

Pertanika Journal of

**SCIENCE &
TECHNOLOGY**

JST

VOL. 25 (2) APR. 2017



PERTANIKA
JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

Journal of Science & Technology

About the Journal

Overview

Pertanika Journal of Science & Technology (JST) is the official journal of Universiti Putra Malaysia published by UPM Press. It is an open-access online scientific journal which is free of charge. It publishes the scientific outputs. It neither accepts nor commissions third party content.

Recognized internationally as the leading peer-reviewed interdisciplinary journal devoted to the publication of original papers, it serves as a forum for practical approaches to improving quality in issues pertaining to science and engineering and its related fields.

JST is a **quarterly** (January, April, July and October) periodical that considers for publication original articles as per its scope. The journal publishes in **English** and it is open to authors around the world regardless of the nationality.

The Journal is available world-wide.

Aims and scope

Pertanika Journal of Science and Technology aims to provide a forum for high quality research related to science and engineering research. Areas relevant to the scope of the journal include: bioinformatics, bioscience, biotechnology and bio-molecular sciences, chemistry, computer science, ecology, engineering, engineering design, environmental control and management, mathematics and statistics, medicine and health sciences, nanotechnology, physics, safety and emergency management, and related fields of study.

History

Pertanika was founded in 1978. A decision was made in 1992 to streamline Pertanika into three journals as Journal of Tropical Agricultural Science, Journal of Science & Technology, and Journal of Social Sciences & Humanities to meet the need for specialised journals in areas of study aligned with the interdisciplinary strengths of the university.

After almost 25 years, as an interdisciplinary Journal of Science & Technology, the revamped journal now focuses on research in science and engineering and its related fields.

Goal of *Pertanika*

Our goal is to bring the highest quality research to the widest possible audience.

Quality

We aim for excellence, sustained by a responsible and professional approach to journal publishing. Submissions are guaranteed to receive a decision within 14 weeks. The elapsed time from submission to publication for the articles averages 5-6 months.

Abstracting and indexing of *Pertanika*

Pertanika is almost 40 years old; this accumulated knowledge has resulted in Pertanika JST being abstracted and indexed in SCOPUS (Elsevier), Thomson (ISI) Web of Knowledge [BIOSIS & CAB Abstracts], EBSCO & EBSCOhost, DOAJ, ERA, Cabell's Directories, Google Scholar, MyAIS, ISC & Rubriq (Journal Guide).

Future vision

We are continuously improving access to our journal archives, content, and research services. We have the drive to realise exciting new horizons that will benefit not only the academic community, but society itself.

Citing journal articles

The abbreviation for *Pertanika Journal of Science & Technology* is *Pertanika J. Sci. Technol.*

Publication policy

Pertanika policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications. It prohibits as well publication of any manuscript that has already been published either in whole or substantial part elsewhere. It also does not permit publication of manuscript that has been published in full in Proceedings.

Code of Ethics

The *Pertanika* Journals and Universiti Putra Malaysia takes seriously the responsibility of all of its journal publications to reflect the highest in publication ethics. Thus all journals and journal editors are expected to abide by the Journal's codes of ethics. Refer to *Pertanika's Code of Ethics* for full details, or visit the Journal's web link at http://www.pertanika.upm.edu.my/code_of_ethics.php

International Standard Serial Number (ISSN)

An ISSN is an 8-digit code used to identify periodicals such as journals of all kinds and on all media—print and electronic. All *Pertanika* journals have ISSN as well as an e-ISSN.

Journal of Science & Technology: ISSN 0128-7680 (*Print*); ISSN 2231-8526 (*Online*).

Lag time

A decision on acceptance or rejection of a manuscript is reached in 3 to 4 months (average 14 weeks). The elapsed time from submission to publication for the articles averages 5-6 months.

Authorship

Authors are not permitted to add or remove any names from the authorship provided at the time of initial submission without the consent of the Journal's Chief Executive Editor.

Manuscript preparation

Refer to *Pertanika's INSTRUCTIONS TO AUTHORS* at the back of this journal.

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words **I**ntroduction, **M**aterials and **M**ethods, **R**esults, **A**nd, **D**iscussion. IMRAD is simply a more 'defined' version of the "IBC" [Introduction, Body, Conclusion] format used for all academic writing. IMRAD indicates a pattern or format rather than a complete list of headings or components of research papers; the missing parts of a paper are: *Title, Authors, Keywords, Abstract, Conclusions, and References*. Additionally, some papers include Acknowledgments and Appendices.

The *Introduction* explains the scope and objective of the study in the light of current knowledge on the subject; the *Materials and Methods* describes how the study was conducted; the *Results* section reports what was found in the study; and the *Discussion* section explains meaning and significance of the results and provides suggestions for future directions of research. The manuscript must be prepared according to the Journal's **INSTRUCTIONS TO AUTHORS**.

Editorial process

Authors are notified with an acknowledgement containing a *Manuscript ID* on receipt of a manuscript, and upon the editorial decision regarding publication.

Pertanika follows a **double-blind peer-review** process. Manuscripts deemed suitable for publication are usually sent to reviewers. Authors are encouraged to suggest names of at least three potential reviewers at the time of submission of their manuscript to Pertanika, but the editors will make the final choice. The editors are not, however, bound by these suggestions.

Notification of the editorial decision is usually provided within ten to fourteen weeks from the receipt of manuscript. Publication of solicited manuscripts is not guaranteed. In most cases, manuscripts are accepted conditionally, pending an author's revision of the material.

As articles are double-blind reviewed, material that might identify authorship of the paper should be placed only on page 2 as described in the first-4 page format in Pertanika's **INSTRUCTIONS TO AUTHORS** given at the back of this journal.

The Journal's peer-review

In the peer-review process, three referees independently evaluate the scientific quality of the submitted manuscripts.

Peer reviewers are experts chosen by journal editors to provide written assessment of the **strengths** and **weaknesses** of written research, with the aim of improving the reporting of research and identifying the most appropriate and highest quality material for the journal.

Operating and review process

What happens to a manuscript once it is submitted to *Pertanika*? Typically, there are seven steps to the editorial review process:

1. The Journal's chief executive editor and the editorial board examine the paper to determine whether it is appropriate for the journal and should be reviewed. If not appropriate, the manuscript is rejected outright and the author is informed.
2. The chief executive editor sends the article-identifying information having been removed, to three reviewers. Typically, one of these is from the Journal's editorial board. Others are specialists in the subject matter represented by the article. The chief executive editor asks them to complete the review in three weeks.

Comments to authors are about the appropriateness and adequacy of the theoretical or conceptual framework, literature review, method, results and discussion, and conclusions. Reviewers often include suggestions for strengthening of the manuscript. Comments to the editor are in the nature of the significance of the work and its potential contribution to the literature.

3. The chief executive editor, in consultation with the editor-in-chief, examines the reviews and decides whether to reject the manuscript, invite the author(s) to revise and resubmit the manuscript, or seek additional reviews. Final acceptance or rejection rests with the Editor-in-Chief, who reserves the right to refuse any material for publication. In rare instances, the manuscript is accepted with almost no revision. Almost without exception, reviewers' comments (to the author) are forwarded to the author. If a revision is indicated, the editor provides guidelines for attending to the reviewers' suggestions and perhaps additional advice about revising the manuscript.
4. The authors decide whether and how to address the reviewers' comments and criticisms and the editor's concerns. The authors return a revised version of the paper to the chief executive editor along with specific information describing how they have answered the concerns of the reviewers and the editor, usually in a tabular form. The author(s) may also submit a rebuttal if there is a need especially when the author disagrees with certain comments provided by reviewer(s).

5. The chief executive editor sends the revised paper out for re-review. Typically, at least one of the original reviewers will be asked to examine the article.
6. When the reviewers have completed their work, the chief executive editor in consultation with the editorial board and the editor-in-chief examine their comments and decide whether the paper is ready to be published, needs another round of revisions, or should be rejected.
7. If the decision is to accept, an acceptance letter is sent to all the author(s), the paper is sent to the Press. The article should appear in print in approximately three months.

The Publisher ensures that the paper adheres to the correct style (in-text citations, the reference list, and tables are typical areas of concern, clarity, and grammar). The authors are asked to respond to any minor queries by the Publisher. Following these corrections, page proofs are mailed to the corresponding authors for their final approval. At this point, **only essential changes are accepted**. Finally, the article appears in the pages of the Journal and is posted on-line.





Pertanika Journal of Science & Technology
Vol. 25 (2) Apr. 2017

Contents

Foreword	i
<i>Nayan Deep S. Kanwal</i>	
Review Article	
Hydrocarbon Sources for the Carbon Nanotubes Production by Chemical Vapour Deposition: A Review	379
<i>Hayder Baqer Abdullah, Irmawati Ramli I, Ismayadi Ismail and Nor Azah Yusof</i>	
Regular Articles	
An Investigation on High Temperature Erosion Behaviour of Plasma Sprayed CoCrAlY/Al ₂ O ₃ /YSZ on Fe and Ni Based Alloys	397
<i>Nithin H. S., Vijay Desai and M. R. Ramesh</i>	
The Use of Infiltration Wells to Reduce the Impacts of Land Use Changes on Flood Peaks: An Indonesian Catchment Case Study	407
<i>Kusumastuti, D. I., Jokowiarno, D., Khotimah, S. N. and Dewi, C.</i>	
A Hybridised Intelligent Technique for the Diagnosis of Medical Diseases	425
<i>Abdu Masanawa Sagir and Saratha Sathasivam</i>	
Ability of <i>Ipomoea aquatica</i> Forssk. to Remediate Phenol in Water and Effects of Phenol on the Plant's Growth	441
<i>Siew-Yi Lee, Siti Aqlima Ahmad, Siti Roslina Mustapha and Janna Ong-Abdullah</i>	
Robust Artificial Bee Colony in the Hopfield Network for 2-Satisfiability Problem	453
<i>Mohd. Shareduwan Mohd. Kasihmuddin, Mohd. Asyraf Mansor and Saratha Sathasivam</i>	
Estimating Lung Cancer Deaths in Thailand Based on Verbal Autopsy Study in 2005	469
<i>Nattakit Pipatjaturon, Phattrawan Tongkumchum and Attachai Ueranantasun</i>	
Rainfall Trends in the Niger-South Basin, Nigeria, 1948-2008	479
<i>Oloruntade, A. J., Mohammad, T. A. and Aimrun, W.</i>	
Numerical Analysis of The Effect of Nozzle Geometry on Flow Parameters in Abrasive Water Jet Machines	497
<i>Deepak, D., Jodel, A. Q., Cornelio, Midhun Abraham, M. and Shiva Prasad, U.</i>	

An Approach for Identification of Copy-Move Image Forgery based on Projection Profiling	507
<i>Mohd Dilshad Ansari, Satya Prakash Ghrera1 and Mohd Wajid</i>	
Novel Impedance Measurement Technique for Soluble Solid Content Determination of Banana	519
<i>Ibrahim, N. U. A., Abd Aziz, S. and Nawi, N. M.</i>	
Phase Transition and Dielectric Properties of $0.9\text{Pb}(\text{Fe}_{1/2}\text{Nb}_{1/2})\text{O}_3-0.1\text{PbTiO}_3$ Modified with Nano ZnO	527
<i>Hassakorn Wattanasarn, Wattana Photankham, Peerapat Pattumma and Rattikorn Yimnirun</i>	
Selected Papers from the 2nd International Conference on Statistics in Science, Business and Engineering (ICSSBE 2015)	
Guest Editors: Yap Bee Wah & Sayang Mohd Deni	
Some New Bivariate Regression Models	537
<i>Faroughi, P. and Ismail, N.</i>	
Identifying the Uncertain Model Parameter of a Steam Turbine System	545
<i>Wan Munirah, W. M., Tahir, A. and Azmirul, A.</i>	
Decision Making Procedure Based on Jaccard Similarity Measure with Z-numbers	561
<i>Mohamad, D. and Ibrahim, S. Z.</i>	
Sample Size and Non-Normality Effects on Goodness of Fit Measures in Structural Equation Models	575
<i>Ainur, A. K., Sayang, M. D., Jannoo, Z. and Yap, B. W.</i>	
A Method for Performing Short Time Series Prediction	587
<i>Pooi, A. H.</i>	
A Goal Programming Model for Portfolio Optimisation Problem in Fuzzy Environment	593
<i>Mokhtar, M., Shuib, A. and Mohamad, D.</i>	
Performance of Variance Targeting Estimator (VTE) under Misspecified Error Distribution Assumption	607
<i>Abdul Rahim, M. A., Zahari, S. M. and Shariff, S. S. R.</i>	
Review of Context-Based Similarity Measure for Categorical Data	619
<i>Nurul Adzlyana, M. S., Rosma, M. D. and Nurazzah, A. R.</i>	
Assessing Stock Market Volatility for Different Sectors in Malaysia	631
<i>Shakila, S., Noryati, A. and Maheran, M. J.</i>	
A Railway Rescheduling Model with Priority Setting	649
<i>Shuib, A. and Alwadood, Z.</i>	

Foreword

Welcome to the **Second Issue 2017** of the Journal of Science and Technology (JST)!

JST is an open-access journal of science and technology published by Universiti Putra Malaysia Press. It is independently owned and managed by the university and run on a non-profit basis for the benefit of the world-wide science community.

In this issue there are a total of **22 articles**. From the 2nd International Conference on Statistics in Science, Business and Engineering (ICSSBE 2015) ten papers have been selected. Included too is a review article and 11 of our usual specialised contributions.

Hayder Baqer Abdullah, Irmawati Ramli, Ismayadi Ismail and Nor Azah Yusof in their review article discuss hydrocarbon sources for carbon nanotubes production by chemical vapour deposition.

The first regular article in this issue is an investigation on the high temperature erosion behaviour of plasma sprayed CoCrAlY/AL₂O₃/YSZ on Fe and Ni based alloys by *Nithin H. S., Vijay Desai and M. R. Ramesh*. This is followed by papers on: the use of infiltration wells to reduce the impact of land use changes on flood peaks: an Indonesian catchment case study by *Kusumastuti, D. I., Jokowinarno, D., Khotimah, S. N. and Dewi, C.*; a hybridised intelligent technique for the diagnosis of medical diseases by *Abdu Masanawa Sagir and Saratha Sathasivam*; the ability of *ipomoea aquatica* forssk to remediate phenol in water and effects of phenol on the plant's growth by *Siew-Yi Lee, Siti Aqlima Ahmad, Siti Roslina Mustapha and Janna Ong-Abdullah*; robust artificial bee colony in the hopfield network for 2-satisfiability problem by *Mohd. Shareduwan Mohd. Kasihmuddin, Mohd. Asyraf Mansor and Saratha Sathasivam*; estimating lung cancer deaths in Thailand based on verbal autopsy study in 2005 by *Nattakit Pipatjaturon, Phattrawan Tongkumchum, and Attachai Ueranantasan*; rainfall trends in the Niger-south basin, Nigeria, 1948-2008 by *Oloruntade, A. J., Mohammad, T. A. and Aimrun, W.*; numerical analysis of the effect of nozzle geometry on flow parameters in abrasive water jet machines by *Deepak, D., Jodel, A. Q., Cornelio, Midhun Abraham, M. and Shiva Prasad, U.*; an approach for identification of copy-move image forgery based on projection profiling by *Mohd Dilshad Ansari, Satya Prakash Ghrera and Mohd Wajid*; novel impedance measurement technique for soluble solid content determination of banana by *Ibrahim, N. U. A., Abd Aziz, S. and Nawli, N. M.*; phase transition and dielectric properties of 0.9Pb(Fe_{1/2}Nb_{1/2})O₃-0.1PbTiO₃ modified with nano ZnO by *Hassakorn Wattanasarn, Wattana Photankham, Peerapat Pattumma and Rattikorn Yimnirun*.

I conclude this issue with 10 papers from the ICSSBE 2015 international conference. They are : some new bivariate regression models by *Faroughi, P. and Ismail, N.*; identifying the uncertain model parameter of a steam turbine system by *Wan Munirah, W. M, Tahir, A. and Azmirul, A.*; decision making procedure based on jaccard similarity measure with z-numbers by *Mohamad, D. and Ibrahim, S. Z.*; sample size and non-normality effects on goodness of fit measures in structural equation models by *Ainur, A. K., Sayang, M. D., Jannoo, Z. and Yap, B. W.*; a method for performing short time series prediction by *Pooi, A. H.*; a goal programming model for portfolio optimisation problem in fuzzy environment by *Mazura, M., Adibah, S. and Daud, M.*; performance of variance targeting estimator

(vte) under misspecified error distribution assumption by *Abdul Rahim, M. A., Zahari, S. M. and Shariff, S. S. R*); review of context-based similarity measure for categorical data by *Nurul Adzlyana, M. S., Rosma, M. D. and Nurazzah, A. R.*; accessing stock market volatility for different sectors in Malaysia by *Shakila, S., Noryati, A. and Maheran, M. J.*; a railway rescheduling model with priority setting by *Shuib, A. and Alwadood, Z.*

I believe readers will find the evidence presented in this issue to be intriguing, thought-provoking, and, useful for further research. Please recommend the journal to your colleagues and students to make this endeavour meaningful.

I would like to take this opportunity to express my gratitude and thanks to all the contributors, reviewer, editors, and editorial assistance of the journal division staff.

JST is currently accepting manuscripts for upcoming issues based on original qualitative or quantitative research that opens new areas of inquiry and investigation.

Chief Executive Editor

Nayan Deep S. KANWAL, [FRSA](#), [ABIM](#), [AMIS](#), Ph.D.

nayan@upm.my

Review Article

Hydrocarbon Sources for the Carbon Nanotubes Production by Chemical Vapour Deposition: A Review

Hayder Baqer Abdullah^{1,3}, Irmawati Ramli^{1,2,*}, Ismayadi Ismail² and Nor Azah Yusof²

¹*Catalysis Science and Technology Research Centre, Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

²*Institute of Advanced Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

³*Department of Chemistry, College of Education for Pure Science, University of Basrah, 61004 Basrah, Iraq*

ABSTRACT

The synthesis of carbon nanotubes (CNTs) using a chemical vapour deposition (CVD) method requires the use of hydrocarbon as the carbon precursor. Among the commonly used hydrocarbons are methane and acetylene, which are both light gas-phase substances. Besides that, other carbon-rich sources, such as carbon monoxide and coal, have also been reportedly used. Nowadays, researches have also been conducted into utilising heavier hydrocarbons and petrochemical products for the production of CNTs, such as kerosene and diesel oil. Therefore, this article reviews the different kind of hydrocarbon sources for CNTs production using a CVD method. The method is used for it allows the decomposition of the carbon-rich source with the aid of a catalyst at a temperature in the range 600-1200°C. This synthesis technique gives an advantage as a high yield and high-quality CNTs can be produced at a relatively low cost process. As compared to other processes for CNTs production such as arc discharge and laser ablation, they may produce high quality CNTs but has a disadvantage for use as large scale synthesis routes.

Keywords: Hydrocarbon source, carbon nanotube, catalyst, chemical vapour deposition

Article history:

Received: 27 April 2016

Accepted: 30 August 2016

E-mail addresses:

haybaiq@gmail.com (Hayder Baqer Abdullah),

irmawati@upm.edu.my (Irmawati Ramli),

kayzen@gmail.com (Ismayadi Ismail),

azahy@upm.edu.my (Nor Azah Yusof)

*Corresponding Author

INTRODUCTION

Carbon nanotubes (CNTs) have received much attention since their discovery by Iijima in 1991 (Iijima, 1991). CNTs display unique mechanical, electrical, optical and chemical characteristics (Eatemadi et al., 2014; Kotsilkova et al., 2015). These particles

are used for clean energy applications, such as a solar cell (Wang et al., 2015), hydrogen storage (Su et al., 2016) and supercapacitor (Lva et al., 2012). CNTs are also applied in micro and nano electromechanical system (Fedorovskaya et al., 2016) and high-performance low-detecting limit sensor (Cheraghi & Taher, 2016). These nanoparticles can be categorised into single-walled CNTs (SWCNTs), with a typical diameter in the range 0.6-3 nm, and multi-walled CNTs (MWCNTs) with a diameter in the range of 10-50 nm. SWCNTs have a rolled of graphene sheet to form a cylinder, whereas MWCNTs consist of multiple cylindrical rolled graphene sheets (Camilli et al., 2011) (Figure 1). CNTs are generally synthesised through arc discharge, laser ablation, and chemical vapour deposition (CVD). Both arc discharge and laser ablation are practical for synthesising SWCNTs (Ka et al., 2012; Maria & Mieno, 2015). The technique involves the use of graphite (Liew et al., 2014), and recently, coal has also been used as a source material that was vapourised and later deposited as CNTs (Moothi et al., 2012). High quality CNTs are produced but the high energy consumption in generating the discharge for the former and laser processing for the later contributed to high costs of equipment and energy. On the other hand, CVD is considered to be the most convenient method to grow all varieties of CNTs. Moreover, CVD is the best choice for large-scale production of CNTs at a relatively low cost under mild growth conditions (Szabó et al., 2010). The utilisation of hydrocarbon sources for the production of high-value chemicals and materials such as CNTs has an obvious impact on sustainable development. As a result, several research groups have explored the use of hydrocarbons as a carbon source for CNT synthesis.

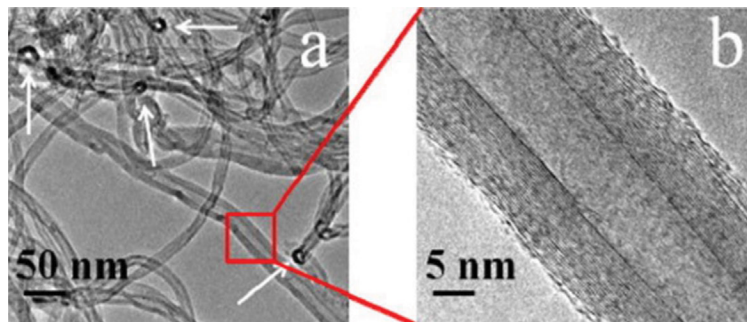


Figure 1. TEM images showing the multi-wall nature of CNTs (Camilli et al., 2011)

SYNTHESIS OF CARBON NANOTUBES BY CHEMICAL VAPOUR DEPOSITION METHOD

CVD is the most commonly used technique to grow all varieties of CNTs. First, a catalyst and a carbon precursor easily react in CVD because this technique only requires an oven designed with a tubular reactor such as a quartz tube and a few gas flow controllers (Nerushev et al., 2001; Hashempour et al., 2014). Second, numerous parameters can be investigated, not only during catalyst treatment but also during CNTs growth, which influences the quality, purity and yield of the CNTs (Cantoro et al., 2006; Inami et al., 2007; Bansal et al., 2010).

Temperature is the vital parameter for CNTs synthesis by CVD. Hydrocarbons are decomposing into reactive radical objects in the temperature range of 600–1200°C during CVD. These reactive species diffuse to a catalyst surface where they remain bonded, forming CNTs. The commonly used energy source is resistive heating (Lee et al., 2001; Gohier et al., 2008; Chiangga et al., 2009).

Two processing system patterns, namely horizontal and vertical, are generally used to synthesise CNTs using CVD. A typical horizontal system has been demonstrated because of its versatility. Floating and fixed-bed catalyst is used in the horizontal system (Figure 2). Floating catalysts are mainly applied when a mixture of reactants and catalyst are in the gas phase in the reactor at an elevated temperature during CVD. The catalyst in the gas phase is transformed by the redox gases and/or the elevated temperature, forming solid-phase nanoparticles where the CVD reaction occurs. This method overcomes the obstacle that prevents the coalescence of nanoparticles and reactants. The solid catalyst nanoparticles can have adequate residence time for CNTs to grow when they hold on to the reactor surfaces. In other words, any unreacted gas precursor and solid catalyst nanoparticles, which cannot react and hold on the reactor surfaces at adequately elevated temperatures, are swept away from the reactor with the carrier gases and unused reactants. Therefore, this phenomenon reduces the process efficiency (Zhou et al., 2003; Chena et al., 2006; Maghrebi et al., 2009).

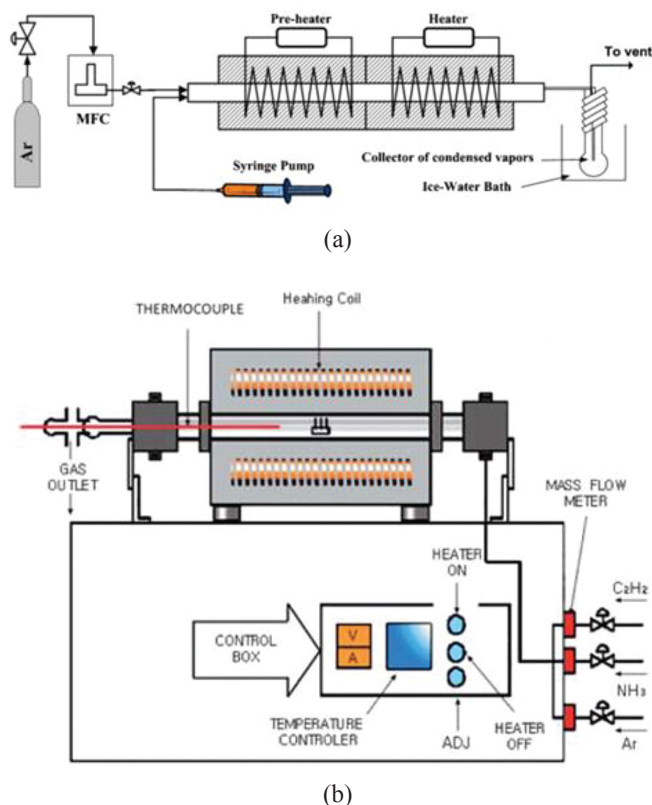


Figure 2. Horizontal CVD: (a) Floating technique; (b) Fixed bed technique (Maghrebi et al., 2009; Nessim, 2010).

By contrast, the catalyst in the fixed bed process is in the solid phase. The catalyst is set in boats and placed inside the reactor. Then, the gas-phase reactant and a carrier are introduced at an elevated temperature where they react during the CVD. The efficiency and productivity of CNTs growth in this system are limited by the heterogeneous contact between the gaseous state and solid reactant as well as the rise in the gradient temperature of the reactor (Koziol et al., 2011). Increasing nanotubes on the catalyst surface leads to a reduction in the diffusion of carbon precursor to the catalyst nanoparticles, thereby reducing the effectiveness of the catalysts (Nessim, 2010; Izadi et al., 2011).

CATALYST USED IN THE CHEMICAL VAPOUR DEPOSITION METHOD

Catalysts play an essential role in the synthesis of CNTs during CVD. The quality and yield of CNTs are improved by enhancing the suitable characteristics of the desired catalyst. Transition metals, such as iron (Fe) (Emmeneggera et al., 2003), cobalt (Co) (Chen et al., 2006), nickel (Ni) (Jehng et al., 2007) and their alloys (Flahaut et al., 2005; Niu & Fang, 2006; Policicchio et al., 2007; Lobiak et al., 2015), are the most effective catalysts for CNT synthesis. These catalysts can grow CNTs in three steps according to the vapour–liquid–solid mechanism. First, a gas precursor produces carbon, which absorbs and dissociates on the surface of the catalyst particles to form elementary carbon atoms. Second, the carbon atoms dissolve in the mass of nanoparticles to form liquid metastable carbide and diffuse within the particles. Finally, solid carbons precipitate at the backside of the nanoparticles to form CNTs.

Two reverse processes cause CNTs growth. The first process is the well-known “root growth model,” which occurs when the catalyst is strongly bound to the substrate, resulting in a closed carbon network. This characteristic then gives rise to the metal in the form of a seamless graphite cylinder. A second process, which is also known as the “tip growth model,” occurs because of the weak contact between the catalyst and the substrate, and CNTs precipitation pushes the metal particles up from the substance. Therefore, diffusion occurs under the metal particle, and the CNTs continues to elongate (Figure 3) (Kukovitsky et al., 2000; Zhao et al., 2006; Kumar, 2011; Bathgate et al., 2012; Purohit et al., 2014). Solid organometallobenes, such as nickelocene, cobaltocene and ferrocene (Malek Abbaslou et al., 2010; Zhang et al., 2011; Zhu et al., 2014), are extensively used as catalysts because they deliver solid metal nanoparticles, which affect the decomposition of the hydrocarbons. Experimentally, catalyst particle size affects the diameter of the CNTs. Figure 3(c) summarises the effects of the diameter of the catalyst on the structure of the utilised nanocarbon. Catalyst nanoparticles with 1 nm diameter mainly produce SWCNTs, whereas MWCNTs are produced from catalyst nanoparticles with diameters in the range of 10–50 nm. The number of MWCNT layers increases with catalyst diameter. Additionally, another nanocarbon structure, i.e. the nano-onion, is produced when the diameter of the catalyst nanoparticles exceeds 50 nm (Gore & Sane, 2011).

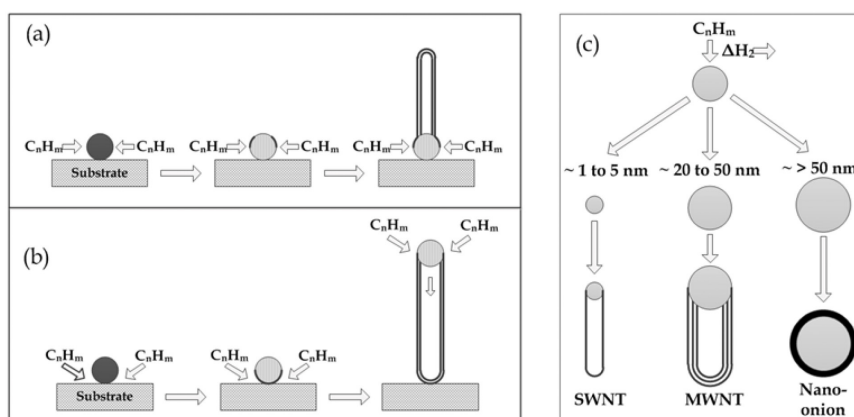


Figure 3. Mechanisms for carbon nanotube growth; (a) Base growth; (b) Tip growth; and (c) Structural dependence on catalyst particle size (Gore & Sane, 2011).

Hydrocarbon decomposition on the catalyst surface, that liberates and prevents aerial decomposition, is the key to obtaining pure CNTs. Furthermore, alloys have been proven to gain higher catalytic activity compared with pure metals (Lobiak et al., 2015). Iron (Fe), cobalt (Co), and nickel (Ni) nanoparticles are effective catalysts for CNTs synthesis. Other metals, such as copper (Cu), gold (Au), palladium (Pd), and platinum (Pt), have also been discovered to be catalysts for CNT growth from a variety of hydrocarbons (Lee et al., 2000; Seung-Yup et al., 2005; Takenaka et al., 2009; Yang et al., 2012; Asokan et al., 2015).

HYDROCARBON DECOMPOSITION AND THERMODYNAMICS

Thermodynamics are useful in elucidating the reaction during CVD. The reaction needed to create solid CNTs is thermodynamically useful under the selected temperature and pressure conditions. Gibbs free energy (ΔG) is a key to extract this information. ΔG depends on the reactivity of the hydrocarbon during pyrolysis. Methane decomposition is thermodynamically preferable above 600 °C because it is the most stable hydrocarbon molecule. By contrast, ΔG for ethylene, acetylene, or benzene is already negative at 200°C (Figure 4). CNTs are frequently synthesised by CVD using methane as a precursor at a temperature above 850°C because of the slow reaction rate. Meanwhile, CVD synthesis of CNTs using acetylene is often performed at 500°C - 750°C, and 650°C - 850°C for ethylene (See & Harris, 2007; MacKenzie et al., 2010; Tessonnier & Su, 2011).

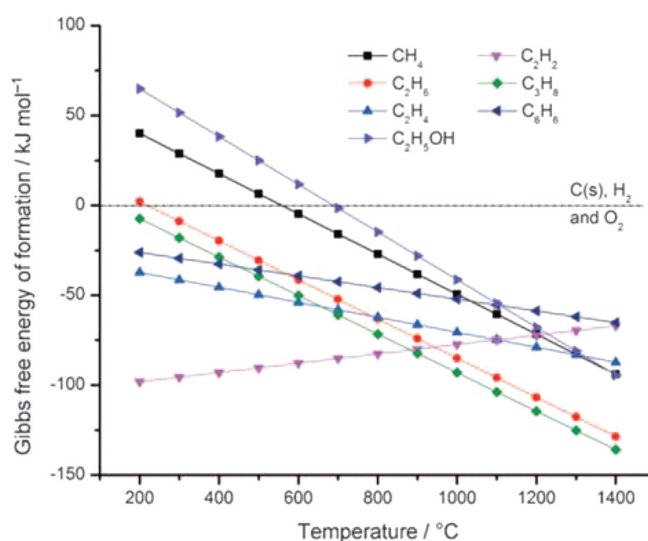


Figure 4. Thermodynamic data calculated using ChemKin database. Gibbs free energies of formation of various carbon precursors. The energies are normalised to the number of carbon atoms in the precursor and correspond to its pyrolysis (Tessonier & Su, 2011).

HYDROCARBON PRECURSORS FOR SYNTHESIS CARBON NANOTUBES

The most common hydrocarbon sources used in CNTs production by CVD are derived from either petrochemical products (Matar & Hatch, 2001), such as methane, ethane, acetylene, and cyclohexane, or petroleum products (Jones, 2006) such as natural gas, kerosene and diesel oil (Table 1).

Table 1
Hydrocarbon sources for the production of CNTs by CVD method

Petrochemical Products						
Hydrocarbon Source	Catalyst	CVD type	Temperature (°C)	Carrier gas	CNT type	Ref.
Methane	Co-Mo-MgO	CVD	1500-1750	N ₂ , H ₂	SWCNT	Izadi et al. (2011)
Methane	Co-Mo/MgO	CCVD	800	N ₂	MWCNT	Hata et al. (2004)
Ethane	Fe/Al ₂ O ₃	CVD	750	H ₂	MWCNT	Maghsoodi et al. (2010)
Acetylene	Fe/ SiO ₂	CVD	690	N ₂ , H ₂	MWCNT	Allaadini et al. (2015)
Acetylene	Co-Mo/MgO	CVD	800	H ₂	SWCNT, MWCNT	Zhan et al. (2011)
Hexane	Ferrocene	Two stage CVD	140, 900	N ₂ , H ₂	MWCNT	Tran et al. (2007)
Cyclohexane	Ferrocene	ACVD	800	N ₂ , H ₂	MWCNT	Kishore et al. (2012)

Table 1 (continue)

Petrochemical Products						
Hydrocarbon Source	Catalyst	CVD type	Temperature (°C)	Carrier gas	CNT type	Ref.
Benzene	Ferrocene	Spray Pyrolysis	900	Ar	MWCNT	Escobar (2007)
Toluene	Ferrocene	CVD	850	Ar	MWCNT	Altalhi et al. (2010)
Xylene	Ferrocene	Floating catalyst	750	Ar, H ₂	MWCNT	Kunadian et al. (2009)
Petroleum Products						
Hydrocarbon Source	Catalyst	CVD type	Temperature (°C)	Carrier gas	CNT type	Ref.
Natural Gas	Fe-Mo/MgO	CCVD	950- 1000	Ar, H ₂	CNT	Tapaszto et al. (2005)
Propane	Fe/ SiO ₂	CCVD	600- 1100	Ar, H ₂	SWCNT, MWCNT	Bachmatiuk et al. (2008)
Liquefied petroleum gas	Ferrocene	Two stage CVD	800	Ar, H ₂	MWCNT	Huang et al. (2008)
Kerosene	Fe, Ni	Two stage CVD	200, 1000	N ₂	CNT	Altalhi et al. (2010)
Diesel oil	-	Modified CVD	950	-	SWCNT	Kunadian et al. (2009)
Coal gas	Ferrocene	Two stage CVD	200, 950	-	SWCNT	Qiu et al. (2004)
Heavy oil residue	Fe, Co, Ni, Au, Pt	CVD	900	Ar, H ₂	SWCNT	Li et al. (2012)
Asphalt	Ferrocene	CVD	1000	Ar, H ₂	CNT	Liu et al. (2007)

The carbon precursor plays an important role in the growth, characteristics and properties of CNTs because of their binding energy, type and the role of reactive groups, as well as thermodynamic properties. CNTs growth efficiency using gaseous carbon precursors depends strongly on the concentration and reactivity of gas-phase intermediates produced simultaneously with reactive radical species as a result of hydrocarbon pyrolysis. Therefore, the most capable intermediates with ability of physisorption or chemisorption on the catalyst surface to initiate CNTs growth must be produced in the gas phase (Mizuno et al., 2005; Prasek et al., 2011). A comparison of CNTs characteristics showed a relationship between the chemical structures of hydrocarbons and CNTs formation (Lee et al., 2003; Li et al., 2004). Hernadi et al. (2000) affirmed that unsaturated hydrocarbons have much higher yield and deposition rate than saturated gases. Additionally, saturated hydrocarbon gases produce highly graphitised filaments with fewer walls compared with unsaturated gases. Consequently, saturated hydrocarbons are favoured for SWCNTs growth, whereas unsaturated hydrocarbons are favoured for MWCNTs (Hernadi et al., 2000). Nevertheless, SWCNTs have been obtained from a highly diluted unsaturated hydrocarbon (Q. Li et al., 2004; Yun et al., 2006; Zheng et al., 2006; Escobar et al., 2007). The growth of clean SWCNTs was observed at relatively low temperatures using

alcohols with various catalysts. Thus, alcohols are much better carbon sources for SWCNTs than hydrocarbons. This phenomenon is probably due to the ability of OH radicals to etch away amorphous carbon deposits (Li et al., 2007; Zhu et al., 2012; Ren et al., 2014).

Most studies have shown that low-temperature CVD at approximately 600°C - 800°C produces MWCNTs, whereas high reaction temperatures of 800°C - 1200°C favour SWCNTs. The results indicate that SWCNTs have the higher energy of formation, caused by their small diameter, and high curvature, which tolerate the high strain energy. Therefore, SWCNTs grow from only selected carbon source, such as carbon monoxide, and methane, which has an equitable stability at high temperature. By contrast, common effective precursors for MWCNTs such as acetylene, benzene, and xylene are unstable at high temperatures, resulting in a large deposit of amorphous carbon (Niu & Fang, 2006; Loebick et al., 2009; Bahgat et al., 2011; Gallego et al., 2011; Monthieux, 2011; Atchudan & Pandurangan, 2012; Abdullahi et al., 2014).

Hata et al. (2004) synthesised highly efficient and pure SWCNTs by water-assisted ethylene on the substrate using CVD. The controllable rate of steam into the CVD reactor functioned as a mild oxidizer, leading to the selective removal of amorphous carbon without adversely affecting the growth of CNTs (Hata et al., 2004). Controlling the relative rate of ethylene and steam was essential to minimize the poison of the catalyst. These studies ascertain that carbon precursors play an important part in CNTs growth. Accurate selection of carbon precursor and water vapour rate can optimise the growth and lifetime of CNTs, thereby increasing the quality and yield of these nanoparticles (Zhong et al., 2009; Saengmee-anupharb et al., 2011).

Hydrocarbons, such as methane (Maghsoodi et al., 2010; Allaedini et al., 2015), ethylene (Tran et al., 2007; Zhan et al., 2011), acetylene (Hayashi et al., 2011; Kishore & Pandurangan, 2012), benzene (Narkiewicz et al., 2010; Liu et al., 2011), cyclohexane (Chena et al., 2006; Bachmatiuk et al., 2008), toluene (Altalhi et al., 2010; Gspann et al., 2014) and xylene (Kunadian et al., 2009; Yah et al., 2011) are among the most commonly used CNTs precursors. MWCNTs are typically grown from the decomposition of benzene at 1100°C and acetylene at 700°C, with Fe nanoparticles as catalyst. MWCNTs have also been grown from various other hydrocarbon precursors such as cyclohexane and fullerene (Maruyama et al., 2003). MWCNTs have also been synthesized from supercritical toluene at 600°C and ferrocene as growth catalysts. Toluene served not only as the carbon source for the nanotube formation but also as the solvent (Lee et al., 2004).

HIGH MOLECULAR WEIGHT CARBON PRECURSOR SOURCES

In addition to the well-defined petrochemical products described in section 5, CNTs have also been successfully and systematically synthesised from local fuels, such as kerosene (Pradhan & Sharon, 2002), liquefied petroleum gas (LPG) (Huang et al., 2008), natural gas (Tapasztó et al., 2005), coal gas (Qiu et al., 2004) and asphalt (Liu et al., 2007).

Coiled and open-ended nanotubes are from commercial kerosene using Ni and Fe catalysts via CVD. These metals assist the growth of straight and coiled nanotubes at 1000°C. Increasing the decomposition temperature to 1100°C leads to the formation of carbon nanobeads, which are a spherical, solid and hollow carbon nanomaterial of size 600-900 nm (Pradhan & Sharon, 2002). These nanobeads are rarely stated in literature (Sharon et al., 1998).

SWCNTs were effectively synthesised from coal gas over ferrocene catalyst in a two-stage catalytic CVD technique. The temperature of the first chamber was maintained at 200°C, whereas that in the second chamber was increased to 850–950°C (Qiu et al., 2004). Another study showed that iron carbide–oxide-filled CNTs were successfully synthesised from coal gas by CVD using ferrocene as a catalyst in a two-stage furnace. The effective reaction temperature was 180°C and 950°C for the first and second furnaces, respectively. Transmission electron micrographs show that the CNTs were filled with iron carbide–oxide (Figure 5) (Qiu et al., 2006).

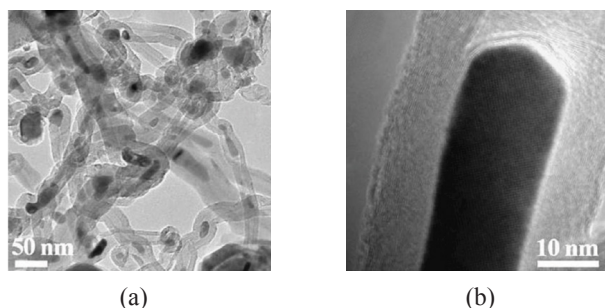


Figure 5. TEM images of iron carbide–oxide filled CNTs obtained via CVD of coal-gas: (a) low-resolution TEM image of partly filled CNTs; (b) high-resolution TEM image of partially filled CNT (Qiu et al., 2006).

Several reports have successfully used LPG as a carbon source for CNT synthesis (Huang et al., 2008; Qian et al., 2002; Zhang et al., 2007; Zhang et al., 2008). Qian et al. (2002) reported the direct synthesis of agglomerated MWCNTs from the pyrolysis of LPG containing sulphur with Fe/Mo/Al₂O₃ catalyst. The synthesised MWCNTs were similar to that attained from propylene, which is a pure carbon source (Figure 6) (Qian et al., 2002). LPG has been employed as a carbon source to produce CNTs arrays on ceramic spherical surfaces in the floating catalyst process in a two-stage furnace. Improved alignments of CNTs have been obtained, with purity as high as 97.5%. Controlling the growth temperature resulted in CNTs with aligned form and a diameter of approximately 13 nm. Thus, large-scale and low-cost production of CNTs arrays are feasible using synthesised industrial fuel as a carbon source and a ceramic substrate (Zhang et al., 2007).

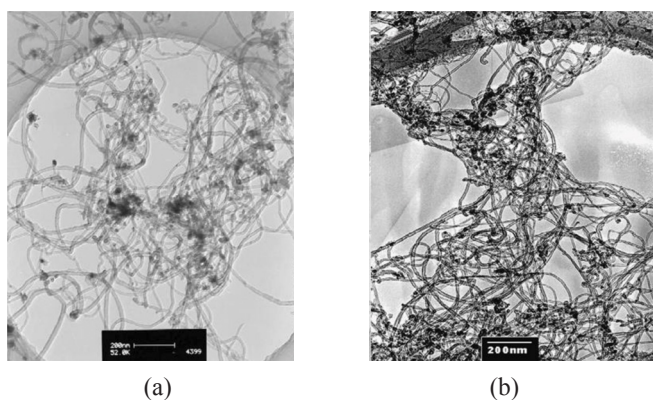


Figure 6. The morphology of as-grown CNTs from: (a) propylene; (b) LPG decomposition (Qian et al., 2002).

CNTs were also reportedly synthesised from heavy petroleum fractions, such as light diesel oil (LDO) and furnace oil (FO) by modifying the CVD method characterised by transmission electron microscopy (TEM) and scanning electron microscopy (SEM). A local reactor was designed to synthesise and collect soot. Collected soot was purified by a Soxhlet extraction apparatus. The purified CNTs were then oxidised with diluted nitric acid. The utilized CNTs were dispersed in different solvents. The dispersed samples were analysed at various temperatures. The results demonstrate that the soot was highly dispersed in distilled water and acetone compared with that in ethanol and methanol. SWCNTs had diameters of approximately 70 nm from FO and 90 nm from LDO (Singh et al., 2011).

In another work, it was reported the synthesis of SWCNTs were synthesised by CVD utilising heavy oil residues as a carbon source. Two categories of metals, namely, magnetic transition metals such as Fe, Co, and Ni, and nonmagnetic metals, namely, Au and Pt, were used as catalysts in the growth of SWCNTs.

The structure and morphology of the synthesised SWCNTs were characterised by Raman spectroscopy, atomic force microscopy, SEM, and TEM. The results demonstrated that high-quality SWCNTs can be synthesised by a CVD method with inexpensive heavy oil residue as the starting material. The diameter distribution of as-grown SWCNTs strongly depends on the type of catalysts. SWCNTs grown from transition metals (e.g., Fe, Co, and Ni) have smaller diameters compared with those synthesised from nonmagnetic catalysts (i.e., Au and Pt). This outcome establishes the feasibility of controlling SWCNT diameters by selecting the catalysts. Additionally, the reaction temperature is the key factor affecting the formation of SWCNTs from oil residues (Li et al., 2012).

The production of CNTs from waste oil has been investigated when Suriani et al. (2015) synthesised for the first time quasi-aligned CNTs from the decomposition of waste engine oil as a carbon source mixed with ferrocene as a catalyst in a two-stage furnace. The synthesis was performed at 500°C and 570°C for precursor and CNTs synthesis, respectively, with a successful ferrocene catalyst ratio of 17.99wt%. Field emission SEM images of quasi-aligned CNTs produced regulated diameters of 18.0 nm to 29.8 nm with lengths from 11.8 µm to 17.1 µm. TEM, SEM, and micro-Raman analysis exposed a dense mixture of quasi-aligned SWCNTs and MWCNTs (Suriani et al., 2015).

CONCLUSION

The synthesis of various CNTs either single-walled or multi-walled is influenced by the type of hydrocarbon source used in CVD method. The process parameters involved, which are the type of catalyst, carrier gas, reaction temperature and reaction time, also contributed to the different structure of CNTs. However, the type of hydrocarbon precursor is conclusively playing an important role in CNTs growth. Therefore, this review paper looked into the effect of utilising different types of hydrocarbon source derived from petroleum and petroleum products in affecting the formation of carbon structures. Through CVD method, a controlled synthesis ambience for the carbon vaporization and nucleation lead to the formation of CNTs with desired quality and improved production yield. The method also favours solubility of hydrocarbon source and provides the conditions and controlled synthesis, which allows for the substrate-catalyst interaction.

There is a clear parallel between petroleum and petrochemical products such as hydrocarbons sources, suggesting that the decomposition of hydrocarbon source at elevated temperature and diffuses through catalytic particle is a key factor in the CNTs growth.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Research Management Centre, Universiti Putra Malaysia, for providing the research grant (GP-IPS/2015/9465500).

REFERENCES

- Abbaslou, R. M. M., Soltan, J., & Dalai, A. K. (2010). The effects of carbon concentration in the precursor gas on the quality and quantity of carbon nanotubes synthesized by CVD method. *Applied Catalysis A: General*, 372(2), 147-152. doi:<http://dx.doi.org/10.1016/j.apcata.2009.10.025>
- Abdullahi, I., Sakulchaicharoen, N., & Herrera, J. E. (2014). Selective synthesis of single-walled carbon nanotubes on Fe–MgO catalyst by chemical vapor deposition of methane. *Diamond and Related Materials*, 41, 84-93. doi:10.1016/j.diamond.2013.11.008
- Allaedini, G., Tasirin, S. M., & Aminayi, P. (2015). Synthesis of CNTs via Chemical Vapor Deposition of Carbon Dioxide as a Carbon Source in the Presence of NiMgO. *Journal of Alloys and Compounds*, 647(0), 809-814. doi:<http://dx.doi.org/10.1016/j.jallcom.2015.06.012>
- Altalhi, T., Ginic-Markovic, M., Han, N., Clarke, S., & Losic, D. (2010). Synthesis of Carbon Nanotube (CNT) Composite Membranes. *Membranes (Basel)*, 1(1), 37-47. doi:10.3390/membranes1010037
- Asokan, V., Myrseth, V., & Kosinski, P. (2015). Effect of Pt and Fe catalysts in the transformation of carbon black into carbon nanotubes. *Journal of Physics and Chemistry of Solids*, 81, 106-115. doi:10.1016/j.jpcs.2015.02.006
- Atchudan, R., & Pandurangan, A. (2012). The use of bimetallic MCM-41 mesoporous catalysts for the synthesis of MWCNTs by chemical vapor deposition. *Journal of Molecular Catalysis A: Chemical*, 355, 75-84. doi:10.1016/j.molcata.2011.11.028
- Bachmatiuk, A., Kalenczuk, R. J., Rummeli, M. H., Gemming, T., & Borowiak-Palen, E. (2008). Preparation of ultra-large-scale catalysts for catalytic vapour deposition of carbon nanotubes. *Materials Science-Poland*, 26(1), 105-111.
- Bahgat, M., Farghali, A. A., El Roubay, W. M. A., & Khedr, M. H. (2011). Synthesis and modification of multi-walled carbon nano-tubes (MWCNTs) for water treatment applications. *Journal of Analytical and Applied Pyrolysis*, 92(2), 307-313. doi:10.1016/j.jaap.2011.07.002
- Bansal, M., Lal, C., Srivastava, R., Kamalasanan, M. N., & Tanwar, L. S. (2010). Comparison of structure and yield of multiwall carbon nanotubes produced by the CVD technique and a water assisted method. *Physica B: Condensed Matter*, 405(7), 1745-1749. doi:<http://dx.doi.org/10.1016/j.physb.2010.01.031>
- Bathgate, G., Iyuke, S., & Kavishe, F. (2012). Comparison of Straight and Helical Nanotube Production in a Swirled Fluid CVD Reactor. *ISRN Nanotechnology*, 2012, 1-10. doi:10.5402/2012/985834
- Camilli, L., Scarselli, M., Del Gobbo, S., Castrucci, P., Nanni, F., Gautron, E., Lefrant, S., & De Crescenzi, M. (2011). The synthesis and characterization of carbon nanotubes grown by chemical vapor deposition using a stainless steel catalyst. *Carbon*, 49(10), 3307-3315. doi:10.1016/j.carbon.2011.04.014

- Cantoro, M., Hofmann, S., Pisana, S., Scardaci, V., Parvez, A., Ducati, C., Ferrari, A. C., Blackburn, A. M., Wang, K. Y., & Robertson, J. (2006). Catalytic Chemical Vapor Deposition of Single-Wall Carbon Nanotubes at Low Temperatures. *Nano Lett*, 6(6), 1107-1112.
- Chen, Y., Ciuparu, D., Lim, S., Haller, G. L., & Pfefferle, L. D. (2006). The effect of the cobalt loading on the growth of single wall carbon nanotubes by CO disproportionation on Co-MCM-41 catalysts. *Carbon*, 44(1), 67-78. doi:10.1016/j.carbon.2005.07.035
- Chena, Y., Suna, Z., Li, Y. N., & Tay, B. K. (2006). Optimization of carbon nanotube powder growth using low pressure floating catalytic chemical vapor deposition. *Materials Chemistry and Physics*, 98(2-3), 256-260. doi:10.1016/j.matchemphys.2005.09.017
- Cheraghi, S., & Taher, M. A. (2016). Fabrication of CdO/single wall carbon nanotubes modified ionic liquids carbon paste electrode as a high performance sensor in diphenhydramine analysis. *Journal of Molecular Liquids*, 219, 1023-1029. doi:http://dx.doi.org/10.1016/j.molliq.2016.04.035
- Chiangga, S., Suttisiri, N., & Nilsaengrat, P. (2009). Effect of temperature on carbon nanotubes growth on thin Iron film by thermal chemical vapor deposition method under the low pressure. *Physics Procedia*, 2(1), 107-111. doi:10.1016/j.phpro.2009.06.017
- Eatemadi, A., Daraee, H., Karimkhanloo, H., Kouhi, M., Zarghami, N., Akbarzadeh, A., Abasi, M., Hanifehpour, Y., & Joo, S. W. (2014). Carbon nanotubes: properties, synthesis, purification and medical applications. *Nanoscale Res Lett*, 9(1), 393-405. doi:10.1186/1556-276x-9-393
- Emmenegger, C., Bonard, J. M., Mauron, P., Sudan, P., Lepora, A., Grobety, B., ... Schlapbach, L. (2003). Synthesis of carbon nanotubes over Fe catalyst on aluminium and suggested growth mechanism. *Carbon*, 41(3), 539-547.
- Escobar, M., Moreno, M. S., Candal, R. J., Marchi, M. C., Caso, A., Polosecki, P. I., ... Goyanes, S. (2007). Synthesis of carbon nanotubes by CVD: Effect of acetylene pressure on nanotubes characteristics. *Applied Surface Science*, 254(1), 251-256. doi:http://dx.doi.org/10.1016/j.apsusc.2007.07.044
- Fedorovskaya, E. O., Apartsin, E. K., Novopashina, D. S., Venyaminova, A. G., Kurenaya, A. G., Bulusheva, L. G., & Okotrub, A. V. (2016). RNA-modified carbon nanotube arrays recognizing RNA via electrochemical capacitance response. *Materials and Design*, 100, 67-72. doi:http://dx.doi.org/10.1016/j.matdes.2016.03.110
- Flahaut, E., Laurent, C., & Peigney, A. (2005). Catalytic CVD synthesis of double and triple-walled carbon nanotubes by the control of the catalyst preparation. *Carbon*, 43(2), 375-383. doi:10.1016/j.carbon.2004.09.021
- Gallego, J., Sierra, G., Mondragon, F., Barrault, J., & Batiot-Dupeyrat, C. (2011). Synthesis of MWCNTs and hydrogen from ethanol catalytic decomposition over a Ni/La₂O₃ catalyst produced by the reduction of LaNiO₃. *Applied Catalysis A: General*, 397(1-2), 73-81. doi:10.1016/j.apcata.2011.02.017
- Gohier, A., Ewels, C. P., Minea, T. M., & Djouadi, M. A. (2008). Carbon nanotube growth mechanism switches from tip- to base-growth with decreasing catalyst particle size. *Carbon*, 46(10), 1331-1338. doi:http://dx.doi.org/10.1016/j.carbon.2008.05.016
- Gore, J. P., & Sane, A. (2011). Flame synthesis of carbon nanotubes. In S. Yellampalli (Ed.), *Carbon Nanotubes - Synthesis, Characterization, Applications* (pp.121-146). Croatia: InTech.
- Gspann, T. S., Smail, F. R., & Windle, A. H. (2014). Spinning of carbon nanotube fibres using the floating catalyst high temperature route: purity issues and the critical role of sulphur. *Faraday Discuss*, 173, 47-65. doi:10.1039/c4fd00066h

- Hashempour, M., Vicenzo, A., Zhao, F., & Bestetti, M. (2014). Effects of CVD direct growth of carbon nanotubes and nanofibers on microstructure and electrochemical corrosion behavior of 316 stainless steel. *Materials Characterization*, *92*, 64-76. doi:http://dx.doi.org/10.1016/j.matchar.2014.03.001
- Hata, K., Futaba, D. N., Mizuno, K., Namai, T., Yumura, M., & Iijima, S. (2004). Water-assisted highly efficient synthesis of impurity-free single-walled carbon nanotubes. *Science*, *306*(5700), 1362-1364.
- Hayashi, Y., Iijima, T., Miyake, M., Satoh, M., Rupesinghe, N. L., Teo, K. B. K., & Tanmemura, M. (2011). Growth evolution of rapid grown aligned carbon nanotube forests without water vapor on Fe/Al₂O₃/SiO₂/Si substrate. *Diamond and Related Materials*, *20*(7), 859-862. doi:http://dx.doi.org/10.1016/j.diamond.2011.02.003
- Hernadi, K., Fonseca, A., Nagy, J. B., Siska, A., & Kiricsi, I. (2000). Production of nanotubes by the catalytic decomposition of different carbon-containing compounds. *Applied Catalysis A: General*, *199*(2), 245-255.
- Huang, J., Zhang, Q., Wei, F., Qian, W., Wang, D., & Hu, L. (2008). Liquefied petroleum gas containing sulfur as the carbon source for carbon nanotube forests. *Carbon*, *46*(2), 291-296. doi:10.1016/j.carbon.2007.11.044
- Iijima, S. (1991). Helical microtubules of graphitic carbon. *Nature*, *354*(6348), 56-58.
- Inami, N., Mohamed, M. A., Shikoh, E., & Fujiwara, A. (2007). Synthesis-condition dependence of carbon nanotube growth by alcohol catalytic chemical vapor deposition method. *Science and Technology of Advanced Materials*, *8*(4), 292-295. doi:10.1016/j.stam.2007.02.009
- Izadi, N., Rashidi, A. M., Horri, B. A., Mosoudi, M. R., Bozorgzadeh, H. R., & Zeraatkar, A. (2011). Growth of single-walled carbon nanotubes on a Co–Mo–MgO supported catalyst by the CVD of methane in a fixed bed reactor: Model setting and parameter estimation. *Solid State Sciences*, *13*(6), 1242-1250. doi:10.1016/j.solidstatesciences.2011.03.016
- Jehng, J. M., Tung, W. C., & Kuo, C. H. (2007). The formation mechanisms of multi-wall carbon nanotubes over the Ni modified MCM-41 catalysts. *Journal of Porous Materials*, *15*(1), 43-51. doi:10.1007/s10934-006-9050-x
- Jones, D. J. S. (2006). Petroleum products and a refinery configuration. In D. J. S. Jones & P. Pujadó (Eds.), *Handbook of Petroleum Processing* (pp. 47-109). Netherlands: Springer.
- Ka, I., Le Borgne, V., Ma, D., & El Khakani, M. A. (2012). Pulsed Laser Ablation based Direct Synthesis of Single-Wall Carbon Nanotube/PbS Quantum Dot Nanohybrids Exhibiting Strong, Spectrally Wide and Fast Photoresponse (Adv. Mater. 47/2012). *Adv Mater*, *24*(47), 6288-6288. doi:10.1002/adma.201290298
- Kishore, S. C., & Pandurangan, A. (2012). Electrophoretic deposition of cobalt catalyst layer over stainless steel for the high yield synthesis of carbon nanotubes. *Applied Surface Science*, *258*(20), 7936-7942. doi:10.1016/j.apsusc.2012.04.138
- Kotsilkova, R., Ivanov, E., Bychanok, D., Paddubskaya, A., Demidenko, M., Macutkevic, J., ... Kuzhir, P. (2015). Effects of sonochemical modification of carbon nanotubes on electrical and electromagnetic shielding properties of epoxy composites. *Composites Science and Technology*, *106*, 85-92. doi:http://dx.doi.org/10.1016/j.compscitech.2014.11.004
- Koziol, K., Boskovic, B. O., & Yahya, N. (2011). Synthesis of Carbon Nanostructures by CVD Method. In N. Yahya (Ed.), *Carbon and Oxide Nanostructures: Synthesis, Characterisation and Applications* (pp. 23-49). Berlin, Heidelberg: Springer Berlin Heidelberg.

- Kukovitsky, E. F., L'Vov, S. G., & Sainov, N. A. (2000). VLS-growth of carbon nanotubes from the vapor. *Chemical Physics Letters*, 317(1–2), 65-70. doi:http://dx.doi.org/10.1016/S0009-2614(99)01299-3
- Kumar, M. (2011). Carbon Nanotube Synthesis and Growth Mechanism. In S. Yellampalli (Ed.), *Carbon Nanotubes - Synthesis, Characterization, Applications* (pp.147-170). Croatia: InTech.
- Kunadian, I., Andrews, R., Mengüç, M. P., & Qian, D. (2009). Multiwalled carbon nanotube deposition profiles within a CVD reactor: An experimental study. *Chemical Engineering Science*, 64(7), 1503-1510. doi:10.1016/j.ces.2008.12.033
- Lee, C. J., Park, J., Huh, Y., & Lee, J. Y. (2001). Temperature effect on the growth of carbon nanotubes using thermal chemical vapor deposition. *Chemical Physics Letters*, 343(1), 33-38.
- Lee, C. J., Park, J., Kim, J. M., Huh, Y., Lee, J. Y., & No, K. S. (2000). Low-temperature growth of carbon nanotubes by thermal chemical vapor deposition using Pd, Cr, and Pt as co-catalyst. *Chemical physics letters*, 327(5), 277-283.
- Lee, D. C., Mikulec, F. V., & Korgel, B. A. (2004). Carbon nanotube synthesis in supercritical toluene. *Journal of the American Chemical Society*, 126(15), 4951-4957.
- Lee, S. Y., Yamada, M., & Miyake, M. (2005). Synthesis of carbon nanotubes over gold nanoparticle supported catalysts. *Carbon*, 43(13), 2654-2663. doi:10.1016/j.carbon.2005.05.045
- Lee, T. Y., Han, J. H., Choi, S. H., Yoo, J. B., Park, C. Y., Jung, T., ... Kim, J. M. (2003). Comparison of source gases and catalyst metals for growth of carbon nanotube. *Surface and Coatings Technology*, 169-170, 348-352. doi:10.1016/s0257-8972(03)00108-7
- Li, Q., Yan, H., Zhang, J., & Liu, Z. (2004). Effect of hydrocarbons precursors on the formation of carbon nanotubes in chemical vapor deposition. *Carbon*, 42(4), 829-835. doi:10.1016/j.carbon.2004.01.070
- Li, Y., Wang, H., Wang, G., & Gao, J. (2012). Synthesis of single-walled carbon nanotubes from heavy oil residue. *Chemical Engineering Journal*, 211-212, 255-259. doi:10.1016/j.cej.2012.09.031
- Li, Y., Zhang, L. H., Zhong, X. H., & Windle, A. H. (2007). Synthesis of high purity single-walled carbon nanotubes from ethanol by catalytic gas flow CVD reactions. *Nanotechnology*, 18(22), 225604–225610. doi:10.1088/0957-4484/18/22/225604
- Liew, S., Ramli, I., Yahya, N., & Shaari, A. H. (2014). Encapsulation of Iron(III) Oxide in Carbon Nanotube Bundles Using Pulsed Laser. *Advances in Natural and Applied Sciences*, 8(2), 69-74.
- Liu, W. W., Aziz, A., Chai, S. P., Mohamed, A. R., & Tye, C. T. (2011). The effect of carbon precursors (methane, benzene and camphor) on the quality of carbon nanotubes synthesised by the chemical vapour decomposition. *Physica E: Low-dimensional Systems and Nanostructures*, 43(8), 1535-1542. doi:http://dx.doi.org/10.1016/j.physe.2011.05.012
- Liu, X., Yang, Y., Liu, H., Ji, W., Zhang, C., & Xu, B. (2007). Carbon nanotubes from catalytic pyrolysis of deoiled asphalt. *Materials Letters*, 61(18), 3916-3919. doi:http://dx.doi.org/10.1016/j.matlet.2006.12.057
- Lobiak, E. V., Shlyakhova, E. V., Bulusheva, L. G., Plyusnin, P. E., Shubin, Y. V., & Okotrüb, A. V. (2015). Ni–Mo and Co–Mo alloy nanoparticles for catalytic chemical vapor deposition synthesis of carbon nanotubes. *Journal of Alloys and Compounds*, 621, 351-356. doi:10.1016/j.jallcom.2014.09.220
- Loebick, C. Z., Derrouiche, S., Fang, F., Li, N., Haller, G. L., & Pfefferle, L. D. (2009). Effect of chromium addition to the Co-MCM-41 catalyst in the synthesis of single wall carbon nanotubes. *Applied Catalysis A: General*, 368(1), 40-49. doi:10.1016/j.apcata.2009.08.004

- Lva, P., Zhanga, P., Lia, F., Lia, Y., Fenga, Y., & Fenga, W. (2012). Vertically aligned carbon nanotubes grown on carbon fabric with high rate capability for super-capacitors. *Synthetic Metals*, 162(13–14), 1090-1096. doi:http://dx.doi.org/10.1016/j.synthmet.2012.04.029
- MacKenzie, K. J., Dunens, O. M., & Harris, A. T. (2010). An Updated Review of Synthesis Parameters and Growth Mechanisms for Carbon Nanotubes in Fluidized Beds. *Industrial and Engineering Chemistry Research*, 49(11), 5323-5338. doi:10.1021/ie9019787
- Maghrebi, M., Khodadadi, A. A., Mortazavi, Y., & Mhaisalkar, S. (2009). Detailed profiling of CNTs arrays along the growth window in a floating catalyst reactor. *Applied Surface Science*, 255(16), 7243-7250. doi:10.1016/j.apsusc.2009.03.074
- Maghsoodi, S., Khodadadi, A., & Mortazavi, Y. (2010). A novel continuous process for synthesis of carbon nanotubes using iron floating catalyst and MgO particles for CVD of methane in a fluidized bed reactor. *Applied Surface Science*, 256(9), 2769-2774. doi:10.1016/j.apsusc.2009.11.026
- Maria, K. H., & Mieno, T. (2015). Synthesis of single-walled carbon nanotubes by low-frequency bipolar pulsed arc discharge method. *Vacuum*, 113, 11-18. doi:http://dx.doi.org/10.1016/j.vacuum.2014.11.025
- Maruyama, S., Miyauchi, Y., Edamura, T., Igarashi, Y., Chiashi, S., & Murakami, Y. (2003). Synthesis of single-walled carbon nanotubes with narrow diameter-distribution from fullerene. *Chemical Physics Letters*, 375(5-6), 553-559. doi:10.1016/s0009-2614(03)00907-2
- Matar, S., & Hatch, L. F. (2001). Chapter One - Primary Raw Materials for Petrochemicals. In S. Matar & L. F. Hatch (Eds.), *Chemistry of Petrochemical Processes* (2nd Ed.) (pp. 1-28). Woburn: Gulf Professional Publishing.
- Mizuno, K., Hata, K., Saito, T., Ohshima, S., Yumura, M., & Iijima, S. (2005). Selective Matching of Catalyst Element and Carbon Source in Single-Walled Carbon Nanotube Synthesis on Silicon Substrates. *The Journal of Physical Chemistry B*, 109(7), 2632-2637. doi:10.1021/jp0454117
- Monthieux, M. (2011). *Carbon Meta-Nanotubes: Synthesis, Properties and Applications*. United Kingdom, UK: John Wiley & Sons Ltd.
- Moothi, K., Iyuke, S. E., Meyyappan, M., & Falcon, R. (2012). Coal as a carbon source for carbon nanotube synthesis. *Carbon*, 50(8), 2679-2690. doi:http://dx.doi.org/10.1016/j.carbon.2012.02.048
- Narkiewicz, U., Podsiadły, M., Jędrzejewski, R., & Pelech, I. (2010). Catalytic decomposition of hydrocarbons on cobalt, nickel and iron catalysts to obtain carbon nanomaterials. *Applied Catalysis A: General*, 384(1–2), 27-35. doi:http://dx.doi.org/10.1016/j.apcata.2010.05.050
- Nerushev, O. A., Sveningsson, M., Falk, L. K. L., & Rohmund, F. (2001). Carbon nanotube films obtained by thermal chemical vapour deposition. *Journal of Materials Chemistry*, 11(4), 1122-1132. doi:10.1039/B009775F
- Nessim, G. D. (2010). Properties, synthesis, and growth mechanisms of carbon nanotubes with special focus on thermal chemical vapor deposition. *Nanoscale*, 2(8), 1306-1323. doi:10.1039/B9NR00427K
- Niu, Z., & Fang, Y. (2006). Effects of synthesis time for synthesizing single-walled carbon nanotubes over Mo–Fe–MgO catalyst and suggested growth mechanism. *Journal of Crystal Growth*, 297(1), 228-233. doi:10.1016/j.jcrysgro.2006.09.003
- Policicchio, A., Caruso, T., Chiarello, G., Colavita, E., Formoso, V., Agostino, R. G., Tsoufis, T., Gournis, D., & La Rosa, S. (2007). Electronic, chemical and structural characterization of CNTs grown by acetylene decomposition over MgO supported Fe–Co bimetallic catalysts. *Surface Science*, 601(13), 2823-2827. doi:10.1016/j.susc.2006.12.051

- Pradhan, D., & Sharon, M. (2002). Carbon nanotubes, nanofilaments and nanobeads by thermal chemical vapor deposition process. *Materials Science and Engineering: B*, *96*(1), 24-28.
- Prasek, J., Drbohlavova, J., Chomoucka, J., Hubalek, J., Jasek, O., Adam, V., & Kizek, R. (2011). Methods for carbon nanotubes synthesis-review. *Journal of Materials Chemistry*, *21*(40), 15872-15884. doi:10.1039/C1JM12254A
- Purohit, R., Purohit, K., Rana, S., Rana, R. S., & Patel, V. (2014). Carbon Nanotubes and Their Growth Methods. *Procedia Materials Science*, *6*, 716-728. doi:http://dx.doi.org/10.1016/j.mspro.2014.07.088
- Qian, W., Yu, H., Wei, F., Zhang, Q., & Wang, Z. (2002). Synthesis of carbon nanotubes from liquefied petroleum gas containing sulfur. *Carbon*, *40*(15), 2968-2970.
- Qiu, J., An, Y., Zhao, Z., Li, Y., & Zhou, Y. (2004). Catalytic synthesis of single-walled carbon nanotubes from coal gas by chemical vapor deposition method. *Fuel Processing Technology*, *85*(8-10), 913-920. doi:10.1016/j.fuproc.2003.11.033
- Qiu, J., Li, Q., Wang, Z., Sun, Y., & Zhang, H. (2006). CVD synthesis of coal-gas-derived carbon nanotubes and nanocapsules containing magnetic iron carbide and oxide. *Carbon*, *44*(12), 2565-2568. doi:10.1016/j.carbon.2006.05.030
- Ren, F., Kanaan, S. A., Majewska, M. M., Keskar, G. D., Azoz, S., Wang, H., ... Pfefferle, L. D. (2014). Increase in the yield of (and selective synthesis of large-diameter) single-walled carbon nanotubes through water-assisted ethanol pyrolysis. *Journal of Catalysis*, *309*, 419-427. doi:http://dx.doi.org/10.1016/j.jcat.2013.10.007
- Saengmee-anupharb, S., Thongpang, S., Bertheir, E. S. P., & Singjai, P. (2011). Growth of Vertically Aligned Carbon Nanotubes on Silicon Using a Sparked Iron-Cobalt Catalyst. *ISRN Nanotechnology*, *2011*, 1-8. doi:10.5402/2011/684748
- See, C. H., & Harris, A. T. (2007). A Review of Carbon Nanotube Synthesis via Fluidized-Bed Chemical Vapor Deposition. *Industrial and Engineering Chemistry Research*, *46*(4), 997-1012. doi:10.1021/ie060955b
- Sharon, M., Mukhopadhyay, K., Yase, K., Iijima, S., Ando, Y., & Zhao, X. (1998). European Materials Research Society 1997 Meeting, Symposium A: Fullerenes and Carbon based Materials Spongy carbon nanobeads—A new material. *Carbon*, *36*(5), 507-511. doi:http://dx.doi.org/10.1016/S0008-6223(98)00060-8
- Singh, J., Kothiyal, N. C., & Pathania, D. (2011). Synthesis of Highly Dispersed Single Walled Carbon Nanotubes from Furnace Oil and Light Diesel Oil by Modified Chemical Vapour. *International Journal of Theoretical and Applied Science*, *3*(2), 15-20.
- Su, W., Zhu, Y., Zhang, J., Liu, Y., Yang, Y., Mao, Q., & Li, L. (2016). Effect of multi-wall carbon nanotubes supported nano-nickel and TiF₃ addition on hydrogen storage properties of magnesium hydride. *Journal of Alloys and Compounds*, *669*, 8-18. doi:http://dx.doi.org/10.1016/j.jallcom.2016.01.253
- Suriani, A. B., Alfariisa, S., Mohamed, A., Isa, I. M., Kamari, A., Hashim, N., ... Rusop, M. (2015). Quasi-aligned carbon nanotubes synthesised from waste engine oil. *Materials Letters*, *139*, 220-223. doi:10.1016/j.matlet.2014.10.046
- Szabó, A., Perri, C., Csató, A., Giordano, G., Vuono, D., & Nagy, J. B. (2010). Synthesis Methods of Carbon Nanotubes and Related Materials. *Materials*, *3*(5), 3092-3140. doi:10.3390/ma3053092

- Takenaka, S., Iguchi, T., Tanabe, E., Matsune, H., & Kishida, M. (2009). Formation of carbon nanotubes through ethylene decomposition over supported Pt catalysts and silica-coated Pt catalysts. *Carbon*, 47(5), 1251-1257. doi:10.1016/j.carbon.2008.12.051
- Tapasztó, L., Kertész, K., Vértesy, Z., Horváth, Z. E., Koós, A. A., Osváth, Z., ... Biró, L. P. (2005). Diameter and morphology dependence on experimental conditions of carbon nanotube arrays grown by spray pyrolysis. *Carbon*, 43(5), 970-977. doi:http://dx.doi.org/10.1016/j.carbon.2004.11.048
- Tessonnier, J. P., & Su, D. S. (2011). Recent Progress on the Growth Mechanism of Carbon Nanotubes: A Review. *ChemSusChem*, 4(7), 824-847. doi:10.1002/cssc.201100175
- Tran, K. Y., Heinrichs, B., Colomer, J. F., Pirard, J. P., & Lambert, S. (2007). Carbon nanotubes synthesis by the ethylene chemical catalytic vapour deposition (CCVD) process on Fe, Co, and Fe-Co/Al₂O₃ sol-gel catalysts. *Applied Catalysis A: General*, 318, 63-69. doi:10.1016/j.apcata.2006.10.042
- Wang, X., Li, Z., Xu, W., Kulkarni, S. A., Batabyal, S. K., Zhang, S., ... Wong, L. H. (2015). TiO₂ nanotube arrays based flexible perovskite solar cells with transparent carbon nanotube electrode. *Nano Energy*, 11, 728-735. doi:http://dx.doi.org/10.1016/j.nanoen.2014.11.042
- Yah, C. S., Iyuke, S. E., Simate, G. S., Unuabonah, E. I., Bathgate, G., Matthews, G., & Cluett, J. D. (2011). Continuous synthesis of multiwalled carbon nanotubes from xylene using the swirled floating catalyst chemical vapor deposition technique. *Journal of Materials Research*, 26(05), 640-644. doi:10.1557/jmr.2010.69
- Yang, X., Shi, C., Liu, E., He, C., Du, X., Li, J., & Zhao, N. (2012). Low-temperature synthesis of multi-walled carbon nanotubes over Cu catalyst. *Materials Letters*, 72, 164-167. doi:10.1016/j.matlet.2011.12.112
- Yun, Y., Shanov, V., Tu, Y., Subramaniam, S., & Schulz, M. J. (2006). Growth Mechanism of Long Aligned Multiwall Carbon Nanotube Arrays by Water-Assisted Chemical Vapor Deposition. *The Journal of Physical Chemistry B*, 110(47), 23920-23925. doi:10.1021/jp057171g
- Zhan, Z. Y., Zhang, Y. N., Sun, G. Z., Zheng, L. X., & Liao, K. (2011). The effects of catalyst treatment on fast growth of millimeter-long multi-walled carbon nanotube arrays. *Applied Surface Science*, 257(17), 7704-7708. doi:10.1016/j.apsusc.2011.04.013
- Zhang, J., Tu, R., & Goto, T. (2011). Preparation of carbon nanotube by rotary CVD on Ni nano-particle precipitated cBN using nickelocene as a precursor. *Materials Letters*, 65(2), 367-370. doi:http://dx.doi.org/10.1016/j.matlet.2010.09.074
- Zhang, Q., Huang, J., Wei, F., Xu, G., Wang, Y., Qian, W., & Wang, D. (2007). Large scale production of carbon nanotube arrays on the sphere surface from liquefied petroleum gas at low cost. *Chinese Science Bulletin*, 52(21), 2896-2902. doi:10.1007/s11434-007-0458-8
- Zhang, Q., Liu, Y., Hunang, J., Qian, W., Wang, Y., & Wei, F. (2008). Synthesis of Single-walled Carbon Nanotubes from Liquefied Petroleum Gas. *Nano*, 03(02), 95-100. doi:10.1142/S179329200800085X
- Zhao, N., He, C., Jiang, Z., Li, J., & Li, Y. (2006). Fabrication and growth mechanism of carbon nanotubes by catalytic chemical vapor deposition. *Materials Letters*, 60(2), 159-163. http://dx.doi.org/10.1016/j.matlet.2005.08.009
- Zheng, L., Liao, X., & Zhu, Y. T. (2006). Parametric study of carbon nanotube growth via cobalt-catalyzed ethanol decomposition. *Materials Letters*, 60(16), 1968-1972. doi:10.1016/j.matlet.2005.12.060

- Zhong, G., Hofmann, S., Yan, F., Telg, H., Warner, J. H., Eder, D., Thomsen, C., Milne, W. I., & Robertson, J. (2009). Acetylene: A Key Growth Precursor for Single-Walled Carbon Nanotube Forests. *The Journal of Physical Chemistry C*, *113*(40), 17321-17325. doi:10.1021/jp905134b
- Zhou, Z., Ci, L., Chen, X., Tang, D., Yan, X., Liu, D., ... Xie, S. (2003). Controllable growth of double wall carbon nanotubes in a floating catalytic system. *Carbon*, *41*(2), 337-342.
- Zhu, H. L., Bai, Y. J., Cui, H. Z., & Liu, L. (2014). Facile synthesis of carbon nanotubes via low temperature pyrolysis of ferrocene. *Journal of Crystal Growth*, *404*, 44-47. doi:http://dx.doi.org/10.1016/j.jcrysgro.2014.06.048
- Zhu, J., Jia, J., Kwong, F. L., & Ng, D. H. L. (2012). Synthesis of bamboo-like carbon nanotubes on a copper foil by catalytic chemical vapor deposition from ethanol. *Carbon*, *50*(7), 2504-2512. doi:http://dx.doi.org/10.1016/j.carbon.2012.01.073



An Investigation on High Temperature Erosion Behaviour of Plasma Sprayed CoCrAlY/Al₂O₃/YSZ on Fe and Ni Based Alloys

Nithin H. S.*, Vijay Desai and M. R. Ramesh

Department of Mechanical Engineering, National Institute of Technology Karnataka, Mangalore 575 025, Karnataka, India

ABSTRACT

Parts of aircraft and gas turbines used for power production are subjected to severe erosion damage since aircrafts frequently operate in sandy environment. Low cost fuel such as poor quality coal is used in gas turbines which produce suspended hard particle in the exhaust. In the past, researchers have worked on minimising the erosion by using certain coatings. Development of new coatings is necessary in order to explore further in improving resistance against erosion process under high operating temperature of gas turbine, aero engines and other components. In the present work, the investigation of elevated temperature erosion behaviour of CoCrAlY/Al₂O₃/YSZ coatings synthesised by plasma spraying on two different base metals, namely, Hastelloy X (Superni 76) and AISI 321 (MDN 321) was carried out. The coated samples were subjected to erosion test at 600°C with the impact angles of 30° and 90° under steady state condition. Alumina powder was used as erodent material of uneven angular shape of 50 µm particle size. The morphology and phase formed on eroded surface are characterised using SEM and X-ray diffraction to determine the erosion mechanism. The rate of erosion is determined by weight loss method and the CoCrAlY/Al₂O₃/YSZ coating showed up to about 25% lower erosion rate than the substrate alloy. It was observed that the erosion resistance of CoCrAlY/Al₂O₃/YSZ coating on both MDN 321 and Superni 76 gave almost similar erosion resistance which shows that the erosion behaviour of coating is not influenced by substrate unless spray parameter and substrate roughness is changed.

Keywords: Ceramics, erosion mechanism, high temperature erosion, plasma spray, substrate

Article history:

Received: 02 February 2016

Accepted: 17 June 2016

E-mail addresses:

nithinshiv1989@gmail.com (Nithin H. S.),

desai@nitk.ac.in (Vijay Desai),

ramesdmt@gmail.com (M. R. Ramesh)

*Corresponding Author

INTRODUCTION

Solid Particle erosion is a major material degradation process in many industries like power production, chemical plant, mining machineries and also aircraft engines. Gas turbines are the power source for these industries and engines, which operate in harsh environment. This imposes continuous

development of new materials. Aircraft engine parts (compressor, turbine and propeller blades) are exposed to severe erosion media (Sundararajan & Roy, 1997). Thermal spray coatings can provide greater resistance to surface degradation process, which enhances the lifecycle of component by maintaining the mechanical strength of the base metal.

There are many thermal spray processes available; among that plasma spraying is a well-established and versatile technique for producing metals and ceramics coatings. It is recorded as a promising technology by providing reliable and cost-effective solution for many industrial problems (Sidhu & Prakash, 2006). In plasma spraying, the coating material in the form of powder is introduced into a high temperature plasma jet and the particle gets melted and accelerated towards the substrate. Upon impact, the molten particle is flattened and quenched on the substrate asperities. This results in a layered microstructure, which may contain typical defects such as pores, splat boundaries and micro cracks. Plasma spraying provides coatings with thickness from tens of micrometres to hundreds of micrometres deposited with high deposition rate, which also gives good quality and good adherent coating.

The plasma sprayed MCrAlY (where M is nickel, cobalt, or a combination of Ni and Co) coatings on Ni-based super alloys and other high strength alloys have been found to enhance the hot corrosion and oxidation resistance, which leads to the formation of oxides, and spinals of nickel, aluminium, chromium and cobalt (Seo et al., 2008). Yttrium and aluminium are the two elements in MCrAlY combination, which oxidise during thermal spraying because of their high affinity to the oxygen. Thus, there are more chances of fine dispersed Al and Y rich oxides in the as-sprayed coatings (Brandl et al., 1998). CoCrAlY bond coat performs excellent hot corrosion and oxidation resistance with high levels of Al and Cr, especially in marine operating condition, although there is a lag in remarkable erosion and wear resistance (Li et al., 1998). Various hard phases used for Ni based coatings as reinforcing additives are WC, Al₂O₃, Cr₃C₂, Cr₂O₃, TiC, SiC, TiN, CeO₂, TiO₂, etc. There are, however, only few reinforcing like MoS₂, Ce, Si, Y₂O₃ etc (Ogawa et al., 2006; Ramesh, et al., 2010; Kim & Korsunsky, 2010) on Co based coating. Ramesh et al. (2010) studied the solid particle erosion resistance of WC–Co/NiCrFeSiB coating deposited using HVOF spraying process and observed that the HVOF spraying led to a high retention of WC in the coating matrix. The average value of microhardness was reported as 1223 HV_{0.3} and the main erosion mechanism of the coating was attributed to brittle mode. It was reported that plasma sprayed NiCrSiB/Al₂O₃ coating on AISI 304 substrate material showed a better performance in 30° impact angles, but coating could protect the substrate at both 30° and 90° of impact angles. The coating shows improved hardness value from 160-673 HV at 200 g load. The presence of grooves, lips, splat removal and detachment of Al₂O₃ hard particles combined to form both ductile and brittle mode of erosion (Praveen, 2015).

In the present study, Al₂O₃ powder is mixed with a CoCrAlY powder and atmospheric plasma spraying was used to deposit the composite coating on Ni and Fe based alloys. Al₂O₃ was selected due to its high hardness, low cost, better thermal stability, better wettability and thermal expansion coefficient similar to Ni alloy (Praveen, 2015). The erosion behaviour of coatings was studied at 30° and 90° of impingement angles and the eroded surfaces were analysed with SEM images to identify the erosion mechanism.

MATERIALS AND METHODS

Substrate material and coating characterisation

Austenitic stainless steel (MDN 321) and Nickel based super alloy (Superni 76), which are used as material for aircraft engine parts, furnaces and many high temperature application, have been used as substrate in the present study. The specimens MDN 321 and Superni 76 are cut into the dimensions of 25x25x4 (mm) and 25x25x3 (mm) respectively from cold rolled plate for erosion studies. Samples were cleaned in acetone and grit blasted using alumina powder of 150µm before the atmospheric plasma spray process for better adhesion between coating and substrate. CoCrAlY/28%Al₂O₃/2%YSZ composite powder of particle sizes -15 to +45µm was deposited using the plasma spray process.

YSZ is a rare earth element which is used mainly for wear and erosion application due to its high hardness. Addition of rare-earth elements (CeO₂, La₂O₃, and Y₂O₃) provides increases in wear, corrosion, oxidation resistance and also improves the hardness, toughness, bond strength, and thermal shock resistance of various coating materials (Sharma, 2012). The plasma spray process was carried out using METCO USA 3MB equipment (Spraymet Surface Technologies Private Limited, Bangalore, India). The spray parameter adapted during deposition is listed in Table 1.

The coated samples were sectioned and polished to determine coating thickness, porosity and morphology using optical microscope supported with image analyser software and SEM. Hardness was carried out using Mitutoyo hardness tester. X-ray diffraction was done to analyse different phases in powder coating.

Table 1
Plasma spray parameters

Argon flow rate	40 l/m
Hydrogen flow rate	5-7 l/m
Current	490 A
Voltage	60-70 V
Powder feed rate	60 g/m
Stand of distance	100-120 mm

Erosion Studies

Erosion tests were carried out as per ASTM G76-13 standard using solid particle Air Jet Erosion Tester TR-471-800 (Ducom Instruments Private Limited, Bangalore, India), which can operate at the maximum temperature of 800°C. The erosion studies were performed both on the substrate and coating at 600°C for comparison. The testing conditions of erosion in the present study are listed in Table 2.

Before testing, the samples were cleaned with acetone, followed by drying and later weighed using electronic weighing balance with least count 0.1 mg. The samples were then fixed to the holder with two different angles of 30° and 90° and allowed it to erode for 5 cycles having each cycle of 10 minutes. Then, the eroded samples were cleaned with high pressure air to remove erodent particle and weighed to determine the weight loss.

Table 2
Erosion Test Conditions

Erodent material	Alumina
Erodent average size	50 μm
Particle velocity	35 m/s
Erodent feed rate	2 g/min
Impact angle	30° and 90°
Temperature	600°C
Cycle test time	10 min
Nozzle diameter	1.5 mm
Standoff distance	10 mm

RESULTS AND DISCUSSION

Microstructure and Properties of Coating

Cross section of the coated sample observed in SEM is shown in Figure 1 and the scale bar in the SEM image shows the appropriate coating thickness ranging from 150 μm to 200 μm . Elliptical splats formed due to the impacts of fully melted and partially melted particles could be seen. Marked small round partially melted particles could also be seen in back scattered image in Figure 1(a). The grey region was found to be Al_2O_3 and white region corresponded to Co rich phases, and pores could also be seen with spotted black colour. Pores present between the boundaries of splats can be observed in secondary electron image as shown in Figure 1(b). Surface morphology of coating, distribution of splats and the splat formed during the molten particle impact are shown in the magnified SEM image in Figure 1 (c, d) and d respectively.

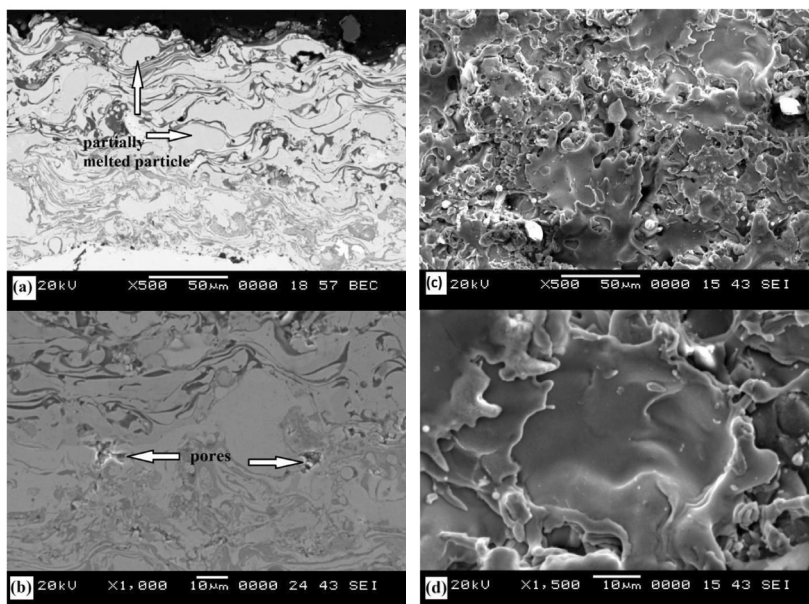


Figure 1. The SEM images of cross section (a, b) and surface (c, d) of as-coated sample.

The X-ray diffraction pattern of powder and plasma sprayed CoCrAlY/Al₂O₃/YSZ shown in Figure 2 shows some similar peaks of Cr, α -Co and Al₃Co. Cr and Al₂O₃ are the major phases in the powder with α -Co and Al₃Co as minor phases having low intensity peaks. Cr oxide has been observed in the coated sample, which was formed due to the occurrence of oxidation during in-flight time, and plasma spray process has higher in-flight time compared to other thermal spray process (Wang et al., 2015). The slight amorphisation may occur during the process leading to change in X-ray diffraction pattern peak height and width, which can be observed in the X-ray diffraction pattern.

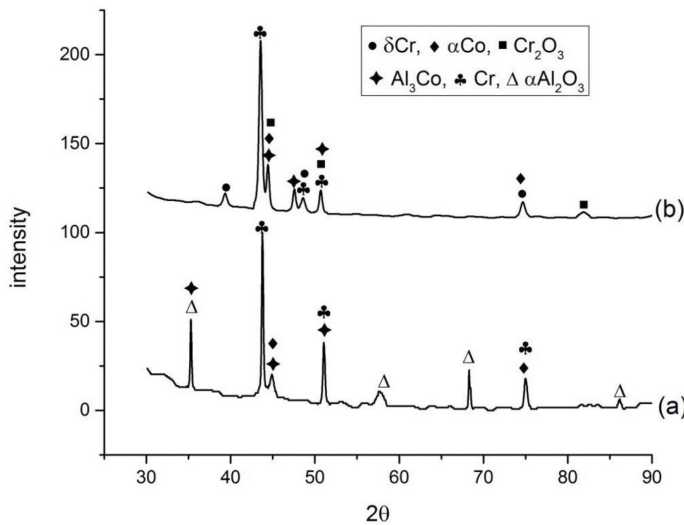


Figure 2. The X-ray diffraction pattern of (a) CoCrAlY/Al₂O₃/YSZ powder; (b) As coated

The variation of the microhardness values of the substrate and coating is shown in Figure 3. Both MDN 321 and Superni 76 substrates obtained the hardness value of 185 HV and 280 HV respectively. The CoCrAlY/Al₂O₃/YSZ coating shows the hardness value of 285 HV.

The increase in the hardness of substrate can be observed near the interface because of work hardening effect during the sand blasting prior to plasma spray process. The variation of the microhardness values of coating across the cross section is due to heterogeneous nature and distribution of Al₂O₃ in CoCrAlY matrix. Similarly, Praveen et al. (2015) reported that the heterogeneous nature of Plasma Sprayed Al₂O₃ with Ni based powder coating with the microhardness value of 673 HV with the load of 300 g.

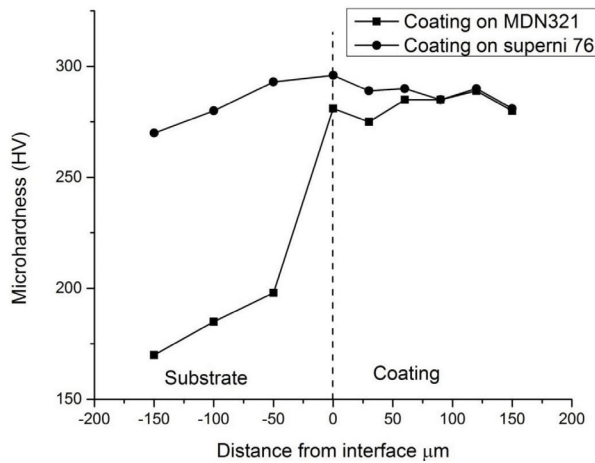


Figure 3. Microhardness variation along the cross section of coating and substrates.

The influence of impact angle on solid particle erosion

The volumetric erosion rate of coating on two substrates with different impact angles of 90° and 30° is represented in a bar chart in Figure 4. At both 90° and 30° of impact angles, the coated samples showed better erosion resistance performance on both Ni and Fe based alloy substrates. By quantifying the volumetric erosion rate, it is clear that coating shows better performance at 90° compared to 30° impact angles. In general, brittle coating material undergoes rapid erosion at higher angle than at smaller angle of impact and ductile coating material shows rapid erosion at lower angle than higher angle of impact (Tu et al., 1997). From the above reference, it can be reported that CoCrAlY/Al₂O₃/YSZ coating follows ductile erosion mechanism. Coating shows less erosion volume loss compared to plasma sprayed NiCrSiB/Al₂O₃ (Dallaire et al., 2001; Ramesh et al., 2010) which has same hardphase Al₂O₃. Figure 5 (a, b) shows the cumulative erosion loss in each cycle of 10 minutes; it was observed that steady state erosion took place in last 3 cycles as compared to the first 2 cycles. At 90° of impact angle, the coating on both substrates follows the same sequence, which has less erosion in the initial cycle and reaches steady state during the last 3 cycles. However, the reverse pattern of erosion, followed 30° of impact angle which has higher value of erosion at initial cycle and attained steady state during the last cycles.

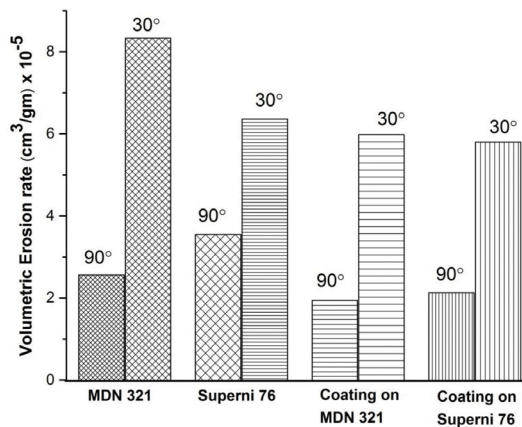
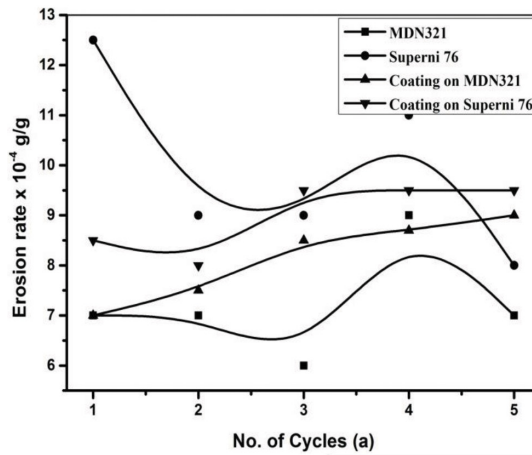
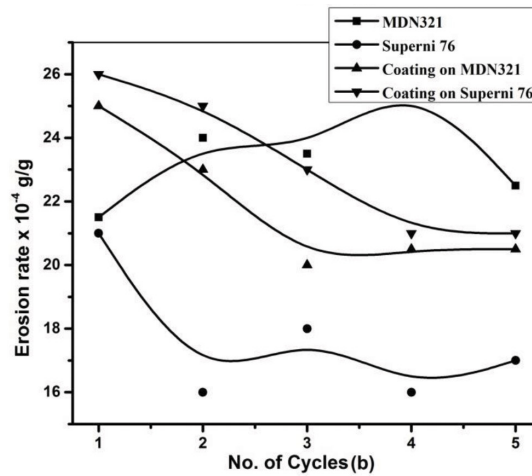


Figure 4. Bar chart illustrating volumetric erosion rate at 90° and 30° of impact angles.



(a)



(b)

Figure 5. The cumulative variation of the erosion rate with the cumulative weight of the erodent for uncoated and CoCrAlY/Al₂O₃/YSZ at 90° (a) and 30°; (b) of impact angles.

Erosion Mechanism

Plasma sprayed CoCrAlY/Al₂O₃/YSZ coating on both MDN 321 and Superni 76 substrates follows a similar erosion mechanism in both 90° and 30° of impact angle. The surface morphology of CoCrAlY/Al₂O₃/YSZ coating at 90° impact angle is shown in Figure 6, which reveals the presence of impact craters, indentation, raised lips and micro cuttings. The indentation on the surface followed by raised lips (Figure 6(a) and (b)), which in turn are plastically deformed from the surface, due to high strain produced by the repeated impacts of erodent particle. Eroder particles, which stick on the surface of coating, can be seen in Figure 6(a). In the feather region, the removal of splats, which forms slightly bigger craters, can be seen. The marked region in 6(b) and 6(d) shows the initial stage of splat removal which loosens the splat boundary by the repeated impact attributed to ductile erosion.

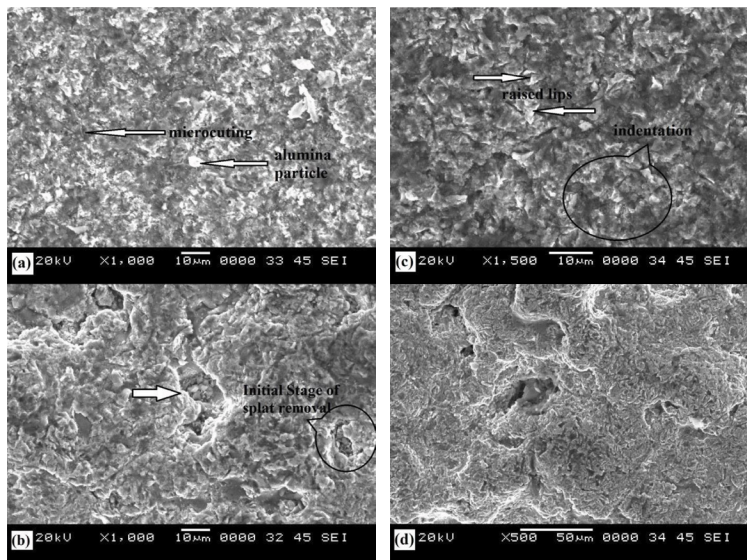


Figure 6. The surface morphology of coating on MDN 321 (a, b) and Superni 76 (c, d) showing erosion on both centre and feather region with impact angle of 90°.

The surface morphology of coating eroded at 30° of impact angle is shown in Figure 7. The presence of grooves and lips represents the material by cutting and ploughing mechanism (Figure 7(c) and 7(d)). Larger grooves are formed on the softer Co-rich region and these grooves are formed by microcutting mechanism. The higher tangential force acts on the hard phases during lower impact angle results in bigger craters, indicating the material removal by chipping mechanism. The number of lips (Figures 7(a) and 7(c)) was observed on the surface, which shows erosion is done by ploughing mechanism.

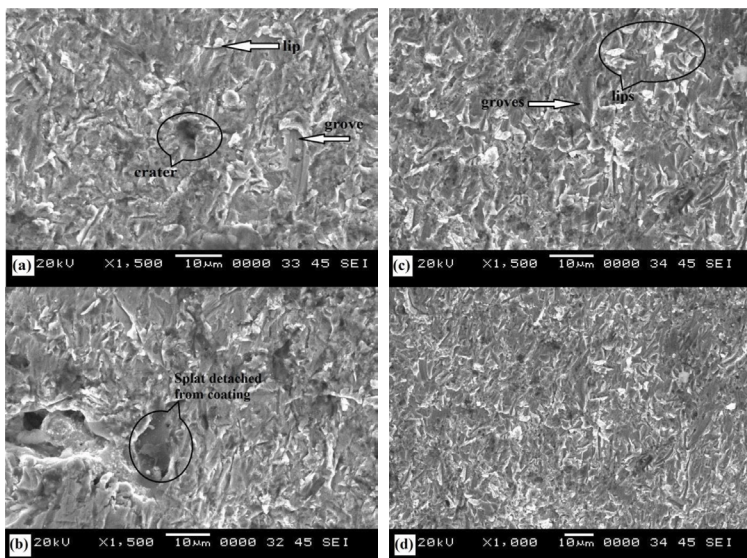


Figure 7. The surface morphology of coating on MDN 321 (a, b) and Superni 76 (c, d) showing erosion on both centre and feather region with impact angle of 30°.

CONCLUSION

The plasma spraying technique was used to successfully deposit CoCrAlY/Al₂O₃/YSZ coating on Fe and Ni based substrate materials. Coating with the porosity less than 1.5% and average hardness of 287 HV at 200 g load has been achieved. The results of solid particle erosion at 600°C indicated that the plasma sprayed CoCrAlY/Al₂O₃/YSZ coating has ability to protect both Fe and Ni based substrate at both 30° and 90° of impact angles. The erosion performance is almost similar for coating on both Ni and Fe based since coating properties are the same, unless coating parameter, substrate temperature and roughness are changed. Coating shows better erosion resistance in 90° compared 30° of impact angles. The presence of indentation with raised lips, groves, craters without cracks and erosion loss is more in lower impact angle than higher impact angle, which concludes that the material removal is by ductile erosion mechanism.

REFERENCES

- Brandl, W., Toma, D., & Grabke, H. J. (1998). The characteristics of alumina scales formed on HVOF-sprayed MCrAlY coatings. *Surface and Coatings Technology*, 108, 10-15.
- Dallaire, S., Levert, H., & Legoux, J. G. (2001). Erosion resistance of arc-sprayed coatings to iron ore at 25°C and 315°C. *Journal of Thermal Spray Technology*, 10(2), 337-350.
- Kim, K., & Korsunsky, A. M. (2010). Dissipated energy and fretting damage in CoCrAlY-MoS 2 coatings. *Tribology International*, 43(3), 676-684.
- Li, S., Langlade, C., Fayeulle, S., & Treheux, D. (1998). Influence of the microstructure of plasma deposited MCrAlY coatings on their tribological behaviour. *Surface and Coatings Technology*, 100, 7-11.
- Ogawa, K., Ito, K., Shoji, T., Seo, D. W., Tezuka, H., & Kato, H. (2006). Effects of Ce and Si additions to CoNiCrAlY bond coat materials on oxidation behavior and crack propagation of thermal barrier coatings. *Journal of Thermal Spray Technology*, 15(4), 640-651.
- Praveen, A. S., Sarangan, J., Suresh, S., & Subramanian, J. S. (2015). Erosion wear behavior of plasma sprayed NiCrSiB/Al₂O₃ composite coating. *International Journal of Refractory Metals and Hard Materials*, 52, 209-218.
- Ramesh, M. R., Prakash, S., Nath, S. K., Sapra, P. K., & Venkataraman, B. (2010). Solid particle erosion of HVOF sprayed WC-Co/NiCrFeSiB coatings. *Wear*, 269(3), 197-205.
- Seo, D., Ogawa, K., Suzuki, Y., Ichimura, K., Shoji, T., & Murata, S. (2008). Comparative study on oxidation behavior of selected MCrAlY coatings by elemental concentration profile analysis. *Applied Surface Science*, 255(5), 2581-2590.
- Sharma, S. (2012). Erosive wear study of rare earth-modified HVOF-sprayed coatings using design of experiment. *Journal of Thermal Spray Technology*, 21(1), 49-62.
- Sidhu, B. S., & Prakash, S. (2006). Performance of NiCrAlY, Ni-Cr, Stellite-6 and Ni 3 Al coatings in Na 2 SO 4-60% V 2 O 5 environment at 900°C under cyclic conditions. *Surface and Coatings Technology*, 201(3), 1643-1654.
- Sundararajan, G., & Roy, M. (1997). Solid particle erosion behaviour of metallic materials at room and elevated temperatures. *Tribology International*, 30(5), 339-359.

- Tu, J. P., Liu, M. S., & Mao, Z. Y. (1997). Erosion resistance of Ni-WC self-fluxing alloy coating at high temperature. *Wear*, 209(1), 43-48.
- Wang, Y., Sun, C., Sun, J., Zhao, W., Dong, L., Li, L., & Meng, F. (2015). Erosion behavior of arc sprayed FeTi/CrB MMC coating at elevated temperature. *Surface and Coatings Technology*, 262, 141-147.

The Use of Infiltration Wells to Reduce the Impacts of Land Use Changes on Flood Peaks: An Indonesian Catchment Case Study

Kusumastuti, D. I.^{1*}, Jokowinarno, D.¹, Khotimah, S. N.¹ and Dewi, C.²

¹Department of Civil Engineering, Faculty of Engineering, Lampung University,
Jl. Sumantri Brojonegoro No.1, Bandar Lampung, 35145 Indonesia

²Department of Geodesy Engineering, Faculty of Engineering, Lampung University,
Jl. Sumantri Brojonegoro No.1, Bandar Lampung, 35145 Indonesia

ABSTRACT

The objective of this study is to investigate the effects of infiltration wells on flood peak reduction and flood frequency. The analysis was carried out with the use of a stochastic rainfall model incorporating a within-storm rainfall distribution of low, medium and high variability. This study was motivated by flooding in a catchment in Indonesia, which is potentially affected by land use changes. The parameterisation of the climate and landscapes was derived from a specific catchment in Indonesia. An analysis of land use changes underlines the importance of green land in reducing flood peaks. With city development, however, land is converted for settlements, industries and trade areas, which will increase flood peaks significantly. An analysis on the use of infiltration wells shows that flood peak reduction of up to 50% compared to without wells. The results also demonstrate that within-storm rainfall distribution affects flood peaks, where the effects of infiltration wells in reducing flood peaks are more observable when incorporating low within-storm variability.

Keywords: Infiltration well, within-storm rainfall distribution, land use, flood peak

INTRODUCTION

During the wet season, flooding occurs frequently at several locations in Indonesia, including Bandar Lampung, the capital city and economic hub of Lampung Province and the entry point to Sumatera. The province is located in the southeast region of Sumatera Island. The city's area is approximately 169.21 km², with a population of 881,801 people (Bappeda, 2010). The population and economy of the city grew significantly over the past two decades. Recently, several flooding events in the city resulted in inundation for several

Article history:

Received: 25 April 2016

Accepted: 02 August 2016

E-mail addresses:

kusumast@gmail.com (Kusumastuti, D. I.),

d.jokowinarno@gmail.com (Jokowinarno, D.),

unun_salim2@yahoo.com (Khotimah, S. N.),

citradewirohana@yahoo.com (Dewi, C.)

*Corresponding Author

hours and caused massive traffic jams and casualties. In addition, a landslide occurred in a canal within Bumi Kedaton Park and caused severe damage to its roads and a bridge (Kusumastuti & Jokowinarno, 2009).

The most common causes of flooding are high rainfall intensity, land use changes, reduced infiltration areas, insufficient drainage systems and reduced river capacities. Combinations of these factors may lead to extreme flooding events. Although flooding is a natural process, the destructive effects of flooding and peak discharge increased due to anthropogenic impacts, such as land use changes (Ashagrie et al., 2006). Significant land use changes, due to rapid development in most portions of Indonesia's larger cities, are considered to be one of the major causes of flooding. Land use changes, including urbanisation, deforestation and industrialisation, are associated with urban development and contribute to flooding in terms of the peak discharge, volume and frequency of floods (Huong & Pathirana, 2013; Zheng et al., 2016).

Recently, the effects of land use changes on hydrologic responses resulted in interest among researchers due to intensified development in developing countries (Doe et al., 1996; Ashagrie et al., 2006). A number of researchers use GIS-based models to evaluate the impacts of land use changes on hydrologic response (Gumindoga, 2010; Hacket, 2009; Juahir et al., 2010; He & Hogue, 2011; Koch et al., 2012; Camorani et al., 2005, Nobert & Jeremiah, 2012). Aside from surface hydrology responses, the effects of land use changes on groundwater hydrology attracted interest (DeSmedt & Batelan, 2001). This study investigated the impacts of land use changes on surface runoff, which contributes to the flood peak.

The flood peak is influenced by many factors, including storm intensity, which is governed by the within-storm rainfall distribution. The within-storm pattern of rainfall intensity is different between spatial locations on a worldwide scale. Several previous studies investigated the effects of within-storm rainfall distributions on floods (Robinson & Sivapalan, 1997; Kusumastuti et al., 2007) and found that the impacts of within-storm variability are most significant for fast-response mechanisms such as surface runoff. Within-storm rainfall patterns represent the variability of small timescales. For a catchment that has experienced significant land use changes such as the Way Kuala Garuntang Catchment, surface runoff can be considered a dominant runoff mechanism. Therefore, the within-storm variability will significantly affect the flood peaks.

To reduce the impacts of flooding, this study investigated the use of an infiltration well, which is used for a water conservation technique that can be used to help reduce surface runoff and peak flow. Sunjoto (1993; 1994) conducted a study of infiltration wells in Indonesia based on the concept of natural water balance as part of the hydrological cycle. When ground surface is permeable, some precipitation will naturally infiltrate and the remaining precipitation will run off along the ground surfaces. However, recent and massive land use changes led to increased runoff because a large portion of the ground area was covered by impervious layers such as roofs and pavement, which prohibit water infiltration. Infiltration wells are ideal in this situation because the water collected from roofs and channels can be directed to wells for infiltration (Sunjoto, 1994) (Figure 1). Several studies on infiltration wells indicated that surface runoff could be significantly reduced when appropriate well dimensions and numbers are used. Furthermore, the use of infiltration wells supports the zero delta discharge policy that limits water discharge to drainage systems (Arafat, 2008; Indriatmoko, 2010; Iriani et al., 2013). Infiltration wells and infiltration techniques, such as basins, trenches, and deeper wells, should be used according to climate conditions to recharge water bodies and convey water from infrastructure and sealed surfaces (Bouwer, 2002).

This study is motivated by flooding problems in the Way Kuala Garuntang Catchment (Kusumastuti & Jokowinarno, 2009), which covers an area of 63.9 km² and lies across the centre of Bandar Lampung City (Figure 2). However, due to the lack of long-term data for this catchment, this study is mainly exploratory and uses a hypothetical catchment with climate and catchment parameters that are representative of the Way Kuala Garuntang Catchment. The main purpose of this study is to investigate the impacts of infiltration wells, combined with within-storm rainfall distribution, in reducing flood peaks. Therefore, the model used in this study is simple and sufficient for capturing the effects of catchment parameters, such as land use and rainfall intensity, governed by within-storm rainfall distribution. This analysis uses a synthetic realisation of rainfall time series, combined with simple rainfall-runoff models.

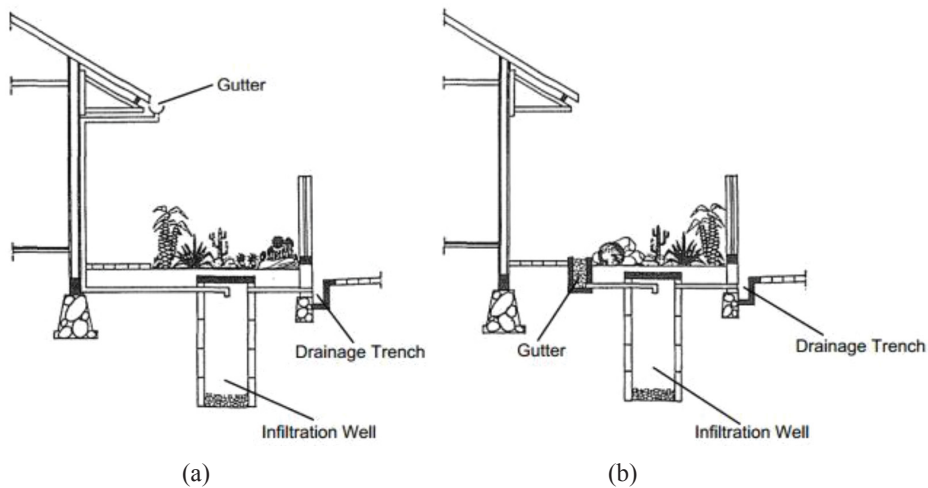


Figure 1. Illustration of infiltration well applications; (a) house with gutter at the eaves, and (b) house with gutter at the ground (Source: Sunjoto, 1994)

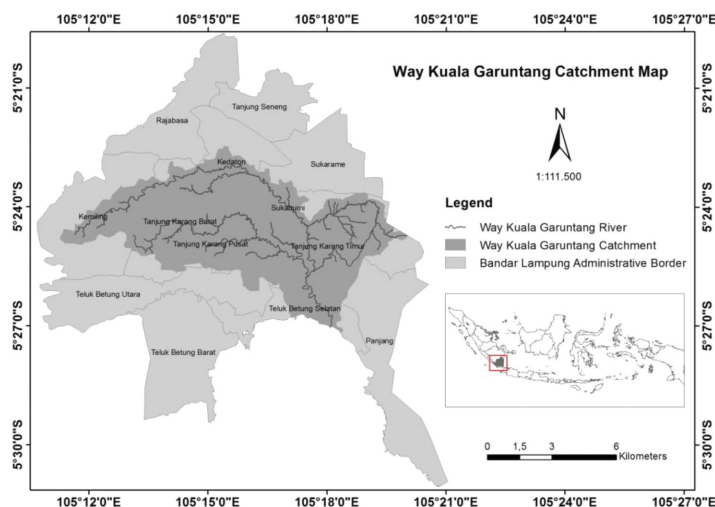


Figure 2. Way Kuala Garuntang Catchment within the city of Bandar Lampung

METHODOLOGY

Rainfall model that includes within-storm rainfall distribution

A long-term rainfall time series at hourly or sub-hourly time steps can be generated using the stochastic rainfall model developed by Sivandran (2002), which is an extension of the model developed by Robinson and Sivapalan (1997). The rainfall model accounts for the seasonal variability of dominant storm types and includes separate synoptic components that are assumed to operate year-round, as well as being a cyclonic component that only operates in the summer months. In this study, only synoptic components are considered; cyclonic components are set to zero. The synoptic component considers that each year consists of 12 months, with storm durations and inter-storm periods that are estimated from observed rainfall data obtained from the Meteorological and Climatology Berueu Panjang in Lampung from 2001 to 2015, which showed an annual average rainfall of between 1,500 and 2,500 mm (BMKG Panjang, 2016). Detailed descriptions of the rainfall model used in this study were included in previous works (Robinson & Sivapalan, 1997; Kusumastuti et al., 2007) and are not repeated in this study. Only the derivation of the parameters that are different from those described in the previous studies are explained below.

The characteristics of the storm data analysed in this study are presented in Figure 3. Storm durations are grouped into several ranges, including 3-6 h, 6-10 h, 10-17 h, 17-30 h, and 30-100 h. These durations represent very short, short, medium, long, and very long durations, respectively. The analysed storm data include the minimum total rainfall, maximum total rainfall, average total rainfall and average rainfall intensity. The minimum total rainfall generally increased as the storm duration increased (Figure 3(a)). In contrast with the minimum total rainfall trends, the maximum total rainfall generally decreased as the storm duration increased (Figure 3(b)). In addition, the average total rainfalls generally increased as the storm duration increased (Figure 3(c)) and the maximum average rainfall intensity generally decreased as the storm duration increased (Figure 3(d)).

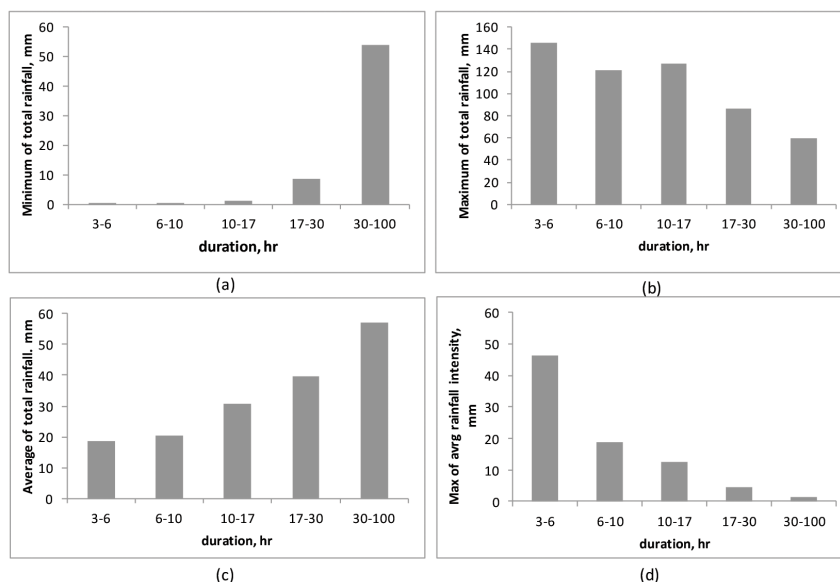


Figure 3. Characteristics of the rainfall data

The rainfall duration classes were statistically analysed to determine the number of events for each duration range. Meanwhile, the mean and coefficient of the variance of the rainfall intensity for the given storm durations are presented in Table 1. Next, the exponent and coefficient values of $E[i|t_r]$ are assigned as variables a1 and b1, and the exponent and coefficient values of $CV^2[i|t_r]$ are assigned as variables a2 and b2, respectively (Table 2).

Table 1
Statistics from the rainfall duration classes

Class	Duration Range	Events	$E[i tr]$, mm/h	$CV^2[i tr]$
1	3 <t<6	291	4.059	1.332
2	6<t<10	102	2.196	1.230
3	10<t<17	44	1.915	1.227
4	17<t<30	10	1.687	0.937
5	30<t<100	3	1.280	1.040

Table 2
Values of the exponents and coefficients for $E[i|t_r]$ and $CV^2[i|t_r]$

Parameter	$E[i t_r]$, mm/h	$CV^2[i t_r]$
Coefficient (a)	5.926	1.5392
Coefficient (b)	-0.3909	-0.1097

The within-storm distribution of rainfall intensity is stochastically governed by an indicator of variability, where the temporal patterns may include small, medium, or high variability. The within-storm pattern of rainfall intensity is represented by the variable, η . Greater η values correspond to the values that are more centred on the median. However, if small values are used, the resulting random variables are distributed at the extremes. The η values used in this study were 0.5, 1, and 3. This choice of η values is based on the analysis of within-storm patterns from several years of storm data in some regions (Robinson & Sivapalan, 1997; Hipsey et al., 2003; Kusumastuti et al., 2007). Figure 4 presents a typical rainfall hyetograph generated by the model for different values of η . The average intensity and total rainfall volume are the same in all the three cases. The simulated patterns demonstrate that lower η values produce highly variable and intermittent rainfall patterns whereas higher η values generate less variable rainfall with nearly uniform rainfall intensities. The parameters used in the rainfall model in this study are presented in Table 3.

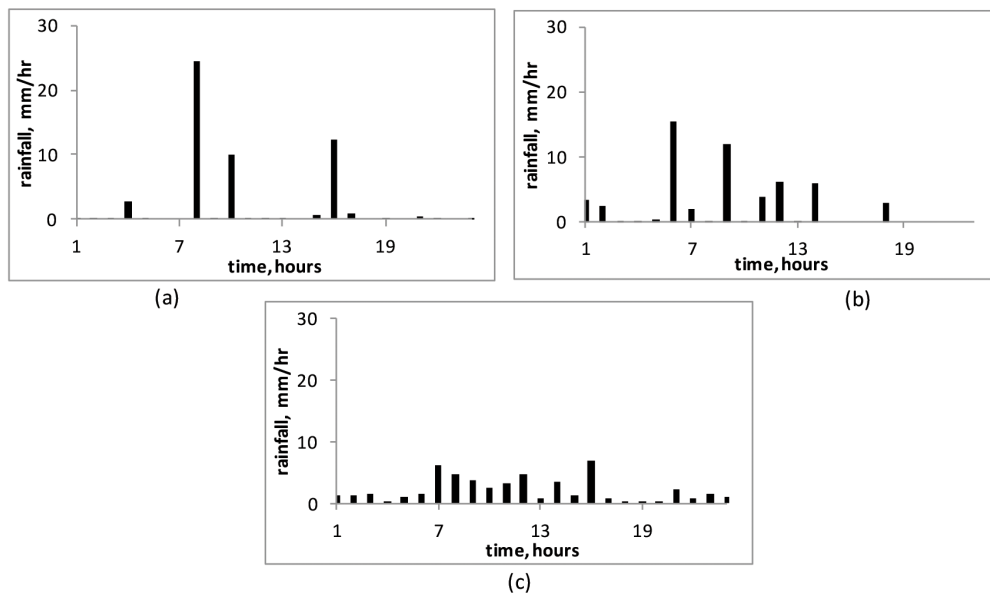


Figure 4. Variability of individual storms in the rainfall time series for (a) $\eta=0.5$; (b) $\eta=1.5$; and (c) $\eta=3.0$

Table 3
Rainfall model parameters

Parameter	Equation	Value	Units
δ_r	(3)	11	hours
α_r	(3)	4.4	hours
γ_b	(4)	100	hours
α_b	(4)	69	hours
$\tau_r = \tau_b$	(3),(4)	0	month
ω	(3),(4)	8760	hours
a_1	(6)	-	-
a_{1m}	(6)	5.9	-
a_{1a}	(6)	-5.9	-
b_1	(5)	-0.39	-
a_2	(5)	1.5	-
b_2	(5)	-0.1	-

Rainfall – runoff model

This study uses the rational method (ASCE, 1992), a simple rainfall-runoff model that is widely used to calculate the runoff-producing potential of a catchment. The rational method is based on a simple formula that relates the runoff producing potential of the watershed for rainfall intensity for a particular length of time (the time of concentration), in addition to the watershed drainage area (Viessman & Lewis, 2003; Young et al., 2009).

$$Q = C_u \cdot C \cdot I \cdot A \quad [1]$$

Here, Q is the design discharge (m^3/sec), C_u is the unit conversion coefficient, C is the runoff coefficient (dimensionless), I is the design rainfall intensity (mm/hour), and A is the catchment area (km^2).

The runoff coefficient (C) plays an important role in catchment management because it can be used as a surface runoff indicator in a catchment and for flood estimation. In order to estimate the runoff coefficient for the Way Kuala Garuntang Catchment, a runoff coefficient must be estimated for each land use in the Way Kuala Garuntang Catchment. Land uses in the Way Kuala Garuntang Catchment (Bappeda, 2010) and their corresponding runoff coefficients (Triatmodjo, 2008) are presented in Table 4.

Scenarios for the land use sensitivity analysis

The existing distribution of land uses in the Way Kuala Garuntang Catchment is presented in Figure 5. In order to understand the impacts of land use on the flood peak, this study considers four scenarios in the land use sensitivity analysis performed using Quantum GIS (Ramadana & Kusnanto, 2010). In Scenario 1, all vacant lands are converted into green lands. Based on the Regional Spatial Plan of Bandar Lampung 2010-2030, approximately 30% of the area is reserved for green land. In Scenario 2, all vacant lands are converted into green lands (as in Scenario 1), and half of the agriculture area is used for settlement or industrial purposes.

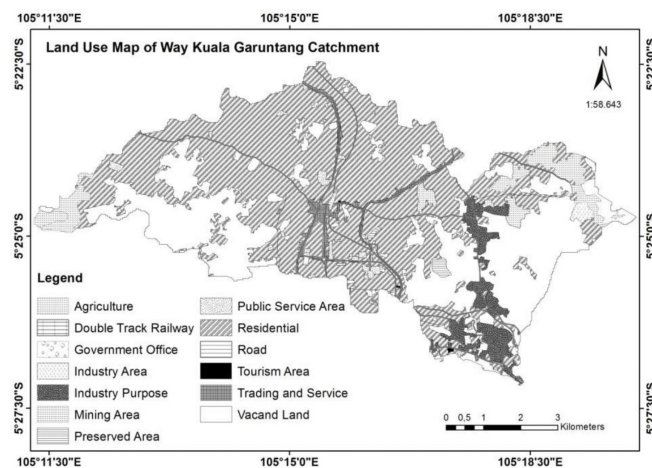


Figure 5. Land use map of the Way Kuala Garuntang Catchment showing its existing conditions (Source: Bappeda, 2010)

Scenarios 3 and 4 consider land use changes that are different from those of Scenarios 1 and 2. In Scenario 3, all vacant lands are converted into industry, and half of the settlement is converted into trading and service. There is a trend in Bandar Lampung of building *ruko*, a term in Indonesia that means homes and stores used for living, trading and services. A *ruko*

typically consists of three or more storeys, where the first storey is used as a store and the second and third storeys are used as home. Scenario 4 considers the conversion of all vacant lands into industry areas and a quarter of the settlement areas into trading and service areas. Scenarios 3 and 4 assume that more lands will not be required with the projected increase in population by 2030 because more people will live in apartments and *ruko*. As significant industrial growth requires more space, the policy to preserve 30% of the areas for green land cannot be maintained. The land use descriptions and corresponding percentage of the catchment area for each scenario above are presented in Table 4.

Table 4
Runoff coefficient (C) and percentage of land use in the catchment area

Land Use	Runoff Coefficient (C)	Percentage of Land Use in the Catchment Area				
		Existing	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Mining Area	0.95	0.40	0.40	0.40	0.40	0.40
Residential	0.65	54.14	54.14	56.50	27.07	40.60
Industrial Purposes	0.55	4.11	4.11	-	4.11	4.11
Vacant Land	0.40	31.23	-	-	-	-
Trading and Service	0.80	1.79	1.79	1.79	28.86	15.33
Government Office	0.75	0.23	0.23	0.23	0.23	0.23
Tourism Area	0.80	0.05	0.05	0.05	0.05	0.05
Industry Area	0.85	0.38	0.38	4.49	31.60	31.60
Preserved Area	0.30	0.46	0.46	0.46	0.46	0.46
Agriculture	0.35	4.72	4.72	2.35	4.72	4.72
Public Service Area	0.75	1.27	1.27	1.27	1.27	1.27
Road	0.90	1.16	1.16	1.16	1.16	1.16
Double Track Railway	0.35	0.07	0.07	0.07	0.07	0.07
Green Land	0.05	-	31.23	31.23	-	-

Source. analysis results

Requirements and number of infiltration wells

During application, the natural condition of the area should be checked to determine whether an area meets the following criteria: (1) land permeability should be more than 2 cm/hr, (2) shallow ground water is greater than 3 m below the ground level during the wet season, and (3) the land slope is less than 30° (Public Work Department, 1991). In this study, conditions of the Way Kuala Garuntang Catchment were established by using maps of land permeability, shallow groundwater levels and land slopes with existing land use maps to determine suitable areas for the application. The number of infiltration wells can be calculated as follows (Suripin, 2004):

$$\begin{aligned}
 \text{The number of wells} &= \frac{Q_{\text{runoff}}}{Q_{\text{well}}} \\
 &= \frac{Q_{\text{runoff}}}{((A_{\text{well base}} \times k) + (A_{\text{well wall}} \times k))} \\
 &= \frac{Q_{\text{runoff}}}{((\pi \cdot r^2 \times k) + (2 \cdot \pi \cdot r \cdot h \times k))}
 \end{aligned}
 \tag{2}$$

Where, Q_{runoff} is the daily peak flow (m^3/day), Q_{well} is the total runoff collected in a well during 1 day (m^3/day), Q_{well} is calculated based on infiltration and well discharge (m^3/day) and wall infiltration well discharge (m^3/day), k is the land permeability coefficient (m/day), r is the well radius (m), $A_{\text{well base}}$ is the well base area (m^2), $A_{\text{well wall}}$ is the well wall area (m^2) and h is the well depth (m). For analysis, the land permeability coefficient was set to 2 cm/hr, and the well depth was 3 m.

RESULTS AND DISCUSSION

Impacts of land use on flood peaks

Land use simulations in the Way Garuntang Catchment are presented in Figure 6. Based on the runoff coefficients for various land uses in the Way Garuntang Catchment and the proportion of catchment area associated with the particular land uses presented in Table 1, the composite runoff coefficients for the existing condition, Scenarios 1 (Figure 6(a)), 2 (Figure 6(b)), 3 (Figure 6(c)) and 4 (Figure 6(d)) were 0.56, 0.45, 0.47, 0.74 and 0.72, respectively.

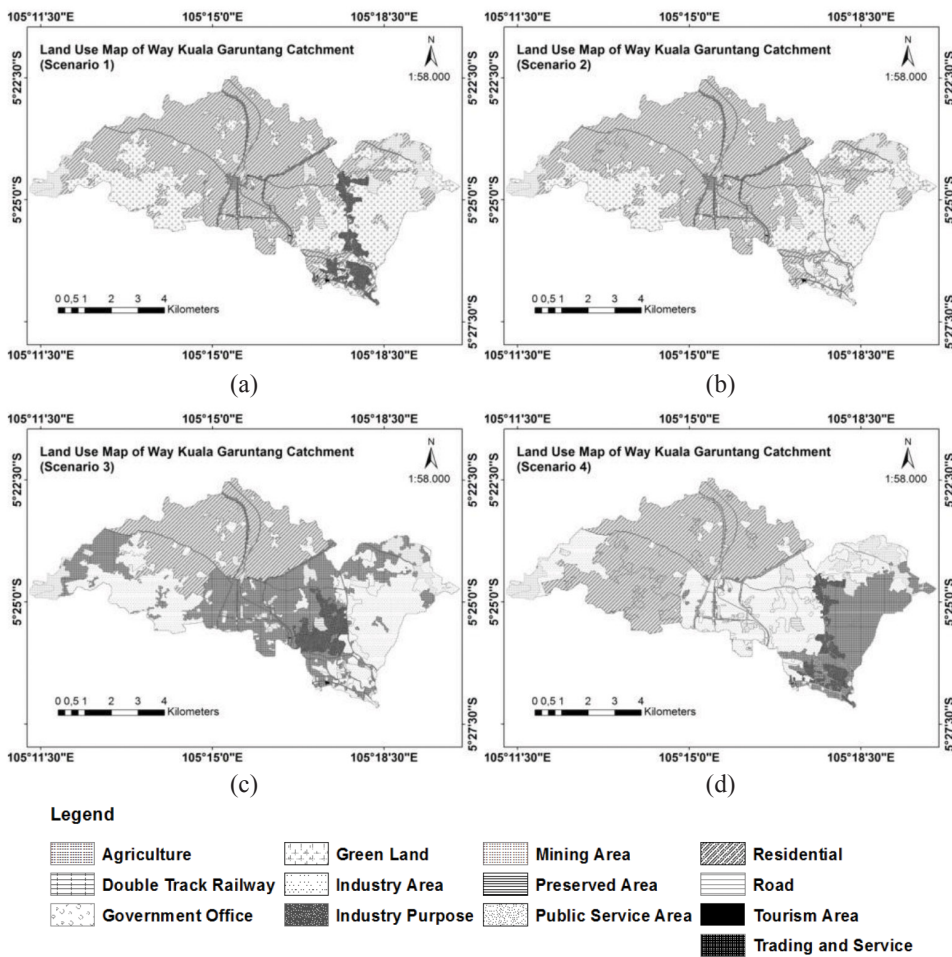


Figure 6. Land uses in the four land use change scenarios; (a) Scenario 1 ($C=0.45$); (b) Scenario 2 ($C=0.47$); (c) Scenario 3 ($C=0.74$), and (d) Scenario 4 ($C=0.72$)

The flood frequency curves, generated by all the scenarios incorporated with the selected within-storm patterns, are presented in Figure 7. The results can be categorised into two groups. The simulation results obtained using Scenarios 1 and 2 had lower flood peaks, whereas Scenarios 3 and 4 resulted in greater flood peaks (Figures 7(a), (b) and (c)). The decrease in flood peaks from Scenarios 1 and 2 resulted from considering approximately 30 percent of the catchment area as green land. A preliminary study by Yuniarti et al. (2013) on the impacts of land use changes in this catchment highlighted the importance of green land. Green land is necessary for providing sufficient space for infiltration, which is good for groundwater recharge and flood control. Scenario 2 includes green land and the conversion of some agriculture areas into settlement. Therefore, the flood peaks from Scenario 2 are greater than those from Scenario 1 because the land use conversion tends to change the recharge potential area into semi-impervious areas. In contrast with the simulation results from the first two scenarios, Scenarios 3 and 4 resulted in significant increases in flood peaks because the land use was converted from recharge potential areas into semi-impervious areas and semi-impervious areas were converted to impervious areas.

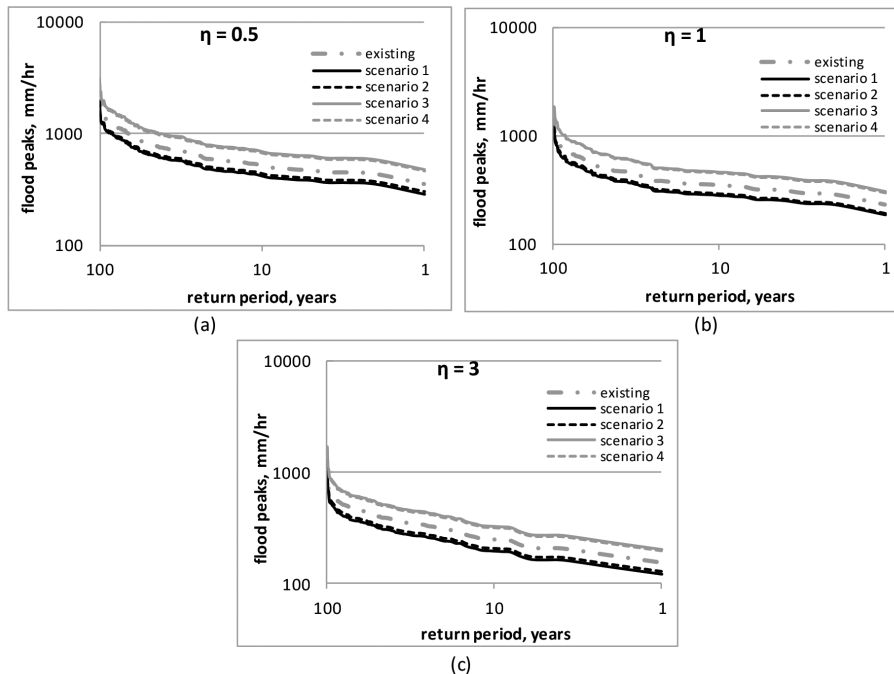


Figure 7. Flood frequency related to land use and the within-storm rainfall pattern

Within-storm patterns were found to significantly affect flood peaks for the same land use scenario because higher degrees of within-storm variability resulted in higher flood peaks (Figures 7(a), (b) and (c)). For example, the flood peaks resulted for Scenario 1 for return period of 10 years are 244.8 mm/hr, 346.5 mm/hr and 518.6 mm/hr for η equals 3, 1 and 0.5 respectively, or as the degree of within-storm variability increases. In addition, similar behaviours were observed for the return periods, and the flood peaks resulting from all scenarios also increased as the return period increased.

Impacts of infiltration wells on flood peaks

When aiming to understand the possible impacts of land use changes presented above, efforts should be made to reduce the impacts of land use changes on flooding. Infiltration is an alternative solution for conserving water and reducing floods. Before analysing the impacts of infiltration wells on flood peaks, the area that is suitable for applying infiltration must be defined. Figure 8 presents a ground water level map (Figure 8(a)), a soil permeability map (Figure 8(b)), a land slope map (Figure 8(c)) and a map of the suitable areas for infiltration wells (Figure 8(d)), which were obtained by overlying the three other maps. Overall, 52.06 km², or 81.47% of the catchment area, was identified to be suitable or effective for applying infiltration wells.

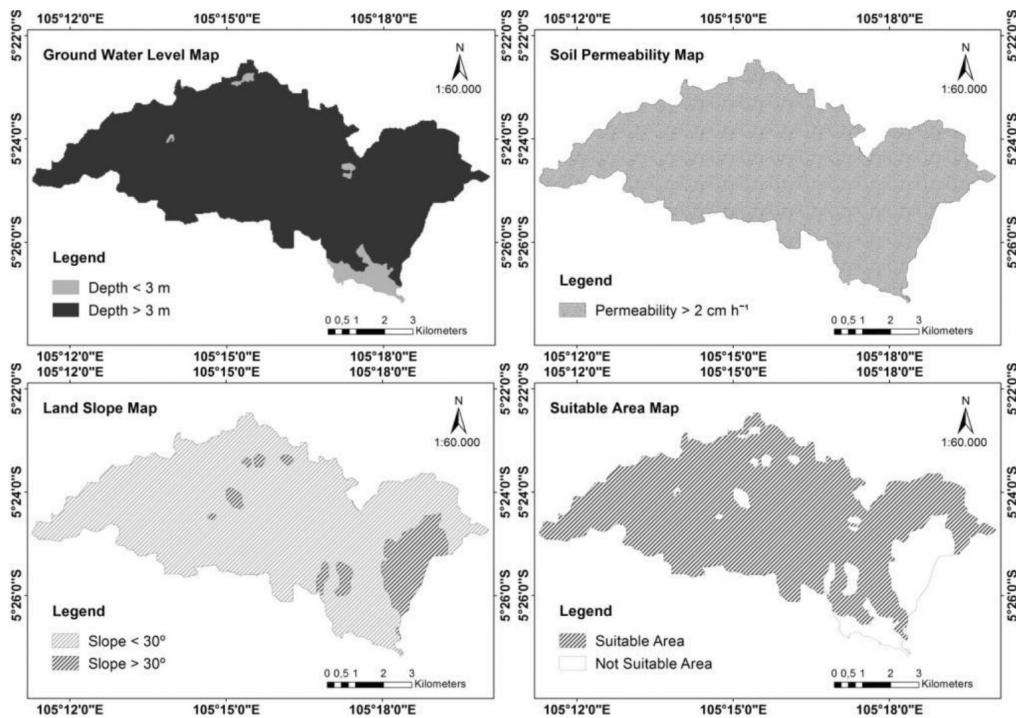


Figure 8. Suitable area for applying infiltration wells

The density of infiltration wells was applied over an average land area and the effects of one infiltration well on the flood peak while incorporating the within-storm pattern were investigated, and these results are presented in Figure 9. The density of infiltration was represented by No Well, Well 1, Well 2, Well 3, and Well 4 symbols, which correspond to no well applied and one well applied within 1,000, 2,000, 5,000, and 10,000 m², respectively. The results showed that denser applications of the infiltration wells resulted in lower flood peaks (Figures 9(a) to (f)). In addition to the application density of the infiltration wells, the magnitude of the flood peaks was determined based on the well diameter. Two well diameters, 1 and 1.4 m, were analysed. The results illustrate that larger-diameter wells result in lower flood peaks. This finding can be observed when comparing the flood peaks that were generated using the same within-storm pattern for the same infiltration well densities. For example, flood peaks that were simulated using a well diameter of 1.4 m (Figure 9(b)) were lower than those using a well diameter of 1.0 m (Figure 9(a)). Similarly, the flood peaks in Figure 9(d) are lower than those in Figure 9(c), and flood peaks in Figure 9(f) are lower than those in Figure 9(e).

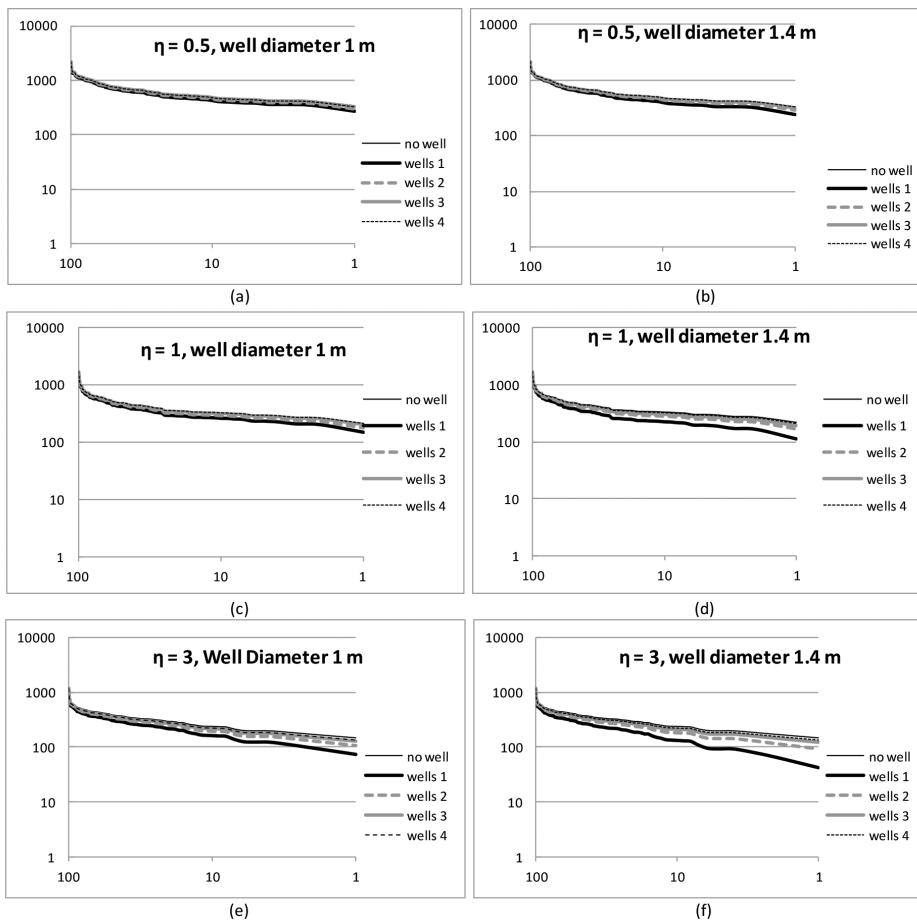
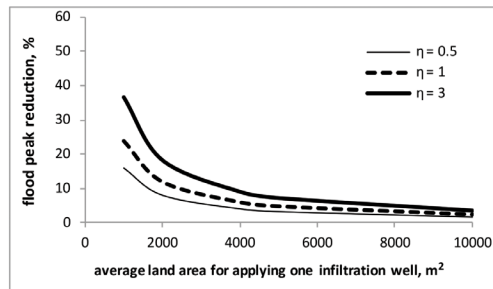


Figure 9. Flood frequency for applying infiltration wells

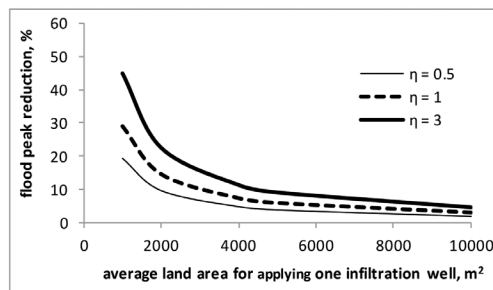
The impacts of within-storm patterns on flood peaks with and without infiltration wells can be observed in Figure 9. The high variability of the within-storm patterns demonstrates small differences in flood peaks between those that resulted from simulations, with and without infiltration wells. For the smaller well diameter, the difference of flood peaks between those generated from simulation, with and without infiltration wells, is insignificant when the high variability of the within-storm pattern is considered (Figure 9(a)). The difference in flood peaks becomes more significant as the within-storm variability decreases. This result can be observed when comparing the flood peak differences in Figures 9(a), 9(c) and 9(e). When using the larger well diameter, the difference in flood peaks becomes more significant as the variability of the within-storm pattern decreases, as shown in Figures 9(b), 9(d) and 9(f).

The results presented in Figure 9 suggest that the effects of within-storm patterns on reducing flood peaks are more significant at lower return periods. Therefore, when using flood peaks with a low return period such as at 5 years, further investigations can be used to examine the effects of within-storm patterns and infiltration wells on flood peak reduction. Figure 10 presents flood peak reductions and corresponding infiltration well densities for various within-

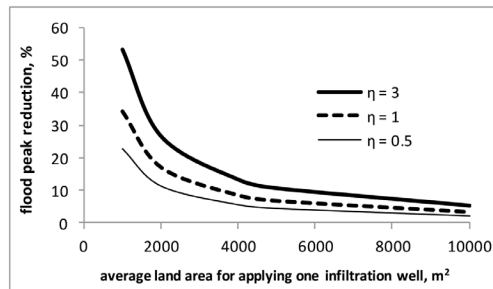
storm patterns. It is shown that the use of infiltration well results in flood peak reduction up to 50%. Well diameters of 1, 1.2, and 1.4 m are used in these simulations—and the results are presented in Figures 10(a), (b) and (c), respectively. Overall, the results confirmed that larger well diameters resulted in more significant reductions in a flood peak. Furthermore, these results strengthen the idea that lower within-storm in regards to storm variability results in more significant flood peak reduction impacts, as shown in Figures 10(a), (b) and (c).



(a)



(b)



(c)

Figure 10. Flood peak reduction vs. infiltration well density for well diameters of (a) 1 m; (b) 1.2 m; and (c) 1.4 m

The use of infiltration wells as one water conservation technique was selected because it was relatively cost-effective and can be applied by individuals or communities. In addition, this approach is similar to the introduction of the so-called green infrastructures in cities that conducted trials around the world. The concept of green infrastructure is related to storm water runoff management at the local level through the use of natural systems or engineered

systems that mimic natural systems. Green infrastructure corresponds to integrated catchment management and is used to address climate change, water quality, flooding and other urban issues. In addition, green infrastructure can be implemented with sequential infrastructures that contribute to storm water before it leaves the site. Through sequential infrastructures, storm water is collected by distributing green infrastructure technologies, such as rain barrels, infiltration wells, infiltration trenches and bio-retention. Subsequently, the excess water from those infrastructures is directed to a centralised outlet before the excess water is added to the urban drainage.

This study demonstrates that infiltration wells are capable of reducing the impacts of land use changes on flood peaks. However, there are some limitations in this study in that the use of the method and corresponding results only cope with a hypothetical catchment. Furthermore, the model parameterisation is limited to the catchments that have similar climatic and catchment characteristics. Climatic and hydraulic data from the studied catchment is also limited, therefore, a validation of the model cannot be carried out. In addition, there were some straightforward assumptions used in the analysis. For example, the well walls are of the same type, which are screened continuously along the bore. In practice, the type of well walls may be different from one location to another depending on geology, soil type and soil slope. The well walls may be screened continuously, screened at specific depth intervals or a cement wall. Similarly, the depth of the walls is considered to be the same in the analysis, while in implementation—ground water level and rainfall intensity are the factors in determining the depth of the wall. In order to simplify the analysis, this study considered soil permeability to be the same in vertical and horizontal directions, which may not be the case in the field. Therefore, to implement an infiltration well in such a location, each of the factors mentioned above need to be considered when designing an infiltration well. Despite its limitations, the results are valuable for understanding the impact of within-storm rainfall patterns—with the presence and absence of infiltration wells on flood peaks.

Several issues regarding the implementation of infiltration wells include the willingness of people to provide land and costs for constructing wells. With little participation from people for implementation, the number of infiltration wells will be low. Consequently, the application of infiltration wells may have less significant effects for flood peak reduction. Required commitment and relevant policies from the government could be used to encourage the use of these applications as well as community participation to address flood reduction. In addition, the implementation of infiltration wells with other green infrastructure techniques, such as rain barrels, porous trenches, retention ponds, and bio-retention, would have a greater impact on flood reduction.

CONCLUSIONS

Urbanisation, population growth and economic increases require more settlements, trading areas and industries, which may result in land use changes. Several scenarios were applied to demonstrate the effects of land use changes on flood peaks. The results indicate that the conversion of lands to obtain space for more settlements, industries and trading areas will increase flood peaks significantly. This study analysed the use of infiltration wells to reduce

flood peaks. The volume and densities of the wells determine the degree to which runoff will be reduced. Greater application densities of infiltration wells in a catchment result in greater reductions in runoff and lower flood peaks. Similarly, larger well diameters allow a greater proportion of runoff to be retained in infiltration wells. Flood peak reduction using infiltration wells is up to 50% compared to without wells. When incorporating the within-storm pattern in the presence and absence of infiltration wells, high variability in the within-storm pattern indicates only a small proportion of the excessive runoff generated by the corresponding individual storm that can be stored within the infiltration wells. This study demonstrates that the effects of infiltration wells in reducing flood peaks are more observable in regions with rainfall patterns exhibiting low within-storm variability.

ACKNOWLEDGEMENTS

This study was financially supported by National Research Funding under the Ministry of Research, Technology and Higher Education, Indonesia, awarded to the authors. This support is gratefully acknowledged.

REFERENCES

- Arafat, Y. (2008). Reduction of surface drainage load using infiltration well. *SMARTek*, 6(3), 144-153.
- ASCE. (1992). Design and construction of urban stormwater management systems. *ASCE Manuals and Reports on Engineering Practice No. 77*. New York: Amer Society of Civil Engineers.
- Ashagrie, A. G., De Laat, P. J. M., De Wit, M. J. M., Tu, M., & Uhlenbrook, S. (2006). Detecting the influence of land use changes on Floods in the Meuse River Basin? The predictive power of a ninety-year rainfall-runoff relation. *Hydrology and Earth System Sciences Discussions*, 3(2), 529-559.
- Bappeda. (2010). *Bandar Lampung spatial plan* [Data File]. Bandar Lampung, Indonesia.
- BMKG Panjang. (2016). *Rainfall data* [Data File]. Bandar Lampung, Indonesia.
- Bouwer, H. (2002). Artificial recharge of groundwater: hydrogeology and engineering. *Hydrogeology Journal*, 10(1), 121-142.
- Camorani, G., Castellarin, A., & Brath, A. (2005). Effects of land-use changes on the hydrologic response of reclamation systems. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(8), 561-574.
- DeSmedt, F., & Batelan, O. (2001). The impact of land use changes on the groundwater in the Grote Nete river basin, Belgium. In L. Ribeiro (Ed.), *Proceeding of the Conference Future of Groundwater Resources* (pp. 151-158). Lisbon, Portugal.
- Doe III, W. W., Saghafian, B., & Julien, P. Y. (1996). Land-use impact on watershed response: the integration of two-dimensional hydrological modelling and geographical information systems. *Hydrological Processes*, 10(11), 1503-1511.
- Gumindoga, W. (2010). *Hydrology impact of land use change in the Upper Gilgel Abay river basin, Ethiopia, TOP MODEL Application*. (Master Thesis). Enschede: International Institute for Geo-Information Science and Earth Observation.

- Hacket, W. R. (2009). *Changing land use, climate, and hydrology in the Winooski river basin, Vermont*. (Master Thesis). Vermont: The University of Vermont.
- He, M., & Hogue, T. S. (2011). Integrating hydrologic modeling and land use projections for evaluation of hydrologic response and regional water supply impacts in semi-arid environments. *Environmental Earth Sciences*, 65(6), 1671-1685. doi: 10.1007/s12665-011-1144-3.
- Hipsey, M., Sivapalan, M., & Menabde, M. (2003). The incorporation of risk in the design of engineered catchments for rural water supply in semi-arid Western Australia. *Hydrological Sciences Journal*, 48(5), 709-727.
- Huong, H. T. L., & Pathirana, A. (2013). Urbanization and climate change impacts on future urban flooding in Can Tho city, Vietnam. *Hydrology and Earth System Sciences*, 17(1), 379-394.
- Indriatmoko, R. H. (2010). Application of zero delta Q principle in regional development. *Jurnal Air Indonesia*, 6(1), 77-83.
- Iriani, K., Gunawan, A., & Besperi. (2013). Design of rainwater infiltration well to conserve groundwater in residential area (case study in Lingkar Timur Residence RT II, III and IV, Bengkulu), *Inersia*, 5(1), 9-22.
- Juahir, H., Zain, S. M., Aris, A. Z., Yusof, M. K., Samah, M. A. A., & Mokhtar, M. (2010). Hydrological trend analysis due to land use changes at Langat river basin. *Environment Asia*, 3(special issue), 20-31.
- Koch, F. J., Griensven, A. V., Uhlenbrook, S., Tekleab, S., & Teferi, E. (2012). The effects of land use change on hydrological response in the Choke Mountain Range (Ethiopia) – a new approach addressing land use dynamics in the model SWAT. In *Proceeding International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Sixth Biennial Meeting*. Leipzig, Germany.
- Kusumastuti, D. I., & Jokowinarno, D. (2009). *Development of the first measured unit hydrograph in Lampung; innovation in flood mitigation* (Research Report). Bandar Lampung: The University of Lampung.
- Kusumastuti, D. I., Struthers, I., Sivapalan, M., & Reynolds, D. A. (2007). Threshold effects in catchment storm response and the occurrence and magnitude of flood events: Implications for flood frequency. *Hydrology and Earth System Sciences*, 11, 1515 – 1528.
- Nobert, J., & Jeremiah, J. (2012). Hydrological Response of Watershed Systems to Land Use/Cover Change. A Case Study of Wami River Basin. *The Open Hydrology Journal*, 6(1), 78-87.
- PWD. (1991). *Technical Design Procedure for Rainwater Infiltration Wells in Yards (SNI No.03-2405-1991)*. Public Work Department. Bandung: LPMB Foundation.
- Ramadana, A. L., & Kusnanto, H. (2010). *Open Source GIS: Quantum GIS Application for Environment Information System*. Yogyakarta: PSLH-UGM Press.
- Robinson, J. S., & Sivapalan, M. (1997). Temporal scales and hydrological regimes: Implications for flood frequency scaling. *Water Resources Research*, 33(12), 2981-2999.
- Sivandran, G. (2002). *Effect of rising water tables and climate change on annual and monthly flood frequencies*. (B. Eng. Thesis). Crawley, Australia: Centre for Water Research, Univ. of West. Australia.
- Sunjoto, S. (1993). Sustainable urban drainage. In *International Conference on Management Geo-Water and Engineering Aspect*. Wollongong, Australia

- Sunjoto, S. (1994). Infiltration well and urban drainage concept, future groundwater resources at risk. In *Proceedings of the Helsinki Conference* (pp. 527-532). Yogyakarta, Indonesia.
- Suripin, S. (2004). *Urban Drainage*. Yogyakarta: Andi Offset.
- Triatmodjo, B. (2008). *Applied Hydrology*. Yogyakarta: Beta Offset.
- Viessman, Jr., W. & Lewis, G. L. (2003). *Introduction to Hydrology* (5th Ed.). New York: Harper Collins College Publishers.
- Young, C. B., McEnroe, B. M., & Rome, A. C. (2009). Empirical determination of rational method runoff coefficients. *Journal of Hydrologic Engineering*, 14(12), 1283-1289.
- Yuniarti, F., Kusumastuti, D. I., & Jokowiarno, D. (2013). Geospatial analysis of land use and land cover changes for discharge at Way Kuala Garuntang watershed in Bandar Lampung. In *Proceedings 2nd International Conference on Engineering and Technology Development*. Bandar Lampung, Indonesia.
- Zheng, J., Sun, G., Li, W., Yu, X., Zhang, C., Gong, Y., & Tu, L. (2016). Impacts of land use change and climate variations on annual inflow into the Miyun Reservoir, Beijing, China. *Hydrology and Earth System Sciences*, 20, 1561 – 1572.

A Hybridised Intelligent Technique for the Diagnosis of Medical Diseases

Abdu Masanawa Sagir* and Saratha Sathasivam

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

ABSTRACT

Medical diagnosis is the process of determining which disease or medical condition explains a person's determinable signs and symptoms. Diagnosis of most diseases is very expensive as many tests are required for predictions. This paper aims to introduce an improved hybrid approach for training the adaptive network based fuzzy inference system (ANFIS). It incorporates hybrid learning algorithms least square estimates with Levenberg-Marquardt algorithm using analytic derivation for computation of Jacobian matrix, as well as code optimisation technique, which indexes membership functions. The goal is to investigate how certain diseases are affected by patient's characteristics and measurement such as abnormalities or a decision about the presence or absence of a disease. In order to achieve an accurate diagnosis at this complex stage of symptom analysis, the physician may need efficient diagnosis system to classify and predict patient condition by using an adaptive neuro fuzzy inference system (ANFIS) pre-processed by grid partitioning. The proposed hybridised intelligent technique was tested with Statlog heart disease and Hepatitis disease datasets obtained from the University of California at Irvine's (UCI) machine learning repository. The robustness of the performance measuring total accuracy, sensitivity and specificity was examined. In comparison, the proposed method was found to achieve superior performance when compared to some other related existing methods.

Keywords: Adaptive neuro fuzzy inference system, classification, Levenberg-Marquardt algorithm, diagnosis of medical diseases

Article history:

Received: 09 March 2016

Accepted: 02 August 2016

E-mail addresses:

ams13_mah013@student.usm.my (Abdu Masanawa Sagir),

saratha@usm.my (Saratha Sathasivam)

*Corresponding Author

INTRODUCTION

Heart attack disease remains the main cause of death worldwide. The World Health Organisation estimated that 17.5 million people died from cardiovascular diseases in 2012, representing 31% of all deaths globally. An estimate of about 16 million deaths under the age of 70 was due to non-communicable

diseases, 82% of which occurred in low- and middle-income countries. Meanwhile, about 7.4 million deaths were due to coronary heart disease, whereas 6.7 million were due to stroke (World Health Organisation, 2015).

Hepatitis is an injury to the liver with inflammation of the liver cell. There are five main different types of the hepatitis virus, which are referred to as type A, B, C, D, E and possibly G. Hepatitis A and E are symptomatically acute types, whereas Hepatitis B, C and D lead to chronic diseases (Davis, 2016). The chronic hepatitis types lead to cirrhosis, which causes destruction of liver parenchymal cells.

These five types are of the greatest concern because of the burden of illness and death they cause and the potential for outbreaks and epidemic spread. Infection from these viruses results in approximately 1.45 million deaths each year (World Health Organisation [WHO], 2015).

Eighty one per cent (81%) of world's infants are vaccinated and protected from hepatitis B infection and about 2 million people with hepatitis B and C infections occur yearly through unsafe injections. These viruses are transmitted through contaminated water and food, as well as by contact with blood or bodily fluids through unsafe injections or transfusions. Infection also occurs from a mother to a child, or through sexual contact (WHO, 2005).

In order to investigate the misfortune of medical diseases, certain features must be observed. Some factors make the physician's work even more difficult to be analysed when evaluating existing test results of patients because some complicated measures are not easy to perform when considering a large number of factors. As described by Anooj (2012), Hedeshi and Abadeh (2014), the decision about the presence or absence of a patient with certain diseases depends on the physician's intuition, experience and skill for comparing current indicators with previous ones than on knowledge-rich data hidden in a database. This measure is a challenging task with regards to the large number of factors that must be considered. To achieve an accurate diagnosis at this complex stage of symptom analysis, the physician may need accurate and efficient hybridised intelligent systems that can classify and predict the likelihood of a patient getting a medical disease problem and to help in diagnosing disease.

Fuzzy logic was conceived by Zadeh (1965); it is a form of many valued logics in which a true value of variables may be any real number between 0 and 1. In fuzzy logic, everything is allowed to be a matter of degree, imprecise, linguistic and perception based. Fuzzy logic provides a foundation for the development of new tools for dealing with natural languages and knowledge representation. Its aim is formalisation of the modes of reasoning that are approximate rather than exact. Fuzzy logic has four principal facets: logical, set theoretic, relational and epistemic (Zadeh, 2004).

The literature in the medical field includes a number of research on the use of intelligent methods, among which, the Takagi-Sugeno-Kang fuzzy inference system was applied.

Nahato, Harichandran, and Arputharaj (2015) designed an intelligent system using rough set theory with a backpropagation neural network to classify the clinical dataset. The 10-fold cross validation technique is used in their research for validation of the training phase of the classifier, including 10 attributes that were selected but then reduced to six attributes in order to produce better results. An accuracy of 90.4% was achieved in their proposed work.

Subbulakshmi and Deepa (2015) developed a hybrid methodology based on a machine learning paradigm that integrates the self-regulated learning capability of particle swarm optimisation (SRLPSO) with an extreme learning machine (ELM) classifier. Five benchmarked datasets of the UCI machine learning repository were used for handling medical data classification, namely, Heart-Statlog, Cleveland Heart Disease, Hepatitis, Wisconsin Breast Cancer and Pima Indians Diabetes. For Heart-Statlog, number of instances, number of features and number of classes used are 270, 13 and 2, respectively. The parameters used for the proposed algorithms are 500 for the maximum iterations for both SRLPSO and ELM. The classification accuracy achieved was 85.88%.

Jaganathan and Kuppuchamy (2013) described feature relevant measures based on fuzzy entropy values and devised three feature selection strategies known as Mean Selection, Half Selection and Neural Network Threshold Selection with an RBF Network classifier. The fuzzy entropy with Neural Network Threshold Selection has achieved the maximum accuracy. Five benchmark datasets are used for evaluation. These data sets are: Breast Cancer, Pima Indians Diabetes, Heart-Statlog, Hepatitis and Cleveland heart diseases datasets from the UCI Machine Learning Repository. Feature selection that is currently being proposed results in accuracies that are as good as or better than when using the entire feature set without any feature selection.

Anooj (2012) developed a heart disease risk prediction technique using a weighted fuzzy rule. The researcher used three heart disease datasets from the UCI repository, namely, Cleveland data, Hungarian data and Switzerland data. The researcher used a pre-processing stage for removing missing values and other noisy information from the selected dataset. The researcher grouped the instances based on the class label and then changed numerical data type to categorical type data based on the equi-width technique. After discretisation, the researcher selected attributes for the fuzzy rule base based on the occurrence of the attribute value in each class.

Neshat, Sargolzaei, Nadjaran Toosi, and Masoumi (2012) developed a hybrid method for diagnosis of hepatitis disease that predicts or determines whether patients with hepatitis will either live or die. The case base weighted cluster algorithm for clustering and Particle Swarm Optimisation (PSO) for classification are combined to diagnose the risky hepatitis disease. The Case – Based Reasoning (CBR) was used to diagnose the disease, while PSO was adopted to assemble a decision-making system based on the selected features and disease recognised. The method CBR-PSO has been tested to ascertain the best accuracy results. The obtained results showed that the lowest, average and best accuracy results are 77.16%, 92.83% and 94.58%, respectively.

Uzer, Yilmaz, and Inan (2013) described the feature selection method based on a hybrid approach, which are the Artificial Bee Colony (ABC) algorithm and Support Vector Machine (SVM) for three different datasets, namely, Hepatitis, Liver disorders and Diabetes obtained from UCI Machine learning repository for classification purposes. A 10-fold cross validation method was used for a reliable performance of the classifier. The classification accuracy of the proposed systems obtained was 94.92%, 74.81% and 79.29% for Hepatitis dataset, Liver disorders dataset and Diabetes dataset, respectively.

This research work introduces a modified hybrid approach for training the adaptive network based fuzzy inference system (ANFIS). Sagir and Saratha (2015) emphasised the gradient base method and Levenberg-Marquardt algorithm using finite difference. In this study, the incorporation of the hybrid learning algorithm least square estimates with the Levenberg-Marquardt algorithm using analytic derivation and chain rule for the computation of a Jacobian matrix is used.

The remaining part of this paper is organised as follows: In section 2, the design of the newly hybridised intelligent system is presented. This leads us to section 3, in which simulation results are described. Meanwhile, discussion and conclusion of the work are given in section 4.

METHODS AND MATERIALS

According to Jang, Sun, and Mizutani (1997), if $f(x,y)$ is a first order polynomial, then the Takagi Sugeno Kang Fuzzy model is given as:

$$IF \ x = A_i \text{ and } y \text{ is } B_i \text{ THEN } z_i = f(x, y) \tag{1}$$

where A and B are fuzzy sets in the rule antecedent part, while $z = px+qy+r = f(x,y)$ is a crisp function in the rule consequent part, and p, q & r are the optimal consequent parameters. Usually $f(x,y)$ is a polynomial in the input variables x and y. When $f(x,y)$ is a first-order polynomial, the resulting fuzzy inference system is called a first – order model, which was employed in this research work.

Degree the input matches' number of rule is typically computed using min operator,

$$w_i = \min(\mu_{A_i}(x), \mu_{B_i}(y)) \tag{2}$$

where μ_{A_i} & μ_{B_i} are membership functions that define the fuzzy sets A and B, respectively, on the universe x and $i = 1, \dots, n$.

The final output of the system is the weighted average of all rule outputs, computed as,

$$Final\ output = \frac{\sum_{i=1}^N w_i z_i}{\sum_{i=1}^N w_i} \tag{3}$$

These rules are combined to get a function, which is defined as:

$$R(x) = \frac{A_1(x)f_1(x) + A_2(x)f_2(x) + \dots + A_n(x)f_n(x)}{A_1(x) + A_2(x) + \dots + A_n(x)} \tag{4}$$

This Takagi Sugeno Kang (TSK) fuzzy model produces a real-valued function.

ANFIS was first introduced by Jang (1993). The ANFIS is a framework of adaptive techniques to assist learning and adaptation. To illustrate the ANFIS structure, two fuzzy IF-THEN rules, according to a first order Sugeno model, are to be considered for simplicity (Sagir & Saratha, 2015).

Image of the ANFIS Structure

Figure 1 shows the structure of the Adaptive Neuro Fuzzy Inference System (ANFIS). The structure of the proposed model contains five layers: the input and output layers and three hidden layers that represent membership functions and fuzzy rules.

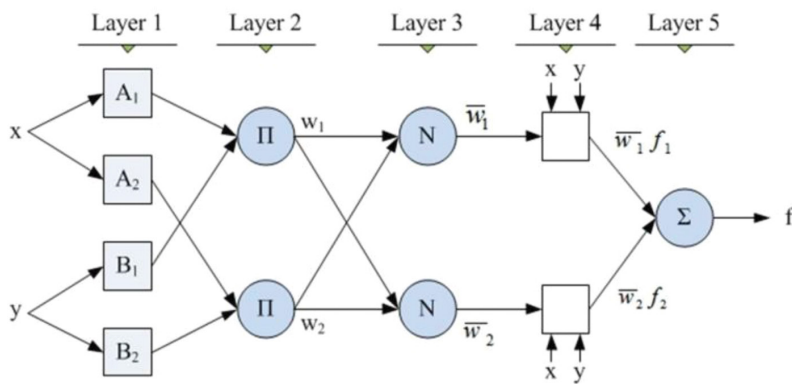


Figure 1. ANFIS-LMAD Structure

Rules Index Vector

Index Membership function is the index vector that keeps track of the unique Membership Functions (MF_s). This function determines the unique MFs in the ANFIS structure and indexes them row-wise. Therefore, the final index vector collects the indices found in MF “row-wise” according to rules. For example,

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix} \tag{5}$$

$$MF = [4 \ 4 \ 2 \ 3 \ 3 \ 3 \ 3] \tag{6}$$

Where, R is the rule list matrix (Jamsandekar & Mudholkar, 2014).

Finally, the index vector is:

$$ix = [1 \ 5 \ 9 \ 11 \ 14 \ 17 \ 20 \ 26 \ 10 \ 12 \ 15 \ 18 \ 21] \tag{7}$$

The Proposed ANFIS-LMAD Model Design

In designing this new ANFIS model, a hybrid learning technique, based on Least squares estimate and the Levenberg-Marquardt algorithm using the analytic difference method for computing the Jacobian Matrix, was used.

Forward Pass

Least squares estimate (LSE) was used at the very beginning to get the initial values of the conclusion parameters and then at backward pass for the Levenberg-Marquardt algorithm to update all parameters. Detail of this is found in Sagir and Saratha (2015).

Backward Pass

For the Levenberg-Marquardt algorithm, the performance index to be optimised is defined as,

$$F(w) = \frac{1}{2} E^T E \tag{8}$$

Where, E is the total error function

Error signals are propagated and the premise parameters are to be updated by the Levenberg-Marquardt algorithm:

$$W_k(t + 1) = W_k - (J_k^T J_k + \mu I)^{-1} J_k^T E(w) \tag{9}$$

$$\Delta W_k = (J_k + \mu I)^{-1} J_k^T E(w) \tag{10}$$

Get the parameters of unique MF_s of current FIS
Obtain Cumulative Current Error vector and MSE

$$\Delta W_k = (J_k + \mu I)^{-1} J_k^T E(w) \tag{11}$$

where $e_{kp} = d_{kp} - o_{kp}$, $k = 1, 2, \dots, K$, $p = 1, \dots, P$

e_{kp} is the training error at output p when applying k pattern, d_{kp} is the desire output vector & o_{kp} is the actual output vector.

$$mse = \sqrt{\frac{E_p}{p}} \tag{12}$$

If the current total error is increased as a result of the update, then retract the step and increase the learning parameter.

Built up Jacobian matrix column – wise, which contains 1st order partial derivatives of network error by establishing the derivatives with respect to MF_s parameters ρ using analytical derivation and chain rule.

$$\frac{\partial y_i}{\partial \sigma_i} \text{ and } \frac{\partial y_i}{\partial \beta_i} \quad (13)$$

where y_i is the number of output of the network, σ_i & β_i are parameters of MF_s, $i = 1, \dots, n$

Therefore,

$$J_{i,j} = \frac{\partial f_i}{\partial \rho_j} \quad (14)$$

Transform Jacobian into sparse matrix to speed things up

$$J_{i,j} = \text{Sparse} \left(\frac{\partial f_i}{\partial \rho_j} \right) \quad (15)$$

Approximate Hessian matrix which contains 2nd order partial derivative of network error using the cross product of Jacobian

$$H \approx J^T J \quad (16)$$

Therefore,

$$H_{i,j} = \frac{\partial^2 f_i}{\partial \rho_i \partial \rho_j} \quad (17)$$

Compute the error gradient

$$g = J^T E \quad (18)$$

Update the Hessian matrix

$$H^* = [H + \Phi I] \quad (19)$$

where I is the sparse identity matrix and $\Phi = 0.1$ is the learning parameter, and the network parameter needs to be updated using (10).

Then, recalculate the mse using the updated parameters

$$mse = \sqrt{\frac{E_p}{p}} \tag{20}$$

Adjust the learning parameters, if the total error is decreased as a result of the update then go to the next epoch. Then accept, otherwise reject.

Description of Input attributes for Datasets

The Statlog-Heart and Hepatitis datasets were obtained from the University of California at Irvine (UCI) machine learning repository (Bache & Lichman, 2013). Detailed information about the input variables is shown in Table 1 and Table 2.

Table 1
Information about input variables for the Statlog-Heart Disease dataset

Variable Name	Min	Max	No. of MF	Description of Input Variable	Type
CP	1	4	4	Chest pain type (1-typical angina, 2- atypical angina, 3- non-anginal pain, 4-asymtomatic)	Nominal
CHOL	126	564	4	Cholesterol (low, medium, high & very high)	Real
FBS	0	1	2	Resting blood sugar (0=false, 1=true) it is true when fbs>120	Binary
RESTECG	0	2	3	Resting electrocardiographic (0-normal, 1-having ST-T & 2-showing definite left VH)	Nominal
OPK	0	6	3	Oldpeak (low, risk, terrible)	Real
CA	0	3	3	No. of major Vessels (one, two, three)	Real
THAL	3	7	3	Thal (3 =normal, 6 = fixed defect, 7 = reverse defect)	Nominal

Table 2
Information about input variables for the Hepatitis Disease dataset

Variable Name	Min	Max	No. of MF	Description of Input Variable	Type
MAL	1	2	2	Malaise (no, yes)	Integer
SPI	1	2	2	Spiders (no, yes)	Integer
ASC	1	2	2	Ascites (no, yes)	Integer
VAR	1	2	2	Varices (no, yes)	Integer
BRB	0.3	8	3	Bilirubin (low, medium, high)	Continuous
ALB	2.1	6.4	3	Albumin (low, medium, high)	Real
HIS	2	2	2	Histology (no, yes)	Integer

RESULTS AND DISCUSSION

Membership function curves

In the Rule Based applications of Fuzzy Logic, membership functions are associated with terms that appear in the antecedents or consequents of rules. The Gaussian membership function provides smooth and non-linear functions that can be used by the learning systems like Neural Networks (Hagras, 2004).

The membership function curves for the input variables of optimised model (ANFIS-LMAD) of two datasets are shown in Figure 2 and Figure 3 respectively. The curves define how each point in the input space is mapped to the membership value (or degree of membership) between 0 and 1.

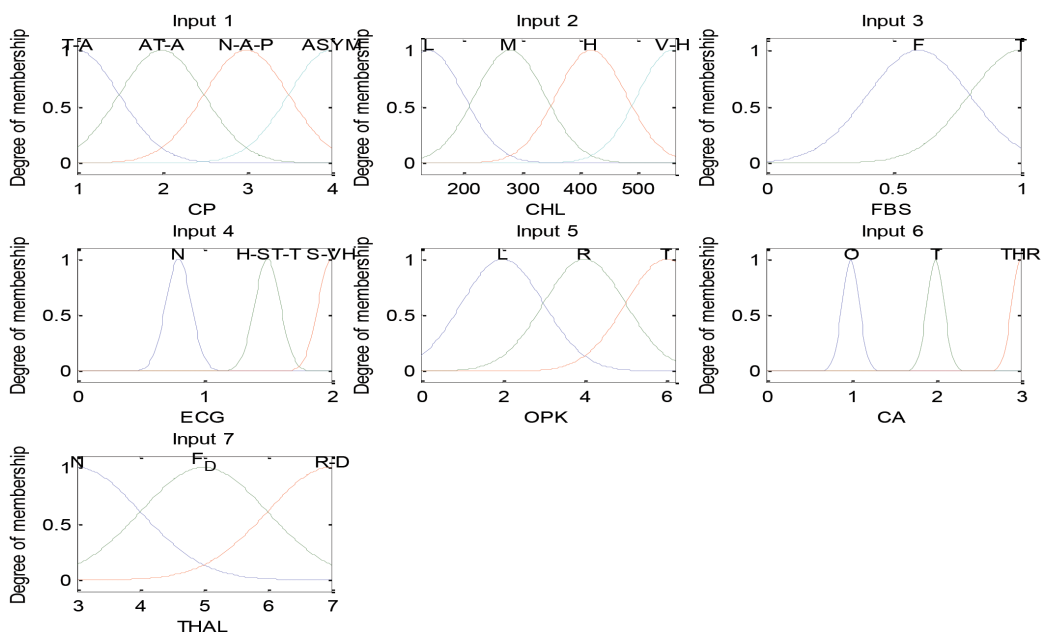


Figure 2. Membership Function Curve of Optimised Model for the Statlog-Heart Disease Dataset

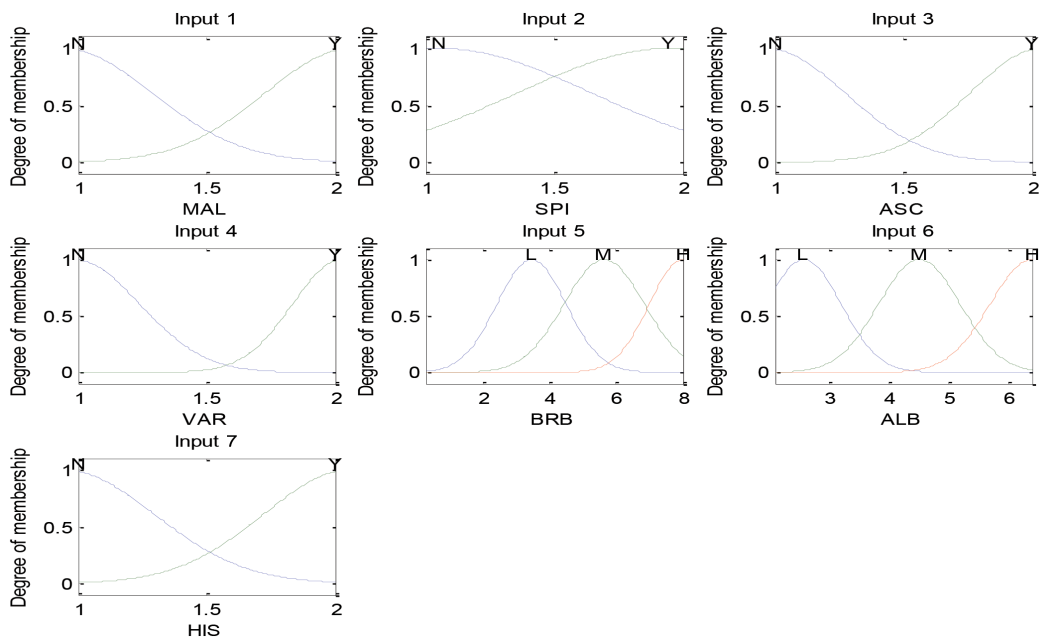


Figure 3. Membership Function Curve of Optimised Model for the Hepatitis Disease Dataset

Accuracy, Sensitivity and Specificity

The measure of the ability of the classifier to produce an accurate diagnosis is determined by accuracy. The measure of the ability of the model to identify the occurrence of a target class accurately is determined by sensitivity. The measure of the ability of the model to separate the target class is determined by specificity (Kahramanli & Allahverdi, 2008).

Table 3
Performance measure of the proposed model

	Statlog-Heart Disease Data set		Hepatitis Disease Data set	
	Train (%)	Test (%)	Train (%)	Test (%)
Sensitivity	81.71	84.21	96.97	97.67
Specificity	87.76	90.38	70.83	100.00
F-measure	83.23	85.33	93.43	98.82
Precision	84.81	86.49	90.14	100.00
Accuracy	85.00	87.78	90.00	97.86

Performance Error

Mean Square Error is one of the most acceptable indicators that describe the differences between the actual data and the predicted values. The values of premise and consequent parameters can be obtained after network training by directly minimising the MSE performance criterion (Ho, Tsai, Lin, & Chou, 2009).

$$MSE = \frac{1}{N} \sum_{i=1}^N (y - y')^2 \tag{20}$$

where y and y' are the desired output and predicted output respectively; and N is the number of total points.

Hence, in line with equation (21), the result for the values of performance error (mean square error), number of iterations and elapsed time of the proposed method for Statlog-Heart disease and Hepatitis disease data sets were found to be 0.10167, 1800 & 18.56762 seconds and 0.07263, 1300 & 11.30284 seconds, respectively.

Comparison of the Results

According to Kathy (2013), the train or test error rates can be obtained based on the calculation of the ratio of number of incorrect predicted values to the total number of train or test of dataset.

Table 4
Comparison of the results based on our recent work for the Statlog-Heart Disease Data set (Sagir & Saratha, 2015)

Author	Methodology Adopted	Accuracy (%)	Performance Error	Elapsed Time (second)
Proposed method	ANFIS-LMAD	87.78	0.10167	18.56762
Sagir & Saratha, (2015)	ANFIS_LSLM	76.67	0.40327	260.13590
Sagir & Saratha, (2015)	ANFIS_LSGD	75.56	0.40344	15.67020

Table 5
Comparison of the results with previous works based on performance measures and selected features for the Statlog-Heart Disease Dataset.

Methodology adopted	Accuracy (%)	Sensitivity (%)	Specificity (%)	Selected features
Proposed method	87.78	84.21	90.38	3,5,6,7,10,12, 13
PSO + ELM, Subbulakshmi & Deepa (2015)	85.88	86.00	86.03	3,11,12,13
ESUNN, Subbulakshmi & Deepa (2015)	83.22	84.32	81.65	3,8,9,11,12
EPUNN, Subbulakshmi & Deepa (2015)	81.89	83.67	84.91	8,9,11,12
NN for threshold selection, Jaganathan & Kuppuchamy (2013)	85.19	85.00	86.00	3,11,12,13
MS method, Jaganathan & Kuppuchamy (2013)	84.44	85.00	84.00	3,8,9,11,12,13
WFR+FL, Anooj (2012)	57.85	45.22	68.75	1,4,5,8,10,13

Table 6

Comparison of results with previous works based on performance measures and selected features for the Hepatitis Disease Dataset

Methodology adopted	Accuracy (%)	Sensitivity (%)	Specificity (%)	Selected features
Proposed method	97.86	100.00	98.82	7,12-15,18,20
PSO + ELM, Subbulakshmi & Deepa (2015)	97.43	93.65	95.71	6,11-14,17,19
RS-BPNN, Nahato et al. (2015)	97.30	98.32	97.28	9-17
ABC-SVM, Uzer et al. (2015)	94.92	97.13	88.33	1,5,12-15,17,18
CBR-PSO, Neshat et al. (2012)	94.58	—	—	Full features
PSO, Neshat et al. (2012)	89.46	—	—	Full features

Discussion of the Results

This research introduces an improved hybrid approach for training the adaptive network based fuzzy inference system (ANFIS). The approach incorporates hybrid learning algorithms least square estimates with the Levenberg-Marquardt algorithm, in addition to using analytic derivation for computation of Jacobian matrix instead of using the central difference method as applied in our previous research work (Sagir & Saratha, 2015).

In our previous work (Sagir & Saratha, 2015), ANFIS_LSLM classifier yields better results when compared with the ANFIS_LSGD classifier but with slightly lower convergence speed, as presented in Table 4. In the present research, the ANFIS-LMAD classifier yields a convergence time faster than the previous classifiers presented in Table 4 through Table 6.

In Table 5 and Table 6, each number in column 5 represents the selected feature for the datasets used. Detailed descriptions of the selected features of the two benchmark data sets can be found in Bache and Lichman (2013).

The present work evaluates the performance of the proposed model using two benchmark datasets for Medical diseases diagnosis. The proposed classifier was justified by the performance measures used in the present study. The test performance of the classifier was determined by the computation of sensitivity, specificity, performance error and total classification accuracy, as shown in Table 3 through Table 6, for the Statlog-Heart Disease and Hepatitis Disease datasets. The developed ANFIS-LMAD classifier yields better results than other classifiers, as presented in Table 5 and Table 6. Based on a comparison of the results, the developed technique produced reasonable results in diagnosing the possible presence of medical disease inpatients.

Hold-out validation is used by dividing the training and data sets into 70% to 30%, respectively (Schneider, 1997; Thacker et al., 2004; Fortmann-Roe, 2012). The test result is better than the training because we have not repeatedly used a holdout set to test a model during development. In other words, test data were not used during training. This shows that the classifier has learned things well and given a better desired output. The results provided good performance in accuracy and generalisation.

Compared to the literature discussed of the existing methods described above, the proposed hybridised intelligent system outperformed other existing methods. This assertion is based on the following observations: in Table 5, it is clearly confirmed that none of the research studies has success rates higher than 85.88% and selected features greater than six. In the grid partition method, the higher the input variables are, the less accuracy is to be obtained because it generates rules by enumerating all possible combinations of membership functions of all inputs. Nonetheless, our proposed method was able to produced better accuracy results. Employing an indexing membership function in a row-wise vector, which is a new way of code optimisation, however was used to make convergence time faster, and the related mentioned existing algorithms lacks this new technique. Nonetheless, unlike other existing methods, the Jacobian matrix is computed via analytic derivation and chain rule, which also contributed to a faster convergence speed by using sparse structure and produced results with better accuracy.

There are limitations associated with this study that constrain the generalisation of the results. The main limitations are the selection of two benchmark medical data sets from UCI machine learning repository and the use of the grid partitioning method, which caused by a high partition of fuzzy rules. Other approaches were taken in developing and testing the proposed model in this study such as the feature selection, learning algorithm, validation method and performance measure.

The proposed hybridised intelligent technique could be enhanced in the future by applying another means of derivation in computation of Jacobian matrix in order to increase convergence speed and accuracy of the results. The use of intelligent algorithms such as Genetic Algorithm, Particle Swarm Optimization and Ant Bee Colony algorithm with k-fold cross validation method could be used.

CONCLUSION

The objective of this study is to design a hybridised intelligent technique for the diagnosis of medical diseases. This study proposed three major novelty techniques by using and modifying the Levenberg-Marquardt algorithm, indexing membership functions using the vectorisation technique and employed an effective way to compute a sparse Jacobian matrix using analytical derivation.

The proposed technique provides a good representation of the training and test samples, input and output mapping and convergence time by using two different datasets, which make the model more robustness.

The classifier was able to learn how to classify input data by class level and correctly maximise classified data, while minimising instances of incorrectly classified patterns. Its performance was evaluated based on machine learning process.

The applicability of the proposed technique in data classification using two benchmark datasets in the area of medical diagnosis was demonstrated. The results of the proposed technique are better than the results of existing methods in the literatures and have potential in diagnosing medical diseases.

ACKNOWLEDGEMENTS

This research is partly financed by FRGS grant (203/ PMATHS/6711368) from the Ministry of Higher Education, Malaysia.

REFERENCES

- Anooj, P. (2012). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University-Computer and Information Sciences*, 24(1), 27-40.
- Bache, K., & Lichman, M. (2013). *UCI machine learning repository*. School of Information and Computer Sciences, University of California, Irvine.
- Davis, C. P. (2016). Hepatitis (Viral Hepatitis, A, B, C, D, E, G). *MedicineNet*. Retrieved June 1, 2016, from http://www.medicinenet.com/viral_hepatitis/article.htm
- Fortmann-Roe, S. (2012). *Accurately measuring model prediction error*. Retrieved December 05, 2015, from <http://scott.fortmann-roe.com/docs/MeasuringError.html>
- Hagras, H. A. (2004). A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. *Fuzzy Systems, IEEE Transactions on*, 12(4), 524-539.
- Hedeshi, N. G., & Abadeh, M. S. (2014). Coronary artery disease detection using a fuzzy-boosting PSO approach. *Computational Intelligence and Neuroscience*, 2014(6), 1-12.
- Ho, W. H., Tsai, J. T., Lin, B. T., & Chou, J. H. (2009). Adaptive network-based fuzzy inference system for prediction of surface roughness in end milling process using hybrid Taguchi-genetic learning algorithm. *Expert Systems with Applications*, 36(2), 3216-3222.
- Jaganathan, P., & Kuppuchamy, R. (2013). A threshold fuzzy entropy based feature selection for medical database classification. *Computers in Biology and Medicine*, 43(12), 2222-2229.
- Jamsandekar, S. S., & Mudholkar, R. R. (2014). Fuzzy Classification System by Self Generated Membership Function Using Clustering. *International Journal of Information Technology*, 6(1), 697-704.
- Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3), 665-685.
- Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*. Prentice Hall, Upper Saddle River.
- Kahramanli, H., & Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. *Expert Systems with Applications*, 35(1), 82-89.
- Kathy, L. T. (2013). *Data Mining Concepts*. Retrieved September 10, 2015, from http://docs.oracle.com/cd/E11882_01/datamine.112/e16808/classify.thm#DMCON035
- Nahato, K. B., Harichandran, K. N., & Arputharaj, K. (2015). Knowledge mining from clinical datasets using rough sets and backpropagation neural network. *Computational and Mathematical Methods in Medicine*, 2015, 1-13.
- Neshat, M., Sargolzaei, M., Toosi, A. N., & Masoumi, A. (2012). Hepatitis disease diagnosis using hybrid case based reasoning and particle swarm optimization. *ISRN Artificial Intelligence*, 2012, 1-6.

- Sagir, A. M., & Sathasivam, S. (2017). A Novel Adaptive Neuro Fuzzy Inference System Based Classification Model for Heart Disease Prediction. *Pertanika Journal of Science and Technology*, 25(1), 43-56.
- Schneider, J., & Moore, A. W. (2000). *A locally weighted learning tutorial using vizier 1.0*. Carnegie Mellon University, the Robotics Institute.
- Subbulakshmi, C., & Deepa, S. (2015). Medical Dataset Classification: A Machine Learning Paradigm Integrating Particle Swarm Optimization with Extreme Learning Machine Classifier. *The Scientific World Journal*, 2015(2015), 1-12.
- Thacker, B. H., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., & Rodriguez, E. A. (2004). *Concepts of model verification and validation*. SciTech Connect. United States, USA: Los Alamos National Lab.
- Uzer, M. S., Yilmaz, N., & Inan, O. (2013). Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification. *The Scientific World Journal*, 2013(2013), 1-10.
- WHO. (2015). *Fact sheets*. World Health Organisation. Retrieved August 28, 2015, from <http://www.who.int/mediacentre/factsheets/en/>
- Zadeh, L. A. (1965). Information and control. *Fuzzy sets*, 8(3), 338-353.
- Zadeh, L. A. (2004). *Fuzzy Logic Systems: Origin, Concepts, and Trends*. Distinguished Lecture at Hong Kong Baptist University, Hong Kong.



Ability of *Ipomoea aquatica* Forssk. to Remediate Phenol in Water and Effects of Phenol on the Plant's Growth

Siew-Yi Lee^{1,3}, Siti Aqlima Ahmad², Siti Roslina Mustapha¹ and Janna Ong-Abdullah^{1,3*}

¹Department of Microbiology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

²Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

³Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Despite wide applications in industries, phenol pollution leads to many health effects, and one of the technologies used to clean up phenol pollution is phytoremediation. The aim of this research was to assess the remediation ability of *Ipomoea aquatica* Forssk., which is easy to handle and has a fast growth rate. Plantlet was grown in water spiked with 0.05, 0.10, 0.20, 0.30 and 0.40 g/L phenol, followed by daily observation of the plantlets morphology and tracking of phenol concentration in the water and plantlet extracts via 4-aminoantipyrine (4-AAP) assay. Plantlet's roots in 0.10 g/L phenol (57.42 ± 1.41 mm) were significantly longer ($p < 0.05$) than those of the control plantlets (43.57 ± 3.87 mm) in contrast to other phenol concentrations which had stunted roots growth. *I. aquatica* Forssk. was able to survive with 0.30 g/L phenol despite exhibiting yellowing of leaves and increased sensitivity to scarring on the stems. The plantlets were able to completely remove the phenol from the water spiked with phenol at 0.05 g/L after 12 days of growth. However, the highest average rate of phenol removal was 0.021 g/L/day from water spiked with 0.30 g/L phenol. Phenol analysis on the plantlets' extracts revealed that *I. aquatica* Forssk. had degraded the absorbed phenol. This observation is of significant interest as it highlights the potential of *I. aquatica* Forssk. for use as a phytoremediator to clean up phenol contaminated water.

Keywords: 4-aminoantipyrine assay, *Ipomoea aquatica* Forssk., phenol, phytoremediator

Article history:

Received: 01 April 2016

Accepted: 04 July 2016

E-mail addresses:

siewyi91@gmail.com (Siew-Yi Lee),

aqlima@upm.edu.my (Siti Aqlima Ahmad),

roslina.sr@gmail.com (Siti Roslina Mustapha),

janna@upm.edu.my (Janna Ong-Abdullah)

*Corresponding Author

INTRODUCTION

Phenol has wide applications in several industries such as the manufacturing of dyes, plastics, fertilisers, pesticides, polymeric resins and in oil refineries (McCall et al.,

2009). As phenol has high water affinity and low volatility (Prpich & Daugulis, 2005), exposure to untreated phenol released from these industries into the aquatic environment has been reported to be extremely harmful (Huang et al., 2014). Besides being a contributing factor to non-potable and non-palatable drinking water (Sihem, Lehocine, & Miniai, 2012), excessive exposure could also cause health effects on skin, digestive system, brain, kidney, heart, and may also result in genetic damages (Al-Muhtaseb et al., 2011). McCall et al. (2009) reported that oral ingestion might cause vomiting, mouth sores, bloody diarrhoea, nausea, dark urine, and even death in certain cases.

There are currently a few technologies that are being applied to clean up phenol pollution sites. These include (a) extracted enzyme (Deva, Arun, Arthanareeswaran, & Sivashanmugam, 2014), (b) microbial biodegradation (Basha, Rajendran, & Thangavelu, 2010), (c) soil removal (Xia, Ma, Liu, & Fan, 2012) and (d) chemical oxidation (Sihem et al., 2012). These technologies, however, require high operating costs and consist of other limitations, which as a result, reduce their effectiveness. Microbial biodegradation can be easily inhibited and has low microbial survival rate under extended phenol exposure. Both soil removal and chemical oxidation are destructive to the sites.

Phytoremediation is a technology that utilises plants to remove or transform toxic chemicals present in soils, sediments, ground water, surface water, and even the atmosphere to clean up or reduce the level of pollution (Susarla, Medina, & Mccutcheon, 2002). There are many advantages of phytoremediation over other technologies, as reported in various studies. The main advantages of this technology include being environmental friendly due to the usage of plants as a main component (Ahemad, 2015), low cost (Rungwa, Arpa, Sakulas, Harakuwe, & Timi, 2013), ability to remove a wide range of pollutants simultaneously (Batty & Dolan, 2011), as well as being an *in situ* bioprocess, where neither soil nor water is removed from the sites. Thus, the original structure of treated sites remains undisturbed (Ho et al., 2012). Both plants and residuals from phytoremediation are valuable components, which can be used later for biofuel production or carbon sequestration upon completion of the process (Batty & Dolan, 2011).

After optimisation of major parameters such as the phenol concentration, sunlight and treatment time (Wang, et al., 2014a), plants have the potential to remediate phenol efficiently. In the phenol degradation pathway, plants utilise peroxidase (Jha et al., 2013) instead of phenol hydroxylase, as used by other microorganisms (Kwon & Yeom, 2009). According to Jha et al. (2013), plants degrade phenol through the catechol-cleavage pathways, where peroxidase hydroxylates phenol and forms catechol. Catechol will then cleave to form muconic acid and fumaric acid which enter the Krebs cycle, completing the phenol degradation pathway. Above tolerable level, phenol might cause damages such as reduction in root length (Ratsch, 1983), root browning (Jha et al., 2013) and seed germination inhibition (Ibanez, Alderete, Medina, & Agostini, 2012).

Aquatic plants play important roles in biochemical cycles, nutrient mobilisation and sediment dynamics in freshwater ecosystems (Wang, et al., 2014b). They also readily remediate toxic pollutants in freshwater and seawater (Rahman & Hasegawa, 2011). Some aquatic plants can concentrate metals inside the plants to concentrations much higher than the surroundings (Mishra & Tripathi, 2008). The aquatic plant to be studied in this

research is *Ipomoea aquatica* Forssk. of the family Convolvulaceae. It is commonly known as water spinach, “nalanibhaji” and “kalmisag” (Manvar, 2011), or “kangkung” and “ong choy” in Malaysia. It is a semi-aquatic, fast-growing vegetable with hollow stems and arrowhead-shaped leaves (Chen et al., 2010). It has high medicinal values with potential as an anti-tumour agent (Ullah et al., 2013) and the ability to lower the mortality rate in coronary heart disease (Sivaraman & Muralidaran, 2010). This plant has also been used in treating iron-deficiency anaemia (Islam et al., 2013), arsenic poisoning (Alkiyumi et al., 2012) and bronchitis (Manvar, 2011). *I. aquatica* Forssk. is a native plant in sites usually contaminated with excessive nutrients (Chen et al., 2010) and it is often studied as a phytoremediator to remove excessive nitrogen (Jampeetong, Brix, & Kantawanichkul, 2012), chromium (Chen et al., 2010), cadmium (Hseu, Jien, Wang, & Deng, 2013) and manganese (Das Sharma, Ravnang, & Nazia, 2014). Thus, the objectives of this study were to study the effects of phenol on *I. aquatica* Forssk. and investigate the ability of the plantlets to remove phenol from phenol-spiked water in order to determine the compatibility of this plant to phytoremediate phenol in water.

MATERIALS AND METHODS

Plant material

The seeds of *I. aquatica* Forssk. were bought from an agricultural seeds shop in Negeri Sembilan, Malaysia (Coordinates 2°43'50.58" N, 101°56'11.76" E). The seeds were kept dry to prevent early germination and spoilage. Germinated plantlets were authenticated by a botanist and deposited at the herbarium of the Biodiversity Unit, Universiti Putra Malaysia, with the specimen voucher number SK2675/15.

Experimental Set-up

Phenol (purchased from Merck, Darmstadt, Germany) concentrations tested in this experiment were 0.05, 0.10, 0.20, 0.30 and 0.40 g/L. This phenol concentrations range was selected based on the reported environmental phenol pollution range of 0.05 to 0.60 g/L (Paisio, Agostini, Gonzalez, & Bertuzzi, 2009), and *I. aquatica* Forssk. could not survive above 0.40 g/L phenol (our unpublished data). Germinated seedlings were grown in containers measuring 30 cm (length), 23 cm (width) and 10 cm (depth). Young plantlets were used as they were more sensitive to pollutants and any morphological changes can be easily noted. Each container held 4 litres of phenol-spiked distilled water, and the water circulation system was supported by a Life Tech AP1000 water pump (Eng Hing Aquatics, Selangor, Malaysia) to circulate the water. Distilled water was used to avoid potential contamination by other microorganisms. No nutrients were added into the system as newly germinated seedlings sustained their growth from nutrients in the cotyledons, while the plantlets were able to photosynthesise. The whole container containing the plantlets was placed on a bench next to a window. All the plantlets were placed equidistant (2.54 cm apart) inside each container to minimise competition and were ensured that roots were always in contact with water, while both leaves and stems were above water level.

Seeds were rinsed to remove dirt on the seed coats, and then soaked in distilled water for 72 hours before transferring to the container. This germination step before placing the seedlings into the container enables selection of healthy seedlings with similar root size to start the experiment. A total of thirty germinated seedlings, with roots measuring approximately 2.00 cm, were placed in the container with phenol-spiked water (designated as Day 0) and grown for 14 days. The experiment was conducted in triplicate for each of the phenol concentration used, which involved a total of ninety seeds distributed equally in three different containers. Growth parameters such as lengths of root, stem, primary leaf and petiole were measured every day for 14 days. Extra phenotypic information such as lengths of secondary leaf, root hair, thickness of root and formation of roots at the nodes was also recorded at the end of the experiment. Phenol remaining in the water and retained in the plantlets were quantified daily and at the end of the experiment, respectively. The pH of the water was not monitored in this experiment since plant peroxidase responsible for phenol degradation was reported to be capable of functioning in a wide pH range from pHs 4 to 9 (Gonzalez et al., 2006).

Phenol assay: 4-aminoantipyrine (AAP) assay

Quantification of phenol remaining in the water was carried out daily throughout the 14 days of growth by removing 1 ml of the water from each container for the 4-aminoantipyrine (4-AAP) assay, following the protocol of Ettinger, Ruchloft, and Lishka (1951). Likewise, the 4-AAP assay was also used to measure absorbed phenol within the plantlets. All the surviving plantlets were pooled and ground in a pestle and mortar, followed by centrifugation at $10,625 \times g$ for 15 minutes. The supernatant was then subjected to the 4-AAP assay. Extract from *I. aquatica* Forssk., grown in distilled water without phenol, was used as a control in the 4-AAP assay. Each sample was assayed in triplicate to obtain the mean concentration of phenol present. All the chemicals used were purchased from Sigma Aldrich, St. Louis, USA.

Statistical Analysis

Measurements of the plantlet's length and phenol assay were carried out daily throughout the 14-day experiments. Roots, stems, leaves, and petioles from all surviving plantlets were measured daily and presented as mean and standard deviation as error bars in the graph. Statistical analysis of the data by one way analysis of variance (ANOVA) was conducted using GraphPad Prism version 5.00 for Windows. Significant differences between means were determined using Tukey's Multiple Comparison Test, where p value less than 0.05 was regarded as significant.

RESULTS AND DISCUSSION

Plantlets growth in distilled water

Ninety germinated plantlets were transferred and grown in containers with phenol-free distilled water for 14 days, and observations on the changes of the roots, stems, leaves, and petioles were recorded daily, as shown in Figure 1(a). Throughout the 14 days of growth, all the plantlets exhibited rapid growth during the early stage, but plateaued as they matured towards the end

of the growth period (Figure 2). The roots were visible and increased significantly in length from Day 1 until Day 10. Meanwhile, measurements of leaves were taken only starting on Day 3 after most of the cotyledons had withered and fallen off, exposing the young green leaves. The leaf length increased until Day 7, with an average increment of 7.66 ± 1.35 mm daily, which remained constant after Day 8.

The stems, however, exhibited a different growth pattern. While both leaf and petiole lengths increased from Day 0 until Day 14, the stem length reached its maximum increment of 9.69 ± 0.62 mm on Day 4, following which, no significant increment was observed. It was also observed that root hairs continued to form and grow throughout the stems.

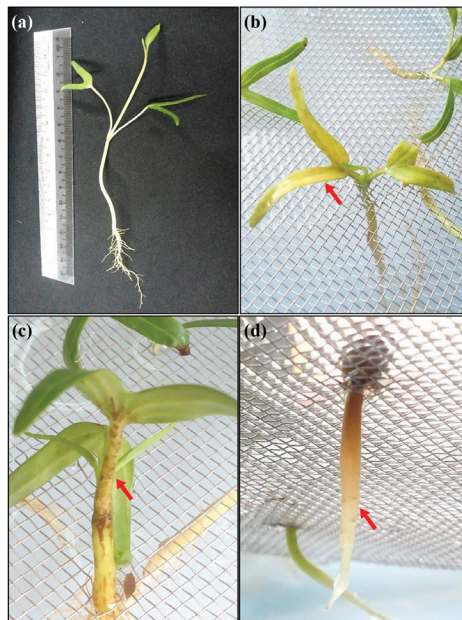


Figure 1. Daily observation of the *I. aquatica* Forssk. including (a) Measurement of structure sizes, and physical effects of phenol on the plantlets (arrows) such as (b) Yellowish leaves, (c) Scars on stems, and (d) Soggy roots.

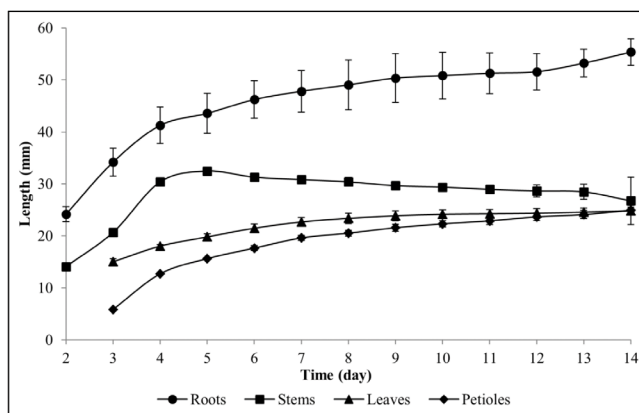


Figure 2. Growth of Control plantlets in distilled water. Values plotted are means \pm standard deviation (SD).

Plantlets subjected to phenol-spiked water

Figure 3 shows that all the plantlet parts exhibited similar growth profiles as that of the control. This observation indicated that at different concentrations, phenol affected the growth rate but not the growth profiles of treated plantlets.

From the data obtained in this study, it was evident that *I. aquatica* Forssk. was able to grow in all the phenol concentrations, except for 0.40 g/L. As phenol concentration increased, the plantlets' growth slowed and the plantlets were mostly stunted. Interestingly, the plantlets grown in 0.05 g/L phenol were always shorter than those in 0.10 g/L phenol, including their root length. This condition could be due to a basal growth induction triggered by phenol concentration of 0.10 g/L. For instance, as displayed in Figure 3(a), the roots of plantlets grown in 0.10 g/L phenol were significantly longer ($p < 0.05$) than all other concentrations, including the Control plantlets. The highest effect was achieved on Day 5 as the plantlets grown in 0.10 g/L have root lengths (57.42 ± 1.41 mm) which were 13.85 ± 1.11 mm longer ($p < 0.05$) than the control plantlets (43.57 ± 3.87 mm). Upon termination of growth (Day 14), plantlets in 0.10 g/L (61.02 ± 1.50 mm) were 5.67 ± 3.11 mm longer than the Control plantlets (55.36 ± 2.55 mm). Growth induction by phenol could occur as plant peroxidase degrades the phenol through the catechol-cleavage pathway (Jha et al., 2013) to generate fumaric acid for subsequent production of useful metabolites such as pyruvic acid, acetic acid and acetyl-CoA in the Krebs cycle.

In this study, plantlets exposed to phenol up to 0.20 g/L exhibited green and fully expanded leaves. However, plantlets in 0.30 g/L phenol were withered and the leaves were yellowish, as shown in Figure 1(b), postulating a compromised on the photosynthesis apparatus of the plantlets. This result is similar to the recent study by Wang et al. (2014a), where the leaves of *Polygonum orientale* turned yellow and brown upon treatment with phenol above 0.10 g/L. The effects that phenol had on the photosynthetic mechanism of plants were discussed in the recent studies of Hang and Zhao (2010), and Li, Zhang, Xiew, Li, and Zhang (2015). Hang and Zhao (2010) reported that phenol stress inactivated the Photosystem II, which is the photosynthetic reaction centre, and caused a restriction on excited electrons transfer resulting in a reduction of the actual photochemical reactions. This could probably explain the yellowing of leaves observed in this study. Li et al. (2015) reported that phenol concentration above 0.20 g/L inhibited the photosynthetic activities of *Salix babylonica* besides reducing their growth and repair capabilities. Nonetheless, further studies need to be done to investigate the actual effects of phenol on *I. aquatica* Forssk. photosynthetic system during phenol treatment.

Phenol was also observed to have significantly reduced stem length ($p < 0.05$) by all except 0.10 g/L phenol in comparison with the Control, as shown in Figure 3(b). Scars were frequently observed on the stems of plantlets exposed to 0.20 g/L and above phenol despite precautions taken during the handling of plantlets throughout the experiment [Figure 1(c)]. Increased phenol concentrations also led to the formation of shorter petioles [Figure 3(d)].

Plantlets exposed to phenol above 0.30 g/L were fatal, whereby 27.8% of them in 0.30 g/L phenol died on Day 14 and 87.8% in 0.40 g/L phenol died on Day 6 (Table 1). The highest death rate in 0.40 g/L phenol was achieved on Day 4. The physical effects observed prior to death in 0.40 g/L phenol were the roots became soggy, transparent and brittle to touch as depicted

in Figure 1(d). Root was the first plant's part to establish direct contact and responded to the phenol (Zhou, Tang, & Wu, 2013; Park et al., 2016), explaining the apparent damages observed prior to death of the whole plantlet.

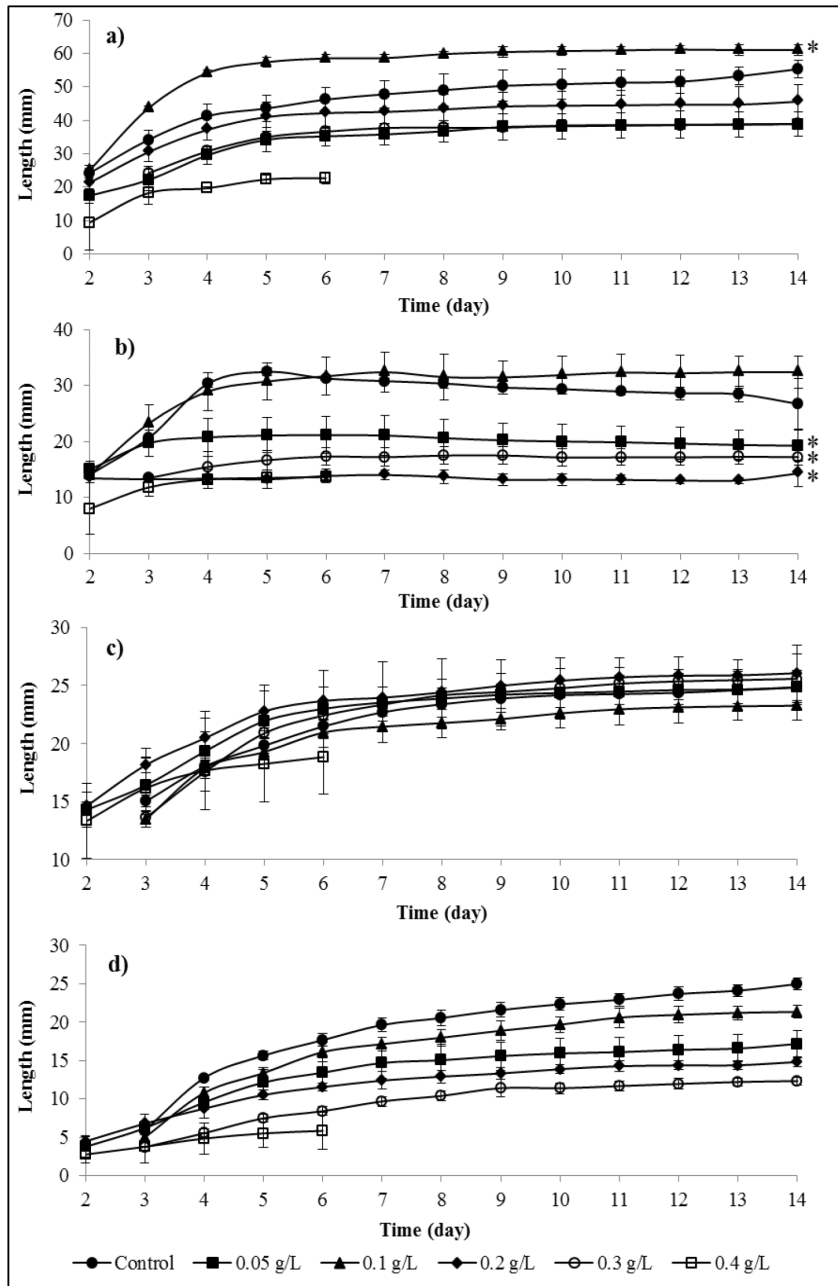


Figure 3. Growth of *I. aquatica* Forssk. in different phenol concentrations: (a) Roots, (b) Stems, (c) Leaves, and (d) Petioles. Values plotted are means \pm standard deviation (SD). *Growth was of significant difference ($p < 0.05$).

Table 1
Total plantlets survived and phenol assay on plantlet extracts

Phenol Concentration (g/L)	Number of Plantlets Survived	Time of Assay on Absorbed Phenol	Phenol in Plantlet Extract (g/L)
0*	90	Day 14	0.02 ± 0.00
0.05	76	-	-
0.1	58	i) Day 9 ii) Day 11	i) 0.10 ± 0.01 ii) 0.08 ± 0.01
0.2	60	-	-
0.3	65	Day 14	0.07 ± 0.01
0.4	11	-	-

*Control plantlets (no phenol exposure) were grown in distilled water only.

Phenol Degradation

I. aquatica Forssk. was capable of gradually removing phenol starting from the first day of growth (Figure 4), with over 80.0% of the phenol removed from the water in all the phenol concentrations below 0.40 g/L. Total removal of phenol (100%) from water spiked with 0.05, 0.10 and 0.30 g/L phenol took 14 days. However, only 0.17 ± 0.01 g/L (84.6%) of phenol was removed from the 0.20 g/L phenol-spiked water, while 0.14 ± 0.01 g/L (34.2%) was removed from the 0.40 g/L phenol-spiked water (the experiment was terminated on Day 6 for 0.4 g/L phenol as almost all the plantlets had died). Compared to *I. aquatica* Forssk., *Vetiveria zizanoides* L. Nash plants removed 0.05 and 0.10 g/L phenol in 4 days (Singh, Melo, Eapen & D'Souza, 2008), *Vicia sativa* removed 0.10 g/L phenol in 7 days, while alfalfa removed 0.10 g/L phenol in 30 days (Flocco et al., 2002).

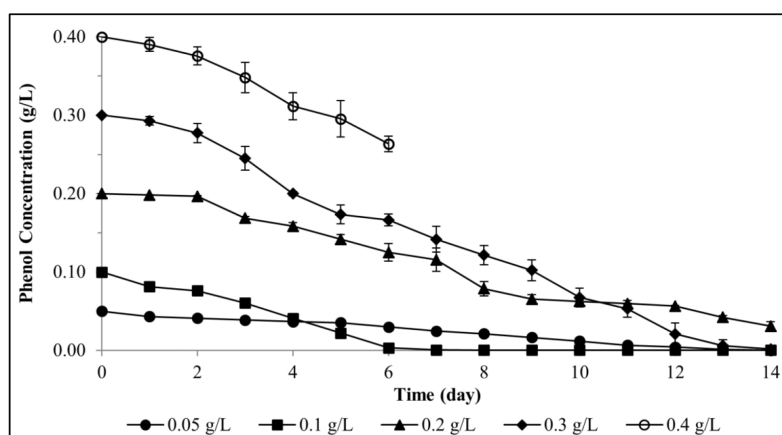


Figure 4. Daily analysis of phenol remaining in the water. Aliquot (1.00 mL) of water from each container was subjected to 4-AAP assay to determine daily concentration of phenol remaining in the water. Data plotted are means ± standard deviation (SD).

Phenol removal activity was calculated as quantity of phenol removed from the water daily throughout the experimental growth period. The highest rate was attained in 0.30 g/L phenol, averaging around 0.021 g/L per day (Figure 5). Although *I. aquatica* Forssk. achieved the highest phenol degrading activity in 0.30 g/L, the growth of the plantlets was compromised (signs of inhibition), as discussed in the previous section (plantlets subjected to phenol-spiked water). *I. aquatica* Forssk. was the least affected by 0.10 g/L phenol treatment, exhibiting a phenol removal rate of 0.013 g/L per day, which was significantly lower compared to 0.30 g/L phenol treatment ($p < 0.05$). Nonetheless, phenol removal activity at 0.20 g/L was 0.012 g/L per day, while much a lower activity was observed in 0.05 g/L phenol (0.004 g/L per day).

Further analysis of the crude extracts (Table 1) indicated that *I. aquatica* Forssk. metabolised or transformed the absorbed phenol instead of accumulating it in the plantlet tissues. Theoretically, a complete removal of 0.30 g/L phenol from the water and its subsequent accumulation in the plantlets without degradation would total up to be 0.32 g/L. However, our analysis on the plantlets extract detected only 0.07 ± 0.01 g/L phenol as compared with 0.02 g/L phenol in the control (plantlets grown in distilled water only). This observation implies that the absorbed phenol has either been biodegraded or biotransformed within the plantlets. Similar results were obtained when plantlets exposed to 0.10 g/L phenol were extracted and assayed on both Day 9 and Day 11, indicating a decrease in phenol concentration over two days of growth. Biodegradation of phenol by *I. aquatica* Forssk. was supported by Cataldo et al. (1987), which reported that plants had the tendency to metabolise phenol readily instead of accumulating it due to their high respiratory decomposition rate. Meanwhile, necrotic lesions and leaves fall were observed if there was accumulation of pollutants in plants (Hawrylak-Nowak, Kalinowska, & Szymanska, 2012). Such leaf defects were absent in this study.

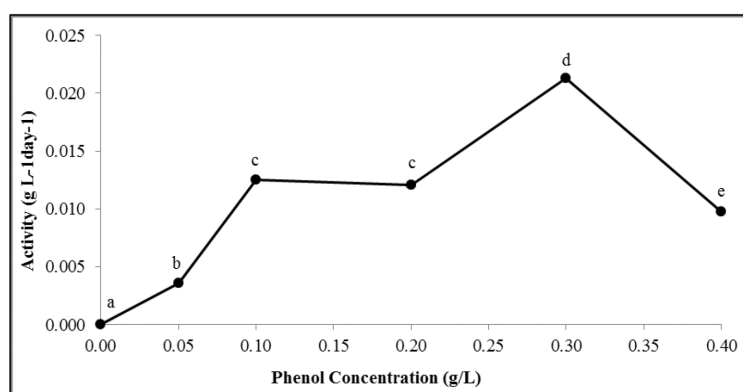


Figure 5. Rate of phenol removal by *I. aquatica* Forssk. Daily removal of phenol from the water by plants was assayed. Data plotted represent means of triplicates. Rates with different letters indicate significant differences ($p < 0.05$). The standard deviation bars are too small to be visible.

CONCLUSION

The maximum tolerable phenol concentration for *I. aquatica* Forssk. was 0.30 g/L, with a survival rate of 72.2% despite various physical and physiological damages exhibited by the plantlets. The positive effect of 0.10 g/L phenol on plantlets root lengths was observed and peaked at Day 5, with a value of 57.42 ± 1.41 mm, which was significantly longer ($p < 0.05$) than the control plantlets by 13.85 ± 1.11 mm. The maximum phenol removal rate by the plantlets was achieved with 0.30 g/L phenol at 0.021 g/L/day. The findings also showed that *I. aquatica* Forssk. was capable of removing phenol from water and possibly biodegrading it with zero post-treatment, and it is a natural low cost, easy to handle and fast growing phenol phytoremediator.

ACKNOWLEDGEMENT

We thanked Prof. Dr. Mohd. Syed Arif, for accessibility to his laboratory to conduct all the experiments.

REFERENCES

- Ahemad, M. (2015). Enhancing phytoremediation of chromium-stressed soils through plant-growth-promoting bacteria. *Journal of Genetic Engineering and Biotechnology*, 13(1), 51-58.
- Alkiyumi, S. S., Abdullah, M. A., Alrashdi, A. S., Salama, S. M., Abdelwahab, S. I., & Hadi, A. H. A. (2012). *Ipomoea aquatica* extract shows protective action against thioacetamide-induced hepatotoxicity. *Molecules*, 17(5), 6146-6155.
- Al-Muhtaseb, A. H., Ibrahim, K. A., Albadarin, A. B., Ali-Khashman, O., Walker, G. M., & Ahmad, M. N. (2011). Remediation of phenol-contaminated water by adsorption using poly (methyl methacrylate) (PMMA). *Chemical Engineering Journal*, 168(2), 691-699.
- Basha, K. M., Rajendran, A., & Thangavelu, V. (2010). Recent advances in the biodegradation of phenol: A review. *Asian Journal of Experimental Biological Sciences*, 1(2), 219-234.
- Batty, L. C., & Dolan, C. (2011). The potential use of phytoremediation for sites with mixed organic and inorganic contamination. *Critical reviews in environmental science and technology*, 43(3), 217-259.
- Cataldo, D., A., Bean, R., M., & Fellows, R., J. (1987). Uptake and fate of phenol, aniline and quinoline in terrestrial plants. In R. H., Gray, E. K., Chess, P. J., Mellinger, R. G. Riley, & D. L. Springer (Eds), *Hanford life sciences symposium and environmental research on complex organic mixtures* (pp. 631-640). Richland, WA: Pacific Northwest Lab.
- Chen, J. C., Wang, K. S., Chen, H., Lu, C. Y., Huang, L. C., Li, H. C., ... Chang, S. H. (2010). Phytoremediation of Cr(III) by *Ipomoea aquatica* (water spinach) from water in the presence of EDTA and chloride: Effects of Cr speciation. *Bioresource Technology*, 101(9), 3033-3039.
- Dassharma, K., Ravnang, P., & Nazia, Q. (2014). *Eichhornia* and *Ipomoea*: Efficient phytoremediators of manganese. *International Journal of Life Sciences*, 2(2), 143-147.
- Deva, A. N., Arun, C., Arthanareeswaran, G., & Sivashanmugam, P. (2014). Extraction of peroxidase from waste Brassica oleracea used for the treatment of aqueous phenol in synthetic waste water. *Journal of Environmental Chemical Engineering*, 2(2), 1148-1154.

- Ettinger, M., Ruchhoft, C., & Lishka, R. (1951). Sensitive 4-aminoantipyrine method for phenolic compounds. *Analytical Chemistry*, 23(12), 1783-1788.
- Flocco, C. G., Lo Balbo, A., Carranza, M. P., & Giulietti, A. M. (2002). Removal of phenol by alfalfa plants (*Medicago sativa* L.) grown in hydroponics and its effect on some physiological parameters. *Engineering in Life Sciences*, 22(1-2), 43-54.
- González, P. S., Capozucca, C. E., Tigier, H. A., Milrad, S. R., & Agostini, E. (2006). Phytoremediation of phenol from wastewater, by peroxidases of tomato hairy root cultures. *Enzyme and microbial technology*, 39(4), 647-653.
- Hang, G., & Zhao, Z. (2010). Light response characteristics of photosynthesis of four xerophilous shrubs under different soil moistures. *Acta Ecologica Sinica*, 30(15), 4019-4026.
- Hawrylak-Nowak, B., Kalinowska, M., & Szymańska, M. (2012). A study on selected physiological parameters of plants grown under lithium supplementation. *Biological trace element research*, 149(3), 425-430.
- Ho, Y. N., Mathew, D. C., Hsiao, S. C., Shih, C. H., Chien, M. F., Chiang, H. M., & Huang, C. C. (2012). Selection and application of endophytic bacterium *Achromobacter xylosoxidans* strain F3B for improving phytoremediation of phenolic pollutants. *Journal of Hazardous Materials*, 219, 43-49.
- Hseu, Z. Y., Jien, S. H., Wang, S. H., & Deng, H. W. (2013). Using EDDS and NTA for enhanced phytoextraction of Cd by water spinach. *Journal of Environmental Management*, 117, 58-64.
- Huang, Z. Z., Wang, P., Li, H., Lin, K. F., Lu, Z. Y., Guo, X. J., & Liu, Y. D. (2014). Community analysis and metabolic pathway of halophilic bacteria for phenol degradation in saline environment. *International Biodeterioration and Biodegradation*, 94, 115-120.
- Ibanez, S., G., Alderete, L., G., S., Medina, M., I., & Agostini, E. (2012). Phytoremediation of phenol using *Vicia sativa* L. plants and its antioxidative response. *Environmental Science and Pollution Research*, 19(5), 1555-1562.
- Islam, M. M. T., Talukder, A., Rahman, T., Islam, J. M. M., Hossain, S., & Shekhar, H. U. (2013). Food functionality of popular and commonly consumed indigenous vegetables and fruits from Bangladesh. *Food and Nutrition Sciences*, 4(07), 741-747.
- Jampeetong, A., Brix, H., & Kantawanichkul, S. (2012). Effects of inorganic forms on growth, morphology, nitrogen uptake capacity and nutrient allocation of four tropical aquatic macrophytes (*Salvinia cucullata*, *Ipomoea aquatica*, *Cyperus involucratus* and *Vetiveria zizanioides*). *Aquatic Botany*, 97(1), 10-16.
- Jha, P., Jobby, R., Kudale, S., Modi, N., Dhaneshwar, A., & Desai, N. (2013). Biodegradation of phenol using hairy roots of *Helianthus annuus* L. *International Biodeterioration and Biodegradation*, 77, 106-113.
- Kwon, K. H., & Yeom, S. H. (2009). Optimal microbial adaptation routes for the rapid degradation of high concentration of phenol. *Bioprocess and biosystems engineering*, 32(4), 435-442.
- Li, H., Zhang, G. C., Xiew, H. C., Li, K., & Zhang, S. Y. (2015). The effects of the phenol concentrations on photosynthetic parameters of *Salix babylonica* L. *Photosynthetica*, 53(3), 430-435.
- Manvar, M. N. (2011). Pharmacognostical investigations on *Ipomoea aquatica* Forssk. *International Journal of Pharmaceutical Sciences and Research*, 2(11), 2812-2815.

- McCall, I. C., Betanzos, A., Weber, D. A., Nava, P., Miller, G. W., & Parkos, C. A. (2009). Effects of phenol on barrier function of a human intestinal epithelial cell line correlate with altered tight junction protein localization. *Toxicology and applied pharmacology*, 241(1), 61-70.
- Mishra, V. K., & Tripathi, B. D. (2008). Concurrent removal and accumulation of heavy metals by the three aquatic macrophytes. *Bioresource technology*, 99(15), 7091-7097.
- Park, J. H., Yoon, J. H., Depuydt, S., Oh, J. W., Jo, Y. M., Kim, K. T., Brown, M. T., & Han, T. J. (2016). The sensitivity of an hydroponic lettuce root elongation bioassay to metals, phenol and wastewaters. *Ecotoxicology and Environmental Safety*, 126, 147-153.
- Paisio, C., Agostini, E., Gonzalez, P., & Bertuzzi, M. (2009). Lethal and teratogenic effects of phenol on *Bufo arenarum* embryos. *Journal of Hazardous Materials*, 167(1), 64-68.
- Prpich, G. P., & Daugulis, A. J. (2005). Enhanced biodegradation of phenol by a microbial consortium in a solid-liquid two phase partitioning bioreactor. *Biodegradation*, 16(4), 329-339.
- Rahman, M. A., & Hasegawa, H. (2011). Aquatic arsenic: phytoremediation using floating macrophytes. *Chemosphere*, 83(5), 633-646.
- Ratsch, H. C. (1983). Interlaboratory root elongation testing of toxic substances on selected plant species. *Environmental Protection Agency*, 600, 3-85.
- Rungwa, S., Arpa, G., Sakulas, H., Harakuwe, A., & Timi, D. (2013). Phytoremediation- An eco-friendly and sustainable method of heavy metal removal from closed mine environments in Papua New Guinea. *Procedia Earth and Planetary Science*, 6, 269-277.
- Sihem, A., Lehocine, M. B., & Miniai, H. A. (2012). Batch adsorption of phenol from industrial waste using cereal by-products as a new adsorbent. *Energy Procedia*, 18, 1135-1144.
- Singh, S., Melo, J., Eapen, S., & D'Souza, S. (2008) Potential of vetiver (*Vetiveria zizanoides* L. Nash) for phytoremediation of phenol. *Ecotoxicology and Environmental Safety*, 71(3), 671-676.
- Sivaraman, D., & Muralidaran, P. (2010). Hypolipidemic activity of *Ipomoea aquatica* Forsk. leaf extracts on lipid profile in hyperlipidemic rats. *International Journal of Pharmaceutical and Biological Archives*, 1(2), 175-179.
- Susarla, S., Medina, V. F., & McCutcheon, S. C. (2002). Phytoremediation: an ecological solution to organic chemical contamination. *Ecological Engineering*, 18(5), 647-658.
- Ullah, M. O., Haque, M., Urmi, K. F., Zulfiker, A. H. M., Anita, E. S., Begum, M., & Hamid, K. (2013). Anti-bacterial activity and brine shrimp lethality bioassay of methanolic extracts of fourteen different edible vegetables from Bangladesh. *Asian Pacific journal of tropical biomedicine*, 3(1), 1-7.
- Wang, K., Cai, J., Feng, J., & Xie, S. (2014a). Phytoremediation of phenol using *Polygonum orientale*, including optimized conditions. *Environmental monitoring and assessment*, 186(12), 8667-8681.
- Wang, C., Zheng, S. S., Wang, P. F., & Qian, J. (2014b). Effects of vegetations on the removal of contaminants in aquatic environments: a review. *Journal of Hydrodynamics*, 26(4), 497-511.
- Xia, C., Ma, X., Liu, S., & Fan, P. (2012). Studies on remediation on DDT-contaminated soil and dechlorination of DDT. *Procedia Environmental Sciences*, 16, 289-292.
- Zhou, M. L., Tang, Y. X., & Wu, Y. M. (2013). Plant hairy roots for remediation of aqueous pollutants. *Plant Molecular Biology Reporter*, 31(1), 1-8.

Robust Artificial Bee Colony in the Hopfield Network for 2-Satisfiability Problem

Mohd. Shareduwan Mohd. Kasihmuddin*, Mohd. Asyraf Mansor and Saratha Sathasivam

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

ABSTRACT

Swarm intelligence is a research area that models the population of swarm that is able to self-organise effectively. Honey bees that gather around their hive with a distinctive behaviour is another example of swarm intelligence. In fact, the artificial bee colony (ABC) algorithm is a swarm-based meta-heuristic algorithm introduced by Karaboga in order to optimise numerical problems. 2SAT can be treated as a constrained optimisation problem which represents any problem by using clauses containing 2 literals each. Most of the current researchers represent their problem by using 2SAT. Meanwhile, the Hopfield neural network incorporated with the ABC has been utilised to perform randomised 2SAT. Hence, the aim of this study is to investigate the performance of the solutions produced by HNN2SAT-ABC and compared it with the traditional HNN2SAT-ES. The comparison of both algorithms has been examined by using Microsoft Visual Studio 2013 C++ Express Software. The detailed comparison on the performance of the ABC and ES in performing 2SAT is discussed based on global minima ratio, hamming distance, CPU time and fitness landscape. The results obtained from the computer simulation depict the beneficial features of ABC compared to ES. Moreover, the findings have led to a significant implication on the choice of determining an alternative method to perform 2SAT.

Keywords: Artificial Bee Colony Algorithm, exhaustive search, hopfield network, satisfiability, logic programming, 2SAT

Article history:

Received: 04 April 2016

Accepted: 21 July 2016

E-mail addresses:

iwanmaidin@gmail.com (Mohd. Shareduwan Mohd. Kasihmuddin),

asyrafalvez@live.com (Mohd. Asyraf Mansor),

saratha@usm.my (Saratha Sathasivam)

*Corresponding Author

INTRODUCTION

Artificial Neural Networks (ANNs) have been successfully applied in solving infinite applications such as classification, function approximation, optimisation and associative memory. Nonetheless, the success of neural networks in constraint optimisation problem largely depends on their architecture and

solution searching techniques (Haykin, 1999). In some approaches, the integration of neural network, logic programming, satisfiability problem and neuro-searching techniques has been proven to minimise the complexity of the network. One of the earliest neural networks that resembles how human brain actually works is the Hopfield neural network. The Hopfield neural network, which was introduced by Hopfield and Tank (1985), is a simple recurrent network that can serve as an efficient associative memory and store definite memories with exceptional retrieval power (Rojas, 1999; Sathasivam et al., 2013). Moreover, it is a branch of the neural networks that has been applied in vast combinatorial problems such as Travelling Salesperson problem (TSP) and hard satisfiability problem (Haykin, 1992). For instance, logic programming can be treated as a problem in combinatorial optimisation perspective (Kowalski, 1979). Previously, logic programming has been implemented and assimilated in a neural network to search desired solutions (Hamadneh et al., 2013). In this paper, the advantages of the Hopfield network, logic programming and neuro-searching methods are combined to solve satisfiability problem.

Moreover, some neuro-searching methods, such as Exhaustive Search (ES) and meta-heuristic, can be implemented as a mechanism in solving satisfiability problems. The most widely used technique is the ES because it is a simple algorithm (Hooker, 2005). Conventionally, this particular technique considers all possible search spaces in order to verify clauses satisfaction for satisfiability problems. However, the ES can be applied only if the problem size or the number of clause is limited (Mark & Lee, 1992). Another limitation is that the ES typically consumes more time to complete the whole searching process (Tobias & Walter, 2004; Kaushik, 2012). Over the past decade, the growing interest in computational swarm intelligence has caused the emergence of several new optimisation algorithms. Recently, an algorithm based on the model of bee foraging behaviour was developed. Artificial bee colony (ABC) was first introduced by Karaboga (2005). Since ABC is simple in concept, easy for implementation and has fewer parameters (Pan et al., 2011), it has attracted the attention of many researchers to solve constraint optimisation.

In this paper, a meta-heuristic approach from the ABC is proposed to obtain the possible satisfied assignments within acceptable timescales. More specifically, the motivation of this paper is to apply the ABC as a searching technique incorporated with the Hopfield neural network in performing logic programming. A good searching technique during neural network simulation is vital to improve the convergence of the algorithm (Siddique et al., 2013).

This paper has been organised as follows. In Section 2, the fundamental theory of random 2-satisfiability (2SAT) problems is discussed. Moving on, Section 3 presents the neuro-searching methods employed in this research, including ES and ABC. Next, Section 4 provides a brief discussion of the neuro-logic, which explains the Hopfield neural network, content addressable memory (CAM) and logic programming in neuro symbolic integration. Meanwhile, the theory implementation of the networks is discussed in Section 5. Finally, Sections 6 and 7 enclose the experimental results and the conclusion of this exploration.

SATISFIABILITY (SAT) PROBLEM

Satisfiability or SAT is a significant problem in the computer science field. This problem revolves in determining if a truth assignment to variables that appears in a Boolean formula ϕ is satisfied (Kowalski, 1979). The Boolean formula is said to be satisfiable if an assignment of true and false values renders the entire expression as true. One way to solve SAT would be by trying out every possible truth of the assignment. For a problem of size n , there will be 2^n such assignments and l literals to set for each assignment (Sathasivam & Sagir, 2014), in which such an approach requires $O(l \cdot 2^n)$ operations (Gu, 1999). Hence, SAT is an NP-complete problem in general. For instance, the satisfiability problem concerns Boolean variables or expressions in conjunctive normal form (CNF). CNF comprises of conjunction of clauses, where the clauses are disjunctions of literal (Sathasivam et al., 2013). Meanwhile, literal is a variable or its negation.

For example:

$$(x_1 \vee x_2) \wedge (\neg x_2 \vee x_3 \vee \neg x_5) \wedge (\neg x_1 \vee x_4) \quad [1]$$

Here x_1, x_2, x_3, x_4 are Boolean variables to be assigned, \neg refers to negations (logical NOT), \vee means negations (logical OR), and \wedge means negations (logical AND). Thus, the formula above is satisfied when $x_1 = \text{true}, x_2 = \text{false}, x_3 = \text{false}, x_4 = \text{true}$, where it takes on the value of true.

2-Satisfiability (2SAT)

2SAT is the problem of deciding the satisfiability of sets of clauses with at most two literals per clause (2-CNF formulas). It is a special case of general Boolean satisfiability, which can involve constraints on two variables (Kowalski, 1979). Besides, the 2SAT paradigm can allow two choices for the value of each variable. Normally, 2SAT problem can be expressed as 2-CNF (2-Conjunctive Normal Form) or Krom formula (Fernandez, 2011). In contrast, randomised 2SAT problem is considered as an NP problem or a non-deterministic problem. The three components of 2SAT are summarised in the following:

1. A set of m variables, x_1, x_2, \dots, x_m
2. A set of literals. A literal is a variable or a negation of a variable.
3. A set of n distinct clauses: C_1, C_2, \dots, C_n . Each clause consists of only literals combined by just logical OR (\vee). Each clause must consist of 2 variables.

The Boolean values are $\{1, -1\}$. In fact, researchers have emphasised the True and False in the neural networks by 1 and -1. Due to this, the goal of the 2SAT problem is to determine if an assignment of truth values to variables does exist, which makes the following formula satisfiable.

$$P = \bigwedge_{i=1}^n C_i \quad [2]$$

Where \bigwedge is a logical AND connector, and P denotes the entire Boolean formula for 2SAT. C_i is a clausal form of DNF with 2 variables. Each clause in 2SAT has the following form:

$$C_i = \bigvee_{i=1}^n (x_i, y_i) \quad [3]$$

$x_i \in \{k_i, \neg k_i\}$ and $y_i \in \{r_i, \neg r_i\}$ $\neg k_i$ and $\neg r_i$ are negations of the literals.

NEURO-SEARCH TECHNIQUES

Exhaustive Search (ES)

Generally, ES is the simplest algorithm, but it can be computationally super expensive. In this algorithm, it exhaustively searches for the entire possible clause even though the search space gets larger. The main advantage of this algorithm is the guarantee to obtain a solution (satisfied clause) by taking into consideration the entire search space (Rojas, 1999). Furthermore, the modification of ES (Ata & Coban, 2015) might increase the efficiency of the algorithm through numerous assumptions and conditions. Moreover, the ES has been proven to consume more computation time or CPU time to hunt for the satisfied interpretation completely (Kaushik, 2012; Asrar & Aiman, 2015). In ES, the clause satisfaction is determined directly for the randomised 2SAT problem until a satisfying one is found (Tobias & Walter, 2004). Meanwhile, in this particular case, the satisfied clause has been sought during the training phase for 2SAT.

The complexity of the network has a positive correlation to a number of neurons. As the number of candidate solution is increased, the complexity of the searching technique will increase. In some cases, it may lead to combinatorial explosion (Gagneur & Klamt, 2004). Generally, for randomised 2SAT problem, there are potentially 2^n candidate solution and run-time complexity, which are equivalent to $O(2^n)$. Additionally, the satisfied assignments are obtained after performing trial-and-error processes. If the algorithm successfully finds an incorrect assignment, the algorithm will reset the whole search space. The correct assignment will be stored as a graded pattern as CAM. In this paper, this algorithm, together with the Hopfield neural network, logic programming, and satisfiability problem, has been implemented.

Artificial Bee Colony (ABC)

The ABC algorithm is a swarm-based meta-heuristic algorithm that was popularised by Karaboga (2005) in order to optimise numerical problem. It was inspired by the intelligent foraging behaviour of honey bees (Muthiah & Rajkumar, 2014). In fact, some papers published by Civicioglu and Besdok (2013) and Karboga and Basturk (2008) discovered that ABC had outperformed some meta-heuristic methods such as cuckoo-search and particle swarm optimisation. The model consists of three essential components, namely, employed and unemployed bees, as well as scout bees. The first two components, employed and unemployed bees, search for rich food sources. These two components are necessary for self-organising

and collective intelligence. Besides, recruitment of foragers to rich sources that results in positive feedback can help us to achieve our desired solution. If the solution obtained from the first two components does not meet the criteria set, the algorithm will spawn scout bees. Scout bees will try to find alternative food source (solution). The act of scout bees will reset the whole search algorithm. This strategy prevents the algorithm from having local maxima (non-improving) solution.

General Artificial Bee Colony (ABC) Algorithm

- 1) Initially, food sources are produced for all employed bees.
- 2) Each employed bee group goes to a food source and checks a neighbour source. They will evaluate its nectar amount and dance in the hive.
- 3) Each onlooker bees group will watch the dance of the employed bees and choose one of their sources, depending on the dances and then, goes to that source. After choosing a neighbour around them, it will evaluate its nectar amount. The best food source will be registered.
- 4) If the food source is not the desired food source, it will be abandoned. The abandoned food sources are determined and replaced with a new food source. The scout bee will reset the food source. The new set of food sources will be discovered by the scout bees.
- 5) Repeat steps 2 until 4. The best output will be recorded.

The difference between employed bees, onlooker bees and scout bees is that the employed bees share their food source information with the onlooker bees, while waiting in the hive (Karboga & Basturk, 2007). From the information retrieved from the employed bees, the onlooker bees would probably choose their food sources depending on this information. Moreover, these onlooker bees choose a food source depending on the probability values calculated by using the fitness values provided by the employed bees. Meanwhile, the unemployed bees that choose their food sources randomly are called scouts. These employed bees, whose solution cannot improve through a pre-determined number of trials specified by user, are called “limit”. These employed bees will be converted to bee scouts and their solutions are abandoned. The scout bees will search for a new set of solution randomly.

Binary Artificial Bee Colony (ABC) Algorithm

Binary ABC has been implemented by several researchers (Kashan et al., 2012) for binary optimisation. As for this study, ABC was utilised to hunt the fittest assignment or the highest satisfied clause given at any randomised 2SAT clause. Since only binary value (1 and -1) had been considered for this representation, the traditional ABC was improved to produce binary solutions. Therefore, the fittest assignment gave the maximum number of satisfied clauses, which depended on the number of satisfied clauses, and can be calculated as follows:

$$fit_i(x) = c_1(x) + c_2(x) + c_3(x) \dots + c_{total\ NC}(x) \quad [4]$$

Basically, the employed bees will find solution (assignments) in a search space. The journey to find the food source (solution) can be evaluated by using the following equation (Jia et al., 2014).

$$v_{ij} = x_{ij} \vee \left(\phi_{ij} \otimes (x_{ij} \wedge x_{kj}) \right) \tag{5}$$

where

ϕ_{ij} parameter where

$$\phi_{ij} = \begin{cases} 1, & rand(0,1) < 0.5 \\ -1, & rand(0,1) \geq 0.5 \end{cases}$$

\otimes is a 'XOR' operator

\wedge is an 'AND' operator

\vee is an 'OR' operator

Once all the employed bees have returned to their hive, they will dance. The information transfer occurs during the dance. Each employed bee has its own fitness. Fitness is evaluated based on the number of (satisfied) clauses. The dancing, on the other hand, is observed by the onlooker bees. The onlooker bees will choose the information based on the following probability of roulette wheel selection (Goldberg, & Deb, 1991).

$$p_i = \frac{fit_i}{\sum_{i=1}^{SN} fit_i} \tag{6}$$

Where $\sum_{i=1}^{SN} fit_i$ is the desired fitness, while SN is denoted by the group size of the bees.

The onlooker bees will find the solution by using equation (5). Thus, the best solution (desired fitness) is generated until the number of trial is equal to the limit. Consequently, if the solution from the onlooker bees cannot be further improved, the onlooker bees will change to scout bees. As a result, the scout bees will abandon the search space (Singh & Alok, 2009). Note that, if the algorithm finds a solution with desired fitness, the solution will exit the algorithm easily and print the best solution. The correct assignments will be stored into Hopfield's brain as CAM.

Binary Artificial Bee Colony (ABC) Pseudocode

Step 1

Initialise parameters, size N. employed bees group size SN (Number of clauses), maximum allowed generations g_{max} (10 Generations) and trial number limit. Initialise all bees X .

Step 2

for $g = 1$ to g_{\max} (Number of Generation)

Calculate the fitness for each bee X (one group of bee) and then evaluate them (we take the best two bee groups as x_i and x_k).

Step 3

{Employed bees phase}

for $i=1$ to SN (Group size)

Produce a new food source v_i using equation [5]

Check the fitness of v_i .

If v_i beats x_i ,

then replace x_i with v_i in next generation and $trial_i = 0$

else $trial_i = trial_i + 1$

end for

Calculate the probability values p_i by using the equation [6] below.

Step 4

{Onlooker bees phase}

$t = 0, i = 1$

while $t < SN$

if $random < p_i$,

then produce a new food source v_i by using the following equation [5].

If v_i beats x_i ,

then replace x_i with v_i in next generation and $trial_i = 0$,

else $trial_i = trial_i + 1$ $BL_{success,i} = BL$

end if

$t = t + 1$ (until $t = 100$)

end while

Step 5

{Scout bees phase}

if $\max(trial_i > limit)$ then

Reset x_i within the search space

end if

Record the best solution founded so far

end for

Output the final solution

NEURO-LOGIC IN HOPFIELD NEURAL NETWORK

Hopfield Model

The Hopfield model was popularised by John Hopfield in 1982 (Hopfield, & Tank, 1985), which was used to solve vast pattern and combinatorial optimisation problems. The model comprised of interconnected units called neurons to form a network (Sathasivam, 2010). Thus, it is easy to hybridise the Hopfield network with other algorithms. The units in Hopfield nets are called the binary threshold unit (Haykin, 1992), which can only take binary values such as 1 and -1. The promising definitions for the activation of unit i , a_i , are:

$$a_i = \begin{cases} 1 & \text{if } \sum_j W_{ij} S_j > \xi_i \\ -1 & \text{Otherwise} \end{cases} \quad [7]$$

where W_{ij} is the connection strength from unit j to i . S_j is the state of unit j and ξ_i is the threshold of unit i . Normally, the connection in the Hopfield net has no connection with itself, but $W_{ii} = 0$ and some of the connections are symmetric or bidirectional (Sathasivam et al., 2013). The network consists of N recognised neurons, each is described by an Ising spin variable. The neurons are basically bipolar. Furthermore, $S_i \in \{1, -1\}$ follows the dynamics $S_i \rightarrow \text{sgn}(h_i)$, where the local field is h_i . The connection model can be generalised to embrace higher order connection. This modifies the field to:

$$h_i = \sum_j W_{ij}^{(2)} S_j + J_i^{(1)} \quad [8]$$

The weight or the connection strength in the Hopfield network is always symmetrical. The updated rule is maintained as:

$$S_i(t+1) = \text{sgn}[h_i(t)] \quad [9]$$

These properties guarantee that the energy will decrease monotonically while following the activation system. The following equation represents the energy for the Hopfield network.

$$E = \dots - \frac{1}{2} \sum_i \sum_j W_{ij}^{(2)} S_i S_j - \sum_i W_i^{(1)} S_i \quad [10]$$

This energy function is significant because it establishes the degree of convergence of the network (Ionescu et al., 2010). Thus, energy value is vital in order to reach the global solutions.

Content Addressable Memory (CAM)

The Hopfield model is a standard model for CAM (Holland, 1975). It is one of the remarkable features of Hopfield, which was inspired by the way the biological brain works. In layman's term, CAM can be defined as common memories where the retrieved data, from a given

address of the memory location, are stored (Ionescu et al., 2010). In this paper, the satisfied randomised 2SAT assignment was stored in the CAM and retrieved after the training process. Basically, the CAM was used to the implementation of search algorithms such as ES and ABC.

Logic Programming in Hopfield Network

Strictly speaking, logic programming can be treated as a problem in combinatorial optimisation and it can be carried out on a Hopfield network (Sathasivam et al., 2013). The most celebrated logic programming model was the Wan Abdullah's logic paradigm, which implemented the Hopfield network with Horn clauses (Wan Abdullah, 1993). Therefore, Pinkas and Dechter (1995) emphasised on a bi-directional mapping between the propositional logic formula and the Hopfield network by integrating the energy minimisation scheme to attain global convergence. Hence, the logic programming was implemented in Hopfield for 2SAT clause with some local search algorithms, namely, ES and ABC.

Implementation of 2SATABC in Hopfield Neural Network (HNN-2SATABC)

- i. Translate all the random 2SAT clauses into Boolean algebra.
- ii. Recognise a neuron to each ground neuron. Initialise all connection strengths to zero.
- iii. Derive a cost function that is related with negation of all 2SAT clauses. For example,

$$X = \frac{1}{2}(1 + S_X) \text{ and } \bar{X} = \frac{1}{2}(1 - S_X). \quad S_X = 1 \text{ (True) and } S_X = -1 \text{ (False).}$$

Multiplication is represented conjunction and addition represents disjunction of clauses.

- iv. The cost function is compared with energy, E in order to obtain the values of the connection strengths or weights (Sathasivam & Wan Abdullah, 2008).
- v. Check 2SAT clauses satisfaction by using ABC algorithm. Satisfied assignments will be stored in CAM.
- vi. Randomise the states of the neurons. The network undergoes a series of network relaxation. Calculate the corresponding local field $h_i(t)$ of the state. If the final state is stable for 5 runs, it is considered as the final state.
- vii. Find the corresponding final energy, E , of the final state by using Lyapunov equation. Verify whether the final energy obtained is global minimum energy or local minima. Calculate the corresponding hamming distance. The time taken to complete the process is recorded. The fitness value is computed by using the Kauffman model (Imada, & Araki, 1997):

$$f = \frac{1}{t_0 p} \sum_{t=1}^{t_0} \sum_{v=1}^p m^v(t) \quad \text{whereby, } m^v(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^v S_i^v(t) \quad [11]$$

THEORY IMPLEMENTATION

The simulations were performed on Microsoft Visual Studio 2013 for Windows. First of all, random 2SAT clauses were generated. After that, the initial states for the neurons were initialised in the 2SAT clauses. The network evolved until it reached the final state. Once the programme had reached the final state, the neuron state was updated via equation (8). As soon as the network relaxed to an equilibrium state, the final state obtained for the relaxed neuron was tested to determine if it was in a stable state. Additionally, if the state had remained unchanged for five runs, a stable state would have been considered. According to Pinkas and Dechter (1995), permitting an ANN to evolve will eventually lead to a stable state where the energy function obtained would not change further. Subsequently, the corresponding final energy for the stable state was computed. If the difference between the final energy and the global minimum energy had been within the tolerance value, then the solution would be considered as a global solution. Both algorithms were repeated 100 times with 100 neuron combinations. The tolerance value for the final energy was 0.001. In fact, according to Sathasivam et al. (2013), 0.001 was selected because it gave a better performance than other values did, besides successfully dodging statistical errors. Moreover, global minima ratio, Hamming distance, fitness landscape value and CPU time obtained from both ES (HNN-2SATES) and ABC (HNN-2SATABC) had also been compared.

RESULTS AND DISCUSSION

Global Minima Ratio and Hamming Distance

In this study, global minima ratio is defined as the ratio between the global solutions over the number of total runs. Meanwhile, Hamming distance is demarcated as the closeness of bits between the training state and the global state (retrieved state) of the neurons upon relaxation process (Sathasivam, 2010).

Table 1

Global minima ratio and hamming distance for HNN- 2SATES and HNN-2SATABC

Number of Neurons (NN)	Global Minima Ratio		Hamming Distance	
	HNN-2SATES	HNN-2SATABC	HNN-2SATES	HNN-2SATABC
10	0.9972	0.9985	0.00887	0.002980
20	0.9915	0.9922	0.01899	0.017852
30	0.9859	0.9897	0.02902	0.028978
40	0.9634	0.9742	0.03876	0.029685
50	0.9537	0.9713	0.04780	0.030669
60	0.9411	0.9699	0.06580	0.037976
70	-	0.9672	-	0.049651
80	-	0.9615	-	0.086186
90	-	0.9598	-	0.086991
100	-	0.9534	-	0.099981

Table 1 delineates the obvious variation in the global minima ratio and the Hamming distance obtained from HNN-2SATABC and HNN-2SATES. Thus, one can observe clearly from Table 1 that the global minima ratios obtained for HNN-2SATABC are nearly 1 from NN=10 until NN=100 compared to those of HNN-2SATES. Moreover, when the number of global minima ratio approached 1, the network produced more global solutions. Hence, the solutions were obtained by implementing HNN-2SATABC, which always converged towards global minima even though the complexity increased. The food source (solution) found by employed bee can be improved further via onlooker bees. When the search space reached local maxima for randomised 2-SAT fitness, scout bees would reset the search space. Through this point of view, the network always effectively finds optimum solutions. Nevertheless, the problem with HNN-2SATES has been the nature of ES that deploys tedious training process in searching the correct neuron states. Hence, the updated rule for HNN-2SATES generated more abrupt energy surface for the program to obtain more local minima instead of global minima. In addition, the neurons have to jump enormous energy barrier to reach the global solutions.

On top of that, the HNN-2SATABC consistently performed better than HNN-2SATES in terms of Hamming distance. When the value of Hamming distance is close to zero, the distance between the stable states and the global states will be almost zero. Thus, the accuracy of the output will be almost 100%. In this case, the solutions were nearly optimal and stable. Besides, Table 1 depicts that HNN-2SATABC outperformed HNN-2SATES in the Hamming distance standpoint. Furthermore, HNN-2SATABC was able to recall the correct states that contributed to the lower hamming distance compared to HNN-2SATES. Contrariwise, the ES algorithm emphasised the trial-and-error process during the clause satisfaction process.

Furthermore, the proposed paradigm, HNN-2SATES, was able to sustain up to 100 neurons. Hence, the ability to sustain a huge number of neurons was due to the special ability of ABC algorithms that reduced the computation burden in hunting the correct states. Thus, this enhanced the capability and the ability to control the energy relaxation process although the network got larger. Besides, providing more relaxation time for the network helped the network to retrieve the state more effectively. Thus, less relaxation time generated spurious minima, which caused the retrieved solution to achieve local minima energy (Sathasivam, & Sagir, 2014). The spurious minima, however, had been discovered in HNN-2SATES because it could only sustain up to 60 neurons.

Fitness Landscape Value

Figure 1 portrays the differences detected in the fitness landscape values obtained for HNN-2SATABC and HNN-2SATES. As observed, the differences in fitness value is zero. Hence, one can conclude that HNN-2SATABC is highly unlikely to get trapped in the local minima and always hunts for global solutions. The smoothness in the figure illustrates the consistency of the landscape fitness values, supported by the Hamming distance values discussed earlier on, which is also zero. The ruggedness of the fitness landscape for HNN-2SATES indicates the existence of error in obtaining global solutions.

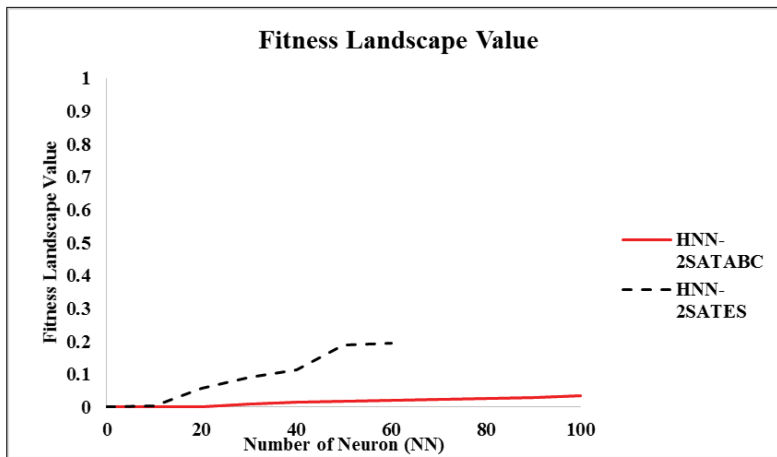


Figure 1. Fitness landscape value for HNN- 2SATES and HNN-2SATABC

Computation Time

The CPU time can be delineated as the time taken for the network to generate global solutions, including the training process. In this study, the analysis was limited up to 100 neurons.

Table 2
Computation time for HNN- 2SATES and HNN-2SATABC

Number of Neurons	CPU Time (in seconds)	
	HNN-2SATES	HNN-2SATABC
10	4	4.28
20	72	14.37
30	208	31.48
40	759	56.77
50	8036	90.16
60	75472	131.6
70	-	194.2
80	-	238
90	-	305.8
100	-	386

Table 2 depicts the CPU time for the proposed paradigm, HNN-2SATABC, in comparison to the conventional technique, HNN-2SATES. In terms of CPU time, the ES consumed has a rather relatively high CPU time (running time) compared to the ABC algorithm. Technically, the training process via ES frequently devoured more training time due to the trial-and-error procedure in achieving satisfied solutions. On the contrary, when the ABC algorithms were employed, the CPU time was faster due to the interaction between the onlooker and employed bees in hunting satisfied 2SAT assignments systematically. Additionally, this occurred as the

HNN-2SATABC incurred less computation burden during the training process, as compared to that of HNN-2SAT. For instance, the complexity of the network increased as the network began to grow massive. One could also note that the computational time increased when the number of neurons got higher. This trend had been consistent for HNN-2SATABC, even though the complexity of the network increased from NN=10 to NN=100, except for HNN-2SAT, which was only capable to sustain up to 60 neurons. This was because when the network became larger and more complex, the network devoured more computation time (Rojas, 1999). As a consequence of these arguments, extra time was needed to relax the global solution as the number of neurons increased.

CONCLUSION

Inspired by foraging intelligence behaviour displayed by honey bee swarm and engaging concept in artificial intelligence, a hybrid paradigm has been proposed; ABC algorithm incorporated with Hopfield neural network (HNN-2SATABC), in performing random 2SAT logic programming. Later, the proposed model was compared with a conventional technique; ES with the Hopfield neural network (HNN-2SAT). The work, as reported in this paper, reveals the tremendous differences in the performance of both the paradigms in terms of global minima ratio, Hamming distance, fitness landscape value and CPU time. Moreover, based on the experimental results, the proposed paradigm offers us a global minima ratio of approximately 1, faster computation time, as well as Hamming distance and fitness landscape values of approximately 0 compared to HNN-2SAT. Hence, the HNN-2SATABC has been unequivocally established to be more robust than the HNN-2SAT in certain aspects, which include better global minima ratio, lower Hamming distance, consistent fitness landscape value and faster CPU time, in performing random 2SAT logic programming. For future work, HNN-2SATABC can be extended to other satisfiability problems such as MAX-SAT, MIN-SAT and other SAT problems. This work can be utilised to solve traditional optimisation method such as travelling salesman and n-queen's problem.

REFERENCES

- Abdullah, W. A. T. W. (1993). The logic of neural networks. *Physics Letters A*, 176(3-4), 202-206.
- Aiman, U., & Asrar, N. (2015). Genetic Algorithm Based Solution to SAT-3 Problem. *Journal of Computer Sciences and Applications*, 3(2), 33-39.
- Ata, B., & Coban, R. (2015). Artificial Bee Colony Algorithm Based Linear Quadratic Optimal Controller Design for a Nonlinear Inverted Pendulum. *International Journal of Intelligent Systems and Applications in Engineering*, 3(1), 1-6.
- Brueggemann, T., & Kern, W. (2004). An improved deterministic local search algorithm for 3-SAT. *Theoretical Computer Science*, 329(1-3), 303-313.
- Civicioglu, P., & Besdok, E. (2013). A conceptual comparison of the Cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms. *Artificial Intelligence Review*, 39(4), 315-346.

- Fernandez, W. (2001). Random 2-SAT: Result and Problems. *Theoretical Computer Science*, 265(1), 131-146.
- Gagneur, J., & Klamt, S. (2004). Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5(1), 175-195.
- Goldberg, D. E., & Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. In G. J. E. Rawlins (Ed.), *Foundations of genetic algorithms*, (pp. 69-93). California: Morgan Kaufmann Publishers.
- Gu, J. (1999). The Multi-SAT algorithm. *Discrete Applied Mathematics*, 96, 111-126.
- Hamadne, N., Sathasivam, S., Tilahun, S. L., & Ong, H. C. (2013). Prey-Predator Algorithm as a New Optimization Technique Using in Radial Basis Function Neural Network. *Research Journal of Applied Sciences*, 8(7), 383-387.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. New York: Macmillan College Publishing.
- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Michigan: The University of Michigan Press.
- Hooker, J. N. (2005). Unifying local and exhaustive search. In L. Villasenor & A. I. Martinez (Eds.), *In Proceeding of ENC 2005- Sixth Mexican International Conference on Computer Science* (pp. 237-243). IEEE Press.
- Hopfield, J.J., & Tank, D.W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, 52(3), 141-152.
- Imada, A., & Araki, K. (1997, April). Application of an evolution strategy to the Hopfield model of associative memory. In *IEEE International Conference on Evolutionary Computation, 1997* (pp. 679-683). IEEE.
- Ionescu, L. M., Mazare, A. G., & Serban, G. (2010). VLSI Implementation of an associative addressable memory based on Hopfield network model. *IEEE Semiconductor Conference*, 2, 499-502.
- Jia, D., Duan, X., & Khan, M. K. (2014). Binary Artificial Bee Colony optimization using bitwise operation. *Computers & Industrial Engineering*, 76, 360-365.
- Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization* (Vol. 200). Technical report-tr06, Erciyes University, Engineering Faculty, Computer Engineering Department.
- Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of Global Optimization*, 39(3), 459-471.
- Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing*, 8(1), 687-697.
- Kashan, M. H., Nahavandi, N., & Kashan, A. H. (2012). DisABC: a new artificial bee colony algorithm for binary optimization. *Applied Soft Computing*, 12(1), 342-352.
- Kaushik, M. (2012). Comparative analysis of exhaustive search algorithm with ARPS algorithm for motion estimation. *International Journal of Applied Information Systems*, 1(6), 16-19.
- Kowalski, R.A. (1979). *Logic for Problem Solving*. New York: Elsevier Science Publishing.

- Mark, W. G., & Lee C. G. (1992). Routing in random multistage interconnections networks: Comparing exhaustive search, greedy, and neural network approaches. *International Journal of Neural System*, 2(3), 125-142.
- Muthiah, A., & Rajkumar, R. (2014). A Comparison of Artificial Bee Colony algorithm and Genetic Algorithm to Minimize the Makespan for Job Shop Scheduling. *Procedia Engineering*, 97, 1745-1754.
- Pan, Q. K., Tasgetiren, M. F., Suganthan, P. N., & Chua, T. J. (2011). A discrete artificial bee colony algorithm for the lot-streaming flow shop scheduling problem. *Information Sciences*, 181(12), 2455-2468.
- Rojas, R. (1999). *Neural Networks: A Systematic Introduction*. Berlin: Springer.
- Sathasivam, S. (2010). Upgrading Logic Programming in Hopfield Network. *Sains Malaysiana*, 39(1), 115-118.
- Sathasivam, S., Ng, P. F., & Hamadneh, N. (2013). Developing agent based modelling for reverse analysis method. *Journal of Applied Sciences, Engineering and Technology*, 6(22), 4281-4288.
- Sathasivam, S., & Sagir, A. M. (2014). An Overview of Hopfield Network and Boltzmann Machine. *International Journal of Computational and Electronics Aspects in Engineering*, 1(1), 20-26.
- Siddique, N., & Adeli, H. (2013). *Computational Intelligence Synergies of Fuzzy Logic, Neural Network and Evolutionary Computing*. United Kingdom: John Wiley and Sons.
- Singh, A. (2009). An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem. *Applied Soft Computing*, 9(2), 625-631.



Estimating Lung Cancer Deaths in Thailand Based on Verbal Autopsy Study in 2005

Nattakit Pipatjaturon^{1,2}, Phattrawan Tongkumchum^{2*} and Attachai Ueranantasun²

¹The office of Diseases Prevention and Control 2nd Phitsanulok, Phitsanulok 65000, Thailand

²Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, 94000, Thailand

ABSTRACT

Information on the causes of death obtained from death certificates in Thailand is incomplete and inaccurate. Therefore, mortality statistics from death registrations (DR) remains unreliable. Accurate mortality statistics is essential for national policies on intervention and care and resource allocation. Verbal Autopsy (VA) is a more reliable source for cause of deaths than the DR. In this study, the classification of lung cancer deaths in Thailand from 1996 to 2009 was investigated based on a logistic regression model of lung cancer deaths with demographic and medical factors from the 2005 VA data. The estimated proportions of lung cancer deaths from the model were applied to the DR data. The goodness of fit of the model was assessed using the ROC curve. The resulting estimates of lung cancer deaths were higher than those reported with inflation factors 1.54 for males and 1.44 for females. Meanwhile, misclassified cases were reported mainly as other cancer types. There is no evidence of regional variation for lung cancer. The methods enable health professionals to estimate specific cause of deaths in countries where low quality of causes of death in the DR database and reliable data such as the VA data is available. The findings provide useful information on death statistics for policy interventions related to lung cancer prevention and treatment.

Keywords: Adjusted percentage, lung cancer deaths, logistic regression model, ROC

Article history:

Received: 06 April 2016

Accepted: 09 August 2016

E-mail addresses:

nattakit@hotmail.com (Nattakit Pipatjaturon),
phattrawan.t@psu.ac.th; phattrawan@gmail.com
(Phattrawan Tongkumchum),
attachai@gmail.com (Attachai Ueranantasun)

*Corresponding Author

INTRODUCTION

An accurate statistics of causes of death is essential for monitoring the health of a nation and identifying priorities. Data on the causes of death obtained from death registration (DR) in Thailand are of low quality (Mathers et al., 2005) because 35-40% of deaths are ill-defined (Patarachachai et al., 2010; Rao et al.,

2010). Extensive misclassification of causes of death (Tangcharoensathien et al., 2006) makes it necessary for mortality studies in Thailand to estimate the number of deaths using other data sources.

The VA in Thailand was conducted by Setting Priorities using Information on Cost-Effectiveness analysis (SPICE) project in 2005 to verify registered causes of death. This was the first national application of this WHO methodology to Thailand. Mortality estimates derived from making adjustments to the DR data in 2005 based on the VA using the simple cross-referencing method have been published (Porapakkham et al., 2010; Rao et al., 2010; Pattaraarchachai et al., 2010; Polprasert et al., 2010). However, this simple cross-referencing method ignored the effects of gender-age groups and location of the decease. That could give incorrect estimates due to confounding. This study offered an alternative approach based on statistical methods applied to a large-scale VA study focusing on lung cancer death.

The reasons for misclassification of the causes of death include a lack of properly trained physician to identify the causes of death for chain of illnesses and a lack of medical knowledge by head of the village (Kijsanayotin et al., 2013). Extensive misclassification of causes of death (Tangcharoensathien et al., 2006) makes it necessary for mortality studies in Thailand to estimate the valid number of deaths for improved DR database and thus vital statistics system in Thailand.

VA is a research method initiated by World Health Organization (WHO) to determine probable causes of death in case that there is no medical record or formal medical attention given. When DR cause of death is misclassified, the VA survey can be used to determine individuals' cause of death.

In 2005, the causes of death in Thailand were re-identified and reviewed using a VA questionnaire and a survey conducted by a physician with a training certificate in specifying causes of death based on International Classification of Diseases (ICD) (Polprasert et al., 2010). This was the first national application of this WHO methodology to find a solution for the low quality causes of death in Thailand.

The mortality estimations derived from making adjustments to the DR data in 2005 based on the VA have been published (Porapakkham et al., 2010; Rao et al., 2010; Pattaraarchachai et al., 2010; Polprasert et al., 2010). In order to reduce costs from conducting VA study for the whole country, an analysis of the VA data using appropriate statistical methods is an alternative approach to a large-scale VA survey such as in the case of HIV (Chutinantakul et al., 2014).

In Thailand, lung cancer contributed to 3.7% of all deaths for males in 2005, whereas it was 3.3% in 1999 (Porapakkham et al., 2010). The rising number of lung cancer deaths and rates for both sexes have been observed (Kamnerdsupaphon et al., 2008). The lung cancer incidence rates among Thai women exceeded those of women from many European countries such as Germany and Finland (Jemal et al., 2010).

Thus, this study aimed to estimate number of lung cancer deaths obtained from the DR during 1996-2009 using the VA data from 2005 with a statistical model of lung cancer deaths taking into account demographic and medical factors. Thus, after correction for the misclassified lung cancer deaths, a more accurate estimate of the proportion (or percentage or number) of deaths due to lung cancer could be obtained.

MATERIALS AND METHODS

Data source and management

This study used secondary data from a 2005 VA survey, which assessed the causes of death based on a sample of 9,644 cases (3,316 in-hospital deaths and 6,328 outside-hospital deaths) from 28 districts in nine provinces (Rao et al., 2010). The nine provinces selected were Bangkok and two provinces from each of the four regions in Thailand. The selected provinces were those whose numbers of reported deaths were above (one province) and below (one province) the median. Similarly, twenty-eight districts were selected from the provinces. Approximately 50% of the death certificates were selected from all the villages and urban areas within the 28 selected districts using simple random sampling.

Since no lung cancer deaths occurred in those aged below five years, the study sample was reduced to 9,495 cases aged five years and older (3,212 in-hospital deaths and 6,283 outside-hospital deaths). Data obtained from each case were the province, gender, age, location of death (in or outside hospital), DR-reported International Statistical Classification of Diseases (ICD-10) code reported on the death certificate and VA-assessed ICD-10 code.

Data Analysis

We analysed the VA data in this study using the chapter-block classification for ICD-10 codes based on mortality tabulation (World Health Organization, 2004), creating 21 major cause groups of deaths. The groups had to be large enough for statistical analysis. The 21 groups are described elsewhere (Chutinantakul et al., 2014; Waeto et al., 2014).

The outcomes of interest were the VA-assessed ICD-10 codes for lung cancer deaths (C30-C39) or others. The determinants were province, gender, age, location of death and DR-reported ICD-10 code. The VA-assessed ICD-10 codes and the DR-reported ICD-10 codes were cross tabulated to give five cause groups (namely, lung cancer, ill-defined, other cancer, respiratory disease and other), where lung cancer deaths are often misreported. The location of death and DR-reported ICD-10 codes were categorised into 10 groups: 5 DR-reported ICD-10 code groups each for the two locations (in and outside the hospitals). Gender and age were classified into 7 groups by gender: ages 5-29, 30-39, 40-49, 50-59, 60-69, 70-79 and 80+ years for each sex. Nine provinces (namely, Bangkok, Nakhon Nayok, Suphan Buri, Ubon Ratchathani, Loei, Phayao, Chiang Rai, Chumphon, and Songkhla) were included in the VA study.

Logistic Regression Model

We estimated the logit of the probability that a person died from lung cancer as a linear function of the determinant factors using logistic regression (McNeil, 1996; Venables & Ripley, 2002; Hosmer & Lemshow, 2004). The simple model is formulated as,

$$\log \left[\frac{p_i}{1 - p_i} \right] = \mu + \alpha_i \quad [1]$$

Where, p_i is the probability of death due to lung cancer, μ is a constant, and α_i is the parameter of DR cause location i . The simple model was compared with the full model (2), which included an additive linear function of further determinant factors. The full model is formulated as,

$$\log\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = \mu + \alpha_i + \beta_j + \gamma_k \quad [2]$$

Where, p_{ijk} is the probability of death due to lung cancer, μ is a constant, and α_i , β_j , and γ_k are parameters specifying DR cause location i , gender-age group j , and province k , respectively. This equation may be inverted to give an expression for the probability p_{ijk} as,

$$p_{ijk} = 1 / (1 + \exp(-(\mu + \alpha_i + \beta_j + \gamma_k))) \quad [3]$$

We fitted logistic regression model using sum contrasts (Venables & Ripley, 2002; Tongkumchum & McNeil, 2009; Kongchouy & Sampantarak, 2010; Sampantarak et al., 2011) instead of conventional treatment contrast, where the first level is left out from the model to be the reference. This model allows us to compute the 95% confidence intervals of lung cancer deaths for each of the covariate levels in the VA data.

Goodness of fit of the model

We used the Receiver Operating Characteristic (ROC) curve (Chongsuvivatwong, 2007) to show how well the simple and full model predicts a binary outcome. It plots sensitivity (proportion of positive outcomes correctly predicted by the model) against the false positive rate (proportion of all outcomes incorrectly predicted). Sensitivity and specificity of the model is a cut-off point in the curve, where the predicted number of lung cancer death is in agreement with the observed value in the VA data. Area under the curve (AUC) represents model accuracy (Sarkar & Midi, 2010).

Spatial triangulation method

The full model gave 10 coefficients for DR-cause location, 14 coefficients for gender-age group and 9 coefficients for province. The province coefficients were used to interpolate coefficients for remaining 67 provinces outside the VA study using a spatial triangulation method based on the latitude and longitude of their central point. The spatial triangulation method is described elsewhere (Chutinantakul et al., 2014).

Extension to DR data

Coefficients for province, gender-age group and DR cause-location were applied to all deaths in the DR data for each year. Assuming the models were correct for 1996-2009, the VA-estimated lung cancer deaths from 1996 to 2009 were thus obtained. Graphical displays and statistical analyses were performed using the R programme version 3.0.1 (R Core Team, 2013).

RESULTS

Of the 9,495 deaths, the VA-assessment gave 320 lung cancer deaths (117 in-hospital deaths and 203 outside-hospital deaths). Only 164 lung cancer deaths were correctly DR-reported. The rest were reported as ill-defined (89), other cancer (32), respiratory (17) and others (18).

The DR-cause-location factor was found to be highly statistically significant in the simple model. Table 1 shows all p-values from the full model. The DR-cause location and gender-age-group factors were highly statistically significant, but there was no significant evidence of a province effect. Although province was not significant, it was retained in the model as a basis for estimating lung cancer deaths for every province in the country.

Table 1
P-values of the estimated coefficients

Factor	Deviance reduction	df	p-value
DR cause-location	1005.23	9	<0.0000001
gender-age group	61.60	13	<0.0000001
province	10.81	8	0.2124
error	468.98	903	

Figure 1 shows ROC curves for both the simple and the full models. The cut-off point in the ROC curve gives the predicted number of lung cancer deaths (319) agreement of the observed value in the VA data set (320). The red lines, drawn from the cut-off point to the x-axis and y-axis, show the model sensitivity and specificity (1-false positive rate). The full model gives 55.3% sensitivity, 98.5% specificity and AUC 0.80, whereas the simple model gives 51.2% sensitivity, 99.4% specificity and AUC of 0.70.

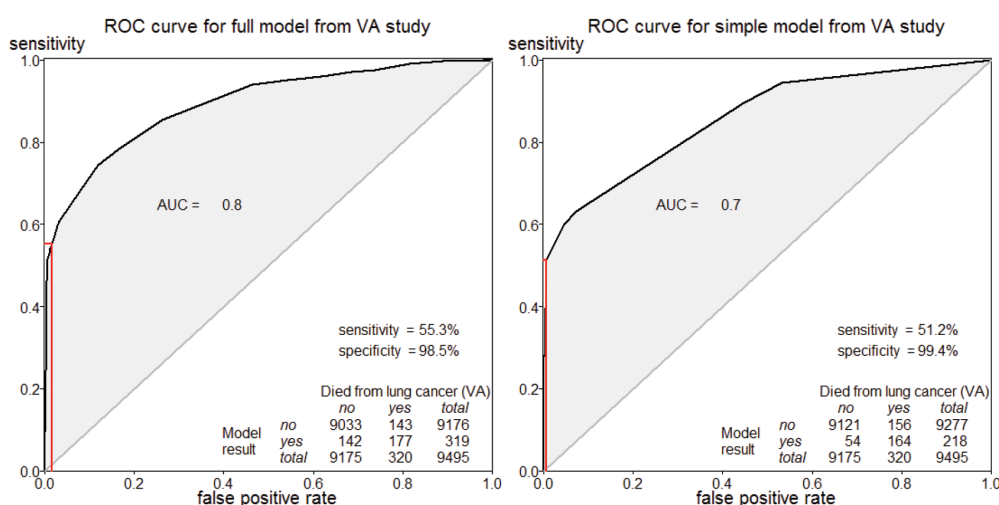


Figure 1. ROC curve for the full fitted model and simple fitted model from the VA study

Adjusted Percentages of Lung Cancer Deaths

The results derived for the model are presented as a graph of adjusted percentages and their corresponding 95% confidence intervals. Figure 2 shows the crude percentages of lung cancer deaths superimposed with adjusted percentages and their corresponding 95% confidence intervals. The horizontal red line is the average percentage of lung cancer deaths (3.4%). In order to distinguish the bar chart and 95% confidence interval, a non-linear vertical axis scale was used.

There is no evidence of province effects. The 95% confidence intervals above average were found in the age groups 60-79 for males. These age groups were more likely to have high levels of under-reporting. Meanwhile, the 95% confidence intervals for lung cancer and other cancer (outside-hospital) are above average. Other cancer outside hospital is the group in which lung cancer deaths were often misclassified.

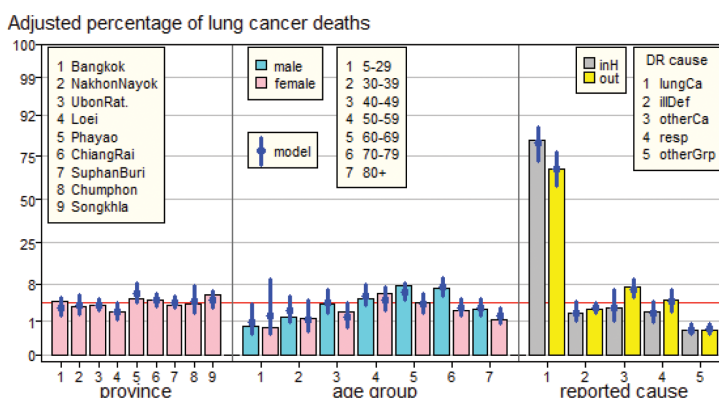


Figure 2. Adjusted percentages of lung cancer death by province, gender-age group and DR

Figure 3 shows the DR estimate of lung cancer deaths by gender-age group in 2005. The numbers of lung cancer deaths from DR reports were 5,887 and 2,549 cases for males and females, respectively. The simple model estimated the number of lung cancer deaths of 7,549 for males and 4,796 for females. The full model estimated the number of lung cancer deaths to be 8,503 in males and 3,433 in females, respectively. These were 44.4% and 34.7% higher than the corresponding DR reports.

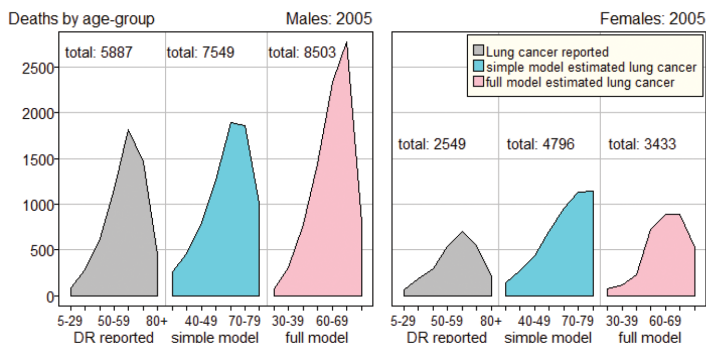


Figure 3. DR reports of lung cancer deaths and estimates from the simple and full models in 2005 by age groups

Figure 4 shows DR reported, the simple model estimated and the full model estimated the numbers of lung cancer deaths by age group in the years from 1996 to 2009. Apart from the drop in 1997-1998, when data are known to be incomplete in the DR database and a correction for temporally lost data from 2004 to 2005, the curves based on the statistical models are quite smooth and thus provide a credible basis for forecasting.

The numbers of lung cancer deaths rose rapidly with year, especially in males. Lung cancer deaths at ages 40+ years tended to increase in both sexes over the 14-year period, whereas deaths at ages 5-39 years tended to decrease. The total numbers of lung cancer deaths reported for 14 years were 64,819 in males and 28,491 in females. The estimated total numbers of lung cancer deaths from the simple model were 89,877 in males and 58,152 in females, whereas the estimates from the full model were 99,671 in males and 40,980 in females. The resulting estimates of lung cancer deaths from the full model were higher than those reported with inflation factors 1.54 for males and 1.44 for females.

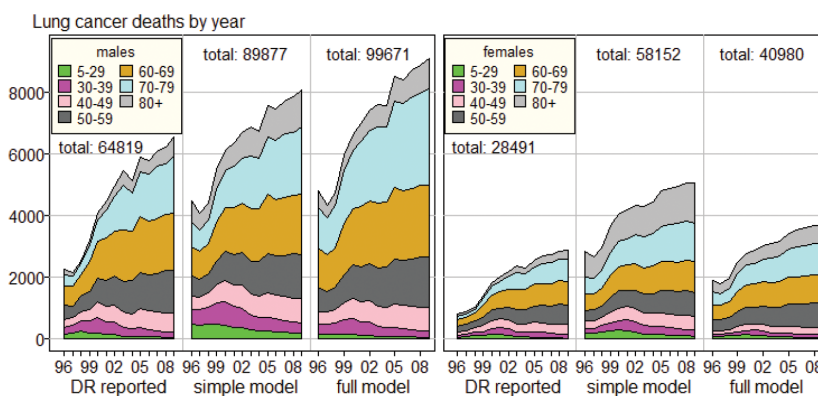


Figure 4. DR reports of lung cancer deaths and estimates from the simple and full models of lung cancer deaths by age groups and years

DISCUSSION

This study has shown that a logistic regression model of lung cancer deaths, with gender-age group, DR cause location and province from the VA data, can be used to adjust the number of deaths in the DR database. However, the accurate cause of death for individuals is uncertain, particularly for causes being reported as ill-defined or unknown cause. Meanwhile, goodness of fit of the model, as assessed by the ROC curve, indicates that the model adequately separates lung cancer deaths from other causes.

The finding from this study is that the medical and demographic factors in the model are highly statistically significant, but there is no evidence of any regional effect. The adjusted percentages of lung cancer deaths were high among the elderly males. Most lung cancer deaths were correctly reported. Misreported cases were mostly observed for deaths outside hospitals, and they were often reported as other cancer groups (C, D00-D48). The estimated numbers of lung cancer deaths from the models were higher than those being reported.

Logistic regression is commonly used in health studies. There are advantages in using the logistic regression model. The method gives confidence intervals for percentages of lung cancer deaths for levels of each risk factor adjusted for other risk factors using methods developed by Tongkumchum and McNeil (2009) and Kongchouy and Sampantarak (2010). These confidence intervals, when compared with bar charts of sample percentages, provide evidence of confounding bias. Moreover, the model can be extended to the larger target population comprising all deaths in Thailand for longer periods of time. Additionally, it can also be used to forecast lung cancer deaths and other specific causes.

However, the method also has some limitations. First, bias may have arisen in the sampling design. The VA study used a clustered sample design, but this sample did not include many subjects from rural places, and none at all from the many Muslim majority districts. Moreover, our model assumes that the patterns of misreporting of deaths in 1996-2009 are the same as in 2005 when the VA study was undertaken. This assumption is questionable, particularly for the years before 2005 when reporting practices were distorted by the HIV epidemic.

Our findings of high lung cancer deaths in elderly males are not surprising. It is well known that lung cancer is common among the elderly, and it more pronounced in males. A previous study in Thailand reported that lung cancer was common in patients aged 50 years or more (Deesomchok et al., 2005).

Although no evidence of regional effect was found in this study, a study on cancer control in Thailand using cancer registration data found high incidence rates of lung cancer in the northern region (Vatanasapt et al., 2002). Geographical variation on lung cancer deaths in 2000 have been observed with high rates in Bangkok (Faramnuayphol et al., 2008). This inconsistency is difficult to explain and there are not many studies on lung cancer deaths in Thailand. The findings on having no evidence of regional effects reported in this study will be useful for research in lung cancer mortality meta analyses.

This study found high percentages of lung cancer deaths, especially deaths in hospitals that were correctly reported and some misclassifications due to other cancers. This finding agrees with a previous study, where lung cancer deaths were observed to not contributing significantly to ill-defined cancer coding (Porapakkham et al., 2010).

CONCLUSION

This method enables public health researchers to estimate the percentage of specific causes of death in countries where there is low quality for recorded causes of deaths, but reliable sample data such as a VA study are available.

ACKNOWLEDGEMENTS

We are grateful to Prof. Don McNeil for his guidance, support and assistance. We are also thankful to Dr. Kanitta Bundhamcharoen from the Bureau of Policy and Strategy, Ministry of Public Health, Thailand, for providing us the data. Finally, we thank to the Graduate School of the Prince of Songkla University for the supported scholarship for Nattakit Pipatjaturon.

REFERENCES

- Chutinantakul, A., Tongkumchum, P., Bundhamcharoen, K., & Chongsuvivatwong, V. (2014). Correcting and estimating HIV mortality in Thailand based on 2005 verbal autopsy data focusing on demographic factors, 1996-2009. *Population Health Metrics*, 12(1), 25-32. <http://dx.doi.org/10.1186/s12963-014-0025-x>.
- Chongsuvivatwong, V. (2007). *Graphs, Tables and Equations for Health Research*. Bangkok: Chulalongkorn University Press.
- Core R Team. (2012). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org> [Cited 2013-05-16].
- Deesomchok, A., Dechayonbancha, N., & Thongprasert, S. (2005). Lung cancer in Maharaj Nakorn Chiang Mai Hospital: Comparison of the clinical manifestations between the young and old age groups. *Journal of the Medical Association of Thailand*, 88(9), 1236-1241.
- Faramnuayphol, P., Chongsuvivatwong, V., & Panarunothai, S. (2008). Geographical variation of mortality in Thailand. *Journal of the Medical Association of Thailand*, 91(9), 1455-1460.
- Hosmer, D. W., & Lemeshow, S. (2004). *Applied logistic regression*. New York: John Wiley & Sons.
- Jemal, A., Center, M. M., DeSantis, C., & Ward, E. M. (2010). Global Patterns of Cancer Incidence and Mortality Rates and Trends. *Cancer Epidemiology Biomarkers & Prevention*, 19(8), OF1-15.
- Kamnerdsupahon, P., Srisukho, S., Sumitsawan, Y., Lorvidhaya, V., & Sukthomya, V. (2008). Cancer in Northern Thailand. *Biomedical Imaging and Intervention Journal*, 4(3), e46-e52. Retrieved from <http://dx.doi.org/10.2349/bij.4.3.e46>.
- Kijisanayotin, B., Ingun, P., & Sumpattanon, K. (2013). *Rapid Assessment of National Civil Registration and Vital Statistics Systems: A case study of Thailand*. Nonthaburi, Thailand: Thai Health Information Standards Development Center, Health Systems Research Institute.
- Kongchouy, N., & Sampantarak, U. (2010). Confidence intervals for adjusted proportions using logistic regression. *Modern Applied Science*, 4(6), 2-6.
- Mathers, C. D., Fat, D. M., Inoue, M., Rao, C., & Lopez, A. D. (2005). Counting the dead and what they died from: and assessment of the global status of cause of death data. *Bulletin of the World Health Organization*, 83(3), 171-177.
- McNeil, D. (1996). *Epidemiological Research Methods*. New York: John Wiley & Sons.
- Pattaraarchachai, J., Rao, C., Polprasert, W., Porapakkham, Y., Pao-in, W., Singwerathum, N., & Lopez, A. D. (2010). Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification. *Population Health Metrics*, 8(1), 1-12. <http://dx.doi.org/10.1186/1478-7954-8-12>.
- Polprasert, W., Rao, C., Adair, T., Pattaraarchachai, J., Porapakkham, Y., & Lopez, A. D. (2010). Cause-of-death ascertainment for deaths that occur outside hospitals in Thailand: application of verbal autopsy methods. *Population Health Metrics*, 8(1), 13-27. <http://dx.doi.org/10.1186/1478-7954-8-13>.
- Porapakkham, Y., Rao, C., Pattaraarchachai, J., Polprasert, W., Vos, T., Adair, T., & Lopez, A. D. (2010). Estimated causes of death in Thailand, 2005: implications for health policy. *Population Health Metrics*, 8(1), 14-24. Retrieved from <http://dx.doi.org/10.1186/1478-7954-8-13>.
- Rao, C., Porapakkham, Y., Pattaraarchachai, J., Polprasert, W., Swampunyalert, N., & Lopez, A. D. (2010). Verifying causes of death in Thailand: rationale and methods for empirical investigation. *Population Health Metrics*, 8(1), 11-23. Retrieved from <http://dx.doi.org/10.1186/1478-7954-8-11>.

- Sampantarak, U., Kongchouy, N., & Kuning, M. (2011). Democratic confidence intervals for adjusted means and incidence rates. *American International Journal of Contemporary Research*, 1(3), 38-43.
- Sarkar, S. K., & Midi, H. (2010). Importance of Assessing the Model Adequacy of Binary Logistic Regression. *Journal of Applied Sciences*, 10(6), 479-486.
- Tangcharoensathien, V., Faramnuayphol, P., Teukul, W., Bundhamcharoen, K., & Wibulpholprasert, S. (2006). A critical assessment of mortality statistics in Thailand: potential for improvements. *Bulletin of the World Health Organization*, 84(3), 233-239.
- Tongkumchum, P., & McNeil, D. (2009). Confidence interval using contrasts for regression model. *Songklanakar Journal of Science and Technology*, 31(2), 151-156.
- Vatanasapt, V., Sriamporn, S., & Vatanasapt, P. (2002). Cancer control in Thailand. *Japanese Journal of Clinical Oncology*, 32(suppl 1), S82-S91.
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th Ed.). New York: Springer-Verlag.
- Waeto, S., Pipatjaturon, N., Tongkumchum, P., Choonpradub, C., Saelim, R., & Makaje, N. (2014). Estimating liver cancer deaths in Thailand based on verbal autopsy study. *Journal of Research in Health Sciences*, 14(1), 18-22.
- WHO. (2004). *ICD-10 International Statistical Classification of Diseases and Related Health Problems*. Geneva: World Health Organization.

Rainfall Trends in the Niger-South Basin, Nigeria, 1948-2008

Oloruntade, A. J.^{1,3*}, Mohammad, T. A.² and Aimrun, W.¹

¹Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

²Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia 43400 UPM, Serdang, Selangor, Malaysia

³Department of Agricultural and Bio-Environmental Engineering Technology, Rufus Giwa Polytechnic, Owo, Ondo State, Nigeria

ABSTRACT

Understanding rainfall trend can be a first step in the planning and management of water resources especially at the basin scale. In this study, standard tests are used to examine rainfall trends based on monthly, seasonal and mean annual series at the Niger-South Basin, Nigeria, between 1948 and 2008. Rainfall variability index showed that the decade 2000s was the driest (-2.1), while 1950s was the wettest (+0.8), with the decade 1980s being the driest in the second half of the last century, whereas the year 1983 was the driest throughout the series. Over the entire basin, rainfall variability was generally low, but higher intra-monthly than inter-annually. Annual rainfall was dominated by August, contributing about 15%, while December contributed the least (0.7%). On a seasonal scale, July-August-September (JJA) contributed over 40% of the annual rainfall, while rainfall was lowest during December-January-February (DJF) (4.5%). The entire basin displayed negative trends but only 15% indicated significant changes ($\alpha < 0.1$), while the magnitudes of change varied between -3.75 and -0.25 mm/yr. Similarly, only JJA exhibited insignificant upward trend, while the rest showed negative trends. About eight months of the year showed reducing trends, but only January trend was significant. Annual downward trend was generally observed in the series. The trend during 1948–1977 was negative, but it was positive for the 1978–2008 period. Hence, water resources management planning may require construction of water storage facilities to reduce summer flooding and prevent possible future water scarcity in the basin.

Keywords: Rainfall, trend analysis, statistical tests, variability index, Niger-South Basin

Article history:

Received: 21 April 2016

Accepted: 02 August 2016

E-mail addresses:

johnntades1@yahoo.com (Oloruntade, A. J.),

thamer@upm.edu.my (Mohammad, T. A.),

aimrun@upm.edu.my (Aimrun, W.)

*Corresponding Author

INTRODUCTION

Adequate understanding of the behaviour of climate variables is important in climate research and agricultural water resources management given that climate change is

now one of the most debated environmental issues all over the world. This is especially so at basins where the phenomenon of climate changes has the potential to confound water resources planners and make management of resources a herculean task. However, despite the diversity of the field of climate change, change in rainfall pattern has received serious and systematic attention as an area within this field, owing to its impacts on both the availability of freshwater and food production (Dore, 2005). According to Oguntunde et al. (2011), detecting trends in rainfall time series provides vital information for the understanding of climate. Hence, detection of trends in rainfall becomes an important study capable of giving a clear direction of change in future climate in basins, apart from providing crucial information useful in planning and designing regional water resources management (Karpouzou et al., 2010).

Many studies have been carried out in different parts of the world to assess the impacts of climate change on hydro-meteorological processes. However, trend analyses provide the first step towards blaming such changes on factors such as variability, greenhouse gases and changes in the municipal environment (Blake et al., 2011). Meanwhile, studies of time series data across the globe have shown either decreasing or increasing rainfall trend (Mondal et al., 2012). In Europe, for example, Brunetti et al. (2006) studied the trend in Italian long precipitation records and reported a strong decreasing trend, with a decrease in rainfall of about 135 mm in the southern regions over the last 50 years. However, a study by Janssen (2013) over the contiguous United States detected an increasing trend in precipitation with variation among seven sub-regions in the area. Zhang et al. (2011) also observed significantly increasing precipitation trends over Canada since 1950 in the majority of stations used for their study; trend toward increasing precipitation was in addition to the rises in the amount of extreme daily precipitation during the growing season.

Meanwhile, mixed results have also been obtained by scientists in China and other parts of Asia. Zhao et al. (2008) studied monotonic trends and abrupt changes for major climatic variables in the headwater catchment of Yellow River basin, China, and observed an insignificant trend in annual precipitation, though a dry tendency was also detected. In a related study in China by Yang et al. (2012), a not so obvious downward trend in annual precipitation, with large decreases in summer, was observed. However, there was a strong continuity in the annual mean precipitation series in the basin, indicating a similar trend between the future and the past. In India, Kumar et al. (2010) studied the trend in rainfall data for the period 1871-2005 and found insignificant trend for monthly, seasonal and annual series over the entire landscape. While both monsoon and annual rainfall reduced, there was increased pre-monsoon, post-monsoon and winter rainfall over the years, with the highest increase in the pre-monsoon season. Meanwhile, there was a reduction in rainfall during the monsoon months of June, July and September, while August revealed an upward trend.

Studies which dealt with spatial and temporal rainfall trend and variability are not many within Africa, especially in the West Africa sub-region. However, in the context of Nigeria, diverse results have been documented by many researchers (see for instance, Ikhile, 2007; Abaje et al., 2010; Salami et al., 2014), depending on the geographical zone and the temporal

scale covered. Many of the studies have also reported extreme weather events like floods and droughts, attributable to the current climate change. Nonetheless, Oguntunde et al. (2011) found a decrease in rainfall that was significant over the last three decades of the previous century, with a spatial distribution that was highly dependent on latitude. Granted that the foregoing study dealt broadly on the trend of rainfall over the country, but given the geographical spread of Nigeria, analysis of rainfall trend within a specific geographical zone can be a valuable effort. Again, the study was limited to the last century even as IPCC (2007) has predicted increased drying for the 21st century. In addition, recent flood events within the present study area have attracted the attention of both local and foreign experts, considering the attendant human and material losses. Thus, to confirm if this is due to climate change or human activities and for the purpose of planning, further investigation is needed. Therefore, the aim of the study is to detect the trend of rainfall in the Niger-South Basin, Nigeria.

THE STUDY AREA

The Niger-South Basin (NSB), located in Nigeria between Latitudes 5.8° – 8.0° N and Longitudes 6.0° – 7.8° E, has a total area of about 26,324 km². It is the last part of the Niger River Basin (NRB) before its final exit to the Atlantic Ocean and falls within the Nigerian Hydrological Area V, excluding the Niger Delta (Figure 1). The basin covers three Nigerian agro-ecological zones, namely, Tall grass Savanna, Rain forest and Fresh water swamp, but with the greater part in the rainforest. The climate of the area is highly influenced by three major wind currents; the tropical maritime (MT) air mass, the tropical continental (CT) air mass and the equatorial easterlies (Ojo, 1977). The onset of rainfall is usually around March/April, attaining its highest between July and September and finally ceases in November/December. Mean annual rainfall ranges between 1000 mm and 2000 mm with a maximum temperature of 37.9 °C, and a relative humidity of 60%. Since the early 1990s, both environmental and hydrological processes in the area have been severely altered by anthropogenic factors, which amongst others include urbanisation, river dredging and road constructions. Recently, devastating floods of high magnitudes have been experienced in the area which many have attributed to climate and perhaps, rapid changes in land use. The population of the people living under the basin is about 12.9 million (NPC, 2006), while their major means of livelihood is agriculture, including fish farming with maize and other grain crops being the dominant products. There are about 20 dams of different sizes constructed for various purposes within the study area and the possibilities of additional ones in future as the population continues to rise and demands for water for different purposes increase.

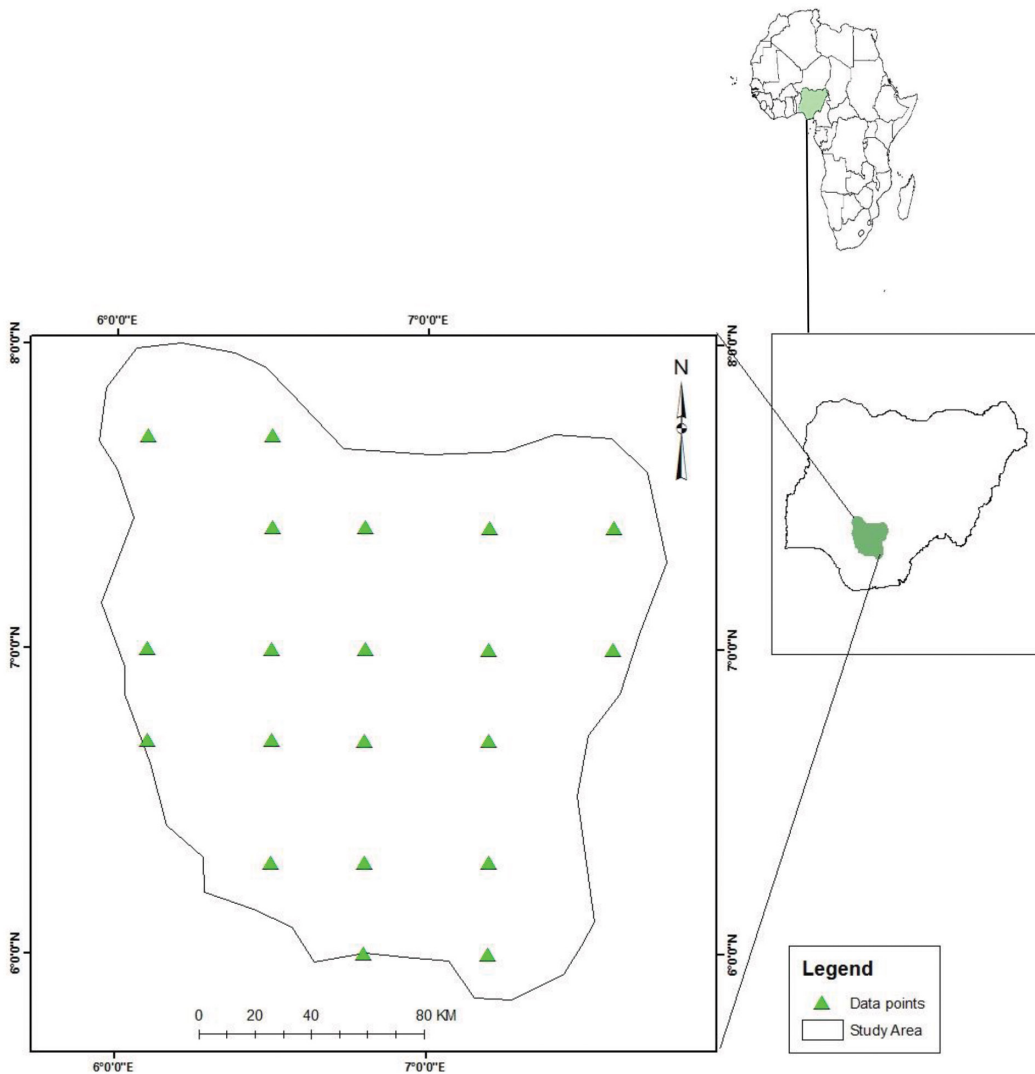


Figure 1. Location of the Niger-South Basin in Nigeria and Africa with the data points used for the study

MATERIALS AND METHODS

Rainfall data for this study were acquired from Global Gridded Climatology (CRU TS 2.1) obtainable at a new high resolution and made available by the Climate Impacts LINK project, Climate Research Unit, University of East Anglia, Norwich, UK (Mitchell & Jones, 2005). The Climatic Research Unit (CRU) data set comprises of monthly 0.5° latitude/longitude gridded series of climatic parameters over the periods of 1901-2008. There is a general poor coverage of meteorological stations in Africa, therefore full information on quality control and interpretation of CRU data are available in related publications (e.g., New et al., 2000; Conway et al., 2009). Kahya and Kalayci (2004) argued that a time series of 30-year is long enough for the evaluation of climatic trend so as to arrive at any realistic conclusion. In this

regard, the use of 61-year data (1948-2008) for this analysis is assumed to be long enough to arrive at a realistic conclusion for the present study. The characteristics of the data points used in terms of spatial spread are presented in Table 1.

Table 1
Description of the dataset

Grid no	Code	Latitude	Longitude	Altitude (m)
1	Prec_47	6.0	6.8	28.32
2	Prec_48	6.0	7.2	113.51
3	Prec_59	6.3	6.5	257.73
4	Prec_60	6.3	6.8	23.89
5	Prec_61	6.3	7.2	42.40
6	Prec_77	6.7	6.1	331.07
7	Prec_78	6.7	6.5	138.45
8	Prec_79	6.7	6.8	26.02
9	Prec_80	6.7	7.2	221.51
10	Prec_97	7.0	6.1	178.28
11	Prec_98	7.0	6.5	51.99
12	Prec_99	7.0	6.8	32.53
13	Prec_100	7.0	7.2	129.47
14	Prec_101	7.0	7.6	395.86
15	Prec_123	7.4	6.5	186.08
16	Prec_124	7.4	6.8	92.91
17	Prec_125	7.4	7.2	285.14
18	Prec_126	7.4	7.6	396.98
19	Prec_148	7.7	6.1	431.29
20	Prec_149	7.7	6.5	145.32

DATA ANALYSIS

Exploratory Data Analysis and Descriptive Statistics

The use of exploratory data analysis (EDA) can help in gaining an insight into the direction and pattern of change in hydro-climatic variables. Aside the formal mathematical methods, it is often included in comprehensive trend detection (Anghileri et al., 2014). EDA refers to any technique of data analysis besides formal statistical methods. It uses graphical tools such as time series plots and scatter plots, and it is aimed at better understanding of the available data and underlying processes. In addition, simple descriptive statistics like the mean, range, standard deviation (SD), coefficient of variation (CV) and determination of maximum and minimum values are also employed to gain preliminary understanding of the data.

Rainfall Variability Index

Rainfall index is normally calculated as the standardised precipitation departure and it provides the metric to classify the rainfall time series into various climatic regimes such as very dry climatic year, normal climatic year and wet or very wet climatic years. Rainfall variability index was computed as:

$$\delta_i = (P_i - \mu) / \sigma \quad [1]$$

where, δ_i is rainfall variability index for year i , P_i is annual rainfall for year i , μ and σ are the mean annual rainfall and standard deviation for the period between 1948 and 2008. According to World Meteorological Organisation (WMO, 1975), rainfall time series can be grouped into different climatic regimes as indicated below.

$P < \mu - 2. \sigma$ – extremely dry

$\mu - 2. \sigma < P < \mu - \sigma$ – dry

$\mu - \sigma < P < \mu + \sigma$ – normal

$P > \mu + \sigma$ – wet

Trend Analysis

The rank-based non-parametric Mann–Kendall (M-K) test (Mann, 1945; Kendall, 1975) has been widely used to assess the significance of monotonic trends in hydro-meteorological time series. The use of M-K test has also been suggested for detecting trends in climatic data by WMO (1988). It is preferred over other techniques because of its benefits which include (1) ability to handle non-normality, outliers and missing values in series or seasonality, and (2) its high asymptotic efficiency (Fu et al., 2004). Therefore, the M-K test is used in this study. For the estimation of an existing trend slope, the Sen's non-parametric method, which has been widely commended for its strength (Salmi et al., 2002; Kahya & Kalayci, 2004; Zhao et al., 2008; Gocic & Trajkovic, 2013a), is used and further detail can be found in Xu et al. (2007).

Trend free pre-whitening

A major requirement of the Mann-Kendall (M–K) test is that time series should be without serial correlation. If serial correlation is positively significant, the strength of M–K is hampered and thus leads to uncertainty. In order to remove or reduce this effect, it is suggested that the original dataset be pre-whitened before applying the M–K test (Abdul Aziz & Burn, 2006). The procedure has been well-documented in Kumar et al. (2010) and many other authors. The final (or pre-whitened) series is then subjected to the M–K test to detect the presence of trend.

Change Point Detection

First, we plotted the residual mass curve, a statistical technique that is widely used in the studies of climatic variations. It is a plot of cumulative deviations from a given reference such as mean, against time or date. In addition, detection of abrupt change point in the series was done using the cumulative sum (CUSUM) technique. The change points in the rainfall series for monthly, seasonal and annual time scales were detected following Kiely (1999), as follows:

$$S_i = \sum_{i=1}^N (y_i - \bar{y}) \quad [2]$$

where, \bar{y} is the average value of the time series. With this, possible change has occurred when S_i is at the maximum. When the CUSUM chart follows a fairly straight line, it signals a period during which the average does not change, whereas an abrupt shift in the average is indicated by an abrupt change in the direction of the CUSUM (Gocic & Trajkovic, 2013a).

RESULTS AND DISCUSSION

Summary of EDA and Descriptive Statistics

The scatter plots of the annual and seasonal series showed the evolution of rainfall over the entire basin for the whole period of study (Figure 2). A visual inspection of the plots shows a decreasing mean annual rainfall. Similarly, on the basis of four season's classification-December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON) - except for JJA which exhibited marginal increase, rainfall declined in all other seasons. However, the rate of decrease was highest in annual series followed by SON, MAM and DJF, respectively.

A summary of the descriptive statistics computed for the long-term (temporal) series is presented in Table 2. Mean annual rainfall varied from 1418.3 mm/yr to 2108.0 mm/yr, while the mean for the entire basin was 1719.2 mm/yr. Similarly, standard deviation (SD) ranged between 162.7 and 286.7mm/yr, with basin-wide SD obtained as 198.0 mm/yr. Maximum and minimum rainfall ranges were 1806.9-2235.0 mm/yr and 1038.4-1207.0 mm/yr, respectively, while the maximum and minimum rainfall for the entire basin were 1255.2 and 2199.5 mm/yr, respectively. For the coefficient of variation (CV), the highest was 15.2%, while the lowest was 11.5%, with the basin-wide CV of 11.5%. This shows that rainfall variability was relatively low within the basin. Similar statistics computed on annual and monthly bases are shown in

Table 3. The minimum rainfall of 0.0 mm/yr was received in December and January, while the maximum rainfall of 461.4 mm/yr was received in August. However, annual minimum and maximum were 1255.2 and 2202.4 mm/yr, respectively. Mean rainfall was highest in August (263.4 mm/yr) and lowest in December (11.2 mm/yr), while the annual mean was 1719.8 mm/yr. SD was highest in July (70.4 mm/yr) and lowest in December (21.5 mm/yr), with the annual SD obtained as 200.7 mm/yr; CV was highest in December (192.9%) and lowest in June (23.1%), whereas annual CV was (11.7%). The results show that rainfall variability was higher on intra-monthly basis than inter-annual in the basin.

In addition, the distribution of the rainfall series plotted as cumulative distribution function (cdf) revealed an overall reduction in rainfall over the basin (Figure 3a). It shows that close to 17% of the entire basin received about 1500 mm/yr of rainfall, 40% experienced about 1750 mm/yr, 90% received about 2000 mm/yr, while just another 20% had above 2000 mm/yr. Further analysis was done to characterise the period of decrease using the cdf by dividing the data set to the periods of 1948-1970 and 1971-2008, based on the assumption that drying in the West African sub-region actually started in the early 1970s (Nicholson et al., 2000). The results also showed that for the period of 1948-1970, less than 15% of the whole basin received about 1500 mm/yr of rainfall, 40% experienced about 1800 mm/yr, 95% received about 2000 mm/yr (Figure 3b), while just 10% had above 2000 mm/yr. However, a decline of the foregoing was noticed for the period of 1971-2008 as the results showed that 18% of the whole basin experienced about 1500 mm/yr of rainfall, 40% experienced about 1700 mm/yr, 99% received about 2000 mm/yr, while less than 2% had above 2000 mm/yr (Figure 3c). These findings actually suggest that rainfall of high amount may have reduced since 1970s in the study area.

Likewise, the percentage contributions on monthly basis are shown in Figure 4. Expectedly, the highest was August (15.3%), followed by July (14.6%) and September (14.2%). Surprisingly, the contribution of June (12.9%) was slightly less than that of May (13.1%), whereas January (1.2%) and February (2.6%) which are generally taken as the driest months in the geographical zone, each contributed more than December (0.7%). On a seasonal basis (not shown), JJA had about 42.9%, MAM, 28.7%; DJF contributed the least (4.5), while SON shared (23.9%). Additional analysis based on two-time slices of 1948-1977 and 1978-2008, showed similar pattern of seasonal contributions except that while only JJA increased from 40.3 to 45.5%, MAM (29.6 to 27.9%), DJF (5.4 to 3.5%) and SON (24.7 to 23.2) had reduced contributions. In general, the long-term rainfall of the basin was dominated by the JJA seasonal intensity.

Rainfall Trends in the Niger-South Basin, Nigeria

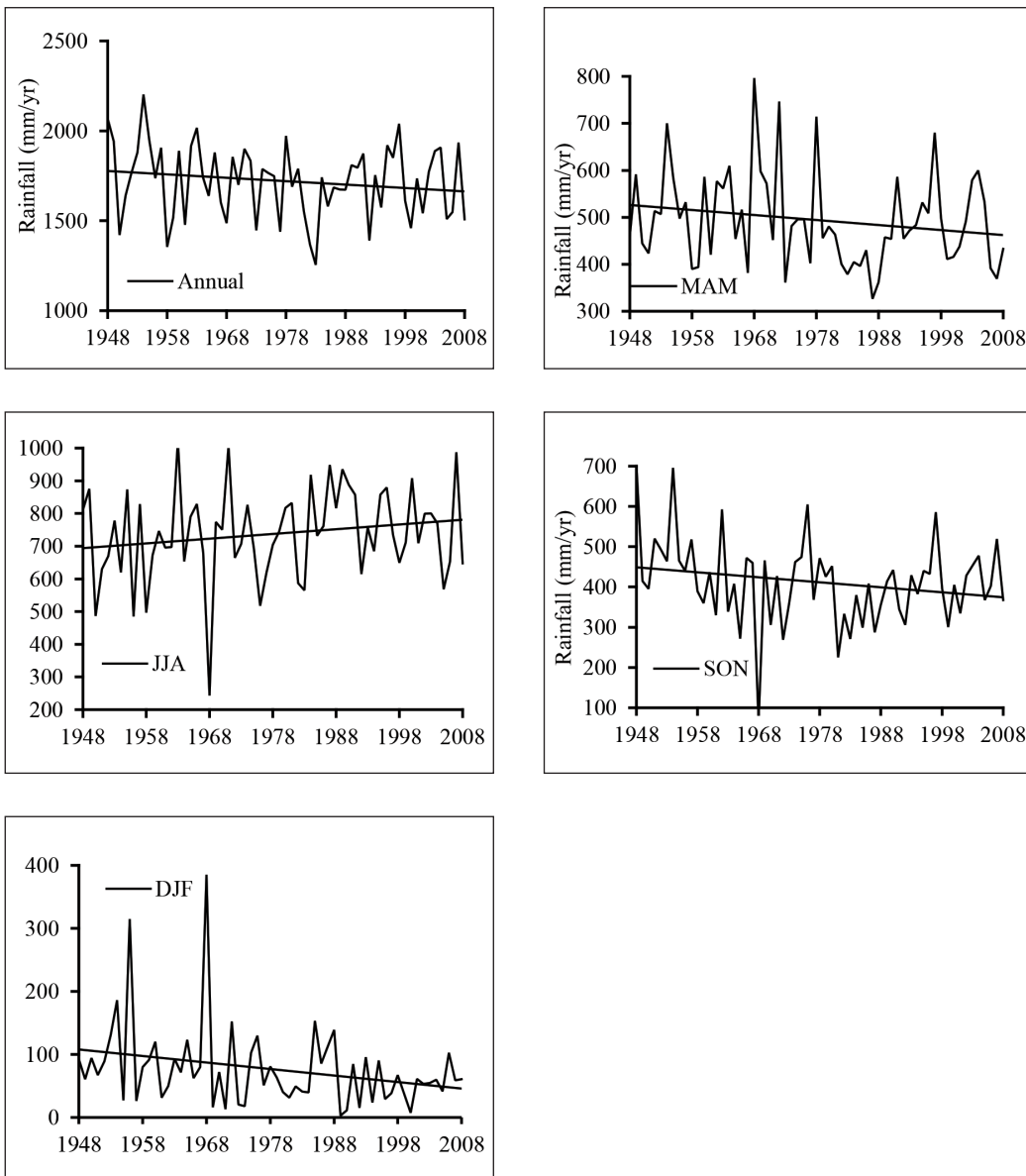


Figure 2. Times series plots of annual and seasonal rainfall data averaged over the basin

Table 2
Descriptive statistics and trend analysis (M-K) for the entire basin

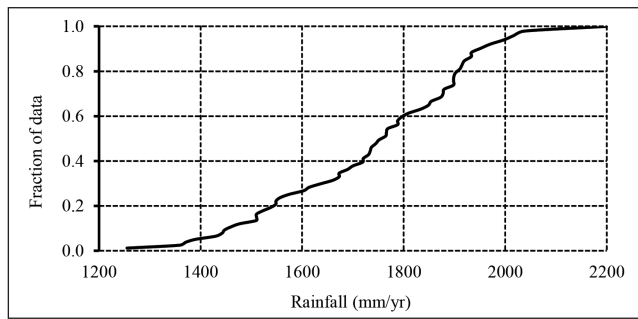
Code	Minimum (mm/yr)	Maximum (mm/yr)	Mean (mm/yr)	SD (mm/yr)	CV (%)	Z (mm/yr)	Slope (mm/yr)
Prec_47	1386.9	2677.2	2101.7	286.0	13.6	-1.25	-2.98
Prec_48	1424.7	2712.6	2108.0	286.7	13.6	-1.46	-3.44
Prec_59	1333.2	2523.0	1996.9	257.6	12.9	-0.85	-1.96
Prec_60	1281.0	2548.6	1981.1	269.2	13.6	-0.94	-2.22
Prec_61	1331.1	2541.7	1981.8	276.0	13.9	-1.69	-3.62
Prec_77	1231.5	2285.7	1791.1	234.4	13.1	-0.50	-1.03
Prec_78	1235.5	2286.0	1771.1	223.7	12.6	-0.68	-1.18
Prec_79	1256.4	2264.9	1742.7	217.2	12.5	-0.85	-1.23
Prec_80	1299.1	2253.4	1727.2	230.5	13.3	-1.57	-2.56
Prec_97	1194.7	2145.7	1677.7	219.1	13.1	-0.27	-0.57
Prec_98	1207.5	2130.4	1656.0	206.2	12.5	-0.57	-0.61
Prec_99	1205.6	2099.9	1627.5	199.3	12.2	-0.57	-1.16
Prec_100	1202.9	2158.0	1617.6	216.9	13.4	-1.57	-2.56
Prec_101	1128.7	2301.3	1632.5	247.5	15.2	-1.92	-3.75
Prec_123	1118.9	1983.3	1547.2	190.7	12.3	-0.47	-0.59
Prec_124	1089.8	1943.6	1518.3	184.5	12.2	-0.70	-1.04
Prec_125	1056.4	2023.3	1514.5	204.7	13.5	-1.60	-2.63
Prec_126	1053.9	2169.1	1526.8	232.2	15.2	-1.81	-3.37
Prec_148	1072.9	1846.5	1446.4	170.9	11.8	-0.18	-0.24
Prec_149	1038.4	1806.9	1418.3	162.7	11.5	-0.38	-0.54
Basin	1255.2	2199.5	1719.2	198.0	11.5	-1.19	-2.07

Bold font significant at 90%

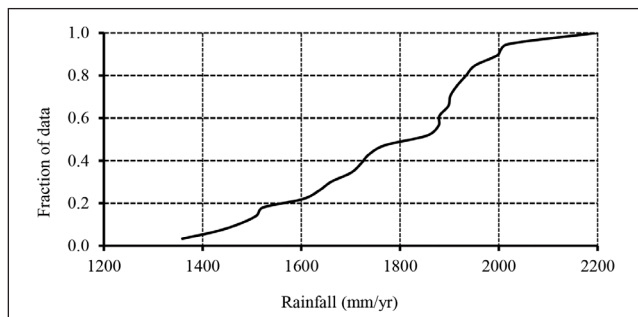
Table 3
Descriptive statistics of basin average monthly and annual time series

Series	Minimum (mm/yr)	Maximum (mm/yr)	Mean (mm/yr)	SD (mm/yr)	CV (%)
January	0.0	107.2	21.0	24.4	116.2
February	0.8	220.1	44.5	38.2	85.8
March	21.0	315.7	97.3	50.0	51.4
April	38.5	408.0	171.2	63.1	36.9
May	110.3	392.0	225.6	53.2	23.6
June	108.9	381.3	222.5	51.5	23.1
July	67.9	443.2	251.6	70.4	28.0
August	0.7	461.4	263.4	69.0	26.2
September	0.3	402.0	246.4	57.2	23.2
October	8.0	253.4	133.8	52.4	39.1
November	0.1	140.9	31.3	31.4	100.5
December	0.0	121.5	11.2	21.5	192.9
Annual	1255.2	2202.4	1719.8	200.7	11.7

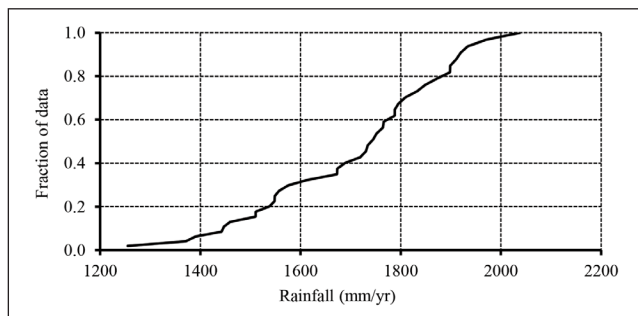
Rainfall Trends in the Niger-South Basin, Nigeria



(a)



(b)



(c)

Figure 3. Rainfall distribution (a) 1948-2008; (b) 1948-1977; and (c) 1978-2008 plotted as cumulative distribution function over the basin

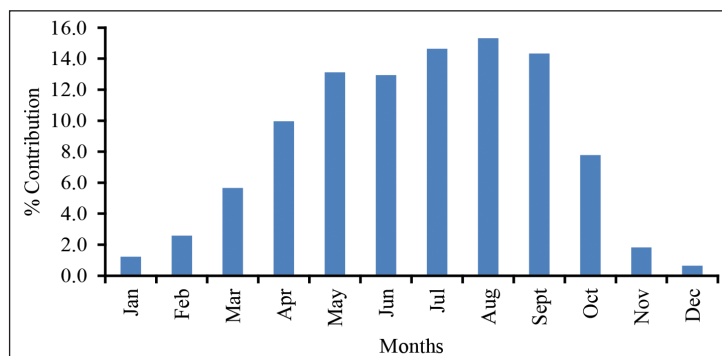
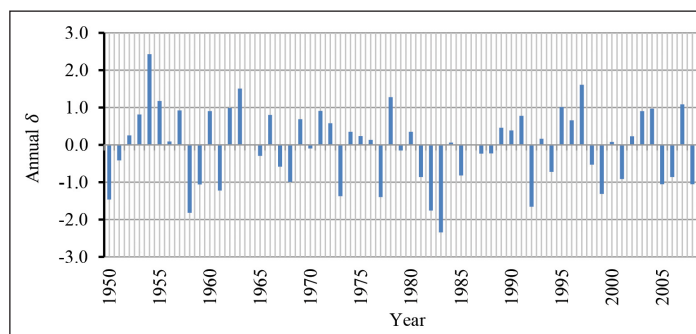


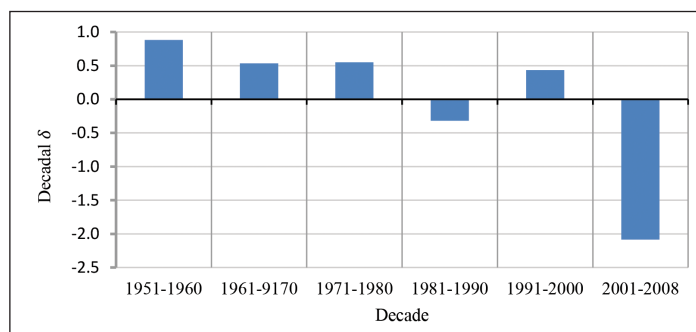
Figure 4. Percentage monthly contribution to annual rainfall

Rainfall Variability

Rainfall variability indices on annual and decadal bases covering 1950 to 2008 are presented in Figures 5a & 5b. The results show three series of distinctive periods which may be described for the basin as: (1) from 1950 to 1965 (16 years), a seemingly random succession of six dry years, five “normal” years and 3 wet years; (2) from 1966 to 1980 (15 years), a series of 4 dry years, 8 “normal” years, and one wet year; (3) from 1981 to 2008 (28 years) of 12 dry years, 11 “normal” years and two wet years. Throughout the second half of the 20th century, the decade 1980s was the driest, while decade 1950s the wettest (Figure 5b). Nevertheless, for the entire 61-years period of study, the decade 2000s was the driest. Moreover, following WMO (1975) classification, as applied by Gocic and Trajkovic (2013b), only 12 years can be classified as “dry”. The dry years were 1950, 1958, 1959, 1961, 1968, 1973, 1977, 1982, 1992, 2005 and 2008, amounting to about 18% of the entire series. On the other hand, eight other years exhibited “normal” conditions; the years were 1948, 1949, 1954, 1955, 1963, 1978, 1997 and 2007, while the rest were “wet”. Throughout the whole length of period under study, “extremely dry” condition was only experienced in 1983.



(a)



(b)

Figure 5. Rainfall variability (a) annual; and (b) decadal

Mann-Kendall (M-K)

The M-K trend test was done using a spreadsheet (Makesen 1.0) developed at the Finnish Meteorological Institute (Salmi et al., 2002). The results presented in Table 2 for all the 20 data points and the entire basin showed negative trends with only three significant ($\alpha < 0.1$). Trend slopes ranged from -3.75 to -0.24 mm/yr and for the entire basin were -1.29 mm/yr and -2.23 mm/yr, respectively. Analysis based on the monthly series also showed predominantly decreasing trends that are not significant (Table 4, columns 2, 3 & 4). For example, out of the 7 months which showed downward trends - January, February, March, April, June, September, October, only January (-0.18 mm/yr), March (-0.64 mm/yr) and September (-0.73 mm/yr) trends were significant. Except for August that showed a significant increasing trend (0.79 mm/yr), the other four months exhibited positive insignificant trends of varied magnitudes. The rate of change ranged between -0.73 mm/yr (September) and 0.78 mm/yr (August). On the seasonal basis, aside from DJF which showed a significant downward trend ($\alpha < 0.1$), trends were generally not significant; however, JJA recorded an upward trend (1.43 mm/yr), whereas slope magnitude ranged between -1.04 mm/yr (SON) and 1.42 mm/yr (JJA).

Moreover, analysis based on the two time slices (1948-1977 and 1978-2008) showed an insignificant decreasing trends (-7.03 mm/yr) with the magnitude -210.8 mm/yr between 1948 and 1977, while an insignificant increasing trend (2.64 mm/yr) was experienced during 1978-2008, with the magnitude of change reaching 81.69 mm/yr. Nevertheless, over the whole series length, an insignificant decreasing trend (slope = -2.25 mm/yr) with the magnitude of -137.07 mm/yr was observed. Similarly, while the trends were generally downward in all the seasons between 1948 and 1977, it was only significant ($\alpha < 0.1$) during SON (not shown). However, between 1978 and 2008, there was no significant trend in all seasons; nonetheless, while both DJF and JJA depicted downward trends, MAM and SON showed upward trends. Obviously, there have been sorts of redistribution in the annual rainfall amongst the seasons between the two time periods which favoured MAM and SON. This agrees with the finding of Wang et al. (2011) who stated that in the event of decreased annual precipitation, which is probably due to the change in global climate or elevation; monthly rainfall would not just decrease as a consequence, but is re-distributed to smoothen the seasonality. This result has implications for dry season farming as farmers may be able to extend their farming season to the usually dry months of SON if the trend persists in the area. With respect to the change point for all cases, the results determined using the CUSUM plot are shown in column 5 of Table 4 below.

Table 4
Change points and M-K trend results based on monthly, annual, seasonal and two time slices

Series	Test Z	M-K	Change magnitude	Change .Point
January	-1.70	-0.18 ⁺	-10.88	1970
February	-1.52	-0.34 ⁺	-21.03	1986
March	-1.65	-0.63	-38.72	1972
April	-1.13	-0.44	-26.97	1981
May	0.13	0.07	4.35	1972
June	-0.27	-0.13	-7.84	1949
July	0.94	0.49	29.93	1968
August	1.92	0.79 ⁺	47.90	1983
September	-1.90	-0.73 ⁺	-44.78	1962
October	-1.11	-0.43	-26.49	1955
November	0.09	0.01	0.74	1963
December	0.30	0.00	0.28	1968
Annual	-1.29	-2.25	-137.07	1957
DJF	-2.00	-0.62 ⁺	-37.51	1968
MAM	-1.35	-0.94	-57.56	1972
JJA	1.19	1.43	87.17	1962
SON	-1.57	-1.04	-63.68	1983
1948-1977	-1.53	-7.03	-210.79	1957
1978-2008	0.41	2.64	81.69	1980

⁺ significant at 90%

DISCUSSION

In the present study, a complete picture of rainfall analysis covering the period between 1948 and 2008 over the NSB was undertaken. Generally, there have been reductions in rainfall over the entire landscape since the beginning of the period under study (Figure 2(a)). Unlike the results of Oguntunde et al. (2011), the average annual rainfall series between 1948 and 1970 were not significantly higher than that between 1971 and 2008, while variability was higher temporally than spatially. However, the result of the low spatial variability might be due to the smaller spatial coverage of the present study as compared with the previous. Conversely, the observed annual variability (Figure 5a) was not significantly different from that of L'Hôte et al. (2002). Based on the cdf, a greater portion of the basin also received mean annual rainfall higher than 2000 mm/yr between 1948 and 1977 as compared to between 1978 and 2008, and for the whole length of period. The increasing amount of mean annual rainfall combined with the rapid change in land cover in the basin could be the reason for floods that have been ravaging the area recently.

Rainfall variability also showed that there were many more dry years between 1981 and 2008, with only 1983 being extremely dry as compared with the period between 1948 and 1980. With respect to decadal variability, the decade 2000s was the driest, while decades 1950s was

the wettest. The results agree with the findings of earlier studies (e.g., Nicholson et al., 2000; Oguntunde et al., 2011); however, the decade 1980s was reported as the driest in the study by the second authors. Apparently, the emergence of the last decade as the driest in the present study is in agreement with the projection of increased drying during the 21st century (IPCC, 2001; 2007). Nonetheless, it should be noted that the temporal scale covered by the present study differs from that of the earlier authors, although the study still showed the decade 1980s as the driest of the last century.

On the seasonal time scale, rainfall has been declining in all seasons except in JJA, which showed a relative increase. Similar to the findings of Miao et al. (2012), who found an increase in summer rainfall in their study of seasonal precipitation for Beijing in China, there was also an increase of about 5% in the contribution of JJA during 1978 to 2008 as compared with 1948 to 1977. The increasing rainfall amount in JJA was mostly caused by the rise in the rainfall amount in August. In addition, of the 12 months of the year, December has been the driest and most varied while August was the wettest. Thus, this suggests that flood events which have been witnessed in the area during September months may have stemmed from the heavy rainfall in August and could persist intermittently for a while in the coming years.

In comparison with the findings of Oguntunde et al. (2011) in their study over Nigeria, decreasing trend dominated the entire basin and mean annual rainfall. Similar insignificant downward trend in rainfall was found by Salami et al. (2014) in their application of M-K test to hydro-meteorological variables at Lokoja, a location in the basin. Elsewhere in India, Jain and Kumar (2012) reported that basin-wise trend analysis carried out over the country showed that 15 basins experienced downward trend in annual rainfall, while one basin depicted a significant decreasing trend at 95% confidence level. Similarly, Longobardi and Villani (2010) also found a predominantly decreasing trend in their study in the Mediterranean area, both at the annual and monthly scale, with over 97% of the total stations significant.

Seasonally, an upward trend that is insignificant was observed in JJA, while other seasons showed a downward trend. In the same vein, 8 of the 12 months in the basin have shown downward rainfall trends, while the other four showed increasing trends. Once again, this also agrees with the results of Hess et al. (1995), who linked rainfall reduction in the northeast arid zone of Nigeria between 1961 and 1990 to decreases in August and September rainfalls, while Adefolalu (2007) attributed the destructive flood events in southern Nigeria to the rapid rise in August rainfall over the region during the last 50 years of the century.

CONCLUSION

The study examined trends in monthly, seasonal and annual rainfall time series between 1948 and 2008 over the NSB. The overall results of this study showed that the general drying in the entire basin, which started since the last 30 years of the 20th century, has intensified into the present. Generally, rainfall in the summer months, especially August, has also been persistent leading to floods in the basin. Meanwhile, given the projection of the IPCC (2007) for the present century, the drying that has been observed in this study might be linked directly or indirectly with climate change. However, considering the rapid rate of urbanisation in the basin, impacts of land cover/use change might have also contributed to the intermittent floods

being witnessed in the area. Nonetheless, water resources management efforts may require the construction and expansion of water storage facilities within the basin to reduce the impacts of summer flooding and possible water scarcity in the future.

ACKNOWLEDGEMENTS

The first author is grateful to the Management of Rufus Giwa Polytechnic, Owo, and the Tertiary Education Trust Fund (TETFUND) Abuja, Nigeria, for sponsoring the PhD research which produced this paper.

REFERENCES

- Abaje, I. B., Ati, O. F., & Iguisi, E. O. (2010). An Analysis of Rainfall Trends in Kafanchan, Kaduna State, Nigeria. *Research Journal of Environmental and Earth Sciences*, 2(2), 89-96.
- Aziz, O. I. A., & Burn, D. H. (2006). Trends and variability in the hydrological regime of the Mackenzie River Basin. *Journal of Hydrology*, 319(1), 282–294.
- Adefolalu, D. O. (2007). Climate change and economic sustainability in Nigeria. In *International Conference on Climate Change and Economic Sustainability*. Nnamdi Azikiwe University, Enugu, Nigeria.
- Anghileri, D., Pianosi, F., & Soncini-Sessa, R. (2014). Trend detection in seasonal data: from hydrology to water resources. *Journal of Hydrology*, 511, 171–179.
- Blake, R., Grimm, A., Ichinose, T., Horton, R., Gaffin, S., Jiong, S., ... Cecil, L. D. (2011). Urban climate: Processes, trends, and projections. *Climate change and cities: First assessment report of the Urban Climate Change Research Network*, 43-81.
- Brunetti, M., Maugeri, M., Monti, F., & Nanni, T. (2006). Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. *International journal of climatology*, 26(3), 345-381.
- Conway, D., Persechino, A., Ardoin-Bardin, S., Hamandawana, H., Dieulin, C., & Mahé, G. (2009). Rainfall and water resources variability in sub-Saharan Africa during the twentieth century. *Journal of Hydrometeorology*, 10(1), 41-59.
- Dore, M. H. (2005). Climate change and changes in global precipitation patterns: what do we know? *Environment international*, 31(8), 1167-1181.
- Fu, G., Chen, S., Liu, C., & Shepard, D. (2004). Hydro-climatic trends of the Yellow River basin for the last 50 years. *Climatic Change*, 65(1), 149-178.
- Gocic, M., & Trajkovic, S. (2013a). Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. *Global and Planetary Change*, 100, 172–182.
- Gocic, M., & Trajkovic, S. (2013b). Analysis of precipitation and drought data in Serbia over the period 1980–2010. *Journal Hydrology*, 494, 32–42.
- Hess, T. M., Stephens, W., & Maryah, U. M. (1995). Rainfall trends in the north east arid zone of Nigeria 1961–1990. *Agricultural and Forest Meteorology*, 74(1-2), 87-97.

- Ikhile, C. I. (2007). *Impacts of climate variability and change on the hydrology and water resources of the Benin-Owena River basin*. (Unpublished PHD Thesis). Department of Geography and Regional Planning, University of Benin, Benin City, Nigeria.
- IPCC (2007). *IPCC Fourth Assessment Report: Climate Change 2007*. Geneva, Switzerland.
- Jain, S. K., & Kumar, V. (2012). Trend analysis of rainfall and temperature data for India. *Current Science*, 102(1), 37-49.
- Jassen, E. E. (2013). *Trends and Projections of Extreme Precipitation over the Contiguous United States*. (Unpublished MSC Thesis). University of Illinois at Urbana-Champaign, USA.
- Kahya, E., & Kalayci, S. (2004). Trend analysis of streamflow in Turkey. *Journal of Hydrology*, 289(1), 128-144.
- Karpouzou, D. K., Kavalieratou, S., & Babajimopoulos, C. (2010). Trend Analysis of Precipitation Data in Pieria region (Greece). *European Water*, 30, 31-40.
- Kendall, M. G. (1975). *Rank Correlation Methods*. Charles Griffin, London.
- Kiely, G. (1999). Climate change in Ireland from precipitation and streamflow observations. *Advances in water resources*, 23(2), 141-151.
- Kumar, V., Jain, S. K., & Singh, Y. (2010). Analysis of long-term rainfall trends in India. *Hydrological Sciences Journal*, 55(4), 484-496.
- L'Hôte, Y., Mahe, G., Some, B., & Triboulet, J. P. (2002). Analysis of a Sahelian annual rainfall index from 1896 to 2000; the drought continues. *Hydrological Science Journal* 48(3), 489-496.
- Longobardi, A., & Villani, P. (2010). Trend analysis of annual and seasonal rainfall time series in the Mediterranean area. *International journal of Climatology*, 30(10), 1538-1546. doi: 10.1002/joc.2001.
- Mann, H. B. (1945). Nonparametric tests against trend. *Econometrical*, 13(3), 245-259.
- Miao, L. I., Jun, X., & Dejuan, M. (2012). Long-term Trend Analysis of Seasonal Precipitation for Beijing, China. *Journal of Resources & Ecology*, 3(1), 064-072.
- Mitchell, T. D., & Jones, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International journal of climatology*, 25(6), 693-712.
- Mondal, A., Kundu, S., & Mukhopadhyay, A. (2012). Rainfall Trend Analysis by Mann-Kendall Test: A Case Study of North-Eastern Part of Cuttack District, Orissa. *International Journal of Geology, Earth and Environmental Sciences*, 2(1), 70-78.
- New, M., Hulme, M., & Jones, P. (2000). Representing twentieth-century space-time climate variability. Part II: development of 1901-96 monthly grids of terrestrial surface climate. *Journal of climate*, 13(13), 2217-2238.
- Nicholson, S. E., Some, B., & Kone, B. (2000). An analysis of recent rainfall conditions in West Africa, including the rainy seasons of the 1997 El Nino and the 1998 La Nina years. *Journal of Climate* 13(14), 2628-2640.
- NPC. (2006). The Nigeria Population Commission Official Result for 2006 Population and Housing Census Figures, Bureau for National Statistics. *National Population Commission, Abuja*. Retrieved from <http://www.nigerianstat.gov.ng>.
- Oguntunde, P. G., Abiodun, B. J., & Lischeid, G. (2011). Rainfall trends in Nigeria, 1901-2000. *Journal of Hydrology*, 411(3), 207-218.

- Ojo, O. (1977). *The Climates of West Africa*. Heineman, London.
- Salami, A. W., Mohammed, A. A., Abdulmalik, Z. A., & Olanlokun, O. K. (2014). Trend Analysis of Hydro-Meteorological Variables using The Mann-Kendall Trend Test: Application to The Niger River and The Benue Sub-Basins in Nigeria. *International Journal of Technology*, 5(2), 100-110. doi: <http://dx.doi.org/10.14716/ijtech.v5i2.406>.
- Salmi, T., Maatta, A., Anttila, P., Ruoho-Airola, T., & Amnell, T. (2002). Detecting Trends of Annual Values of Atmospheric Pollutants by the Mann-Kendall Test and Sen's Slope Estimates. *Publications on Air Quality, No. 31*. Helsinki, Finland.
- Sonali, P., & Kumar, D. N. (2013). Review of trend detection methods and their application to detect temperature changes in India. *Journal of Hydrology*, 476, 212–227.
- Wang, D., Hejazi, M., Cai, X., & Valocchi, A. J. (2011). Climate change impact on meteorological, agricultural, and hydrological drought in central Illinois. *Water Resources Research*, 47(9), 1-13. doi:10.1029/2010WR009845.
- WMO (1975). *Drought and Agriculture*. WMO/ TN 138, Geneva.
- WMO (1988). *Analyzing long time series of hydrological data with respect to climate variability*. World Meteorological Organization (WMO): WCAP-3, WMO/TD-No: 224, Switzerland, pp. 1–12.
- Xu, Z. X., Li, J. Y., & Liu, C. M. (2007). Long-term trend analysis for major climate variables in the Yellow River basin. *Hydrological Processes*, 21(14), 1935-1948.
- Yang, X. L., Xu, L. R., Liu, K. K., Li, C. H., Hu, J., & Xia, X. H. (2012). Trends in temperature and Precipitation in the Zhangweinan River Basin during the last 53 Years. *Procedia Environmental Sciences*, 13, 1966 – 1974.
- Zhang, X., Brown, R., Vincent, L., Skinner, W., Feng, Y., & Mekis, E. (2011). *Canadian climate trends, 1950-2007. Canadian Biodiversity: Ecosystem Status and Trends 2010, Technical Thematic Report No. 5*. Canadian Councils of Resource Ministers. Ottawa, ON. iv + 21p. <http://www.biodivcanada.ca/default.asp?lang=En&n=137E1147-0>
- Zhao, F. F., Xu, Z. X., Huang, J. X., & Li, J. Y. (2008). Monotonic trend and abrupt changes for major climate variables in the headwater catchment of the Yellow River basin. *Hydrological Processes*, 22(23), 4587-4599.

Numerical Analysis of The Effect of Nozzle Geometry on Flow Parameters in Abrasive Water Jet Machines

Deepak, D.*¹, Jodel, A. Q., Cornelio¹, Midhun Abraham, M.¹ and Shiva Prasad, U.^{2#}

¹Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal University, Manipal 576104, Karnataka, India

²Department of Aeronautical and Automobile Engineering, Manipal Institute of Technology, Manipal University, Manipal 576104, Karnataka, India

ABSTRACT

The nozzle of the AWJ machine is a critical component which has direct influence on the jet force developed. In the present scenario, commercially available nozzles have conical section followed by focus section. The critical section is where the cross section of the nozzle changes from conical to straight and suffers severe wall shears stress leading to flow loss. In considering this, computational analysis has been carried on the jet flow through AWJ nozzle with different nozzle geometries. The geometric variation of nozzle profiles show that, reduction in radius of curvature (radius 20 mm) of nozzle geometry produced higher jet velocity and force, as well as lower pressure drop compared to other geometric dimensions.

Keywords: AWJ nozzle, jet force, SOD, CFD analysis, garnet abrasive

INTRODUCTION

Technological advancements have made Abrasive Water Jet (AWJ) as a one of the promising field of non-conventional machining. Presently, AWJ machines are developed to operate with pressures up-to 1000 MPa for cutting and drilling applications on various materials. The

critical component of the AWJ machine is the nozzle which has direct influence on the jet force developed. In the present scenario, commercially available nozzles have conical section followed by focus section. The critical section, where cross section of the nozzle changes from conical to straight suffers severe wall shears stress which leads to flow loss and jet energy (Deepak et al., 2012). There are cited literatures on nozzle flow based on experimental and numerical methods. Jiang

Article history:

Received: 19 May 2016

Accepted: 30 August 2016

E-mail addresses:

deepak.d@manipal.edu (Deepak, D.),

jodel902@yahoo.com (Jodel, A. Q., Cornelio),

mathewsmathew.mm@gmail.com (Midhun Abraham, M.),

shivaprasad047@gmail.com (Shiva Prasad, U.)

*Corresponding Author

Author's Current Affiliation:

Department of Aeronautical Engineering,

Institute of Aeronautical Engineering, 500043 Hyderabad, India

et al. (2013) studied the influence of grid size on the accuracy of the numerical simulations. The effect of inlet pressure and particle size on the jet flow field was studied by Deepak et al. (2011). Due to the geometrical structure, the velocity of the jet was found to increase up-to critical section of the nozzle and thereafter it decreased slightly in focus section due to Wall Shear Stress (WSS). Optimum focus length of nozzle was determined by Hu et al. (2008).

Meanwhile, Taggart et al. (2002) investigated the effects of abrasive flow rate, length, cone angle and diameter of nozzle, as well as inlet water pressure on nozzle wear by experimental method. The nozzle length exhibited a direct influence on the exit jet diameter, which delays the development of wear profile from reaching the exit. Flow pattern within the nozzle was unaffected by its length. Increase in the inlet angle exhibited non-linear increase in bore of the nozzle. The optimum diameter ratio of the nozzle was found to be between 0.3 and 0.4. Increase in the abrasive flow rate resulted in increased nozzle wear rate without changing the wear pattern, but increase in water pressure exhibited maximum wear rate. Anand and Katz (2003) developed porous nozzle which reduced nozzle wear by providing lubrication. The oil film was also found to improve the coherence of the jet. Erosion rate of the boron carbide nozzle was found to increase with increase in abrasive hardness (Deng, 2005). The nozzle entrance area suffered from severe abrasive impact under large impact angles, at the nozzle centre wall section, most of the particles travelled parallel to the nozzle wall, and showed minimum tensile stress. The wear mode in this area was found to be changed from impact to sliding erosion and the wear mechanisms appeared to be the lateral cracking owing to a surface fatigue fracture.

Shimizu et al. (2008) studied the effect of abrasive particle size on jet structure that was formed after exiting the nozzle. For higher abrasive particle sizes, the jet diameter at the exit of the nozzle was found to increase with the increase in distance from the nozzle tip. The jet became unstable at a distance of about $x = 25 \times d_f$ and the jet broke up at a distance of about $x = 80 \times d_f$. In case of finer abrasive particle (mesh size: 220), stable compact jet was formed up to distance of about $x = 80 \times d_f$. This indicated that abrasive particles of larger size contributed to flow turbulence and promote the jet breakup at comparatively shorter distances. Vu and Nguyen (2009) developed model to predict the nozzle life at various operating conditions. Numerical investigation was made by Mabrouki et al. (2000) to study the effect of abrasive water jet impact on aluminium sheets.

Kasmol et al. (2001) developed a model using FEA to predict the depth of jet penetration onto work-piece. Simulations show that pure water jet has less interaction on target material, whereas with AWJ, a strong interaction with target material in terms of localised deformation was found. The developed model was validated with experimental results which were in good agreement. Yang et al. (2009) investigated the effect of SOD and fluid viscosity on jet performance. The higher the viscosity of fluid, the more coherent is the velocity distributions at the impacting region. A velocity distribution along the length of traverse till the target material was found to be different at different positions. Waviness in velocity distribution was seen on target material for SOD greater than 3 mm. Since the nozzle geometry also has an influence on the flow properties of the jet, the present work aims at jet force, pressure drop, WSS developed in nozzle having different profiles.

THEORITICAL FORMULATION

Numerical Model

Computational domain consists of converging of nozzle having length 4 mm along the axis of the nozzle. The length of the converging section of the nozzle is 10mm with focus section of length 5mm. The profile of the nozzle is varied with radius 20 mm, 40 mm, 60mm and their flow characteristics are compared with the nozzle having straight conical section.

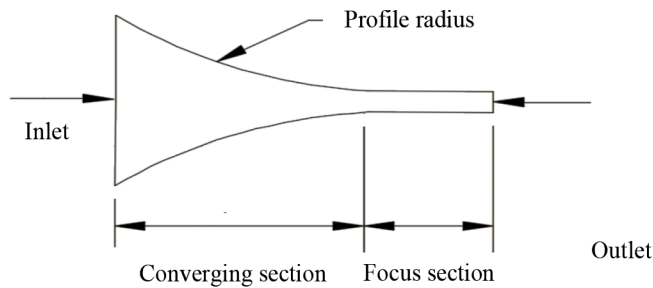


Figure 1. Profile of the AWJ nozzle for flow analysis

The inlet and exit diameters are maintained constant for all nozzles at 8mm and 1mm respectively. The mixture of abrasive and water is let into the nozzle at the inlet and is carried down through the converging cone to the focus tube and exits as coherent jet at the nozzle exit, in which the focus tube is used for stabilising the flow. Figure 1 shows the numerical region for flow analysis. The governing equations for mass and momentum conservation are solved for the steady incompressible flow. The coupling between velocity and pressure has been attempted through the phase coupled SIMPLE algorithm. The standard k-ε turbulence model is used to predict the flow physics with standard wall functions. The governing partial differential equations for mass and momentum conservations are detailed below.

Continuity Equation

$$\frac{1}{\rho_{pq}} \left[\frac{\partial}{\partial t} (\alpha_q \rho_q) + \nabla \cdot (\alpha_q \rho_q v_q) \right] = \sum_{p=1}^N (m_{pq} - m_{qp}) \quad [1]$$

Fluid-Solid Momentum Equation

The conservation of momentum equation for the solid phase is as follows.

$$\frac{\partial}{\partial t} (\alpha_s \rho_s v_s) + \nabla \cdot (\alpha_s \rho_s v_s^2) = -\alpha_s \nabla p - \nabla p_s + \nabla \cdot \tau_s + \alpha_s \rho_s g + \sum_{l=1}^N [k_{ls} (v_l - v_s) + (m_{ls} v_{ls} - m_{sl} v_{sl})] + (F_s + F_{lift,s} + F_{vm,s}) \quad [2]$$

The conservation of momentum equation for the fluid phase is as follows.

$$\frac{\partial}{\partial t}(\alpha_q \rho_q v_q) + \nabla \cdot (\alpha_q \rho_q v_q^2) = -\alpha_q \nabla P + \nabla \cdot \tau_q + \alpha_q \rho_s g + \sum_{p=1}^N [k_{pq}(v_p - v_q) + (m_{pq} v_{pq} - m_{qp} v_{qp})] + (F_q + F_{lift,q} + F_{vm,q}) \quad [3]$$

Assumptions of the Numerical Analysis

Following assumptions are made for the formulation of numerical solution.

- The primary liquid phase is continuous and incompressible.
- Two-phase flow assumed is steady and characterized by turbulent flow.
- The Coanda effect is not considered on the flow.
- Flow is taken to be two-phase flow in which the primary liquid phase mixes homogeneously with the particles of equal diameter, constituting the solid phase.

Numerical Scheme

Computational domain is discretised by quad cells. Conservation equations are solved for each control volume to obtain the velocity and pressure fields. Convergence is affected when all the dependent variable residuals fall below 0.00001 at all grid points. Wall region in the flow domain is closely meshed using the boundary layer mesh concepts for extracting high velocity gradients near the boundary walls as shown in Figure 2. According to the structure of nozzle and jet characteristics, computational domain is built as axis-symmetric model.

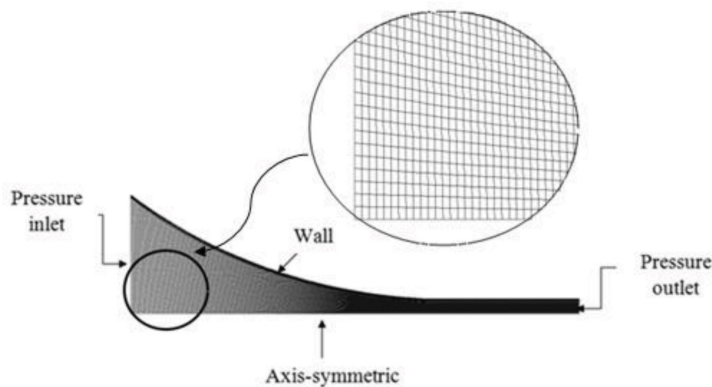


Figure 2. Computational domain with boundary conditions

Grid Independence Test and Validation of the Results

Grid test is performed to check the quality of mesh for solution convergence. Figure 3 shows the velocity distribution along the length of the nozzle for grids of different sizes. It is clear from the graph that there is almost negligible variation in the velocity for the grids of different sizes. Considering the accuracy of the results and computational time, the geometry with appropriate number of grids are chosen for the study. Figure 3 also shows the velocity distribution along the nozzle axis, as obtained by the analytical formulations. It shows that results obtained from numerical simulations are in good agreement with the theoretical calculations.

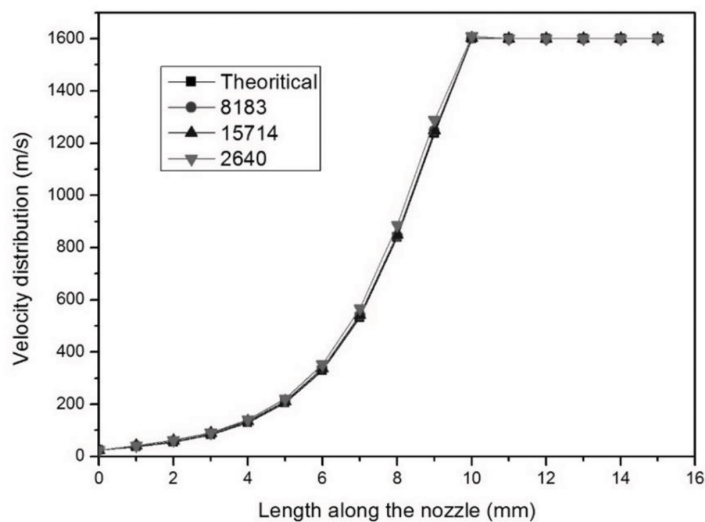


Figure 3. Velocity distribution for different mesh sizes

RESULTS AND DISCUSSION

Effect of nozzle profile on flow velocity

The profile of the nozzle is varied with radius of 20 mm, 40 mm and 60 mm as shown in Figure 1. Numerical simulations were carried out for the flow domain consisting of two phase mixture of water and garnet abrasive of diameter 160 μm . The inlet pressure is varied from 100 to 400 MPa and the corresponding velocities are shown in Figures 4(a) and 4(b) respectively.

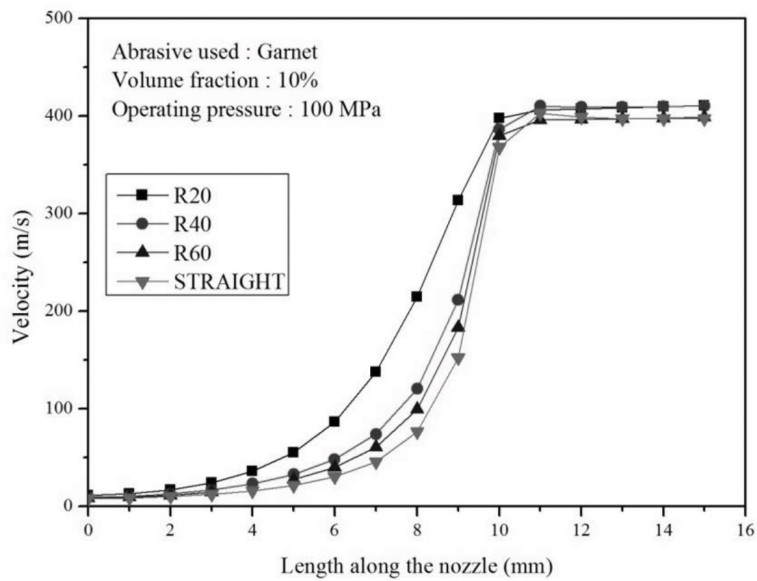


Figure 4(a). Velocity distribution for nozzle profiles at inlet pressure 100 MPa

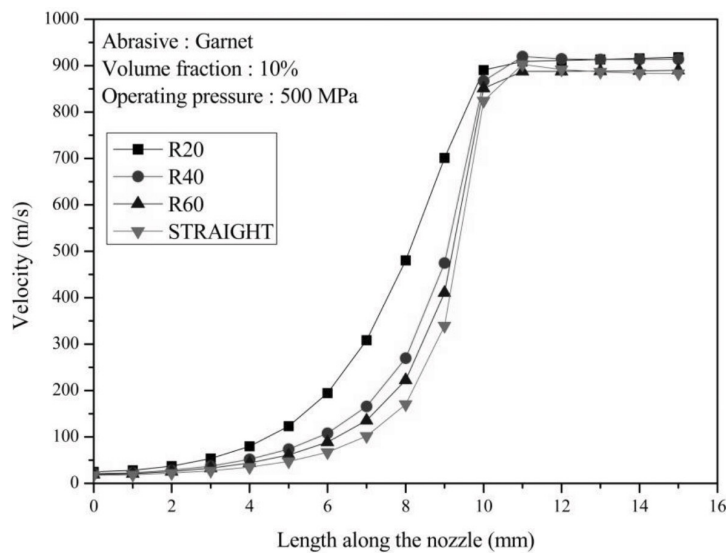


Figure 4(b). Velocity distribution for nozzle profiles at inlet pressure 400 MPa

It was observed that the jet velocity increased along the axis of the nozzle and it reached peak exit velocities ranging from 400 m/s to 900 m/s for different inlet pressures. The jet velocity is found to rapidly increase for a axial distance of 10 mm, i.e. up to the region where cross-section changes to straight circular duct of length 5 mm at the end of the nozzle. The variation in the jet velocity remains almost constant in the focus section. From the geometrical variation of the nozzle profile, it was observed that there is an improvement in the jet velocity

produced by the nozzle with curved inlet section compared to conical section. Further it is seen from Figure 4(a) and Figure 4(b) that the jet velocity increases with the reduction in radius of curvature from straight profile. The nozzle with profile radius 20 mm is found to produce maximum jet velocity compared to other geometries having 40 mm and 60 mm profile radius. Variation of the profile below 20 mm radious did not show any appreciable variations in the jet flow. The higher the radious of curvature, the higher the velocity gradiants are, leading to the corresponding flow loss particularly at the critical section. The nozzle with 20 mm profile radious produced smooth transion in the jet flow which is almost parallel to wall surface of the nozzle, thus producing peak jet velocity.

Effects of nozzle profile on jet force

The machinability of the material depends on cutting or jet force developed by the nozzle. Hence, in this study the cutting force developed by nozzles with different profiles were analysed at different operating pressures ranging from 50 MPa to 500 MPa. Figure 5 shows the cutting force developed by various nozzles corresponding to different inlet pressures. The cutting force is found to be increased linearly with the increase in inlet pressure. This is due to the increase in kinetic energy which increases with operating pressure (Deepak et al., 2011). Further, it is seen that, the cutting force is increasing with the decrease in the profile radius of the nozzle for the same operating conditions. The jet force developed in the nozzle depends on exit velocity. As explained earlier, nozzle with profile radius 20 mm produce higher jet velocity, the cutting force developed is also higher.

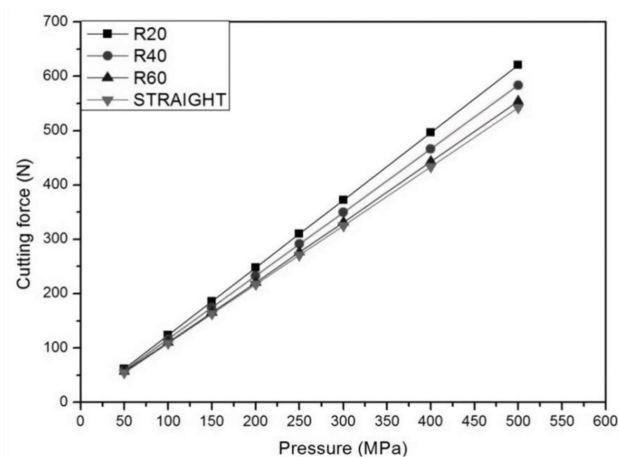


Figure 5. Cutting force distribution for various nozzle profiles

Effects of nozzle profile on pressure drop

The total pressure change represents the potential energy loss in the nozzle. Figure 6 shows the pressure drops between jet entry and exit section of the nozzle for different geometric models at inlet pressure ranging between 50 to 500 MPa. It is observed that the profile of the

nozzle greatly affects the jet flow. The pressure drop is found to increase with the increase in operating pressure for all geometries of the nozzle considered in the study. This is due to a higher wall shear stress at a higher operating pressures. Further, nozzle with profile radius 20 mm exhibits lower pressure drop compared to other models due to smooth transition of flow in the nozzle which resulted in generating higher jet force. Nozzles which have straight converging section and profile with increased radius leads to jet dispersion and hence resulting in a higher pressure drop.

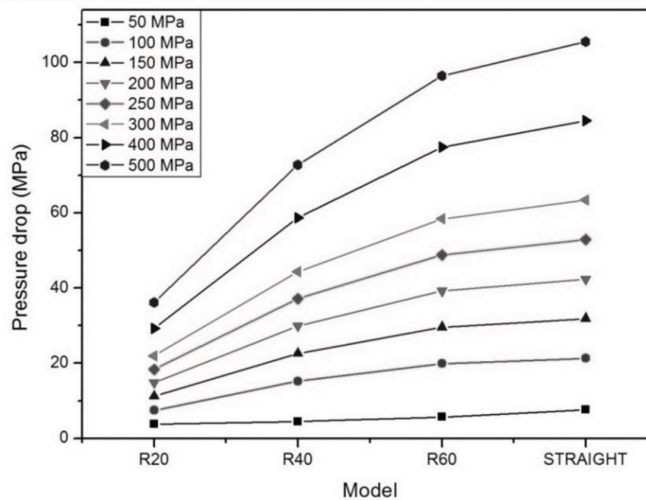


Figure 6. Pressure drop between jet entry and exit section

CONCLUSIONS

A computational analysis has been carried on the jet flow through AWJ nozzle with different nozzle geometries. Based on the simulated results, the following conclusions are drawn. The geometric variation of nozzle profiles shows that, reduction in radius of curvature (radius 20 mm) of nozzle geometry produced higher jet velocity and cutting force as well as lower pressure drop compared with other geometric dimensions. The study also shows that irrespective of the nozzle geometry the jet velocity and force increased with increase in inlet pressure and it remains almost constant in the focus section of the nozzle. Hence, it is recommended to manufacture AWJ nozzle with the profiles having curvature against the current practice of having straight section.

ACKNOWLEDGEMENT

The authors would like to acknowledge the laboratory support rendered by Manipal Institute of Technology, Manipal University, Manipal, India.

REFERENCES

- Anand, U., & Katz J. (2003). Prevention of Nozzle Wear in Abrasive Water Suspension Jets Using Porous Lubricated Nozzles. *Journal of Tribology*, 125(1), 168-180.
- Deepak, D., Anjaiah, D., & Sharma, N. Y. (2011) Numerical Analysis of Flow through Abrasive Water Suspension Jet: The Effect of Abrasive Grain Size and Jet Diameter Ratio on Wall Shear. *International Journal of Earth Sciences and Engineering*, 04, 78-83.
- Deepak, D., Anjaiah, D., Karanth, K. V., & Sharma, N. Y. (2012). CFD Simulation of Flow in an Abrasive Water Suspension Jet the Effect of Inlet operating Pressure and Volume Fraction on Skin Friction and Exit Kinetic Energy. *Advances in Mechanical Engineering*, 4, 1-8.
- Deepak, D., Anjaiah, D., & Sharma, N. Y. (2012). Numerical Analysis of Flow through Abrasive Water Suspension Jet the Effect of Garnet, Aluminium Oxide and Silicon Abrasive on Skin Friction Coefficient Due To Wall Shear and Jet Exit Kinetic energy. *Journal of World Academy of Science, Engineering and Technology*, 6(10), 2178-2183
- Deepak, D., Anjaiah D., & Sharma, N. Y. (2011). Numerical Analysis of Flow through Abrasive Water Suspension Jet: The Effect of Inlet Pressure on Wall Shear and Jet Exit Kinetic Energy. In *World Congress Engineering-ICME-2011*. London, U.K.
- Deng, J., Zhang, X., Niu, P., Liu, L., & Wang, J. (2005). Wear of ceramic nozzles by dry sand blasting. *Tribology International*, 39(3), 274-280.
- Hu, G., Zhu, W., Yu, T., & Yuan, J. (2008, July). Numerical simulation and experimental study of liquid-solid two-phase flow in nozzle of DIA jet. In *International Conference on Industrial Informatics, 2008. INDIN 2008*. 6th IEEE (pp. 1700-1705). IEEE.
- Kosmol, J. & Hassan, A. I. (2001). Dynamic elastic and plastic analysis of 3-D deformation in AWJ machining. *Journal of Materials Processing Technology*, 113(1-3), 337-341.
- Mabrouki, T., Raissi, K., & Cornier, A. (2000). Numerical simulation and experimental study of the interaction between a pure high-velocity waterjet and targets: contribution to investigate the decoating process. *Wear*, 239(2), 260-273.
- Momber, A. W., & Kovacevic R. (1998). *Principles of Abrasive Water Jet Machining*. UK: Springer-Verlag London Ltd.
- Nanduri, M., Taggart, D. G., & Kim, T. J. (2002). The effects of system and geometric parameters on abrasive water jet nozzle wear. *International Journal of Machine Tools and Manufacture*, 42(5), 615-623.
- Pi, V. N., & Tuan, N. Q. (2009). A study on nozzle wear modeling in abrasive waterjet cutting. In *Advanced Materials Research* (Vol. 76, pp. 345-350). Trans Tech Publications.
- Shimizu, Z., Shinoda, Y., & Peng, G. (2008). Flow characteristics of water jet issuing from a fan jet nozzle. In *Proceedings of 19th International conference on water jetting*. Nottingham, U.K.
- Yang, F. C., Shiah, S. W., & Heh, T. Y. (2009). The effect of orifice lead cutting edge distance and fluid viscosity on jet performance in high-velocity waterjet cutting systems. *The International Journal of Advanced Manufacturing Technology*, 40(3), 332-341.
- Yuying, J., Liehang, G., Runlian, J., & Xinlin, X. (2013). Influence of Nozzle Grid Precision in Abrasive Water Jet Numerical Simulation. *Advances in Computer Science and its Applications (ACSA)*, 2(3), 386-389.





An Approach for Identification of Copy-Move Image Forgery based on Projection Profiling

Mohd Dilshad Ansari^{1*}, Satya Prakash Ghrera¹ and Mohd Wajid^{2#}

¹Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat, Solan 1732 34, India

²Department of Electronics and Engineering, Jaypee University of Information Technology, Waknaghat, Solan 1732 34, India

ABSTRACT

An image forgery is a common problem which causes the negative impact on society. In the earlier period it did not affect the general public because the sophisticated image processing software and editing tools were not easily available. Thus, the rapid growth of the image processing software makes this task pretty easy. If it is done with care then it is very difficult for humans to recognize visually whether the image is original or forged. Therefore, the authenticity of an image is a necessity of today's digital era. The copy-move image forgery is the most common type of image forgery in which an area or object is copied and pasted at some other places within the same image in order to hide some important features of the image. In this paper, we have proposed copy-move image forgery detection technique based on the image projection profiling. First, we convert the input image into binary image. The horizontal and vertical projection profiles, which are used in estimating the rectangular regions of copy-move image forgery, are then calculated. The experimental result shows that the proposed approach is able to detect copy-move region successfully and significant improvements have been suggested in computational time compared to other reported algorithms. The performance of proposed approach is demonstrated on various forged images.

Keywords: Copy-move forgery detection, image forgery, image projection profiling, tampering

Article history:

Received: 29 April 2016

Accepted: 30 August 2016

E-mail addresses:

m.dilshadse@gmail.com (Mohd Dilshad Ansari),

sp.ghrera@juit.ac.in (Satya Prakash Ghrera),

wajidiitd@ieee.org (Mohd Wajid)

*Corresponding Author

Author's Current Affiliation:

Department of Electronics Engineering,

Zakin Husain College of Engineering and Technology,

Aligarh Muslim University, 202002 Aligarh, India

INTRODUCTION

If image forgery is done with care without leaving any traces, then it is very difficult for a human visionary system to recognise whether the digital image is original or forged. The fast growth of the digital image processing software and internet makes this task pretty easy where anybody can easily

doctored digital image with the help of these available tools (Ansari & Ghrera, 2014). This tendency indicates serious vulnerabilities and decreases the trustworthiness of the digital images. Therefore, new techniques should be developed to authenticate the integrity and the genuineness of digital images. This is extremely important in today’s digital era, especially taking into consideration that these images can be presented as evidences in a court of law, as news items, as parts of a medical records, as financial documents, or can be used in some other significant places. Therefore, digital image forgery detection is critical.

The most common type of image forgery is copy-move image forgery that is frequently used for doctoring digital image. In this type of image forgery, an object or an area of an image is copied and pasted onto another part of the same image to cover some important features of the image. The task of detecting image forgery becomes more complicated because the copied area will have similar characteristics of the image such as noise component, color palette, texture, etc. This indicates that detection approaches that search for copy-move forgery regions using inconsistencies in statistical measures will fail. There are a number of approaches that give solutions for copy-move image forgery detection. Each of these approaches gives a solution under a set of circumstances or assumptions; these techniques will not work if these assumptions are not realised (Fridrich et al., 2003; Huang et al., 2008; Lin et al., 2011).

Digital image forgery detection techniques are categorised into two approaches: active and passive. In the active approach, the digital image requires preprocessing of image such as watermark embedding or signature generation, which limits their application in practice (Ansari & Ghrera, 2014; Farid 2009). Unlike the watermark and signature-based methods, the passive techniques do not need any digital signature to be generated or to embed any watermark. The passive or blind image forgery detection techniques roughly can be divided into five categories as shown in Figure 1. These are (a) Pixel-based techniques that detects statistical anomalies introduced at the pixel level, (b) the format-based techniques leverage the statistical correlations introduced by a specific lossy compression scheme, (c) camera-based techniques exploit artifacts introduced by the camera lens, sensor, or on-chip post-processing, (d) the physical environment-based techniques explicitly model and detect anomalies in the three-dimensional interaction between physical objects, light, and the camera, and (e) the geometry-based techniques make measurements of objects in the world and their positions relative to the camera (Ansari & Ghrera, 2014; Farid 2009).

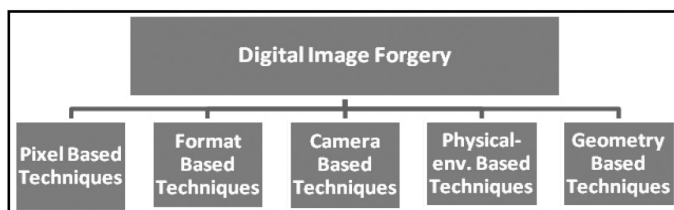


Figure 1. Types of image forgery techniques

The projection profiling of an image contains significant characteristics of an image, which can be used in detecting and estimating many features of an image. In addition to that, it is less computational compared to other image characterising techniques. The image profiling

has been used in recognising Braille character pattern (Wajid et al., 2011; Wajid & Kumar 2014). In this paper, we have also used image projection profiling, which helps in detecting copy-move forgery regions in an image.

The remaining part of the paper is organised as follows. In the Background section, we discuss the various copy-move image forgery detection techniques. In the Proposed Method and Algorithm section, we have proposed an algorithm for detecting copy-move image forgery. In the Result and Discussion section, the results are discussed. Finally, the Conclusion section provides the conclusion of this paper.

BACKGROUND

In the direct method for detecting forgeries the image is divided into small overlapping blocks and then all the blocks are compared with one another. If two blocks are found to be similar, then these blocks represent potential forgeries. Suppose the image contains $n \times n$ pixels, the block size $b \times b$ pixels and the blocks are compared pixel by pixel, then there are $((n-b + 1) \times (n-b)/2) \times b^2$ comparisons.

If $n = 256$ and $b = 16$, then there are 7.4 million comparisons. That is why these direct approaches are too slow for detecting forgeries from the digital images. In order to make the computation fast and overcome direct method for detecting forgeries, Fridrich et al. (2003) have proposed an effective block-based method, in which the image blocks are represented by quantized Discrete Cosine Transform (DCT) coefficients and a lexicographic sort is applied to detect the similar blocks. Popescu and Farid (2004) have developed a similar approach, applied the Principal Component Analysis (PCA) to yield a reduced dimension representation and computed eigenvalues, eigenvectors of each block for detecting copy-move regions. Ansari and Ghrera (In Press) have proposed noble method for extracting texture feature from the input image, further extracted feature vector can be used to detect copy-move forgery detection. Li et al. (2007) have presented a sorted neighborhood method based on Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD). In this approach, first DWT is applied to the input image and then SVD is used on low-frequency components to reduce their dimensions. Singular value vectors are then lexicographically sorted, similar blocks will be nearby in the sorted list and forgery region is detected. The DWT and SVD methods suffer from the disadvantage that the computation of SVD takes lot of time and it is computationally complex. Bayram et al. (2009) used a scale and rotation invariant Fourier Mellin Transform (FMT) and the notion of bloom filters to detect copy-move image forgery. Their approach is computationally proficient and able to detect forgery in highly compressed images.

Small amounts of work have been reported on copy-move image forgery detection based on “*key point matching*”. Huang et al. (2008) have proposed Scale Invariant Feature Transform (SIFT) descriptors for detecting copy-move forgery detection. After extracting the descriptors of different regions, they match them with each other to find possible forgery region in the images. Lin et al. (2011) proposed an integrated method for copy-move forgery and splicing detection. First, image is converted into the YCbCr color space. For splice detection, the image is divided into sub-blocks and DCT is used for feature extraction, SURF is used for copy-move detection.

Most of the above methods suffer from false positives and are not computationally efficient (Fridrich et al., 2003; Kang et al., 2010; Wang et al., 2011). Therefore, the goal of the image forensics is to provide accurate results with less computational time. We have proposed an efficient algorithm for detecting copy-move region successfully with a tool, i.e. projection profiling, which suggests a new improvement in computational time, as shown in Table 1.

PROPOSED METHOD AND ALGORITHM

The algorithm of the proposed approach is given below.

1. Convert an Image **I** (color/gray) of order (M, N) into binary image, **J**.
2. Let $h(j)$ and $v(i)$ be the horizontal and vertical projection profile of **J** respectively, they are defined as:

$$h(j) = \sum_{i=1}^M J(i, j)$$

and,

$$v(i) = \sum_{j=1}^N J(i, j).$$

3. $r(n) = \frac{1}{L} \sum_{l=0}^{L-1} \delta(n-l)$. Consider a moving average filter of order L with impulse response assume, $L \ll \min(M, N)$

4. Pass the horizontal and vertical projection profile h and v through this filter and let h' and v' be the outputs of this filter as given below,

$$h'(j) = \sum_{k=1}^L r(k)h(j-k)$$

and,

$$v'(i) = \sum_{k=1}^L r(k)v(i-k)$$

5. Find the means of these outputs of the filter

$$m_h = \frac{1}{N} \sum_{j=1}^N h'(j)$$

$$m_v = \frac{1}{M} \sum_{i=1}^M v'(i)$$

6. Search for the corner indexes of the rectangular regions of the copy-move part.

Let, $h_s = h_e = v_s = v_e = \phi$, where ϕ is a null set.

for $j \rightarrow 1:N$

if $h'(j) < m_h$

$h'(j) = 1$

else

$h'(j) = 0$

$h''(j) = h'(j) - h'(j-1)$

if $h''(j) = 1$

$hs = [hs, j]$

if $h''(j) = -1$

$he = [he, j]$

end

for $i \rightarrow 1:M$

if $v'(i) < m_v$

$v'(i) = 1$

else

$v'(i) = 0$

$v''(i) = v'(i) - v'(i-1)$

if $v''(i) = 1$

$vs = [vs, i]$

if $v''(i) = -1$

$ve = [ve, i]$

end

7. Identify the rectangular regions of the copy-move part of the image **I**

$\mathbf{R}_1 = \mathbf{I}(hs(1,1): he(1,1), vs(1,1): ve(1,1))$

and,

$\mathbf{R}_2 = \mathbf{I}(hs(1,2): he(1,2), vs(1,2): ve(1,2))$

Where, \mathbf{R}_1 and \mathbf{R}_2 are the matched regions.

RESULT AND DISCUSSION

The experimental results were carried out on MATLAB, 4GB RAM and core i3 processor. To check the robustness and feasibility of proposed algorithm, we have tested algorithm on more than 20 forged images with various image sizes, 128×128 , 256×256 and 512×512 , etc.

Figure 2(a) shows original image which is forged by copy-move techniques shown in Figure 2(b), where an object is copied and pasted on the same image at some other location. The forge image is converted into binary image; thereafter the horizontal and vertical projection profile of image is taken, shown in Figures 3 and 4, respectively. Further, moving average filter has been applied on these projection profiles, which has smoothen these curves (also shown in Figure 3 and Figure 4). From the filtered horizontal projection profile, the x-coordinate of the corners of the rectangle boxes have been calculated, as given in algorithm. Similarly, the y-coordinate has been calculated using filtered vertical projection profile. The algorithm is also evaluated on other forged images as shown in Figures 5-16. The proposed technique is robust even though multiple objects are available in the image, which is based on the repetitions of the valleys in the projection profile. From the projection profile, we can identify the number of copy-move objects in the forged image. If another distinct object is available in the image, it can contribute to a valley in the projection profile but will not match with any other valley in the same projection profile. However, the proposed technique is limited to the case when the object is having contrast to the background of the image. The computational time efficiency is compared with the existing method shown in Table 1. It has been shown that computation time is relatively less for the proposed method for different sizes of images. For the proposed algorithm, there is a tremendous improvement in the computation time for the bigger image sizes.



Figure 2. (a) Original image; (b) Forged image; (c) Detected region of copy-move forgery

Copy-Move Image Forgery Identifications based on Projection Profiling

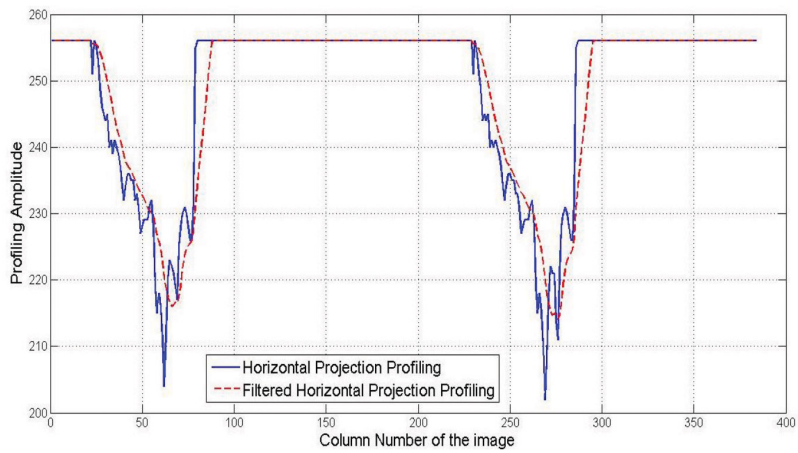


Figure 3. Horizontal projection profile and filtered horizontal projection profile for the image shown in Figure 2(b)

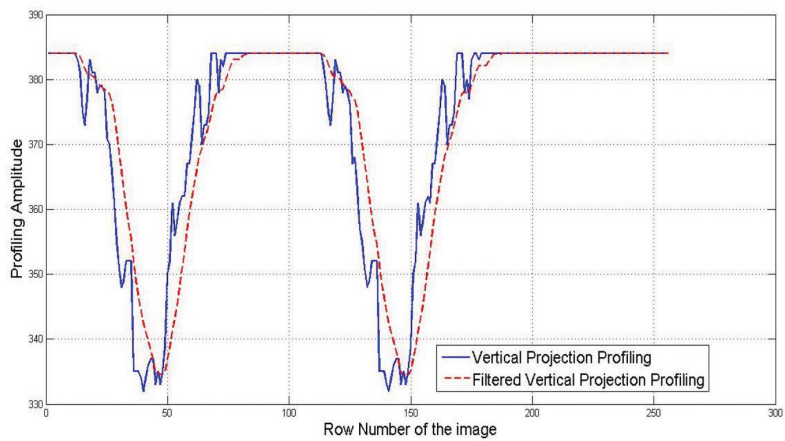


Figure 4. Vertical projection profile and filtered vertical projection profile for the image shown in Figure 2(b)



Figure 5. (a) Original image; (b) Forged image; and (c) Detected region of copy-move forgery

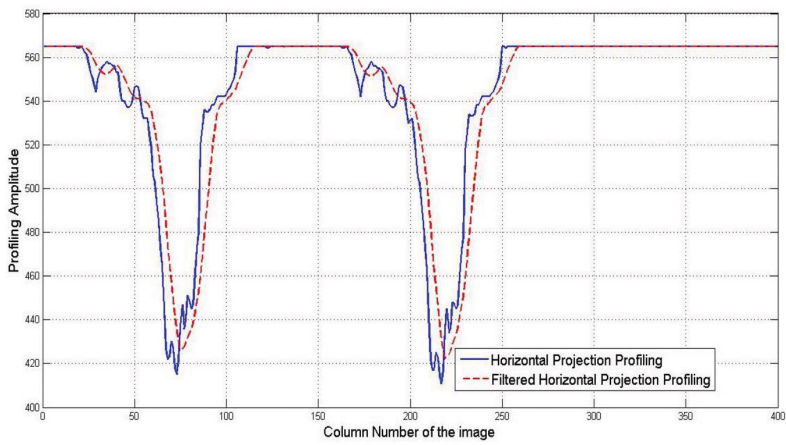


Figure 6. Horizontal projection profile and filtered horizontal projection profile for the image shown in Figure 5(b)

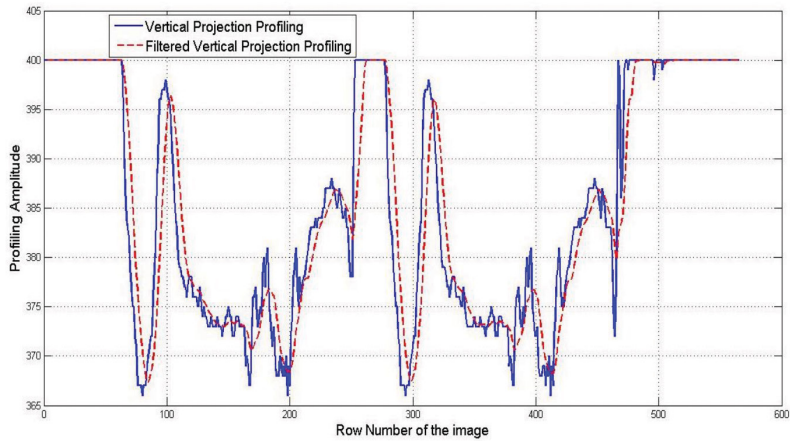


Figure 7. Vertical projection profile and filtered vertical projection profile for the image shown in Figure 5(b)

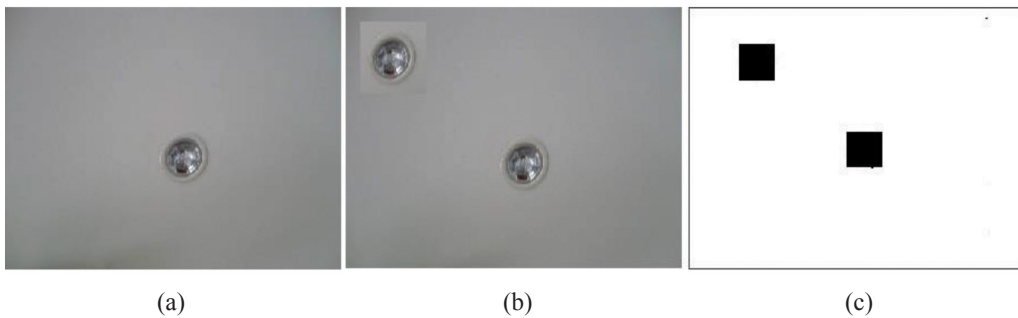


Figure 8. (a) Original image; (b) Forged image; and (c) Detected region of copy-move forgery

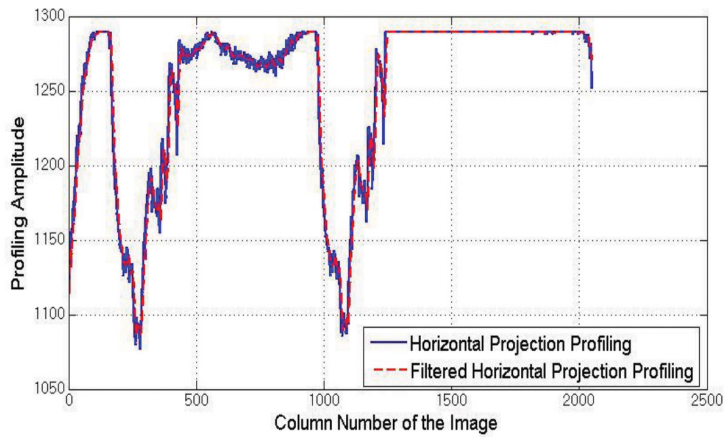


Figure 9. Horizontal projection profile and filtered horizontal projection profile for the image shown in Figure 8(b)

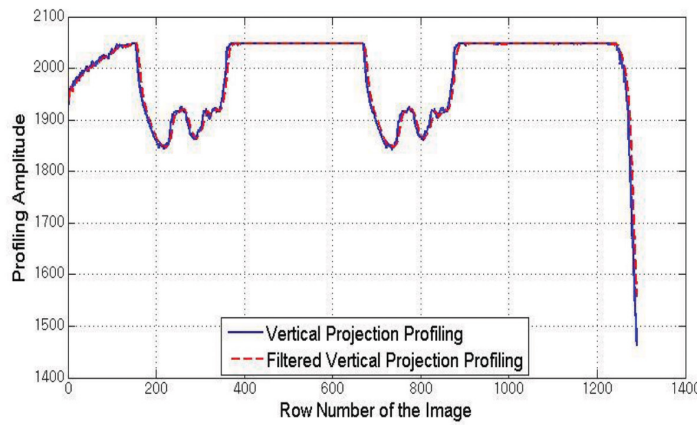


Figure 10. Vertical projection profile and filtered vertical projection profile for the image shown in Figure 8(b)

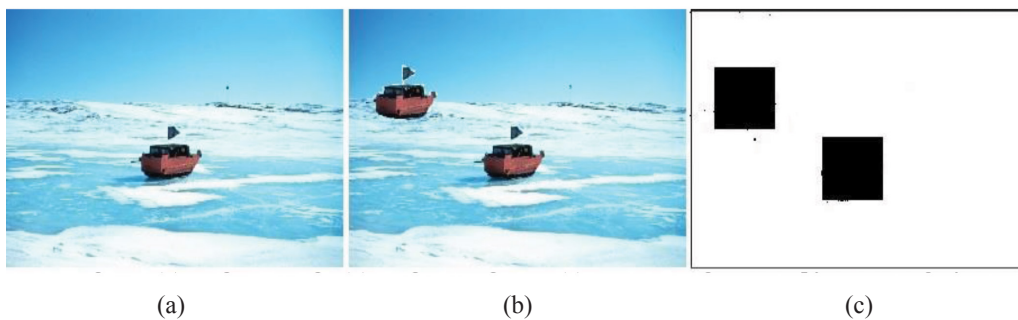


Figure 11. (a) Original image; (b) Forged image; and (c) Detected region of copy-move forgery

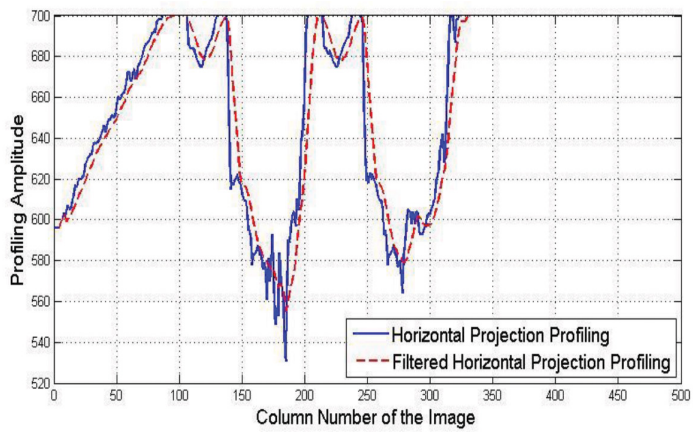


Figure 12. Horizontal projection profile and filtered horizontal projection profile for the image shown in Figure 11(b)

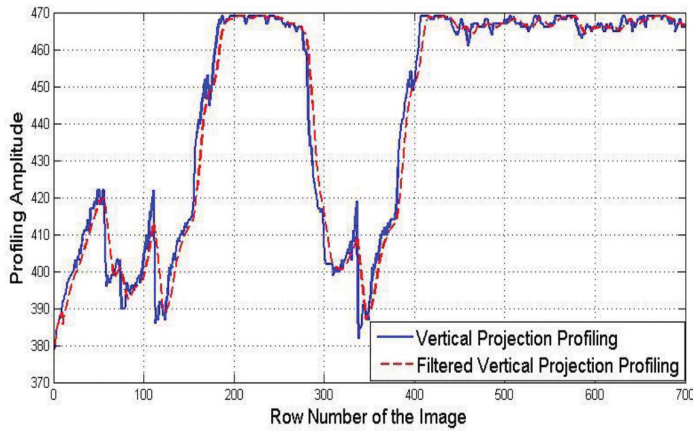


Figure 13. Vertical projection profile and filtered vertical projection profile for the image shown in Figure 11(b)

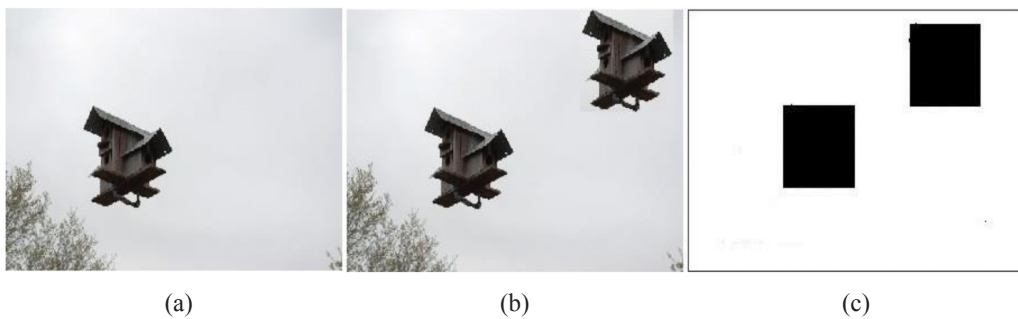


Figure 14. (a) Original image; (b) Forged image; and (c) Detected region of copy-move forgery

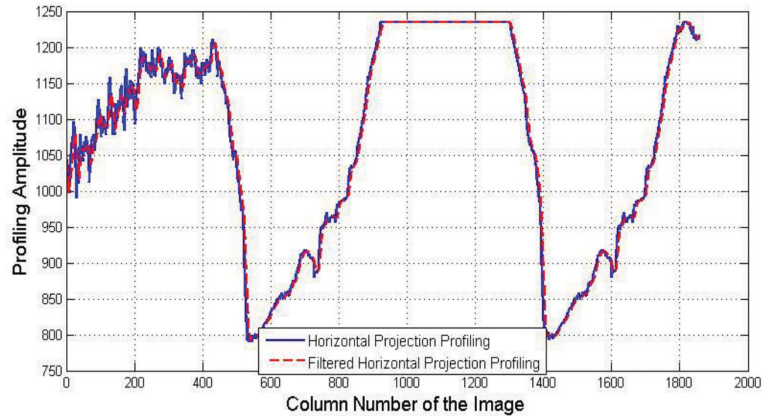


Figure 15. Horizontal projection profile and filtered horizontal projection profile for the image shown in Figure 14(b)

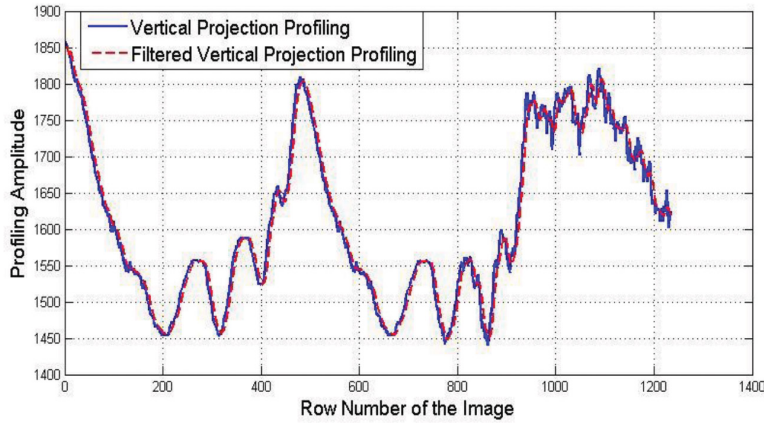


Figure 16. Vertical projection profile and filtered vertical projection profile for the image shown in Figure 14(b)

Table 1
Copy-move forgery detection execution time in seconds

Image Name	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7
Image Size	128×128	256×256	384×384	424×424	512×512	656×656	832×832
Xiaofeng Wang et al.	4.96	39.70	171.09	239.48	478.109	1259.35	1308.25
Proposed Method	3.29	4.27	4.81	5.33	5.69	6.12	6.84

CONCLUSION

We have proposed a new method for image forensics based on projection profiling. As the block-based copy-move forgery detection methods are computationally intensive any improvement in execution time will help in checking forgery from the given image. The proposed method

successfully addresses this issue and is noticeably faster than the existing methods. To check the robustness of our proposed approach it has been performed on various forgery images. The results show that the proposed approach is able to detect copy-move region effectively and new improvements have been suggested in computational time shown in Table 1. In the future, we can improve the copy-move forgery detection approach by applying various attacks such as rotation, scaling, added Gaussian noise, JPEG Compression, etc.

REFERENCES

- Ansari, M. D., Ghrera, S. P., & Tyagi, V. (2014). Pixel-Based Image Forgery Detection: A Review. *IETE Journal of Education*, 55(1), 40-46.
- Ansari, M. D., Ghrera, S. P. (n.d.). Intuitionistic fuzzy local binary pattern for features extraction. *International Journal of Information and Communication Technology*. In Press
- Bayram, S., Sencar, H. T., & Memon, N. (2009, April). An efficient and robust method for detecting copy-move forgery. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 1053-1056). IEEE.
- Fridrich, A. J., Soukal, B. D., & Lukáš, A. J. (2003). Detection of copy-move forgery in digital images. In *in Proceedings of Digital Forensic Research Workshop*.
- Farid, H. (2009). Image forgery detection. *Signal Processing Magazine, IEEE*, 26(2), 16-25.
- Huang, H., Guo, W., & Zhang, Y. (2008, December). Detection of copy-move forgery in digital images using SIFT algorithm. In *Computational Intelligence and Industrial Application, 2008. PACIIA'08. Pacific-Asia Workshop on* (Vol. 2, pp. 272-276). IEEE.
- Kang, Li., & Cheng, X. P. (2010). Copy-move forgery detection in digital image. In *Image and Signal Processing (CISP), 2010 3rd International Congress on* (pp. 2419-21). IEEE.
- Li, G., Wu, Q., Tu, D., & Sun, S. (2007, July). A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD. In *Multimedia and Expo, 2007 IEEE International Conference on*(pp. 1750-1753). IEEE.
- Lin, S. D., & Wu, T. (2011, October). An integrated technique for splicing and copy-move forgery image detection. In *Image and Signal Processing (CISP), 2011 4th International Congress on* (Vol. 2, pp. 1086-1090). IEEE.
- Popescu, A. C., & Farid, H. (2004). *Exposing digital forgeries by detecting duplicated image region* [Technical Report 2004-515]. Hanover, Department of Computer Science, Dartmouth College, USA.
- Wajid, M., Abdullah, M. W., & Farooq, O. (2011, November). Imprinted Braille-character pattern recognition using image processing techniques. In *Image Information Processing (ICIIP), 2011 International Conference on* (pp. 1-5). IEEE.
- Wajid, M., & Kumar, V. (2014). E-Braille Documents: Novel Method For Error Free Generation. *Image Processing and Communications*, 19(4), 21-26.
- Wang, X., Zhang, X., Li, Z., & Wang, S. (2011, November). A DWT-DCT based passive forensics method for copy-move attacks. In *2011 Third International Conference on Multimedia Information Networking and Security* (pp. 304-308). IEEE.



Novel Impedance Measurement Technique for Soluble Solid Content Determination of Banana

Ibrahim, N. U. A., Abd Aziz, S.* and Nawi, N. M.

¹Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Soluble solid content (SSC) is one of the important traits that indicate the ripeness of banana fruits. Determination of SSC for banana often requires destructive laboratory analysis on the fruit. An impedance measurement technique was investigated as a non-destructive approach for SSC determination of bananas. A pair of electrocardiogram (ECG) electrode connected to an impedance analyser board was used to measure the impedance value of bananas over the frequency of 19.5 to 20.5 KHz. The SSC measurement was conducted using a pocket refractometer and data was analysed to correlate SSC with impedance values. It was found that the mean of impedance, Z decreased from 10.01 to 99.93 $K\Omega$ at the frequency of 20 KHz, while the mean value of SSC increased from 0.58 to 4.93 % Brix from day 1 to day 8. The best correlation between impedance and SSC was found at 20 KHz, with the coefficient of determination, R^2 of 0.87. This result indicates the potential of impedance measurement in predicting SSC of banana fruits.

Keywords: Banana ripeness, impedance measurement, non-destructive, soluble solid content (SSC)

INTRODUCTION

Banana (*Musa paradisiaca L.*) is the most preferred fruit for healthy eating diet. A finger of a banana provides 23% of the potassium that we need on a daily basis and it also provides an excellent source of vitamin A, B6, C and D (Kumar et al., 2012). The nutritional values of bananas vary between degrees of ripeness. Bananas also rich in carbohydrates and in unripe stage, bananas mostly starches. The starches converted into sugar during ripening, a fully ripe banana has only 1-2% starch (Forsyth, 1980).

In the post-harvest production, banana is cut at the mature-green stage to preserve a firm pulp texture, good colour and flavour (Soltani et al., 2010) and then the fruit is artificially ripened under ethylene treatment. In this case, quality assessment is important to make sure that the bananas are ready to

Article history:

Received: 26 April 2016

Accepted: 13 September 2016

E-mail addresses:

samsuzana@upm.edu.my (Ibrahim, N. U. A.),

nurulatikah.ibrahim@gmail.com (Abd Aziz, S.),

nazmimat@upm.edu.my (Nawi, N. M.)

*Corresponding Author

be marketed or exported. Soluble solid content (SSC) is often used as a quality indicator for bananas. Usually, SSC of banana is increased as it reaches the ripening stage. Traditionally, determination of SSC requires a flesh of banana fruit for destructive measurement using refractometer (Harker et al., 2002). The destructive test is not practical for quality assessment because it will cause damage to the fruit.

Therefore, many studies were done to replace the destructive test in SSC determination. Several non-destructive methods to assess the quality of bananas optical chlorophyll sensing system (Li et al., 1997), neural network-based electronic nose (Llobet et al., 1999) and near infrared (NIR) spectroscopy (Liew & Lau, 2012). Li et al. (1997) studied the use of optical chlorophyll sensing system to measure the chlorophyll content of banana peels and the results showed a high correlation to other peel colour analysis such as spectral analysis. However, this sensing system is not reliable for predicting peel colour changes of banana at the ripeness stage above four. Llobet et al. (1999) reported that fuzzy-adaptive resonance theory mapping (ARTMAP) based electronic nose provides an attractive means for classifying the ripeness of bananas. This system was able to predict the state of ripeness of unknown sets of bananas with 90% accuracy. Liew and Lau (2012) used NIR spectroscopy combined with multiple linear regression (MLR) to determine the ripeness stage of Cavendish bananas and found that NIR had the potential for non-destructive determination of internal quality of Cavendish bananas such as SSC and firmness.

In addition, some researchers proposed the use of nuclear magnetic resonance (NMR) and proton magnetic resonance (PMR) as non-destructive methods to determine levels of soluble solid (Cho et al., 1990; Wai et al., 1995). However, all the studies are only fundamental approaches that are far from reaching the practical application and some of the techniques, like NMR and PMR, rely upon expensive equipment and required skilled operators.

Since 1980's, dielectric properties measurement of several fruits and vegetables have been studied as a non-destructive approach. Nelson (1980) measured dielectric properties of peaches at microwave frequency and found that no differences in dielectric properties at 2450 MHz were distinguishable between mature green and full-ripe peaches. Then, permittivity-based maturity indices of peaches, which combine the dielectric constant values at 200 MHz and the loss factor values at 10 GHz, were suggested; however, a lot more work is required to establish their practicality (Nelson et al., 1995). A linear relationship complex-plane plot for permittivity components divided by SSC for external surface measurements at 40 MHz and 24°C was found on watermelon (Nelson et al., 2007). However, it is not useful in predicting SSC values from the dielectric properties because of the poor relationship between SSC and the dielectric properties of R^2 of 0.4.

Another study on the feasibility of capacitance measurement for under ripe and ripe banana was carried out using parallel plate capacitor at different plate separation by Zulhusin et al. (2008). They found that the difference between maturity levels depends on the plate separation and the best separation of the plate was found at 4 cm. Later, Soltani et al. (2010) also investigated the capacitive property of banana fruits by using parallel plate capacitor and they found that permittivity for green-ripe and full-ripe banana was changed from 1.74 to 1.64, respectively, at 1 MHz. Linear regression at 1 MHz for both SSC and firmness with relative permittivity provides an acceptable correlation in predicting banana quality during the ripening treatment.

In recent decades, electrical impedance measurement has been broadly used to study the electrical properties of plant tissue (Zywica et al., 2005; Liu, 2006; Fang et al., 2007). Electrical Impedance Spectroscopy (EIS) is suitable for non-destructive techniques which provide easy, inexpensive and rapid assessment (Bauchot et al., 2000; Zywica et al., 2005; Juansah et al., 2012; Caravia et al., 2015). Electrical impedance measurement has been used in quality assessment of agriculture produce to determine the soluble solid content of fruits (Zywica et al., 2005; Fang et al., 2007; Guo et al., 2011), or evaluate the bruising damage in fruits, assess the moisture and salt content in smoked fish (Karásková et al., 2011), evaluate the quality of meat (Freywald et al., 1995) and determine the freshness of fish and shrimp (Marshall & Wiese-Lehigh, 1997; Niu & Lee, 2000).

Therefore, this study aims to use impedance spectroscopy technique as a potential measurement method in assessing SSC of bananas. The objectives of this study are to investigate the electrical impedance of banana from *Berangan* variety during ripening and to find the correlation between impedance and SSC measurement.

MATERIAL AND METHODS

Banana samples preparation

Six hands of green bananas, *Musa paradisiaca L.*, from *Berangan* variety were purchased from a local market at ripening stage one with approximately 15 or 20 fingers in each hand. The bananas were cleaned with tap water to remove any foreign material from its skin and dried by using tissues. After that, they were kept in a box at a room temperature of 26°C for eight days. Each day, four fingers of bananas were taken randomly from the box and their impedance and SSC were measured. A total of 32 fingers of banana (4 fingers of banana x 8 days = 32 fingers) were used for impedance and SSC measurements.

Impedance measurement

The impedance measurement was carried out using a pair of electrocardiogram (ECG) electrodes which were connected to an impedance analyser board, AD5933 (Analog Devices Inc., MA, USA) and a computer. The ECG electrodes were composed of a small metal plate surrounded by an adhesive pad, which was coated with conducting gel to help transmit electrical signal (Figure 1). Impedance measurement was obtained along a sweep frequency from 19.5 to 20.5 KHz. The magnitude and phase of the impedance at each frequency point along the sweep were calculated. The range of sweep frequency from 19.5 to 20.5 KHz was chosen after several trials were done and this sweep frequency was found to provide sensitive reading to distinguish the level of SSC content in banana fruit. The impedance converter system was connected to a computer and data were acquired using AD5933 evaluation software (Analogue Devices Inc., MA, USA). The ECG electrodes were discarded after each measurement. The measurements were conducted at a room temperature of 26°C. After completion of the impedance measurements, the SSC measurements were performed.

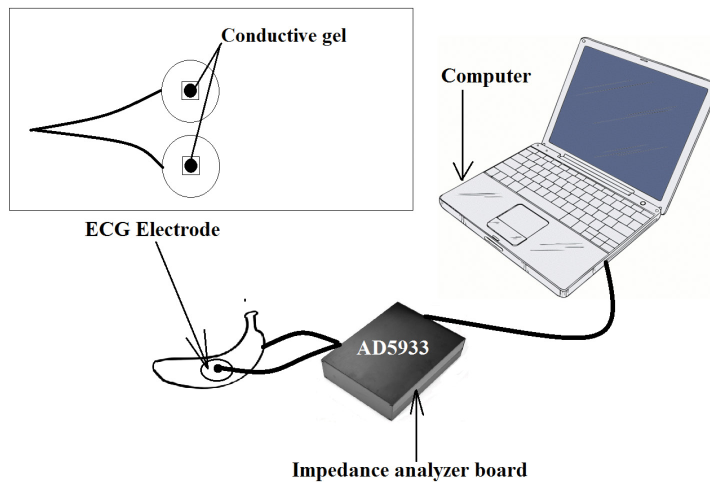


Figure 1. Experimental set-up for impedance measurement

SSC measurement

SSC value of each finger of banana was determined from its juice. For each banana finger, the juice was expressed from 10 g of banana flesh using a blender. The flesh was blended with 40 ml distilled water for one minute until sufficiently homogenised and then the juice was collected in a 50 ml beaker. The SSC was measured with a Pocket Refractometer PAL- α (Atago Co., Ltd, Tokyo, Japan) by placing a drop of the juice on the prism glass. The refractometer readings are referred to as Brix readings, which expressed in percent total solids by weight (Harker et al., 2002). Three replicate of SSC readings were recorded for each finger of banana.

Statistical analysis

Relationships between impedance and SSC were analysed using Analysis ToolPak option in Microsoft Excel 2007 (MS Corporation, Redmond, WA, USA). Significant differences of impedance and SSC during eight days of measurement were determined using analysis of variance (ANOVA) and mean separation by Duncan Multiple Range Test (DMRT) at the 5% probability level. This analysis was performed using SAS 9.1 (SAS Institute, Inc., Cary, NC, USA).

RESULTS AND DISCUSSIONS

The mean of impedance and SSC values over the eight days of measurement were shown in Table 3. The SSC values were found to increase steadily day by day. This increment had been expected as SSC usually increases as the banana ripens due to the conversion of starch to soluble sugar. The increase of SSC is an important trait of hydrolysis of starch into soluble sugars such as glucose, sucrose and fructose (Marriot et al., 1981). For impedance, a decrease occurred with the most frequency. Results indicated that impedance decreased when SSC increased during ripening.

Changes in impedance and SSC value during ripening

ANOVA showed that there were highly significant differences found between impedance and SSC during ripening with α level less than 0.0001 (Table 1 and Table 2). To further investigate the differences, mean separation by DMRT was performed. The mean value of SSC and impedance data at five frequencies (19.5, 19.9, 20, 20.1 and 20.5 KHz) over eight ripening days were summarised in Table 3. Based on the results in Table 3, there was a significant difference found in SSC during ripening, except for day 1 and day 2. The mean of SSC increased from 0.58% Brix to 4.93% Brix. The results also clearly indicated that from day 1 to day 8 a decrement trend was found in the mean of impedance from 10.01 to 99.93 K Ω at 20 KHz (Table 3). However, between day 5 and day 6, as well as for day 7 and day 8, there was no significant difference in the mean of impedance (Table 3).

Table 1

ANOVA for impedance measurement

Source of Variation	F	p-value	
Day	21.18	<.0001	***

Note. *** Highly significant difference in the means of impedance during ripening.

Table 2

ANOVA for SSC measurement

Source of Variation	F	p-value	
Day	194.27	<.0001	***

Note. *** Highly significant difference in the means of impedance during ripening.

Table 3

Impedance and SSC measurements over the frequency range from 19.5 to 20.5 KHz during eight days

Day	SSC (%Brix)	Frequency (KHz)				
		19.5	19.9	20	20.1	20.5
1	0.58 a*	10017.80 a	10012.98 a	10011.88 a	10008.03 a	10001.38 a
2	0.78 a	10021.13 a	10010.43 a	10010.20 b	10008.24 a	10000.11 a
3	1.70 b	10015.10 a	10008.39 b	10006.77 c	10003.78 b	9997.56 a
4	2.15 c	10016.84 a	10009.20 b	10005.62 d	10005.33 a	9996.66 a
5	3.05 d	10015.42 a	10006.66 c	10001.67 e	10004.25 b	9997.85 a
6	3.69 e	10015.10 a	10006.33 c	10001.16 e	10004.58 b	9997.19 a
7	4.42 f	10010.07 b	10004.29 d	9996.06 f	10002.59 c	9997.16 a
8	4.93 g	10003.50 c	10001.38 d	9993.98 f	9999.13 c	9997.72 a

Note. *Means within a column followed by a common letter are not significantly different at the 5% probability level

Correlation between impedance and SSC measurement

The conversion of starch during ripening process usually increases the content of soluble sugar in banana. The increase of sugar content in fruit pulp with the loss of water from fruit during ripening enhances the concentration of ions (Liu, 2006). High concentration of ions in the electrolyte of banana leads to the low value of impedance (Zywica et al., 2005). As discussed earlier, SSC was found to have increased while impedance decreased during ripening. Their correlation was investigated using linear regression. The result from linear regression showed that SSC exhibited negative linear correlation with impedance (Figure 2) with the highest R² found at the frequency of 20 KHz (Table 4). This result is similar to the result reported by Fang et al. (2007), where SSC has linear correlation with impedance in predicting the quality of apple from a different variety. The relationship can be expressed in the regression equation, as follows:

$$SSC = -0.22 Z + 2224.48 \quad [1]$$

where Z is the impedance in ohm (Ω). This equation can be used to predict SSC from impedance measurements.

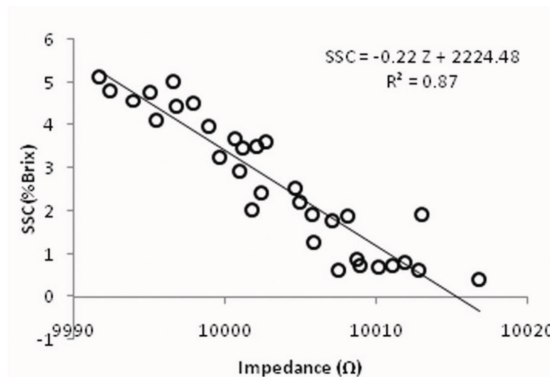


Figure 2. Relationship between impedance and SSC value of bananas at 20 KHz

Table 4

Linear regression equation to predict the SSC of banana fruits using impedance value

Regression equation	R ²	Frequency (KHz)
SSC = -0.20 Z+ 2035.90	0.63	19.5
SSC = -0.30 Z+ 3008.50	0.67	19.9
SSC = -0.22 Z+ 2224.48	0.87	20
SSC = -0.27 Z+ 2700.30	0.46	20.1
SSC = -0.17 Z+ 1739.10	0.13	20.5

CONCLUSION

Measurement of impedance on 32 fingers of banana during eight ripening days provided useful information in developing a non-destructive method for quality assessment of banana. The impedance values from the measurements decreased when the SSC increased from day 1 to day 8. A negative linear relationship was found between impedance and SSC, with $R^2 = 0.87$ at 20 KHz. The correlation between impedance and SSC showed good potential in developing a non-destructive method for SSC determination and online quality sensing. Further research is needed to validate the reliability of the equation.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Biological and Agricultural Engineering, Universiti Putra Malaysia (Selangor, Malaysia) for providing the research facilities. The authors would also like to thank the anonymous reviewers for their invaluable comments and suggestions to improve the quality of this paper.

REFERENCES

- Aziz, A. H. A., & Ahmad, R. B. (2008, December). Feasibility study of a non-destructive fruit maturity testing system on banana utilising capacitive properties. In *Electronic Design, 2008. ICED 2008. International Conference on* (pp. 1-4). IEEE.
- Bauchot, A. D., Harker, F. R., & Arnold, W. M. (2000). The use of electrical impedance spectroscopy to assess the physiological condition of kiwifruit. *Postharvest Biology and Technology*, 18(1), 9–18.
- Caravia, L., Collins, C., & Tyerman, S. D. (2015). Electrical impedance of Shiraz berries correlates with decreasing cell vitality during ripening. *Australian Journal of Grape and Wine Research*, 21(3), 430–438.
- Cho, S. I., Kurtz, G. W., Gibson, H. G., & Haghighi, K. (1990). Magnet console design of an NMR-based sensor to detect ripeness of fruit. *Transactions of the ASAE*, 33(4), 1043–1050.
- Fang, Q., Liu, X., & Cosic, I. (2007). Bioimpedance Study on Four Apple Varieties. In *13th International Conference on Electrical Bioimpedance and the 8th Conference on Electrical Impedance Tomography* (pp. 114–117). Springer Berlin Heidelberg.
- Forsyth, W. G. C. (1980). Banana and Plantain. In S. Nagy & P. E. Shaw (Eds.), *Tropical and Subtropical Fruits* (pp. 258–278). Westport, CT: AVI Publishing.
- Freywald, K., Pliquet, F., Schoberlein, L., & Pliquet, U. (1995). Passive electrical properties of meat as a characterisation of its quality. In *IX International Conference on Electrical Bio-Impedance* (pp. 366–369). Heidelberg.
- Guo, W., Zhu, X., Nelson, S. O., Yue, R., Liu, H., & Liu, Y. (2011). Maturity effects on dielectric properties of apples from 10 to 4500 MHz. *LWT - Food Science and Technology*, 44(1), 224–230.
- Harker, F. R., Marsh, K. B., Young, H., Murray, S. H., Gunson, F. A., & Walker, S. B. (2002). Sensory interpretation of instrumental measurements 2: Sweet and acid taste of apple fruit. *Postharvest Biology and Technology*, 24(3), 241–250.

- Juansah, J., Budiastira, I. W., & Dahlan, K. (2012). The Prospect of Electrical Impedance Spectroscopy as Non-Destructive Evaluation of Citrus Fruit Acidity. *International Journal of Emerging Technology and Advanced Engineering*, 2(11), 58–64.
- Karášková, P., Fuentes, A., Fernández-Segovia, I., Alcañiz, M., Masot, R., & Barat, J. M. (2011). Development of a low-cost non-destructive system for measuring moisture and salt content in smoked fish products. *Procedia Food Science*, 1, 1195–1201.
- Kumar, K. P. S., Bhowmik, D., Umadevi, M., & Duraivel, S. (2012). Traditional and Medicinal Uses of Banana. *Journal of Pharmacognosy and Phytochemistry*, 1(3), 51–63.
- Li, M., Slaughter, D. C., & Thompson, J. F. (1997). Optical chlorophyll sensing system for banana ripening. *Postharvest Biology and Technology*, 12(3), 273–283.
- Liew, C. Y., & Lau, C. Y. (2012). Determination of quality parameters in Cavendish banana during ripening by NIR spectroscopy. *International Food Research Journal*, 19(2), 751–758.
- Liu, X. (2006). *Electrical impedance spectroscopy applied in plant physiology studies*. (Master Dissertation). School of Electrical and Computer Engineering, RMIT University.
- Llobet, E., Hines, E. L., Gardner, J. W., & Franco, S. (1999). Non-destructive banana ripeness determination using a neural network-based electronic nose. *Measurement Science and Technology*, 10(6), 538.
- Marriot, J., Robinson, M., & Karikari, S. K. (1981). Starch and sugar transformation during the ripening of plantains and bananas. *Journal of the Science of Food and Agriculture*, 32(10), 1021–1026.
- Marshall, D. L., & Wiese-Lehigh, P. L. (1997). Comparison of impedance, microbial, sensory, and pH methods to determine shrimp quality. *Journal of Aquatic Food Product Technology*, 6(2), 17–31.
- Nelson, S. O. (1980). Microwave dielectric properties of fresh fruits and vegetables. *Transactions of the ASAE*, 23(5), 1314–1317.
- Nelson, S. O., Forbus, W. R., & Lawrence, K. C. (1995). Assessment of microwave permittivity for sensing peach maturity. *American Society of Agricultural Engineers*, 38(2), 579–585.
- Nelson, S. O., Guo, W. C., Trabelsi, S., & Kays, S. J. (2007). Dielectric spectroscopy of watermelons for quality sensing. *Measurement Science and Technology*, 18(7), 1887–1892.
- Niu, J., & Lee, J. Y. (2000). A new approach for the determination of fish freshness by electrochemical impedance spectroscopy. *Journal of Food Science*, 65(5), 780–785.
- Soltani, M., Alimardani, R., & Omid, M. (2010). Prediction of banana quality during ripening stage using capacitance sensing system. *Australian Journal of Crop Science*, 4(6), 443–447.
- Wai, W. K., Strohshine, R. L., & Kurtz, G. W. (1995). A Modified Hahn Echo Pulse Sequence for Proton Magnetic Resonance (1H-MR) measurements of percent soluble solids of fruits. *Transactions of the ASAE*, 38(3), 849–855.
- Zywica, R., Pierzynowska-Korniak, G., & Wójcik, J. (2005). Application of food products electrical model parameters for evaluation of apple purée dilution. *Journal of Food Engineering*, 67(4), 413–418.



Phase Transition and Dielectric Properties of $0.9\text{Pb}(\text{Fe}_{1/2}\text{Nb}_{1/2})\text{O}_3-0.1\text{PbTiO}_3$ Modified with Nano ZnO

Hassakorn Wattanasarn^{1,2*}, Wattana Photankham^{1,2}, Peerapat Pattumma^{1,2} and Rattikorn Yimnirun³

¹Thermoelectrics Research Centre, Research and Development Institution, Sakon Nakhon Rajabhat University, 680 Nittayo Road, Mueang District, Sakon Nakhon, 47000 Thailand

²Program of Physics, Faculty of Science and Technology, Sakon Nakhon Rajabhat University, 680 Nittayo Road, Mueang District, Sakon Nakhon, 47000 Thailand

³School of Physics, Institute of Science, and COE-NANOTEC-SUT on Advanced Functional Nanomaterials, Suranaree University of Technology, Nakhon Ratchasima, 30000 Thailand

ABSTRACT

Samples of $0.9\text{Pb}(\text{Fe}_{1/2}\text{Nb}_{1/2})\text{O}_3-0.1\text{PbTiO}_3$ were mixed with ZnO at 0, 1, 2, and 3 wt.%, and were synthesised by mixed oxide through a two-step sintering method. Phase transition of the samples was analysed by using X-ray diffractometer (XRD) and Fourier transform infrared spectroscopy (FTIR). The dielectric properties were determined by LCR meter at 1 kHz, 10 kHz, and 100 kHz. The results showed that the morphotropic phase boundary of $0.9\text{PFN}-0.1\text{PT}$ shifting to tetragonal phase and suppressed pyrochlore phase when ZnO was added into $0.9\text{PFN}-0.1\text{PT}$. In addition, FTIR spectra peak showed zinc and oxygen bond bonding vibration at frequencies range $3,452\text{ cm}^{-1}$ and $3,792\text{ cm}^{-1}$ after level doping ZnO of 3 wt.%. The samples exhibited the maximum dielectric constant at temperature for 144°C . The dispersion of dielectric constantly decreased with increasing ZnO contents. The relaxor ferroelectric of $0.9\text{PFN}-0.1\text{PT}$ ceramic shifted to normal ferroelectric with increasing ZnO contents.

Keywords: Dielectric properties, ferroelectric material, FTIR, microstructure, XRD, ZnO

Article history:

Received: 13 May 2016

Accepted: 09 August 2016

E-mail addresses:

w_hussakorn@hotmail.com (Hassakorn Wattanasarn),

dagon_wat@hotmail.com (Wattana Photankham),

peerapatpongpan2522@gmail.com (Peerapat Pattumma),

rattikorn@g.sut.ac.th (Rattikorn Yimnirun)

*Corresponding Author

INTRODUCTION

Many applications in memories and sensors are applied from piezoelectric devices which are produced from relaxor ferroelectric ceramics with lead-based complex perovskite structure, and are favoured by researchers in physics field (Cross, 1996; Vladimir et al., 2003). $\text{Pb}(\text{Fe}_{1/2}\text{Nb}_{1/2})\text{O}_3-\text{PbTiO}_3$ (PFN-PT) ceramic is well known for ferroelectric

materials that have been developed by researchers for portable microelectronic applications (Kim et al., 2011). Generally, the perovskite structure of ferroelectrics ceramic has been obtained on the basis of complex lead niobate perovskite $\text{Pb}(\text{B}'\text{Nb})\text{O}_3$ (where $\text{B}' = \text{Mg}^{2+}, \text{Zn}^{2+}, \text{Fe}^{3+}$) (Mackeviciute, et al., 2015). Meanwhile, the regarded pyrochlore phase exists in microstructure of PFN–PT ceramics because of Pb^{2+} vacancy. Pyrochlore phase was suppressed by either using columbite method or two-steps sintering into mixed oxide process (Moetakef & Nemati, 2007) which had an effect to produce piezoelectric properties. Singh et al. (2008) found that the structure of $(1-x)$ PFN– x PT was in monoclinic phase for $x < 0.05$ and predominantly tetragonal phase for $x > 0.10$, whereas the two phases coexist in the intermediate composition range. The morphotropic phase boundary (MPB) depends on composition, which is a technique for showing maximum dielectric properties (Park & Shrout, 1997) $(1-x)$ PFN– x PT is a maximum peak of the dielectric constant at $x = 0.08$ (Singh et al., 2008). The $(1-x)$ PFN– x PT system is reported on the dielectric properties of highly leaky samples (Wang et al., 2005), which cannot attain the phase transition at a higher temperature. To improve the physical and electrical properties for applications lead-based relaxor ceramics, which were produced by doping with transition oxides (ZnO, CuO, La_2O_3 , Fe_2O_3 , MnO_2 , etc.), have been intensively investigated. For example, Zhang et al. (2009) found that Ce doped PNW–PMN–PZT ceramics were able to optimize piezoelectric and dielectric properties with stable temperature. In addition, the researchers (Kang et al., 2004; Feng, Rongzi et al., 2009) revealed that PMN–PNN–PZT ceramics modified with Zn^{2+} and Li^+ ions exhibited excellent electrical properties and Curie temperature ($T_c = 251^\circ\text{C}$), at 960°C for low sintering temperature. In addition, Yan et al. (2012) studied the effect of MnO_2 on the phase structure and ferroelectric behaviour of PMN–PZT ceramics found that the Mn^{2+} ion substitute on Zr^{4+} and Ti^{4+} sites induced a hardening effect. In addition, Wattanasarn and Seetawan (Wattanasarn & Seetawan, 2014) reported Zn atom has affected the tetragonal structure and improved the highest temperature that can be achieved due to a single normal vibration (Debye temperature) of PbTiO_3 with first principle calculation.

The synthesised ferroelectric material plays a crucial role for responding to the microstructure and electrical properties. In this work, the $0.9\text{Pb}(\text{Fe}_{1/2}\text{Nb}_{1/2})\text{O}_3-0.1\text{PbTiO}_3$ ($0.9\text{PFN}-0.1\text{PT}$) ceramics occupy into MPB of $(1-x)$ PFN– x PT system which is prepared by columbite mixed oxide. The ZnO was added into $0.9\text{PFN}-0.1\text{PT}$ ceramics with solid state reaction. The samples were investigated in microscale for traces of atomic vibration by Fourier transform infrared spectroscopy (FTIR) method. In the phase transition, the temperature is dependent on the dielectric behaviour that corresponded to ferroelectric properties of $0.9\text{PFN}-0.1\text{PT}$ and $0.9\text{PFN}-0.1\text{PT}: x\text{Zn}$ ceramics. By adding ZnO into the $0.9\text{PFN}-0.1\text{PT}$ it is expected to decrease the pyrochlore phase, as well as affect the dielectric properties of $0.9\text{PFN}-0.1\text{PT}$.

MATERIALS AND METHOD

The $0.9\text{PFN}-0.1\text{PT}: x\text{Zn}$ were obtained by solid state reaction technique, limiting Zn atoms substituted in any vacancies and also protecting the temperature of maximum dielectric constant which would be shifted when ZnO contents were increased, and the researcher chose $x = 0, 1, 2,$ and 3 wt.%. In this study, the $0.9\text{PFN}-0.1\text{PT}: x\text{Zn}$ ceramics were synthesised by columbite method. Firstly, iron niobate ($\text{Fe}_2\text{Nb}_2\text{O}_4$: FN) was prepared from the stoichiometric ratio for

precursor, which composed of Fe₂O₃ (98%), and Nb₂O₅ (99.9%). The materials were mixed with ball milling, and fired at 1000°C for 6 h. Secondly, 0.9PFN-0.1PT: *x*Zn was prepared by mixing oxide materials that were composed of FN, PbO (99%) excess with 4 mol% for compensate PbO which volatilised during the heat treatment, TiO₂ (98%) and ZnO (< 100 nm). These powders were mixed in stoichiometric ratio of the samples, and then the milling process in deionize water took 24 h and was vaporised by hotplate. The mixed powder was pressed through the following two-steps sintering process (Moetakef & Nemati, 2007). In the first step, the sample was pressed to a pellet for calcination, treated in PbO atmosphere by covering the sample with Pb(Zr_{1-x}Ti_x)O₃ (PZT) powder in an alumina crucible, and heated to 850°C with the heating rate of 10°C min⁻¹. Then it was soaked at that temperature for 4 h, and then cooled down at room temperature. After that, the sample was crushed into powder and then mixed with polyvinyl alcohol solution (5 wt.%), and pressed at 320 MPa to form disc shapes with 10 mm in diameter and 1.2 mm in thickness. The second step for sintering at a high temperature was for the disc sample to be maintained under lead atmosphere during the firing process. The sample pellets were embedded in the PZT powder which was derived from the previous step. The sample was then sintered at 1100°C for 2 h with a heating rate of 10°C min⁻¹. After that, the sintered sample was polished to remove the lead-rich layer on the surface of the sample.

The plane patterns of sintered samples were determined with an X-ray diffractometer (Shimadzu XRD-6100). The XRD patterns were recorded at room temperature with Cu K α radiation. Diffraction intensity was measured at 2θ in range of 20° to 60° with a step up of 0.02°. The bulk density of the sample was measured using the Archimedes method in order to compare it with theoretical density. The FTIR absorption in the spectral range 400–4,000 cm⁻¹ was obtained. The IR absorption measurements were done using the KBr pellet technique. The samples were crushed in an agate mortar to obtain particles. This procedure was applied every time to fragments of a sample to avoid structural modifications due to ambient moisture. The samples were coated with silver paint on the sample surfaces as electrode for dielectric measurements. The capacitances were obtained using Chen Hwa 1061 LCR-Meter connected to the sample electrode in a chamber that was heated at a temperature from 30–300°C. In this study, the dielectric constants were measured at a discrete frequency range of 1 kHz to 100 kHz. The capacitances were used to investigate the dielectric constants from $\epsilon_r = Ct / \epsilon_0 A$; where C is the capacitance of the sample, t and A are the thickness and the area of the electrode, respectively, and ϵ_0 is the dielectric permittivity in vacuum.

RESULTS AND DISCUSSION

The XRD technique was run on CuK α (wavelength 1.54056 Å) for investigated phase patterns of 0.9PFN-0.1PT: *x*Zn ($x = 0, 1, 2,$ and 3 wt.%) as shown in Figure 1. The samples were identified to the characteristic patterns of complex perovskite structure (A(B'B'')O₃). Usually, 0.9PFN-0.1PT exhibits morphotropic phase boundary (MPB) and then gradually shifts to tetragonal phase with increasing PT contents (Ciomega, et al., 2013). In Figure 1, the ZnO content doped into 0.9PFN-0.1PT. It observed that pyrochlore phase was suppressed with increasing ZnO contents as presented in Figure 1 (inset). The phase transformation from

the MPB phase eventually shifted to tetragonal phase with increasing ZnO contents. Using 2θ data from 20° to 60° of the XRD results evaluated the percentage of perovskite structure, lattice parameters a , c , tetragonality (c/a), bulk density and relative density are listed in Table 1. This result implied that not only Zn^{2+} ion can be a reducing Pb^{2+} vacancy but it also increases tetragonality of 0.9PFN–0.1PT. Furthermore, Peak of 0.9PFN–0.1PT is shifted due to substitution in B site in perovskite structure and defect structure, which Zn^{2+} (ionic radii 0.8 \AA) substituted Fe^{3+} (0.75 \AA) and Nb (0.86 \AA), lead to compressing lattice strain in microstructure and effect to shifted peak (Zhang, et al., 2016).

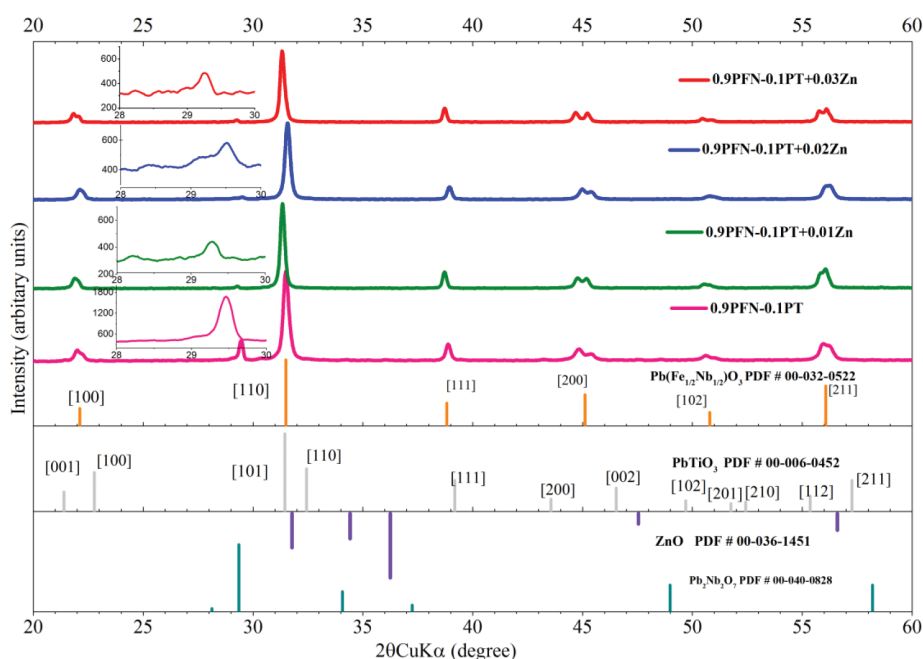


Figure 1. The illustrated of XRD patterns of 0.9PFN–0.1PT: xZn ($x = 0, 1, 2,$ and 3 wt.%) ceramics

Table 1

Lattice parameters a and c , tetragonality (c/a), Rietveld refinement using data from 20° to 60° of the XRD results in Figure 1. Experiment Density ($g\ cm^{-3}$), and relative density (%)

0.9PFN– 0.1PT: xZn	% Perovskite	a (\AA)	c (\AA)	c/a	Experiment Density	Relative density
$x = 0$	80.11	0.4008	0.4063	1.013	7.701	92.57
$x = 1$	91.61	0.3986	0.4023	1.009	7.687	89.55
$x = 2$	92.28	0.3987	0.4021	1.008	7.764	90.59
$x = 3$	92.36	0.3975	0.4019	1.011	7.567	87.87

The (200) plane is a single peak for the cubic, while (200)T and (002)T plane can be found in tetragonal phase (Kang, et al., 2004). It is noted that the broad peak exhibited the tetragonal phase split into two peaks at (200) plane. The major differences of XRD pattern are analysed on the peaks of (200) plane which ranged from 44°–46°. In order to determine the phase transition of various samples accurately, the broad peak of (002) plane lines of three samples were obtained by extracting from XRD pattern range 44°–46°, as presented in Figure 2. The (200) reflections splits to (200)T and (002)T were found by fitting Gaussian peaks (red and green lines). The MPB phase for $x = 0$. The coexistence of equal quantities of the MPB region, in the tetragonal and cubic phases is well known (Feng et al., 2009; Yan et al., 2012). Consequently, X-ray diffractometer is the evidence clearly of splitting on plane (200)T as shown in Figure 2. In addition, lattice parameters and tetragonality (c/a) are the reasons that justify the shift of MPB phase to tetragonal phase, that increased tetragonality (c/a) with increasing ZnO contents as listed in Table 1.

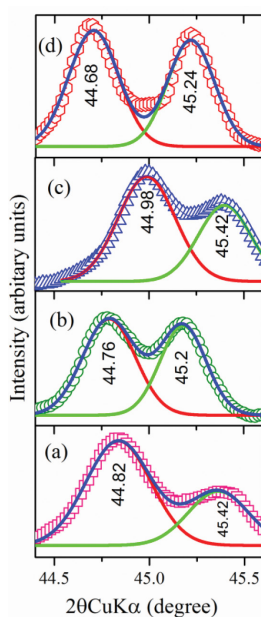


Figure 2. The phase transition of MPB to tetragonal phase on plane (200)T and (002)T ($2\theta = 44^\circ\text{--}46^\circ$) of 0.9PFN–0.1PT: x Zn, (a) $x = 0$ wt.%, (b) $x = 1$ wt.%, (c) $x = 2$ wt.%, and (d) $x = 3$ wt.%

In order to study the vibrational properties of ZnO the sample was doped into 0.9PFN–0.1PT. Fourier transform infrared spectroscopy (FTIR) spectra were in the region of 400–4,000 cm^{-1} at room temperature, as depicted in Figure 3. It observed that all samples exhibited quite similarly to peaks confirming 0.9PFN–0.1PT: x Zn. Absorption bands was around 605 cm^{-1} for all samples which attributed to O–Ti–O and Nb–O bending vibrations of TiO₆ and NbO₆ octahedral groups respectively (Eastel & Udy, 1972; Shimizu et al., 1977). Furthermore, FTIR peak showed intensity of Zn at 3,412 cm^{-1} and 3,792 cm^{-1} for 2 wt.% and 3 wt.%. In contrast, 0.9PFN–0.1PT: 1Zn removed any peak on this range due to the complexity of zinc with the

substitution in complex perovskite structure (Kwon et al., 2002; Silva & Zaniquelli, 2002; Paterson, et. al, 2015). In addition, FTIR peaks shifted due to some structural modifications that occurred during the dioxygen intercalation process (Laureano, et al., 1998). It is noted that Zn atom has affected the ferroelectric materials that is phase transition corresponding with XRD data.

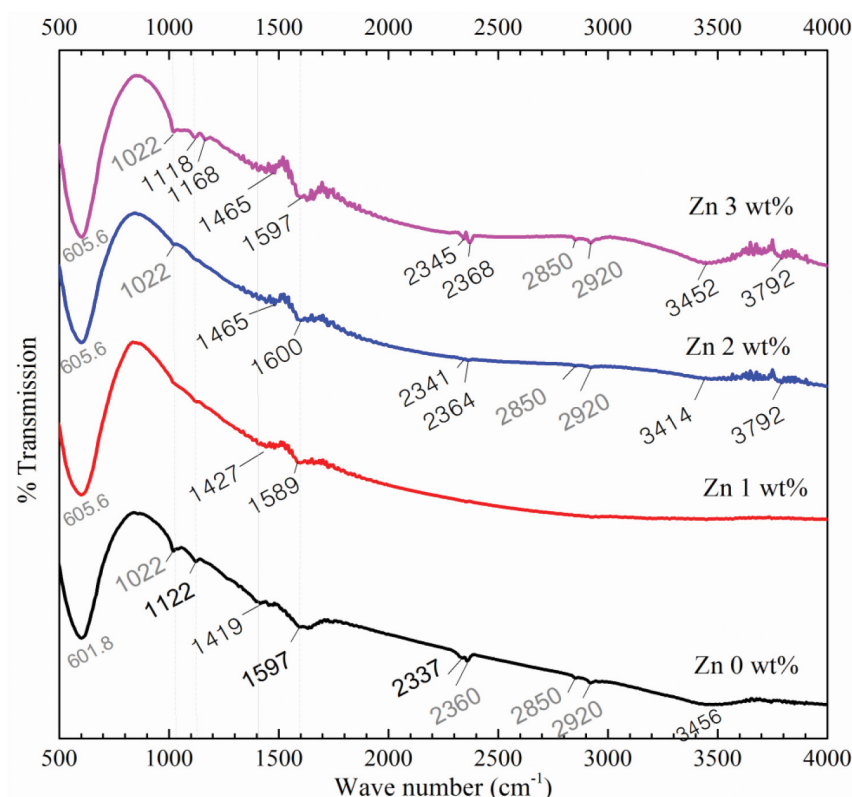


Figure 3. Fourier transform infrared spectra (FTIR) of 0.9PFN-0.1PT: x Zn ($x = 0, 1, 2,$ and 3 wt.%) ceramics

Generally, dielectric behaviour is favoured to explain the existence in microscopic and nanoscopic scales. Dielectric constant (ϵ_r) and dielectric loss ($\tan\delta$) as a function of temperature at various frequencies of 0.9PFN-0.1PT: x Zn was found to have the maximum dielectric constant at temperature 144°C, as shown in Figure 4. The ϵ_r of 0.9PFN-0.1PT was found to be very high at 1 kHz see in Figure 4(a) (inset), and then rapidly decreasing. The broad wide peaks of dielectric dispersion appeared with increasing frequency, as shown in Figure 4(a). Similarly, 0.9PFN-0.1PT: 1Zn exhibited broad wide peaks of dielectric dispersion and increased when the temperature was higher than 150°C is depicted in Figure 4(b). It is evident that 0.9PFN-0.1PT exhibited a diffuse ferroelectric phase transition with a transition temperature that caused B site ions disorder in complex perovskite. Figure 4(c) and Figure 4(d) showed a ferroelectric behaviour of 0.9PFN-0.1PT:2Zn, and 0.9PFN-0.1PT:3Zn respectively,

with a broad wide peak transformed to sharp peak and dielectric constant, which was observed to significantly decrease when ZnO was added. It was assumed that the spontaneous polarization (domain wall) dynamically changed the polar nanoregions' direction at the above average transition temperature (Mackeviciute, et al., 2015). It exhibited that increasing ZnO content into 0.9PFN-0.1PT would be affected to relaxor ferroelectric behaviours that were ordered structure of a short range of heterogeneous in mixed oxide powder.

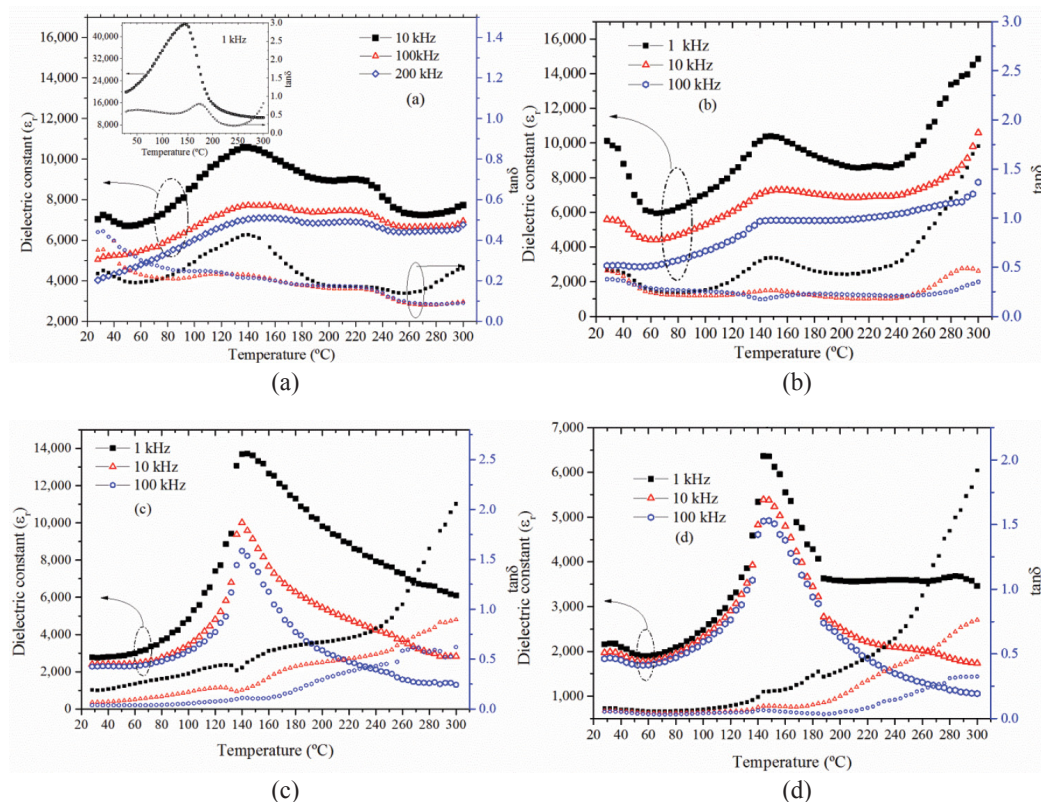


Figure 4. Illustrating the temperature and frequency dependence of dielectric constant and dielectric losses of 0.9PFN-0.1PT: x Zn; (a) $x = 0$ wt.%; (b) $x = 1$ wt.%; (c) $x = 2$ wt.%; and (d) $x = 3$ wt.%

The dielectric loss ($\tan\delta$) of 0.9PFN-0.1PT: x Zn ($x = 0, 1, 2,$ and 3 wt.%) ceramics are illustrated in Figure 4(a)–(e), the $\tan\delta$ shows dependence with frequencies of 0.9PFN-0.1PT: x Zn at different temperatures. It was found that $\tan\delta$ rose with increasing temperatures, as well as the $\tan\delta$ decreased with increasing frequency of the applied alternative current. This may be due to the hopping of electron that cannot follow the frequencies of applied current. Moreover, the dispersion in $\tan\delta$ at higher temperatures was observed for the samples. This is attributed to the conductivity of the ceramics. In addition, the oxygen vacancies and PbO evaporated during higher temperature sintering. Besides, appearing peak of $\tan\delta$ is suggested that long range order in relaxor ferroelectric phase affected by heating (Ni, Luo, Pan, Zhang, & Chen, 2012).

CONCLUSION

Modified ZnO into 0.9PFN–0.1PT: x Zn ($x = 0, 1, 2,$ and 3 wt.%) ceramics were prepared by the two-step sintering process. The phase transition was demonstrated by using X-ray diffraction technique. The structures of the 0.9PFN–0.1PT experienced a gradual transition process from MPB phase to tetragonal phase with ZnO contents added. FTIR peak showed transmission of bonding vibration of Zn at $3,412\text{ cm}^{-1}$ and $3,792\text{ cm}^{-1}$ for 2 wt.% and 3 wt.%, 0.9PFN–0.1PT: 1Zn disappeared at any peak on these ranges due to homogeneous substitution of zinc atom into complex perovskite structure. 0.9PFN–0.1PT: x Zn exhibited the maximum dielectric constant at a temperature of 144°C , the broad wide peak of dielectric constant was transformed to sharp peak and dielectric constant significantly decreased by adding ZnO. The dielectric loss increased at higher temperatures, as well as decreased with increasing frequency of the applied alternative current. These results confirmed the effect of ZnO to micro-scale and nano-scale transition into 0.9PFN–0.1PT. These results can also be attributed to ferroelectric and relaxor ferroelectric behaviour, and will be useful for electronic applications.

ACKNOWLEDGEMENTS

We would like thank the Research and Development Institute of Rajabhat Sakon Nakhon for the financial support given to us to conduct this research. We would also acknowledge the Science and Technology Department of Rajabhat Sakon Nakhon for FTIR. In addition, we are thankful to Sakon Nakhon Rajabhat University International Conference 2015 (SNRU-IC 2015) team for preparing the documents and for commenting and improving the English language presentation of this paper.

REFERENCES

- Ciomega, C. E., Neagu, A. M., Pop, M. V., Airimioaei, M., Tascu, S., Schileo, G., & Mitoseriu, L. (2013). Ferroelectric and dielectric properties of ferrite–ferroelectric ceramic composites. *Journal of Applied Physics*, *113*, 074103_1-074103_4.
- Cross, L. E. (1996). Ferroelectric materials for electromechanical transducer applications. *Materials chemistry and physics*, *43*(2), 108-115.
- Eastel, A. J., & Udy, D. J. (1972). Preparation and properties of potassium titanates. *High Temperature Science*, *4*(6), 487-495.
- Feng, G., Rongzi, H., Jiaji, L., Zhen, L., & Chang, S. T. (2009). Effects of ZnO/ Li_2O co doping on micro structure and piezoelectric properties of low-temperature sintered PMN–PNN–PZT ceramics. *Ceramics International*, *35*(5), 1863-1869.
- Gao, Y., Xu, H. Y., Wu, He, T., Xu, G., & Luo, H. (2001). Growth and Dielectric Properties of 0.48 $\text{Pb}(\text{Fe}_{1/2}\text{Nb}_{1/2})\text{O}_3$ –0.52 PbTiO_3 Single Crystal. *Japanese Journal of Applied Physics*, *40*(8R), 4998-4999.
- Kang, S. H., Lee, D. S., Lee, S. Y., Kim, I. W., Kim, J. S., Park, E. C., & Lee, J. S. (2004). Pyroelectric and piezoelectric properties of yttrium-doped 0.15 $[\text{Pb}(\text{Ni}_{1/3}\text{Nb}_{2/3})\text{O}_3]$ –0.85 $[\text{Pb}(\text{Zr}_{1/2}\text{Ti}_{1/2})\text{O}_3]$ ceramics. *Ceramics International*, *30*(7), 1453-1457.

- Kim, H. S., Kim, J. H., & Kim, J. (2011). A Review of Piezoelectric Energy Harvesting Based on Vibration. *International Journal of Precision Engineering and Manufacturing*, 12(6), 1129-1141.
- Kwon, Y. J., Kim, K. H., Lim, C. S., & Shim, K. B. (2002). Characterization of ZnO nanopowders synthesized by the polymerized complex method via an organo chemical route. *Journal of Ceramic Processing Research*, 3, 146-149.
- Laureano, M. R., Enrique, R. L., Miguel, A. G., Maria, M. L., Bruque, S., & Gabas, M. (1998). A Peroxonioium Phosphate Derived from NbOPO₄ -3H₂O. *Journal of Solid State Chemistry*, 137(2), 289-294.
- Mackeviciute, R., Goian, V., Greicius, S., Grigalaitis, R., Nuzhnyy, D., Hol, J., & Banys, J. (2015). Lattice dynamics and broad-band dielectric properties of multiferroic Pb(Fe_{1/2}Nb_{1/2})O₃ ceramics. *Journal of Applied Physics*, 117(8), 084101.
- Moetakef, P., & Nemati, Z. A. (2007). Synthesis of PMN pyrochlore free ceramics via a modified mixed oxide method. *Journal of Electroceramics*, 19(2-3), 207-213.
- Ni, F., Luo, L. H., Pan, X. Y., Zhang, Y. P., & Chen, H. B. (2012). Piezoelectric and dielectric properties of Bi_{0.5}Na_{0.5}TiO₃-Bi_{0.5}K_{0.5}TiO₃-Ba_{0.77}Ca_{0.23}TiO₃ lead-free piezoelectric ceramics. *Journal of Materials Science*, 47(7), 3354-3360.
- Park, S. E., & Shrout, T. R. (1997). Ultrahigh strain and piezoelectric behavior in relaxor based ferroelectric single crystals. *Applied Physics Letters*, 82(4), 1804-1811.
- Paterson, A., Wong, H. T., Liu, Z., Ren, W., & Ye, Z. G. (2015). Synthesis, structure and electric properties of a new lead-free ferroelectric solid solution of (1-x)BaTiO₃-xBi(Zn_{2/3}Nb_{1/3})O₃. *Ceramics International*, 41, S57-S62.
- Shimizu, T., Morita, T., Yanagida, H., & Hashimoto, K. (1977). Crystallographic Study of Potassium Hexa-titanate. *Journal of the Ceramic Association Japan*, 85(980), 189-193.
- Silva, R. F., & Zaniquelli, M. E. (2002). Morphology of nanometric size particulate aluminium doped zinc oxide films. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 198-200, 551-558.
- Singh, S., Pandey, D., Yoon, S., Baik, S., & Shin, N. (2008). Resolving the characteristics of morphotropic phase boundary in the (1-x)Pb(Fe_{1/2}Nb_{1/2})O₃-xPbTiO₃ system: a combined dielectric and synchrotron x-ray diffraction study. *Applied Physics Letter*, 93(18), 182910_1-182910_3.
- Vladimír, K. A., Carlos, Jaroslav, B., Helena, B., & Karol, S. (2003). Effect of PMN modification on structure and electrical response of xPMN-(1-x)PZT ceramic systems. *Journal of the European Ceramic Society*, 23(7), 1157-1166.
- Wang, J., Tang, X., Chan, H. L., & Choy, C. L. (2005). Dielectric relaxation and electrical properties of 0.94Pb(Fe_{1/2}Nb_{1/2})O₃-0.06PbTiO₃ single crystals. *Applied Physics Letters*, 86(15), 152907_1-152907_3.
- Wattanasarn, H., & Seetawan, T. (2014). Elastic Properties and Debye Temperature of Zn Doped PbTiO₃ from First Principles Calculation. *Integrated Ferroelectrics*, 155(1), 59-65.
- Yan, Y. K., Correa, M., Cho, K. H., Katiyar, R. S., & Priya, S. (2012). Phase transition and temperature stability of piezoelectric properties in Mn modified Pb(Mg_{1/3}Nb_{2/3})O₃-PbZrO₃-PbTiO₃ ceramics. *Applied Physics Letters*, 100(15), 152902_1-152902_3.

- Zhang, C. H., Chi, Q., He, X., Lin, J. Q., Chen, Y., Liu, L. Z., & Lei (2016). Microstructure and electric properties of Nb doping $x(\text{Ba}_{0.7}\text{Ca}_{0.3})\text{TiO}_3-(1-x)\text{Ba}(\text{Zr}_{0.2}\text{Ti}_{0.8})\text{O}_3$ ceramics. *Journal of Alloys and Compounds*, 685, 936-940.
- Zhang, R., Yang, Z., Chao, X., & Kang, C. (2009). Effects of CeO_2 addition on the piezoelectric properties of PNW–PMN–PZT ceramics. *Ceramics International*, 35(1), 199-204.



Some New Bivariate Regression Models

Faroughi, P.^{1*} and Ismail, N.²

¹*Department of Statistics, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran*

²*School of Mathematical Sciences, Faculty of Science and Technology Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*

ABSTRACT

This paper introduces new forms of bivariate generalized Poisson (BGP) and bivariate negative binomial (BNB) regression models which can be fitted to bivariate and correlated count data with covariates. The BGP and BNB regression models can be fitted to bivariate count data with positive, zero or negative correlations. Applications of new BGP and BNB regression models are illustrated on Australian health survey data.

Keywords: Generalized Poisson, negative binomial, bivariate, correlation, overdispersion

INTRODUCTION

Poisson regression is used for modelling count data with covariates. One of the issues with count data is over dispersed. Also, negative binomial (NB) regression can be used to manage over dispersion whilst generalized Poisson (GP) regression may be fitted for under dispersed or over dispersed count data. GP is achieved by the limiting form of a generalized NB distribution (Consul & Jain, 1973). Focusing on previous studies, different forms of GP regressions have been offered for various parameterization of GP regression (Consul & Famoye, 1992; Wang & Famoye, 1997; Famoye, Wulu & Singh, 2004; Zamani & Ismail, 2012; Karimi, Faroughi & Rahim, 2015; Zamani, Faroughi & Ismail, 2015).

When we have bivariate count data, there are some forms of bivariate models which can be fitted to bivariate Poisson (BP) (Campbell, 1934). Kocherlakota and Kocherlakota (1992) using trivariate reduction method. Bivariate generalized Poisson (BGP) distribution which follows the trivariate reduction method was obtained in Famoye and Consul (1995).

BP distribution which applies for different levels of correct as presented by Lakshminarayana Lakshminarayana, Pandit & Rao (1999) where distance is obtained from the product of two Poisson marginals with a multiplicative factor parameter. This study was continued by Famoye (2010a) presented BGP distribution, Famoye (2010b) who introduced BNB regression, and Famoye (2012) who defined BGP regression.

Article history:

Received: 24 February 2016

Accepted: 30 September 2016

E-mail addresses:

pouya.faroughi@gmail.com (Faroughi, P.),

ni@ukm.edu.my (Ismail, N.)

* Corresponding author

MATERIALS AND METHODS

The result of the joint p.m.f. of BP distribution is flexible structure which obtained from the product of two Poisson marginals with a multiplicative factor parameter (Lakshminarayana et al., 1999)

$$P(y_1, y_2) = e^{-\mu_1 - \mu_2} \frac{\mu_1^{y_1} \mu_2^{y_2}}{y_1! y_2!} \{1 + \alpha[(g_1(y_1) - \bar{g}_1)(g_2(y_2) - \bar{g}_2)]\}$$

$$y_1, y_2 = 0, 1, 2, \dots, \quad \mu_1, \mu_2 > 0 \tag{1}$$

where $g_t(y_t)$ and \bar{g}_t are bounded functions in y_t and y_2 . To confide non-negativity in the value of $\{.\}$ in (1),

$$g_t(y_t) = e^{-y_t}, \quad \bar{g}_t = E[g_t(Y_t)] = E(e^{-Y_t}), \quad t = 1, 2. \tag{2}$$

Suppose Y_{i1} and Y_{i2} ($i = 1, 2, \dots, n$) are count response variables. Following (1)-(2), the joint p.m.f. of BP regression model is

$$P(y_{i1}, y_{i2}) = \left[\prod_{t=1}^2 e^{-\mu_{it}} \frac{\mu_{it}^{y_{it}}}{y_{it}!} \right] [1 + \alpha \prod_{t=1}^2 (e^{-y_{it}} - e^{-d\mu_{it}})] \tag{3}$$

where $d = 1 - e^{-1}$ and α is the correlation parameter. The covariates can be included using log links, $\log(\mu_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta}_t$, where \mathbf{x}_{it} are vectors of explanatory variables and $\boldsymbol{\beta}_t$ are vectors of regression parameters. The marginal means and variances are $E(Y_{it}) = Var(Y_{it}) = \mu_{it}$, $t = 1, 2$, and the covariance is $Cov(Y_{i1}, Y_{i2}) = \alpha \mu_{i1} \mu_{i2} d^2 e^{-d(\mu_{i1} + \mu_{i2})}$. When α is zero, it means random variables Y_{i1} and Y_{i2} are not dependent, which they are distributed as a marginal Poisson regression model. If $\alpha > 0$, there are positive and if $\alpha < 0$, there are negative correlations.

The p.m.f. of GP distribution (Consul & Famoye, 1992) is

$$P(y) = \frac{\theta(\theta + \nu y)^{y-1}}{y!} e^{-\theta - \nu y} \tag{4}$$

Where $\theta > 0$, and ν called dispersion parameter with $\max(-1, -\frac{\theta}{4}) < \nu < 1$. The mean is $E(Y) = \mu = \theta(1 - \nu)^{-1}$ and variance is $Var(Y) = \theta(1 - \nu)^{-3}$. The model changes to Poisson when $\nu = 0$, and manages under-dispersion or over-dispersion when $\nu < 0$ or $\nu > 0$. The p.g.f. of GP distribution is $\varphi_Y(u) = E(U^Y) = e^{\theta(t-1)}$ where $t = ue^{\nu(t-1)}$, and the m.g.f. is

$$M_Y(u) = E(e^{uY}) = e^{\theta(e^t - 1)} \tag{5}$$

In (5), $e^t = e^{\nu(e^t - 1) + u}$. If we put $u = -1$ in (5), we gain

$$E(e^{-Y}) = e^{\theta(s-1)} \tag{6}$$

where $\ln s - \nu(s - 1) + 1 = 0$. By differentiating m.g.f. in (5) with respect to u and putting $u = -1$, we obtain

$$\left. \frac{\partial}{\partial u} M_Y(u) \right|_{u=-1} = E(Ye^{-Y}) = \frac{\theta}{1-\nu s} e^{(\theta+\nu)(s-1)-1} \tag{7}$$

where $\ln s - \nu(s-1) + 1 = 0$. GP-1 models gained when $\theta_i = (1-\nu)\mu_i$ in (4) generate

$$P(y_i) = \frac{(1-\nu)\mu_i [(1-\nu)\mu_i + \nu y_i]^{y_i-1}}{y_i!} e^{-(1-\nu)\mu_i - \nu y_i} \tag{8}$$

By using the same method which offered by Lakshminarayana et al. (1999) for obtaining BP distribution, the joint p.m.f. of BGP regression will be equalled

$$\prod_{t=1}^2 \frac{(1-\nu_t)\mu_{it} [(1-\nu_t)\mu_{it} + \nu_t y_{it}]^{y_{it}-1} e^{-(1-\nu_t)\mu_{it} - \nu_t y_{it}}}{y_{it}!} [1 + \alpha \prod_{t=1}^2 (e^{-y_{it}} - c_{it})] \tag{9}$$

From (6), $c_{it} = E(e^{-Y_{it}}) = e^{\mu_{it}(1-\nu_t)(s_t-1)}$, where $\ln s_t - \nu_t(s_t-1) + 1 = 0$, $t=1,2$. The mean, variance and covariance for BGP regression are $E(Y_{it}) = \mu_{it}$, $Var(Y_{it}) = \mu_{it}(1-\nu_t)^2$, $t=1,2$, and $Cov(Y_{i1}, Y_{i2}) = \alpha(c_{i11} - c_{i1}\mu_{i1})(c_{i22} - c_{i2}\mu_{i2})$. From (7), $c_{it} = E(Y_{it}e^{-Y_{it}}) = \frac{(1-\nu_t)\mu_{it}}{1-\nu_t s_t} e^{[(1-\nu_t)\mu_{it} + \nu_t](s_t-1)-1}$

where $\ln s_t - \nu_t(s_t-1) + 1 = 0$, $t=1,2$. When, random variables Y_1 and Y_2 are independent, each is distributed as a marginal GP regression. When $\alpha > 0$ and $\alpha < 0$, we have positive and negative correlations respectively. BGP regression reduces to BP regression when $\nu_1 = \nu_2 = 0$, and handles under- and over dispersion when $\nu_t < 0$ and $\nu_t > 0$, $t=1,2$, respectively.

This article introduces a BGP regression which is based on GP-1 regression exist in (Zamani & Ismail, 2012; Zamani, Faroughi & Ismail, 2016). GP-1 regression is produced by putting ν with $\frac{a}{1+a}$ and θ_i with $\frac{\mu_i}{1+a}$ in (4). The mean is for GP-1 regression is $E(Y_i) = \mu_i$ and variance is $Var(Y_i) = \mu_i(1+a)^2$, where a is the dispersion parameter. The model decline to Poisson regression when $a = 0$, and handles under- and over dispersion when $a < 0$ and $a > 0$ respectively. The p.g.f. of GP-1 distribution is $\varphi_{Y_i}(u) = E(U^{Y_i}) = e^{\frac{\mu_i}{1+a}(t-1)}$ where $t = ue^{\frac{a}{1+a}(t-1)}$, so that the m.g.f. is $M_{Y_i}(u) = E(e^{uY_i}) = e^{\frac{\mu_i}{1+a}(e^t-1)}$ where $e^t = e^{\frac{a}{1+a}(e^t-1)+u}$. Therefore, $E(e^{-Y_i}) = e^{\frac{\mu_i}{1+a}(s-1)}$ and $E(Y_i e^{-Y_i}) = \frac{\mu_i}{1-a(s-1)} e^{\frac{\mu_i+a}{1+a}(s-1)-1}$ where $\ln s - \frac{a}{1+a}(s-1) + 1 = 0$.

The joint p.m.f. of BGP regression is

$$P(y_{i1}, y_{i2}) = \prod_{t=1}^2 \frac{\mu_{it} (\mu_{it} + a_t y_{it})^{y_{it}-1}}{(1+a_t)^{y_{it}} y_{it}!} e^{-\frac{\mu_{it}+a_t y_{it}}{1+a_t}} [1 + \alpha \prod_{t=1}^2 (e^{-y_{it}} - c_{it})] \tag{10}$$

where $c_{it} = e^{\frac{\mu_{it}}{1+a_t}(s_t-1)}$ and $\ln s_t - \frac{a_t}{1+a_t}(s_t-1) + 1 = 0, t=1,2$. The mean, variance and covariance for BGP-1 regression are $E(Y_{it}) = \mu_{it}, Var(Y_{it}) = \mu_{it}(1+a_t)^2, t=1,2$, and $Cov(Y_{i1}, Y_{i2}) = \alpha(c_{i11} - c_{i1}\mu_{i1})(c_{i22} - c_{i2}\mu_{i2})$, where $c_{iit} = \frac{\mu_{it}}{1-a_t(s_t-1)} e^{\frac{\mu_{it}+a_t}{1+a_t}(s_t-1)}$ and $\ln s_t - \frac{a_t}{1+a_t}(s_t-1) + 1 = 0, t=1,2$. When $\alpha = 0$, random variables Y_1 and Y_2 are independent, each is distributed as a marginal GP-1 regression. When $\alpha > 0$ and $\alpha < 0$, we have positive correlation and negative correlation. BGP-1 regression decline to BP regression when $a_1 = a_2 = 0$, and manages under dispersion and over dispersion when $a_t < 0$ and $a_t > 0, t=1,2$, respectively.

The p.m.f. of NB regression is

$$P(y_i) = \frac{\Gamma(y_i + v_i)}{y_i! \Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i} \right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i} \right)^{y_i}, \quad y_i = 0, 1, 2, \dots, n \tag{11}$$

where $v_i^{-1} = a$ is the dispersion parameter. The mean and variance of NB regression are $E(Y_i) = \mu_i$ and $V(Y_i) = \mu_i(1 + v_i^{-1}\mu_i) = \mu_i(1 + a\mu_i)$. NB regression in (11) is also referred as NB-2 regression. NB-2 regression reduces to Poisson regression in the limit as $a \rightarrow 0$, and display over dispersion when $a > 0$.

If we replace $v_i = a^{-1}\mu_i$ in p.m.f. (4), NB-1 regression is obtained. The p.m.f. is (Cameron & Trivedi, 2013; Greene, 2008)

$$P(y_i) = \frac{\Gamma(y_i + a^{-1}\mu_i)}{y_i! \Gamma(a^{-1}\mu_i)} \left(\frac{1}{a^{-1} + 1} \right)^{y_i} \left(\frac{a^{-1}}{a^{-1} + 1} \right)^{a^{-1}\mu_i}, \quad y_i = 0, 1, 2, \dots, n \tag{12}$$

where a is the dispersion parameter. The mean and variance of NB-1 regression are $E(Y_i) = \mu_i$ and $V(Y_i) = \mu_i(1 + a)$.

By use the same method offered by Lakshminarayana et al. (1999), BNB regression model can be derived from NB-1 marginals and a multiplicative variable parameter. The p.m.f. of BNB regression model is

$$P(y_1, y_2) = \prod_{t=1}^2 \frac{\Gamma(y_{it} + a_t^{-1}\mu_{it})}{y_{it}! \Gamma(a_t^{-1}\mu_{it})} \left(\frac{1}{a_t^{-1} + 1} \right)^{y_{it}} \left(\frac{a_t^{-1}}{a_t^{-1} + 1} \right)^{a_t^{-1}\mu_{it}} \left[1 + \alpha(e^{-y_{i1}} - c_{i1})(e^{-y_{i2}} - c_{i2}) \right] \tag{13}$$

where $c_{it} = \left[(1 - \theta_{it}) / (1 - \theta_{it}e^{-1}) \right]^{a_t^{-1}\mu_{it}}, \theta_{it} = 1 / (a_t^{-1} + 1)$, a is the dispersion parameter and α is a multiplicative factor (or correlation) parameter.

From p.m.f. (13), Y_{i1} and Y_{i2} are independent if $\alpha = 0$. When $\alpha < 0$, the correlation between the response variables is negative and when $\alpha > 0$, the correlation between Y_{i1} and

Y_{i2} is positive. If $a \rightarrow 0$, BNB regression reduces to BP regression in the limit and if $a > 0$, the variance exceeds the mean and BNB regression allows over dispersion. The correlation between Y_{i1} and Y_{i2} can be defined, and it is equal to

$$\prod_{i=1}^2 \alpha(1-e^{-1})^2 \sqrt{\mu_{ii}(1+a_i^{-1}\mu_{ii}^2)} \left[1+(1-e^{-1})a_i^{-1}\mu_{ii}^2\right]^{-1-a_i^{-1}\mu_{ii}}$$

A two-sided likelihood ratio test (LRT) can be performed to test the dispersion (over- or under dispersion) in BP against BGP alternatives where the hypothesis is $H_0 : a_1 = a_2 = 0$. The LRT is $T = 2(\ln L_1 - \ln L_0)$, where $\ln L_1$ and $\ln L_0$ are the models' log likelihood under their respective hypothesis. Since BP model is nested within BGP model, the statistic is asymptotically distributed as a chi-square with two degrees of freedom.

Likelihood ratio test can be performed to test over dispersion in BP regression against BNB regression. where L_0 and L_1 are the likelihood functions when H_0 and H_1 are true respectively. Since BNB regressions reduce to BP regression in the limit when $a \rightarrow 0$, the null hypothesis is $H_0 : a_1 = a_2 = 0$. The LRT statistic is approximately distributed as a probability of 0.25 at point zero, a 0.5 of chi-square with one degree of freedom and a 0.25 of chi-square with two degrees of freedom (Famoye, 2010a).

Akaike Information Criteria (AIC) is defined as $AIC = 2 \dim(\theta) - 2 \ln(L)$, where $\dim(\theta)$ is the number of parameters and $\ln(L)$ is the log likelihood of the estimated model. The model with the smallest AIC is the best model.

FINDINGS AND DISCUSSION

The health survey Australian data (Cameron, Trivedi, Milne & Piggott, 1988) is used for fitting different types of distributions such as BP, BGP as well as BNB regression models. Cameron and Johansson (1997) applied These data for fitting some univariate models, bivariate generalized negative binomial (BGNB) regression model was defined by Gurm and Elder (2000). Famoye (2012) defined BGP-2 regression model. The health survey data includes 5190 single-person households for 1977–1978 Australian Health Survey.

In this article, we focus on two possibly dependent and negatively correlated response variables called Y_1 , the total number of prescribed medications consumed in two days ago (PRESCRIBED), and Y_2 is the number of non-prescribed medications used in the same period (NON-PRESCRIBED). The mean for prescribed medications is 0.863 and standard deviation for prescribed medications is 1.415, the mean for non-prescribed medications is 0.356 and standard deviation for non-prescribed medications is 0.712 and the correlation between response variables is -0.043. The negative correlation illustrates possible negative dependency between the two variables. Cameron & Trivedi (2013) founded more information on the explanatory variables.

Table 1 shows the estimates and standard errors for BP, BGP, BNB regression models which are fitted jointly to both data.

The LRT for testing BP against BGP regressions is 381.46 and the LRT for testing BP against BNB regressions is 379.74, which is show over dispersion in data sets.

The estimates of correlation parameter for all models are negative. The negative estimates of correlation parameter indicating negative dependence between the two response variables.

The t-ratio for correlation parameter under BP model is 6.66, the t-ratio for correlation parameter under BGP model is 6.94 as well the correlation parameter under BNB model is 6.94 indicating that the two response variables are significantly dependent. Hence, the response variables are better to be fitted under BGP and BNB regression. Based on the AIC the best model is BGP, although the difference between AIC for BGP and BNB is small. BGP and BNB are more suitable than BP.

Table 1
BP, BGP and BNB regression models

Parameter	BP		BGP		BNB	
	est.	s.e.	est.	s.e.	est.	s.e.
PRESCRIBED						
Intercept	-2.70	0.13	-2.66	0.15	-2.66	0.15
Sex	0.48	0.04	0.55	0.04	0.55	0.04
Age	2.41	0.62	2.33	0.71	2.27	0.71
Agesq	-0.64	0.64	-0.65	0.74	-0.56	0.74
Income	0.00	0.06	-0.01	0.06	0.00	0.06
Levyplus	0.29	0.05	0.27	0.06	0.27	0.06
Freepoor	-0.05	0.12	-0.09	0.14	-0.09	0.14
Freerepa	0.30	0.06	0.28	0.07	0.27	0.07
Illness	0.20	0.01	0.20	0.01	0.20	0.01
Actdays	0.03	0.01	0.03	0.01	0.03	0.00
Hscore	0.02	0.01	0.02	0.01	0.02	0.01
Chcond1	0.77	0.05	0.76	0.05	0.75	0.05
Chcond2	1.01	0.05	1.00	0.06	0.99	0.06
NOPRESCRIBED						
Intercept	-2.03	0.17	-1.95	0.19	-2.02	0.19
Sex	0.27	0.05	0.26	0.06	0.27	0.06
Age	2.86	0.95	2.83	1.05	3.08	1.05
Agesq	-3.90	1.07	-3.89	1.19	-4.19	1.19
Income	0.17	0.08	0.11	0.09	0.13	0.09
Levyplus	-0.03	0.06	-0.05	0.06	-0.03	0.06
Freepoor	0.00	0.12	-0.08	0.14	-0.04	0.14
Freerepa	-0.29	0.09	-0.29	0.10	-0.26	0.10
Illness	0.20	0.02	0.20	0.02	0.20	0.02
Actdays	0.01	0.01	-0.00	0.01	-0.00	0.01
Hscore	0.03	0.01	0.03	0.01	0.03	0.01
Chcond1	0.15	0.06	0.14	0.06	0.13	0.06
Chcond2	0.02	0.08	0.03	0.09	0.03	0.09
a_1 , dispersion	-	-	0.18	0.02	0.39	0.04
a_2 , dispersion	-	-	0.14	0.02	0.29	0.03
α , correlation	-0.89	0.13	-0.91	0.13	-0.91	0.13
Log likelihood	-9522.59		-9331.86		-9332.744	
AIC	19099.18		18721.72		18723.49	

CONCLUSIONS

This article has introduced new types of BGP and BNB regression models. Because new form of BGP and BNB regression models have transformative mean and variance relationship, therefore could be fitted to bivariate count data with different levels of correlations, and admit over dispersion of the response variables.

New forms of BGP and BNB regression models were fitted to the Australian health survey and shown to have a negative correlation. The best model is BGP regression model based on AIC, for BNB and BP regression models. The estimates of correlation parameter for all models are significantly negative.

REFERENCES

- Cameron, A. C., Trivedi, P. K., Milne, F., & Piggott, J. (1988). A microeconomic model of the demand for health care and health insurance in Australia. *The Review of economic studies*, 55(1), 85-106.
- Cameron, A. C., & Johansson, P. (1997). Count data regression using series expansions: with applications. *Journal of Applied Econometrics*, 12(3), 203-223.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). United States of America, USA: Cambridge university press.
- Campbell, J. T. (1934). The Poisson correlation action. *Proceedings of the Edinburg Mathematical Society*, 4, 18-26.
- Consul, P. C., & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15(4), 791-799.
- Consul, P. C., & Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 21(1), 89-109.
- Famoye, F., & Consul, P. C. (1995). Bivariate generalized Poisson distribution with some applications. *Metrika*, 42(1), 127-138.
- Famoye, F., Wulu, J. T., & Singh, K. P. (2004). On the generalized Poisson regression model with an application to accident data. *Journal of Data Science*, 2(2004), 287-295.
- Famoye, F. (2010). A new bivariate generalized Poisson distribution. *Statistica Neerlandica*, 64(1), 112-124.
- Famoye, F. (2010). On the bivariate negative binomial regression model. *Journal of Applied Statistics*, 37(6), 969-981.
- Famoye, F. (2012). Comparisons of some bivariate regression models. *Journal of Statistical Computation and Simulation*, 82(7), 937-949.
- Greene, W. (2008). Functional forms for the negative binomial model for count data. *Economics Letters*, 99(3), 585-590.
- Gurmu, S., & Elder, J. (2000). Generalized bivariate count data regression models. *Economics Letters*, 68(1), 31-36.
- Karimi, A., Faroughi, P., & Rahim, K. A. (2015). Modeling and Forecasting of International Tourism Demand in ASEAN Countries. *American Journal of Applied Sciences*, 12(7), 479-486.

- Kocherlakota, S., & Kocherlakota, K. (1992). Regression in the bivariate Poisson distribution. *Communication Statistics-Theory and Methods*, 30(5), 815-825.
- Lakshminarayana, J., Pandit, S. N. N., & Srinivasa Rao, K. (1999). On a bivariate Poisson distribution. *Communications in Statistics-Theory and Methods*, 28(2), 267-276.
- Wang, W., & Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics*, 10(3), 273-283.
- Zamani, H., & Ismail, N. (2012). Functional form for the generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 41(20), 3666-3675.
- Zamani, H., Faroughi, P., & Ismail, N. (2015). Bivariate Poisson-Lindley distribution with application. *Journal of Mathematics and Statistics*, 11(1), 1-6.
- Zamani, H., Faroughi, P., & Ismail, N. (2016). Bivariate generalized Poisson regression model: applications on health care data. *Empirical Economics*, 51(4), 1607-1621.



Identifying the Uncertain Model Parameter of a Steam Turbine System

Wan Munirah, W. M.¹, Tahir, A.^{1,2*} and Azmirul, A.¹

¹Department of Mathematical Science, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Johor, Malaysia

²Centre for Sustainable Nanomaterials, Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, 81310 UTM, Skudai, Johor, Malaysia

ABSTRACT

The transformation method (TM) of fuzzy arithmetic is aimed at simulation and analysis of a system. The aim of this paper is to use fuzzy arithmetic based on the TM on a state space of a steam turbine system. The model is then used to identify the degree of influence of each parameter on the system. Simulation and analysis of the system are presented in this paper.

Keywords: Fuzzy arithmetic, uncertain model parameter, steam turbine system

INTRODUCTION

A power generation plant consists of a of a steam turbine, generator, condenser, boiler and feed water system (Wagner& Priluck, 1982; Mukherjee, 1984; De Mello, 1991a). A combined cycle power plant is a large-scale electric power generation plant in which electricity is obtained from both gas and steam turbines. Energy is transferred in the form of heat or gaseous flow in each section of the turbine (Ordys et al, 2012). A combination of gas and a steam turbine gives the best output for a high efficiency thermal process (Kehlhofer et al., 2009).

There are several ways to increase output of a steam turbine, namely by increasing the amount of energy that goes into the steam turbine or improving its effectiveness. The

former method is easier using a large turbine with huge wheels and vanes. However, the drawback is the efficiency of the steam turbine itself. Therefore, identifying inputs which have the highest degree of influences is paramount in order to maximise energy production.

The general transformation method (TM) of fuzzy arithmetic has proven to be a realistic approach in determining these parameters. It

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail addresses:

wanmunirah@ymail.com (Wan Munirah, W. M.),

tahir@ibnusina.utm.my (Tahir, A.),

mierul2000@gmail.com (Rattikorn Yimnirun)

*Corresponding Author

is a selection method to reduce the number of parameters of a given system. In this way, the influence of the uncertain parameters can be determined, thus simplifying and enhancing the system simulations and analysis.

The model inputs are represented by fuzzy numbers \tilde{p} , σ_1 and σ_r , which denote the worst deviations from the mean value in quasi-Gaussian membership function distribution form.

$$\tilde{p} = gfn^*(\bar{x}, \sigma_1, \sigma_r) \tag{1}$$

The quasi-Gaussian fuzzy number $\tilde{p} \in \tilde{P}'(R)$ with the membership function is expressed (Hanss, 2005) by:

$$\mu_{\tilde{p}}(x) = \begin{cases} 0 & \text{for } x \leq \bar{x} - 3\sigma_1 \\ \exp\left[-\frac{(x-\bar{x})^2}{2\sigma_1^2}\right] & \text{for } \bar{x} - 3\sigma_1 < x < \bar{x} \\ \exp\left[-\frac{(x-\bar{x})^2}{2\sigma_r^2}\right] & \text{for } \bar{x} \leq x < \bar{x} + 3\sigma_r \\ 0 & \text{for } x \geq \bar{x} + 3\sigma_r \end{cases} \quad \forall x \in R \tag{2}$$

The quasi-Gaussian fuzzy number consists of Gaussian fuzzy number that is truncated as and $x < \bar{x} - 3\sigma_1$ and $x > \bar{x} + 3\sigma_r$ respectively.

The state space modelling of steam turbine is presented in Section 2, followed by the transformation method of fuzzy arithmetic in Section 3, and its implementation in Section 4.

STATE SPACE MODELLING OF A STEAM TURBINE

The mathematical modelling of a steam turbine was first presented by Ray (1980). The model considered impulses and reaction stages of the steam turbine. De mello(1991b) and Report (1973) meanwhile presented simple linear models of the steam turbine and the importance of non-linearity. Below are the assumptions of these models:

- Superheated steam is considered as an ideal gas.
- The steam turbine is divided into three sections: high pressure (HP), intermediate pressure (IP), and low pressure (LP).
- Kinetic energy at inlets of each stage is negligible.
- The energy stored at each stage is lumped.

The complete turbine stage (HP, IP or LP) comprises the number of impulse and reaction stages in series. The impulse or reactions at each stage in this connection are expressed in the following equation (Khan et al., 2012).

$$\frac{d}{dt}(X_o) = \frac{1}{v}(w_{in}h_{in} - w_{ou}h_o) \tag{3}$$

$$\frac{d}{dt}(w_{ou}) = \frac{1}{\tau_s}(w_{in} - w_{ou}) \tag{4}$$

$$\frac{d}{dt}(\rho_{ou}) = \frac{1}{v}(w_{in} - w_{ou}) \quad [5]$$

The ideal gas and nozzle are presented by Equations [6-9]:

$$T_o = \frac{h_o - h_{in}}{c_p} + T_{in} \text{ and } p_o = R\rho_o T_o \quad [6]$$

$$\left(r^{\left(\frac{2}{m}\right)} - r^{\left(\frac{m+1}{m}\right)}\right) = \frac{w_{ou}^2}{A^2 \rho_o p_o} \left(\frac{m-1}{2\eta_{\infty m}}\right) \quad [7]$$

$$\text{where } m = \frac{\gamma}{\gamma - \eta_{\infty}(1 - \gamma)}$$

$$\frac{T_{ou}}{T_o} = \left[\frac{P_{ou}}{P_o}\right]^{\eta_{\infty} \left[\frac{\gamma-1}{\gamma}\right]} \text{ where } \gamma = \frac{c_p}{c_p - R} = \frac{c_p}{c_v} \quad [8]$$

$$\Delta h_I = c_p T_o \left(r^{\frac{R}{c_p}} - 1\right) \quad [9]$$

The state space model is formulated as below:

$$\frac{d}{dt}(X_o) = \frac{w_{in} c_p}{v} T_{in} + \frac{w_{in} h_o}{v} - \frac{w_{in}}{v} \cdot \frac{c_p}{R\rho_o} \cdot \frac{P_{in} T_{in}}{T_{ou}} - \frac{w_{ou} h_o}{v} \quad [10]$$

$$\frac{d}{dt}(w_{ou}) = \frac{1}{\tau_s} \left[w_{in} - \frac{AX_o \sqrt{\frac{2\eta_{\infty m} \rho_o \left(r^{\left(\frac{2}{m}\right)} - r^{\left(\frac{m+1}{m}\right)}\right)}{(m-1)\rho_o}}}{h_o} \right] \quad [11]$$

$$\frac{d}{dt}(\rho_{ou}) = \frac{1}{v} \left[w_{in} - \frac{A\rho_{ou} \sqrt{\frac{2\eta_{\infty m} \rho_o \left(r^{\left(\frac{2}{m}\right)} - r^{\left(\frac{m+1}{m}\right)}\right)}{(m-1)\rho_o}}}{r} \right] \quad [12]$$

$$\rho_{ou} = \frac{rRT_o X_o}{h_o} \quad [13]$$

$$T_{ou} = \frac{(c_p T_{ou} - h_{ou}) \rho_{ou} r^{\eta_{\infty} \left(\frac{\gamma-1}{\gamma}\right)}}{r c_p \rho_o} + \frac{X_o r^{\eta_{\infty} \left(\frac{\gamma-1}{\gamma}\right)}}{c_p \rho_o} \quad [14]$$

$$P = \eta c_p T_o \left(r_{hp}^{\left(\frac{R}{c_p}\right)} - 1 \right) w_{ou} \tag{15}$$

From equation [3-15] the state space equation is expressed as below:

$$\begin{bmatrix} X_o' \\ w_{ou}' \\ \rho_{ou}' \end{bmatrix} = \begin{bmatrix} 0 & -\frac{h_o}{v} & 0 \\ A \sqrt{\frac{2\eta_{\infty} m \rho_o \left(r^{\left(\frac{2}{m}\right)} - r^{\left(\frac{m+1}{m}\right)} \right)}{(m-1)\rho_o}} & 0 & 0 \\ 0 & 0 & -\frac{A \sqrt{\frac{2\eta_{\infty} m \rho_o \left(r^{\left(\frac{2}{m}\right)} - r^{\left(\frac{m+1}{m}\right)} \right)}{(m-1)\rho_o}}}{rV} \end{bmatrix} \begin{bmatrix} X_o \\ w_{ou} \\ \rho_{ou} \end{bmatrix} + \begin{bmatrix} \frac{h_o}{v} & -\frac{w_{in}}{v} \cdot \frac{c_p}{R\rho_o} \cdot \frac{T_{in}}{T_{ou}} & -\frac{w_{in}c_p}{v} \\ \frac{1}{\tau_s} & 0 & 0 \\ \frac{1}{v} & 0 & 0 \end{bmatrix} \begin{bmatrix} w_{in} \\ P_{in} \\ T_{in} \end{bmatrix} \tag{16}$$

$$\begin{bmatrix} \rho_{ou} \\ T_{ou} \\ P \end{bmatrix} = \begin{bmatrix} \frac{rRT_o}{h_o} & 0 & 0 \\ r^{\eta_{\infty}\left(\frac{\gamma-1}{\gamma}\right)} & 0 & \frac{(c_p T_{ou} - h_{ou}) r^{\eta_{\infty}\left(\frac{\gamma-1}{\gamma}\right)}}{r c_p \rho_o} \\ 0 & \eta c_p T_o \left(r_{hp}^{\left(\frac{R}{c_p}\right)} - 1 \right) & 0 \end{bmatrix} \begin{bmatrix} X_o \\ w_{ou} \\ \rho_{ou} \end{bmatrix} \tag{17}$$

Equations [16 and 17] form the state space model of a section of the steam turbine. The model can be applied to all sections of the steam turbine.

TRANSFORMATION METHOD OF FUZZY ARITHMETIC

The advanced fuzzy arithmetic can be considered as an elimination of restrictions in its area of applications by Hanss (2005). Therefore, it can be used to evaluate fuzzy rational expressions and can also be applied to simulate static or dynamic systems with fuzzy valued parameters.

Transformation method is basically available in two forms: general and in reduced form. Both methods are practical instruments for simulation and analysis of any system with uncertain model parameters (Hanss, 2002).

The problem with uncertainties can be solved with the implementation of transformation method of fuzzy arithmetic, since the fuzzy valued result of the problem only shows the overall influence of all the uncertain parameters. However, the degree of influence of the different uncertain model parameters are certainly not equal in general. In this regard, the percentages to which the n uncertain parameters of the system contribute to the overall uncertainty of the system can be determined. The coefficients for general transformation method are proposed by Hanss and Nehls (2000, 2001), and Hanss (2002, 2005).

Simulation of a system with uncertain parameters: General transformation method

The uncertain parameters can be represented by fuzzy numbers \tilde{p}_i with $i = 1, 2, \dots, n$.

$$P_i = \{X_i^{(0)}, X_i^{(1)}, \dots, X_i^{(m)}\} \tag{18}$$

$$X_i^{(j)} = [a_i^{(j)}, b_i^{(j)}], \quad a_i^{(j)} \leq b_i^{(j)}, \tag{19}$$

$$i = 1, 2, \dots, n, \quad j = 0, 1, \dots, m.$$

A fuzzy parameterised model is expected to show non-monotonic behaviour with respect to n , for fuzzy $n > 1$ value parameter \tilde{p}_i with $i = 1, 2, \dots, n$. The intervals $X_i^{(j)}$, $i = 1, 2, \dots, n, j = 0, 1, \dots, m-2$, are considered for the transformation scheme. The intervals are now transformed into arrays $\hat{X}_i^{(j)}$ as below:

$$X_i^{(j)} = \underbrace{\left((\gamma_{1,i}^{(j)}, \gamma_{2,i}^{(j)}, \dots, \gamma_{(m+1-j),i}^{(j)}), \dots, (\gamma_{1,i}^{(j)}, \gamma_{2,i}^{(j)}, \dots, \gamma_{(m+1-j),i}^{(j)}) \right)}_{(m+1-j)^{i-j}(m+1-j)\text{-tuples}} \tag{20}$$

$$\gamma_{l,i}^{(j)} = \underbrace{\left(c_{l,i}^{(j)}, \dots, c_{l,i}^{(j)} \right)}_{(m+1-j)^{n-i} \text{ elements}} \tag{21}$$

$$c_{l,i}^{(j)} = \begin{cases} a_i^{(j)} \text{ for } l = 1 \text{ and } j = 0, 1, \dots, m, \\ \frac{1}{2} (c_{l-1,i}^{(j+1)} + c_{l,i}^{(j+1)}) \text{ for } l = 2, 3, \dots, m-j \text{ and } j = 0, 1, \dots, m-2, \\ b_i^{(j)} \text{ for } l = m-j+1 \text{ and } j = 0, 1, \dots, m, \end{cases} \tag{22}$$

The arithmetical expression of F is expressed as:

$$\tilde{q} = F(\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_n) \tag{23}$$

The expression is evaluated separately at each 2^n positions of the combination using conventional arithmetic for crisp numbers. The result of the problem can be expressed in its decomposed and transformed form by the combination $\hat{Z}^{(j)}, j = 0, 1, \dots, m$, the k^{th} element ${}^k\hat{Z}^{(j)}$ of the array $\hat{Z}^{(j)}$ as below:

$${}^k\hat{Z}^{(j)} = F({}^k\hat{x}_1^{(j)}, {}^k\hat{x}_2^{(j)}, \dots, {}^k\hat{x}_n^{(j)}) \quad [24]$$

$$k = 1, 2, \dots, 2^n,$$

Finally, the fuzzy-valued result \tilde{q} of the expression can be achieved in its decomposed form

$$Z^{(j)} = [a^{(j)}, b^{(j)}], j = 0, 1, \dots, m \quad [25]$$

By retransforming $\hat{Z}^{(j)}$ with a correction procedure namely by the recursive formulae

$$a^{(j)} = \min_k(a^{(j+1)}, {}^k\hat{Z}^{(j)}), \quad [26]$$

$$b^{(j)} = \max_k(b^{(j+1)}, {}^k\hat{Z}^{(j)}), j=0, 1, \dots, m-1, \quad [27]$$

$$a^{(m)} = \min_k({}^k\hat{Z}^{(m)}) = \max_k({}^k\hat{Z}^{(m)}) = b^{(m)} \quad [28]$$

Analysis of system with uncertain parameters: General transformation method

The coefficients $\eta_i^{(j)}, i = 1, 2, \dots, n, j = 0, 1, \dots, m-1$, are determined using the following equation:

$$\eta_i^{(j)} = \frac{1}{(m-j+1)^{n-1}(b_1^{(j)} - a_1^{(j)})} \sum_{k=1}^{(m-j+1)^{n-i}} \sum_{l=1}^{(m-j+1)^{i-1}} (s_2\hat{Z}^{(j)} - s_1\hat{Z}^{(j)}) \quad [29]$$

$$s_1(k, l) = k + (l - 1)(m - j + 1)^{n-i+1} \quad [30]$$

$$s_2(k, l) = k + [(m - j + 1)l - 1](m - j + 1)^{n-i} \quad [31]$$

The values $a_i^{(j)}$ and $b_i^{(j)}$ denote the lower and upper bound of the interval $X_i^{(j)}$, and ${}^k\hat{Z}^{(j)}$ is the k^{th} elements of the array $\hat{Z}^{(j)}$. The coefficients $\eta_i^{(j)}$ can be interpreted as gain factors which express the effect of the uncertainty of the i^{th} parameter on the uncertainty of the output z of the problem at the membership level μ_j . The mean gain factor $\eta_i^{(j)}$ is computed such that:

$$\bar{\eta}_i = \frac{\sum_{j=1}^{m-1} \mu_j \eta_i^{(j)}}{\sum_{j=1}^{m-1} \mu_j} \quad [32]$$

Finally, the degree of influence, namely the normalised values ρ_i , is determined for $i=1, 2, \dots, n$, using

$$\rho_i = \frac{\sum_{j=1}^{m-1} \mu_j |\eta_i^{(j)} (a_i^{(j)} + b_i^{(j)})|}{\sum_{q=1}^n \sum_{j=1}^{m-1} \mu_j |\eta_q^{(j)} (a_q^{(j)} + b_q^{(j)})|} \quad [33]$$

$$\sum_{i=1}^n \rho_i = 1 \quad [34]$$

Thus, the values ρ_i give the influence of the i^{th} varying parameter $\tilde{\rho}_i$ on the overall variation \tilde{q} of the system output z , assuming that every parameter is to be varied relatively to the same extent.

From the geometrical point of view, it can be described as multiple evaluations of the system for different set of crisp parameter with the coordinates of points on the $(n-1)$ -dimensional hyper surfaces of a number of $m+1$ n -dimensional cuboids. Only 2^n vertex points of the n -dimensional cuboids are considered for the evaluation of the problem and additional point on the hyper surfaces are taken into account for general transformation method (Hanss, 2002).

APPLICATION TO A STEAM TURBINE SYSTEM

Simulation for the parameter of HP, IP and LP of the steam turbine are considered as independent parameters. The uncertain model parameters and the model inputs are represented by quasi-*Gaussian* distribution fuzzy numbers (Table 1).

Table 1
Fuzzified input parameters for HP, IP and LP of the steam turbine

Parameter		0	0.2	0.4	0.6	0.8	1
High Pressure							
$\tilde{\rho}_1 = w_{in}$	a	10	10.9	11.2	11.4	11.6	12
	b	14	13.1	12.8	12.6	12.4	12
$\tilde{\rho}_2 = P_{in}$	a	4515000	4520000	4521000	4522000	4523000	4525100
	b	4535000	4530000	4529000	4528000	4527000	4525100
$\tilde{\rho}_3 = T_{in}$	a	715	716.6	716.9	717.1	717.3	717.72
	b	720	718.8	718.5	718.3	718.1	717.72
$\tilde{\rho}_4 = h_{in}$	a	3301700	3306000	3308000	3309000	3310000	3311700
	b	3321700	3317000	3316000	3315000	3314000	3311700
$\tilde{\rho}_5 = \rho_{in}$	a	11	12.4	12.7	13	13.2	13.662
	b	16	14.9	14.6	14.4	14.1	13.662

Table 1 (continue)

Parameter		0	0.2	0.4	0.6	0.8	1
Intermediate Pressure							
$\tilde{p}_1 = w_{inLP}$	a	9	9.74	9.92	10.05	10.19	10.459
	b	12	11.18	11	10.86	10.73	10.459
$\tilde{p}_2 = P_o$	a	1200000	1255000	1267000	1276000	1285000	1303400
	b	1400000	1352000	1340000	1331000	1321000	1303400
$\tilde{p}_3 = T_o$	a	720	725.8	726.2	726.4	726.7	727.25
	b	730	728.7	728.3	728.1	727.8	727.25
$\tilde{p}_4 = h_o$	a	3320000	3330000	3331000	3331000	3332000	3333400
	b	3340000	3337000	3336000	3335000	3335000	3333400
$\tilde{p}_5 = \rho_o$	a	0	2.096	2.529	2.879	3.223	3.8835
	b	10	5.681	5.237	4.89	4.55	3.8835
Low Pressure							
$\tilde{p}_1 = w_{inLP}$	a	9	9.74	9.916	10.06	10.19	10.459
	b	12	11.018	11	10.86	10.72	10.459
$\tilde{p}_2 = P_{inLP}$	a	400000	441700	450600	457600	464400	477700
	b	550000	513800	504600	498000	491200	477700
$\tilde{p}_3 = T_{inLP}$	a	590	604.1	605.5	606.5	607.5	609.54
	b	620	614.9	613.2	612.6	611.5	609.54
$\tilde{p}_4 = h_{inLP}$	a	3060000	3065000	3067000	3068000	3069000	3070900
	b	3085000	3076000	3075000	3074000	3073000	3070900
$\tilde{p}_5 = \rho_{inLP}$	a	0	0.8797	1.056	1.195	1.333	1.5982
	b	3	2.32	2.143	2.001	1.866	1.5982

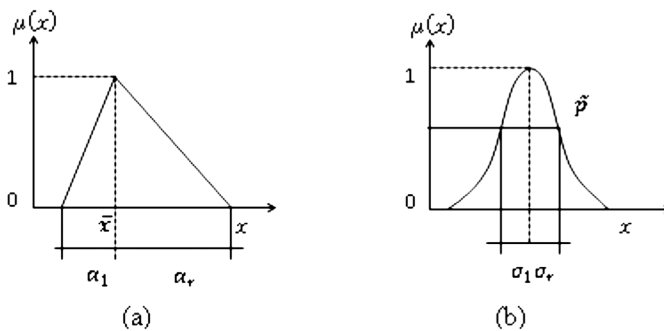


Figure 1. (a) Triangular Fuzzy Number; (b) Gaussian Fuzzy Number

The uncertain steam turbine of HP, IP and LP can be simulated using transformation method in general form and considering ρ_{ou} , T_{ou} as the uncertain output $\tilde{q}(t)$. The results of the simulation of the model with a decomposition number $m = 5$, can be obtained as a fuzzy-valued output of the model. The graph is plotted against time and simulated by MATLAB. The contour plot of the uncertain steam turbine of HP, IP and LP model with the membership μ as a contour parameter in steps of $\Delta\mu = 0.2$ with (a) ρ_{ou} (b) T_{ou} and (c) P is presented by Figures 2-4.

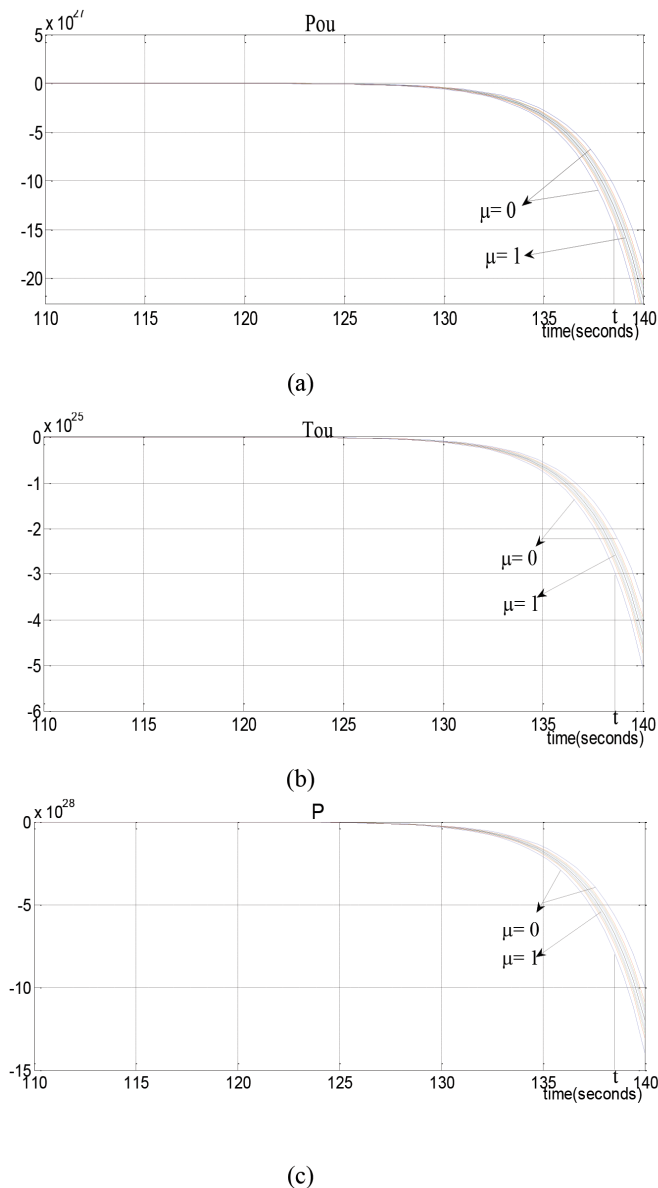


Figure 2. Contour Plot of The Uncertain Steam Turbine of HP Model with μ as a Contour Parameter in Steps of $\Delta\mu = 0.2$ with (a) P_{ou} ; (b) T_{ou} ; and (c) P

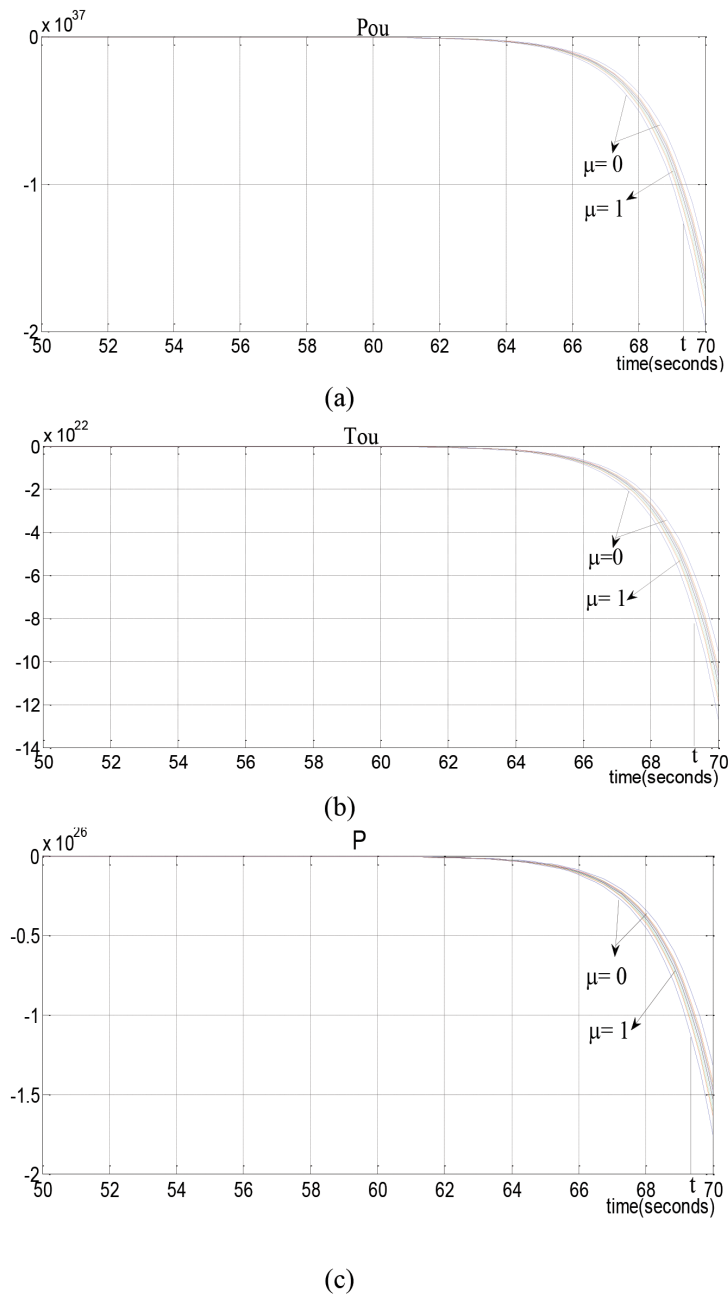
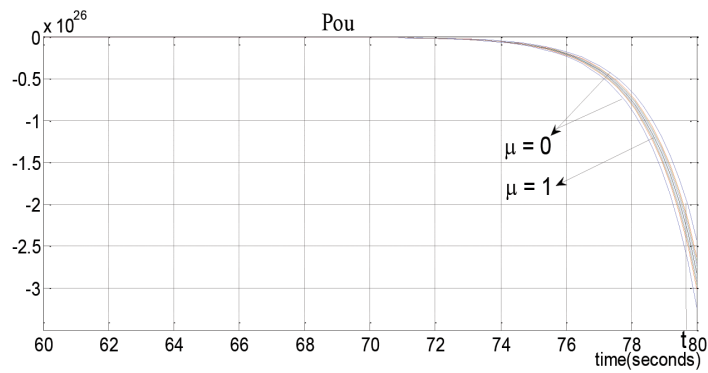
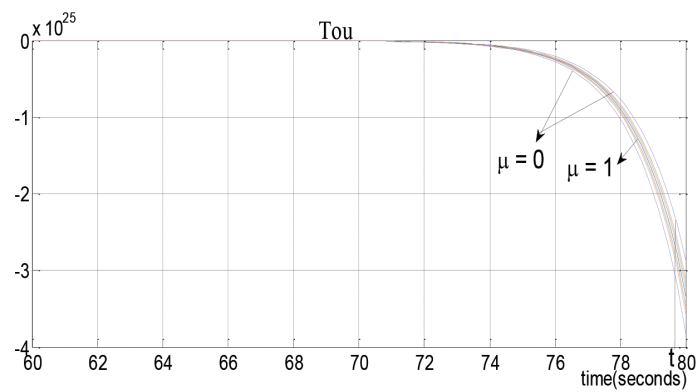


Figure 3. Contour Plot of the Uncertain Steam Turbine of IP Model with μ as a Contour Parameter in Steps of $\Delta\mu = 0.2$ with (a) P_{Ou} ; (b) T_{Ou} ; and (c) P

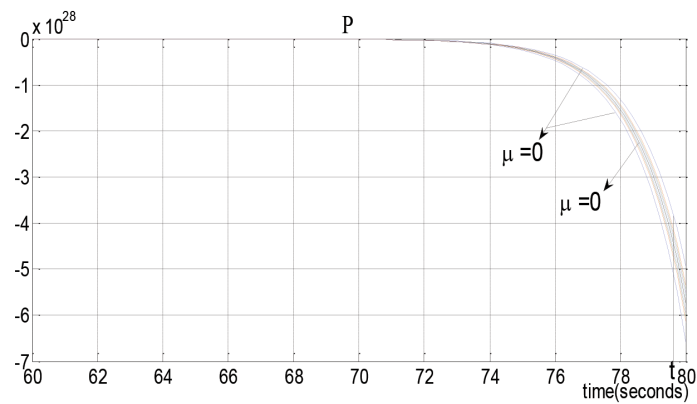
Identification of the Uncertain Model Parameter of a Steam Turbine System



(a)



(b)



(c)

Figure 4. Contour Plot of the Uncertain Steam Turbine of LP Model with μ as a Contour Parameter in steps of $\Delta\mu = 0.2$ with (a) P_{Ou} ; (b) T_{Ou} ; and (c) P

On closer examination (Figure 2), the worst case of interval for HP section of steam turbine at $t = 138.5$ seconds is (a) $\rho_{ou} = 4.15 \times 10^{27} Pa$ (b) $T_{ou} = 8.07 \times 10^{24} \text{ }^\circ K$ and (c) $P = 2.245 \times 10^{28} W$. The initial condition for HP section of steam turbine is $t = 110$ seconds.

Based on Figure 3, the worst case of interval for IP section of steam turbine at $t = 69.35$ seconds is (a) $\rho_{ou} = 3.151 \times 10^{36} Pa$ (b) $T_{ou} = 2.043 \times 10^{22} \text{ }^\circ K$ and (c) $P = 2.822 \times 10^{25} W$. The initial condition for IP section of steam turbine is $t = 50$ seconds.

The worst case of interval for LP section (see Figure 4) of steam turbine at $t = 79.67$ seconds is (a) $\rho_{ou} = 6.53^{25} Pa$ (b) $T_{ou} = 7.81 \times 10^{24} \text{ }^\circ K$ and (c) $P = 1.314 \times 10^{28} W$. The initial condition for LP section of steam turbine is $t = 60$ seconds.

As a result of analysing the overall steam turbine model, the parameters $\rho_1, \rho_2, \dots, \rho_5$ to be identified in the final state space model with $n = 5$ are independent uncertain parameters.

The settings for the mean value (\bar{x}_i) and the standard deviation (σ_i) of the fuzzy parameters are listed in Table 2. The high values for the standard deviations are assumed to be a large uncertainty of the model parameters (Hanss & Nehls, 2001). The following is the settings for the mean values and the standard deviations of the uncertain model parameters.

Table 2
Settings for the mean values (\bar{x}_i) and the standard deviations (σ_i) of the uncertain model parameters

Parameter	\bar{x}	$\sigma 1$	$\sigma 2$	Unit
High Pressure				
\tilde{p}_1	12	0.06	0.06	kg/s
\tilde{p}_2	4525016.67	0.000738	0.00074	Pa
\tilde{p}_3	717.67	0.0013	0.0011	$^\circ K$
\tilde{p}_4	3311816.67	0.0019	0.001	J/kg
\tilde{p}_5	13.64	0.065	0.06	kg/m ³
Intermediate Pressure				
\tilde{p}_1	10.47	0.047	0.049	kg/s
\tilde{p}_2	1302816.67	0.0264	0.0249	Pa
\tilde{p}_3	726.88	0.0032	0.0015	$^\circ K$
\tilde{p}_4	3332816.67	0.0013	0.00072	J/kg
\tilde{p}_5	4.07	0.33	0.49	kg/m ³
Low Pressure				
\tilde{p}_1	10.45	0.047	0.05	kg/s
\tilde{p}_2	477275	0.054	0.051	Pa
\tilde{p}_3	608.74	0.011	0.0062	$^\circ K$
\tilde{p}_4	307115	0.00122	0.00151	J/kg
\tilde{p}_5	1.58	0.33	0.3	kg/m ³

The non-monotonic properties are likely to appear as a consequence of linear elements in the model equations [3-17]. The uncertainty of the steam turbine model is then analysed using an appropriate form of general transformation method. The analysis can be evaluated and the degree of influence can be obtained.

Finally, the result of the analysis model shows the relative influence ρ_i of the uncertain parameters $\rho_i, i= 1, 2, \dots, 5$ on the uncertainty of $\tilde{q}(t) = T_{ou}(t)$ (see Table 3) for every section in the steam turbine model.

Table 3
Percentage the degree of influence of $(t), i= 1, 2, \dots, 5$, of HP, IP and LP of steam turbine

i	High Pressure	Intermediate Pressure	Low Pressure
1	1.16	5.1335	12.4126
2	47.00	3.7212	4.3130
3	36.05	39.3510	0.0250
4	15.58	51.7488	82.8612
5	0.21	0.0455	0.3883

The evaluation of the degree of influence $\rho_i, i= 1, 2, \dots, 5$, for all sections of the steam turbine indicates that each of the parameters are considered to be varied to the same extent relative to its peak value.

For the HP section, the overall uncertainty of the model output is induced by the uncertainty of ρ_2, ρ_3 , and ρ_4 . It shows a moderate impact. The rest of the uncertainties ρ_1 and ρ_5 have a low impact and can be considered negligible.

Next, for the IP section, about half of the overall uncertainty of the model output are induced by the uncertainty of ρ_4 and followed by the uncertainties ρ_3 of which show moderate impact. The rest of the uncertainties of ρ_1, ρ_2 and ρ_5 only show low impact and can be considered negligible.

Moreover, for the LP section the uncertain parameter of ρ_4 has a high degree of influence on an overall uncertainty of the output model. However, the uncertain parameters ρ_1 and ρ_2 only have a low impact and ρ_3 and ρ_5 are negligible.

CONCLUSIONS

The transformation method was applied to the steam turbine system. It showed significant reliability with minimal computing time.

ACKNOWLEDGEMENTS

This research was supported by Ibnu Sina Institute, Universiti Teknologi Malaysia (Q.J130000.3026.00M30).

REFERENCES

- De Mello, F. P. (1991a). Boiler models for system dynamic performance studies. *Power Systems, IEEE Transactions on*, 6(1), 66-74.
- De Mello, F. P. (1991b). Dynamic-Models for Fossil Fueled Steam Units in Power-System Studies. *IEEE Transactions on Power Systems*, 6(2), 753-761.
- Hanss, M., & Nehls, O. (2000). Simulation of the human glucose metabolism using fuzzy arithmetic. In *Fuzzy Information Processing Society, 2000. NAFIPS. 19th International Conference of the North American* (pp. 201-205). IEEE.
- Hanss, M., & Nehls, O. (2001, July). Enhanced parameter identification for complex biomedical models on the basis of fuzzy arithmetic. In *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th* (Vol. 3, pp. 1631-1636). IEEE.
- Hanss, M. (2002). The transformation method for the simulation and analysis of systems with uncertain parameters. *Fuzzy Sets and Systems*, 130(3), 277-289.
- Hanss, M. (2005). *Applied fuzzy arithmetic*. Netherlands: Springer-Verlag Berlin Heidelberg.
- Khan, I., Ahmad, T., & Maan, N. (2012). Feedback Fuzzy State-Space Modeling and Optimal Production Planning for Steam Turbine of a Combined Cycle Power Generation Plant. *Research Journal of Applied Science*, 7(2), 100-107
- Kehlhofer, R., Hannemann, F., Rukes, B., & Stirnimann, F. (2009). *Combined-cycle gas and steam turbine power plants*. United States of America, USA: Pennwell Books.
- Mukherjee, D. (1984). *U.S. Patent No. 4,424,668*. Washington, DC: U.S. Patent and Trademark Office.
- Ordys, A., Pike, A. W., Johnson, M. A., Katebi, R. M., & Grimble, M. J. (2012). *Modelling and simulation of power generation plants*. London: Springer Science & Business Media.
- Ray, A. (1980). Dynamic modelling of power plant turbines for controller design. *Applied Mathematical Modelling*, 4(2), 109-112.
- Report, I. E. E. E. (1973). Dynamic models for steam and hydro turbines in power system studies. *Power Apparatus and Systems, IEEE Transactions on, PAS-92*(6), 1904-1915.
- Wagner, J. B., & Priluck, D. M. (1982). *U.S. Patent No. 4,329,592*. Washington, DC: U.S. Patent and Trademark Office.

APPENDIX

Parameter	Unit	Definition
High Pressure		
W_{in}	kg/s	Inlet steam flow from boiler superheat section
P_{in}	Pa	Inlet steam pressure
T_{in}	$^{\circ}K$	Inlet steam flow temperature
h_{in}	J/kg	Inlet steam flow enthalpy
ρ_{in}	Kg/m^3	Inlet steam flow density
Intermediate Pressure		
W_{inIP}	kg/s	Inlet steam flow from boiler reheater section
P_o	Pa	Inlet steam pressure
T_o	$^{\circ}K$	Inlet steam flow temperature
h_o	J/kg	Inlet steam flow enthalpy
ρ_{o}	Kg/m^3	Inlet steam flow density
Low Pressure		
W_{inLP}	kg/s	Inlet steam flow from IP section
P_{inLP}	Pa	Inlet steam pressure
T_{inLP}	$^{\circ}K$	Inlet steam flow temperature
h_{inLP}	J/kg	Inlet steam flow enthalpy
ρ_{inLP}	Kg/m^3	Inlet steam flow density





Decision Making Procedure Based on Jaccard Similarity Measure with Z-numbers

Mohamad, D.* and Ibrahim, S. Z.

Department of Mathematics, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

ABSTRACT

Fuzzy set with similarity measure approaches are known to be effective in handling imprecise and subjective information to solve decision making problems. Many methods have been introduced based on these two concepts. However, most methods do not take into account the reliability factor of the imprecise information in the evaluation process. In 2010, Zadeh coined the idea of Z-number that has the ability to consider the reliability factor or the level of confidence of human's information expression. Since then, some decision-making methods have included this concept. In this paper, we present a new fuzzy decision making procedure by integrating the Jaccard similarity measure with Z-number to solve a multi criteria decision making problem. The conversion method of the Z-number based linguistic value to trapezoidal fuzzy numbers is used and the Jaccard similarity measure of the expected intervals of trapezoidal fuzzy numbers is applied to obtain the final decision. The feasibility of the methodology is demonstrated by investigating the preference factors that could influence customers to buy their preferred choice of car. The proposed methodology is applicable to solving decision making with a fuzzy environment to achieve a reliable and optimal decision.

Keywords: Decision making, jaccard similarity measure, Z-number, multicriteria group decision making, expected interval of fuzzy numbers

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail addresses:

daud@tmsk.uitm.edu.my (Mohamad, D.),

decu262@yahoo.com (Ibrahim, S. Z.)

*Corresponding Author

INTRODUCTION

Decision making can be defined as a process of problem solving which results in an action (Easton, 1976). The evaluation process becomes difficult due to uncertainty or incomplete information but decision makers are capable of successfully developing and representing the subjective information. Methods are likely used to arrive at their

judgment linguistically rather than provide a finite crisp value. Emergence of fuzzy set theory (Zadeh, 1965) and methods for solving decision making problems based on the fuzzy number have been acknowledged. Fuzzy Analytic Hierarchy Process (FAHP) (Kaboli et al., 2007; Bozbura et al., 2007) and Fuzzy TOPSIS (Chen, 2000; Goli, 2013), known to be effective tools for solving complicated multi-criteria decision making problems, have been applied successfully to numerous problems. Another tool for solving similar problems which is also widely applied in fuzzy multicriteria decision making (FMCDM) problem is the similarity measure approach.

Similarity Measure

Essentially, similarity measure (SM) calculates the degree of similarity between objects or sets through comparisons. In cases of imprecisions, similarity measure of fuzzy numbers is used to measure the degree of similarity between the evaluations. This approach has been successfully applied in solving, for instance, emergency management problem (Xu et al., 2012) and quality of river water (Wang et al., 2013).

Different researchers have developed different similarity measure techniques for diverse purposes such as decision making, clustering, risk analysis, image processing and pattern recognition or information retrieval. Hsieh et al. (1999) proposed similarity measure for generalised fuzzy number based on graded mean integration representation distance. Chen and Chen (2003) presented similarity measure based on centre of gravity points of the generalised fuzzy numbers. Yong et al. (2004) introduced a new similarity measure based on radius of gyration to address the shortcomings of Chen and Chen (2003). Sridevi and Nadarajan (2009) proposed a similarity measure for generalised fuzzy numbers which later became a basis for another similarity measure introduced by Farhadinia and Ban (2013) which is suitable for situations where there are two different generalised fuzzy numbers with the same centre of gravity points.

In recent years, the vector based similarity measure has been adopted by several researchers to solve fuzzy decision making problems. Ye (2012a) applied the Dice similarity measure in the form of a vector which incorporates inner products to find similarity between query and document in order to retrieve the document that represent the query. In the same year, Ye (2012b,c) proposed the vector similarity measure between trapezoidal intuitionistic fuzzy numbers for multi-criteria group decision making, in which the criteria weights and the evaluation values in a decision matrix are expressed by trapezoidal intuitionistic fuzzy numbers, followed by the Dice similarity measure based on the expected interval of trapezoidal fuzzy number with unknown criteria weight. Wu and Mendel (2014) proposed a vector similarity measure for interval type-2 and type-1 fuzzy sets to solve the linguistic approximation problem. Through the weight expected similarity measure between each alternative and the ideal alternative, the ranking order of all the alternatives can be determined and the best one (s) is easily identified.

Z-number

Since uncertainty phenomenon is perceived to exist in human decision making, the reliability of information in decision making can be attributed using the Z-number concept. Fundamentally, a Z-number concept incorporates confidence, certainty, strength of belief or sureness factor in the fuzzy environment evaluation (Zadeh, 2011). By definition, a Z-number, is an ordered pair of fuzzy numbers, (A, R) . A Z-number is associated with a real-valued uncertain variable, X with the first component, A , is a fuzzy restriction on the values of X . A is referred to possibilistic restriction of X . The second component, R , is the certainty that X is A (reliability factor). A and R are perception-based values which are described in a natural language (Zadeh, 2011; Kang et al., 2012b). Zadeh (2011) remarked that the Z-number is more general than the concept of confidence intervals in probability theory. It has been applied in solving some decision-making problems with multiple uncertain environments (Kang et al., 2012a; Zeinalova, 2014; Gardashova, 2014).

In this paper, a new decision making procedure will be presented by using the weighted expected Jaccard similarity measure based on Z-numbers which will be used to solve multi-criteria group decision making (MCGDM) problems. A case study investigates the effectiveness of the proposed decision making procedure.

PRELIMINARIES

This section briefly introduces some definitions and basic concepts related to fuzzy sets, fuzzy numbers, Z-number and expected Jaccard similarity measure of the trapezoidal fuzzy number.

Definition 1

(Bojadziev and Bojadziev, 2007) A fuzzy set, A , is defined by a set or ordered pairs, binary relation on a universe X which may be given as:

$$\{(x, \mu_A(x); x \in X)\}$$

where $\mu_A(x)$ is known as membership function in the interval $[0,1]$ that specifies the grade or degree to which any element x in X belongs to the fuzzy set A .

Definition 2

(Allahviranloo et al., 2012) A generalised trapezoidal fuzzy number GTpFN is defined as $A = (a_1, a_2, a_3, a_4, w_A)$ with membership function $\mu_A(x) : R \rightarrow [0, 1]$ denoted as:

$$\mu_A(x) = \begin{cases} 0, & x < a_1 \\ w_A(x - a_1)/a_2 - a_1, & a_1 \leq x \leq a_2 \\ w_A, & a_2 \leq x \leq a_3 \\ w_A(a_3 - x)/a_4 - a_3, & a_3 \leq x \leq a_4 \\ 0, & x < a_4 \end{cases}$$

where, $a_1 \leq a_2 \leq a_3 \leq a_4$, $a_1, a_2, a_3, a_4 \in R$ and $w_A \in [0, 1]$. When $w_A = 1$, the GTpFN becomes a regular trapezoidal fuzzy number. If $w_A = 1$ and $a_2 = a_3$, then it becomes a triangular fuzzy number.

Definition 3

(Kang et al., 2012). The expectation of a fuzzy number is denoted as:

$$E_A(x) = \int_x x \mu_A(x) dx$$

which is considered as the information strength supporting the fuzzy set A . Note that this value is not the same as the meaning of the expectation of probability space.

Definition 4

(Zadeh, 2011). A Z-number is an ordered pair of two fuzzy numbers denoted as $Z=(A,R)$. The first component of the fuzzy number, A , is a restriction on the values of real-valued uncertain variable, X . Meanwhile, the second component of the fuzzy number, R , is a measure of reliability such as confidence, sureness, strength of belief, probability or possibility for the first component.

Expected Interval of Generalised Trapezoidal Fuzzy Number (GTpFN) (Ye, 2012c)

Let $A = (a_1, a_2, a_3, a_4; w_A)$ be a GTpFN with strictly monotonic left sides and right sides given as $f_A(x) = w_A(x - a_1)(a_2 - a_1)$, $g_A(x) = w_A(x - a_4)(a_3 - a_4)$ respectively. Their inverse functions are defined as $f_A^{-1}(x) = a_1 + (a_2 - a_1)(y/w_A)$, $g_A^{-1}(x) = a_4 + (a_3 - a_4)(y/w_A)$ respectively where $y \in [0, w_A]$. The expected interval of the GTpFN A is a crisp interval $E(A)$ given by

$$EI(A) = [E_L(A), E_U(A)] = \left[\int_0^{w_A} f_A^{-1}(x) dy, \int_0^{w_A} g_A^{-1}(x) dy \right] .$$

It can be shown that

$$E_L(A) = \int_0^{w_A} f_A^{-1}(x) dy = \int_0^{w_A} \left[a_1 + (a_2 - a_1) \frac{y}{w_A} \right] dy = a_1 w_A + \frac{a_2 - a_1}{2} w_A = \frac{w_A(a_1 + a_2)}{2} .$$

Similarly, we have $E_U(A) = \frac{w_A(a_3 + a_4)}{2}$. Hence, the expected interval of GTpFN $A = (a_1, a_2, a_3, a_4; w_A)$ is $E(A) = [w_A(a_1 + a_2)/2, w_A(a_3 + a_4)/2]$ and the expected value is defined as the centre of the expected interval of A given by

$$E(A) = (E_L(A) + E_U(A)) / 2 = w_A(a_1 + a_2 + a_3 + a_4) / 4 .$$

Jaccard Similarity Measure between Vectors

Let $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$ be the two vectors of length n , where all the coordinates are positives. The Jaccard index (Wu and Mendel, 2014) that measures the similarity of these vectors is defined as:

$$J = \frac{X \cdot Y}{\|X\|_2^2 + \|Y\|_2^2} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i y_i}$$

where $X \cdot Y = \sum_{i=1}^n x_i y_i$ is the inner product of the two vectors X and Y , and $\|X\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$, $\|Y\|_2 = \sqrt{\sum_{i=1}^n y_i^2}$ are the Euclidean norms (L_2) of X and Y . Let $A_1 = (a_{11}, a_{12}, a_{13}, a_{14}; w_1)$ and $A_2 = (a_{21}, a_{22}, a_{23}, a_{24}; w_2)$ be two GTFNs. The expected intervals of two GTPFNs A_1 and A_2 are $EI(A_1) = [E_L(A_1), E_U(A_1)]$, $EI(A_2) = [E_L(A_2), E_U(A_2)]$ respectively. Hence, we define the Jaccard similarity in vector space as:

$$J(A_1, A_2) = \frac{[E_L(A_1)E_L(A_2) + E_U(A_1)E_U(A_2)]}{\left\{ \begin{array}{l} (E_L(A_1))^2 + (E_U(A_1))^2 + (E_L(A_2))^2 + (E_U(A_2))^2 \\ - [E_L(A_1)E_L(A_2) + E_U(A_1)E_U(A_2)] \end{array} \right\}}$$

$$= \frac{w_1 w_2 [(a_{11} + a_{12})(a_{21} + a_{22}) + (a_{13} + a_{14})(a_{23} + a_{24})]}{\left\{ \begin{array}{l} w_1^2 [(a_{11} + a_{12})^2 + (a_{13} + a_{14})^2] + w_2^2 [(a_{21} + a_{22})^2 \\ + (a_{23} + a_{24})^2] - w_1 w_2 [(a_{11} + a_{12})(a_{21} + a_{22}) \\ + (a_{13} + a_{14})(a_{23} + a_{24})] \end{array} \right\}} \quad (1)$$

DECISION MAKING PROCEDURE BASED ON WEIGHTED EXPECTED JACCARD SIMILARITY MEASURE USING Z-NUMBER

Let $O = \{O_1, O_2, \dots, O_m\}$ be a set of alternatives and let $C = \{C_1, C_2, \dots, C_n\}$ be a set of criteria. Decision makers (DMs) express their rating evaluation on each alternative O_i ($i=1, 2, \dots, m$) based on the criteria set C_j ($j=1, 2, \dots, n$) using the Z-number evaluation by employing the linguistic variables in the linguistic terms set (Ye, 2012c) as in Table 1 for A (evaluation restriction value) and Table 2 is the linguistic terms set (Kang et al., 2012) for R (reliability measure).

The linguistic value for linguistic terms of restriction value is represented by a trapezoidal fuzzy number (TpFN) and linguistic value for linguistic terms of reliability is represented by a triangular fuzzy number (TFN).

Table 1
Linguistic terms for evaluation (restriction values)

Linguistic terms	Linguistic values
Absolutely Low (AL)	(0.00, 0.00, 0.00, 0.00)
Very Low (VL)	(0.00, 0.00, 0.02, 0.07)
Low (L)	(0.04, 0.10, 0.18, 0.23)
Medium Low (ML)	(0.17, 0.22, 0.36, 0.42)
Medium (M)	(0.32, 0.41, 0.58, 0.65)
Medium High (MH)	(0.58, 0.63, 0.80, 0.86)
High (H)	(0.72, 0.78, 0.92, 0.97)
Very High (VH)	(0.93, 0.98, 1.00, 1.00)
Absolutely High (AH)	(1.00, 1.00, 1.00, 1.00)

Table 2
Linguistic terms for reliability measure

Linguistic terms	Linguistic values
Very Low (VL)	(0.00, 0.00, 0.25)
Low (L)	(0.00, 0.25, 0.50)
Medium (M)	(0.25, 0.50, 0.75)
High (H)	(0.50, 0.75, 1.00)
Very High (VH)	(0.75, 1.00, 1.00)

The procedure for decision making is as follows:

Step 1. Evaluation Process

Fuzzy information (fuzzy rating) as expressed by a DM which denoted by the Z-number (Z_{ij}) can be shown as in the decision matrix D below:

$$D = \begin{matrix} & \begin{matrix} C_1 & & C_n \end{matrix} \\ \begin{matrix} O_1 \\ \vdots \\ O_m \end{matrix} & \begin{bmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & \ddots & \vdots \\ Z_{m1} & \cdots & Z_{mn} \end{bmatrix} \end{matrix}$$

where $Z_{ij} = (A_{ij}, R_{ij})$ involves two parts of the evaluation: the restriction value and the reliability of the evaluation for each alternative $O_i (i = 1, 2, \dots, m)$ with respect to each criterion $C_j (j = 1, 2, \dots, n)$.

Step 2. Z-number conversion to fuzzy number (Kang et al., 2012)

Using the fuzzy expectation, all Z-numbers obtained in Step 1 are converted to fuzzy numbers as in the following procedure:

- a) Convert the reliability of a Z-number, R_{ij} , to a crisp value using the following equation:

$$\alpha = \frac{\int x\mu_R(x)dx}{\int \mu_R(x)dx}$$

- b) Embed the converted value of R_{ij} (crisp value, α) to the first part (restriction). The evaluation is now in the form of weighted Z-number and denoted as:

$$Z^{\alpha} = \left\{ (x, \mu_{A^{\alpha}}(x) \mid \mu_{A^{\alpha}}(x) = \alpha\mu_A(x), x \in [0,1]) \right\}$$

where

$$Z_{ij}^{\alpha} = (a_{ij}^1, a_{ij}^2, a_{ij}^3, a_{ij}^4; \alpha) .$$

- c) Convert the weighted Z-number to a regular fuzzy number using Theorem 3 in Kang et al. (2012) as:

$$\begin{aligned} Z'_{ij} &= (\sqrt{\alpha} \times a_{ij}^1, \sqrt{\alpha} \times a_{ij}^2, \sqrt{\alpha} \times a_{ij}^3, \sqrt{\alpha} \times a_{ij}^4) \\ &= (a'^1_{ij}, a'^2_{ij}, a'^3_{ij}, a'^4_{ij}) \end{aligned}$$

Step 3. Aggregate individual evaluation to group decision matrix (Ye, 2012c)

The individual preference rating of the weighted Z-number is aggregated to the group consensus decision matrix by employing the following aggregation operator:

$$a'_{ij} = \left(\sum_{q=1}^d \lambda_q a'_{ijq}{}^1, \sum_{q=1}^d \lambda_q a'_{ijq}{}^2, \sum_{q=1}^d \lambda_q a'_{ijq}{}^3, \sum_{q=1}^d \lambda_q a'_{ijq}{}^4 \right)$$

where q is a DM, d is the total number of DMs $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_d)$ is DMs' weight vector, $\lambda_d \in [0,1]$ and $\sum_{q=1}^d \lambda_q = 1$.

Step 4. Obtain the normalised decision matrix $B = (b_{ij})_{m \times n}$

Two types of criteria will be considered: benefit and cost. In order to transform the various criteria dimension into non-dimensional criteria, each criterion value a'_{ij} ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$) needs to be normalised into the corresponding comparable element in the normalised decision matrix $B = (b_{ij})_{m \times n}$.

For benefit criterion:
$$b_{ij} = \frac{p_j - a'_{ij}}{\sqrt{\sum_{i=1}^m (E(p_j - a'_{ij}))^2}}$$

And for cost criterion:
$$b_{ij} = \frac{a'_{ij}}{\sqrt{\sum_{i=1}^m (E(a'_{ij}))^2}}$$

where $p_j = \max_{1 \leq i \leq m} \sup \{x_{ij} \mid \mu_{ij}(x_{ij}) > 0\}$. Thus, $B = (b_{ij})_{m \times n}$.

Step 5. Determine the criteria weight, w_j (Ye, 2012b)

Since the criteria weight in the MCDM problem is completely unknown, then a weight model is established based on the standard deviation of the expected values of trapezoidal fuzzy numbers in order to determine a criteria weight vector from the normalised decision matrix. The procedure is as follows:

- a) Obtain the matrix of expected values from the normalised matrix B as $E = (E_{ij})_{m \times n}$ where:

$$E_{ij} = \frac{b_{ij1} + b_{ij2} + b_{ij3} + b_{ij4}}{4}$$

- b) Calculate the standard deviation of expected values for different alternatives with respect to a criterion as follows:

$$f_j(C_j) = \sqrt{\frac{1}{m} \sum_{i=1}^m \left(E_{ij} - \frac{1}{m} \sum_{i=1}^m E_{ij} \right)^2}$$

c) Calculate the criteria weight, w_j using weight model:

$$w_j = \frac{f_j(C_j)}{\sum_{j=1}^n f_j(C_j)}$$

where $w_j \geq 0$ and $\sum_{j=1}^n w_j = 1$. The weight vector is in the form $w = (w_1, w_2, \dots, w_n)$.

Step 6. Calculate the weighted expected Jaccard similarity measure for an alternative O_i .

A weighted expected Jaccard similarity measure between an alternative O_i and the ideal alternative $O^* = (1, 1, 1, 1)$, is defined as absolutely high and is calculated as:

$$W(O^*, O_i) = \sum_{j=1}^n w_j \frac{1.6(b_{ij}^1 + b_{ij}^2 + b_{ij}^3 + b_{ij}^4)}{5.12 + (b_{ij}^1 + b_{ij}^2)^2 + (b_{ij}^3 + b_{ij}^4)^2 - 1.6(b_{ij}^1 + b_{ij}^2 + b_{ij}^3 + b_{ij}^4)}$$

which is obtained using equation (1).

Step 7. Ranking order preferences

Rank the alternatives and select the best one(s) in accordance to the weighted expected similarity measure value. The larger the value of a weighted expected Dice similarity measure $W(O^*, O_i)$, the higher the ordering of the alternative $O_i (i=1, 2, \dots, m)$ in which the rank is based on the descending order.

ILLUSTRATIVE EXAMPLE

The proposed decision making procedure has been employed in investigating factors that influence customers' purchasing decision of four types of cars O_1, O_2, O_3 and O_4 which are from the same manufacturer of the family car categories with the engine cubic capacity range of 1600-2000 cc and prices in the range of RM75,000 to RM100,000. Seven decision makers (DMs) with different backgrounds in automotive industries evaluated each alternative with respect to 15 criteria which are value for money (C_1), fuel consumption (C_2), power (C_3), riding comfort (C_4), performance (C_5), safety (C_6), equipment and interior (C_7), exterior design/size (C_8), after sales maintenance (C_9), environmental friendly (C_{10}), brand image (C_{11}), resale value (C_{12}), advancing technology (C_{13}), delivery time (C_{14}) and promotion (C_{15}).

Evaluations are given by the seven DMs using linguistic terms (for restriction value) and confidence level (reliability) as shown in Table 1 and Table 2 respectively. Table 3, 4, 5 and 6 show the full evaluations given by the decision makers in the form of Z-numbers.

Table 3
Evaluation by DMs on O_1

	DM1	DM2	DM3	DM4	DM5	DM6	DM7
C1	(MH,M)	(VH,VH)	(M,M)	(MH,H)	(MH,H)	(L,H)	(MH,H)
C2	(MH,H)	(L,VH)	(H,H)	(MH,H)	(MH,VH)	(L,H)	(H,H)
C3	(MH,H)	(L,VH)	(ML,M)	(H,H)	(ML,VH)	(ML,H)	(VH,H)
C4	(H,VH)	(ML,VH)	(H,H)	(M,H)	(ML,VH)	(M,L)	(VH,H)
C5	(M,VH)	(ML,VH)	(M,H)	(MH,H)	(M,VH)	(ML,L)	(VH,H)
C6	(H,H)	(H,VH)	(MH,H)	(MH,H)	(M,VH)	(M,L)	(AH,H)
C7	(MH,VH)	(L,VH)	(M,H)	(M,H)	(M,H)	(ML,H)	(VH,H)
C8	(AH,VH)	(H,VH)	(M,H)	(MH,H)	(M,VH)	(ML,L)	(VH,H)
C9	(VH,VH)	(M,VH)	(H,H)	(MH,H)	(H,H)	(M,L)	(VH,H)
C10	(VH,H)	(M,VH)	(ML,H)	(H,H)	(ML,VH)	(M,L)	(AH,H)
C11	(MH,VH)	(H,VH)	(H,H)	(M,H)	(H,VH)	(ML,M)	(AH,H)
C12	(HH,M)	(VH,VH)	(H,H)	(MH,H)	(H,VH)	(ML,L)	(VH,H)
C13	(MH,VH)	(M,VH)	(M,H)	(M,H)	(MH,VH)	(L,M)	(H,H)
C14	(M,H)	(M,VH)	(H,H)	(M,H)	(MH,VH)	(H,H)	(AH,H)
C15	(MH,M)	(M,VH)	(H,H)	(MH,H)	(H,VH)	(L,H)	(H,H)

Table 4
Evaluation by DMs on O_2

	DM1	DM2	DM3	DM4	DM5	DM6	DM7
C1	(H,H)	(AH,VH)	(M,M)	(VH,H)	(H,VH)	(H,H)	(H,H)
C2	(H,H)	(L,VH)	(H,H)	(H,H)	(H,VH)	(ML,H)	(VH,H)
C3	(VH,H)	(MH,VH)	(H,H)	(H,H)	(H,VH)	(MH,H)	(VH,H)
C4	(H,H)	(VH,VH)	(M,H)	(H,H)	(VH,VH)	(H,M)	(AH,H)
C5	(H,VH)	(M,VH)	(H,H)	(H,H)	(H,VH)	(MH,H)	(AH,H)
C6	(H,H)	(H,VH)	(MH,H)	(H,H)	(VH,VH)	(VH,H)	(AH,H)
C7	(H,H)	(MH,VH)	(M,H)	(H,H)	(AH,VH)	(M,H)	(VH,H)
C8	(H,H)	(H,VH)	(M,H)	(H,H)	(AH,VH)	(H,H)	(VH,H)
C9	(VH,H)	(M,VH)	(H,H)	(H,H)	(H,H)	(MH,H)	(VH,H)
C10	(VH,H)	(M,VH)	(ML,H)	(H,H)	(H,H)	(M,H)	(AH,H)
C11	(H,VH)	(H,VH)	(H,H)	(H,H)	(H,VH)	(H,H)	(AH,H)
C12	(MH,VH)	(VH,VH)	(H,H)	(H,H)	(H,VH)	(H,H)	(VH,H)
C13	(H,H)	(H,VH)	(M,H)	(MH,H)	(MH,VH)	(M,H)	(H,H)
C14	(H,M)	(M,VH)	(H,H)	(M,H)	(H,H)	(M,H)	(H,H)
C15	(VH,H)	(M,VH)	(H,H)	(MH,H)	(MH,H)	(M,H)	(H,H)

Table 5
Evaluation by DMs on O_3

	DM1	DM2	DM3	DM4	DM5	DM6	DM7
C1	(MH,M)	(H,VH)	(M,M)	(MH,H)	(MH,H)	(ML,H)	(MH,H)
C2	(MH,H)	(L,VH)	(H,H)	(M,H)	(ML,H)	(L,H)	(H,H)
C3	(H,H)	(L,VH)	(ML,M)	(H,H)	(ML,H)	(ML,H)	(VH,H)
C4	(H,H)	(M,VH)	(H,H)	(MH,H)	(ML,H)	(M,L)	(VH,H)
C5	(H,M)	(L,VH)	(M,M)	(H,H)	(M,VH)	(M,L)	(VH,H)
C6	(H,H)	(H,VH)	(MH,H)	(H,H)	(M,VH)	(M,M)	(AH,H)
C7	(MH,H)	(ML,VH)	(M,H)	(MH,H)	(M,VH)	(ML,L)	(VH,H)
C8	(M,H)	(L,VH)	(M,H)	(H,H)	(M,VH)	(L,H)	(VH,H)
C9	(VH,H)	(M,VH)	(H,H)	(H,H)	(H,H)	(M,L)	(VH,H)
C10	(H,H)	(H,VH)	(H,H)	(MH,H)	(MH,VH)	(M,M)	(AH,H)
C11	(H,VH)	(H,VH)	(H,H)	(H,H)	(H,VH)	(H,H)	(AH,H)
C12	(MH,M)	(VH,VH)	(H,H)	(MH,H)	(H,H)	(M,L)	(VH,H)
C13	(H,M)	(L,VH)	(M,H)	(MH,H)	(MH,VH)	(M,H)	(H,H)
C14	(M,M)	(M,VH)	(H,H)	(M,H)	(H,H)	(M,H)	(AH,H)
C15	(H,VH)	(M,VH)	(H,H)	(MH,H)	(MH,H)	(L,H)	(H,H)

Table 6
Evaluation by DMs on O_4

	DM1	DM2	DM3	DM4	DM5	DM6	DM7
C1	(AH,VH)	(AH,VH)	(H,H)	(VH,H)	(H,VH)	(M,H)	(VH,H)
C2	(VH,VH)	(VL,VH)	(H,H)	(H,H)	(H,VH)	(H,H)	(VH,H)
C3	(AH,VH)	(M,VH)	(M,M)	(H,H)	(H,VH)	(M,H)	(VH,H)
C4	(AH,VH)	(H,VH)	(H,H)	(MH,H)	(VH,VH)	(H,H)	(VH,H)
C5	(AH,VH)	(M,VH)	(H,H)	(H,H)	(H,VH)	(MH,H)	(VH,H)
C6	(H,VH)	(H,VH)	(MH,H)	(MH,H)	(MH,VH)	(VH,H)	(AH,H)
C7	(AH,VH)	(M,VH)	(M,,H)	(MH,H)	(M,H)	(ML,H)	(VH,H)
C8	(AH,VH)	(H,VH)	(M,H)	(MH,H)	(VH,VH)	(H,H)	(VH,H)
C9	(AH,VH)	(M,VH)	(H,H)	(H,H)	(H,H)	(H,H)	(VH,H)
C10	(AH,VH)	(M,VH)	(ML,H)	(H,H)	(H,H)	(M,H)	(AH,H)
C11	(AH,VH)	(H,VH)	(H,H)	(H,H)	(VH,H)	(H,H)	(AH,H)
C12	(VH,VH)	(VH,VH)	(H,H)	(H,H)	(VH,H)	(H,H)	(VH,H)
C13	(AH,VH)	(M,VH)	(M,H)	(M,H)	(M,H)	(AH,H)	(VH,H)
C14	(VH,VH)	(M,VH)	(H,H)	(M,H)	(MH,H)	(M,H)	(AH,H)
C15	(AH,VH)	(M,VH)	(H,H)	(MH,H)	(MH,H)	(M,H)	(VH,H)

Using equation in Step 5 of the procedure, the criteria weights $C_j(j=1,2,\dots, 15)$ are obtained as the following weight vector $w_j = (w_1, w_2, \dots, w_{15})$.

$$\begin{aligned} w_1 &= 0.1003, w_2 = 0.0651, w_3 = 0.0969, w_4 = 0.0960, \\ w_5 &= 0.0850, w_6 = 0.0426, w_7 = 0.0892, w_8 = 0.0939, \\ w_9 &= 0.0453, w_{10} = 0.0474, w_{11} = 0.0431, w_{12} = 0.0500, \\ w_{13} &= 0.0639, w_{14} = 0.0421, w_{15} = 0.0391. \end{aligned}$$

A weighted expected Jaccard similarity measure between an alternative O_i and the ideal alternative O^* is calculated as

$$\begin{aligned} W(O^*, O_1) &= 0.5906, W(O^*, O_2) = 0.7948, \\ W(O^*, O_3) &= 0.5963, W(O^*, O_4) = 0.7967. \end{aligned}$$

Hence, the ranking of the alternatives of cars with respect to the 15 criteria under consideration is as follows:

$$O_4 > O_2 > O_3 > O_1.$$

From the ranking order, the most favoured car is O_4 . Moreover, from the criteria weight vector, the most influential factor in influencing decision to purchase a car is its value for money while the least important is the promotion factor.

CONCLUSION

Decision making in a fuzzy environment leads to vague expressions of evaluation. The fuzzy set theory has the ability via fuzzy numbers to represent vague phrases or languages which involve subjectivity of values. Utilising the Z-number concept, decision making process with the expected Jaccard similarity measure technique provides a better decision, owing to the inclusion of the reliability factor, particularly in fuzzy evaluation of multi-criteria decision making problem. The reliability factor in the Z-number representation is a useful tool in the evaluation process as the background of the decision makers usually defers to one another in terms of experience, knowledge, authority level and so on. As an extension of this effort, different similarity measures may be used in the proposed procedure and comparison of results may be made. Furthermore, the proposed procedure can be applied in other decision making problems in various areas with fuzzy environment that also considers the level of confidence of evaluators.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Teknologi MARA (UiTM) Malaysia, in particular to the Faculty of Computer & Mathematical Sciences, for its financial support to undertake this research.

REFERENCES

- Allahviranloo, T., Abbasbandy, S., & Hajighasemi, S. (2012). A new similarity measure for generalized fuzzy numbers, *Neural Computing and Applications*, 21(1), 289-294.
- Bojadziev, G., & Bojadziev, M. (2007). *Fuzzy logic for business, finance and management* (2nd Ed.). United States, USA: World Scientific Publishing Company.
- Bozbura, F. T., Beskese, A., & Kahraman, C. (2007). Prioritization of human capital measurement indicators using fuzzy AHP. *Expert Systems with Application*, 32(4), 1100-1112.
- Chen, C. T. (2000). Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets and Systems*, 114(1), 1-9.
- Chen, S. J., & Chen, S. M. (2003). Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *IEEE Transactions on fuzzy systems*, 11(1), 45-56.
- Chen, S. H., & Hsieh, C. H. (1999). Ranking of generalized fuzzy numbers with graded mean integration representation. In *Proceedings of the Eight International Fuzzy Systems Association World Congress* (pp. 899-902). Republic of China.
- Dongyan, L., & Xuan, B. (2008). *Car purchasing behavior in Beijing: An empirical investigation*. (Master Thesis). University of Umeå: Umeå School of Business and Economics, Sweden.
- Easton, A. (1976). *Decision making: A short course for professional*. New York: Wiley.
- Farhadinia, B., & Ban, A. I. (2013). Developing new similarity measures of generalized intuitionistic fuzzy numbers and generalized interval-valued fuzzy numbers from similarity measures of generalized fuzzy numbers. *Mathematical and Computer Modelling*, 57(3-4), 812-825.
- Gardashova, L. A. (2014) Application of operational approaches to solving decision making problem using Z-numbers. *Applied Mathematics*, 5(9), 1323-1334.
- Goli, D. (2013). Group fuzzy TOPSIS methodology in computer security software selection. *International Journal of Fuzzy Logic Systems*, 3(2), 29-47.
- Kaboli, A., Aryanezhad, M. B., Shahanaghi, K., & Niroomand, I. (2007). A new method for plant location selection problem: A fuzzy-AHP. In *International Conference on Systems, Man and Cybernetics* (pp. 582-586). Institute of Electrical and Electronics Engineers, IEEE.
- Kang, B., Wei, D., Li, Y., & Deng, Y. (2012a). Decision making using Z-numbers under uncertain environment. *Journal of Information and Computational Science*, 8(7), 2807-2814.
- Kang, B., Wei, D., Li, Y., & Deng, Y. (2012b). A method of converting Z-number to classical fuzzy number. *Journal of Information and Computational Science*, 9(3), 703-709.
- Sridevi, B., & Nadarajan, R. (2009). Fuzzy similarity measure for generalized fuzzy numbers. *International Journal of Open Problems in Computational Mathematics*, 2(2), 239-253.
- Wang, X., Zou, Z., & Zou, H. (2013). Water quality evaluation of Haihe River with fuzzy similarity measure methods. *Journal of Environmental Sciences*, 25(10), 2041-2046.
- Wu, D., & Mendel, J. M. (2008). A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets. *Information Sciences*, 178(2), 381-402.
- Xu, X., Zhang, L., & Wan, Q. (2012). A variation coefficient similarity measure and its application in emergency group decision-making. *Systems Engineering Procedia*, 5, 119-124.

- Ye, J. (2012a). Multicriteria decision-making method using the Dice similarity measure between expected intervals of trapezoidal fuzzy numbers. *Journal of Decision Systems*, 21(4), 307-317.
- Ye, J. (2012b). Multicriteria group decision-making method using vector similarity measures for trapezoidal intuitionistic fuzzy numbers. *Group Decision and Negotiation*, 21(4), 519-530.
- Ye, J. (2012c). The Dice similarity measure between generalized trapezoidal fuzzy numbers based on the expected interval and its multicriteria group decision making method. *Journal of the Chinese Institute of Industrial Engineers*, 29(6), 375-382.
- Yong, D., Wenkang, S., Feng, D., & Qi, L. (2004). A new similarity measure of generalized fuzzy numbers and its application to pattern recognition. *Pattern Recognition Letters*, 25(8), 875-883.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information Control*, 8(3), 338-353.
- Zadeh, L. A. (2011). A note on z-numbers. *Information Sciences*, 181(14), 1-10.
- Zeinalova, L. M. (2014). Expected Utility Based Decision Making Under Z-Information. *Intelligent Automation and Soft Computing*, 20(3), 419-431.

Sample Size and Non-Normality Effects on Goodness of Fit Measures in Structural Equation Models

Ainur, A. K.^{1*}, Sayang, M. D.¹, Jannoo, Z.² and Yap, B. W.¹

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

²Faculty of Social Studies and Humanities, University of Mauritius, Reduit 80837 Mauritius

ABSTRACT

A Structural Equation Model (SEM) is often used to test whether a hypothesised theoretical model agrees with data by examining the model fit. This study investigates the effect of sample size and distribution of data (normal and non-normal) on goodness of fit measures in structural equation model. Simulation results confirm that the GoF measures are affected by sample size, whereas they are quite robust when data are not normal. Absolute measures (GFI, AGFI, RMSEA) are more affected by sample size while incremental fit measures such as TLI and CFI are less affected by sample size and non-normality.

Keywords: Structural equation model, goodness-of-fit, non-normality, simulation

INTRODUCTION

Structural equation model, also known as simultaneous equation model, is one of the advanced multivariate regression models. Structural equation modeling (SEM) is a statistical technique that combines elements in traditional multivariate models, such as regression analysis, factor analysis, and simultaneous equation modeling. In contrast with the more traditional multivariate linear model, in SEM, the response variable in one regression equation can appear as a predictor in another regression equation. The variables in SEM may influence other variables either directly or through other variables as intermediaries (Fox, 2002). The aim of SEM is to test whether the hypothesised

theoretical model is consistent with the data in reflecting the theory by examining the model fit (Hair Jr, Black, Babin, & Anderson, 2010). The model fits the data when the covariance matrix is equal or approaches the sample covariance matrix (Lei & Wu, 2007).

The most common softwares used for modeling data are LISREL, AMOS, EQS and MPLUS. The Maximum Likelihood

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail addresses:

memiera87@yahoo.com (Ainur, A. K.),
sayang@tmsk.uitm.edu.my (Sayang, M. D.),
z.jannoo@uom.ac.mu (Jannoo, Z.),
beewah@tmsk.uitm.edu.my (Yap, B. W.)
*Corresponding Author

(ML) is the default parameter estimation method, while other estimation methods include the Generalized Least Square (GLS), Weighted Least Square (WLS) and Asymptotically Distribution Free (ADF). SEM is widely used in many fields such as business research (Lei & Wu, 2007; Hooper, Coughlan, & Mullen, 2008), psychology (Mulaik et al., 1989; Curran et al., 1996; Bandalos, 2002; Schermelleh-engel, Moosbrugger, & Müller, 2003) and marketing (Bearden, Sharma, & Teel, 1982).

In SEM, the measurement model and the structural model need to be tested. The Goodness of Fit (GoF) measures commonly used to confirm if the model fit the data are Chi-Square (χ^2), Root Mean Square Error of Approximation (RMSEA), Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Comparative Fit Index (CFI), Normed Fit Index (NFI), and Non-Normed Fit Index (NNFI) or also known as Tucker-Lewis Index (TLI). Some GoF measures are affected by the sample size (Marsh, Balla, & McDonald, 1988; Rocha & Chelladurai, 2012), as well as the distribution of data (Schermelleh-engel et al., 2003; Sharma, Mukherjee, Kumar, & Dillon, 2005).

Computer-intensive procedures are used in simulation study to generate random data that closely represent the real-world data. Monte Carlo simulation is one of the techniques applied for generating data using random sampling from probability distribution of interest in order to fit the model in SEM.

Bearden et al., (1982) performed simulation to examine the effect of sample size on the Chi-Square statistic, NFI, Normalized Residual Index and on the construct shared variance as well as reliability estimators. The simulation study was conducted using two-construct and four-construct models which consist of three indicators for each construct. The simulation was conducted for sample sizes of 25, 50, 75, 100, 500, 1,000, 2,500, 5,000 and 10,000. It was found that the overall fit statistic provided by LISREL is accurate for large sample sizes under simple model and the interpretation of Chi-Square can be affected by the sample size and model complexity. However, this study was limited to two causal models and used the assumption of multivariate normality assumption. Hence, recommendations for future research include assessing the effects of the violation of normality assumption and conducting studies on more complex models by varying the number of constructs and indicators.

The number of indicators per factor can affect the fit measures in structural equation modeling. A simulation study (Ding, Velicer, & Harlow, 1995) using the EQS computer program was conducted by including five different numbers of indicators (2, 3, 4, 5, and 6), three levels of factor loading size (0.5, 0.7 and 0.9), four levels of sample size (50, 100, 200 and 500), and two estimation methods (ML and WLS). The fit measures considered in the study were Chi-Square per degree of freedom ratio (χ^2/df), NFI, NNFI, Centrality m Index, RNI, and CFI. It was found that GLS produced more improper solutions than ML, while the GoF was less affected by the improper solutions except for NFI; NNFI was found to be unaffected by the estimation method. Improper solution (zero or negative error variances) occurs when a combination of the number of indicators per factor is less than three, the sample size is small and the factor loading is low.

A study by Sharma et al. (2005) applied simulation analysis to investigate the effect of sample size, number of indicators, factor loading sizes and factor correlation sizes on the mean value and percentage of models acceptance when pre-specified cutoff values were compared

for Normed Version (NNCP), Relative Non-Centrality Index (RNI), TLI, RMSEA and GFI. Correct and misspecified models of correlated two-factor, four-factor, six-factor and eight-factor confirmatory factor models with four indicators for each factor were considered. Data was generated using GGNSM procedure in Fortran programming involving 100, 200, 400 and 800 sample sizes with 100 replications for each sample size. Three factor loadings (0.3, 0.5 and 0.7) and three correlations among the factors (0.3, 0.5 and 0.7) were employed. This study found that GFI, NNCP, RNI and TLI were sensitive to sample sizes corresponding with the number of indicators, while RMSEA was dependent on sample size but independent of number of indicators. Based on the performance and sensitivity of misspecification, RNI and TLI were found to be better than NNCP, followed by RMSEA and GFI. These findings suggest that GFI should not be used to evaluate model fit. RNI and TLI are recommended for factor loading of 0.5 and above while NNCP are recommended to be used in conjunction with RNI and TLI. Additionally, RMSEA is recommended to be used in conjunction with NNCP, TLI and RNI.

A simulation study was conducted to test the effect of sample size on certain fit measures such as Chi-Square, Standardized Root Mean Square (SRMR) and CFI by Iacobucci (2010). The study consisted of six different sample sizes (30, 50, 100, 200, 500 and 1000) with 2000 replications for each sample size. It was reported that size of the sample should be more than 50 since the GoF measures improve when sample size increases. The ideal number of indicators for each construct was three and Maximum Likelihood estimation performed better because it was relatively robust to the multivariate normality assumption.

A Monte Carlo simulation procedure to compare the performance of Covariance-Based SEM (CB-SEM) and Partial Least Square SEM (PLS-SEM) under normality and non-normality conditions was carried out by Jannoo, Yap, Auchoybur, & Lazim (2014). Under the normality condition, data was generated from a normal distribution for the latent constructs and residual. The model consisted of one endogenous and four exogenous constructs with three indicators for each. Under the non-normality condition, data was generated from Chi-Square distribution for the indicators with 2 degrees of freedom. The skewness as reported at 1.983 while kurtosis was at 4.375. The sample sizes generated for both data sets were 20, 40, 90, 150 and 200 and replication was done for 500 times for both CB-SEM and PLS-SEM by using R programming language procedure. Their results revealed that when condition is normal and sample size is small, CB-SEM becomes inaccurate while PLS-SEM estimates are closer to the true parameter value. On the other hand, when the sample is large, CB-SEM is better than PLS-SEM due to its lower variability. Under non-normality condition, when the sample is small, CB-SEM is inaccurate, but has better accuracy compared with PLS-SEM when sample size is large.

Previous simulation studies involved different sample sizes and GoF measures. However, these studies could not provide a comprehensive and conclusive findings on six common GoF measures.

Thus, this study examined the effect of different sample sizes and distributions on GoF measures in the structural equation model. The objectives of this study were:

- (a) to identify the effect of different sample sizes on GoF measures
- (b) to identify the effect of different distributions on GoF measures

METHODS

Various types of model fit measures reported by Hair Jr et al. (2010) can be applied to examine the model. There are three GoF categories: absolute fit measures, incremental fit measures and parsimony fit measures.

The first category, absolute fit measures, determine how well a priori model fits the sample data and demonstrates which proposed model has the most superior fit. The methods used in this category are Chi-Square test (χ^2), RMSEA, Root Mean Square Residual (RMR), SRMR and GFI. The second category is incremental fit measures or known as comparative fit measures. These GoF measures compare the values to a baseline model. They include NFI, TLI and CFI. The third category is parsimony fit measures which includes AGFI, Parsimony Goodness-of-Fit Index (PGFI) and Parsimonious Normed Fit Index (PNFI). Parsimony fit measures adjusts the number of parameters in the estimated model. However, this study only covered six common GoF measures which are GFI, AGFI, RMSEA, NFI, TLI and CFI.

Maximum Likelihood (ML) is the method of parameter estimation most commonly used and had been set as a default parameter estimation technique in most of the statistical software. It is obtained by comparing the actual covariance matrices representing the relationship between variables with the estimated covariance matrices of the best fitted model. The minimum fit function is:

$$F[\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})] = \log|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})| - \log|\mathbf{S}| + tr[\mathbf{S}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}] - p \quad (1)$$

where, $F[\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})]$ is the minimum value of the fit function, $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ is the estimated covariance matrix, \mathbf{S} is the sample covariance matrix, $tr[\]$ is the trace of a matrix and p is the number of observed variables.

The RMSEA is based on the difference between the \mathbf{S} and $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ to indicate the equivalence between the structural equation model and empirical data. $RMSEA(\hat{\varepsilon}_a)$ is the square root of the estimated discrepancy due to the approximation over the degree of freedom (Schermelleh-engel et al., 2003), as shown in the equation below:

$$\hat{\varepsilon}_a = \sqrt{\max\left\{\left(\frac{F[\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})]}{df} - \frac{1}{N-1}\right), 0\right\}} \quad (2)$$

where, $df_t = s - t$ is the number of degree of freedom and N is the sample size.

The GFI and AGFI are comparison measures where the model of interest is compared with a baseline model. The baseline model is the null model in which all measured variables are unrelated to each other. The GFI and AGFI are obtained using the following equations (Schermelleh-engel et al., 2003) [8] :

$$GFI = 1 - \frac{F_t}{F_n} = 1 - \frac{\chi_t^2}{\chi_n^2} \quad (3)$$

where, χ_t^2 is the Chi-Square of the target model, χ_n^2 is the Chi-Square of the null model (baseline model) and F is the corresponding minimum fit function value.

$$AGFI = 1 - \frac{df_n}{df_t} (1 - GFI) = 1 - \frac{\chi_t^2 / df_t}{\chi_n^2 / df_n} \tag{4}$$

where, $df_t = s - t$ is the number of degree of freedom for the target model, $df_n = s = p(p + 1) / 2$ is the number of freedom for the null model.

NFI and NNFI, also known as TLI, are obtained using the formula below (Schermelleh-engel et al., 2003):

$$NFI = \frac{\chi_i^2 - \chi_t^2}{\chi_i^2} = 1 - \frac{\chi_t^2}{\chi_i^2} = 1 - \frac{F_t}{F_i} \tag{5}$$

where, χ_i^2 is the Chi-Square of the independent model (baseline model),

$$TLI = \frac{(\chi_i^2 / df_i) - (\chi_t^2 / df_t)}{(\chi_i^2 / df_i) - 1} = \frac{(F_i / df_i) - (F_t / df_t)}{(F_i / df_i) - 1 / (N - 1)} \tag{6}$$

The CFI is introduced to overcome the problem of underestimation of fit when the sample is too small for NFI (Bentler, 1990). The CFI can be obtained using the following equation (Schermelleh-engel et al., 2003):

$$CFI = 1 - \frac{\max[(\chi_t^2 - df_t), 0]}{\max[(\chi_i^2 - df_i), (\chi_t^2 - df_t), 0]} \tag{7}$$

where, max is the maximum of the values in brackets.

The recommended levels to be used to represent an acceptable model fit for the GoF measures are presented in Table 1.

Table 1
Acceptable Threshold Levels

Fit Indices	Acceptable Threshold Levels
GFI	>0.90 (Hooper et al., 2008)
AGFI	>0.85 (Schermelleh-engel et al., 2003) >0.95 (Hooper et al., 2008)
RMSEA	<0.06 (Hu & Bentler, 1999)
NFI	>0.90 (Arbuckle, 1995)
TLI	>0.95 (Hu & Bentler, 1999)
CFI	>0.95 (Hu & Bentler, 1999)

SIMULATION DESIGN

This study investigated the performance of GoF measures under normal and non-normal conditions in SEM. Data were generated using a specified model given by Goodhue, Lewis, & Thompson (2012). The model included five constructs (one endogenous variable and four exogenous variables) with three indicators for each constructs. For each of the three indicators, standardised loading was fixed at 0.70, 0.80 and 0.90 respectively. Data was generated from Normal distribution $N(0,1)$, and non-normal distributions which are Chi-Square distribution with 2 degrees of freedom, $\chi^2(2)$ and Chi-Square distribution with 0.5 degrees of freedom, $\chi^2(0.5)$. The distribution of $N(0,1)$ is a symmetric normal distribution, while $\chi^2(2)$ and $\chi^2(0.5)$ are skewed distribution with skewness of 2 and 4, respectively.

For each data set, various sample sizes were considered to investigate their effect on GoF measures, such as $n= 100, 150, 200, 250, 300, 500, 1000, 1500$ and 2000. The simulation involved 500 replications for each sample size and condition by using R, open source programming language. The minimum sample size of 100 was suggested by Hair Jr et al. (2010) for model containing constructs of less than six with indicators of more than three for each construct and possessing high item communalities (≥ 0.6). The theoretical model used in this study is shown in Figure 1.

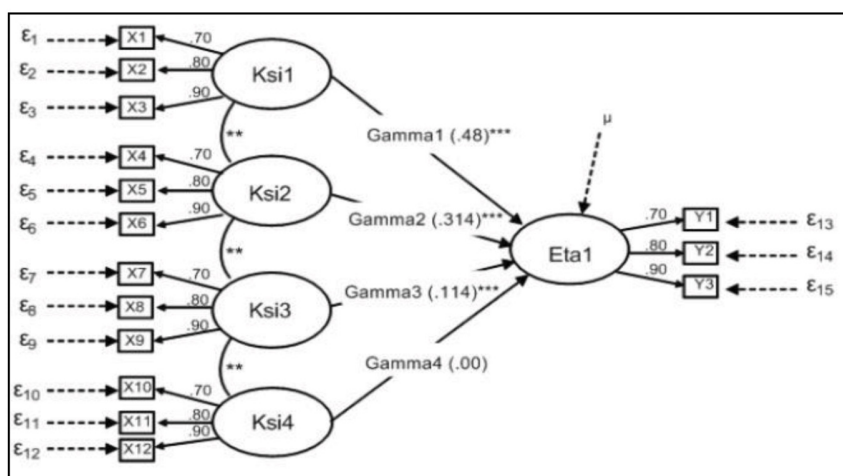


Figure 1. Theoretical model - (Source: Goodhue, 2012)

RESULTS AND DISCUSSIONS

This section reports the results of GoF measures under normal and non-normal distributions. Box-plots were drawn to represent the GoF measures for various sample sizes under normal and non-normal conditions.

The performance of GoF measures for various sample sizes under normal and non-normal conditions are shown in Table 2. The GoF measures improved as the sample size

increased for all distributions. The TLI and CFI showed better fit compared with GFI and AGFI for all sample sizes. Thus, incremental fit measures were found to be less affected by sample sizes.

Table 2
Mean of GoF measures

n	GFI	AGFI	RMSEA	NFI	TLI	CFI
^a 100	0.904	0.856	0.023	0.897	0.991	0.989
^b 100	0.902	0.853	0.028	0.896	0.986	0.986
^c 100	0.903	0.855	0.026	0.898	0.990	0.986
^a 150	0.933	0.900	0.018	0.929	0.995	0.993
^b 150	0.932	0.899	0.020	0.928	0.994	0.992
^c 150	0.933	0.900	0.019	0.930	0.995	0.992
^a 200	0.949	0.923	0.015	0.946	0.997	0.995
^b 200	0.949	0.923	0.015	0.946	0.996	0.995
^c 200	0.949	0.924	0.015	0.947	0.997	0.994
^a 250	0.959	0.939	0.012	0.957	0.999	0.997
^b 250	0.959	0.939	0.013	0.957	0.998	0.996
^c 250	0.959	0.938	0.013	0.957	0.998	0.996
^a 300	0.966	0.948	0.011	0.964	0.999	0.997
^b 300	0.965	0.948	0.011	0.964	0.999	0.997
^c 300	0.966	0.948	0.011	0.964	0.999	0.997
^a 500	0.979	0.968	0.008	0.978	0.999	0.998
^b 500	0.979	0.968	0.008	0.978	0.999	0.998
^c 500	0.979	0.969	0.008	0.978	1.000	0.998
^a 1000	0.989	0.984	0.005	0.989	1.000	0.999
^b 1000	0.989	0.984	0.006	0.989	1.000	0.999
^c 1000	0.989	0.984	0.006	0.989	1.000	0.999
^a 1500	0.993	0.989	0.004	0.993	1.000	0.999
^b 1500	0.993	0.989	0.004	0.993	1.000	0.999
^c 1500	0.993	0.989	0.004	0.993	1.000	0.999
^a 2000	0.995	0.992	0.004	0.994	1.000	1.000
^b 2000	0.995	0.992	0.004	0.994	1.000	1.000
^c 2000	0.995	0.992	0.004	0.994	1.000	1.000

Note: ^aNormal N(0,1) ^bChi-Square (2) ^cChi-Square (0.5)

The results in Table 2 were represented in box-plots as shown in Figure 2 for comparison purpose. Only the box-plot for GFI, RMSEA, TLI and CFI measures were presented because the pattern for AGFI and NFI measures were found to be similar with GFI.

Results show that GFI and AGFI are affected by sample size. These results contradicted the findings of Joreskog & Sorbom (1982) and Bagozzi & Yi (1988) who reported that GFI and AGFI are independent of sample size. However, our results confirm the findings of Fan, Thompson, & Wang (1999) who reported that GFI and AGFI are overly affected by sample size. GFI and AGFI are found to be robust against non-normality and these results are consistent with the findings by Joreskog & Sorbom (1982) and Bagozzi & Yi (1988). The behaviour of NFI is similar to GFI and AGFI.

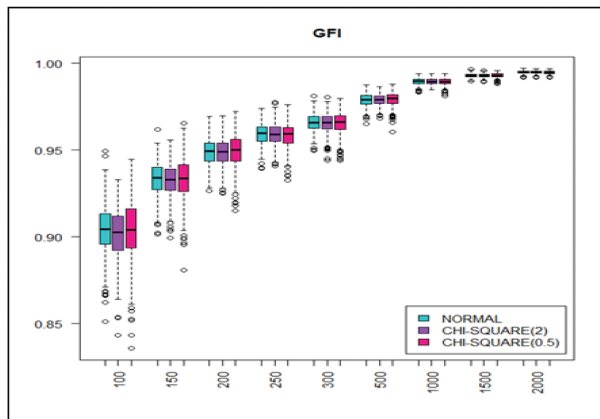


Figure 2(a). Boxplots for GFI

Figure 2(b) and Figure 2(c) show the results of RMSEA and TLI for various sample sizes under normal and non-normal conditions. The boxplot (Figure 2(b)) shows that RMSEA is affected by sample size and less affected by non-normality when sample size is large. Figure 2(c) shows that TLI is less affected by sample size and this result supports the findings of Fan et al. (1999). However, TLI is not affected by distribution of data. According to Bentler (1990), TLI is difficult to interpret since the value can exceed the range (0-1).

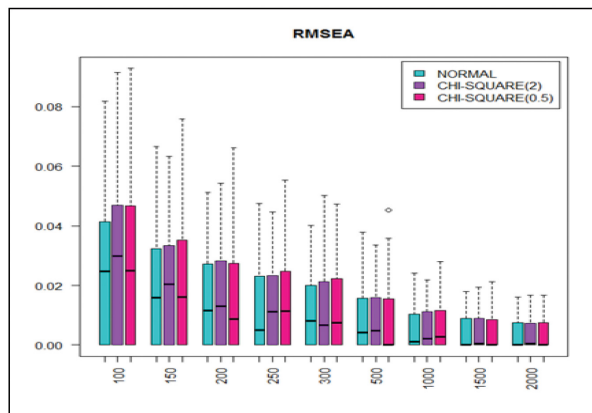


Figure 2(b). Boxplots for RMSEA

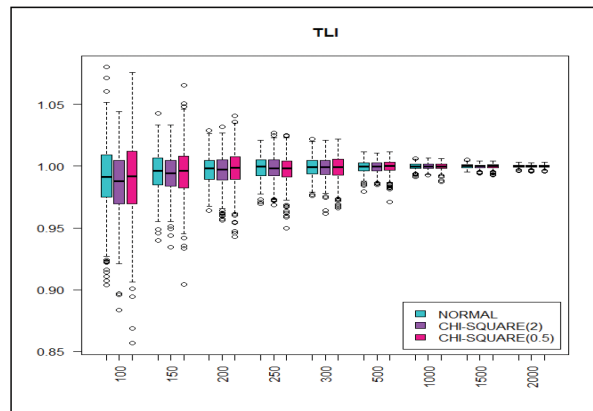


Figure 2(c). Boxplots for TLI

Figure 2(d) displays the performance of CFI for various sample sizes under normal and non-normal conditions. The boxplot show that CFI is less affected by sample size and this result supports the findings of Fan et al. (1999). CFI is also less affected by distribution of data.

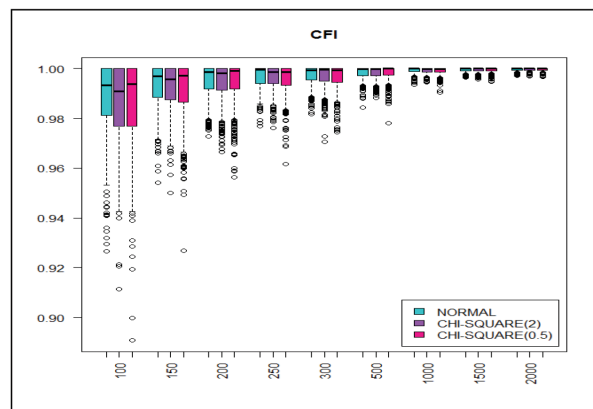


Figure 2(d). Boxplots for CFI

These boxplots show that GoF measures improved with the increase in sample size as expected. However, the simulation results indicated that GoF measures are not severely affected if distributions are non-normal.

CONCLUSIONS

This simulation study compared the effect of sample size and distribution of data on GoF measures for SEM. We found that the incremental fit measures are less affected by sample size and when distribution of data is not normal. However, severe non-normality due to the presence of outliers might affect GoF for SEM. Future simulation study can investigate the GoF measures in the presence of outliers and when data is non-normal with high multivariate skewness and kurtosis.

REFERENCES

- Arbuckle, J. L. (1995). *Amos 17 User 's Guide* (pp. 1995-2005). Chicago, IL: SmallWaters Corporation.
- Bagozzi, R. R., & Yi, Y. (1988). On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*, 16(1), 74–94.
- Bandalos, D. L. (2002). The Effects of Item Parceling on Goodness-of-Fit and Parameter Estimate Bias in Structural Equation Modeling. *Structural Equation Modeling*, 9(1), 78–102.
- Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample Size Effects on Chi Square and Other Statistics Used in Evaluating Causal Models. *Journal of Marketing Research*, 19(4), 425–430.
- Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, 107(2), 238–246.
- Curran, P. J., West, S. G., Finch, J. F., Aiken, L., Bentler, P., & Kaplan, D. (1996). The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis. *Psychological Methods*, 1(1), 16–29.
- Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of Estimation Methods, Number of Indicators per Factor, and Improper Solutions on Structural Equation Modeling Fit Indices. *Structural Equation Modeling*, 2(2), 119–144.
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6(1), 56–83. doi:10.1080/10705519909540119
- Fox, J. (2002). Structural Equation Models. *Appendix to An R and S-PLUS Companion to Applied Regression*, 1–20.
- Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS Have Advantages for Small Sample Size or Non-Normal Data? *MIS Quarterly*, 36(3), 981–A16.
- Hair JR, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling : Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, 6(1), 53–60.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi:10.1080/10705519909540118
- Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, 20(2010), 90–98. doi:10.1016/j.jcps.2009.09.003
- Jannoo, Z., Yap, B. W., Auchoybur, N., & Lazim, M. A. (2014). The Effect of Nonnormality on CB-SEM and PLS-SEM Path Estimates. *International Journal of Mathematical, Computational, Physical and Quantum Engineering*, 8(2), 285–291.
- Joreskog, K. G., & Sorbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, 19(4), 404–416.
- Lei, P. W., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: issues and practice*, 26(3), 33–43.

- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size. *Psychological Bulletin*, *103*(3), 391–410. doi:10.1037//0033-2909.103.3.391
- Mulaik, S. A., James, L. R., Alstine, J. Van, Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of Goodness-of-Fit Indices for Structural Equation Models. *Psychological Bulletin*, *105*(3), 430–445.
- Rocha, C. M., & Chelladurai, P. (2012). Item Parcels in Structural Equation Modeling: an Applied Study in Sport Management. *International Journal of Psychology and Behavioral Sciences*, *2*(1), 46–53. doi:10.5923/j.ijpbs.20120201.07
- Schermelleh-engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models : Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, *8*(2), 23–74.
- Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, *58*(7), 935-943.



A Method for Performing Short Time Series Prediction

Pooi, A. H.

Sunway University Business School, Sunway University, 47500 Bandar Sunway, Malaysia

ABSTRACT

This paper offers a technique to construct a prediction interval for the future value of the last variable in the vector \mathbf{r} of m variables when the number of observed values of \mathbf{r} is small. Denoting $\mathbf{r}(t)$ as the time- t value of \mathbf{r} , we model the time- $(t+1)$ value of the m -th variable to be dependent on the present and $l-1$ previous values $\mathbf{r}(t)$, $\mathbf{r}(t-1)$, ..., $\mathbf{r}(t-l+1)$ via a conditional distribution which is derived from an $(ml+1)$ -dimensional power-normal distribution. The $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points of the conditional distribution may then be used to form a prediction interval for the future value of the m -th variable. A method is introduced to estimate the above $(ml+1)$ -dimensional power-normal distribution such that the coverage probability of the resulting prediction interval is nearer to the target value $1-\alpha$.

Keywords: Multivariate power-normal distribution, prediction interval, coverage probability

INTRODUCTION

Time series models usually require big data set in order to produce reasonably good out-of-sample prediction of future values. In practice, there are many situations in which the data size is small. For example, when a new product or service is launched, the newly recorded series is likely to be a short time series. Short time series may also be an outcome when a corporation has undergone a business process re-engineering so that most past data become irrelevant.

When the short time series exhibits stable seasonal pattern, several authors attempted to predict the end-of-season total of the $(n+1)$ -th season, given the observation series of the first n seasons and part of the observation series in the $(n+1)$ -th season (see Hertz & Schaffir (1960); Murray & Silver (1966); Chang & Fyffe (1971); Green & Harrison (1973); Box & Jenkins (1976); Oliver (1987); Guerrero & Elizondo (1997); Chen & Fomby (1999); Alba & Mendoza (2001), (2006)).

This paper offers a method to perform the one-step prediction of the future value of a short time series. Initially let us assume that we have a small number of observations of a vector $\mathbf{r} = (r_1, r_2, \dots, r_m)$ of m variables which

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail address:

ahhinp@sunway.edu.my (Pooi, A. H.)

have been recorded at evenly spaced time intervals. Let t be the present time, and $\mathbf{r}(t)$ the value of \mathbf{r} at time t . The time- $(t+1)$ value of the m -th variable r_m of \mathbf{r} is modelled to be dependent on the present and $l-1$ previous values $\mathbf{r}(t), \mathbf{r}(t-1), \dots, \mathbf{r}(t-l+1)$ via a conditional distribution which is derived from an $(ml+1)$ -dimensional power-normal distribution. A prediction interval for the value of r_m at time $t+1$ may be formed from the $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points of the conditional distribution.

The above $(ml+1)$ -dimensional power-normal distribution may be estimated by using

- (A) the values of $\mathbf{r}(u)$ for $u \leq t$ or
- (B) the values of $\mathbf{r}(u)$ for $u \leq t$ and an estimated value of r_m at time $t+1$

The estimation of the value of r_m at time $t+1$ in (B) may be performed by using an extrapolation based on a low degree polynomial fitted to a small number n_r of values of r_m at time $t, t-1, \dots, t-n_r+1$.

The prediction interval based on the $(ml+1)$ - dimensional power-normal distribution estimated by using the values in (A) (or B) may be referred to as a Type A (or B) prediction interval.

Type A prediction interval has a coverage probability which may be fairly low. On the other hand, the coverage probability of Type B prediction interval tends to be larger than that of Type A prediction interval. When we take a union of the Type B prediction interval based on linear extrapolation with that based on quadratic extrapolation, the resulting prediction interval may have a coverage probability which is 1.5 to 2 times of the coverage probability of Type A prediction interval.

The paper is organised as follows: Section 2 provides a brief description of the method based on multivariate power-normal distribution for finding prediction intervals. In Section 3, Type B prediction intervals and their union are compared with Type A prediction interval. Section 4 concludes the paper.

METHOD BASED ON MULTIVARIATE POWER-NORMAL DISTRIBUTION

Let us begin with the non-normal distribution given in Yeo and Johnson (2000). The authors in Yeo and Johnson (2000) have introduced the following power transformation

$$\tilde{\varepsilon} = \psi(\lambda^+, \lambda^-, z) = \begin{cases} [(z+1)^{\lambda^+} - 1] / \lambda^+, & (z \geq 0, \lambda^+ \neq 0) \\ \log(z+1), & (z \geq 0, \lambda^+ = 0) \\ -[(-z+1)^{\lambda^-} - 1] / \lambda^-, & (z < 0, \lambda^- \neq 0) \\ -\log(-z+1), & (z < 0, \lambda^- = 0) \end{cases} \quad (1)$$

Let \mathbf{y} be a vector consisting of k correlated random variables. The vector \mathbf{y} is said to have a k -dimensional power-normal distribution with parameters $\boldsymbol{\mu}, \mathbf{H}, \lambda_i^+, \lambda_i^-, \sigma_i, 1 \leq i \leq k$ if

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{H}\boldsymbol{\varepsilon} \quad (2)$$

where $\boldsymbol{\mu} = E(\mathbf{y})$, \mathbf{H} is an orthogonal matrix, $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_k$ are uncorrelated,

$$\varepsilon_i = \sigma_i [\tilde{\varepsilon}_i - E(\tilde{\varepsilon}_i)] / \{\text{var}(\tilde{\varepsilon}_i)\}^{1/2}, \tag{3}$$

$\sigma_i > 0$ is a constant, and $\tilde{\varepsilon}_i$ has a power-normal distribution with parameters λ_i^+ and λ_i^- .

When the values of y_1, y_2, \dots, y_{k-1} are given, we may find an approximation for the conditional probability density function (pdf) of y_k by using the numerical procedure given in Pooi (2012).

We may choose the variables y_1, y_2, \dots, y_k to be those given by the values of components of $\mathbf{r}(t-l+1), \dots, \mathbf{r}(t-1), \mathbf{r}(t)$ together with the value of $r_m(t+1)$ in the lag- $(l-1)$ model.

From the data which span over T units of time, we can form a table of $T-l$ rows with each row representing an observed value of (y_1, y_2, \dots, y_k) . From the table, we can form the i_w -th moving window of size n_w from the i_w -th row till the $(i_w + n_w - 1)$ -th row. We can form a total of $T-l-n_w$ such windows of size n_w . We next find a k -dimensional power-normal distribution for (y_1, y_2, \dots, y_k) using the data in the i_w -th window.

Letting y_1, y_2, \dots, y_k be given by the first $k-1$ values in the $(i_w + n_w)$ -th row immediately after the i_w -th window, we may now find a conditional distribution for y_k when y_1, y_2, \dots, y_{k-1} are given. The mean $\hat{y}_k^{(i_w)}$ of the conditional distribution is then an estimate of the value of the last component at the next unit of time. On the other hand, the $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points of the conditional distribution may be regarded as the lower and upper limits of the nominally $100(1-\alpha)\%$ out-of-sample prediction interval for the value of the last component at the next unit of time. This prediction interval may be referred to as a Type A prediction interval.

The mean absolute percentage error (MAPE) is given by

$$\text{MAPE} = \left[\frac{1}{T-l-n_w} \sum_{i_w=1}^{T-l-n_w} \left| \hat{y}_k^{(i_w)} - y_k^{(i_w)} \right| / y_k^{(i_w)} \right] \times 100\% \tag{4}$$

where $y_k^{(i_w)}$ is the observed value of the last component at the next unit of time. The value of MAPE which is small is an indication that the predictive power of the model is good.

The coverage probability of the prediction interval may be estimated by the proportion of prediction intervals which include the observed value of the last component at the next unit of time. Meanwhile, the expected length of the prediction interval may be estimated by the average length of the prediction intervals. When the estimated coverage probability is close to the target value $1-\alpha$, a small value of the average length is indicative of good predictive power of the model.

When the size T of the data is small, the size n_w of the window that we can form will also be small. The first $k-1$ values in the $(i_w + n_w)$ -th row immediately after the i_w -th window may be then not within the feasible range specified by the $(k-1)$ -dimensional power-normal distribution fitted to the data in the first $k-1$ columns in the i_w -th window. The prediction of the k -th component y_k in the light of the first $k-1$ components would thus be unreliable.

A remedy for the problem caused by the non-concordance of the first $k-1$ values in the $(i_w + n_w)$ -th row with the fitted $(k-1)$ -dimensional distribution is to augment the i_w -th window by another row formed by the first $k-1$ values in the $(i_w + n_w)$ -th row and the value obtained by an extrapolation of a low degree polynomial fitted to a small number n_r of values in the lower portion of the last column in the i_w -th window, and fit a k -dimensional power-normal distribution to the augmented i_w -th window. A prediction interval for the k -th component in the light of the first $k-1$ components may then be found as before. The resulting prediction interval shall be called a Type B prediction interval.

We may take a union of the Type B prediction interval based on linear extrapolation with that based on quadratic extrapolation to form yet another prediction interval. The estimate associated with the interval formed by the union operation for the value of the k -th component y_k one unit of time ahead may be taken to be the average of the means of the conditional distributions used for forming the two Type B prediction intervals.

PERFORMANCE OF PREDICTION INTERVALS

Monthly data from January 2006 to December 2012 for six selected macroeconomic variables shall be used to investigate Type A and Type B prediction intervals. The six selected variables are Gross Domestic Product, Money Supply (M2), Inflation Rate, Oil Price, Gold Price and Kuala Lumpur Composite Index (KLCI).

From the Malaysian data based on six variables, we can obtain a total of $T-1 = 83$ observed values of the vector

$$[r_1(t), r_2(t), \dots, r_6(t), r_6(t+1)] \tag{5}$$

where $r_1(t), r_2(t), \dots, r_6(t)$ are the values of the six variables in the t -th month, $t = 1, 2, \dots, 83$.

Setting $k = 7, y_i = r_i(t), 1 \leq i \leq 6, y_7 = r_6(t+1), n_w = 13, 15, 17, 20, 25, 30, 40, \text{ or } 50$, we apply the method in Section 2 to find

1. Type A prediction interval
2. Type B prediction interval based on linear extrapolation
3. Type B prediction interval based on quadratic extrapolation
4. Union of the Type B prediction intervals in 2 and 3.

The measures of the performance of the four prediction intervals are shown in Tables 1-3.

Table 1
Coverage probability of prediction interval ($l = 1, k = 7, n_r = 6, \alpha = 0.05$)

n_w	Type A	Type B (Linear)	Type B (Quadratic)	Union
13	0.442857	0.500000	0.442857	0.685714
15	0.397059	0.602941	0.529412	0.794118
17	0.439394	0.606061	0.530303	0.757576
20	0.539683	0.650794	0.603175	0.809524
25	0.534483	0.706897	0.655172	0.793103
30	0.679245	0.735849	0.735849	0.792453
40	0.813953	0.837209	0.813953	0.883721
50	0.787879	0.818182	0.878788	0.909091

Table 1 shows that the coverage probability of Type A prediction interval is about 0.4 which is fairly low when the size n_w of the window is small, and this probability increases to about 0.8 as n_w increases to 50. It appears that further increase in n_w may not improve the coverage probability very much. The coverage probability which falls short of the target value 0.95 may be attributed to the situations when the KLCI next month may not be concordant with the historical distribution due possibly to sudden changes in the economic conditions.

Table 1 also reveals that Type B prediction intervals tend to have larger coverage probabilities than Type A prediction intervals. Furthermore, when n_w is small, the union of Type B prediction intervals has a coverage probability which is 1.5 to 2.0 times of the coverage probabilities of Type A prediction interval.

We observe from Table 2 that among the four prediction intervals, the fourth interval which is based on the union operation has the longest average length. This observation is consistent with the finding that the fourth interval has the largest coverage probability. Table 3 shows that among the four prediction intervals, the interval based on the union operation tends to have the smallest MAPE.

Table 2
Average length of prediction interval ($l = 1, k = 7, n_r = 6, \alpha = 0.05$)

n_w	Type A	Type B (Linear)	Type B (Quadratic)	Union
13	64.71819	81.34884	81.11899	125.72483
15	70.75342	91.85950	88.93333	130.89046
17	80.20313	99.27582	95.95662	137.14357
20	91.92571	108.42993	106.34482	145.10693
25	100.69606	112.54155	113.33079	142.61797
30	107.40846	120.31234	122.07450	144.63468
40	123.55432	129.32158	129.46505	143.97366
50	132.58058	137.31274	137.93529	148.67818

Table 3
Mean absolute percentage error of prediction (MAPE) ($l = 1, k = 7, n_r = 6, \alpha = 0.05$)

n_w	Type A	Type B (Linear)	Type B (Quadratic)	Union
13	3.89217	3.80479	4.18914	3.52733
15	4.92098	3.56945	4.00412	3.41588
17	4.09067	3.49706	3.90302	3.39779
20	4.23346	3.30180	3.68185	3.10027
25	4.63374	3.10077	3.60971	3.11114
30	3.69365	2.96345	2.97683	2.75132
40	2.58016	2.38707	2.32127	2.21094
50	2.31802	2.15529	2.02773	2.06152

CONCLUSION

The event which may occur after the short time series is difficult to predict. The union of Type B prediction intervals which encompass more likely events, is found to have better performance in terms of coverage probability. The better performance of the related point estimate for the future value, as measured by MAPE, may be attributed to the averaging process which tends to produce a more satisfactory estimate.

REFERENCES

- Alba, E. D., & Mendoza, M. (2001). Forecasting an Accumulated Series Based on Partial Accumulation: A Bayesian Method for Short Series with Seasonal Patterns. *Journal of Business and Economic Statistics*, 19(1), 95-102.
- Alba, E. D., & Mendoza, M. (2006). Forecasting an Accumulated Series Based on Partial Accumulation: A New Bayesian Method for Short Series with Seasonal Patterns. *International Journal of Forecasting*, 22(4), 781-798.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Chang, S. H., & Fyffe, D. E. (1971). Estimation of Forecast Errors for Seasonal-Style-Goods Sales. *Management Science*, 18(2), 89-96.
- Chen, R., & Fomby, T. B. (1999). Forecasting with Stable Seasonal Pattern Models with an Application to Hawaiian Tourism Data. *Journal of Business and Economic Statistics*, 17(4), 497-504.
- Green, M., & Harrison, P. J. (1973). Fashion Forecasting for a Mail Order Company Using a Bayesian Approach. *Operational Research Quarterly*, 24(2), 193-205.
- Guerrero, V. M., & Elizondo, J. A. (1997). Forecasting a Cumulative Variable Using Its Partially Accumulated Data. *Management Science*, 43(6), 879-889.
- Hertz, D. B., & Schaffir, K. H. (1960). A Forecasting Method for Management of Seasonal Style-Goods Inventories. *Operations Research*, 8(1), 45-52.
- Murray, Jr. G. R., & Silver, E. A. (1966). A Bayesian Analysis of the Style Goods Inventory Problem. *Management Science*, 12(11), 785-797.
- Oliver, R. M. (1987). Bayesian forecasting with stable seasonal patterns. *Journal of Business and Economic Statistics*, 5(1), 77-85.
- Pooi, A. H. (2012). A Model for Time Series Analysis. *Applied Mathematical Sciences*, 6(115), 5735-5748.
- Yeo, I. K., & Johnson, R.A. (2000). A New Family of Power Transformations to Improve Normality or Symmetry. *Biometrika*, 87(4), 954-959.



A Goal Programming Model for Portfolio Optimisation Problem in Fuzzy Environment

Mokhtar, M.^{1*}, Shuib, A.² and Mohamad, D.²

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 27600 MARA, Raub, Pahang, Malaysia

²Faculty of Computer and Mathematical Sciences Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia

ABSTRACT

Portfolio optimisation is one of the most crucial issues in investment decision-making and has received considerable attention from researchers and practitioners. Traditionally, the portfolio optimisation models are formulated based on the assumption that investors have complete information on the distribution of random returns. However, in real life case, this is not possible since decisions have to be made under uncertainty. This paper deals with a fuzzy portfolio optimisation problem in which returns and turnover rates of securities are represented by fuzzy variables. A goal programming model is proposed to optimise three objectives: maximisation of portfolio return, maximisation of liquidity and minimisation of the portfolio risk. The cardinality constraints, floor and ceiling constraints are also taken into consideration. Finally, a numerical experiment using real data is conducted to demonstrate the applicability of the model.

Keywords: Portfolio optimisation, goal programming, fuzzy portfolio, multi-objective

INTRODUCTION

Over the past few decades, portfolio optimisation problem has become one of the most interesting topics in the field of financial management and investment. Its basic formulation is based on selecting a set of securities among a number of available ones which can best meet the investor's goal. The first portfolio selection model was introduced by Markowitz in 1952. The

model, also known as mean-variance model, has become a basis for the development of modern portfolio theory. In Markowitz's approach, the problem is formulated based upon two criteria: the portfolio return and the portfolio risk. The portfolio return is described by the mean return of the securities while the portfolio risk is quantified by the variance

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail addresses:

mazura_mokhtar@pahang.uitm.edu.my (Mokhtar, M.),

adibah@tmsk.uitm.edu.my (Shuib, A.),

daud@tmsk.uitm.edu.my (Mohamad, D.)

*Corresponding Author

of returns between the securities. The mean-variance model assumes that a rational investor wishes to maximise the portfolio expected return for a given level of risk, or alternatively, the investor wants the lowest portfolio risk for a given level of expected return.

Although the mean-variance model has been widely accepted for its theoretical reputations, it has not been applied extensively to construct a large-scale portfolio (Konno & Yamazaki, 1991). This is mainly due to the difficulty of handling a large-scale quadratic programming problem with a dense variance-covariance matrix. In order to alleviate this difficulty, several attempts have been made to transform the quadratic problem into a linear one. Konno and Yamazaki (1991) proposed mean absolute deviation as a new measure of risk to replace the variance in Markowitz's model. The authors showed that the mean absolute deviation model is equivalent to the mean-variance model when returns are assumed to have a normal distribution. Motivated by Konno and Yamazaki's work, Speranza (1993) proposed mean semi-absolute deviation to evaluate risk in portfolio selection model. This risk measure describes the preferences of investors in a more realistic way since it only considers return below the mean. From a computational point of view, Speranza (1993) showed that the semi-absolute deviation reduced the number of required constraints by half in comparison with the mean absolute deviation.

In the above-cited works, portfolio return and risk are considered as the only main factors that impact an investor's decision. However, many have argued that some of the relevant information for selecting a portfolio can never be completely captured in terms of these two criteria. There are other considerations that might be important to investors. As a result, numerous portfolio optimisation models that consider criteria other than risk and return have been developed in recent years. Ehrgott et al. (2004), for example, extended the mean-variance model by formulating a hierarchy of objectives, which breaks down risk and return into five sub-objectives and employed a multi-criteria decision-making method to solve the problem. Gupta et al. (2008) proposed a portfolio optimisation model based on semi-absolute deviation function. Their model considers multiple objectives which are short term return, long term return, annual dividend, risk, and liquidity. Li and Xu (2013) presented a multi-objective portfolio selection model with fuzzy random returns for investors. Their model optimizes three objectives, namely, return, risk and liquidity.

One of the most popular and promising techniques to handle portfolio optimisation problem with multiple and conflicting objectives is goal programming. Developed by Charnes and Cooper in 1961, it focuses on minimising deviations between the realised goal and the desired target. The minimisation process can be achieved using different approaches, each one leads to several variants of goal programming. Unlike linear programming that seeks for an optimal solution, goal programming attempts to look for a satisfactory solution that comes as close as possible to the desired goals. According to Arenas-Parra et al. (2010), the main advantage of goal programming approach is it provides decision makers with enough flexibility to include numerous variations of constraints and goals into a model.

Goal programming approach which is a branch of multi-objective optimisation has been extensively and successfully applied to formulate portfolio selection problems (see Pendaraki et al., 2005; Sharma & Sharma, 2006; Stoyan & Kwon, 2011; Tamiz et al., 2013). Most of the models however, assume that the future returns of security are dependent on random variables.

In practical investment, many other uncertain factors affect the stock markets such as economy, policies, laws and regulations (Liu & Zhang, 2013). Under such situations, fuzzy set theory initiated by Zadeh (1965) has become a useful tool in managing the vagueness and ambiguity of security returns. In recent years, much research has been done on portfolio optimisation problems in fuzzy environment. For instance, Carlsson et al. (2002) introduced a possibilistic approach for selecting portfolios with highest utility score based on the assumptions that the security returns are characterised by trapezoidal fuzzy numbers. Vercher et al. (2007) presented portfolio optimisation models under downside risk approach using interval-valued probabilistic and possibilistic means. Zhang et al. (2007) proposed two portfolio selection models based on lower and upper possibilistic means and possibilistic variances of fuzzy numbers. In addition, Huang (2011) proposed two credibility-based minimax mean-variance models for fuzzy portfolio selection problem in the situation that each security return belongs to a certain class of fuzzy variables.

This paper deals with a portfolio optimisation problem in fuzzy environment. We propose a goal programming model by considering three objectives which are minimisation of portfolio risk, maximisation of portfolio return and maximisation of liquidity. The returns and turnover rates of securities are characterised by fuzzy variables. In addition, the model includes practical constraints such as cardinality constraints, floor and ceiling constraints.

The remaining part of this paper is organised as follows: Section 2 introduces some basic concepts of fuzzy numbers while Section 3 presents the formulation of a fuzzy goal programming model for portfolio optimisation problem with practical constraints. Section 4 illustrates a numerical example of the proposed model along with the corresponding results. The final section (Section 5), ends with some concluding remarks.

PRELIMINARIES

In this section, we define some basic concepts about fuzzy numbers and the expected value of a fuzzy number which will be used in the following sections. A fuzzy number A is a fuzzy set of the real line \mathfrak{R} , characterised by means of a membership function $\mu_A(x)$ which is upper semi-continuous and satisfies the condition $\sup_{x \in \mathfrak{R}} \mu_A(x) = 1$, and whose γ -cuts, for $0 \leq \gamma \leq 1$: $[A]^\gamma = \{x \in \mathfrak{R} : \mu_A(x) \geq \gamma\}$, are convex sets.

A fuzzy number A are called trapezoidal with tolerance interval $[a; b]$, left width α and right width β if its membership function has the following form:

$$\mu_A(x) = \begin{cases} 1 - \frac{a-x}{\alpha} & \text{if } a - \alpha \leq x \leq a \\ 1 & \text{if } a \leq x \leq b \\ 1 - \frac{x-b}{\beta} & \text{if } a \leq x \leq b + \beta \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and it can be written as $A = (a, b, \alpha, \beta)$. Then, the γ -level sets of A can easily be calculated as

$$[A]^\gamma = [a - (1 - \gamma)\alpha, b + (1 - \gamma)\beta], \quad \forall \gamma \in [0,1],$$

Let A be a fuzzy number with $[A]^\gamma = [a_1(\gamma), a_2(\gamma)]$, $\gamma \in [0,1]$. Carlsson and Fullér (2001) defined the lower and upper possibilistic expected value of fuzzy number A as

$$\tilde{E}_*(A) = 2 \int_0^1 \gamma a_1(\gamma) d\gamma \tag{2}$$

$$\tilde{E}^*(A) = 2 \int_0^1 \gamma a_2(\gamma) d\gamma \tag{3}$$

The interval-valued and crisp possibilistic expected values of fuzzy number A are defined as (Carlsson & Fullér, 2001)

$$\tilde{E}(A) = [\tilde{E}_*(A), \tilde{E}^*(A)] \tag{4}$$

$$E(A) = \frac{\tilde{E}_*(A) + \tilde{E}^*(A)}{2} \tag{5}$$

MODEL FORMULATION

Assume that there are N securities available for trading in stock market. The return rates of the N securities are denoted as trapezoidal fuzzy numbers. The reason is that it can represent quite well the empirical distribution of security returns which have asymmetric and fat tails.

Following Vercher et al. (2007), the tolerance interval $[a, b]$, left width α and right width β of the fuzzy return for every security j are computed based on the percentiles of data distribution of the security return. Using this approach, the imprecise estimation of the expected return of each security can be simply and clearly represented based on actual data. For ease of description, basic notations used in this study are defined in Table 1.

Table 1
Basic Notations

Index	
j	Set of securities, $j = 1, \dots, N$
Decision variables	
x_j	Proportion invested in security j
q_j	Binary variable which will be 1 if any of security j is held and 0 otherwise and 0 otherwise
Parameters	
r_j	Return rate of security j
L_j	Turnover rate of security j
δ_j	Upper limits of the budget that can be invested in security j
ν_j	Lower limits of the budget that can be invested in security j
K	Desired number of securities to be included in the portfolio

Model's Objective

In the proposed multi-objective portfolio optimisation problem, the following objectives were considered.

Portfolio return. The portfolio return is the most practical objective which is usually used in portfolio optimisation models. In this paper, the return rates of the j^{th} asset are characterised by the trapezoidal fuzzy numbers $r_j = (a_j, b_j, \alpha_j, \beta_j)$ whose γ -level cuts are $[r_j]^\gamma = [a - (1 - \gamma)\alpha, b + (1 - \gamma)\beta]$ for $\forall \gamma \in [0,1]$. Therefore, the total return of a portfolio $x = (x_1, x_2, \dots, x_n)$ is the following trapezoidal fuzzy number

$$\begin{aligned}
 P &= \sum_{j=1}^N r_j x_j \\
 &= \left(\sum_{j=1}^N a_j x_j, \sum_{j=1}^N b_j x_j, \sum_{j=1}^N \alpha_j x_j, \sum_{j=1}^N \beta_j x_j \right) \\
 &= (P_l(x), P_u(x), C(x), D(x))
 \end{aligned}$$

The γ -level cuts of P are computed as

$$[P]^\gamma = [P_l(x) - (1 - \gamma)C(x), P_u(x) + (1 - \gamma)D(x)]$$

Based on the equation (2), (3), (4) and (5), we can obtain the lower and upper possibilistic means, the interval-valued and crisp possibilistic mean values of the portfolio return as follows:

$$\begin{aligned} \tilde{E}_*(P) &= 2 \int_0^1 \gamma(P_l(x) - (1 - \gamma)C(x))d\gamma \\ &= P_l(x) - \frac{1}{3} C(x) \end{aligned}$$

$$\begin{aligned} \tilde{E}^*(A) &= 2 \int_0^1 \gamma(P_u(x) + (1 - \gamma)D(x))d\gamma \\ &= P_u(x) + \frac{1}{3} D(x) \end{aligned}$$

$$\tilde{E}(P) = \left[P_l(x) - \frac{1}{3} C(x), P_u(x) + \frac{1}{3} D(x) \right]$$

$$E(P) = \frac{1}{2} (P_l(x) + P_u(x)) + \frac{1}{6} (D(x) - C(x))$$

Thus, the possibilistic mean value of the return on portfolio $x = (x_1, x_2, \dots, x_n)$ is given by

$$E(P) = \sum_{j=1}^N \left(\frac{1}{2} (a_j + b_j) + \frac{1}{6} (\beta_j - \alpha_j) \right) x_j \tag{6}$$

Portfolio risk. The portfolio risk is measured by the possibilistic semi-absolute deviation which is defined by Vercher et al. (2007) as

$$\tilde{w}(P) = E(\max\{0, \tilde{E}(P) - P\}) \tag{7}$$

Proposition 1: (Carlsson & Fullér, 2001) Let $r_j = (a_j, b_j, \alpha_j, \beta_j)$ be the trapezoidal return on the j^{th} asset, $j = 1, \dots, N$, and let $P = (P_l(x), P_u(x), C(x), D(x))$ be the total return of the portfolio $x = (x_1, x_2, \dots, x_n)$, then

$$\text{a) } \max\{0, \tilde{E}(P) - P\} = \left(0, P_u(x) - P_l(x) + \frac{D(x)}{3}, 0, C(x) \right)$$

$$\text{b) } \tilde{W}(P) = E(\max\{0, \tilde{E}(P) - P\}) = \left[0, P_u(x) - P_l(x) + \frac{C(x) + D(x)}{3} \right]$$

Based on the Proposition 1, the interval-valued possibilistic semi-absolute deviation can be expressed as follows:

$$\tilde{w}(P) = \left[0, P_u(x) - P_l(x) + \frac{C(x) + D(x)}{3} \right]$$

Thus, the crisp possibilistic semi-absolute deviation of the return associated with the portfolio $x = (x_1, x_2, \dots, x_n)$ is given by

$$\begin{aligned} w(P) &= \frac{P_u(x) - P_l(x)}{2} + \frac{C(x) + D(x)}{6} \\ &= \sum_{j=1}^N \left(\frac{1}{2} (b_j - a_j) + \frac{1}{6} (\alpha_j + \beta_j) \right) x_j \end{aligned} \tag{8}$$

Liquidity. In practical investment, liquidity is also one of the main concerns for investors. Liquidity can be defined as the ability to easily sell a security without affecting its price in the market and without incurring a significant loss. It can be computed using turnover rate which is the proportion between the average stock traded at the market in the most recent month and the outstanding shares of that stock for that month. In this study, the turnover rates of security j are denoted by trapezoidal fuzzy numbers $l_j = (la_j, lb_j, l\alpha_j, l\beta_j)$ since they cannot be accurately predicted. The γ -level cuts of l_j are $[l_j]^\gamma = [la_j - (1 - \gamma)l\alpha_j, lb_j + (1 - \gamma)l\beta_j]$ for $\forall \gamma \in [0,1]$. Thus, the liquidity of a portfolio $x = (x_1, x_2, \dots, x_n)$ is the following trapezoidal fuzzy number

$$\begin{aligned} L &= \sum_{j=1}^N l_j x_j \\ &= \left(\sum_{j=1}^N la_j x_j, \sum_{j=1}^N lb_j x_j, \sum_{j=1}^N l\alpha_j x_j, \sum_{j=1}^N l\beta_j x_j \right) \\ &= (Q_l(x), Q_u(x), F(x), G(x)) \end{aligned}$$

Based on the equation (2), (3), (4) and (5), the interval-valued and crisp possibilistic mean values of the portfolio liquidity can be expressed as

$$\begin{aligned} \tilde{E}(L) &= \left[Q_l(x) - \frac{1}{3} F(x), Q_u(x) + \frac{1}{3} G(x) \right] \\ E(L) &= \frac{1}{2} (Q_l(x) + Q_u(x)) + \frac{1}{6} (G(x) - F(x)) \end{aligned}$$

Therefore, the crisp possibilistic mean value of the turnover rate of the portfolio can be expressed as follows:

$$E(L) = \sum_{j=1}^N \left(\frac{1}{2} (h_j + b_j) + \frac{1}{6} (l\beta_j - l\alpha_j) \right) x_j \quad (9)$$

Constraints

There are two types of constraints: theoretical and practical. The theoretical constraint is necessary in order to define the feasibility of a solution.

Budget constraint. Budget constraint is imposed in order to normalise the solution. It ensures that all the available capital are invested. This constraint can be written as follows:

$$\sum_{j=1}^N x_j = 1$$

Floor and ceiling constraints. Floor constraints are used in practice to avoid the cost of administrating very small portions of securities which will have a negligible influence on the portfolio's performance. Ceiling constraints are imposed to limit the excessive concentration of the portfolio to a specific security. By introducing a binary variable q_j , (equal to 1 if security j is in the portfolio and 0 otherwise) the constraint can be expressed as follows:

$$\nu_j q_j \leq x_j \leq \delta_j q_j \quad j = 1, 2, \dots, N$$

Cardinality constraints. Cardinality constraints limit the total number of securities held in a portfolio. This constraint is imposed to simplify the management of the portfolio and to reduce transaction costs. This constraint is formulated as follows:

$$\sum_{j=1}^N q_j = K$$

Goal Programming

Basically, there are three variants in goal programming which are lexicographic, weighted and MinMax goal programming. Lexicographic goal programming approach assigns pre-emptive priority to different goals in order to minimise the sum of the unwanted deviation variables. The weighted goal programming approach assigns weights to goal deviations based on their relative importance and seeks to minimise the total weighted deviations of the goals. Finally, the MinMax goal programming attempts to minimise the largest unwanted deviation from the desired goals.

In this paper, we employed the lexicographic goal programming approach which requires ranking the goals in order of importance. A goal placed at the higher priority level is infinitely more important than a goal placed at the lower priority level. The problem is then solved by meeting the highest priority goal first as closely as possible before proceeding to the next priority goal. The procedure continues until the goal placed at the lowest priority level is solved. Using this approach, the solution achieved by a higher priority goal is never degraded by a lower priority goal.

The framework of pre-emptive goal programming model can be formulated as follows (Jones & Tamiz, 2010):

$$\text{Lex min } a = [h_1(d_i^-, d_i^+), \dots, h_p(d_i^-, d_i^+)]$$

subject to,

$$f_i(x) + d_i^- - d_i^+ = g_i \quad i = 1, \dots, m$$

$$x \in F$$

$$x \geq 0$$

$$d_i^-, d_i^+ \geq 0 \quad i = 1, \dots, m$$

where d_i^- and d_i^+ are the negative and positive deviational variables attached to the goal i , ($i = 1, \dots, m$), $f_i(x)$ is the mathematical expression of the i^{th} goal, g_i is the target value of goal i , x is the vector of the decision variables, F is a set of hard constraints that may exist in the model, h_s is the index set of goals placed in the s^{th} priority level and a is the lexicographic optimisation process.

Formulation of the Goal Programming Model

The first goal is to obtain the maximum possible return on the investment. This goal is formulated to minimise negative deviation. From equation (6), this goal can be expressed as follows:

$$\sum_{j=1}^N \left(\frac{1}{2} (a_j + b_j) + \frac{1}{6} (\beta_j - \alpha_j) \right) x_j + d_1^- - d_1^+ = g_1$$

where g_1 is the desired portfolio return. The second goal is to reduce the portfolio risk to a certain level. To achieve this goal, the positive deviation is minimised. Thus, from equation (8),

$$\sum_{j=1}^N \left(\frac{1}{2} (b_j - a_j) + \frac{1}{6} (\alpha_j + \beta_j) \right) x_j + d_2^- - d_2^+ = g_2$$

where g_2 is an acceptable level of risk. The third goal is to maximise the portfolio liquidity. This goal is formulated by minimising the negative deviation. From equation (9), this goal is given by

$$\sum_{j=1}^N \left(\frac{1}{2} (k_j + b_j) + \frac{1}{6} (l\beta_j - l\alpha_j) \right) x_j + d_3^- - d_3^+ = g_3$$

where g_3 is an acceptable level of liquidity. Based on the above discussion, the goal programming model for portfolio optimisation problem is formulated as follows:

Lex min $a = (d_1^-, d_2^+, d_3^-)$
 subject to,

$$\begin{aligned} \sum_{j=1}^N \left(\frac{1}{2} (a_j + b_j) + \frac{1}{6} (\beta_j - \alpha_j) \right) x_j + d_1^- - d_1^+ &= g_1 \\ \sum_{j=1}^N \left(\frac{1}{2} (b_j - a_j) + \frac{1}{6} (\alpha_j + \beta_j) \right) x_j + d_2^- - d_2^+ &= g_2 \\ \sum_{j=1}^N \left(\frac{1}{2} (k_j + b_j) + \frac{1}{6} (l\beta_j - l\alpha_j) \right) x_j + d_3^- - d_3^+ &= g_3 \\ \sum_{j=1}^N x_j &= 1 \\ v_j q_j \leq x_j \leq \delta_j q_j & \quad j = 1, 2, \dots, N \\ \sum_{j=1}^N q_j &= K \\ x_j \geq 0 & \quad j = 1, 2, \dots, N \\ d_i^-, d_i^+ \geq 0 & \quad j = 1, 2, \dots, N \\ q_j \in \{0,1\} & \quad j = 1, 2, \dots, N \end{aligned}$$

NUMERICAL EXAMPLE

Data

In this study, the model was tested on 100 Shariah-compliant stocks listed on the main board of Bursa Malaysia. The distributions of the selected securities in the corresponding sectors are shown in Table 2.

Table 2
Number of stocks in the sample per sector

Sector	Number of Securities	Decision Variables
Consumer	14	$x_1 - x_{14}$
Industrial	33	$x_{15} - x_{47}$
Technology	4	$x_{48} - x_{51}$
Plantation	6	$x_{52} - x_{57}$
Construction	5	$x_{58} - x_{62}$
Properties	11	$x_{63} - x_{73}$
Trading and Services	23	$x_{74} - x_{96}$
Infrastructure	3	$x_{97} - x_{99}$
Finance	1	x_{100}
Total	100	

The historical data consist of closing prices and turnover rates of all securities starting from January 2008 until December 2012. All data are collected from Datastream software. Data of monthly closing prices were transformed to monthly rate of returns using the formula below:

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$$

where p_t and p_{t-1} are the stock prices at time t and $t - 1$ respectively.

In this study, since the future return rates and turnover rates of the securities were assumed to be trapezoidal fuzzy numbers, the tolerance interval, left width and right width of the fuzzy numbers were approximated using sample percentiles method introduced by Vercher et al. (2007). First, we calculate the 5th, 40th, 60th and 95th percentiles of the samples using historical data. Then, we set the interval $[P_{40}, P_{60}]$ as the core and the quantities $P_{40} - P_5$ and $P_{95} - P_{60}$ as the left and right spreads respectively, where P_w is the w th percentile of the sample. Thus, the possibility distribution of security j is obtained, that is, $a_j = P_{40}$, $b_j = P_{60}$, $\alpha_j = P_{40} - P_5$ and $\beta_j = P_{95} - P_{60}$.

In this example, the minimum investment in each stock is 3% and the maximum investment must be of 20%. We also assume that the acceptance level of b_1 , b_2 and b_3 are 0.03, 0.05 and 0.10 respectively. In order to illustrate the effect of number of securities on the portfolio selection, we vary the value of K from 10 to 15. The problems were solved using MATLAB R2014a and the results are presented in Table 3.

Table 3
Return, Risk and Liquidity of Portfolio

K	Objectives			Selected stocks
	Return	Risk	Liquidity	
10	0.03	0.0495	0.10	$x_4 = 0.03, x_{10} = 0.03, x_{34} = 0.03,$
				$x_{40} = 0.0455, x_{52} = 0.03, x_{60} = 0.2,$
				$x_{61} = 0.0553, x_{69} = 0.2, x_{75} = 0.1792,$
11	0.03	0.05	0.10	$x_{78} = 0.2$
				$x_4 = 0.03, x_{10} = 0.03, x_{16} = 0.03,$
				$x_{32} = 0.0320, x_{52} = 0.03, x_{60} = 0.2,$
12	0.03	0.0492	0.10	$x_{61} = 0.03, x_{62} = 0.0784, x_{69} = 0.1715,$
				$x_{75} = 0.2, x_{78} = 0.1681$
				$x_1 = 0.03, x_{10} = 0.03, x_{14} = 0.0355,$
13	0.03	0.05	0.10	$x_{21} = 0.03, x_{34} = 0.03, x_{40} = 0.0354,$
				$x_{47} = 0.03, x_{60} = 0.2, x_{61} = 0.1492,$
				$x_{69} = 0.03, x_{75} = 0.2, x_{78} = 0.2$
14	0.03	0.0494	0.10	$x_1 = 0.03, x_{14} = 0.03, x_{16} = 0.0645,$
				$x_{22} = 0.03, x_{34} = 0.1829, x_{40} = 0.0560,$
				$x_{60} = 0.2, x_{61} = 0.03, x_{62} = 0.03,$
				$x_{69} = 0.0865, x_{72} = 0.03, x_{75} = 0.2,$
				$x_{82} = 0.03$
				$x_1 = 0.03, x_{14} = 0.03, x_{16} = 0.0748,$
				$x_{21} = 0.03, x_{34} = 0.2, x_{40} = 0.0336,$
				$x_{60} = 0.2, x_{61} = 0.03, x_{62} = 0.03,$
				$x_{69} = 0.0516, x_{72} = 0.03, x_{75} = 0.2,$
				$x_{78} = 0.03, x_{82} = 0.03$

Table 3 (continue)

K	Objectives			Selected stocks
	Return	Risk	Liquidity	
				$x_{10} = 0.03, x_{15} = 0.03, x_{32} = 0.0748,$ $x_{34} = 0.03, x_{40} = 0.0306, x_{47} = 0.03,$
15	0.03	0.05	0.1012	$x_{60} = 0.0886, x_{61} = 0.03, x_{69} = 0.2,$ $x_{72} = 0.03, x_{75} = 0.2, x_{77} = 0.03,$ $x_{78} = 0.1808, x_{80} = 0.03, x_{85} = 0.03$

The first column in Table 3 shows the desired number of securities. Columns 2 to 4 contain the return, risk and liquidity of the optimal solutions while the last column shows the selected securities of the optimal portfolio. From this table, it can be seen that the return goal for all portfolios are completely satisfied which indicates that the target value set for this goal is realistic and achievable to the degree that there is no room for possible further improvement. The second goal of minimising the portfolio risk is fully achieved when the numbers of securities are 11, 13 and 15. However, the risk for portfolio with 10, 12 and 14 securities has been reduced by 0.0005, 0.0008 and 0.0006. Finally, if we vary the number of securities in a portfolio from 10 to 14, the liquidity goal is completely satisfied. For portfolio with 15 securities, the liquidity goal is shown to be about 0.0012, more than the target value of 0.10.

CONCLUSION

This paper is concerned with multi-objective portfolio optimisation problem in a fuzzy environment. Based on semi-absolute deviation as the risk measure, a goal programming model is proposed in which three objectives are optimised in a lexicographic order. The three objectives are: minimisation of portfolio risk, maximisation of portfolio return and maximisation of liquidity. Additionally, the model also considers three practical constraints which are cardinality, floor and ceiling constraints. A numerical example using real data from Bursa Malaysia is also presented to illustrate the modeling concept. The results indicate that the proposed model could generate satisfying portfolio selection strategies to investors.

Finally, future research may be conducted to investigate the impact of inclusion of other real-life constraints such as minimum transaction lot and sector diversification constraint in the model.

REFERENCES

- Arenas-Parra, M., Bilbao-Terol, A., Pérez-Gladish, B., & Rodríguez-Uría, M. V. (2010). A new approach of romero's extended lexicographic goal programming: fuzzy extended lexicographic goal programming. *Soft Computing*, 14(11), 1217–1226.
- Carlsson C., & Fullér, R. (2001). On possibilistic mean value and variance of fuzzy numbers. *Fuzzy Sets and Systems*, 122(2), 315–326.
- Carlsson, C., Fullér, R., & Majlender, P. (2002). A possibilistic approach to selecting portfolios with highest utility score. *Fuzzy Sets and Systems*, 131(1), 13–21.
- Ehrgott, M., Klamroth, K., & Schwehm, C. (2004). An mcdm approach to portfolio optimization. *European Journal of Operational Research*, 155(3), 752–770.
- Gupta, P., Mehlawat, M. K., & Saxena, A. (2008). Asset portfolio optimization using fuzzy mathematical programming. *Information Sciences*, 178(6), 1734–1755.
- Huang, X. (2011). Minimax mean-variance models for fuzzy portfolio selection. *Soft Computing*, 15(2), 251–260.
- Jones D., & Tamiz, M. (2010). *Practical goal programming*, 141. New York, NY: Springer.
- Konno, H. & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management science*, 37(5), 519–531.
- Li, J., & Xu, J. (2013). Multi-objective portfolio selection model with fuzzy random returns and a compromise approach-based genetic algorithm. *Information Sciences*, 220, 507–521.
- Liu Y. J., & Zhang, W. G. (2013). Fuzzy portfolio optimization model under real constraints. *Insurance: Mathematics and Economics*, 53(3), 704–711.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91.
- Pendaraki, K., Zopounidis, C., & Doumpos, M. (2005). On the construction of mutual fund portfolios: A multicriteria methodology and an application to the Greek market of equity mutual funds. *European Journal of Operational Research*, 163(2), 462–481.
- Sharma H. P., & Sharma, D. K. (2006). A multi-objective decision making approach for mutual fund portfolio. *Journal of Business & Economics Research*, 4(6), 13-24.
- Speranza, M. G. (1993). Linear programming models for portfolio optimization. *Finance*, 14, 107–123.
- Stoyan S. J., & Kwon, R. H. (2011). A stochastic goal mixed integer programming approach for integrated stock and bond portfolio optimization. *Computers and Industrial Engineering*, 61(4), 1285–1295.
- Tamiz, M., Azmi, R. A., & Jones, D. F. (2013). On selecting portfolio of international mutual funds using goal programming with extended factors. *European Journal of Operational Research*, 226(3), 560–576.
- Vercher, E., Bermúdez, J. D., & Segura, J. V. (2007). Fuzzy portfolio optimization under downside risk measures. *Fuzzy sets and systems*, 158(7), 769–782.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3), 338–353.
- Zhang, W. G., Wang, Y. L., Chen, Z. P., & Nie, Z. K. (2007). Possibilistic mean–variance models and efficient frontiers for portfolio selection problem. *Information Sciences*, 177(13), 2787–2801.

Performance of Variance Targeting Estimator (VTE) under Misspecified Error Distribution Assumption

Abdul Rahim, M. A.*, Zahari, S. M. and Shariff, S. S. R.

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40500 UiTM, Shah Alam, Selangor, Malaysia

ABSTRACT

Parameter estimation in Generalized Autoregressive Conditional Heteroscedastic (GARCH) model has received much attention in the literature. Commonly used quasi maximum likelihood estimator (QMLE) may not be suitable if the model is misspecified. Alternatively, we can consider using variance targeting estimator (VTE) as it seems to be a better fit for misspecified initial parameters. This paper extends the application to see how both QMLE and VTE perform under error distribution misspecifications. Data are simulated under two error distribution conditions: one is to have a true normal error distribution and the other is to have a true student-t error distribution with degree of freedom equals to 3. The error distribution assumption that has been selected for this study are: normal distribution, student-t distribution, skewed normal distribution and skewed student-t. In addition, this study also includes the effect of initial parameter specification. The analyses are divided into two case designs. Case 1 $\omega_0 = 0.1, \alpha_0 = 0.05, \beta_0 = 0.85$ is when to represent the well specified initial parameters while Case 2 is when $\omega_0 = 1, \alpha_0 = 0, \beta_0 = 0$ to represent misspecified initial parameters. The results show that both QMLE and VTE estimator performances for misspecified initial parameters may not improve in well specified error distribution assumptions. Nevertheless, VTE shows a favourable performance compared to QMLE when the error distribution assumption is not the same as true underlying error distribution.

Keywords: GARCH, variance targeting, parameter estimation, error distribution

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail addresses:

muie.asmui@gmail.com (Abdul Rahim, M. A.),

mariam@tmsk.uitm.edu.my (Zahari, S. M.),

radiah@tmsk.uitm.edu.my (Shariff, S. S. R.)

*Corresponding Author

INTRODUCTION

Autoregressive conditional heteroscedastic (ARCH) model was first introduced by Engle (1982) and later extended as generalized ARCH (GARCH) by Bollerslev (1986). Its ability to mirror clustering characteristic in financial data makes this model popular, especially GARCH model for modelling and

forecasting financial market volatility. Clustering means high volatility tends to be followed by large changes, of either positive or negative signs or vice versa as mentioned by Mandelbrot (1963). This model is important because it has important roles in several financial applications such as option pricing, asset allocation and hedging.

Many studies have discussed the ability of the GARCH model in forecasting volatility. Poon and Granger (2003) have conducted an extensive review of the GARCH model but their findings were inconclusive which could have been due to the different sample period, sample frequency, forecast horizon, loss function and the proxy used for the ex post variance (Wilhelmsson, 2006). As the matter of fact, our study investigate further on these specifications issue as well as the error distribution assumption selection.

In discussing the GARCH model, two types of distribution have to be considered: the marginal, also known as the unconditional distribution, and the conditional distribution. As the GARCH model itself depends on time, it is known that the conditional error distribution has to follow a certain distribution even though the true underlying error distribution is naturally unknown. As a result, practitioners always assume the financial datasets under study to follow certain available distribution. The most preferable assumption is that the financial data follows a normal distribution because the commonly used estimator, which is a quasi-maximum likelihood estimator (QMLE), works well under such conditions. However, the financial data series is actually not normally distributed as shown by Mikosch and Stărică (2004), thus motivating our study to address this issue.

Among the studies that have discussed about the error distribution assumption specification are Hamilton and Susmel (1994), Franses and Ghijssels (1999), Lopez (2001), and Wilhelmsson (2006). Hamilton and Susmel (1994) applied Markov switching GARCH model and allowed the error term to be distributed according to a normal, Student's-t or generalized error distribution. By using weekly stock market data, it is found that the GARCH model with a Student's-t distribution performs best, followed by the generalized error distribution when the forecast performance of the one-week horizon is evaluated. On the other hand, Franses and Ghijssels (1999) drew different conclusion when using weekly European stock market data to evaluate the out-of-sample forecast. It is found that the GARCH model with t-distributed error is the worst model. Lopez (2001) checked the performance if GARCH(1,1) model fitted with the normal, Student's-t and generalized error distribution on four daily exchange rate series. It is shown that the performance of the models in-sample and out-of-sample are different and highlighting the importance of out of sample results as a model selection criteria. The results are mixed depending on the data series and loss function. Meanwhile, Wilhelmsson (2006) found that when the models were estimated, allowing for a skewed and excess kurtosis to be taken into account, it improved the log-likelihood. On the other hand, the out-of-sample results showed that allowing for skewness does not lead to any improvement over the normal distribution.

Mixed results, as the above example, might be due to several possible specifications that can be applied to the GARCH model which may lead to misspecification problem. One particular aspect of error misspecification impact is it will reduce the estimator performance for GARCH model parameters. Engle and Gonzalez-Rivera (1991) found that QMLE can suffered a 84% loss of efficiency due to misspecification of the error density. There is one particular estimator that could provide a good model even though there might be a misspecification which

is variance targeting estimator (VTE) by Engle and Mezrich (1996). The VTE is a two-step estimator based on reparameterization of the volatility equation where the intercept is replaced by the returns unconditional variance.

Francq et al. (2009) detailed the asymptotic properties of the VTE and list the method's advantages and disadvantages by studying the performance of VTE and QMLE toward modelling the simple univariate GARCH(1,1). Besides numerical simplicity, it is found that VTE can ensure the estimated unconditional variance of the GARCH model is equal to the sample variance. Hence, it is possible that in case of misspecification when the true underlying process is not GARCH, it can provide a superior result than QMLE (Francq et al., 2009). It is useful for prediction over long horizons. The main drawback of VTE is it needs a finite fourth order moment to retain its efficiency. Vaynman and Beare (2014) confirm it by investigating the VTE performance under infinite fourth moment with a heavy tailed distribution. It is found that in heavier tail condition, the finite fourth order moment is likely to be infinite and concluded that VTE should be used with caution in application when the distribution is heavy tailed.

The VTE may serve as a good alternative to QMLE especially because of its robustness towards model misspecification and seems to ease the numerical process. The main focus of this study is the application of VTE toward error distribution assumption misspecification, addressing gap in the literature. It has to be noted that this study is limited to in-sample model fitted to evaluate how well VTE performs in this scenario.

MODEL

This section presents GARCH (1,1) parameter estimation using two different estimators which are QMLE and VTE. GARCH(1,1), QMLE and VTE are explained as below;

Univariate GARCH

For GARCH (1,1), the model that has been used in this research is expressed as:

$$\text{GARCH (1,1)} = \begin{cases} \epsilon_t = \sqrt{h_t}\eta_t \\ h_t = \omega + \alpha_0\epsilon_{t-1}^2 + \beta_0h_{t-1} \end{cases} \quad (1)$$

where η_t is a sequence of iid with unit variance, $\omega > 0$, $\alpha_0 \geq 0$, $\beta_0 \geq 0$ and $\alpha_0 + \beta_0 < 1$.

QMLE Estimation

The asymptotic behaviour of QMLE is the reason why. Thus, it is important to use QMLE as a benchmark to compare the performance of proposed estimator technique. Using GARCH (1,1), the estimators which need to be estimated are $\theta_0 = (\omega, \alpha, \beta)'$. QMLE under assumption of derived as any measurable solution $\hat{\theta}_n$ of $\hat{\theta}_n = \arg \max_{\theta \in \Theta} L(\theta)$ of where

$$L_n(\theta) = L_n(\theta; \epsilon_1, \dots, \epsilon_n) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\tilde{\sigma}_t^2}\right) \quad (2)$$

$$\tilde{\sigma}_t^2 \text{ defined as; } \tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) = \omega + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \tilde{\sigma}_{t-1}^2$$

Thus, implementing the logarithm gives the maximising of the likelihood equal to maximising \tilde{I}_n with respect to θ ,

where,

$$\tilde{I}_n(\theta) = \frac{1}{2n} \sum_{t=1}^n \tilde{\ell}_t \text{ where, } \tilde{\ell}_t = \tilde{\ell}_t(\theta) = -\left(\frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \ln \tilde{\sigma}_t^2\right)$$

VTE Estimation

The VTE consists of two steps. First, the unconditional variance of the observed data is estimated by a moment estimator and next, the remaining parameters are estimated by the QMLE. The steps are explained below.

Consider GARCH (1,1) as (1) where $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$ is an unknown parameter (η_t) and a sequence of independent in variance of identically distributed (i.i.d) random variables such that $E\eta_t^2 = 1$ and $\omega_0 > 0, \alpha_0 > 0, \beta_0 \geq 0$. Under condition $\alpha_0 + \beta_0 < 1$, this model admits a second-order stationary solution, which the unconditional variance is given by,

$$\gamma_0 := \sigma^2(\omega_0, \alpha_0, \beta_0) = \frac{\omega_0}{1 - \alpha_0 - \beta_0} := \frac{\omega_0}{\kappa_0}$$

A reparameterisation of the model with $\vartheta_0 = (\gamma_0, \alpha_0, \kappa_0)'$ yields;

$$\begin{aligned} \epsilon_t &= \sqrt{h_t} \eta_t, \\ h_t &= h_{t-1} + \kappa_0(\gamma_0 - h_{t-1}) + \alpha_0(\epsilon_{t-1}^2 - h_{t-1}) \end{aligned}$$

and κ_0 is the speed of mean reversion.

Writing (6) as $h_t = \kappa_0 \gamma_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 h_{t-1}$ where $\kappa_0 + \alpha_0 + \beta_0 = 1$, we can interpret the volatility h_t at time t as weighted average of the long-run variance γ_0 , the square of the last return ϵ_{t-1}^2 and the previous volatility h_{t-1} . In this average, κ_0 is the weight of the long-run variance. This reparametrisation limits; $\kappa_0, \gamma_0 > 0, \alpha_0 \geq 0$ and $\kappa_0 + \alpha_0 \leq 1$.

Let $(\epsilon_1, \dots, \epsilon_n)$ be a realisation of length n of the unique nonanticipative second-order stationary solution (ϵ_t) to model (1). In this framework, VTE involves (i) reparametrising the model as in (6), and (ii) estimating γ_0 by the sample variance using moment estimator and then $\lambda_0 := (\alpha_0, \kappa_0)'$ by the QMLE. The QMLE of θ_0 is denoted by $\hat{\theta}_n^* := (\hat{\omega}_n^*, \hat{\alpha}_n^*, \hat{\beta}_n^*)'$. Two consistent estimator of λ_0 are the sample variance and the QML-based estimator given by $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^n \epsilon_t^2$ (7) and $\sigma^2(\hat{\theta}_n^*) = \frac{\hat{\omega}_n^*}{1 - \hat{\alpha}_n^* - \hat{\beta}_n^*}$.

Consider a parameter space $\Lambda \subset \{(\alpha, \kappa) | \alpha \geq 0, \kappa > 0, \alpha + \kappa \leq 1\}$. All the vectors are considered as column vectors written as row vectors. In particular, we write $\vartheta_0 = (\gamma_0, \lambda_0)'$ and at the point $\vartheta = (\gamma, \lambda) \in (0, \infty) \times \Lambda$, the QMLE of the sample given by

$$\tilde{L}_n(\vartheta) = \tilde{L}_n(\gamma, \lambda) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2(\vartheta)}} \exp\left\{-\frac{\epsilon_t^2}{2\tilde{\sigma}_t^2(\vartheta)}\right\} \tag{3}$$

where the $\tilde{\sigma}_t^2(\vartheta)$ are defined as recursively, for $t \geq 1$, by $\tilde{\sigma}_t^2(\vartheta) = \kappa\gamma + \alpha\epsilon_{t-1}^2 + (1 - \kappa - \alpha)\tilde{\sigma}_{t-1}^2(\vartheta)$ with the initial values ϵ_0 and $\tilde{\sigma}_t^2(\vartheta) := \sigma_0^2$. Since the parameter λ_0 is estimated by the sample variance $\hat{\sigma}_n^2$, the variance targeting version of the QMLE function is

$$L(\lambda) = L(\hat{\sigma}_n^2, \lambda) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_{t,n}^2}} \exp\left\{-\frac{\epsilon_t^2}{2\sigma_{t,n}^2}\right\} \tag{4}$$

where the $\sigma_{t,n}^2 := \sigma_{t,n}^2(\lambda) = \kappa_0\hat{\sigma}_n^2 + \alpha_0\epsilon_{t-1}^2 + (1 - \kappa - \alpha)\hat{\sigma}_{t,n}^2$ with $\sigma_{0,n}^2 = \sigma_0^2$. A VTE of λ_0 is defined as any measurable solution $\tilde{\lambda}_n$ of;

$$\lambda = \arg \max_{\lambda \in \Lambda} L_n(\lambda) = \arg \min_{\lambda \in \Lambda} \tilde{I}_n(\lambda)$$

and

$$\tilde{I}_n(\lambda) = \frac{1}{2n} \sum_{t=1}^n \ell_{t,n}, \ell_{t,n} := \ell_{t,n}(\lambda) = -\left(\frac{\epsilon_t^2}{\sigma_{t,n}^2} + \ln \sigma_{t,n}^2\right) \tag{5}$$

METHODOLOGY

A simulation study is considered to investigate the performance of the two estimators, QMLE and VTE.

First, random variable of ϵ_t is simulated under two different underlying error distributions, one with normal distribution that has a finite moment across all orders and the other under the Student's-t distribution with degrees of freedom equal to 3 to represent the infinite fourth order moment. Datasets are simulated into three different sample size, in which $n=500, 1000$ and 5000 for both true error distribution setup.

Next, using “rugarch” package in R-programming, different error distribution assumption is used to model GARCH (1,1). The distribution assumption is normal, Student-t, skewed normal and skewed Student's-t. Skewed distributions for both normal and Student's-t are considered to represent the assumption of a heavy tailed error.

Two case designs, Case 1 and Case 2, are established to differentiate initial parameter specifications. Case 1 is when $\omega_0 = 0.1, \alpha_0 = 0.05, \beta_0 = 0.85$ are used to represent the well specified initial parameters while Case 2 is when $\omega_0 = 1, \alpha_0 = 0, \beta_0 = 0$ are used as the representative of misspecified initial parameters.

RESULTS AND DISCUSSION

For Case 1 (well specified initial parameters), both QMLE and VTE under well specified error distribution (error distribution assumption is the same with the true underlying error distribution) outperform other distribution assumption setting. This result applies if the true distribution is normally distributed. At the same time as shown in Table 1 VTE perform better than QMLE if the error distribution assumption is misspecified. However, the result is different if the simulated data true underlying error is Student-t distributed. It seems that a well

specified error distribution does not help in improving the likelihood values for this scenario. Estimators under normal and skewed normal assumption perform better than Student-t and skewed Student-t based on Table 2. Comparing the results in Table 1 and 2, we can conclude that the performance of both QMLE and VTE is reduced when the true underlying distribution is heavy-tailed.

For Case 2, QMLE and VTE under misspecification error distribution assumption (error distribution assumption is different with the true underlying error distribution) perform better than when the estimators are under well specified condition for both true error distributions (refer Table 3 and 4.). The VTE only outperforms QMLE when $n=500$ and $n=5000$.

Based on the results, there are several important findings as well. One of it is, VTE needs less processing time than QMLE. Furthermore, for Case 1, the sample size must be greater than 1000 to produce the significant parameters for all. The significance value must be less than 0.05 in order to conclude the significance of the parameter.

Under infinite fourth order moment, as suspected, the VTE performance is reduced. The QMLE outperforms most of the VTE under all levels if we compare the likelihood and standard error produce for each parameter. But still, VTE produces more significant parameters than QMLE.

Table 1
Parameters estimation under true normal error distribution for Case 1

Sample Distribution Assumption	n=500					n=1000					n=5000					
	ω	α	β	Likelihood	Times	ω	α	β	Likelihood	Times	ω	α	β	Likelihood	Times	
Normal	QMLE	0.2882*	0.0327*	0.9391	-1282.75	0.11	0.6941	0.0788	0.8554	-2577.31	0.15	1.1443	0.0589	0.8242	-12761.09	0.70
	VTE	0.2843	0.0190	0.0407	0.2978	0.0211	0.0406	0.0406	0.3609	0.0121	0.3609	0.0121	0.0452	0.0452	0.0452	0.0452
Student-t	QMLE	0.2893	0.0323*	0.9389	-1282.76	0.08	0.6940	0.0784	0.8555	-2577.31	0.08	1.1437	0.0588	0.8242	-12761.09	0.23
	VTE	0.2914*	0.0687	0.1503	0.0195	0.0404	0.0404	0.0404	0.1118	0.0118	0.1118	0.0118	0.0452	0.0452	0.0452	0.0452
Skewed-Normal	QMLE	0.2941	0.0192	0.9393	-1282.66	0.30	0.6861	0.0785	0.8566	-2577.23	0.48	1.1189	0.0582	0.8277	-12760.41	1.84
	VTE	0.2941	0.0192	0.0415	0.3003	0.0213	0.0409	0.0409	0.3707	0.0124	0.3707	0.0124	0.0465	0.0465	0.0465	0.0465
Skewed-Student	QMLE	0.2929	0.0319*	0.9390	-1282.67	0.21	0.6865	0.0781	0.8565	-2577.23	0.31	1.0880	0.0568	0.8319	-12760.42	0.86
	VTE	0.0498	0.1105	0.0001	0.0001	0.0406	0.0406	0.0406	0.0120	0.0120	0.0120	0.0120	0.0456	0.0456	0.0456	0.0456
Skewed-Normal	QMLE	0.2950*	0.0328*	0.9382	-1282.64	0.14	0.6700	0.0779	0.8588	-2576.82	0.19	1.1527	0.0592	0.8231	-12760.99	0.76
	VTE	0.2923	0.0193	0.0418	0.2907	0.0209	0.0399	0.0399	0.3608	0.0121	0.3608	0.0121	0.0452	0.0452	0.0452	0.0452
Skewed-Student	QMLE	0.2958	0.0324*	0.9381	-1282.65	0.08	0.6699	0.0773	0.8589	-2576.82	0.11	1.1520	0.0590	0.8231	-12760.99	0.44
	VTE	0.0414	0.0919	0.0192	0.0192	0.0394	0.0394	0.0394	0.0118	0.0118	0.0118	0.0118	0.0452	0.0452	0.0452	0.0452
Skewed-Student	QMLE	0.2969*	0.0324*	0.9385	-1282.57	0.58	0.6615	0.0777	0.8599	-2576.77	0.68	1.1262	0.0584	0.8266	-12760.25	2.41
	VTE	0.2998	0.0195	0.0424	0.2935	0.0212	0.0403	0.0403	0.3687	0.0124	0.3687	0.0124	0.0463	0.0463	0.0463	0.0463
Skewed-Student	QMLE	0.2981	0.0321*	0.9382	-1282.58	0.24	0.6620	0.0770	0.8599	-2576.78	0.40	1.1297	0.0582	0.8262	-12760.26	1.61
	VTE	0.0367	0.0820	0.0194	0.0194	0.0398	0.0398	0.0398	0.0121	0.0121	0.0121	0.0121	0.0465	0.0465	0.0465	0.0465

*Not Significant (p-value>0.05)

Table 2
Parameters estimation under true student's-t ($df=3$) error distribution for Case 1

Sample Distribution Assumption	n=500					n=1000					n=5000					
	ω	α	β	Likelihood	Times	ω	α	β	Likelihood	Times	ω	α	β	Likelihood	Times	
Normal	QMLE	0.0033*	0.0024*	0.9965	-1305.17	0.14	1.3205*	0.0651	0.8010	-2558.57	0.20	0.6506	0.0349	0.8980	-12758.33	0.46
	VTE	0.0033	0.0026	0.0003	0.6887	0.10	1.3255	0.0658	0.8008	-2558.57	0.09	0.6505	0.0349	0.8981	-12758.33	0.21
Student-t	QMLE	0.0047*	0.0017*	0.9972	-1301.24	0.19	1.3090*	0.0659	0.8018	-2558.95	0.39	0.6527	0.0351	0.8977	-12758.02	1.69
	VTE	0.0289	0.0028	0.0002	0.6901	0.15	1.3112	0.0663	0.8018	-2558.95	0.35	0.6511	0.0351	0.8979	-12757.99	1.12
Skewed-Normal	QMLE	0.0034*	0.0022*	0.9967	-1305.10	0.15	1.3216*	0.0653	0.8007	-2558.56	0.27	0.6509	0.0350	0.8980	-12758.32	0.75
	VTE	0.0258	0.0025	0.0002	0.6907	0.13	1.3261	0.0660	0.8006	-2558.57	0.12	0.6508	0.0349	0.8980	-12758.32	0.28
Skewed-Student	QMLE	0.0047*	0.0016*	0.9973	-1301.23	0.26	1.3046*	0.0667	0.8019	-2559.25	0.49	0.6525	0.0352	0.8977	-12758.15	2.44
	VTE	0.0289	0.0028	0.0002	0.6954	0.21	1.3031	0.0668	0.8020	-2559.25	0.41	0.6522	0.0351	0.8977	-12758.15	2.06

*Not Significant (p-value>0.05)

Table 3
Parameters estimation under true normal error distribution for Case 2

Sample Distribution Assumption	n=500					n=1000					n=5000					
	ω	α	β	Likelihood	Times	ω	α	β	Likelihood	Times	ω	α	β	Likelihood	Times	
Normal	QMLE	0.0009	0.0000*	0.9989	-714.24	0.13	0.0036	0.0000*	0.9965	-1440.90	0.23	0.0234	0.0070	0.9690	-7111.04	0.50
	VTE	0.0001	0.0000	0.0001	-714.31	0.08	0.0003	0.0000	0.0003	-1440.46	0.08	0.0033	0.0029	0.0000	-7111.04	0.37
Student-t	QMLE	0.0009	0.0000*	0.9989	-714.42	0.20	0.0010	0.0000*	0.9989	-1441.13	0.34	0.0240	0.0070	0.9690	-7112.29	1.31
	VTE	0.0001	0.0000	0.0001	-714.48	0.17	0.0001	0.0000	0.0001	-1440.71	0.28	0.0033	0.0032	0.0009	-7112.37	0.98
Skewed-Normal	QMLE	0.0009	0.0000*	0.9989	-713.83	0.16	0.0594*	0.0104*	0.9324	-1440.27	0.34	0.0230	0.0070	0.9699	-7110.90	0.72
	VTE	0.0001	0.0000	0.0001	-713.83	0.11	0.0707	0.0130	0.0729	-1440.27	0.09	0.0029	0.0029	0.0000	-7110.91	0.31
Skewed-Student	QMLE	0.0009	0.0000*	0.9989	-714.23	0.31	0.0010	0.0000*	0.9989	-1441.19	0.58	0.0242	0.0071	0.9688	-7113.37	2.23
	VTE	0.0001	0.0001	0.0001	-714.28	0.28	0.0001	0.0001	0.0001	-1440.78	0.37	0.0034	0.0033	0.0009	-7113.37	1.29

*Not Significant (p-value>0.05)

Table 4
Parameters estimation under true student's-t ($df=3$) error distribution for Case 2

Sample Distribution Assumption	n=500					n=1000					n=5000					
	ω	α	β	Likelihood	Times	ω	α	β	Likelihood	Times	ω	α	β	Likelihood	Times	
Normal	QMLE	0.0086	0.0000*	0.9990	-1086.07	0.24	0.3944	0.0324	0.8109	-1872.88	0.32	0.1258*	0.0094	0.9449	-9633.68	0.64
	VTE	0.0019	0.0000	0.0007			0.0975	0.0124	0.0430			0.0749	0.0028	0.0310		
Student-t	QMLE	Na	Na	Na	Na	Na	0.0118	0.0000*	0.9957	-1726.82	0.26	0.0533	0.0014*	0.9824	-8915.99	1.31
	VTE	0.5473	0.0000*	0.8816*	-1092.39	0.10	0.3954	0.0326	0.8108	-1872.89	0.09	0.1267	0.0095	0.9447	-9633.70	0.23
Skewed-Normal	QMLE	Na	Na	Na	Na	Na	0.0142	0.0000	0.9957	-1726.82	0.26	0.0533	0.0014*	0.9824	-8915.99	1.31
	VTE	0.5866	0.0000*	0.8732	-870.87	0.14	0.4142	0.0148*	0.8213	-1726.01	0.14	0.0365	0.0010	0.9857	-8918.30	0.61
Skewed-Student	QMLE	0.0061	0.0000*	0.1551	-1069.67	0.22	0.0232	0.0020*	0.9889	-1878.72	0.21	0.1123	0.0095	0.9496	-9628.03	0.84
	VTE	0.0004	0.0000				0.0051	0.0017	0.0007			0.0219	0.0026	0.0088		
Normal	QMLE	0.0439	0.0000*	0.9811	-867.38	0.27	0.0129	0.0000*	0.9954	-1726.44	0.37	0.0529	0.0015*	0.9824	-8915.64	2.03
	VTE	0.8649	0.0000*	0.8130*	-870.86	0.18	0.4039	0.0142*	0.8258	-1725.63	0.20	0.0368	0.0010	0.9856	-8917.96	1.05
									0.0138					0.0003		

*Not Significant (p-value>0.05)

CONCLUSION

In conclusion, VTE shows promising performance when dealing with the misspecification model for both initial parameters and error distribution. But under misspecified initial parameters, a well specified error distribution assumption does not fix the estimators performance suggesting that these two factors are to be treated differently, but more evidences are needed to arrive at this conclusion. Future research should examine the performance of these two estimators in managing volatility forecasting in the presence of leverage effect. Asymmetric GARCH can also be used in future research. More research is needed on parameter estimation so that the most efficient model can be built and helping in reducing the risk faced in financial data series. In addition, out-of-sample forecast evaluation of real datasets might help in finding more conclusive evidence of VTE effectiveness.

REFERENCES

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4), 987-1007.
- Engle, R. F., & Gonzalez-Rivera, G. (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics*, 9(4), 345-359.
- Engle, R., & Mezrich, J. (1996). Garch for Groups: A round-up of recent developments in Garch techniques for estimating correlation. *Risk-London-Risk Magazine Limited*, 9, 36-40.
- Franco, C., Horvath, L., & Zakořan, J. M. (2011). Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics*, 9(4), 619-656.
- Franses, P. H., & Ghijssels, H. (1999). Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting*, 15(1), 1-9.
- Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1), 307-333.
- Lopez, J. A. (2001). Evaluating the predictive accuracy of volatility models. *Journal of Forecasting*, 20(2), 87-109.
- Mandelbrot, B. B. (1997). *The variation of certain speculative prices* (pp. 371-418). New York, NY: Springer.
- Poon, S. H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of economic literature*, 41(2), 478-539.
- Vaynman, I., & Beare, B. K. (2014). Stable limit theory for the variance targeting estimator. In Y. Chang, T. B. Fomby & J. Y. Park (Eds.), *Essays in Honor of Peter CB Phillips Advances in Econometrics* (pp. 639-672). United Kingdom, UK: Emerald Group Publishing Limited.
- Wilhelmsson, A. (2006). GARCH forecasting performance under different distribution assumptions. *Journal of Forecasting*, 25(8), 561-578.



Review of Context-Based Similarity Measure for Categorical Data

Nurul Adzlyana, M. S.*, Rosma, M. D. and Nurazzah, A. R.

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

ABSTRACT

Data mining processes such as clustering, classification, regression and outlier detection are developed based on similarity between two objects. Data mining processes of categorical data is found to be most challenging. Earlier similarity measures are context-free. In recent years, researchers have come up with context-sensitive similarity measure based on the relationships of objects. This paper provides an in-depth review of context-based similarity measures. Descriptions of algorithm for four context-based similarity measure, namely Association-based similarity measure, DILCA, CBDL and the hybrid context-based similarity measure, are described. Advantages and limitations of each context-based similarity measure are identified and explained. Context-based similarity measure is highly recommended for data-mining tasks for categorical data. The findings of this paper will help data miners in choosing appropriate similarity measures to achieve more accurate classification or clustering results.

Keywords: Categorical data, context-based, data mining, similarity measure

INTRODUCTION

Similarity measure is the measure of how much alike two data objects are. Similarity measure in data mining is usually described as a distance with dimensions representing features of the objects. A small distance means a high degree of similarity and vice versa. Similarity is very subjective and is highly dependent on the application domain (Yong, 2010).

Similarity between two objects plays an important role in data mining tasks such as clustering, classifying, regressing, or finding distance for outlier detection of various types of data (Desai et al., 2011) involving distance computations. The distance of similarity for integer-type data and ratio-scaled data are well defined and understood.

However, devising similarity or distance metrics for classification and clustering of categorical data is found to be more challenging (Alamuri et al., 2014). The

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail addresses:

nurul_adzlyana@yahoo.com (Nurul Adzlyana, M. S.),

rosma@tmsk.uitm.edu.my (Rosma, M. D.),

nurazzah@tmsk.uitm.edu.my (Nurazzah, A. R.)

*Corresponding Author

usual similarity measures for categorical data are binary methods, where each bit indicates the presence or absence of a possible attribute value. The similarity between two objects is determined by the similarity between two corresponding binary vectors (Khorshidpour et al., 2010). Nevertheless, the alteration of data objects into binary vectors is the main problem, as the binary vectors will calculate the similarity between two values to be either 0 or 1 and in the process may eliminate important information about the data.

Earlier similarity measures are context-free. But recently researchers have come up with context-sensitive similarity measures. Hence, similarity measures can be divided into two main categories based on the way they utilise the context of the given attributes. Thus, similarity measures can be context-free or context-sensitive. Context-free similarity measure is used when the distance between two objects in the data is taken as a function of the objects only and does not depend on the relationship between those objects to other data objects. On the other hand, context-sensitive similarity measure is used when the similarity between two data objects depends not only on the two objects alone, but also on the relationship between the objects and the other data objects (Alamuri et al., 2014).

In more recent research, hybrid similarity measures have been introduced which combine two important elements. The first element is the context selection process followed by distance computation (Alamuri et al., 2014). Context-selection is observing the meta-attributes connected with the current attributes of the objects called context attributes. In order to determine the context of each attribute, a data driven method is employed and it is application specific while the distance computation is for the pair of values of an attribute based on context selection. Alamuri has suggested a hybrid similarity measure that combines learning algorithm for context selection and distance computation based on the learned context.

This paper will review three commonly used context-based similarity measures. The advantages and limitations of the methods are described for comparison purposes.

CONTEXT-BASED SIMILARITY MEASURE

There are four techniques in context-based similarity measure for categorical data, namely the Association-Based Similarity Measure, DILCA, CBDL and the hybrid context-based similarity measure. These techniques are described below.

Association-Based Similarity Measure

An association-based similarity measure was proposed by (Le & Ho, 2005). A novel indirect method to measure the dissimilarity for categorical data was introduced. The dissimilarity between two values of an attribute is indirectly estimated by using relations between other related attributes.

The efficiency of the proposed method is investigated in terms of theoretical proofs and the experiments with real data showed that attributes are typically correlated. However, this method is found to be unsuitable for data sets with independent attributes (Le & Ho, 2005). The

Association-Based Similarity Measure comprises two steps which are finding the dissimilarity between two values of attribute followed by finding the dissimilarity between two data objects.

The algorithm of the method is as below:

Step 1: The dissimilarity between two values of attribute. The dissimilarity between two values v_i and v'_i of attribute A_i , denoted by $\emptyset_{A_i}(v_i, v'_i)$, is the sum of dissimilarities between conditional probability distributions of other attributes given that attribute A_i holds values v_i and v'_i :

$$\emptyset_{A_i}(v_i, v'_i) = \sum_{j, j \neq i} \psi \left(cpd(A_j|A_i = v_i), cpd(A_j|A_i = v'_i) \right) \quad (1)$$

where $\psi(\dots)$ is a dissimilarity function for two probability distributions.

The dissimilarity between two values v_i and v'_i of attributes is directly proportional to dissimilarities between the conditional probability distributions of other attributes. Thus, the smaller the dissimilarities between these conditional probability distributions, the smaller the dissimilarity between v_i and v'_i .

Le and Ho (2005) used the popular dissimilarity measure, which is the KullbackLeibler method, (Kullback & Leibler, 1951) that is given by:

$$KL(P, P') \sum_x \left(p(x) \lg \frac{p(x)}{p'(x)} + p'(x) \lg \frac{p'(x)}{p(x)} \right) \quad (2)$$

where \lg is a logarithm of base 2.

Step 2: The dissimilarity between two data objects. The dissimilarity between two data objects x and y , denoted by $\emptyset(x, y)$, is the sum of dissimilarities of their attribute value pairs:

$$KL(P, P') \sum_x \left(p(x) \lg \frac{p(x)}{p'(x)} + p'(x) \lg \frac{p'(x)}{p(x)} \right) \quad (3)$$

If the dissimilarities of attribute value pairs of x and y is smaller, then the dissimilarities between x and y is also smaller.

Distance Learning for Categorical Attributes (DILCA)

Ienco et al. (2012) proposed a context-based similarity measure called Distance Learning for Categorical Attributes (DILCA) to compute the distance between any pair of values of a specific categorical attribute. The method consists of two steps: context selection and distance computation. The context selection step is the process of selecting the relevant subset of the whole attributes set while the distance computation is the process of computing the distance between pair of values of the same attribute using the context defined in the context selection.

Step 1: Context selection. The aim in this step is to select a subset of relevant and not overlapped features. Ienco et al. (2012) proposed several approaches for measuring the correlation between two variables. One of them is the Symmetric Uncertainty. This context selection is a correlation based measure inspired by information theory. Symmetric Uncertainty is derived from entropy as it is a measure of the uncertainty of a random variable. The entropy of a random variable is defined as:

$$H(X) = - \sum_i P(x_i) \log_2(P(x_i)) \tag{4}$$

where $P(x_i)$ is the probability of the value x_i of the X . The entropy of X after having observed the values of another variable Y is defined as:

$$H(X|Y) = - \sum_j P(y_j) \sum_i P(x_i|y_i) \log_2(P(x_i|y_i)) \tag{5}$$

where $P(x_i|y_i)$ is the probability that $X = x_i$ after observing that $Y = y_i$. The information about X provided by Y is given by the information gain, which is defined as follows:

$$IG(X|Y) = H(X) - H(X|Y) \tag{6}$$

When $IG(X|Y) > IG(Z|Y)$ then the feature X is more correlated to Y than Z . Moreover, the information gain is symmetrical for two random variables X and Y . The symmetrical uncertainty is then defined as:

$$SU(X, Y) = 2 \cdot \frac{IG(X|Y)}{H(X) + H(Y)} \tag{7}$$

This measure varies between 0 and 1 where 1 indicates that knowledge of the value of either X or Y and completely predicts the value of the other variable while 0 indicates that X and Y are independent. The advantage of symmetrical uncertainty is that the Information Gain that it measures is not biased by the number of values of an attribute.

Step 2: Distance Computation. The goal of this step is to compute the distance between x_i and x_j where $x_i \in X, x_j \in X$, using this given formulation:

$$d(x_i, x_j) = \sqrt{\sum_{Y \in context(X)} \sum_{y_k \in Y} (P(x_i|y_k) - P(x_j|y_k))^2} \tag{8}$$

The conditional probability for both values x_i and x_j in each context attribute Y is given by the values $y_k \in Y$. Then, the Euclidean distance is applied.

Context-Based Distance Learning (CBDL)

In 2011, Khorshidpour et al. (2010) proposed a method to measure the dissimilarity of categorical data, CBDL. This method consists of two steps. In the first step, a relevant subset of the whole attributes is selected. Then, the dissimilarity between pair of the values of the same attribute is computed using the context defined in the first step. The two steps are described as below:

Step 1: Context Selection. Supervised feature selection is employed in this step. The goal is to select a subset of correlated features based on the given one. The outcome of feature selection is a subset of input variables by eliminating features with given class attribute. Feature selection process can improve the comprehensibility of the resulting classifier models. It often builds a model that generalises better to unseen points. The steps are defined as below:

Entropy can be used to derive Mutual Information. The entropy of a random variable A_i is defined as:

$$H(A_i) = - \sum_k p(a_k^i) \log_2(p(a_k^i)) \quad (9)$$

$$H(A_i|A_j) = - \sum_k p(a_k^i) \sum_l p(a_l^i|a_k^j) \log_2 p(a_l^i|a_k^j)$$

where $p(a_k^i)$ is the probability of the value a_k of A_i and $p(a_l^i|a_k^j)$ is the probability that $A_i = a_l$ after observing $A_j = a_k$. The mutual information is related to the conditional entropy through

$$MI(A_i|A_j) = H(A_i) - H(A_i|A_j) \quad (10)$$

The redundancy, R , is a more useful and symmetric scaled information measure, where:

$$R = \frac{MI(A_i; A_j)}{H(A_i) + H(A_j)} \quad (11)$$

The symmetric uncertainty is another alternative of the symmetrical measure given by:

$$DS(A_i; A_j) = SU(A_i; A_j) = 2R = 2 \frac{MI(A_i; A_j)}{H(A_i) + H(A_j)} \quad (12)$$

This measure varies between 0 and 1 (1 indicates that knowledge of the value of either A_i or A_j completely predicts the value of the other variable while 0 indicates that A_i and A_j are independent).

Finally, the relevance score for each feature A_i , $RS(A_i)$, is computed as the average dependence score between A_i and the rest of the feature:

$$RS(A_i) = \frac{1}{m-1} \sum_{A_j \in F \setminus A_i} DS(A_i; A_j) \tag{13}$$

where m denotes number of features and F is feature set. The lower the value of $RS(A_i)$, the lesser relevant of A_i . The following inequality is used to determine the context of an attribute A_i :

$$context(A_i) = \{A_j \in F \setminus A_i$$

$$DS(A_i; A_j) \geq 0.5RS(A_i)\}$$

Step 2: Distance Computation. The sum of distance of their attribute value pairs reflects distance between two data objects X and Y denoted by $D(X, Y)$, is:

$$D(X, Y) = \sum_{i=1}^m d_{A_i}(x_i, y_i) \tag{14}$$

where $D(x_i, y_i)$ denoted between two values x_i and y_i of attribute A_i :

$$d_{A_i}(x_i, y_i) = \sum_{A_j \in context(A_i)} KL\left(\begin{matrix} cpd(A_j|A_i = x_i) \\ cpd(A_j|A_i = y_i) \end{matrix}\right) \tag{15}$$

Dissimilarity between two values x_i and y_i of attribute A_i is directly proportional to dissimilarities of context's attributes given these values.

KullbackLeibler divergence method is used to compute dissimilarity between probability distributions.

$$KL(p, q) = \sum_i \left(p_i \log \frac{p_i}{q_i} + q_i \log \frac{q_i}{p_i} \right) \tag{16}$$

For each context attribute A_j , the conditional probability is computed for both values x_i and y_i given the values $v_j \in A_j$. Then, KullbackLeibler divergence method is applied. The dissimilarity between x_i and y_i equals to 0 if and only if the conditional probability distributions of other attributes when A_i holds values x_i and y_i are identical since KullbackLeibler dissimilarity between two probability distributions is non-negative, and equal to 0 if and only if the distributions are identical.

HYBRID SIMILARITY MEASURE

Alamuri et al. (2014) introduced a two-step hybrid similarity measure using context selection and distance computation. Context selection considers the meta attributes related to the current attributes called “context attributes”. The determination of the context of every attribute is data-driven and data-specific. Distance computation is made for the pair of values of an attribute. The context selection describes the steps adopted.

Alamuri et al. (2014) proposed a hybrid method based on entropy (Cover & Thomas, 1991) and mutual information (Shannon, 1948). This method is described below:

Let D be the data set with feature set $F = \{A_1, A_2, \dots, A_M\}$ of n data points, where each data object is of m attributes which are categorical. The entropy of a random variable A_i is defined as:

$$H(A_i) = - \sum_{k \in A_i} p(a_k^i) \log_2 p(a_k^i) \quad (17)$$

where $p(a_k^i)$ is the probability of value a_k of attribute A_i . The entropy of random variable can be conditioned on other variables. The conditional entropy A_i given A_j is:

$$H(A_i|A_j) = - \sum_{k \in A_j} p(a_k^j) \sum_{l \in A_i} p(a_l^i|a_k^j) \log_2 p(a_l^i|a_k^j) \quad (18)$$

where $p(a_l^i|a_k^j)$ is the probability that $A_i = a_l$. This means the amount of uncertainty present in A_i after observing the variable A_j . The amount of information shared between A_i and A_j (mutual information) is defined as:

$$I(A_i; A_j) = H(A_i) - H(A_i|A_j) \quad (19)$$

This is the difference between two entropies which can be interpreted as the amount of uncertainty in A_i which is removed by knowing A_j . After observing another variable A_z , the mutual information can also be conditioned as the amount of information still shared between A_i and A_j . The conditional mutual information is:

$$I(A_i; A_j|A_z) = H(A_i|A_z) - H(A_i|A_jA_z) \quad (20)$$

$$I(A_i; A_j|A_z) = \sum_{m \in A_z} p(a_m^z) \sum_{k \in A_i} \sum_{l \in A_j} p(a_k^i a_l^j | a_m^z) \log \frac{p(a_k^i a_l^j | a_m^z)}{p(a_k^i | a_m^z) p(a_l^j | a_m^z)}$$

KullbackLeibler divergence method is applied to calculate the distance between pair of values of an attribute. The formula is given by:

$$d_{A_i}(x_i, y_i) = \sum_{A_i \in context(A_i)} \sum_{v_j \in A_j} \left(p(v_j|x_i) \log \frac{p(v_j|x_i)}{p(v_j|y_i)} + p(v_j|y_i) \log \frac{p(v_j|y_i)}{p(v_j|x_i)} \right) \quad (21)$$

DISCUSSION & CONCLUSION

A review of several context-based similarity measures was conducted. First, components of four context-based similarity measures, including one hybrid similarity measure, were identified. The context-based similarity measures include Association Based Similarity Measure, DILCA, CBDL and the Hybrid Similarity Measure proposed by (Alamuri et al., 2014). The components for each context-based similarity measures and context-free similarity measure may be different in terms of their ability to compute similarity of attributes, similarity of objects, and distance computation. Furthermore, the concepts used to develop the similarity measures are found to be different between one similarity measure and the other. Specific components are shown in Table 1 below. Context-free similarity is based on a very simple concept which uses mainly distance computation using overlap measure. This measure does not take into consideration the relationship between data features. On the other hand, context-based similarity measures compute similarity between attributes and/or between objects. Above all, hybrid similarity measure is found to be the best since it measures similarity by considering all the three components.

Table 1
Components of the context-free in comparison to context-based similarity measure

	Association-based Similarity Measure	DILCA	CBDL	Hybrid Similarity Measure	Context-Free Similarity Measure
Similarity of attributes	Kullback Leibler	Entropy & Mutual Information	Entropy & Mutual Information	Entropy & Mutual Information	
Similarity of objects	Sum of dissimilarities of attributes value pair			Entropy & Mutual Information	
Distance computation		Euclidean distance	Kullback Leibler	Kullback Leibler	Overlap Measure

Second, a thorough comparison of four commonly used context-based similarity measures, including one hybrid similarity measure, was done. The context-based similarity measures include Association Based Similarity Measure, DILCA, CBDL and the Hybrid Similarity Measure proposed by (Alamuri et al., 2014). Four characteristics are discussed, namely algorithm, concepts, strengths and limitations of each method. A description of each characteristic for respective context-based similarity measure is provided in Table 2.

Table 2
Characteristic descriptions of four context-based similarity measures

	Association-based Dissimilarity Measure	DILCA	CBDL	Hybrid Context-Based Similarity Measure
Algorithm	Dissimilarities between two values of an attribute is found as the sum of the dissimilarities between conditional probability distributions of other related attributes. Finding the dissimilarity between two data objects.	Feature Selection is applied to select the relevant subset of the whole attributes in respect to the given one. Euclidean distance is applied to compute the distance between values of the same attribute.	Context Extraction Component is used to extract the relevant subset of feature set for a given attribute. Distance Learning Component is applied to learn distance between each pair of values of an attribute based on extracted context.	Context selection process looks at meta attributes associated with the current attributes. Distance computation is done for the pair of values of an attribute using the context defined in context selection.
Strengths	Experiments show that attributes are typically correlated. Lead to the idea of replacing each of the attribute groups by one or a few attributes that can have more discriminating power. Boost the accuracy of neural network when applied to real data.	Good clustering result is obtained when applied into clustering algorithm. A new methodology to compute a matrix of distances between any pair of values of a specific categorical attribute X The method is independent from the specific clustering algorithm. DILCA is considered a simple way to compute distances for categorical attribute. Attributes that introduce noise are ignored in the value distance computation step.	No sign of degradation when the number of irrelevant attributes increased. Accuracy was significantly higher when compared with the other popular similarity measure called Value Difference Metric (VDM). Improve the classification accuracy by reducing the effects of irrelevant attributes. Can be applied to any data mining task that involves categorical data.	Context selection is done by taking into consideration the meta attributes associated with the current attributes called context attributes.

Table 2 (continue)

	Association-based Dissimilarity Measure	DILCA	CBDL	Hybrid Context-Based Similarity Measure
Limitation	Cannot be applied to databases whose attributes are absolutely independent.	When the size of the dataset is small with respect to the number of attributes, it is expected that the results are biased by the weak representativeness of the samples. In some cases, performances are low for any clustering algorithm. The partitions determined by the class labels are not supported by data.		The context selection algorithm has the tendency to select the complete set of attributes as relevant context for the given attribute.

In summary, all the three context-based similarity measures reviewed above provide high accuracy when applied to clustering tasks. The Association based similarity measure can boost the accuracy of NN in clustering tasks. However, it cannot be applied when attributes are absolutely independent from one another. The strength of DILCA lies in the fact that the method uses simple distance computation and at the same time ignores the noise that exists in the attributes. However, it may produce biased results when dealing with small data set. The CBDL method showed no sign of degradation even as the number of irrelevant attributes increased. Alamuri's hybrid similarity measure has not been fully experimented. Its strength is based on its ability to take into consideration the meta attributes associated with the current attributes called context attributes. However, its context selection algorithm has the tendency to select the complete set of attributes as relevant context for the given attribute. In conclusion, context-based similarity measure is found to be highly recommended for data-mining tasks for categorical data. The findings of this paper will help data miners to choose the most appropriate similarity measure in achieving a more accurate classification or clustering result.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Teknologi MARA for its financial assistance.

REFERENCES

- Alamuri, M., Surampudi, B. R., & Negi, A. (2014, September). A survey of distance/similarity measures for categorical data. In *Conference on Neural Networks (IJCNN), 2014 International Joint* (pp. 1907-1914). IEEE.
- Cover T. M. & Thomas J. A. (1991). *Elements of Information Theory*. United States of America, USA: John Wiley & Sons.
- Desai, A., Singh, H., & Pudi, V. (2011, May). Disc: Data-intensive similarity measure for categorical data. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 469-481). Springer Berlin Heidelberg.
- Ienco D., Pensa. R. G. & Meo R. (2012). From Context to Distance: Learning Dissimilarity for Categorical Data Clustering. *ACM Transaction on Knowledge Discovery from Data*, 6(1), 1-25.
- Khorshidpour, Z., Hashemi, S., & Hamzeh, A. (2010, October). Distance learning for categorical attribute based on context information. In *2010 2nd International Conference on Software Technology and Engineering (ICSTE)*, (Vol. 2, pp. V2-296). IEEE.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1), 79-86.
- Le, S. Q., & Ho, T. B. (2005). An association-based dissimilarity measure for categorical data. *Pattern Recognition Letters*, 26(16), 2549-2557.
- Shannon C. E. (1948). A Mathematical Theory of Communication. *Bell Systems Technical Journal*, 27(3), 379-423.
- Yong J. B. (2010). *Data Mining Portfolio: Similarity Measure*. Retrieved from: humanoriented.com/classes/2010/fall/csci568/portfolio_exports/bhoenes/similarity.htm.



Assessing Stock Market Volatility for Different Sectors in Malaysia

Shakila, S.^{1,2*}, Noryati, A.³ and Maheran, M. J.¹

^{1,2}*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah alam, Selangor, Malaysia*

²*Institute of Engineering Mathematics, Universiti Malaysia Perlis, 02600 UMP, Arau, Perlis, Malaysia*

³*Arshad Ayub Graduate Business School, Universiti Teknologi MARA, 40450 UiTM, Shah alam, Selangor, Malaysia*

ABSTRACT

The study of stock market volatility has been the focus of market participants primarily because most of the applications in financial economics are concerned with volatility. The economic structure in Malaysia is divided into three sectors: primary, secondary and tertiary. As the stability of the stock market is important for businesses, this paper carefully reviews the concept of volatility and analyses how different business sectors in Malaysia are affected by stock market volatility.

Keywords: Historical volatility, stock market volatility, business sector, Malaysia

INTRODUCTION

Volatility estimation is important for financial practitioners, researchers and market participants for several reasons. The unexpected conditions such as changes in economic policy, political shocks, competitors and business rivals affect stock prices resulting in stock market volatility. Greater changes in stock prices indicate a high level of volatility. Commonly, higher volatility means more uncertainty in the stock market and this phenomenon has an impact on the financial stability of businesses.

According to Sill (1993), stock market volatility may affect the economy in terms of how people spend and save money, stock and option price and how investors may hedge against investment risks.

There are variety of methods that can be used to measure volatility such as equally weighted average, an exponentially weighted moving average (EWMA), and sophisticated models such as ARCH and GARCH (Cuthbertson & Nitzsche, 2001; Chang, 2014).

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail addresses:

shakila@unimap.edu.my (Shakila, S.),

noryatia@salam.uitm.edu.my (Noryati, A.),

maheran@tmsk.uitm.edu.my (Maheran, M. J.)

*Corresponding Author

Volatility is commonly used in estimating market risks (Ladokhin, 2009) in addition to valuing financial derivatives. Another important application can be seen in pricing the option since volatility is one of the parameters involved in estimating the option value.

Given the important role of volatility forecasting in the context of finance, this paper analyses volatility on different sectors of business in Malaysia. Economic theories divide economic sectors based on their activity. The main sector, primary sector, refers to industries engaged in the extraction of raw materials. Secondary or manufacturing sector uses raw materials sourced from the primary sector, processes them and produce finished goods. The tertiary sector offers or provides services to customers.

The remaining sections of this paper are organised as follows: Section 2 discusses historical and implied volatility. Definition of preliminary concepts related to volatility and their application in finance are also discussed in this section. Section 3 presents the methodology including data and steps used in measuring volatility. Section 4 provides the result on different sectors' volatility. Section 5 discusses and summarises the paper.

LITERATURE REVIEW

This section provides a brief discussion of approaches involved in estimating volatility which are historical and implied volatility. It also discusses the application of volatility estimation in the finance sector.

Measuring volatility

In most cases, the volatility value cannot be observed directly, thus the estimation of its value depends on the judgment by the analyst. There are two approaches in estimating the volatility value. Volatility value extracted from stock market returns is called historical volatility while the volatility value derived from option pricing model is called implied volatility (Koopman, Jungbacker & Hol, 2005). The difference between historical and implied volatility is discussed below.

Historical volatility

Historical volatility is usually referred to as backward looking volatility and uses past data over some period to estimate the volatility value. The standard and easiest way is by taking the standard deviation σ of the stock price return as below (Chang, 2014):

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (R_i - \bar{R})^2}{n-1}} \quad (1)$$

where R_i is a continuous return at time i , \bar{R} equals to mean daily return and n is the number of observations. R_i and \bar{R} are computed using formula

$$R_i = \ln \left[\frac{S_i}{S_{i-1}} \right] \quad (2)$$

and

$$\bar{R} = \frac{\sum_{i=1}^n R_i}{n}. \quad (3)$$

The variable S_i in (2) denotes the stock price at time i .

Using historical standard deviation, the only important information is past returns and other information is ignored although it might move the markets (Ederington & Guan, 2006). One may predict the movement of stock in the future since its value is based on how fast it has been moving in recent past.

Implied volatility

Implied volatility or forward looking volatility shows what the market “implies” about the future stock’s volatility. It is derived by solving the option pricing model e.g. Black Scholes formula for a call option as the equation below (Anoniemi, 2006):

$$C = SN(d_1) - Ke^{-r(T-t)}N(d_2) \quad (4)$$

where

$$d_1 = \frac{\ln \left(\frac{S}{K} \right) + \left(r + \frac{\sigma^2}{2} \right) T}{\sigma \sqrt{T-t}}$$

and

$$d_2 = d_1 - \sigma \sqrt{T-t}.$$

In the equation above, C is defined as call option value, S is underlying asset, $N(\cdot)$ is cumulative standard normal distribution, K is strike price, r equals to risk-free interest rate, $T-t$ is the time to maturity where T is the date at maturity and t is the current time, and σ is volatility. In order to compute the volatility value, all other inputs required by Black Scholes formula are known except the volatility value.

The implication of volatility in finance

In the field of finance, volatility is referred to as risk and can be measured as standard deviation (Poon & Granger, 2003). Essentially, standard deviation is used to measure the amount of dispersion of a set of data values. The data points tend to be very close to the expected value or mean if the value of standard deviation approaches zero. The large value of standard deviation means that the data points are far from mean. Another way to measure volatility is by using variance since variance and standard deviation are connected by a simple relationship (Ladokhin, 2009). However, this approach is less common compared with estimation using standard deviation.

Groud, Levy and Lubochinsky (2003) state that the most widely used concept for representing risk is the volatility of returns. However, Poon and Granger (2003) and Ladokhin (2009) do not agree. They believe that volatility is not the same as risk although it is related to it. This is because risk refers to negative outcomes of some event and volatility is about the spread of outcomes which can be positive or negative.

An important factor in valuing the option value using Black Scholes Merton model is volatility of stock prices (Rotkowsky, 2011). Modelling volatility is crucial to option pricing model since it is the only variable that is unobservable (Mitra, 2011). The option value relates to the asset's volatility value. Higher volatility leads to higher option value as the probability of the option to be valuable is higher.

METHODOLOGY

This section provides a detailed analysis of stock market volatility of 867 companies from different business sub-sectors in Malaysia. Data was taken from *Datastream*. The companies are grouped together based on their main business activity. There are 40 different sub-sectors which are furthered categorised into three main sectors - primary, secondary and tertiary. Table 1 shows the distribution of sub-sectors according to the main sector.

Data

Stock market prices are the source of volatility forecasts. The most common used price for estimating volatility is closing price. In this paper, the closing price is observed every day for 246 trading days. Data used in this paper was taken from the UiTM *Datastream* for the period of January 2014 to December 2014.

Table 1
Distribution of the sub-sector

Primary sector	Secondary sector	Tertiary sector
Mining & quarrying	Chemicals	Water, gas & multiutilities
Oil equipment & services	Tobacco	Non-life insurance
	Pharmaceuticals & biotechnology	Industrial transportation
	Technology hardware & equipment	Healthcare equipment & services
	Personal goods	Real estate investment & services
	Beverages	Financial services
	Food producers	Software & computer services
	Leisure goods	Real estate investment trusts
	General industrials	Life insurance
	Forestry & papers	Travel & leisure
	Construction & material	Food & drug retailers
	Electronic & electrical equipment	Fixed line telecommunication
	General industrials	Banks
	Automobiles & parts	Mobile telecommunication
	Household goods & home construction	Media
	Aerospace & defence	General retailers
	Industrial metal & mining	Non-equity investment instrument
	Industrial engineering	
	Support services	

Measuring Stock Market Volatility from Historical Data

This sub-section briefly explains in detail the procedure involved in the calculation of stock market volatility. The steps are as follows:

- Calculate the stock market return of each company involved by using the standard logarithmic method as in (2).
- Compute the standard deviation σ of the continuously stock market return by using formula in (1).
- Calculate the annualised volatility $\hat{\sigma}$ by:

$$\hat{\sigma} = \sigma\sqrt{\tau} \quad (5)$$

where τ represents the trading days per year. For this study τ is assumed equal to 246 days. Compute the volatility of each sub-sector by taking the average standard deviation of the companies with similar business activities.

The following section discusses the results based on the calculation above.

RESULT AND DISCUSSION

This section presents the result and discusses the volatility estimation for each sector, starting with the primary sector followed by secondary and tertiary sectors.

For the primary sector, there are only 2 sub-sectors with 24 companies (see Table 2).

Table 2
Distribution of the companies in primary sector

Sub-sector	Number of Companies
Mining & quarrying	2
Oil equipment & services	22
Total	24

Figure 1 shows the volatility in primary sector by sub-sectors.

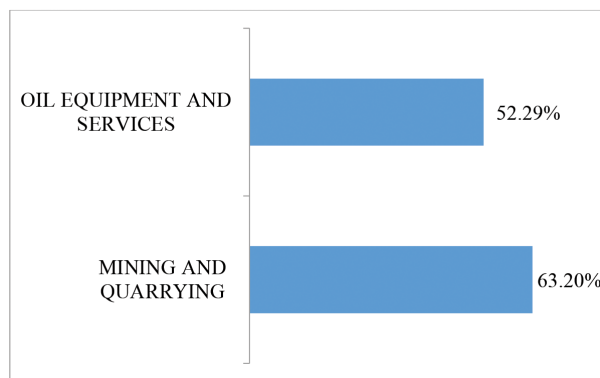


Figure 1. Volatility in primary sector by sub-sector

The volatility of mining and quarrying is 63.20% which is higher than oil equipment and services which is 52.29%. This means the mining and quarrying sub sector is more volatile compared with oil equipment and services. Appendix A presents the volatility value for mining and quarrying sector.

Table 3 shows the distribution of companies based on their sub-sectors. There are 20 sub-sectors with 559 companies.

Table 3
Distribution of companies in secondary sector

Sub-sector	Number of Companies
Chemicals	31
Tobacco	1
Pharmaceuticals & biotechnology	8
Technology hardware & equipment	25
Personal goods	21
Beverages	9
Food producers	77
General industrials	3
Forestry & papers	14
Construction & material	106
Leisure goods	6
Electronic & electrical equipment	40
General industrials	32
Automobiles & parts	19
Household goods & home construction	39
Aerospace & defence	1
Industrial metal & mining	33
Oil & gas producers	8
Industrial engineering	53
Support services	33
Total	559

The volatility for secondary sector is presented in Figure 2. The highest volatility is seen in the technology hardware and equipment sub sector, 63.72% and the lowest volatility in the tobacco sub sector, 20.33%. From the result, it is known that technology hardware and equipment sub sector is more volatile compared with the tobacco sub sector – about three times as volatile.

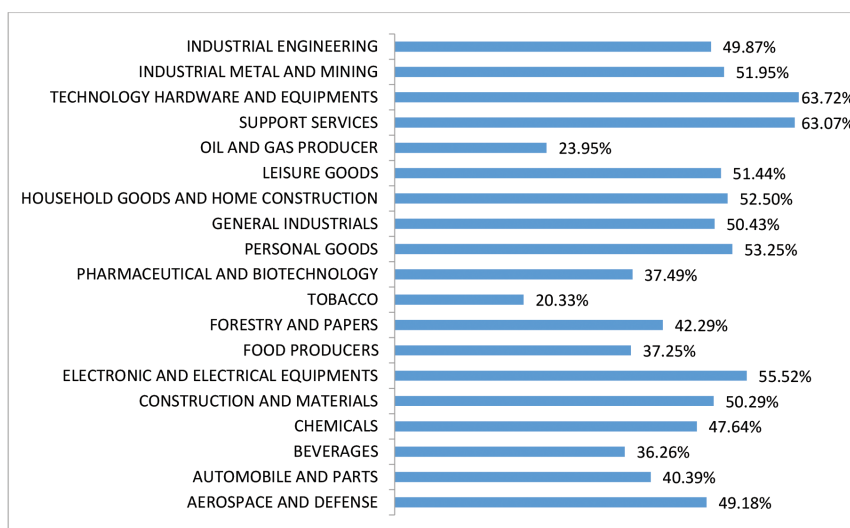


Figure 2. Volatility in secondary sector by sub-sector

The last sector is tertiary which has 18 sub-sectors with 324 companies. Table 4 shows the distribution of companies according to their sub-sectors.

Table 4
Distribution of companies in tertiary sector

Sub-sector	Number of Companies
Water, gas & multiutilities	9
Non-life insurance	8
Industrial transportation	32
Healthcare equipment & services	13
Real estate invest & services	79
Financial services	16
Software & computer services	57
Real estate investment trusts	15
Life insurance	1
Travel & leisure	29
Food & drug retailers	2
Fixed line telecommunication	4
Banks	10
Mobile telecommunication	8
Media	11
General retailers	23
Non-equity investment instrument	6
Equity investment instrument	1
Total	324

Volatility for the tertiary sector is shown in Figure 3. The fixed line telecommunications industry has the highest volatility, 86.49%, which is about 10 times more volatile than the lowest volatility which is non-equity investment instrument, 8.90%.

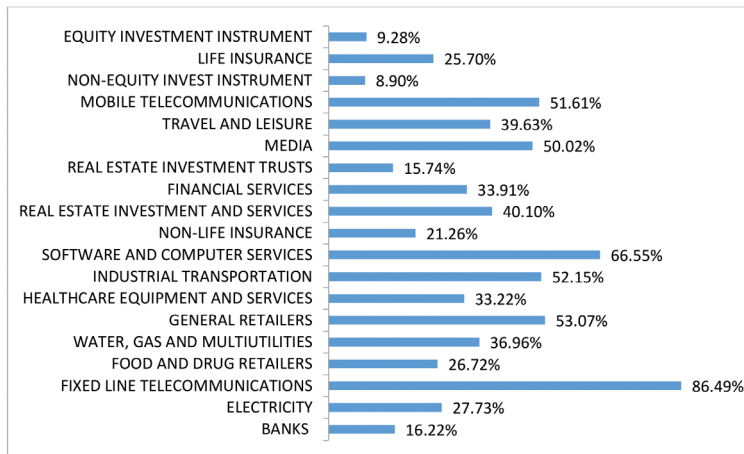


Figure 3. Volatility in tertiary sector by sub-sector

Based on the results, five sectors - mining and quarrying, support services, technology hardware and equipment, fixed line telecommunication, and software and computer services - have a volatility value of more than 60%. Sectors such as equity investment instrument, non-equity investment instrument, banks and real estate investment trusts have a volatility value of less than 20%. The annualised volatility of stocks normally lies between 20% and 60% (Hull, 2006).

The 35 sub-sector have more than 3 companies operating in similar industries except tobacco, aerospace and defence, life insurance, food and drug retailers, and equity investment instrument.

RESULT AND DISCUSSION

According to the results, the volatility values in different sub-sectors vary substantially. This implies the riskiness of individual company in each sub-sector. The greater the risk, the greater the potential for profits. Thus, investors with different risk portfolios can use these findings in re-creating their portfolio. Researchers may access the volatility of their countries' stock market by using the formula and methods used in this study.

The number of companies in each sub-sector should also be considered. For a meaningful comparison, the number of companies should be between 3 and 10 (Rotkowsky, 2011). Future research may examine company returns in each sector.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support received from the Tabung Amanah Pembangunan akademik Pelajar and the Fundamental Research Grant Scheme (FRGS) managed by Universiti Teknologi MARA, Ministry Higher Education of Malaysia and Universiti Malaysia Perlis.

REFERENCES

- Ahoniemi, K. (2006). Modeling and forecasting implied volatility: An econometric analysis of the VIX index. *Helsinki Center for Economic Research Discussion Paper*, 129, 1-32.
- Chang, V. (2014). The business intelligence as a service in the cloud. *Future Generation Computer Systems*, 37, 512-534.
- Cuthbertson, K., & Nitzsche, D. (2001). *Financial engineering: derivatives and risk management*. New Jersey, NJ: John Wiley & Sons, Inc.
- Ederington, L. H., & Guan, W. (2006). Measuring historical volatility. *Journal of Applied Finance*, 16(1), 1-10.
- Grouard, M. H., Lévy, S., & Lubochinsky, C. (2003). Stock market volatility: from empirical data to their interpretation. *FSR, Banque de France*, 2003(2), 57-74.
- Hull, J. C. (2006). *Options, futures, and other derivatives*. India: Pearson Education India.
- Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12(3), 445-475.
- Ladokhin, S. (2009). *Forecasting Volatility in the Stock Market*. (Unpublished Thesis). VU University Amsterdam, Faculty of Science.
- Mitra, S. (2011). A review of volatility and option pricing. *International Journal of Financial Markets and Derivatives*, 2(3), 149-179.
- Rotkowsky, A. M. (2011). Estimating Stock Price Volatility in the Black-Scholes-Merton Model. *A Professional Development Journal for the Consulting Disciplines*, 12-19.
- Sill, D. K. (1993). Predicting stock-market volatility. *Business Review*, (Jan), 15-28.

APPENDIX A

1 MINETECH RESOURCES				2 SINO HUA-AN INTL.			
Day	Date	Closing price	Return	Day	Date	Closing price	Return
1	31-Dec-14	0.12	23.3615%	1	31-Dec-14	0.11	0.0000%
2	30-Dec-14	0.095	0.0000%	2	30-Dec-14	0.11	4.6520%
3	29-Dec-14	0.095	0.0000%	3	29-Dec-14	0.105	-4.6520%
4	26-Dec-14	0.095	0.0000%	4	26-Dec-14	0.11	9.5310%
5	24-Dec-14	0.095	0.0000%	5	24-Dec-14	0.1	-4.8790%
6	23-Dec-14	0.095	0.0000%	6	23-Dec-14	0.105	0.0000%
7	22-Dec-14	0.095	-5.1293%	7	22-Dec-14	0.105	0.0000%
8	19-Dec-14	0.1	10.5361%	8	19-Dec-14	0.105	0.0000%
9	18-Dec-14	0.09	0.0000%	9	18-Dec-14	0.105	4.8790%
10	17-Dec-14	0.09	5.7158%	10	17-Dec-14	0.1	-4.8790%
11	16-Dec-14	0.085	0.0000%	11	16-Dec-14	0.105	-4.6520%
12	15-Dec-14	0.085	-5.7158%	12	15-Dec-14	0.11	-8.7011%
13	12-Dec-14	0.09	-5.4067%	13	12-Dec-14	0.12	0.0000%
14	11-Dec-14	0.095	5.4067%	14	11-Dec-14	0.12	0.0000%
15	10-Dec-14	0.09	0.0000%	15	10-Dec-14	0.12	0.0000%
16	9-Dec-14	0.09	0.0000%	16	9-Dec-14	0.12	-4.0822%
17	8-Dec-14	0.09	-5.4067%	17	8-Dec-14	0.125	0.0000%
18	5-Dec-14	0.095	0.0000%	18	5-Dec-14	0.125	4.0822%
19	4-Dec-14	0.095	-14.6603%	19	4-Dec-14	0.12	-4.0822%
20	3-Dec-14	0.11	0.0000%	20	3-Dec-14	0.125	0.0000%
21	2-Dec-14	0.11	-16.7054%	21	2-Dec-14	0.125	0.0000%
22	1-Dec-14	0.13	0.0000%	22	1-Dec-14	0.125	-3.9221%
23	28-Nov-14	0.13	-3.7740%	23	28-Nov-14	0.13	0.0000%
24	27-Nov-14	0.135	0.0000%	24	27-Nov-14	0.13	0.0000%
25	26-Nov-14	0.135	-7.1459%	25	26-Nov-14	0.13	0.0000%
26	25-Nov-14	0.145	-9.8440%	26	25-Nov-14	0.13	0.0000%
27	24-Nov-14	0.16	-3.0772%	27	24-Nov-14	0.13	0.0000%
28	21-Nov-14	0.165	-2.9853%	28	21-Nov-14	0.13	-3.7740%
29	20-Nov-14	0.17	2.9853%	29	20-Nov-14	0.135	3.7740%
30	19-Nov-14	0.165	3.0772%	30	19-Nov-14	0.13	-3.7740%
31	18-Nov-14	0.16	6.4539%	31	18-Nov-14	0.135	0.0000%
32	17-Nov-14	0.15	-6.4539%	32	17-Nov-14	0.135	-3.6368%
33	14-Nov-14	0.16	-3.0772%	33	14-Nov-14	0.14	0.0000%
34	13-Nov-14	0.165	3.0772%	34	13-Nov-14	0.14	0.0000%
35	12-Nov-14	0.16	6.4539%	35	12-Nov-14	0.14	0.0000%

36	11-Nov-14	0.15	6.8993%		36	11-Nov-14	0.14	0.0000%
37	10-Nov-14	0.14	0.0000%		37	10-Nov-14	0.14	0.0000%
38	7-Nov-14	0.14	3.6368%		38	7-Nov-14	0.14	3.6368%
39	6-Nov-14	0.135	3.7740%		39	6-Nov-14	0.135	0.0000%
40	5-Nov-14	0.13	-20.7639%		40	5-Nov-14	0.135	0.0000%
41	4-Nov-14	0.16	6.4539%		41	4-Nov-14	0.135	0.0000%
42	3-Nov-14	0.15	0.0000%		42	3-Nov-14	0.135	0.0000%
43	31-Oct-14	0.15	3.3902%		43	31-Oct-14	0.135	0.0000%
44	30-Oct-14	0.145	-3.3902%		44	30-Oct-14	0.135	0.0000%
45	29-Oct-14	0.15	0.0000%		45	29-Oct-14	0.135	0.0000%
46	28-Oct-14	0.15	3.3902%		46	28-Oct-14	0.135	0.0000%
47	27-Oct-14	0.145	0.0000%		47	27-Oct-14	0.135	0.0000%
48	24-Oct-14	0.145	0.0000%		48	24-Oct-14	0.135	-3.6368%
49	23-Oct-14	0.145	0.0000%		49	23-Oct-14	0.14	3.6368%
50	21-Oct-14	0.145	3.5091%		50	21-Oct-14	0.135	0.0000%
51	20-Oct-14	0.14	-3.5091%		51	20-Oct-14	0.135	0.0000%
52	17-Oct-14	0.145	0.0000%		52	17-Oct-14	0.135	3.7740%
53	16-Oct-14	0.145	0.0000%		53	16-Oct-14	0.13	-3.7740%
54	15-Oct-14	0.145	-3.3902%		54	15-Oct-14	0.135	0.0000%
55	14-Oct-14	0.15	3.3902%		55	14-Oct-14	0.135	-3.6368%
56	13-Oct-14	0.145	0.0000%		56	13-Oct-14	0.14	0.0000%
57	10-Oct-14	0.145	-3.3902%		57	10-Oct-14	0.14	-3.5091%
58	9-Oct-14	0.15	3.3902%		58	9-Oct-14	0.145	0.0000%
59	8-Oct-14	0.145	0.0000%		59	8-Oct-14	0.145	-3.3902%
60	7-Oct-14	0.145	0.0000%		60	7-Oct-14	0.15	0.0000%
61	3-Oct-14	0.145	0.0000%		61	3-Oct-14	0.15	0.0000%
62	2-Oct-14	0.145	0.0000%		62	2-Oct-14	0.15	0.0000%
63	1-Oct-14	0.145	0.0000%		63	1-Oct-14	0.15	-3.2790%
64	30-Sep-14	0.145	0.0000%		64	30-Sep-14	0.155	0.0000%
65	29-Sep-14	0.145	0.0000%		65	29-Sep-14	0.155	3.2790%
66	26-Sep-14	0.145	0.0000%		66	26-Sep-14	0.15	-3.2790%
67	25-Sep-14	0.145	-3.3902%		67	25-Sep-14	0.155	3.2790%
68	24-Sep-14	0.15	0.0000%		68	24-Sep-14	0.15	0.0000%
69	23-Sep-14	0.15	0.0000%		69	23-Sep-14	0.15	0.0000%
70	22-Sep-14	0.15	3.3902%		70	22-Sep-14	0.15	-3.2790%
71	19-Sep-14	0.145	-3.3902%		71	19-Sep-14	0.155	3.2790%
72	18-Sep-14	0.15	3.3902%		72	18-Sep-14	0.15	-3.2790%
73	17-Sep-14	0.145	0.0000%		73	17-Sep-14	0.155	3.2790%
74	15-Sep-14	0.145	0.0000%		74	15-Sep-14	0.15	0.0000%

Assessing Stock Market Volatility for Different Sectors in Malaysia

75	12-Sep-14	0.145	-3.3902%		75	12-Sep-14	0.15	-3.2790%
76	11-Sep-14	0.15	3.3902%		76	11-Sep-14	0.155	3.2790%
77	10-Sep-14	0.145	-3.3902%		77	10-Sep-14	0.15	0.0000%
78	9-Sep-14	0.15	0.0000%		78	9-Sep-14	0.15	0.0000%
79	8-Sep-14	0.15	0.0000%		79	8-Sep-14	0.15	0.0000%
80	5-Sep-14	0.15	3.3902%		80	5-Sep-14	0.15	0.0000%
81	4-Sep-14	0.145	0.0000%		81	4-Sep-14	0.15	-3.2790%
82	3-Sep-14	0.145	0.0000%		82	3-Sep-14	0.155	0.0000%
83	2-Sep-14	0.145	0.0000%		83	2-Sep-14	0.155	3.2790%
84	29-Aug-14	0.145	-3.3902%		84	29-Aug-14	0.15	-6.4539%
85	28-Aug-14	0.15	3.3902%		85	28-Aug-14	0.16	3.1749%
86	27-Aug-14	0.145	0.0000%		86	27-Aug-14	0.155	3.2790%
87	26-Aug-14	0.145	0.0000%		87	26-Aug-14	0.15	0.0000%
88	25-Aug-14	0.145	0.0000%		88	25-Aug-14	0.15	-6.4539%
89	22-Aug-14	0.145	0.0000%		89	22-Aug-14	0.16	0.0000%
90	21-Aug-14	0.145	3.5091%		90	21-Aug-14	0.16	3.1749%
91	20-Aug-14	0.14	-6.8993%		91	20-Aug-14	0.155	0.0000%
92	19-Aug-14	0.15	6.8993%		92	19-Aug-14	0.155	0.0000%
93	18-Aug-14	0.14	0.0000%		93	18-Aug-14	0.155	0.0000%
94	15-Aug-14	0.14	0.0000%		94	15-Aug-14	0.155	0.0000%
95	14-Aug-14	0.14	-3.5091%		95	14-Aug-14	0.155	0.0000%
96	13-Aug-14	0.145	3.5091%		96	13-Aug-14	0.155	3.2790%
97	12-Aug-14	0.14	0.0000%		97	12-Aug-14	0.15	-3.2790%
98	11-Aug-14	0.14	0.0000%		98	11-Aug-14	0.155	3.2790%
99	8-Aug-14	0.14	-6.8993%		99	8-Aug-14	0.15	-3.2790%
100	7-Aug-14	0.15	0.0000%		100	7-Aug-14	0.155	-3.1749%
101	6-Aug-14	0.15	6.8993%		101	6-Aug-14	0.16	3.1749%
102	5-Aug-14	0.14	0.0000%		102	5-Aug-14	0.155	0.0000%
103	4-Aug-14	0.14	-3.5091%		103	4-Aug-14	0.155	3.2790%
104	1-Aug-14	0.145	0.0000%		104	1-Aug-14	0.15	0.0000%
105	31-Jul-14	0.145	-3.3902%		105	31-Jul-14	0.15	0.0000%
106	30-Jul-14	0.15	3.3902%		106	30-Jul-14	0.15	0.0000%
107	25-Jul-14	0.145	3.5091%		107	25-Jul-14	0.15	3.3902%
108	24-Jul-14	0.14	-3.5091%		108	24-Jul-14	0.145	-3.3902%
109	23-Jul-14	0.145	-3.3902%		109	23-Jul-14	0.15	3.3902%
110	22-Jul-14	0.15	0.0000%		110	22-Jul-14	0.145	0.0000%
111	21-Jul-14	0.15	0.0000%		111	21-Jul-14	0.145	0.0000%
112	18-Jul-14	0.15	14.3101%		112	18-Jul-14	0.145	0.0000%
113	17-Jul-14	0.13	0.0000%		113	17-Jul-14	0.145	0.0000%

114	16-Jul-14	0.13	-3.7740%		114	16-Jul-14	0.145	0.0000%
115	14-Jul-14	0.135	0.0000%		115	14-Jul-14	0.145	0.0000%
116	11-Jul-14	0.135	0.0000%		116	11-Jul-14	0.145	-3.3902%
117	10-Jul-14	0.135	3.7740%		117	10-Jul-14	0.15	0.0000%
118	9-Jul-14	0.13	0.0000%		118	9-Jul-14	0.15	3.3902%
119	8-Jul-14	0.13	-7.4108%		119	8-Jul-14	0.145	-6.6691%
120	7-Jul-14	0.14	3.6368%		120	7-Jul-14	0.155	3.2790%
121	4-Jul-14	0.135	-10.5361%		121	4-Jul-14	0.15	0.0000%
122	3-Jul-14	0.15	0.0000%		122	3-Jul-14	0.15	0.0000%
123	2-Jul-14	0.15	0.0000%		123	2-Jul-14	0.15	0.0000%
124	1-Jul-14	0.15	0.0000%		124	1-Jul-14	0.15	3.3902%
125	30-Jun-14	0.15	-3.2790%		125	30-Jun-14	0.145	0.0000%
126	27-Jun-14	0.155	0.0000%		126	27-Jun-14	0.145	-3.3902%
127	26-Jun-14	0.155	0.0000%		127	26-Jun-14	0.15	0.0000%
128	25-Jun-14	0.155	3.2790%		128	25-Jun-14	0.15	0.0000%
129	24-Jun-14	0.15	6.8993%		129	24-Jun-14	0.15	0.0000%
130	23-Jun-14	0.14	0.0000%		130	23-Jun-14	0.15	0.0000%
131	20-Jun-14	0.14	-3.5091%		131	20-Jun-14	0.15	0.0000%
132	19-Jun-14	0.145	-3.3902%		132	19-Jun-14	0.15	0.0000%
133	18-Jun-14	0.15	0.0000%		133	18-Jun-14	0.15	0.0000%
134	17-Jun-14	0.15	0.0000%		134	17-Jun-14	0.15	3.3902%
135	16-Jun-14	0.15	14.3101%		135	16-Jun-14	0.145	0.0000%
136	13-Jun-14	0.13	0.0000%		136	13-Jun-14	0.145	0.0000%
137	12-Jun-14	0.13	-3.7740%		137	12-Jun-14	0.145	0.0000%
138	11-Jun-14	0.135	0.0000%		138	11-Jun-14	0.145	0.0000%
139	10-Jun-14	0.135	7.6961%		139	10-Jun-14	0.145	0.0000%
140	9-Jun-14	0.125	-3.9221%		140	9-Jun-14	0.145	0.0000%
141	6-Jun-14	0.13	3.9221%		141	6-Jun-14	0.145	0.0000%
142	5-Jun-14	0.125	-3.9221%		142	5-Jun-14	0.145	-3.3902%
143	4-Jun-14	0.13	3.9221%		143	4-Jun-14	0.15	0.0000%
144	3-Jun-14	0.125	0.0000%		144	3-Jun-14	0.15	3.3902%
145	2-Jun-14	0.125	-3.9221%		145	2-Jun-14	0.145	0.0000%
146	30-May-14	0.13	0.0000%		146	30-May-14	0.145	0.0000%
147	29-May-14	0.13	3.9221%		147	29-May-14	0.145	0.0000%
148	28-May-14	0.125	-3.9221%		148	28-May-14	0.145	-3.3902%
149	27-May-14	0.13	3.9221%		149	27-May-14	0.15	6.8993%
150	26-May-14	0.125	-3.9221%		150	26-May-14	0.14	0.0000%
151	23-May-14	0.13	0.0000%		151	23-May-14	0.14	-3.5091%
152	22-May-14	0.13	-3.7740%		152	22-May-14	0.145	0.0000%

Assessing Stock Market Volatility for Different Sectors in Malaysia

153	21-May-14	0.135	0.0000%		153	21-May-14	0.145	0.0000%
154	20-May-14	0.135	0.0000%		154	20-May-14	0.145	0.0000%
155	19-May-14	0.135	0.0000%		155	19-May-14	0.145	3.5091%
156	16-May-14	0.135	0.0000%		156	16-May-14	0.14	-3.5091%
157	15-May-14	0.135	-3.6368%		157	15-May-14	0.145	-3.3902%
158	14-May-14	0.14	0.0000%		158	14-May-14	0.15	6.8993%
159	12-May-14	0.14	0.0000%		159	12-May-14	0.14	0.0000%
160	9-May-14	0.14	-3.5091%		160	9-May-14	0.14	-3.5091%
161	8-May-14	0.145	0.0000%		161	8-May-14	0.145	3.5091%
162	7-May-14	0.145	7.1459%		162	7-May-14	0.14	-6.8993%
163	6-May-14	0.135	0.0000%		163	6-May-14	0.15	3.3902%
164	5-May-14	0.135	-3.6368%		164	5-May-14	0.145	-3.3902%
165	2-May-14	0.14	-3.5091%		165	2-May-14	0.15	0.0000%
166	30-Apr-14	0.145	3.5091%		166	30-Apr-14	0.15	0.0000%
167	29-Apr-14	0.14	-3.5091%		167	29-Apr-14	0.15	-3.2790%
168	28-Apr-14	0.145	-3.3902%		168	28-Apr-14	0.155	-3.1749%
169	25-Apr-14	0.15	0.0000%		169	25-Apr-14	0.16	6.4539%
170	24-Apr-14	0.15	0.0000%		170	24-Apr-14	0.15	-6.4539%
171	23-Apr-14	0.15	3.3902%		171	23-Apr-14	0.16	9.8440%
172	22-Apr-14	0.145	-3.3902%		172	22-Apr-14	0.145	0.0000%
173	21-Apr-14	0.15	0.0000%		173	21-Apr-14	0.145	-3.3902%
174	18-Apr-14	0.15	0.0000%		174	18-Apr-14	0.15	3.3902%
175	17-Apr-14	0.15	6.8993%		175	17-Apr-14	0.145	0.0000%
176	16-Apr-14	0.14	-6.8993%		176	16-Apr-14	0.145	-3.3902%
177	15-Apr-14	0.15	-3.2790%		177	15-Apr-14	0.15	3.3902%
178	14-Apr-14	0.155	0.0000%		178	14-Apr-14	0.145	0.0000%
179	11-Apr-14	0.155	3.2790%		179	11-Apr-14	0.145	0.0000%
180	10-Apr-14	0.15	6.8993%		180	10-Apr-14	0.145	-3.3902%
181	9-Apr-14	0.14	7.4108%		181	9-Apr-14	0.15	6.8993%
182	8-Apr-14	0.13	0.0000%		182	8-Apr-14	0.14	-3.5091%
183	7-Apr-14	0.13	-3.7740%		183	7-Apr-14	0.145	0.0000%
184	4-Apr-14	0.135	-3.6368%		184	4-Apr-14	0.145	0.0000%
185	3-Apr-14	0.14	0.0000%		185	3-Apr-14	0.145	-3.3902%
186	2-Apr-14	0.14	0.0000%		186	2-Apr-14	0.15	3.3902%
187	1-Apr-14	0.14	0.0000%		187	1-Apr-14	0.145	0.0000%
188	31-Mar-14	0.14	-3.5091%		188	31-Mar-14	0.145	0.0000%
189	28-Mar-14	0.145	-3.3902%		189	28-Mar-14	0.145	0.0000%
190	27-Mar-14	0.15	3.3902%		190	27-Mar-14	0.145	0.0000%
191	26-Mar-14	0.145	-3.3902%		191	26-Mar-14	0.145	0.0000%

192	25-Mar-14	0.15	0.0000%		192	25-Mar-14	0.145	-3.3902%
193	24-Mar-14	0.15	-3.2790%		193	24-Mar-14	0.15	3.3902%
194	21-Mar-14	0.155	3.2790%		194	21-Mar-14	0.145	0.0000%
195	20-Mar-14	0.15	0.0000%		195	20-Mar-14	0.145	0.0000%
196	19-Mar-14	0.15	3.3902%		196	19-Mar-14	0.145	0.0000%
197	18-Mar-14	0.145	0.0000%		197	18-Mar-14	0.145	3.5091%
198	17-Mar-14	0.145	0.0000%		198	17-Mar-14	0.14	-3.5091%
199	14-Mar-14	0.145	-3.3902%		199	14-Mar-14	0.145	0.0000%
200	13-Mar-14	0.15	0.0000%		200	13-Mar-14	0.145	0.0000%
201	12-Mar-14	0.15	-6.4539%		201	12-Mar-14	0.145	-6.6691%
202	11-Mar-14	0.16	-3.0772%		202	11-Mar-14	0.155	3.2790%
203	10-Mar-14	0.165	0.0000%		203	10-Mar-14	0.15	3.3902%
204	7-Mar-14	0.165	-2.9853%		204	7-Mar-14	0.145	-3.3902%
205	6-Mar-14	0.17	-2.8988%		205	6-Mar-14	0.15	3.3902%
206	5-Mar-14	0.175	2.8988%		206	5-Mar-14	0.145	3.5091%
207	4-Mar-14	0.17	-5.7158%		207	4-Mar-14	0.14	0.0000%
208	3-Mar-14	0.18	0.0000%		208	3-Mar-14	0.14	0.0000%
209	28-Feb-14	0.18	-2.7399%		209	28-Feb-14	0.14	-6.8993%
210	27-Feb-14	0.185	-2.6668%		210	27-Feb-14	0.15	3.3902%
211	26-Feb-14	0.19	-7.5986%		211	26-Feb-14	0.145	0.0000%
212	25-Feb-14	0.205	0.0000%		212	25-Feb-14	0.145	3.5091%
213	24-Feb-14	0.205	2.4693%		213	24-Feb-14	0.14	-6.8993%
214	21-Feb-14	0.2	-7.2321%		214	21-Feb-14	0.15	0.0000%
215	20-Feb-14	0.215	-24.6133%		215	20-Feb-14	0.15	0.0000%
216	19-Feb-14	0.275	-5.3110%		216	19-Feb-14	0.15	3.3902%
217	18-Feb-14	0.29	-8.2692%		217	18-Feb-14	0.145	-3.3902%
218	17-Feb-14	0.315	3.2261%		218	17-Feb-14	0.15	0.0000%
219	14-Feb-14	0.305	5.0431%		219	14-Feb-14	0.15	3.3902%
220	13-Feb-14	0.29	-5.0431%		220	13-Feb-14	0.145	-3.3902%
221	12-Feb-14	0.305	-3.2261%		221	12-Feb-14	0.15	3.3902%
222	11-Feb-14	0.315	6.5597%		222	11-Feb-14	0.145	3.5091%
223	10-Feb-14	0.295	14.5712%		223	10-Feb-14	0.14	0.0000%
224	7-Feb-14	0.255	0.0000%		224	7-Feb-14	0.14	7.4108%
225	6-Feb-14	0.255	10.3184%		225	6-Feb-14	0.13	0.0000%
226	5-Feb-14	0.23	6.7441%		226	5-Feb-14	0.13	0.0000%
227	4-Feb-14	0.215	17.7681%		227	4-Feb-14	0.13	-3.7740%
228	30-Jan-14	0.18	0.0000%		228	30-Jan-14	0.135	0.0000%
229	29-Jan-14	0.18	0.0000%		229	29-Jan-14	0.135	0.0000%
230	28-Jan-14	0.18	2.8171%		230	28-Jan-14	0.135	-3.6368%

Assessing Stock Market Volatility for Different Sectors in Malaysia

231	27-Jan-14	0.175	-2.8171%		231	27-Jan-14	0.14	0.0000%
232	24-Jan-14	0.18	5.7158%		232	24-Jan-14	0.14	3.6368%
233	23-Jan-14	0.17	-5.7158%		233	23-Jan-14	0.135	-3.6368%
234	22-Jan-14	0.18	-8.0043%		234	22-Jan-14	0.14	3.6368%
235	21-Jan-14	0.195	8.0043%		235	21-Jan-14	0.135	-3.6368%
236	20-Jan-14	0.18	-8.0043%		236	20-Jan-14	0.14	0.0000%
237	16-Jan-14	0.195	-2.5318%		237	16-Jan-14	0.14	0.0000%
238	15-Jan-14	0.2	-9.5310%		238	15-Jan-14	0.14	0.0000%
239	13-Jan-14	0.22	4.6520%		239	13-Jan-14	0.14	-3.5091%
240	10-Jan-14	0.21	10.0083%		240	10-Jan-14	0.145	3.5091%
241	9-Jan-14	0.19	0.0000%		241	9-Jan-14	0.14	0.0000%
242	8-Jan-14	0.19	2.6668%		242	8-Jan-14	0.14	0.0000%
243	7-Jan-14	0.185	0.0000%		243	7-Jan-14	0.14	3.6368%
244	6-Jan-14	0.185	-2.6668%		244	6-Jan-14	0.135	0.0000%
245	3-Jan-14	0.19	2.6668%		245	3-Jan-14	0.135	0.0000%
246	2-Jan-14	0.185			246	2-Jan-14	0.135	
	Volatility (day)		5.12%			Volatility (day)		2.94%
	Volatility (annually)		80.28%			Volatility (annually)		46.12%
	Average (volatility)				63.20%			



A Railway Rescheduling Model with Priority Setting

Shuib, A. and Alwadood, Z.*

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia

ABSTRACT

This paper presents a mathematical approach to solve railway rescheduling problems. The approach assumes that the trains are able to resume their journey after a given time frame of disruption whereby The train that experiences disruption and trains affected by the incident are rescheduled. The approach employed mathematical model to prioritise certain types of train according the railway operator's requirement. A pre-emptive goal programming model was adapted to find an optimal solution that satisfies the operational constraints and the company's stated goals. Initially, the model minimises the total service delay of all trains while adhering to the minimum headway requirement and track capacity. Subsequently, it maximises the train service reliability by only considering the trains with delay time window of five minutes or less. The model uses MATLAB R2014a software which automatically generates the optimal solution of the problem based on the input matrix of constraints. An experiment with three incident scenarios on a double-track railway of local network was conducted to evaluate the performance of the proposed model. The new provisional timetable was produced in short computing time and the model was able to prioritise desired train schedule.

Keywords: Mathematical optimisation model, mixed integer programming, service delays, railway rescheduling

INTRODUCTION

A variety of factors can lead to operational problems in railway transportation network, such as equipment failures, track damage, extraordinary passenger volumes, train accidents or weather conditions. When these unexpected incidents occur, the control managers need to reshuffle train orders, make unplanned stops and break connections, re-route trains and even delay or cancel scheduled services. Changes in the original train departure and arrival schedules can create conflicts in the use of resources, such as tracks, platforms, rolling

Article history:

Received: 27 May 2016

Accepted: 14 November 2016

E-mail addresses:

adibah@tmsk.uitm.edu.my (Shuib, A.),

zuraida794@salam.uitm.edu.my (Alwadood, Z.)

*Corresponding Author

stock and crew. Thus, rescheduling affected train services and managing delays with an objective to minimise traffic disruptions is the duty of the operational section of the railway company.

The main objective of this research is to present an optimal solution for solving post-disruption railway rescheduling problem. An output in the form of new provisional timetable is aimed at minimising the total delays of trains in the whole railway network. In order to achieve this objective, a mixed integer programming (MIP) model for rescheduling railway is proposed.

An earlier paper on this topic have illustrated the important elements for the proposed model construction, including the sets, parameters and variables. We have developed a mathematical model that incorporated all the technical and practical constraints that are subject to resource limitations. In this paper, we present the evaluation of the models by setting priority to a certain type of trains. Our aim is to produce a schedule which satisfies the safety rules and other operating requirements of all railway services, including the prioritised ones. This work is expected to highlight the future direction of the modelling works by bringing new ideas for multiple perspective improvements in delay management, as well as business and quality engineering process.

This paper is outlined as follows: The next section discusses some relevant literatures on the railway rescheduling model and followed by the presentation of the mathematical model. The succeeding section highlights the computational experiment, while the conclusion and further research of the study are drawn at the end of the paper.

RELATED WORKS

Over the last few decades, the topic of real-time rescheduling of railway traffic during disturbances has attracted tremendous attention from many researchers. A wide range of mathematical programming models have been utilized for solving the train rescheduling problem in managing train service delays. Mathematical programming is commonly selected to be the tool for the rail traffic analysis and solution because of its capability of incorporating observed reality of the system into the model's constraints and objective function(s). Furthermore, it is also able to address the highly combinatorial problem and interlinked nature of the rail traffic system.

Some early delay management problems were formulated as an MIP model, for example, the model based on event-activity network (Schobel, 2001). It deals with the delay involved in transportation network. The objective of the model is to minimize the sum over all delays of all customers in the network. The work was later extended by Schachtebeck & Schobel (2008) to a railway service whereby this time a priority decision was added to an integer programming (IP) model. As the rail track system is subject to a limited capacity, the priority decisions would decide the sequence of trains in passing a track. A reduction technique derived for the network is used as the basis to solve the capacitated delay management problem using exact and heuristics approaches. The work contributes a significant finding as the four heuristics suggested are able to yield a faster result than the optimal solution obtained from IP formulation.

Acuna-Agost et al. (2009) proposed two models namely an MIP model and a Constraint Programming (CP) model of which both shares common objective function in minimizing delays, changes of tracks and platform and unplanned stops. Both models differ in terms of

decision variables and constraints. Although utilizing more memory as compared to the CP, the proposed MIP is found to have developed more efficient solution methods when compared to CP. This is because, in CP, a large number of binary variables is needed to model the order of trains. The authors described the complexity of these two models in relation to its allocation of tracks and platforms, connection between trains, bidirectional/multi-track lines and extra time for accelerating and braking.

A recent MIP model which is established by Narayanaswami & Rangaraj (2013) includes disruption and conflicts-resolving constraints in the model itself. The novelty of the method ensures that only disrupted trains will be rescheduled, leaving alone the unaffected trains. This is done by partitioning train sets into conflicting and non-conflicting trains by means of linear constraints. To solve the model, the traveling salesman problem (TSP) approach was applied. The model is NP-Complete as it is reduced to a TSP in a polynomial time. Since the model used a small size fictive data, it should be extended to a larger scale of real data so that the validity of the model can be proven.

A model predictive control approach attempts to reschedule trains by a discrete-time control (Caimi et al., 2012). A set of alternative blocking-stairways is used as the basis for each rescheduling step. This is followed by several planning steps which are linked to each other by different temporal scopes. The concept of bi-level multi-objective formulation means three criteria are considered separately in the first level. They are then aggregated into one objective function as a weighted sum. This method is appropriate when it comes to optimising a multi-criteria objective because the choice of weights depends absolutely on the dispatcher or the experts, based on the importance of the criteria valued.

The possibilities of rerouting trains were considered by Veelenturf et al. (2014) in the attempt to minimise the number of cancelled and delayed trains. The IP model tested on a busy part of Dutch railway network was able to solve most of their cases in two minutes. The method was claimed to be much more flexible and efficient than the current practice which uses contingency plans.

A fuzzy rescheduling model in a double-track railway network was proposed by Yang et al. (2014). They formulate the problem by introducing a space-time network-based representation to capture the uncertainty of incident duration. Taking the duration as a fuzzy variable, a two-stage fuzzy programming model was then formulated to generate the optimised train schedule. In relation to mathematical programming aspects, the proposed model can be viewed as an initial step towards incorporating fuzzy factors into a train rescheduling process. Apart from this, last train timetable rescheduling using genetic algorithm in order to minimise the train running and dwelling times was proposed by Kang et al. (2015). The model modifies the original timetable as little as possible to ensure that the train delay will be minimised while satisfying both hard and soft constraints.

Tornquist & Persson (2005) formulated their combinatorial problem of real-time disturbances in railway traffic rescheduling with the objective of minimising total service delay. An iterative two-level process is used to solve the problem. The order of meeting and overtaking of trains on the track section is done at the upper level using simulated annealing and tabu search while the lower level process determines the start and end time for each train. Local reordering trains are used to obtain good quality solutions.

The study was later extended by using a heterogeneous data set with more interacting traffic, to analyse the practicality of the model and solution approach (Tornquist & Persson, 2007). In this work, the authors presented a model and solution approach for the railway traffic rescheduling problem with a highly complex setting, taking into account the large number of railway tracks and segments and each rail network direction. A mixed integer linear programming model (MILP) representing the disturbed n -tracked network was evaluated and demonstrated on a complex Swedish railway network. The approaches used have shown good computational capabilities and they extended their work in the attempt to improve the computing time taken towards arriving at the rescheduling solutions (Krasemann, 2012). Besides complementing the earlier approach, the new greedy algorithm is able to provide good solutions within the permitted time.

THE MILP MODEL

The MILP model presented in this paper is an extension of the model established in Tornquist & Persson (2007) and uses basically the same variables and parameters. There are two main differences between the current proposed model and theirs. First, our model prioritises electric trains (ETS) over commuter trains whenever a conflict occurs. During disruption, the commuter is set to wait and give way to the ETS while holding to the limited capacity of tracks available at different segments of the railway line. The second difference is in regard to the switches along the tracks, which are modelled explicitly in this formulation.

The rail track between any two switches is defined as a segment. B denotes the set of segments considered in the rescheduling problem where each segment is labelled with index j , where $j \in B$. While i is defined as the index of the set of trains T , the set of events E is denoted by index k . In this model, the final event of the ordered set of train event is referred to as n_i ($i \in T$) and the disrupted event is represented by q , where $q \in E$.

An ordinary train route has a signal switch control located right before the block entrance. Once accepted, no other trains will be allowed on the block or segment without permission from the signal switch. The distance between the signals which varies along the track is based on the geographical safety aspects such as the elevation of ground and the blind spots at curves and turns.

There are two types of segment location l , which are the non-station segment and station segment. For each segment, there is a set of parallel tracks $P_j = \{l, \dots, p_j\}$. A standard railway safety regulation normally imposes a minimum distance between two consecutive trains to avoid accident, which is termed as time headway. It indicates the minimum duration between the time when a train exits from a segment and the subsequent train enters the same segment. The parameter H_j denotes the time headway in cases where one train is following or meeting the other on a track of segment j .

Table 1 presents the elements in mathematical programming model, consisting of variables, sets and parameters of the formulation. There are seven decision variables anticipated from the method. The first two are d_k^R and a_k^R , which represent the departure time and the arrival time of the rescheduled event k , where $k \in E$, $K_i \subseteq E$ and $L_j \subseteq E$. Another four are the common

scheduling disjunctive binary decision variables, namely r_k^t , τ_i , γ_{kk} and λ_{kk} . As a result of the new schedule, the amount of delay that will be recorded from the rescheduling event k will be denoted by the decision variable z_k .

Table 1
Variables, sets and parameters of model

Component	Type	Description
d_k^R	Decision variables	The departure time of the rescheduled event k
a_k^R	Decision variables	The arrival time of the rescheduled event k
Z_k	Decision variables	The amount of delay resulting from the rescheduling event k
r_k^t	Binary variables	The decision if event k is assigned to track t
τ_i	Binary variables	The decision if train i arrives within the delay time window
γ_{kk}	Binary variables	The decision of the order of the occurrence of event k and \hat{k}
λ_{kk}	Binary variables	The decision of the order of the occurrence of event \hat{k} and k
T	Set	The set of trains considered in the rescheduling problem
B	Set	The set of segment considered in the rescheduling problem
K_i	Set	The ordered set of events of train i
L_j	Set	The ordered set of events of segment j
E_k	Set	The set of event k
l_j	Set	The types of segment j
P_j	Set	The set of parallel tracks on segment j
w_i	Parameters	The delay time window for train i
H_j	Parameter	The time headway between two consecutive train on a track of segment j
o_k	Parameter	The direction of event k
A_k	Parameter	The minimum time for an event k running at a segment
d_k^S	Parameter	The scheduled departure time of event k as in timetable
a_k^S	Parameter	The scheduled arrival time of event k as in timetable
d_k^A	Parameter	The actual departure time of event k
a_k^A	Parameter	The actual arrival time of event k
α_k	Constant	Arbitrarily large positive constant
M		

Unlike linear programming model which normally consists of a set of functional constraints and a single objective function to be maximised or minimised, a goal programming model seeks to satisfy several objectives or goals, subject to a set of constraints that are prioritised in some sense. The objective of a goal programming is to find a solution that satisfies the constraints and the stated goals. In some cases, the feasible solutions of the goal programming may not satisfy all the conflicting goals simultaneously. This implies that the set of feasible solutions does meet the restrictions posed by the constraints, but it does not optimise all the objective functions in one go.

The technique of pre-emptive goal programming will be applied to solve the MILP model. In this method, minimising train delays will be the first priority while maximising service reliability will be the subsequent one. This is considered as the optimal solution to the goal programming problem. In this case, we consider the delay time window w_i to be less than or equals to 5 minutes. The MILP model is given in Figure 1.

Minimize $\sum_{i=1}^{n_i} Z_{n_i}$	(1a)
Maximize $\frac{\sum_{i=1}^{n_i} \tau_i}{\sum_{i=1}^{n_i} T_i}$	(1b)
Subject to	
$a_k^R \leq a_{k+1}^R, k \in K_i, i \in T : k \neq n_i,$	(2a)
$a_k^R \geq a_k^R + \Delta_k, k \in E,$	(2b)
$a_q^R \geq a_q^R, q \in E,$	(2c)
$a_k^R = a_{k-1}^R, k \in K_i, i \in T : k \neq n_i,$	(3a)
$a_k^R = a_k^R + \Delta_k, k \in E,$	(3b)
$d_k^R \geq d_k^S, k \in E,$	(4)
$d_k^R = d_k^A, k \in E : d_k^S > 0,$	(5)
$a_k^R = a_k^A, k \in E : a_k^A > 0,$	(6)
$a_k^R - a_k^S \leq z_k, k \in E.$	(7)
$\sum_{i=1}^{p_j} r_k^i = 1, k \in L_j, j \in B,$	(8)
$r_k^i + r_{\hat{k}}^i - 1 \leq \lambda_{kk} + \gamma_{kk}, k, \hat{k} \in L_j, j \in B : k < \hat{k},$	(9)
$\lambda_{kk} + \gamma_{kk} \leq 1, k, \hat{k} \in L_j, j \in B : k < \hat{k},$	(10)
$d_k^R - a_k^R \geq H_j \gamma_{kk} - M(1 - \gamma_{kk}), k, \hat{k} \in L_j, j \in B : k < \hat{k}, o_{\hat{k}} = o_k, l_j = 0$	(11a)
$d_k^R - a_k^R \geq H_j \gamma_{kk} - M(1 - \gamma_{kk}), k, \hat{k} \in L_j, j \in B : k < \hat{k}, o_{\hat{k}} = o_k, l_j = 1$	(11b)
$d_k^R - a_k^R \geq H_j \lambda_{kk} - M(1 - \lambda_{kk}), k, \hat{k} \in L_j, j \in B : k < \hat{k}, o_{\hat{k}} = o_k, l_j = 0$	(12a)
$d_k^R - a_k^R \geq H_j \lambda_{kk} - M(1 - \lambda_{kk}), k, \hat{k} \in L_j, j \in B : k < \hat{k}, o_{\hat{k}} = o_k, l_j = 1$	(12b)
$d_k^R, a_k^R, z_k \geq 0, k \in E,$	(13)
$r_k^i \in \{0, 1\}, k \in L_j, i \in P_j, j \in B,$	(14)
$\gamma_{kk}, \lambda_{kk} \in \{0, 1\}, k, \hat{k} \in L_j, j \in B : k < \hat{k},$	(15)
$\tau_i \in \{0, 1\}, i \in T$	(16)

Figure 1. The mathematical model

The goals of the MIP model are represented by (1a) and (1b). The objective function (1a) is to minimise the sum of delays experienced by all trains when they reach their final destination. The objective function (1b) is aimed to maximise train service reliability. Constraint (2) governs the restrictions posed to commuter trains. Constraint (2a) indicates that a successor of a train event must wait until its predecessor has completed its journey before it can start. The minimum running time for each train event is guaranteed by Constraint (2b) while Constraint (2c) forces the disrupted event to resume journey once it recovers.

Constraints (3a) and (3b) control the strict schedule for the prioritised ETS train. Constraint (3a) ensures that each ETS train event must be directly succeeded by the next one, as far as the original schedule is concerned while Constraint (3b) guarantees that the ETS trains should strictly depart and arrive, according to the planned scheduled. Constraint (4) indicates that the reschedule departure time should never be earlier than the original time scheduled. Constraint

(5) and Constraint (6) ensure that events that have already started before disruption occurs must follow the original timetable. Constraint (7) defines the total delay of all trains as the deviation between the rescheduled and the original arrival times.

Constraint (8) restricts the utilisation of track line as one train per track. Constraint (9) is introduced to ensure that the total concurrent events at a segment must not exceed the track capacity. Constraint (10) checks the sequence between an event and its preceding event, so as to ensure that it is either γ_{kk} or λ_{kk} will take value of '1'. Constraints (11) and Constraints (12) impose a restriction for the minimum headway between two consecutive trains using the same track. It is either the set of Constraint (11) or Constraint (12) that will become active, depending on the value of γ_{kk} and λ_{kk} . In addition to this, the minimum headway H_j equals 5 minutes for both station and non-station segment j . The lateness of train i is denoted by a binary variable τ_i , over the delay time window of 5 minutes while M is an arbitrarily large positive integer. Finally, the Constraint (13) up to Constraint (16) define the domain of the decision variables.

THE COMPUTATIONAL EXPERIMENT

Computational experiments in this study are based on realistic cases drawn from a part of a local railway line in Malaysia. Sample data is composed of 23 segments, including 10 stations along the railway network connecting some major towns in Kuala Lumpur.

The experimental analysis is carried out by considering 3 incident scenarios involving a disruption on a track segment. For all cases, we consider a partial blockage, in which only one track of a segment is blocked. Segment 22 and Segment 23 have four rail tracks while the rest are equipped with a double track system. We specify beforehand that all trains going south and north will be using Track 1 and Track 2 respectively. Any additional tracks available can be used when needed. For partial blockade, which is left with only one single track, any two consecutive trains using the single piece of track need to adhere to the minimum headway of 5 minutes.

In rescheduling cases, the rail operator has established its priority settings in two aspects. First, in any conflict circumstances between types of train, a commuter must always give priority to ETS. The fast train is set to run according to the timetable as long as there is at least one track available. In this research, it is assumed that an ETS train will not experience any service delay or break down. Second, if two commuter trains of the same direction are anticipated to meet each other in less than the minimum headway requirement of 5 minutes, then the recovered disrupted train has to be given priority over the normal-scheduled train.

There are nine commuter trains running in both directions, including one ETS train heading south. Each train is assumed to be able to fit on any track and a maximum of six-car train is assumed. The location of all trains in the network is known at all times. For simplicity, the speed of trains is assumed constant and the dwell time of the trains at stations is embedded in the event duration.

There were 124 events and 1145 decision variables altogether, which technically produce a list of 2018 constraints. The time horizon was limited to 42 minutes and three incident cases were manually created to capture the rescheduling scenario. The cases were randomly

selected with the aim to get solutions which satisfy the track capacity and adhere to the minimum headway requirement. We use preemptive goal programming approach in which the total service delay is computed first and the result obtained will then be used to determine the maximum train service reliability. The service reliability is computed by only considering the trains with delay time window of five minutes or less. The mathematical model is solved by using MATLAB R2014a. The computational tests were run on a 3.00GHz AMD Phenom Processor with 4Gb RAM.

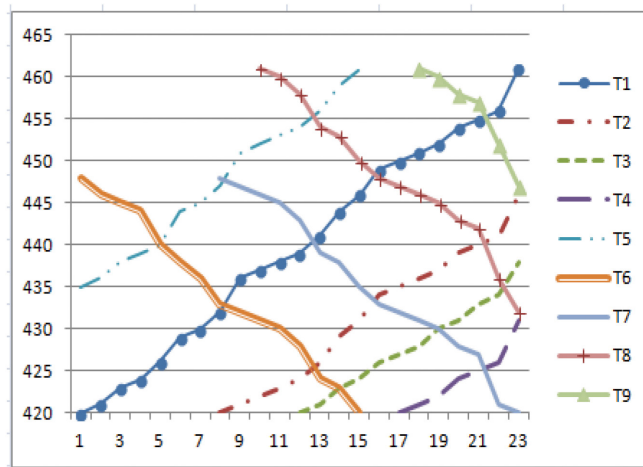


Figure 2. Original schedule

Figure 2 displays the original schedule of nine trains that run along the rail tracks. The vertical axis depicts the time and the horizontal axis indicates the segment. Each curve in the diagram represents a train, which is indicated by the number next to the line. Curves with positive gradient indicate trains heading south, while those with negative gradient indicate the opposite direction. The prioritised ETS train is denoted by Train 3.

Incident Case 1

The first scenario is a 10-minute disruption at a track segment 9 starting at $t=423$, experienced by Train 2. During the blockage period, Train 6 of the opposite direction passes the segment by using the other remaining track. When Train 2 resumes its journey at $t=433$, Train 1 is ahead running at the same direction. To maintain a safe minimum headway distance between them, the priority settings forced the normal-scheduled Train 1 to be delayed and give way to the recovered Train 2. The tiny dots along the TT_T1 and TT_T2 curves in Figure 3 represent the optimal reschedule plan generated by the optimisation model in contrast to the patterned curves which are the original timetable of the respective trains. The optimum solution obtained from the model shows that the minimum delays experienced by Train 1 and Train 2 are 2 minutes and 11 minutes respectively. As a consequence, this has brought the objective function of total delay z to be 13 minutes and its service reliability at 89% percent.

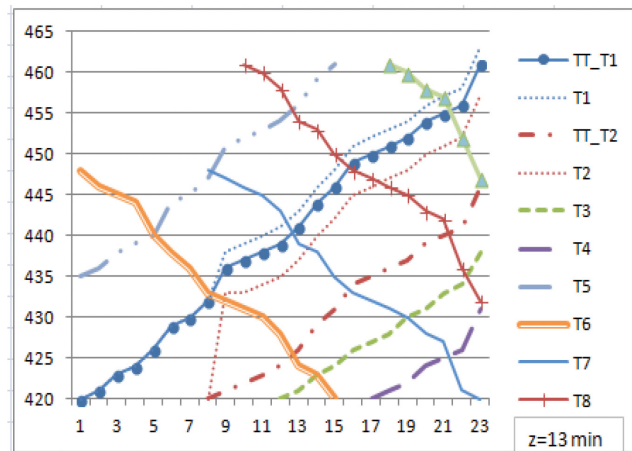


Figure 3. Rescheduling timetable for Case 1

Incident Case 2

In Case 2, a south-bound Train 4 experiences a 7-minute delay at Segment 17. Once recovered, it could not resume its journey because there is an ETS which has just passed by. Therefore, Train 4 is forced to wait for ETS to be 5 minute ahead, in compliance with the minimum headway constraint. When Train 4 is about to resume its journey at $t=435$, again it faces a conflict, now with a normal-scheduled Train 2 which is heading to the same direction. This time Train 4 is granted priority as a recovered disrupted train while Train 2 will be delayed as required by the minimum headway restriction.

On the opposite direction, Train 7 is also affected when there is a track capacity restriction at Segment 17 during which Train 4 was stranded. The tiny dotted lines along the TT_T2, TT_T4 and TT_T7 curves in Figure 4 illustrates the new reschedule plan for the affected trains. The result generated shows that the minimum delays incurred by Train 2, Train 4 and Train 7 are 5 minutes, 14 minutes and 1 minute respectively. To this end, the objective function yields the total delay z to be 20 minutes with a service reliability of 77%.

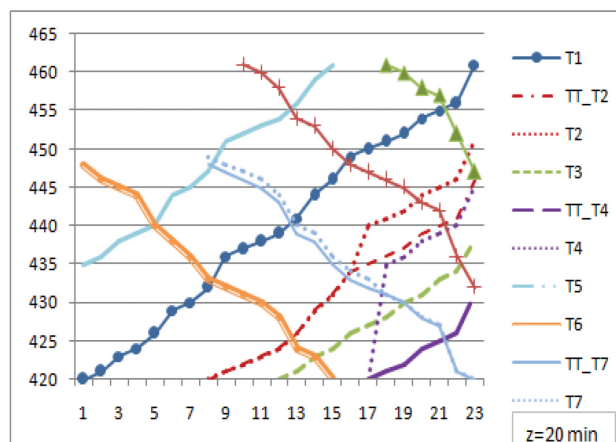


Figure 4. Rescheduling timetable for Case 2

Incident Case 3

In Case 3, Train 4 only experienced a 4-minute disruption at Segment 19. However, the total delay generated was 26 minutes, due to priority settings and capacity constraints that it has to satisfy. Having recovered from disruption, Train 4 has to wait for another 10 minutes before it can travel at a safe distance behind ETS. In the meantime, the opposite direction Train 7 is delayed to meet the capacity constraint at the affected segment. The tiny dotted lines emerging from the TT_T4 and TT_T7 curves in Figure 5 illustrates the new reschedule timetable. Due to the long total delay, service reliability has dropped to only 67%.

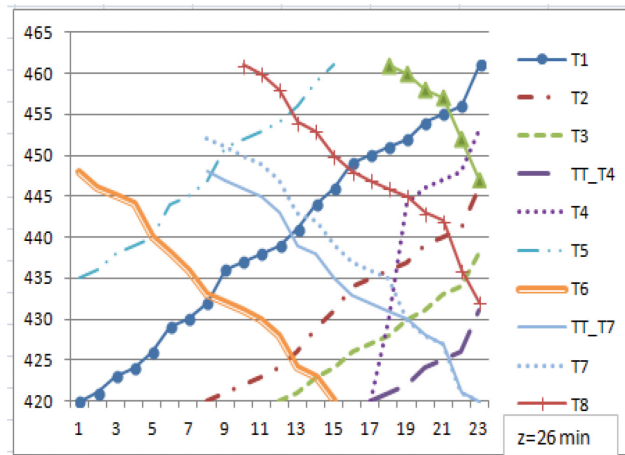


Figure 5. Rescheduling timetable for Case 3

CONCLUSION AND FURTHER RESEARCH

From the three cases that have been analysed, it has been shown that the MILP model successfully generated the rescheduling timetable. The total service delay of each affected train is attainable, together with the service reliabilities and the list of all departure and arrival times for each event. The new rescheduling timetable considers the aspects of track capacity and the minimum headway requirement practically, besides other model constraints. Ultimately, the fast train ETS runs according to the original schedule and does not experience any delay in all cases.

In spite of these results, some aspects have been disregarded in this experiment. The scope of research was limited. A more complex result might have been obtained if more segments were covered, more trains included and the time horizon lengthened. The possibility of one train overtaking the other is also neglected; while in practice, a train is allowed to overtake the damaged train at a segment loop. The model formulation also disregards the importance of connecting trains. In some railway system, a train is forced to wait for other trains to connect their passengers to their final destination.

ACKNOWLEDGEMENT

This study was partly financed by Universiti Teknologi MARA Malaysia which the authors gratefully acknowledge. This financial support together with the data provided by an anonymous railway company made this research possible. The authors record their appreciation to the latter.

REFERENCES

- Acuna-Agost, R., Michelon, P., Feillet, D., & Gueye, S. (2009, May). Constraint programming and mixed integer linear programming for rescheduling trains under disrupted operations. In *International Conference on AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems* (pp. 312-313). Springer Berlin Heidelberg.
- Caimi, G., Fuchsberger, M., Laumanns, M., & Lüthi, M. (2012). A model predictive control approach for discrete-time rescheduling in complex central railway station areas. *Computers and Operations Research*, 39(11), 2578-2593.
- Kang, L., Wu, J., Sun, H., Zhu, X., & Wang, B. (2015). A practical model for last train rescheduling with train delay in urban railway transit networks. *Omega*, 50, 29-42.
- Krasemann, J. T. (2012). Design of an effective algorithm for fast response to the re-scheduling of railway traffic during disturbances. *Transportation Research Part C: Emerging Technologies*, 20(1), 62-78.
- Narayanaswami, S., & Rangaraj, N. (2013). Modelling disruptions and resolving conflicts optimally in a railway schedule. *Computers and Industrial Engineering*, 64(1), 469-481.
- Schachtebeck, M., & Schöbel, A. (2008). IP-based techniques for delay management with priority decisions. In *OASICS-OpenAccess Series in Informatics* (Vol. 9). Schloss Dagstuhl-Leibniz-Zentrum für Informatik.
- Schobel, A. (2001). A model for the delay management problem based on Mixed Integer Programming. *Electronic Notes in Theoretical Computer Science*, 50(1), 1-10.
- Törnquist, J., & Persson, J. A. (2007). N-tracked railway traffic re-scheduling during disturbances. *Transportation Research Part B: Methodological*, 41(3), 342-362.
- Törnquist, J., & Persson, J. A. (2005). Train Traffic Deviation Handling Using Tabu Search and Simulated Annealing: A Swedish Case. In *HICSS-38, Hawaii International Conference on Systems Science*. Hawaii.
- Veelenturf, L. P., Kidd, M. P., Cacchiani, V., Kroon, L. G., & Toth, P. (2014). *A railway timetable rescheduling approach for handling large scale disruptions*. Rotterdam: Erasmus Research Institute of Management.
- Yang, L., Zhou, X., & Gao, Z. (2014). Credibility-based rescheduling model in a double-track railway network: A fuzzy reliable optimization approach. *Omega*, 48, 75-93.



REFEREES FOR THE PERTANIKA JOURNAL OF SCIENCE AND TECHNOLOGY

VOL. 25 (2) APR. 2017

The Editorial Board of the Journal of Science and Technology wishes to thank the following for acting as referees for manuscripts published in this issue of JST.

Abdul Razak Ismail (UTM, Malaysia)	M. Chithirai Pon Selvan (Amity University, India)	Ramakrishnan Ramachanra (Classic Consultants, India)
Aji Prasetya Wibawa (State University of Malang, Indonesia)	Mohd Zalisham Jali (USIM, Malaysia)	Razali Yaakob (UPM, Malaysia)
Aminuddin Ab. Ghani (USM, Malaysia)	Nafarizal Nayan (UTHM, Malaysia)	Rozi Abdullah (USM, Malaysia)
Arrifin Samsuri (UTM, Malaysia)	Nazatul Liana Sukiman (UM, Malaysia)	Sarinder Kaur Kashmir Singh (UM, Malaysia)
Binu K Gopalakrishna (SJEC, India)	Ng Choung Min (UM, Malaysia)	Seripah Awang (UiTM, Malaysia)
Biswajeet Pradhan (UPM, Malaysia)	Noor Azina Ismail (UM, Malaysia)	Siti Rahmah Awang (UTM, Malaysia)
Chua Han Bing (Curtin University, Malaysia)	Nor Ashidi Mat Isa (USM, Malaysia)	Tadashi Chosa (TUAT, Japan)
Cynthia Paul Sebli (UPM, Malaysia)	Noraini Abdullah (UMS, Malaysia)	Tay Chia Chay (UiTM, Malaysia)
Edriyana Abd. Aziz (UMP, Malaysia)	Norazan Mohamed Ramli (UiTM, Malaysia)	Wan Jeffrey Basirun (UM, Malaysia)
Ghazali Osman (UiTM, Malaysia)	Normala Halimoon (UPM, Malaysia)	Wan Nor Azmin Sulaiman (UPM, Malaysia)
Hatijah Basri (UTHM, Malaysia)	Norshahida Shaadan (UiTM, Malaysia)	Wan Ruslan Ismail (USM, Malaysia)
Hazrul Abdul Hamid (UiTM, Malaysia)	Ong Hong Choon (USM, Malaysia)	Wong Voon Hee (UTAR, Malaysia)
Hilary Green (Macquarie University, Australia)	Phang Yook Ngor (UiTM, Malaysia)	Zeinab Abbas Jawad (Curtin University, Malaysia)
Ida Idayu Muhamad (UTM, Malaysia)	Pradeep Kumar Gupta (UP, South Africa)	Zuhaimy Ismail (UTM, Malaysia)
Joseph Sahaya Anand T. (UTeM, Malaysia)	Pusparini Dewi Abd Aziz (UniKL, Malaysia)	

SJEC – St. Joseph Engineering College
TUAT – Tokyo University of Agriculture and Technology
UiTM – Universiti Teknologi MARA
UM – Universiti Malaya
UMP – Universiti Malaysia Pahang
UMS – Universiti Malaysia Sabah
UniKL – Universiti Kuala Lumpur

UP – University of Pretoria
USIM – Universiti Sains Islam Malaysia
USM – Universiti Sains Malaysia
UTAR – Universiti Tunku Abdul Rahman
UTeM – Universiti Teknikal Malaysia Melaka
UTHM – Universiti Tun Hussein Onn Malaysia
UTM – Universiti Teknologi, Malaysia

While every effort has been made to include a complete list of referees for the period stated above, however if any name(s) have been omitted unintentionally or spelt incorrectly, please notify the Chief Executive Editor, *Pertanika* Journals at nayan@upm.my.

Any inclusion or exclusion of name(s) on this page does not commit the *Pertanika* Editorial Office, nor the UPM Press or the University to provide any liability for whatsoever reason.



Pertanika Journals

Our goal is to bring high quality research to the widest possible audience

INSTRUCTIONS TO AUTHORS (Manuscript Preparation & Submission Guide)

Revised: June 2016

Please read the Pertanika guidelines and follow these instructions carefully. Manuscripts not adhering to the instructions will be returned for revision without review. The Chief Executive Editor reserves the right to return manuscripts that are not prepared in accordance with these guidelines.

MANUSCRIPT PREPARATION

Manuscript Types

Pertanika accepts submission of mainly **four** types of manuscripts for peer-review.

1. REGULAR ARTICLE

Regular articles are full-length original empirical investigations, consisting of introduction, materials and methods, results and discussion, conclusions. Original work must provide references and an explanation on research findings that contain new and significant findings.

Size: Generally, these are expected to be between 6 and 12 journal pages (excluding the abstract, references, tables and/or figures), a maximum of 80 references, and an abstract of 100–200 words.

2. REVIEW ARTICLE

These report critical evaluation of materials about current research that has already been published by organizing, integrating, and evaluating previously published materials. It summarizes the status of knowledge and outline future directions of research within the journal scope. Review articles should aim to provide systemic overviews, evaluations and interpretations of research in a given field. Re-analyses as meta-analysis and systemic reviews are encouraged. The manuscript title must start with "Review Article:".

Size: These articles do not have an expected page limit or maximum number of references, should include appropriate figures and/or tables, and an abstract of 100–200 words. Ideally, a review article should be of 7 to 8 printed pages.

3. SHORT COMMUNICATIONS

They are timely, peer-reviewed and brief. These are suitable for the publication of significant technical advances and may be used to:

- (a) report new developments, significant advances and novel aspects of experimental and theoretical methods and techniques which are relevant for scientific investigations within the journal scope;
- (b) report/discuss on significant matters of policy and perspective related to the science of the journal, including 'personal' commentary;
- (c) disseminate information and data on topical events of significant scientific and/or social interest within the scope of the journal.

The manuscript title must start with "*Brief Communication:*".

Size: These are usually between 2 and 4 journal pages and have a maximum of three figures and/or tables, from 8 to 20 references, and an abstract length not exceeding 100 words. Information must be in short but complete form and it is not intended to publish preliminary results or to be a reduced version of Regular or Rapid Papers.

4. OTHERS

Brief reports, case studies, comments, concept papers, Letters to the Editor, and replies on previously published articles may be considered.

PLEASE NOTE: NO EXCEPTIONS WILL BE MADE FOR PAGE LENGTH.

Language Accuracy

Pertanika **emphasizes** on the linguistic accuracy of every manuscript published. Articles must be in **English** and they must be competently written and argued in clear and concise grammatical English. Contributors are strongly advised to have the manuscript checked by a colleague with ample experience in writing English manuscripts or a competent English language editor.

Author(s) **must provide a certificate** confirming that their manuscripts have been adequately edited. A proof from a recognised editing service should be submitted together with the cover letter at the time of submitting a manuscript to Pertanika. **All editing costs must be borne by the author(s)**. This step, taken by authors before submission, will greatly facilitate reviewing, and thus publication if the content is acceptable.

Linguistically hopeless manuscripts will be rejected straightaway (e.g., when the language is so poor that one cannot be sure of what the authors really mean). This process, taken by authors before submission, will greatly facilitate reviewing, and thus publication if the content is acceptable.

MANUSCRIPT FORMAT

The paper should be submitted in one column format with at least 4cm margins and 1.5 line spacing throughout. Authors are advised to use Times New Roman 12-point font and *MS Word* format.

1. Manuscript Structure

Manuscripts in general should be organised in the following order:

Page 1: Running title

This page should **only** contain the running title of your paper. The running title is an abbreviated title used as the running head on every page of the manuscript. The running title should not exceed 60 characters, counting letters and spaces.

Page 2: Author(s) and Corresponding author information.

This page should contain the **full title** of your paper not exceeding 25 words, with name(s) of all the authors, institutions and corresponding author's name, institution and full address (Street address, telephone number (including extension), hand phone number, and e-mail address) for editorial correspondence. First and corresponding authors must be clearly indicated.

The names of the authors may be abbreviated following the international naming convention. e.g. Salleh, A.B.¹, Tan, S.G^{2*}., and Sapuan, S.M³.

Authors' addresses. Multiple authors with different addresses must indicate their respective addresses separately by superscript numbers:

George Swan¹ and Nayan Kanwal²

¹Department of Biology, Faculty of Science, Duke University, Durham, North Carolina, USA.,

²Office of the Deputy Vice Chancellor (R&I), Universiti Putra Malaysia, Serdang, Malaysia.

A **list** of number of **black and white / colour figures and tables** should also be indicated on this page. Figures submitted in color will be printed in colour. See "5. Figures & Photographs" for details.

Page 3: Abstract

This page should **repeat** the **full title** of your paper with only the **Abstract** (the abstract should be less than 250 words for a Regular Paper and up to 100 words for a Short Communication), and **Keywords**.

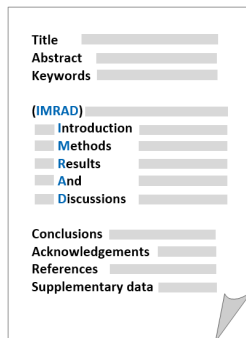
Keywords: Not more than eight keywords in alphabetical order must be provided to describe the contents of the manuscript.

Page 4: Introduction

This page should begin with the **Introduction** of your article and followed by the rest of your paper.

2. Text

Regular Papers should be prepared with the headings *Introduction, Materials and Methods, Results and Discussion, Conclusions, Acknowledgements, References, and Supplementary data* (if available) in this order.



Title _____
 Abstract _____
 Keywords _____
 (IMRAD)
 Introduction _____
 Methods _____
 Results _____
 And _____
 Discussions _____
 Conclusions _____
 Acknowledgements _____
 References _____
 Supplementary data _____

MAKE YOUR ARTICLES AS CONCISE AS POSSIBLE

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words Introduction, Materials and Methods, Results, And, Discussion. It indicates a pattern or format rather than a complete list of headings or components of research papers; the missing parts of a paper are: Title, Authors, Keywords, Abstract, Conclusions, and References. Additionally, some papers include Acknowledgments and Appendices.

The Introduction explains the scope and objective of the study in the light of current knowledge on the subject; the Materials and Methods describes how the study was conducted; the Results section reports what was found in the study; and the Discussion section explains meaning and significance of the results and provides suggestions for future directions of research. The manuscript must be prepared according to the Journal's instructions to authors.

3. Equations and Formulae

These must be set up clearly and should be typed double spaced. Numbers identifying equations should be in square brackets and placed on the right margin of the text.

4. Tables

All tables should be prepared in a form consistent with recent issues of Pertanika and should be numbered consecutively with Roman numerals. Explanatory material should be given in the table legends and footnotes. Each table should be prepared on a new page, embedded in the manuscript.

When a manuscript is submitted for publication, tables must also be submitted separately as data - .doc, .rtf, Excel or PowerPoint files- because tables submitted as image data cannot be edited for publication and are usually in low-resolution.

5. Figures & Photographs

Submit an **original** figure or photograph. Line drawings must be clear, with high black and white contrast. Each figure or photograph should be prepared on a new page, embedded in the manuscript for reviewing to keep the file of the manuscript under 5 MB. These should be numbered consecutively with Roman numerals.

Figures or photographs must also be submitted separately as TIFF, JPEG, or Excel files- because figures or photographs submitted in low-resolution embedded in the manuscript cannot be accepted for publication. For electronic figures, create your figures using applications that are capable of preparing high resolution TIFF files. In general, we require **300 dpi** or higher resolution for **coloured and half-tone artwork**, and **1200 dpi or higher** for **line drawings** are required.

Failure to comply with these specifications will require new figures and delay in publication.

NOTE: Illustrations may be produced in colour at no extra cost at the discretion of the Publisher; the author could be charged Malaysian Ringgit 50 for each colour page.

6. References

References begin on their own page and are listed in alphabetical order by the first author's last name. Only references cited within the text should be included. All references should be in 12-point font and double-spaced.

NOTE: When formatting your references, please follow the **APA reference style** (6th Edition). Ensure that the references are strictly in the journal's prescribed style, failing which your article will **not be accepted for peer-review**. You may refer to the *Publication Manual of the American Psychological Association* for further details (<http://www.apastyle.org/>).

7. General Guidelines

Abbreviations: Define alphabetically, other than abbreviations that can be used without definition. Words or phrases that are abbreviated in the introduction and following text should be written out in full the first time that they appear in the text, with each abbreviated form in parenthesis. Include the common name or scientific name, or both, of animal and plant materials.

Acknowledgements: Individuals and entities that have provided essential support such as research grants and fellowships and other sources of funding should be acknowledged. Contributions that do not involve researching (clerical assistance or personal acknowledgements) should **not** appear in acknowledgements.

Authors' Affiliation: The primary affiliation for each author should be the institution where the majority of their work was done. If an author has subsequently moved to another institution, the current address may also be stated in the footer.

Co-Authors: The commonly accepted guideline for authorship is that one must have substantially contributed to the development of the paper and share accountability for the results. Researchers should decide who will be an author and what order they will be listed depending upon their order of importance to the study. Other contributions should be cited in the manuscript's Acknowledgements.

Copyright Permissions: Authors should seek necessary permissions for quotations, artwork, boxes or tables taken from other publications or from other freely available sources on the Internet before submission to Pertanika. Acknowledgement must be given to the original source in the illustration legend, in a table footnote, or at the end of the quotation.

Footnotes: Current addresses of authors if different from heading may be inserted here.

Page Numbering: Every page of the manuscript, including the title page, references, tables, etc. should be numbered.

Spelling: The journal uses American or British spelling and authors may follow the latest edition of the Oxford Advanced Learner's Dictionary for British spellings.

SUBMISSION OF MANUSCRIPTS

Owing to the volume of manuscripts we receive, we must insist that all submissions be made electronically using the **online submission system ScholarOne™**, a web-based portal by Thomson Reuters. For more information, go to our web page and [click "Online Submission"](#).

Submission Checklist

1. **MANUSCRIPT:** Ensure your MS has followed the Pertanika style particularly the first four pages as explained earlier. The article should be written in a good academic style and provide an accurate and succinct description of the contents ensuring that grammar and spelling errors have been corrected before submission. It should also not exceed the suggested length.

COVER LETTER: All submissions must be accompanied by a cover letter detailing what you are submitting. Papers are accepted for publication in the journal on the understanding that the article is **original** and the content has **not been published** either **in English** or **any other language(s)** or **submitted for publication elsewhere**. The letter should also briefly describe the research you are reporting, why it is important, and why you think the readers of the journal would be interested in it. The cover letter must also contain an acknowledgement that all authors have contributed significantly, and that all authors have approved the paper for release and are in agreement with its content.

The cover letter of the paper should contain (i) the title; (ii) the full names of the authors; (iii) the addresses of the institutions at which the work was carried out together with (iv) the full postal and email address, plus telephone numbers and emails of all the authors. The current address of any author, if different from that where the work was carried out, should be supplied in a footnote.

The above must be stated in the cover letter. Submission of your manuscript will not be accepted until a cover letter has been received

2. **COPYRIGHT:** Authors publishing the Journal will be asked to sign a copyright form. In signing the form, it is assumed that authors have obtained permission to use any copyrighted or previously published material. All authors must read and agree to the conditions outlined in the form, and must sign the form or agree that the corresponding author can sign on their behalf. Articles cannot be published until a signed form (*original pen-to-paper signature*) has been received.

Please do **not** submit manuscripts to the editor-in-chief or to any other office directly. Any queries must be directed to the **Chief Executive Editor's** office via email to nayan@upm.my.

Visit our Journal's website for more details at <http://www.pertanika.upm.edu.my/home.php>.

HARDCOPIES OF THE JOURNALS AND OFF PRINTS

Under the Journal's open access initiative, authors can choose to download free material (via PDF link) from any of the journal issues from Pertanika's website. Under "**Browse Journals**" you will see a link, "*Current Issues*" or "*Archives*". Here you will get access to all current and back-issues from 1978 onwards.

The **corresponding author** for all articles will receive one complimentary hardcopy of the journal in which his/her articles is published. In addition, 20 off prints of the full text of their article will also be provided. Additional copies of the journals may be purchased by writing to the Chief Executive Editor.



Why should you publish in

Pertanika?

BENEFITS TO AUTHORS

PROFILE: Our journals are circulated in large numbers all over Malaysia, and beyond in Southeast Asia. Our circulation covers other overseas countries as well. We ensure that your work reaches the widest possible audience in print and online, through our wide publicity campaigns held frequently, and through our constantly developing electronic initiatives such as Web of Science Author Connect backed by Thomson Reuters.

QUALITY: Our journals' reputation for quality is unsurpassed ensuring that the originality, authority and accuracy of your work are fully recognised. Each manuscript submitted to Pertanika undergoes a rigid originality check. Our double-blind peer refereeing procedures are fair and open, and we aim to help authors develop and improve their scientific work. Pertanika is now over 38 years old; this accumulated knowledge has resulted in our journals being indexed in SCOPUS (Elsevier), Thomson (ISI) Web of Science™ Core Collection, Emerging Sources Citation Index (ESCI), Web of Knowledge [BIOSIS & CAB Abstracts], EBSCO, DOAJ, ERA, AGRICOLA, Google Scholar, ISC, TIB, Journal Guide, Citefactor, Cabell's Directories and MyCite.

AUTHOR SERVICES: We provide a rapid response service to all our authors, with dedicated support staff for each journal, and a point of contact throughout the refereeing and production processes. Our aim is to ensure that the production process is as smooth as possible, is borne out by the high number of authors who prefer to publish with us.

CODE OF ETHICS: Our Journal has adopted a Code of Ethics to ensure that its commitment to integrity is recognized and adhered to by contributors, editors and reviewers. It warns against plagiarism and self-plagiarism, and provides guidelines on authorship, copyright and submission, among others.

PRESS RELEASES: Landmark academic papers that are published in Pertanika journals are converted into press-releases as a unique strategy for increasing visibility of the journal as well as to make major findings accessible to non-specialist readers. These press releases are then featured in the university's UK and Australian based research portal, ResearchSEA, for the perusal of journalists all over the world.

LAG TIME: The elapsed time from submission to publication for the articles averages 3 to 4 months. A decision on acceptance of a manuscript is reached in 3 to 4 months (average 14 weeks).



Address your submissions to:
The Chief Executive Editor
Tel: +603 8947 1622
nayan@upm.my

Journal's Profile: www.pertanika.upm.edu.my/

Call for Papers 2017-18

now accepting submissions...

Pertanika invites you to explore frontiers from all key areas of agriculture, science and technology to social sciences and humanities.

Original research and review articles are invited from scholars, scientists, professors, post-docs, and university students who are seeking publishing opportunities for their research papers through the Journal's three titles; JTAS, JST & JSSH. Preference is given to the work on leading and innovative research approaches.

Pertanika is a fast track peer-reviewed and open-access academic journal published by Universiti Putra Malaysia. To date, Pertanika Journals have been indexed by many important databases. Authors may contribute their scientific work by publishing in UPM's hallmark SCOPUS & ISI indexed journals.

Our journals are open access - international journals. Researchers worldwide will have full access to all the articles published online and be able to download them with zero subscription fee.

Pertanika uses online article submission, review and tracking system for quality and quick review processing backed by Thomson Reuter's ScholarOne™. Journals provide rapid publication of research articles through this system.

For details on the Guide to Online Submissions, please visit http://www.pertanika.upm.edu.my/guide_online_submission.php

About the Journal

Pertanika is an international multidisciplinary peer-reviewed leading journal in Malaysia which began publication in 1978. The journal publishes in three different areas — Journal of Tropical Agricultural Science (JTAS); Journal of Science and Technology (JST); and Journal of Social Sciences and Humanities (JSSH). All journals are published in English.

JTAS is devoted to the publication of original papers that serves as a forum for practical approaches to improving quality in issues pertaining to tropical agricultural research- or related fields of study. It is published four times a year in *February, May, August* and *November*.

JST caters for science and engineering research- or related fields of study. It is published twice a year in *January* and *July*.

JSSH deals in research or theories in social sciences and humanities research. It aims to develop as a flagship journal with a focus on emerging issues pertaining to the social and behavioural sciences as well as the humanities, particularly in the Asia Pacific region. It is published four times a year in *March, June, September* and *December*.



An Award-winning
International-Malaysian Journal
— CREAM AWARD, MoHE
—Sept 2015



An Approach for Identification of Copy-Move Image Forgery based on Projection Profiling	507
<i>Mohd Dilshad Ansari, Satya Prakash Ghrera1 and Mohd Wajid</i>	
Novel Impedance Measurement Technique for Soluble Solid Content Determination of Banana	519
<i>Ibrahim, N. U. A., Abd Aziz, S. and Nawi, N. M.</i>	
Phase Transition and Dielectric Properties of $0.9\text{Pb}(\text{Fe}_{1/2}\text{Nb}_{1/2})\text{O}_3-0.1\text{PbTiO}_3$ Modified with Nano ZnO	527
<i>Hassakorn Wattanasarn, Wattana Photankham, Peerapat Pattumma and Rattikorn Yimnirun</i>	
Selected Papers from the 2nd International Conference on Statistics in Science, Business and Engineering (ICSSBE 2015)	
Guest Editors: Yap Bee Wah & Sayang Mohd Deni	
Some New Bivariate Regression Models	537
<i>Faroughi, P. and Ismail, N.</i>	
Identifying the Uncertain Model Parameter of a Steam Turbine System	545
<i>Wan Munirah, W. M., Tahir, A. and Azmirul, A.</i>	
Decision Making Procedure Based on Jaccard Similarity Measure with Z-numbers	561
<i>Mohamad, D. and Ibrahim, S. Z.</i>	
Sample Size and Non-Normality Effects on Goodness of Fit Measures in Structural Equation Models	575
<i>Ainur, A. K., Sayang, M. D., Jannoo, Z. and Yap, B. W.</i>	
A Method for Performing Short Time Series Prediction	587
<i>Pooi, A. H.</i>	
A Goal Programming Model for Portfolio Optimisation Problem in Fuzzy Environment	593
<i>Mokhtar, M., Shuib, A. and Mohamad, D.</i>	
Performance of Variance Targeting Estimator (VTE) under Misspecified Error Distribution Assumption	607
<i>Abdul Rahim, M. A., Zahari, S. M. and Shariff, S. S. R.</i>	
Review of Context-Based Similarity Measure for Categorical Data	619
<i>Nurul Adzlyana, M. S., Rosma, M. D. and Nurazzah, A. R.</i>	
Assessing Stock Market Volatility for Different Sectors in Malaysia	631
<i>Shakila, S., Noryati, A. and Maheran, M. J.</i>	
A Railway Rescheduling Model with Priority Setting	649
<i>Shuib, A. and Alwadood, Z.</i>	

Contents

Foreword	i
<i>Nayan Deep S. Kanwal</i>	
Review Article	
Hydrocarbon Sources for the Carbon Nanotubes Production by Chemical Vapour Deposition: A Review	379
<i>Hayder Baqer Abdullah, Irmawati Ramli, Ismayadi Ismail and Nor Azah Yusof</i>	
Regular Articles	
An Investigation on High Temperature Erosion Behaviour of Plasma Sprayed CoCrAlY/Al ₂ O ₃ /YSZ on Fe and Ni Based Alloys	397
<i>Nithin H. S., Vijay Desai and M. R. Ramesh</i>	
The Use of Infiltration Wells to Reduce the Impacts of Land Use Changes on Flood Peaks: An Indonesian Catchment Case Study	407
<i>Kusumastuti, D. I., Jokowinarno, D., Khotimah, S. N. and Dewi, C.</i>	
A Hybridised Intelligent Technique for the Diagnosis of Medical Diseases	425
<i>Abdu Masanawa Sagir and Saratha Sathasivam</i>	
Ability of Ipomoea aquatica Forssk. to Remediate Phenol in Water and Effects of Phenol on the Plant's Growth	441
<i>Siew-Yi Lee, Siti Aqlima Ahmad, Siti Roslina Mustapha and Janna Ong-Abdullah</i>	
Robust Artificial Bee Colony in the Hopfield Network for 2-Satisfiability Problem	453
<i>Mohd. Shareduwan Mohd. Kasihmuddin, Mohd. Asyraf Mansor and Saratha Sathasivam</i>	
Estimating Lung Cancer Deaths in Thailand Based on Verbal Autopsy Study in 2005	469
<i>Nattakit Pipatjaturon, Phattrawan Tongkumchum and Attachai Ueranantasun</i>	
Rainfall Trends in the Niger-South Basin, Nigeria, 1948-2008	479
<i>Oloruntade, A. J., Mohammad, T. A. and Aimrun, W.</i>	
Numerical Analysis of The Effect of Nozzle Geometry on Flow Parameters in Abrasive Water Jet Machines	497
<i>Deepak, D., Jodel, A. Q., Cornelio, Midhun Abraham, M. and Shiva Prasad, U.</i>	



Pertanika Editorial Office, Journal Division
Office of the Deputy Vice Chancellor (R&I),
1st Floor, IDEA Tower II,
UPM-MTDC Technology Centre
Universiti Putra Malaysia
43400 UPM Serdang
Selangor Darul Ehsan
Malaysia

<http://www.pertanika.upm.edu.my/>
E-mail: executive_editor.pertanika@upm.my
Tel: +603 8947 1622/1620

PENERBIT
UPM
UNIVERSITI PUTRA MALAYSIA
PRESS

<http://penerbit.upm.edu.my>
E-mail : penerbit@putra.upm.edu.my
Tel : +603 8946 8855/8854
Fax : +603 8941 6172

