for banks to do e-business. Other reasons against this model include lack of legal support, no effective means of making profits, and inherent incompatibility with the Internet.

I think the actual effect of this model needs to be tested in practice, though current statistics from the banks show that the number of customers, trading volume and customer activity don't seem to be satisfactory so far.

4.3 *The model of cooperation with e-businesses*

The online business of banks is inseparable from the development of e-commerce, and banks are also an integral part of e-commerce. Banks as professional financial service providers have been working in extensive cooperation with e-businesses and providing financial services for them. E-commerce is characterized by a four-in-one integration of business flow, information flow, cash flow and logistics. Cash flow has to be handled by banks which serve e-businesses by offering online payment and online financing. Banks are not only important practitioners of e-commerce but also innovators of finance and supporters of e-business. The integration of four flows into one is an indispensable node for the ultimate formation of a closed loop in e-commerce, which plays four major roles: credit intermediary, payment intermediary, financial intermediary, and information intermediary.

There are three patterns of cooperation between banks and e-businesses. The first one is the regulatory pattern, in which customer accounts and funds are all managed through the E-business platform under the bank. This pattern allows banks not only the strongest control but also brings the greatest gains. The second one is the depository pattern, in which customers' e-ledgers and physical accounts, as well as customers' capital accounts are managed by the banks, and account reconciliation is conducted daily between banks and e-businesses. This pattern is a one-to-many relationship --- several banks on the platform on a single e-business. The third pattern is in the form of payment gateway, or the third-party payment. E-business platforms have an account system of customers, who make their payments through the gateways of banks. In this pattern, banks are most passive, with the least control and profits. Now banks have to adopt all the three patterns to accommodate different business platforms and the needs of customers.

Online financing is a hot and difficult spot in electronic financial service, and the difficulty lies

not in the innovation of financial products but in the integration of data about trading platforms, third-party prices and logistics. Banks have to obtain information of customers' transactions, logistics and prices online, then have online electronic pledge contracts signed to ensure the pledge and redemption of goods. The difficulty also consists in the integration of resources, rather than technical support. Currently, part of online financing, like credit authorization and approval of loans, has to be carried out offline. Some banks offer online financing service without any offline processing. This kind of financial service can be classified into three models --- guaranteed securities, mortgages and pledges. E-commerce platforms and online pledge financing under certain banks have already gone online and been put into use.

5 CONCLUSIONS

Prospect of banks – No more bank outlets in the future.

Bill Gates, founder of Microsoft, predicted that traditional commercial banks would be extinct like dinosaurs in the 21st century. To survive and thrive, banks must reform the traditional model of business by using IT in business management and in the change of bank outlets. In fact, the ultimate goal of banks is to eliminate outlets. In 2012, the bank outlets in Australia decreased from 6631 to 6501, with 130 closed. The outlets of Bank of Communications of China, according to its 2013 interim report, have been reduced from 2691 to 2681. Perhaps in the near future, the bank outlets will be fundamentally

changed, and the bank-customer interaction will take place mainly in the cloud.

REFERENCES

- [1] Chen, Zhiwu.(2014). How New to Internet Finance. *The Economic Observer*, Jan. 4.
- [2] Huang, Biao&Qu, Jun.(2013). Development of P2P and Crowd funding in abroad. *China Fores*, No.12, pp.49–51.
- [3] Huang, Hailong. (2013). Research of the Internet Finance with the Core as E-commerce Platform. *Shanghai Finance*, No.8, pp.18–23.
- [4] Song, Wang&Zhong, Zhengsheng. (2010). The Perception of Financial Disintermediation: Based on Financial Intermediary Theory. *Shanghai Finance*, No.6, pp.12–17.
- [5] Wang, Penghu&Zhang, Lianze&Li, Liangfang. (2013). *B2B Electronic Commerce*.Enterprise Management Publishing House..
- [6] Xiao, Benhua.(2013). The Development of Crowdsourcing Mode in United States and the Inspiration to China. *South China Finance*, No.1, pp.25–26.
- [7] Xie, Ping. &Zou, Chuanwei.(2012). A Study on Internet Based Finance Model.*Journal of Financial Research*,No.12, pp.11–22.
- [8] Yang, Daqiang.(2007).The development tendency of financial intermediation theory and theoretical front. *Jinan Finance*,No.12,pp.24–27.
- [9] Zeng, Gang. (2013). Concentrating on the Feature and Development of Internet Finance: Based on the Theory of Monetary and Financial Perspective. *The Chinese Banker*, No.11, pp.11–13.
- [10] Zhang, Jie. (2001). Review on Development in theory of Financial intermediation. *Social Sciences in China*, No.6, pp.74–84.

The empirical study of construction of public rental based tenants' attention

X.M. Yang & L.L. Zhao

Rongzhi College of Chongqing Business and Technology School, Chongqing, China

ABSTRACT: The public rental housing is considered as an important livelihood project and it is promoted by local governments of China. Along with the increasing number of tenant occupancy in public rental housing, the key of governments is to identify the problems of constructing public rental housing from the perspective of tenants and complete them. Through the liner regression analysis of survey data of Chongqing public rental housing tenant, we have identified the causes of abandoned rent and unreasonable management. It can provide a decision-making gist for command personnel.

1 INTRODUCTION

Public rental housing as supplementary improves the dual system of the housing supply system. Through the construction of public rental housing, the state not only improves the housing supply system, but also adjusts the supply structure of the real estate market, thus promoting the healthy development of the market. It has shown a series of problems with the tenant to stay, and local governments have been paying more and more attention to the voice of occupancy talents. How to further improve the construction of public rental housing, according to factors concerned by tenants is a practical problem currently facing by local governments.

At present, the rent check in time is not long. From such a limited number of studies, the existing conclusions only find some superficial problems, belonging to the weak. With the increase of tenant occupancy and with the growth in the tenant's residential, public rental housing tenants will have more and more objective understanding of the problem, so it is necessary to pay attention to the tenant demand from the view of science, and concerned factors also need further investigation, related to mining deep-seated countermeasures. Based on the reasons mentioned above, this paper chooses Chongqing city as the research object, which is at the forefront of the country to exert efforts to implement the public housing. At present, there are more than 4000 public rental housing in Chongqing city construction, and tenants have lived in some public housing estate. Up to September 2013, the city's public rental contract amounts to about 135000 households with 340000 people. This topic investigates tenants occupied in public rental housing of Chongqing, models the factors concerned by public rental housing tenants and proposes some countermeasures as reference for government.

2 AN EMPIRICAL STUDY ON FACTORS CONCERNING FOR PUBLIC RENTAL HOUSING TENANTS

According to the related factors on the domestic grandpa housing construction demand analysis, Research group refers to the relevant information and does a in-depth survey, compiling the "public rental housing tenants use questionnaire". The questionnaire involves questions about the basic situation, application and use of public rental housing tenants, traffic and facilities, management and service, and lease and repurchase. The research group investigates Minxin garden, West City, South City homes gardens, housing, issues 463 questionnaires, expects objective descriptions of public rental housing tenants, in order to find out the relationship between the construction of public rental housing, the thoughts of tenants and the management of public rental housing in order to figure out perfect countermeasures for the construction of public rental housing in the future.

2.1 Correlation test

Due to a series of problems related to the basic situation of the questionnaire, the research group set the problems as 55 variables which are tested through the SPSS18.0 software. We find that there are different degree of 12 variables.

In order to clarify the relationship deeply, research group adopts Pearson product moment correlation of two tailed test, to test the relationship between the two groups of 12 variables, and the results show two groups of relationship (see [Table 1](#) and [Table 2](#)). The result is consistent with the actual survey situation, because of tenant occupancy preferences and the original rental condition by contrast. If the public housing rental price and traffic advantage is not

obvious, tenants will consider whether to find other rental places; in addition, the surveyed tenants in the rental process generally feel that the existing public housing rents are high, and paying the house management costs makes them uncomfortable and it is not a satisfactory service, and public housing area does not meet the rental needs (especially after the replacement housing rental increase of population is very difficult). Also, the waiting time for public rental housing occupancy and primary concerns are not satisfying.

Table 1. The tenants of rental correlation analysis results.

| | B | C | D | E |
|---------------------|--------|-------|--------|--------|
| Pearson Correlation | -.293* | .310* | .354** | .380** |
| A Sig.(2-tailed) | .021 | .014 | .004 | .002 |
| N | 463 | 463 | 463 | 463 |

Variable A: Abandoning rent reasons

Variable B: The number of declaration after occupancy

Variable C: Property management

Variable D: Transportation distance

Variable E: Contrast before environment to the current environment

Table 2. The results of correlation analysis of the cause of unreasonable tenants think of public rental housing.

| | G | H | I | J | K |
|---------------------|-------|--------|--------|-------|---------|
| Pearson Correlation | .275* | -.271* | .382** | .281* | -.425** |
| F Sig. (2-tailed) | .030 | .033 | .002 | .027 | .001 |
| N | 463 | 463 | 463 | 463 | 463 |

Variable F: The cause of unreasonable tenants

Variable G: The rent

Variable H: The time waiting for the arrival

Variable I: View for current rent

Variable J: View for current property management fee

Variable K: The primary concern

2.2 Linear regression analysis

2.2.1 Linear regression analysis of the reason why tenants abandon renting

At each stage of the stepwise selection method for selecting variables, it needs to consider removing a significant independent variable. Take the cause of abandoning for rent as the dependent variable, and take all factors that could cause giving up renting to linear regression analysis as independent variables. Stepwise selection refers to Table 3. Results of two models:

$$\text{Model 1: } A = 0.4 \times L$$

$$\text{Model 2: } A = 0.34 \times L + 0.324 \times D$$

The model 2 illustrates the influence of tenants abandoning rent from the sample. The reflected rent was caused mainly by the transport distance and the living environment with a previous and present contrast. The public housing transportation distance and the living environment are important factors to consider. This is consistent with the actual situation, and the current public housing construction in Chongqing is far away from the city center. The general need for traffic tools is at least 3. The tenant work trip brings great inconvenience, and increases the cost of living. It might also need to consider the housing price based on the tenants' income, largely influencing their giving up public housing rent and thus making the prices cheaper. And it also reminds us that in the early stage of construction of public housing, its traffic planning and environmental planning areas need to be detailed as much as possible.

Table 3. Linear Regression Analysis of the reason that tenants abandon renting.

| Model | Unstandardized Coefficients | | Standardized Coefficients | | t | Sig. |
|--------------|-----------------------------|------------|---------------------------|--|-------|------|
| | B | Std. Error | Beta | | | |
| 1 (Constant) | 3.261 | .450 | | | 7.245 | .000 |
| L | .695 | .208 | .400 | | 3.347 | .001 |
| 2 (Constant) | 2.277 | .551 | | | 4.135 | .000 |
| L | .591 | .200 | .340 | | 2.959 | .004 |
| D | .526 | .187 | .324 | | 2.818 | .007 |

a. Dependent Variable A: abandoning rent reasons

Variable L: Contrast before environment to the current environment

Variable D: Transportation distance

2.2.2 The cause of unreasonable public rental operation of linear regression analysis

Take the current public rental housing tenants without reasonable cause as the dependent variable, and the survey forms all factors that may lead to the current tenants of public rental housing with unreasonable causes as variable linear regression analysis. Through stepwise selection in Table 3, it can be obtained from linear regression in Table 4. The results of three models:

$$\text{Model 1 } F = -0.456 \times K$$

$$\text{Model 2 } F = -0.473 \times K + 0.372 \times J$$

$$\text{Model 3 } F = -0.492 \times K + 0.366 \times J - 0.246 \times M$$

We choose model 3, and the description of the samples reflects the view that the current public rental housing tenants having no reasonable cause is mainly

caused by the selection of public rental housing priority, the current property management charges views and living and working distance. The top concern of the public rental housing tenants to choose is

to consider whether to meet the current property management fees, and whether sexual price is consistent. The living and working distance are an important factors to consider.

Table 4. Linear Regression Analysis of the cause of unreasonable tenants think of public rental housing.

| Model | Unstandardized Coefficients | | Standardized Coefficients | | t | Sig. |
|-------|-----------------------------|------------|---------------------------|-------|--------|------|
| | B | Std. Error | Beta | | | |
| 1 | (Constant) | 5.036 | .349 | | 14.436 | .000 |
| | K | -.299 | .077 | -.456 | -3.906 | .000 |
| 2 | (Constant) | 4.131 | .413 | | 10.005 | .000 |
| | K | -.310 | .070 | -.473 | -4.408 | .000 |
| | J | .646 | .186 | .372 | 3.467 | .001 |
| 3 | (Constant) | 4.832 | .494 | | 9.780 | .000 |
| | K | -.322 | .068 | -.492 | -4.757 | .000 |
| | J | .635 | .179 | .366 | 3.546 | .001 |
| | M | -.074 | .031 | -.246 | -2.383 | .021 |

a. Dependent Variable F: the cause of unreasonable tenants

Variable J: View for current property management fee

Variable K: The primary concern

Variable M: Living and working distance

Table 5. Comprehensive multi-frequency analysis of public rental selecting.

| | | Set Frequencies | | |
|-------|---|-----------------|---------|------------------|
| | | Responses | | Percent of Cases |
| | | N | Percent | |
| N | G | 393 | 42.4% | 84.8% |
| | O | 211 | 22.8% | 45.7% |
| | P | 151 | 16.3% | 32.6% |
| | Q | 111 | 12.0% | 23.9% |
| | M | 40 | 4.3% | 8.7% |
| | R | 20 | 2.2% | 4.3% |
| Total | | 926 | 100.0% | 200.0% |

a. Group

Variable G: The rent

Variable M: Living and working distance

Variable N: Pay attention to public rental housing factors

Variable O: The use of the area

Variable P: Decoration and facilities

Variable Q: supporting facilities

Variable R: The number of rooms

In Table 5, for frequency of multiple public rental housing concerns, analysis of the model is a good proof. According to the survey, the questionnaire in public rental housing concerns a (priority), B (a minor concern) analysis of the frequency of multiple options after the merger of SPSS, and the results are

shown in Table 5. If these tenants' concerns are not satisfied, the tenants will think it is not reasonable; while the tenants in the survey reflect the level of property management services is poor, the service is not timely (especially the repair service), causing tenants generally believe that the fees is charged and should obtain service existence value, which does not match the situation property management. Also in the analysis of the staff age structure of the sampling survey, it shows that in persons over 50 years of age accounts for the largest proportion (38.1% of the total), and their voices to complete and perfect the public rental is also in the largest proportion.

2.3 The analysis of conclusion

First of all, the reason for abandoning renting is caused mainly by the transport distance and the living environment with a previous and present contrast. Therefore, the needs for pre - on the location, transportation, residential environment, residential space, residential facilities planning and other issues in the construction of public rental housing are to be considered as far as possible.

Secondly, the existing public housing tenants think unreasonable causes are mainly caused by the selection of public rental housing priority (housing area use rationality, small indoor decoration and facilities allocation, repair, the surrounding supporting index), the current property management charges

views and living and working distance. Therefore, the construction of public rental housing should pay attention to in the early public housing location, public housing area and room number configuration, interior decoration and facilities configuration. At the same time in the process of operation, much attention should be paid to improving the level of property management, and increasing management fees, and the price rent standards have the space to float downward.

Finally, with a voice concern for a considerable proportion of the elderly in public housing tenants as the aging Chinese population and the one-child policy, the family pension pressure is more and more prominent. According to the experience of developed countries, the community care system will effectively solve the problem of the elderly, and this model combines the traditional family pension and concentrated homes endowment. Therefore, in public rental housing, community construction and service can be used for reference.

3 PUBLIC RENTAL HOUSING CONSTRUCTION COUNTERMEASURES BASED ON THE PERSPECTIVE OF TENANTS

Analysis of the correlation and regression results has referenced significance for Chongqing public rental construction, mainly embodying in two aspects: First of all, by reducing the reasons for leaving, allow tenants to choose to live in public rental housing. Secondly, by reducing the unreasonable operational reason, it can improve the public rental housing tenants' occupancy satisfaction. Therefore, the specific measures of Chongqing public housing construction are as shown in figure 1.

3.1 Many groups scattered layout planning, advocated the development of the construction of public rental housing

Firstly, it needs to consider the planning layout for planning group, and it needs to control the building volume concentration of the groups in public rental housing. Change the original states that focus on one or two locations to build large public housing community and construct small-size public rental housing communities, according to the multi group for site planning. In this way, the overall construction of public rental scale can increase the stationing of public rental. It allows tenants to choose the nearest rental location based on the need of work location, to solve the traffic problems, and also avoids the law and order problem because of the large-scale public rental.

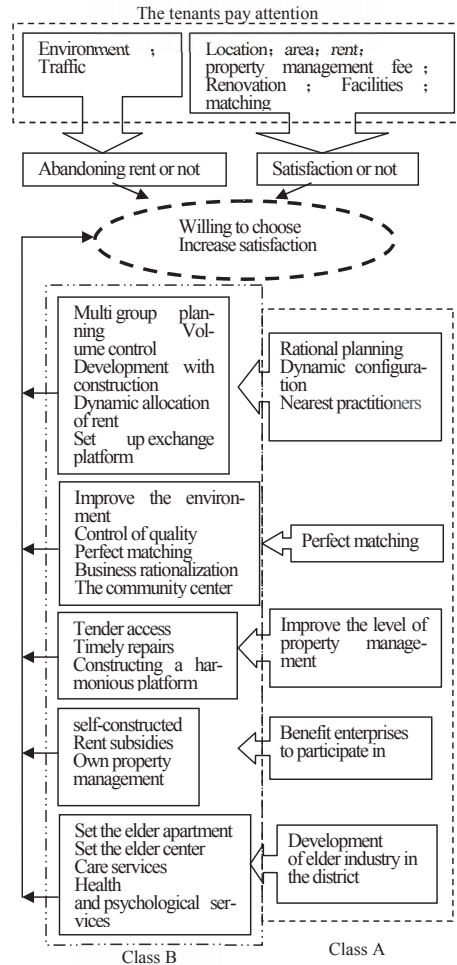


Figure.1 Tenants concerned factors and Solution.

Secondly, the department of Homeland can be combined with planning to require developers to build public rental housing in the residential area at the time of transferring land. Such public rental housing community can enjoy the existing facilities, and can also avoid social problems caused by the partition between the rich and poor because the public rental is separated from commercial housing, and avoids poor site selection because the independent public rental set.

3.2 Implementing of dynamic with rent in the process of renting houses, and building realtor swap information platform

Public rental management agencies should improve the dynamic with rent for public rental. It mainly needs to solve two problems: Traffic problems were created by the changing living and working distance

because tenants' job changes; and housing shortage is caused by the original area due to the change of tenants' demographic structure.

For this situation, tenants can submit an application to the administrative department of public rental in the hiring process. Management can allow tenants to use a larger area housing swap, or swap housing to more recent public rental community from the place of work. This can achieve the dynamic configuration for housing tenants in signing period. Also, it can establish an information platform for swapping public rental housing. It builds bridges for the replacement for both and the supervision of the swap process.

3.3 *Improve community infrastructure, improve service quality property management*

Firstly, for planning ancillary facilities, it should study consumer preferences and consumption levels of tenants. Appropriate commercial activities are introduced, such as the use of pre-signed, the lease contract control and a reasonable guide to the phenomenon for sublease in business operating processes. It not only needs to meet the recreational needs of tenants by configuring the residential area of basic sports facilities, but also needs to carry out a variety of community activities. The rational allocation of the kindergartens, schools, farmers markets is to be made. It should introduce cheap supermarkets, and guide street stalls to centralize premises, etc. It needs to encourage businesses or enterprises to recruit and train staff who live a difficult life to recruit and train laid-off workers in the area of public rental. It creates employment opportunities for public rental community residents.

Second, the emphasis on quality is focused on interior decoration and indoor facilities. Strengthen the construction team screening and dynamic management of the construction process, place an emphasis on intermediate inspection procedures for covert projects, and strengthen procurement management. In order to reduce the "pass the buck" phenomenon, rationalize the bearing of the main and expenditure program of public rental maintenance costs between rental income of public rental agency to a property management company.

Again, keep a high level of property management services, by bidding in social choice approach to property management companies. Public rental management agencies should encourage tenants to set up a "tenant management committee" to participate in community management, increasing the region's property owners and property management companies' trust.

3.4 *Public rental beneficiary enterprises are encouraged to participate in the construction and operation, and guided to social responsibility*

Firstly, encourage eligible medium-sized enterprises and enterprises in the park to participate in the construction of public rental. Encourage enterprises to put money into staff housing construction to build public rental to employees for personal use, and check-qualified staff remains unified and audited by public rental management agencies. The government can give tax concessions related to participation in the construction of public rental companies. Meanwhile, public rental depreciation rate may also be appropriate to improve in order to attract businesses to participate.

Secondly, it draws a conclusion from the foregoing analysis that rent is a most attention factor, and most of the tenants think that the current public rental rent is too high through investigation. According to business conditions and employee's income, medium-sized enterprises in the park should provide rent subsidies for different grades of employees of public rental lease. In this way, one can reduce the pressure on staff rents, and it also reflects the corporate welfare, increasing their sense of belonging.

Thirdly, it can encourage enterprises to bear property management, which means bring property management into enterprise logistics management in this public rental community by enterprise participation, primarily by its own rental unit staff. Thus, one can clearly service, and reduce public rental repairing "passes the buck" phenomenon. On the other hand employees can be subsidized by companies to reduce the property management fee charges. So tenants will be benefited, and it will improve tenant's satisfaction.

3.5 *Concerns older age groups of public rental, the development of the old industrial area of public rental*

Firstly, the elder apartment block is provided to the elderly who are not living with their children and rental is planned to set in public rental community. The psychological and physiological characteristics of the elderly are taken into account in the planning and design time.

Second, social centers or the club for the elderly is set up. It will provide entertainment and social networking sites for the elderly. Meanwhile, the community will organize some activities so that older people will concern about the community, and enrich their lives.

Again, the provision of elderly helpline service, setting versatile service stations, day care for the elderly, home care, cooking, cleaning, health, care,

health, counseling and other services are made possible. The government should give tax breaks to provide low-rent business premises and other policies to support the industry engaging in the elderly subjects.

ACKNOWLEDGMENTS

Foundation item: the Chongqing City Board of Education, Science and technology research project (KJ120729); Research projects in the humanities and Social Sciences at Rongzhi college of Chongqing business and technology school (20128001)

REFERENCES

- [1] Liu Gui-wen, Zhang Yun-sheng, Xu Peng-peng. 2013. Comprehensive evaluation of public rental tenant satisfaction. *Construction Industry* (09) :79–83
- [2] Dong Xin-long, Lin Jin-gui. 2012. High quality public rental: Country Strategy and livelihood choices. *Economists*.(01) :30–40
- [3] Zhang Hongting, Li Xiaoping. 2011. Analysis and recommendations of foreign public rental construction. *Value Engineering* (09) :283–284
- [4] Tian Qiu-sheng, Li Lei. 2011. A package of solutions to the problem of financing the construction of public rental. *Financial Practice* (05) :71–74
- [5] Chen Li-li. 2011. Analysis of Chongqing public rental model. *Housing industry* (4): 70–72
- [6] Yang Ling. 2011. China security status of housing management and perfect -- Taking Chongqing city as an example. *Western forum* (9): 35–43
- [7] Zhang Chuan-hui. 2011. The government role orientation based on the new Keynes doctrine. *Administrative forum* (2): 49–53
- [8] Zhang Yan-qun. 2011. The Enlightenment of German public housing policy of our country. *Chinese Trade Herald*.(7): 32–33

Rhyme of ink: To composite effects of ink painting in after effects

Z.Y. Li & J. Kang

Academy of Fine Arts of Capital Normal University Beijing, China

R.P. Cao

Advertising Department of Zhejiang Wanli University Ningbo, China

ABSTRACT: It is a hot topic which the digital visual effects circles discuss in recent years that our intention is to composite effects of ink painting in After Effects. The abundant synthesis function of After Effects and many third-party tools support for it can satisfy people higher levels spirit demand. Contextualized in Track mattes and Cycore FX HD CC of comprehensive application study, this paper mainly discusses a feasible method of ink painting effects in a dynamic process.

KEYWORDS: Ink penetration, track mattes, plug-ins, after effects.

1 INTRODUCTION

In the TV advertising, the effects of ink painting which can conveniently make similar result by After Effects and its related plug-ins becomes more and more common. In this example, Track mattes, rail mask or orbital matte, has played a very important role in helping to solve the problem of the effects. Matte can be understood as what is used to define another layer of the transparent regions and it is on any channel of layer which may be defined by another layer of transparent regions. As one of the channels associated masking, the alpha-channel, which often plays the role of Matte, is defined opaque area by white color and defined transparent area by black color. If there is a lack of the alpha-channel in the layer where the Matte is being, or the result of making by other channels is better than the effect of making by alpha-channel, then the effect of Matte will be generally chosen. You can use Track mattes to do some works which show the partial content of layer through another layer, such as the videos on display will only see what would happen when the shape of the text is as a matter of the film. Sure enough, you watch the film in fonts. At this time, transparent value of fill layer that is located in the lower layer is depended on the alpha-channel or luminance value of the upper layer that acts as orbital matte. If you want to use a layer that does not contain alpha-channel to define the Matte effect, you had better to use the value of image brightness, that is, the Luma Matte which can make it better. By the way, the effect of Luma Matte is contrary to the effect of Luma Inverted Matte. Certainly, if a layer contains alpha-channel,

you can use the way of Alpha Matte or Alpha Inverted Matte to realize the matte effects. Next we will use the Track mattes and plug-ins to achieve the ink painting effects.

In the beginning before production, please install the plug-ins of Cycore FX HD CC and Genarts Sapphire for After Effects, and among them, the series of CC filters can appear in different menus.

2 INK PENETRATION AND ITS ANIMATION

Bamboo is a common theme of Chinese ink paintings. But we should not discuss the methods of making bamboo in this paper, you might solve the problem of bamboo materials by yourself. In addition, the materials of landscape we will mention below is expectedly bamboo.

The first step is the use of After Effects — primarily used for creating motion graphics and visual effects — to make the effect of ink penetration. Open the software, where you can use command of Composition→New Composition to create a synthesis that is named “ink” in this example. In the properties dialog box, you can choose the TV formats of PAL D1/DV, whose ratio for pixels is 720:576, and length of time for composition is 35 frames.

Now we come to the main item on the Ink droplet. Using command of Layer→New→Solid (or employ the Ctrl + Y key combination to open the dialog) to create a layer of solid and named it for “ink a”. In this composition, meanwhile, the layer’s background is set for white. Now we add the effect of CC Particle

World for layer of “ink a” while click the right mouse vis as follows:

- Grid: off
- Velocity: 0.30
- Gravity: 0.000
- Birth Size: 2.000
- Death Size: 2.000
- Size Variation: 85.0%
- Max Opacity: 66.0%

The next step is to create a new layer. What you need to do is to select the ink a layer and copy it into a new layer by the command of Edit→Duplicate (or use the Ctrl + D of shortcut key combination), and then click the Enter key on the keyboard after selecting the new layer and naming it “ink b”. Now we can add an effect of “S_WarpBubble” for this layer. The settings in the parameters dialog boxes are represented as below:

- Amplitude: 0.500
- Frequency: 42.00

At present, the effect in composition window about the ink penetration is not clear, which may lead to the bad effect of layers. So we have better to apply two kinds of effects that can arise visual appearance of ink with menu command of “Fast Blur” and “Levels” to “ink b” layer and should respectively set their own parameters. After being assigned the effects to the “ink b” layer, we once again set effects on the “ink a” layer. First, we choose and press “S” key on the keyboard to unfold the Scale item of “ink a” layer and scale it as needed to fit (such as 112%). And then press “T” key on the keyboard to unfold the Opacity item of the same layer and set its opacity value to 70%. This time we finally get to set a “Soft Light” mode superposition for “ink a” layer. The general effects is as what is shown in Figure 1. Now you can press the “O” key on the keyboard to examine the effect of ink penetration.

Just now, we will add the materials of bamboo that has been completed the ink penetration, according to the following operations:

First of all, you can make a new Composition (Ctrl + N) whose length of time is required to keep 35 frames, and name it “ink droplet”. The background color of new composition is set to be soft ecru while keeping the other parameters constant. Next, we move the cursor to the project window and double click the mouse which will import materials of bamboo from folders where default parameters are allowed for them. At this time, there are bamboo frame sequences and ink composition that have been completed before the project window, and then these materials could be dragged respectively into the composition of “ink droplet”. But we would like to

remind that the “bamboo” is above the “ink” in this new composition. Now you can choose the Alpha Inverted Matte “bamboo→0000-0075→ .tga” in pull-down menu after clicking the None button of Track matte on the layer of ink. Finally, the position and size of bamboo should get properly settled. If necessary, you can refer to the settings in Figure 2 while are working on the related content that could be good for your works.

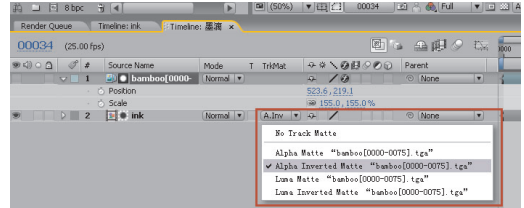


Figure 1. The effect of ink penetration and parameters setting.

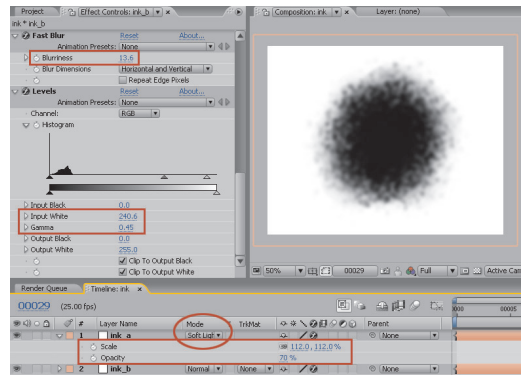


Figure 2. The settings of “ink droplet.”

We require, in order to rich the expression of animation, the other ink for our vision. So the “ink droplet” is dragged into composition again from the Project window. Here we just leave it in another place and leave its size, transparency and spin value set properly. Meanwhile, the new layer is to undergo a “Fast Blur” effect that this ink droplet will have a little fuzzy vision if the Blurriness value sets to approximately 8.5. As what is shown in the Figure 3.

In addition to the impressiveness of the settings, there is a wide use of the words, which at times seems magical to ink painting style. In the layer of ink droplet, the Ctrl + Y key combination will open the solid dialog. In the dialog box, we set the ratio of this solid for 280 x 36 pixels, and name it “strip”. The following is the specific parameters of the strip:

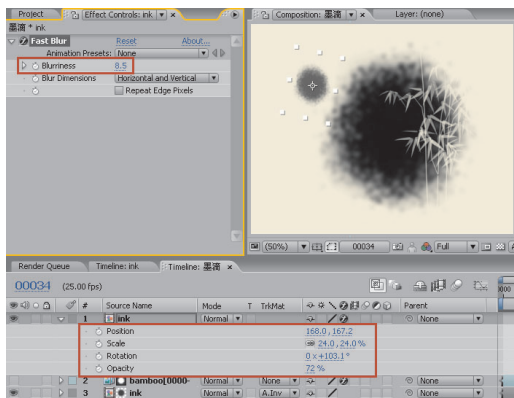


Figure 3. The settings of new “ink droplet”.

- Pixel Aspect Ratio: D1/DV PAL (1.07)
- Color R: 220; G: 98; B: 98

Then we simply input the text you wish to type, Of course, it is best of the ancient poetry of China. The writing on the red strip has been adjusted in the words spacing and its size. At present the red strip which looks abrupt, can be chosen and changed from its blending mode to “Luminosity”. It is the basis on which we continue to add mask to the red strip, an act that is set to subtract mode of Mask, and set a feather

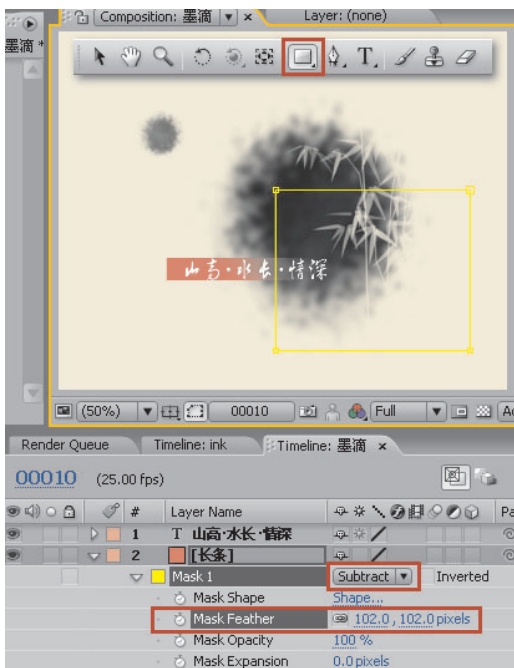


Figure 4. The settings of new “ink droplet”.

value to 102.0 or so for the Mask, which is shown in Figure 4.

If that match up with the ink penetration happens, we have to take it a fade in effect that we’re going to be given in the two layers of “strip” and words. What we need to do is choose them and press the “T” key on the keyboard to open their attributes of transparency, The time line firstly is moved to 10th frame where you can click on the button of key frames to record animation, and it will be moved once again to the first frame where the Opacity of two layers is set to 0%. Finally, the layer of little “ink” is to move 8th frame backward so that it will start to appear in the moment which is as shown in the Figure 5.

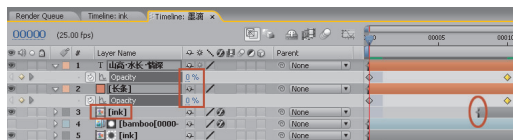


Figure 5. The animation setting of two layers.

3 LANDSCAPE IN THE INK DROPLET

Mountains and water are nature’s dynamic treasures. As the ancient poem goes, “Mountains are being more active if they are surrounded by water and water is being more elegance if they are embraced by mountains”. When the situation that emerges in your mind as the poem describes, you can start to make the landscape better serve your works.

Let us create a new Composition (Ctrl + N) whose length of time is required to keep 125 frames, the ratio is now set for 720 x 576 pixels and name it “lakeside”. As the above says, the process of making landscape in the ink droplet is beautifully handled of some materials by you. If you have completed them or have collected materials in our work which are needed, you will import them into the project window, such as ink droplet, background, lake and mountains. Moreover, if it reminds you when the material in TGA format comes to Project, you have better to choose “Ignore” in interpreting Footage.

Now, the materials first will be delivered into timeline window. After all are done, you’ll perform the following operation:

- Changing the layer of ink droplet type on the “Multiply”.
- Building a new solid layer called “matte” by Ctrl + Y key combination and using the pen to draw a mask for it.
- Pressing the “M” key twice to open the mask attributes of “matte” item. The value of the Mask

Feather is set for 116.0, and adjusts the shape of the Mask.

- Clicking on the None button below the TrkMat of “lake” layer and selecting Alpha Matte “Matte” option in the pull-down menu. As what is shown in the [Figure 6](#).

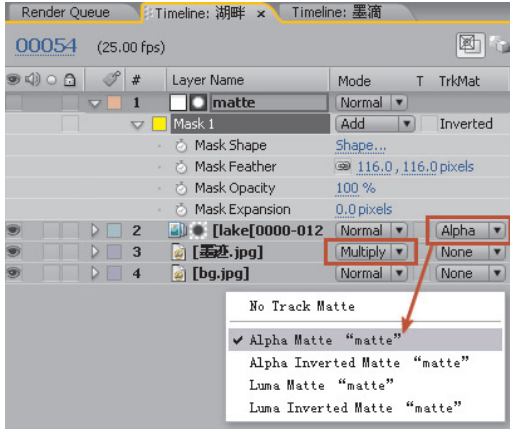


Figure 6. Layer type and mask feather.

After scaling the two layers of “lake” and “ink” as needed to make them fit in the specified size (such as 85.0%), the bamboo sequence frame that is designated as TGA is dragged into the window of Time line from the Project. For the sake of better visual effect, let us add Effect→Color Correction→Curves on the bamboo, and adjust curve in Effect Controls window. At the same time you can modify the bamboo on its size and location if you are not satisfied with them. However, not only is there abate at bamboo’s stereo feeling that it will reduce its vitality, but there is also the threat of dimensional feeling with lake and mountains that act as background. So we should duplicate the bamboo (Ctrl + D) to create a new layer where the effect of Hue/Saturation (Effect→Color Correction→Hue/Saturation) is accepted. You may also checkmark the Colorize option in the Hue/Saturation tool if you want to use a specific color for the new layer. In this example, Colorize Hue is set for green and Colorize Saturation is set for 50. Don’t forget to set the blend mode of this layer to Overlay.

We need a new text in the scene at this time. You can make a similar scene as it is shown in [Figure 7](#) according to the method as mentioned before.

In the following steps, you should animate your work in Timeline of different layer, and here only set up their animation of position, transparency and scale.



Figure 7. Add new text to the scene.

Choose the two layers of ink and matte, then press the “S” key on keyboard to open their attributes of scale, the time line firstly is moved to 20th frame where you can click on the button of key frames to record animation, and it will be moved once again to first frame where the scale of ink layer is to set to 30.0% and matte layer is set to 28.0% which is as shown in the [Figure 8](#).

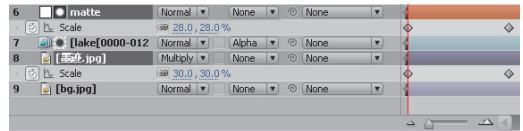


Figure 8. Setting animation of two layers.

Choose the layer of lake and press the “P” key on keyboard to open their attributes of Position, the time line firstly is moved to first frame where you can click on the button of key frames to record animation, and set the position of lake for (300.0, 288.0) — the parameters is only for reference, and the same as below. The time line will be moved once again to end frame where the position of object for (340.0, 288.0).

Choose the two layers of bamboo and draw backward to move them of total 49 frames that also end up animation in the last frame. Press the “P” on keyboard and then press the “Shift + T” key combination to open their attributes of Position and transparency that would be convenient to set up the key frames animation respectively for the two properties:

The time line firstly is moved to 49th frame where you can click on the button of key frames before Position records animation, and set the position of two layers for (262.0, 286.0) then the time line will be moved once again to end frame where the position of them are in a suitable position.

The time line firstly is moved to 49th frame where you can click on the button of key frames before Opacity records animation. Set the opacity of two layers

for 0%, and then the time line will be moved once again to 62th frame where the opacity is set for 100%.

As what is shown in the [Figure 9](#).

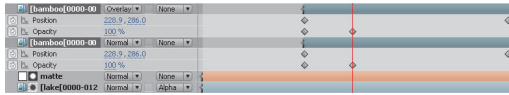


Figure 9. Animation settings of two properties.

The final stage is that set up the translation and fade-in animation for text, you can make the effect be the same as that before. If you are interested in our methods in making the ink painting effect, you could work through the above context that we have discussed.

4 CONCLUSION

We have seen the unique function of Track mattes and two plug-ins in After Effects from the above process of compositing effects of ink painting. There are three tips for working with track mattes at the end of the paper:

First, the contrast between light and dark parts of the matte layer is increased by using the levels effect. This reduces the problem of having a lot of mid-range values, which will translate to partial transparency. In general, mattes are most useful when they define areas as entirely transparent or entirely opaque, except at the edges.

Second, use a channel as a matte other than the alpha channel of the matte layer shall be routed the desired channel's value into the alpha channel.

Finally, if you want to move the layer that is matting, the track which mattes a child of the layer is matted when a track matte is animated.

CycoreFX HD and Genarts Sapphire are the professional version that allows to do the compositing and effect works at higher bit depths, HDTV and film. The plug-ins in the Cycore FX HD package supports 16 bpc and 35 has 32 bpc supports. Genarts Sapphire is used by the film's best content creators to create the looks that engage viewers, grow audiences, and increase the value of content. If you want to know more the information about the plug-ins, web sites are effective places to get information, just like elevators are effective places to get to a particular floor in a building.

REFERENCES

- [1] Mark Christiansen. After Effects CS5 Visual Effects and Compositing Studio Techniques [M]. U.S.: Adobe Press, 2010.
- [2] Michele Yamazaki. Plug-in to After Effects: Third Party Plug-in Mastery [M]. England: Focal Press, 2011.
- [3] The online Help Manual of After Effects [DB/OL]. http://help.adobe.com/en_US/AfterEffects.html. 2012-3-12.
- [4] Aharon Rabinowitz. Force Motion Blur [EB/OL]. http://allbetsareoff.com/?s=Cycore+FX&submit_button=Search.2012-3-20.

Author index

- Ahmad, M., 99
Ameur, S., 159
Amin, Y., 369
Amrita, D., 41
An, J., 295
Ao, M.S., 187
Attia, S., 319
- Bai, D.S., 169, 263, 271
Bannari, A., 23, 173
Berardi, L., 193
Bu, H.M., 111
- Cai, D.W., 353
Caiji, Z.M., 165
Cao, L.L., 133
Cao, R.P., 469
Cao, T., 353
Chen, B., 115
Chen, C.W., 139, 145
Chen, F.F., 259
Chen, G., 169, 263, 271
Cheng, Z.G., 95
Chen, P.F., 307
Chen, Q.X., 329
Chen, X.Y., 63
Chen, Y.C., 145
Chen, Y.J., 335
Cherifi, M., 159
Choi, J., 285
Cimbelli, A., 193
- Dai, J.P., 377
Dalipi, F., 223
Deng, H.L., 295
Ding, K.H., 187
Ding, X., 311
Dong, C.D., 139, 145
Dong, Z.Y., 219
Džunda, M., 239
Du, Y.H., 9
- Fang, Y.J., 105
Fan, Y.L., 381
Feng, J.H., 307
- Feng, P.F., 423
Fu, C.H., 121
Fu, L., 311
- Gebremedhin, A., 223
Ge, W.F., 81
Girouard, G., 173
Gon, C., 315
Gui, Z.B., 411
Gu, J., 105
Guo, L., 9, 165, 205
- Han, D.Y., 385
Han, H.Y., 3
Han, J.S., 51
Han, L.X., 115
Hashiba, H., 179
He, F., 105
He, S.P., 29
Hong, S.G., 303
Hou, Y., 329
Hou, Y.L., 311
Hsia, K.F., 139
Hsu, W.J., 335
Hu, B., 377
Hu, H., 307
Hu, Q., 211
Hu, R.H., 153
Hu, R.J., 77
Husain, T., 99
Hu, W., 391
Hu, Y.J., 153, 187
Hu, Y.Q., 71
Hwang, I.K., 243, 397, 401
- Igolkin, A.A., 417
- Janecka, K., 193
Jia, C.Y., 55
Jiang, S.H., 77
Jin, D.H., 343
Jin, K., 251
- Kang, J., 469
Kang, L.Y., 315
- Kanth, R.K., 369
Kaspin, S., 19
Khalil, A., 23
Khan, M.W., 433
Kholoden, E.E., 59
Khritonenko, D.I., 45
Kim, M.J., 243, 397, 401
Korbelová, Š., 15
Kostina, A.M., 227
Kotianová, N., 239
Kumar, H., 369
Kurumi, S., 255
- Lahdir, M., 159
Lee, J.H., 243
Li, C.W., 381
Li, F., 423
Li, F.X., 427
Li, K.X., 267
Liljeberg, P., 369
Li, L.M., 29
Li, S.Y., 267
Liu, F., 9
Liu, H.G., 81
Liu, H.Y., 359
Liu, J., 169, 263, 271
Liu, L., 51
Liu, M.D., 133
Liu, M.J., 381
Liu, S.W., 343
Liu, W., 55, 247
Liu, Y., 353
Liu, Y.M., 427
Liu, Y.P., 303
Liu, Z.H., 55
Li, W.F., 169, 263, 271
Li, X., 105
Li, X.J., 55, 211
Li, X.Q., 169, 263, 267, 271
Li, Y., 127
Li, Z.J., 427
Li, Z.Y., 469
Li, Z.Z., 71
Lobanov, S.A., 59
Luo, S.S., 29

Lu, T.K., 259
 Lv, L., 205
 Lysenko, A.I., 227

 Ma, F., 77
 Ma, Y., 423
 Mi, C.L., 51
 Mohamad, N., 19
 Morina, O.M., 59

 Oey, E., 405
 Olijár, A., 279
 Oravcová, E., 279, 447

 Park, D.I., 243
 Park, J.H., 41
 Park, J.K., 41
 Park, T.W., 243, 397, 401
 Pol'aková, S., 447
 Pomatilov, F.S., 417
 Pyatakov, A.D., 89

 Qian, L.F., 81
 Qi, H.S., 289
 Qu, J.Q., 347

 Rallapalli, M.V., 373
 Rao, Y., 251
 Reisová, M., 279, 447
 Rodionov, L.V., 417

 Sang, H.T., 71
 Saquaque, A., 173
 Semenkin, E.S., 45
 Shammout, M.W., 85
 Shan, Z.H., 307
 Shih, Y.J., 139
 Shi, P.M., 385
 Sim, K.S., 243, 397, 401
 Šípková, V, 15
 Song, F.B., 81
 Spišák, J., 35, 279, 447
 Su, C.C., 145
 Sun, L.Z., 55
 Sun, T.K., 247
 Sun, X.M., 315
 Su, Y.H., 299
 Suzuki, K., 255

 Tai, P.D., 55
 Tang, H.K., 381

 Tang, L.S., 71
 Tan, P.H., 299
 Tan, Y.A., 347
 Tao, Q.S., 363
 Teng, Y.J., 359
 Tenhunen, H., 369
 Tian, C., 95, 299

 Vodičková, M., 15

 Wang, D., 205
 Wang, D.L., 63
 Wang, D.X., 439
 Wang, H.L., 233, 267, 271
 Wang, N., 77
 Wang, P.H., 457
 Wang, P.P., 311
 Wang, R.R., 133
 Wang, X., 55, 347
 Wang, Y.J., 165
 Wang, Z.B., 289
 Wang, Z.Y., 289
 Wan, H.H., 303
 Wan, X.H., 233
 Wei, H.H., 307
 Wei, S.M., 385
 Wei, W.S., 63
 Wei, Y.B., 121
 Wu, B.S., 205
 Wu, G.S., 63
 Wu, J.Z., 77
 Wu, N., 329
 Wu, Y.C., 63

 Xia, S., 3
 Xing, H., 263
 Xu, D., 363
 Xue, D.Y., 9, 165
 Xu, G.Q., 433
 Xu, S.C., 439
 Xu, W.Y., 303

 Yan, F.F., 115
 Yang, H., 289
 Yang, H.M., 219
 Yang, K., 427
 Yang, X.M., 463
 Yao, G.Q., 295
 Yao, L., 391
 Yastremskaya, E.A., 89
 Yayilgan, S.Y., 223

 Yin, C., 251
 Yonekawa, S., 255
 Yuan, G.W., 363
 Yuan, W.N., 169
 Yu, J., 315, 411
 Yu, X.W., 423
 Yu, X.Y., 153, 187
 Yu, X.Z., 9

 Zahid, T., 433
 Zakaria, H., 85
 Zelko, M., 35
 Zeng, B., 391
 Zhang, C.X., 315
 Zhang, D., 247
 Zhang, F., 81
 Zhang, G.Z., 299
 Zhang, H., 121
 Zhang, H.F., 299
 Zhang, M.B., 423
 Zhang, N., 311, 385
 Zhang, Q.F., 111
 Zhang, Q.X., 347
 Zhang, S.F., 211
 Zhang, S.Y., 127
 Zhang, X., 81, 315, 423
 Zhang, X.F., 95, 299
 Zhang, Y.X., 219
 Zhao, F.F., 439
 Zhao, G.Y., 81
 Zhao, L., 453
 Zhao, L.L., 463
 Zhao, Y., 127, 363
 Zhao, Y.Q., 105
 Zha, Y.P., 427
 Zheng, J., 439
 Zheng, L.R., 369
 Zhen, J.H., 29
 Zhou, A.H., 377
 Zhou, H., 363
 Zhou, J.P., 289
 Zhou, J.X., 303
 Zhou, L.H., 233
 Zhou, Y.M., 433
 Zhu, H.Q., 311
 Zhu, L.H., 77
 Zhu, L.P., 377
 Zhu, Y., 289
 Zhu, Y.X., 311
 Zvereva, V.P., 89, 227