Pertanika Journal of

SCIENCE & TECHNOLOGY

JST

VOL. 26 (1) JAN. 2018

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 Methods, **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.

Journal of Science & Technology
Journal of Science & Technology

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 Edito-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.

JST
Journal of Science & Technology
AN INTERNATIONAL PEER-REVIEWED JOURNAL

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 Serdang, Selangor Malaysia.
Gen Enq.: +603 8947 1622 | 1616
E-mail: executive_editor.pertanika@upm.my
URL: www.journals-jd.upm.edu.my

PUBLISHER
Kamariah Mohd Saidin
UPM Press
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia.
Tel: +603 8946 8855, 8946 8854
Fax: +603 8941 6172
E-mail: penerbit@putra.upm.edu.my
URL: http://penerbit.upm.edu.my

ABSTRACTING/INDEXING

Pertanika is now over 36 years old; this accumulated knowledge has resulted the journals being indexed in SCOPUS (Elsevier), Thomson (ISI) Web of Knowledge [BIOSIS & CAB Abstracts], EBSCO, DOAJ, Google Scholar, AGRICOLA, ISC, Citefactor, Rubriq and MyAIS. JST is also indexed in ERA.

Pertanika Journal of Science & Technology
Vol. 26 (1) Jan. 2018

Contents

Foreword

Welcome to the **First Issue 2018** of the Journal of Science and Technology (JST)!

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

This issue contains **30 articles,** of which **four** are review articles and **13** are regular research articles. This issue also features **eight** selected papers from the 2nd International Conference on Statistics in Science, Business and Engineering (ICSSBE 2015) and **five** papers from the Annual Applied Science and Engineering Conference 2016 (AASEC2016). The authors of these articles hail from several countries namely, **Malaysia**, **Nigeria**, **Australia**, **India**, **China**, **Iran**, **Indonesia**, **Pakistan** and **Thailand**.

The first review article in this issue reports briefly on bioremediation of petroleum hydrocarbon in Antarctica by microbial species (*Syahir Habib, Siti Aqlima Ahmad, Wan Lutfi Wan Johari, Mohd Yunus Abd Shukor* and *Nur Adeela Yasid*), while the second is on improving cost and time control in construction using a Building Information Model (BIM) (*M. M. Tahir, Haron, N. A., Alias, A. H., Harun, A. N., I. B. Muhammad* and *D. L. Baba*). The next review article looks at Android botnets as a serious threat to Android devices (*Shahid Anwar, Mohamad Fadli Zokipli, Zakira Inayat, Julius Odili, Mushtaq Ali* and *Jasni Mohamad Zain*), while the final review article provides an update on Type 1 diabetes treatments, namely, insulin treatment, cell therapy and transplantation (*Homayoun Hani, Mohd-Azmi Mohd-Lila, Rasedee Abdullah, Zeenathul Nazariah Allaudin, Kazhal Sarsaifi* and *Faez Firdaus Jesse Abdullah*).

The regular articles cover a wide range of topics. The first article is on the simulation of a force in micro-scale sensing employing an optical double-ring resonator system (*Youplao, P., Tasakorn, M.* and *Phattaraworamet, T.*). The following articles look at: Distributions, composition patterns, sources and potential toxicity of polycyclic aromatic hydrocarbon (PAH) pollution in surface sediments from the Kim Kim River and Segget River, Peninsular Malaysia (*Mehrzad Keshavarzifard, Mohamad Pauzi Zakaria, Shahin Keshavarzifard* and *Reza Sharifi*); Antiulcer properties of kelulut honey against ethanol-induced gastric ulcer (*Latifah Saiful Yazan, Nurul Amira Zainal, Razana Mohd Ali, Muhamad Firdaus Shyfiq Muhamad Zali, Ong Yong Sze, Tor Yin Sim, Banulata Gopalsamy, Voon Fui Ling, Sarah Sapuan, Nurulaidah Esa, Aminah Suhaila Haron, Fatin Hannani Zakarial Ansar, Ana Masara Ahmad Mokhtar* and *Sharifah Sakinah Syed Alwi*); Revocable and non-invertible multibiometric template protection based on matrix transformation (*Jegede, A., Udzir, N. I., Abdullah, A.* and *Mahmod, R.*); The effects of alkali treatment on the mechanical and chemical properties of banana fibre and adhesion to epoxy resin (*Zin, M. H., Abdan, K., Norizan, M. N.* and *Mazlan, N.*); the use of environmentally friendly bio-oil in the

production of phenol formaldehyde (PF) resin (*Naja Nadiera Omar, Iskandar Shahrim Mustafa, Nurhayati Abdullah* and *Rokiah Hashim*); High intrinsic biosorption efficiency of cattle manure on Cr(VI) as a potential low-cost fibre-rich biosorbent (*Yap, K. L., Lee, C. M., Gan, Y. L., Tang, T. K., Lee, Y. Y., Tee, T. P.* and *Lai, O. M.*); Building a low cost, good quality and safe infinity mirror room for the Suroboyo night carnival amusement park (*Rinda Hedwig, Andrew Nafarin, Ichwanhono Kosasih, Jimmy Linggarjati* and *Wiedjaja Atmadja*); The value of 18F-fluorodeoxyglucose PET (18F-FDG PET) and MRI spectroscopy in underpinning suspicious breast cancer (*Shazreen, S., Shakher, R., Shahrun Niza, A. S.* and *Fathinul Fikri, A. S.*); Vertically integrated moisture flux convergence over Southeast Asia and its relation to rainfall over Thailand (*Kosum Chansaengkrachang, Anirut Luadsong* and *Nitima Aschariyaphotha*); New smoothed location models integrated with PCA and two types of MCA for handling a large number of mixed continuous and binary variables (*Hamid, H., Ngu, P. A. H.* and *Alipiah, F. M.*); The effect of distilled and sea water absorption on mechanical behaviour of short coir fibre epoxy composite/sawdust filler (*Moorthy M Nair, Nagaraja Shetty, Divakara Shetty S.* and *Shambhavi Kamath M*); and Green wall for retention of stormwater (*J. T. Lau* and *D. Y. S. Mah*).

I conclude this issue with 13 articles arising from selected international conferences featuring the following: An investigation into the randomness and duration of PM10 pollution using functional data analysis (*Shaadan, N., Deni, S. M.* and *Jemain, A. A.*); Application of active contours driven by local Gaussian distribution fitting energy to computed tomography images (*Nurwahidah, M., Wan, E. Z. W. A. R.* and *Shaharuddin, C. S.*); A general formula for calculating accumulated amount based on the average lowest balance concept (*Wahab, Z. A., Embong, R., Azmi, A.* and *Isa, N. B. M.*); Feature selection methods, with a focus on the filter and wrapper approaches for maximising classification accuracy (*Yap Bee Wah, Nurain Ibrahim, Hamzah Abdul Hamid, Shuzlina Abdul-Rahman* and *Simon Fong*); The neurocomputing approach for firearm identification (*Nor Azura Md Ghani, Choog-Yeun Liong* and *Abdul Aziz Jemain*); Consolidated backpropagation neural network for Malaysian construction cost indices data with the problem of outliers (*Saadi Ahmad Kamaruddin, Nor Azura Md Ghani* and *Norazan Mohamed Ramli*); Application of the first order of the Markov Chain Model in describing the PM10 occurrences in Shah Alam and Jerantut, Malaysia (*Mohamad, N. S., Deni, S. M.* and *Ul-Saufie, A. Z.*); Application of the system dynamic approach for a family takaful product analysis (*Mohamad, A. I., Tumin, M. H., Noor, N. L. M., Saman, F. M.* and *Amin, M. N. M.*); Monodispersed and size-controllable potassium silicate nanoparticles from rice straw waste produced using a flame-assisted spray pyrolysis (*Asep Bayu Dani Nandiyanto, Rena Zaen, Rosi Oktiani, Ade Gafar Abdullah* and *Ari Arifin Danuwijaya*); Morphodynamics of coastal lagoons, featuring an analysis of multitemporal landsat images of the Segara Anakan lagoon area (*Nandi*); Multi-Hop wireless sensor network performance and energy simulation (*Lilik Hasanah, Heru Yuwono, Ahmad Aminudin, Endi Suhendi, Yuyu Rachmat Tayubi* and *Khairurrijal*); Parallel exponential smoothing using the bootstrap method in R for

forecasting orbital elements of asteroids (*Lala Septem Riza, Judhistira Aria Utama, Syandi Mufti Putra, Ferry Mukharradi Simatupang* and *Eddy Prasetyo Nugroho*); and omnidirectional MIMO antenna with collinear array for LTE applications (*Enjang Ahmad Juanda, Tommi Hariyadi, Abdul Aziz Reguna* and *Arjuni Budi Pantjawati*).

I anticipate that you will find the evidence presented in this issue to be intriguing, thought-provoking and useful in setting new milestones. Please recommend the journal to your colleagues and students to make this endeavour meaningful.

I would also like to express my gratitude to all the contributors, namely, the authors, reviewers and editors for their professional contribution towards making this issue feasible. Last but not least, the editorial assistance of the Journal Division staff is fully appreciated.

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

Pertanika J. Sci. & Technol. 26 (1): 1 - 20 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Bioremediation of Petroleum Hydrocarbon in Antarctica by Microbial Species: An Overview

Syahir Habib[1], Siti Aqlima Ahmad[1], Wan Lutfi Wan Johari[2], Mohd Yunus Abd Shukor[1] and Nur Adeela Yasid[1]*

[1]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Department of Environmental Sciences, Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

The increase of anthropogenic activities and growth of technology in Antarctica is fuelled by the high demand for petroleum hydrocarbons needed for daily activities. Oil and fuel spills that occur during explorations have caused hydrocarbon pollution in this region, prompting concern for the environment by polar communities and the larger world community. Crude oil and petroleum hydrocarbon products contain a wide variety of lethal components with high toxicity and low biodegradability. Hydrocarbon persistence in the Antarctic environment only worsens the issues stemming from environmental pollution as they can be long-term. Numerous efforts to lower the contamination level caused by these pollutants have been conducted mainly in bioremediation, an economical and degrading-wise method. Bioremediation mainly functions on conversion of complex toxic compounds to simpler organic compounds due to the consumption of hydrocarbons by microorganisms as their energy source. This review presents a summary of the collective understanding on bioremediation of petroleum hydrocarbons by microorganisms indigenous to the Antarctic region from past decades to current knowledge.

Keywords: Antarctic region, bioremediation, fuel spills, indigenous microorganisms, petroleum hydrocarbons

ARTICLE INFO

Article history:
Received: 01 August 2016
Accepted: 17 April 2017

E-mail addresses:
syahirhabib@gmail.com (Syahir Habib),
aqlima@upm.edu.my (Siti Aqlima Ahmad),
lutfi@env.upm.edu.my (Wan Lutfi Wan Johari),
mohdyunus@upm.edu.my (Mohd Yunus Abd Shukor),
adeela@upm.edu.my (Nur Adeela Yasid)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Accidental spills of oils and fuels in Antarctica is particularly sensitive as the environment is near pristine, causing pollution and its effects to be more apparent. Cold temperatures also

cause the natural processes that help remove contamination in other parts of the world to happen far more slowly, making it possible for contaminants to build up (AMAP 1998; Det Norske Veritas, 2003). For instance, hydrocarbon compounds can take decades to completely mineralise in this region, while only months are needed for compounds to be removed in temperate regions. Common spills are typically of kerosene and the Special Antarctic Blend diesel. Fuel spills have been acknowledged as being the most common source of pollution; thus have a great potential to cause the greatest environmental harm in and around the continent. Such spills have been reported to occur mainly near the Antarctic research stations and military bases, mainly due to the poor management or regulation of such pollution (Aislabie, Balks, Foght, & Waterhouse, 2004). Fortunately, due to the Antarctic Treaty prohibition on oil exploration and exploitation (Rothwell & Davis, 1997), petroleum hydrocarbon contamination manages to remain at controlled levels to a certain extent compared with other cold regions.

Attempts to clean up Antarctic polluted sites have been done using both physical and chemical methods. However, the estimated budget for the application of physical machinery would be large due to the harsh environment of the Antarctic, while chemical application should be critically considered as the introduction of chemicals may add to the risks of environmental pollution. Bioremediation, which basically manipulates the utilisation of hydrocarbon by microbial species as their energy source, is highly recommended (Das & Chandran, 2011; Jesus, Peixoto, & Rosado, 2015). Successful attempts of Antarctic bioremediation have been widely reported and reviewed several times to date (Azubuike, Chikere, & Okpokwasili, 2016; McWatters et al., 2016). However, the effectiveness of degradation of hydrocarbons by native microorganisms has been reported in laboratory studies but not on a pilot scale.

In this paper, we review the composition of petroleum hydrocarbons and their toxicity; the physical management of oil spill site; hydrocarbon degradation by microbial species; factors that affect the rate of biodegradation; and new bioremediation prospects in Antarctica. The purpose of this review is to fill the gap in the literature in order to direct relevant scientific research on areas to be explored in the future.

PETROLEUM HYDROCARBONS – COMPOSITION, FATE AND TOXICITY

Petroleum is a complex blend consisting of both aliphatic and aromatic hydrocarbons. Small amounts of non-hydrocarbons such as resin and asphalt can also be found in petroleum mixture. Crude oils can be divided into paraffinic, asphaltic and mixed crude oils (WHO, 1982). Paraffins (aliphatic hydrocarbons), wax and high-quality oils belong to the paraffinic crude oils group. Kerosene and naphtha are the lightest among paraffinic fractions. High viscous lubricating oils and cycloaliphatics can be grouped into asphaltic crude oils. These petroleum solvents are the product of crude oil distillation and are generally characterised by boiling point ranges. Lubricants, greases and waxes are high boiling-point portions of crude oils, whereas the heaviest solid fractions of crude are the residuals or bitumen (ATSDR, 1999).

Petroleum hydrocarbons are extensively used in the Antarctic as a source for heating, transportation and generating electricity (Aislabie, Saul, & Foght, 2006). Because of the broad usage of hydrocarbons, contamination of these pollutants is likely to occur (Mohn & Stewart, 2000; Ruberto, Vazquez, Lo Baldo, & Mac Cormack, 2005). As crude oils are made up of many

such components, they have the potential to become environmental contaminants, depending on the nature of release and fate of hydrocarbons in the environment. The structure and components of hydrocarbons do change and transform physically, chemically and biologically directly after crude oil is released into the environment, an action that significantly affects their potential impacts. The succession of transformations may be due to evaporation, dissolution, emulsification, dispersion, adsorption, sedimentation, biodegradation and photo-oxidation, with carbon dioxide and water as the ultimate products (Montagnolli, Lopes, & Bidoia, 2015).

The degradation of hydrocarbons in the Antarctic environment can be divided into two different processes, which are the soil systems and the aqueous systems. According to Aislabie et al. (2004), hydrocarbon spills on Antarctic soils generally undergo naturally occurring processes that decrease the mass of the pollutants. Although mechanisms such as dispersion, volatilisation and microbial degradation typically occur at most spill sites, degrees of the hydrocarbon mass reduction depend on the type and concentration of the spilled fuels and also the soil characteristics itself. Hydrocarbons with high vapour pressure such as jet fuel and MoGas tend to volatilise quickly after spills in Antarctic soil but due to their low viscosity, they are mobile and are able to migrate down through the unfrozen soil active layer (Webster, Webster, Nelson, & Waterhouse, 2003). Meanwhile, there is a limit to the downward movement due to the ice-saturated layer, causing an emergence of oil pools of the less volatile and more viscous engine and lubricating oils (Chuvilin, Naletova, Miklyaeva, Kozlova, & Instanes, 2001a). The freeze-thaw cycle may also influence the hydrocarbon movement in Antarctic soils. It is based on the implication of the melting of the active frozen layer and snow, which make the mobilisation of hydrocarbon compounds in downward movements easier (Chuvilin et al., 2001a). Contamination of the offshore marine environment due to surface runoff has also been reported (Kennicutt, McDonald, Denoux, & McDonald, 1992).

The fate of crude oil in marine ecosystems does not deviate much from oil degradation in soil ecosystems as both ultimately undergo total biodegradation by microbial species (Das & Chandran, 2011). In marine ecosystems, oil is likely to spread on the surface water when spilled and almost directly subjected to many modifications called the weathering process. The weathering process of oil is mainly due to the evaporation of low molecular weight hydrocarbons, dissolution of water soluble components, emulsification of oil droplets, photo-oxidation and lastly, biodegradation (Harayama, Kishira, Kasai, & Shutsubo, 1999; National Research Council, 2003). The succession of the reduction of oil pollutants in marine systems can be stated in the following order: evaporation of low boiling point *n*-alkanes and aromatic compounds > photochemical modification under sunlight (during summer season) > dissolution and dispersion of oil droplets by wave action > emulsification of the oil 'chocolate mousse' > formation of tar balls from petroleum heavy residues > sinking and degradation of tar balls by microbial species at the sediments of the sea (Harayama et al., 1999).

Wang and Bartha (1990) stated that the contamination can cause risks for humans and other living organisms if spilled fuel reaches groundwater reservoirs and water bodies. Compounds derived from petroleum tend to be toxic to a small population of animals and plants in the Antarctic region. Moreover, due to their persistence in the cold environment, they have a long-term carcinogenic potential that could affect this populace (Montagnolli et al., 2015). Long-term toxicity effects associated with petroleum substances represent a major concern. Different types

of hydrocarbon differ in their toxicity upon the local habitat. Generally, the longer the carbon chain and the higher the number of benzene rings possessed by hydrocarbons, the higher the toxicity (Montagnolli et al., 2015). The order of toxicity also tends to correspond to hydrocarbon susceptibility to microbial degradation. The order of hydrocarbon toxicity/susceptibility to microbial degradation can be arranged in decreasing order as follows: polycyclic aromatics hydrocarbons (PAHs) > benzene > heavy oil > light oil > aromatic hydrocarbons > alkanes (Atlas & Bragg, 2009; Paixão et al., 2007).

Polycyclic aromatic hydrocarbons (PAHs), the higher molecular weight aromatic hydrocarbons, are the most toxic components in a hydrocarbon mixture. The acute effects of PAHs on human health will vary due to the degree of exposure, PAHs quantity during exposure, the level of toxicity of PAHs and the route of exposure (ACGIH, 2005). PAHs have also been reported to cause skin irritation and inflammation, nausea and vomiting in low concentrations (Kim et al., 2013). Long-term effects are more severe, such as the production of cancer cells in the vital organs, gene mutation and cardiopulmonary mortality (Abdel-Shafy & Mansour, 2016; Armstrong, Hutchinson, Unwin, & Fletcher, 2004; Kuo, Hsu, & Lee, 2003). Although the probability of health effects due to PAHs is small and remains understudied in the Antarctic environment, deterioration of health can be damaging as the chronic effect of hydrocarbons is more prone in cold temperature. The toxic effects of petroleum hydrocarbons (including in temperate region) on humans and other living organisms are listed in Table 1 and 2.

Table 1
Toxic effects of different types of hydrocarbon on animals and plants

Hydrocarbon	Toxic effect	Reference(s)
Automated diesel	Changes in CD-1 mice biological activity	Singh et al. (2004)
Gasoline vapours	Impaired the levels of monoamine neurotransmitters in CD-1 mice's brain	Kinawy (2009)
Crude oil	Induces reproductive cytotoxicity confined to the differentiating spermatogonia compartment in rats	Obidike, Maduabuchi, & Olumuyiwa (2007)
PAHs	Impaired glucose metabolism and changes in tricarboxylic acid (TCA) cycle intermediates in earthworm *Lumbricus rubellus*	Jones, Spurgeon, Svendsen, & Griffin (2008)
Petroleum effluent	Low reproduction of earthworm *Eudrilus eugebiae*	Oboh, Adeyinka, Awonuga, & Akinola (2007)
TPH	Reduced seed germination and root growth of *Festuca arundinacea* (tall fescue) and *Euclaena mexicana* (Corn grass)	Tang, Wang, Wang, Sun, & Zhou (2011)
Special Antarctic Blend (SAB) diesel	Reduction of photosynthetic efficiency of green and Antarctic moss	Nydahl, King, Wasley, Jolley, & Robinson (2015)
Diesel fuel	Inhibition of shoot and root growth of *Colobanthus muscoides* (sub-Antarctic moss)	Macoustra, King, Wasley, Robinson, & Jolley (2015)

Table 2
Toxic effects of different types of hydrocarbon on human health

Hydrocarbon	Toxic effect	Reference(s)
Crude petroleum	Chronic cough and phlegm	Rodriguez-Trigo et al. (2010)
Crude oil	Lower respiratory tract symptoms	Zock et al. (2007)
Crude oil	Increased rates of bronchial hyperresponsiveness	Zock et al. (2009)
Crude oil	Persistent respiratory symptoms	Zock et al. (2012)
Volatile organic compounds (VOCs)	Elevated rates of genotoxic damage	Aguilera et al. (2010)
PAHs	Elevated toxicity in reproductive organs	ATSDR (1995)
	Skin and lung cancer	ATSDR (1995)
	Chronic ingestion due to contaminated food	Aguilera et al. (2010)
Diesel exhaust	Irritation to eyes, nose, throat and lungs and increase in the frequency or intensity of asthma attacks	OEHHA (2001)
	Potential of carcinogens in humans	IARC (2012)

The physical characteristics of oil itself can lead to health and environmental disturbance. Deposits of oils tend to accumulate in sewage pipe walls and can cause operational and physical damage, leading to bad odours and the promotion of chemical and microbiological corrosion. These oil clumps also stick onto pipes and ultimately cause filter blockage, which can create pathogenicity issues (Xu & Zhu, 2004). Thick black oils also tend to stick to the feathers of penguins and other birds in the event of an oil spill. This ruins the waterproof coating on the birds' feathers, which could lead to their death due to freezing. Oil that they may swallow is poisonous to them (Kalman & Johnson, 2007).

MANAGEMENT OF OIL SPILL SITES

The Antarctic Treaty Consultative Meeting held in 1998 agreed on the decision to adopt guidelines in the management of oil spills that included measures such as fuel oil handling in scientific stations, spill prevention, control of fuel oils, oil spill contingency planning and reporting of oil spills (ATCM, 1998). Although fuel spills still happen, one might expect their occurrence and frequency to reduce with improvements in both infrastructure and practical applications (Aislabie et al., 2004).

Since then, various attempts to remove pollutants from the Antarctic environment via physical, chemical and biological methods have been established. Treatment of hydrocarbon-polluted Antarctic soils are frequently reviewed (Camenzuli & Freidman, 2015; Jesus et al., 2015) as petroleum hydrocarbon contamination in Antarctica is typically focussed around both operating and abandoned research stations (Curtosi, Pelletier, Vodopivez, & Mac Cormack, 2007; Snape, Morris, & Cole, 2001).

The initial response to a fuel spill is to remove mechanically as much contaminated soil as possible, which includes the underlying snow or ice. These contaminated pieces are

then shipped back to the home country to be disposed of. Although this method is still the prevalent way to remove hydrocarbon pollutants, the fate of the spilled fuel and the impact of removal on the ecosystems cannot be estimated directly (Aislabie et al., 2004). Moreover, this kind of methodology may add up to the original damage towards the environment such as altered streamflow, soil shrinkage, land slumping (due to melting of permafrost) and add to contaminants in other places (Campbell, Claridge, & Balks, 1994; Perelo, 2010). Due to this, more refined approaches are required for both preservation of local environmental conditions and for ethical reasons.

As demands for the eco-friendly technique are rising, alternative remediation methods such as bioremediation are highly proposed and currently established for the Antarctic. Aislabie et al. (2004) stated that bioremediation is gradually regarded as a suitable remediation technology for polluted sites in Antarctica. Successful bioremediation attempts in the clean-up of the Exxon-Valdez oil spill in 1989 (Atlas & Bragg, 2009) also sparked remarkable interest in using bioremediation in cold environments. The accomplishment of oil spill bioremediation varies according to the ability to ascertain and maintain settings that boost oil biodegradation rates in the contaminated environment (Das & Chandran, 2011). In general, fuel spill bioremediation can be separated into two forms: biostimulation and bioaugmentation.

In biostimulation, indigenous soil microbial populations are stimulated through the addition of relevant nutrients and the characteristics of the environmental settings such as pH, oxygen content and also temperature. Successful biostimulation has been reported in Antarctica (Delille, Pelletier, Delille, & Coulon, 2003; Dias et al., 2012). Meanwhile, bioaugmentation encompasses the addition of a pre-adapted microbial strain/consortium or introduction of genetically-engineered microorganisms to particular contaminants present on the site (Tyagi, Da Fonseca, & De Carvalho, 2011). This method has been proposed in Antarctic soils with positive outcomes (Ruberto, Vazquez, & Mac Cormack, 2003; Ruberto et al., 2009).

MICROBIAL DEGRADATION OF PETROLEUM HYDROCARBONS

Bioremediation has been characterised as a pleasant method for petroleum hydrocarbon removal using the application of microbes as it is easy to maintain, applicable over wide areas, economical and leads to complete mineralisation of the pollutants. However, as the Antarctic Treaty prohibits the importation of foreign organisms to the Antarctic environment, only local microbes are required for the bioremediation purposes (Aislabie, Foght, & Saul, 2000).

According to the definition given by Morita (1975), cold-adapted microbes can be differentiated as cold-loving (psychrophilic) microbes and cold-tolerant (psychrotolerant) microbes on the basis of their cardinal temperature. Psychrophilic microbes can be described as a microbial group that achieved the ideal growth temperature of 15°C and below, with growth inhibition above 20°C. In contrast, psychrotolerant microbes can grow over a broad range of temperatures, exhibiting the fastest growth rates above 20°C (Lo Giudice, Bruni, De Domenico, & Michaud, 2010).

Bacteria are the most prevalent species in petroleum hydrocarbon degradation as they play a major role in degrading hydrocarbon pollutants. *Rhodococcus, Pseudomonas* and *Sphingomonas* are among the bacterial group that are commonly isolated from the Antarctic environment polluted by hydrocarbons (Ruberto et al., 2005; Aislabie et al., 2006). They are the most diverse group of microorganisms that is strongly reliable in hydrocarbon degradation in Antarctica and has been reported to degrade both alkanes (Bej, Saul, & Aislabie, 2000; Shukor et al., 2009) and/or aromatic hydrocarbons (Aislabie et al., 2000) aerobically as shown in Table 3. *Rhodococcus* spp. are one of the promising group in hydrocarbon degradation due to their outstanding qualities. Owing to the multiple alkane hydroxylase systems in them, the extent of alkane compound mineralisation is broad (Aggarwal, Dawar, Phanindranath, Mutnuri, & Dayal, 2016; Nie et al., 2014; Whyte et al., 2002). Production of biosurfactants by *Rhodococcus* spp. isolated from the Antarctic have also been reported by Gesheva, Stackebrandt and Vasileva-Tonkova (2010) and Malavenda et al. (2015). The combination of these particular properties together with their tolerance for cold environments definitely establishes them as an effective tool for Antarctic soil remediation. Another bacterial group that has shown a capability for utilising hydrocarbons as carbon and energy sources are *Pseudomonas* spp. and *Sphingomonas* spp. Comparable to *Rhodococcus* spp., this group of bacteria also has several distinctive properties such as biosurfactant production for the former (Pacwa-Plociniczak, Plaza, Poliwoda, & Piotrowska-Seget, 2014; Santa Anna et al., 2002) and the ability to degrade both simple aromatic hydrocarbons and PAHs that have low susceptibility towards microbial degradation (Aislabie et al., 2000; Ma, Wang, & Shao, 2006). Based on the collected reports on the hydrocarbon-degrading ability of microbial isolates, hydrocarbon degraders show a high prevalence of psychrotolerant rather than psychrophilic microbes. Lo Giudice et al. (2010) asserted that this finding may suggest the temperature values and fluctuations that characterise the cold habitat of the microbes' origin. Hence, further efforts should be aimed at isolating pure psychrophilic hydrocarbon degraders, especially those from permanent cold environments.

The mechanistic approach of bacteria in degrading petroleum hydrocarbon is commonly predicted under aerobic degradation, as it is the most rapid and offers complete degradation (Das & Chandran, 2011). Enzymes play a major role in biodegradation of recalcitrant compounds. The process of bioremediation basically depends on microbial enzymes that attack the contaminants, thus converting them into harmless compounds such as carbon dioxide and water molecules. The degradation of petroleum hydrocarbons can be mediated by specific enzyme systems. Even though hydrocarbon degradation in the Antarctic by the anaerobic process has been reported (Powell, Ferguson, Snape, & Siciliano, 2006), aerobic degradation is the prevalent mineralisation pathway chosen mostly by bacteria. Monooxygenases, which are the alkane-activating enzymes, incorporate O_2 as a reactant for the activation of alkane molecules. Another group of enzymes, hydroxylases, initiates oxidation of alkane chains (Rojo, 2009). One of the most commonly mentioned hydroxylase systems is the Cytochrome P450 enzyme system, which is able to hydroxylate large numbers of compounds (Sekine et al., 2006). These enzymes act as the key component in initiating hydrocarbon degradation.

The subsequent pathway of alkane degradation after the involvement of hydroxylases is dependent on the mechanism of the bacteria itself. The most generic one is the yielding of *n*-alcanols by alkane hydroxylases, which are then continued with the oxidation of membrane-bound alcohol dehydrogenase to form *n*-alkanals (Harayama et al., 1999). After transformed subsequently to fatty acids, the products are likely to enter the fatty acid metabolism and eventually produce the ultimate products, carbon dioxide and water.

Meanwhile, Harayama et al. (1999) stated that cycloalkane degradation involves the co-oxidation mechanism as formation of cyclic alcohol and ketones has been observed. A monooxygenase then introduces an oxygen atom into the cyclic ketone causing the cyclic ring to be cleaved (Morgan & Watkinson, 1994). The catabolic pathways for aromatic compounds can be achieved through several pathways. Although different microbial species utilise different pathways for mineralisation of these compounds, hydroxylases and monooxygenases still play the major role in the whole picture (Johnson & Olsen, 1997). In general, the biodegradation of PAHs is initiated by dihydroxylation of one of the polynuclear aromatic rings, followed by cleavage of the dihydroxylated ring. Both hydroxylation is catalysed by a multi-component dioxygenase comprising a reductase, a ferredoxin and an iron sulfur protein, while ring cleavage is normally catalysed by an iron-containing meta-cleavage enzyme. The carbon outline produced by the cleavage reaction is then disassembled, before cleavage of the second aromatic ring (Harayama, Kok, & Neidle, 1992; Saito, Iwabuchi, & Harayama, 1999).

The uptake mechanism of alkanes varies depending on the bacterial species, the molecular weight of the alkanes and the physico-chemical characteristics of the polluted surroundings (Wentzel, Ellingsen, Kotlar, Zotchev, & Throne-Holst, 2007). Low-molecular weight alkanes can be uptaken directly, but for medium- and long-chain *n*-alkanes, microbes may gain contact to the compounds by adhering to hydrocarbon precipitates or by surfactant production (Rojo, 2009). Biosurfactants are surface-active biological compounds produced by a wide variety of microorganisms. One of the advantages of biosurfactants in bioremediation is that they have the ability to improve solubilisation and removal of contaminants (Muthusamy, Gopalakrishnan, Ravi, & Sivachidambaram, 2008). In general, biosurfactants act as emulsifying agents by decreasing the surface tension and forming micelles, which encapsulate the contaminants before they are taken up and degraded by cells.

The existence of filamentous fungi has been reported in Antarctic cold environments, while discovery of hydrocarbon-degrading yeast from the Antarctic has never been reported to the present day. Several hydrocarbon-degrading fungi isolated from Antarctica are listed in Table 4. However, Margesin and Schinner (1997) have reported on a yeast strain, identified as *Yarrowia lipolytica*, that effectively degraded hexadecane and dodecane isolated from the Alpine cold-habitat. Although fungi show capability for utilising hydrocarbons, information on their mechanistic approach and full potential is limited compared to that on bacteria.

Table 3
List of several Antarctic hydrocarbon-degrading bacteria

Strain	Substrate degraded	Isolation place	Reference(s)
Rhodococcus DM1-21	C_{12}-C_{30}, crude oil, gas oil	Marambio station	Ruberto et al., 2005
Rhodococcus B11/B15	Diesel oil	Terra Nova Bay, Ross Sea	De Domenico et al., 2004
Rhodococcus JG-3	N/A (grown on TSA/TSB)	Dry Valleys	Goordial et al., 2015
Rhodococcus sp.	Crude oil, diesel oil	South Shetland Islands	Malavenda et al., 2015
Pseudomonas J3	Diesel oil	Jubany station	Shukor et al., 2009
Pseudomonas Ant 9	xylene, benzene, naphthalene	Scott Base	Aislabie et al., 1998
Pseudomonas LCY16	Naphthalene, phenanthrene	Great Wall station	Ma et al., 2006
Sphingomonas Ant 17	Naphthalene, phenanthrene	Scott Base	Baraniecki et al., 2002

*N/A - not available

Table 4
List of several hydrocarbon-degrading fungi from Antarctica

Strain	Substrate degraded	Isolation place	Reference(s)
Arthroderma sp. 1	Monoaromatic hydrocarbon	Macquarie Island	Ferrari et al., 2011
Pseudeurotium bakeri	Diesel oil	Macquarie Island	Ferrari et al., 2011
Mortierella sp. HC8D	Dodecane	Rothera research station	Hughes et al., 2007
Phialophora sp.	N/A (#)	Scott Base, Marble Point	Aislabie et al., 2001
Chrysosporium sp.	N/A (#)	Scott Base, Marble Point	Aislabie et al., 2001
Alternaria sp.	N/A (#)	Scott Base, Marble Point	Aislabie et al., 2001

*N/A - not available, # - isolated from hydrocarbons-contaminated sites

FACTORS AFFECTING HYDROCARBON DEGRADATION IN ANTARCTICA

The Antarctic environment is known for its harsh environment – extremely cold weather, strong winds and brutal storms (Mishra, Yadav, Sharma, Ganju, & Singh, 2014). These conditions become limiting factors in hydrocarbon degradation in this region. Among the factors that affect the hydrocarbon biodegradation rate are temperature and nutrient availability.

Temperature is one of the most important aspects in determining the success of biodegradation. Ma et al. (2006) agreed that low temperatures in Antarctica are the main limiting factor of hydrocarbon biodegradation, as physical nature and composition of spilled oil are more prone to modifications in such temperature. According to Margesin and Schinner (1999), due to cold temperatures, oil tends to increase in viscosity, thus causing hydrocarbon volatilisation to be reduced and slowing down the mineralisation process. As the rate of reaction

is thought to follow the Arrhenius relationship, biodegradation will be assumed to decrease as the temperature decreases, which is a disadvantage of the Antarctic cold environment. However, contaminants can be reduced in the Antarctic cold environment as bioremediation treatments are usually done and proposed in the summertime, when temperatures are higher, where soils are slightly thawed and water is more accessible (Atlas, 1986). Hydrocarbon degradation at low temperatures has been investigated in the laboratory in the range of 4-30°C (Aislabie et al., 2006; Muangchinda et al., 2015), but there have been no reports on degradation in subzero temperature.

Necessary nutrients are also one of the important factors in dealing with biodegradation of petroleum hydrocarbon. Fuel spills on Antarctic soil (or water bodies) can raise the levels of soil organic carbon, which may either serve as substrates for microbial growth or be growth-retardant due to toxic effects (Bossert & Bartha, 1984). However, the increased proportions of carbon in the environment reduce the availability of other nutrients like nitrogen and phosphorus to the microorganisms, thus disturbing nutrient balance in the Antarctic environment. Due to low nutrient levels in polluted Antarctic sites, several attempts have been made to increase nutrient concentration such as addition of oleophilic fertilisers to sub-Antarctic soils (Coulon, Pelletier, St. Louis, Gourhant, & Delille, 2004); optimisation of the biostimulation strategy (Alvarez, Lo Balbo, Mac Cormack, & Ruberto, 2015); and supplementation of nutrients/water (Stallwood, Shears, Williams, & Hughes, 2005).

Furthermore, availability of oil waste and oxygen in the environment to microorganisms does influence the rate of biodegradation. For the former, Barathi and Vasudevan (2001) stated that hydrocarbon compounds tend to bind to soil particles, causing the degradation and removal of hydrocarbons to be harder. These problems, however, can be countered in certain ways – usage of biosurfactant-producing microorganisms, biopiling and soil heating. Atlas (1991), meanwhile, stated that the availability of oxygen in soils, sediments, and aquifers are often limiting and reliant on the type of soil and whether the soil is clogged, thus disturbing the aerobic degradation of hydrocarbon compounds.

NEW BIOREMEDIATION PROSPECTS IN ANTARCTICA

The outcomes of applying microbial species for treatment of petroleum hydrocarbon pollution in Antarctica have been tremendous due to the success of great ideas and their effective application in recent decades. Although microbial degradation of petroleum hydrocarbons in Antarctica is well understood nowadays, extra work and effort are needed to optimise and improve the existing approach. Emerging works should be considered and studied to attain the promise they carry. Several developing technologies in microbial bioremediation that show great potential are phytoremediation and the use of genetically-modified bacteria.

According to Greipsson (2011), phytoremediation is the use of plants and associated soil microbes to reduce the concentration or toxic effects of contaminants in the environment. Some advantages of phytoremediation are cost-effective, eco-friendly and generates recyclable energy in ecosystems. Phytoremediation technologies can be categorised as phytostabilisation, phytodegradation, phytovolatilisation and phytoextraction, with comparable results among the systems (Greipsson, 2011). Recent decades have favoured phytoremediation although

there are fewer reports on its use in Antarctica (Ali, Khan, & Sajad, 2013; Mitton, Gonzalez, Monserrat, & Miglioranza, 2016).

Successful hydrocarbon remediation through plants has been reported by Jones, Sun, Tang and Robert (2004) and Sun, Lo, Robert, Ray and Tang (2004). Yet, the majority of research with respect to phytoremediation of petroleum hydrocarbons has concentrated on the method of rhizodegradation, due to the cold climate of Antarctica. Rhizodegradation, in general, is defined as the transformation of pollutants in the soil proximal to the roots (rhizosphere) by organisms associated with vegetative species such as bacteria. Though plants and microorganisms can mineralise petroleum hydrocarbons independently of one another, Atlas and Bartha (1998) implied that it is the interaction between plants and microorganisms that is the principal means responsible for petrochemical degradation in phytoremediation efforts. Daryabeigi Zand and Hoveidi (2016) supported the theory of the synergetic energy between plants and microorganisms for the increased degradation effect of recalcitrant compounds based on the positive outcomes in recent studies by Liu, Meng, Tong and Chi (2014a) and Xiao, Liu, Jin and Dai (2015).

In most higher and vascular plants like grasses, several aspects do need to be evaluated for a potential phytoremediator such as their root architecture, exudate patterns of the root, cell wall components and the genetic composition of the plant involved (Balasubramaniyam, 2015). However, certain limitations in vascular plants might occur as plant roots play the most vital part in phytoremediation, specifically rhizodegradation. The presence of hydrocarbons in soil, for instance, may pose a challenge to the growth of plant roots for their toxicity and water stress (Balasubramaniyam, Chapman, & Harvey, 2015), automated hindrance and nutrient deficiency (Brandt, Merkl, Schultze-Kraft, Infante, & Broll, 2006; Merkl, Schultze-Kraft, & Infante, 2005). Meanwhile, studies on phytoremediation using lower plants such as algae, lichens and moss also have increased in recent decades due to their key role in carbon dioxide fixation and immense growth and biomass (Chekroun, Sanchez, & Baghour, 2014; Jacques & McMartin, 2009). While their importance as a potential degrader can be considered as a branch of phytoremediation, most studies on certain lower plants are classified as degradation by microorganisms. In general, the application of phytoremediation in Antarctica might be less practical due to the harsh climate, lack of native plants and soil conditions (Camenzuli, Freidman, Statham, Mumford, & Gore, 2013). Overall, no data have been reported to date on the application of phytoremediation in Antarctic hydrocarbon-polluted sites, except for the mass study by Bramley-Alves, Wasley, King, Powell and Robinson (2014) on the degradation of petroleum hydrocarbons in sub-Antarctic soils by native tussock grass (*Poa foliosa*). Therefore, it is highly encouraged that researchers consider a new branch of subjects regarding this topic.

The idea of employment of genetically-modified (GM) microorganisms (especially bacteria) in hydrocarbon bioremediation has sparked enormous attention among researchers to improve the degradation of hazardous wastes. Past decades have shown several studies on hydrocarbon degradation by GM bacteria such as *Alcaligenes eutrophus* H850 (Van Dyke, Lee, & Trevors, 1996), *Pseudomonas fluorescens* HK44 (Sayler & Ripp, 2000), Sinorhizobium meliloti (Dutta, Hollowell, Hashem, & Kuykendall, 2003), *Pseudomonas putida* PaW 340/pDH5 (Massa et al., 2009) and *Rhodococcus* sp. strain RHA1 (Rodrigues et al., 2001). However,

the Antarctic Treaty of 1998 prohibits the introduction of foreign organisms into the Antarctic environment. Hence, GM bacterial species indigenous to the Antarctic are hugely in demand, as no information has been reported on this subject.

CONCLUSION

Recovering and improving Antarctic polluted environments is a real-world concern, as the Antarctic continent can be deemed to be one of the last pristine environments on Earth. In treating pollution ranging from oil spills to soil contamination, bioremediation stands out as an effective approach in the attempt to restore environments to their original conditions. Exploiting the ability of microorganisms to degrade petroleum hydrocarbons, together with the introduction of nutrients, oxygen and other biological means has brought bioremediation to a new level. However, managing this biological means is a meticulous challenge in a harsh environmental condition such as Antarctica.

Improved understanding of the present knowledge such as petroleum hydrocarbon composition and toxicity; management of oil spills; microbial degradation and factors affecting degradation are needed for successful bioremediation. Furthermore, new ideas such as the promising use of phytoremediation (rhizodegradation) and genetically-modified (GM) indigenous bacteria must be looked into for future consideration.

ACKNOWLEDGEMENT

The authors would like to acknowledge Universiti Putra Malaysia for awarding the IPM grant (9476900) and (9422700) that enabled the writing of this review article.

REFERENCES

Abdel-Shafy, H. I., & Mansour, M. S. M. (2016). A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egyptian Journal of Petroleum, 25*(1), 107–123.

ACGIH. (2005). *Polycyclic aromatics hydrocarbons (PAHs) biologic exposure indices (BEI)*. American Conference of Governmental Industrial Hygienists, Cincinnati, USA.

Aggarwal, R. K., Dawar, C., Phanindranath, R., Mutnuri, L., & Dayal, M. (2016). Draft genome sequence of a versatile hydrocarbon-degrading bacterium, *Rhodococcus pyridivorans* strain KG-16, collected from oil fields in India. *Genome Announcements, 4*(1), 1704–1715.

Aguilera, F., Mendez, J., Pasaro, E., & Laffon, B. (2010). Review on the effects of exposure to spilled oils on human health. *Journal of Applied Toxicology, 30*(4), 291–301.

Aislabie, J., Balks, M., Foght, J., & Waterhouse, E. (2004). Hydrocarbon spills on Antarctic soils: Effects and management. *Environmental Science and Technology, 38*(5), 1266–1274.

Aislabie, J., Foght, J., & Saul, D. (2000). Aromatic hydrocarbon-degrading bacteria from soil near Scott Base, Antarctica. *Polar Biology, 23*, 183–188.

Aislabie, J., Fraser, R., Duncan, S., & Farrell, R. L. (2001). Effects of oil spills on microbial heterotrophs in Antarctic soils. *Polar Biology, 24*, 308–313.

Aislabie, J., McLeod, M., & Fraser, R. (1998). Potential of biodegradation of hydrocarbons in soil from the Ross Dependency, Antarctica. *Applied Microbiology and Biotechnology, 49*, 210–214.

Aislabie, J., Saul, D. J., & Foght, J. M. (2006). Bioremediation of hydrocarbon-contaminated polar soils. *Extremophiles, 10*, 171–179.

Ali, H., Khan, E., & Sajad, M. A. (2013). Phytoremediation of heavy metals – Concepts and applications. *Chemosphere, 91*(7), 869–881.

Alvarez, L. M. M., Lo Balbo, A., Mac Cormack, W. P., & Ruberto, L. A. M. (2015). Bioremediation of a petroleum hydrocarbon-contaminated Antarctic soil: Optimization of a biostimulation strategy using response surface methodology (RSM). *Cold Regions Science and Technology, 119*, 61–67.

AMAP. (1998). AMAP Assessment Report. *Arctic Pollution Issues, Arctic Monitoring and Assessment Programme (AMAP) (p. 859)*. Oslo, Norway.

Armstrong, B. G., Hutchinson, E., Unwin, J., & Fletcher, T. (2004). Lung cancer risk after exposure to polycyclic aromatic hydrocarbons: A review and meta-analysis. *Environmental Health Perspectives, 112*(9), 970–978.

ATCM. (1998). *Final report of the twenty-second Antarctic Treaty Consultative Meeting*. Antarctic Treaty Consultative Meeting: Tromso, Norway.

Atlas, R., & Bragg, J. (2009). Bioremediation of marine oil spills: When and when not – The *Exxon Valdez* experience. *Microbial Biotechnology, 2*(2), 213–221.

Atlas, R. M. (1986). Fate of petroleum pollutants in Arctic ecosystems. *Water Science and Technology, 18*(2), 59–67.

Atlas, R. M. (1991). Microbial hydrocarbon degradation – Bioremediation of oil spills. *Journal of Chemical Technology and Biotechnology, 52*, 149–156.

Atlas, R. M., & Bartha, R. (1998). *Microbial ecology: Fundamentals and applications* (4th Ed.). USA: Benjamin Cummings.

ATSDR. (1995). *Toxicological profile for polycyclic aromatic hydrocarbons (PAHs)*. Agency for Toxic Substances and Disease Registry. Atlanta, Georgia, US Department of Health and Human Services, Public Health Service.

ATSDR. (1999). *Toxicological profile for total petroleum hydrocarbons (TPH)*. Agency for Toxic Substances and Disease Registry. Atlanta, Georgia, US Department of Health and Human Services, Public Health Service.

Azubuike, C. C., Chikere, C. B., & Okpokwasili, G. C. (2016). Bioremediation techniques – Classification based on site of application: Principles, advantages, limitations and prospects. *World Journal of Microbiology and Biotechnology, 32*(180), 1–18.

Balasubramaniyam, A. (2015). The influence of plants in the remediation of petroleum hydrocarbon-contaminated sites. *Pharmaceutical Analytical Chemistry, 1*, 105. doi: 10.4172/2471-2698.1000105.

Balasubramaniyam, A., Chapman, M. M., & Harvey, P. J. (2015). Responses of tall fescue (*Festuca arundinacea*) to growth in naphthalene-contaminated sand: Xenobiotic stress versus water stress. *Environmental Science and Pollution Research, 22*, 7495–7507.

Baraniecki, C. A., Aislabie, J., & Foght, J. M. (2002). Characterization of *Sphingomonas* sp. Ant 17, an aromatic hydrocarbon-degrading bacterium isolated from Antarctic soil. *Microbial Ecology, 43*, 22–54.

Barathi, S., & Vasudevan, N. (2001). Utilization of petroleum hydrocarbons by *Pseudomonas fluorescens* isolated from a petroleum-contaminated soil. *Environmental International, 20*(5–6), 413–416.

Bej, A. K., Saul, D., & Aislabie, J. (2000). Cold-tolerant alkane-degrading *Rhodococcus* species from Antarctica. *Polar Biology, 23*, 100–105.

Bosset, I., & Bartha, R. (1984). The fate of petroleum in soil ecosystems. In R. M. Atlas, (Ed.), *Petroleum microbiology* (pp. 435–476). New York, NY: Macmillan Publishing Company.

Bramley-Alves, J. (2009). *Phytoremediation of hydrocarbon contaminants in sub-antarctic soils*. (Unpublished bachelor's dissertation). School of Biological Sciences, University of Wollongong, New South Wales, Australia.

Brandt, R., Merkl, N., Schultze-Kraft, R., Infante, C., & Broll, G. (2006). Potential of vetiver [*Vetiveria zizanioides* (L.) Nash] for phytoremediation of petroleum hydrocarbon-contaminated soils in Venezuela. *International Journal of Phytoremediation, 8*, 273–284.

Camenzuli, D., & Freidman, B. L. (2015). On-site and in situ remediation technologies applicable to petroleum hydrocarbon contaminated sites in the Antarctic and Arctic. *Polar Research, 34*, 24492, http://dx.doi.org/10.3402/polar.v34.24492.

Camenzuli, D., Freidman, B. L., Statham, T. M., Mumford, K. A., & Gore, D. B. (2013). On-site and in situ remediation technologies applicable to metal-contaminated sites in Antarctica and the Arctic: A review. *Polar Research, 32*(1), 21522. doi: http://dx.doi.org/10.3402/polar.v33.21522.

Campbell, I. B., Claridge, G. C. G., & Balks, M. R. (1994). The effect of human activities on moisture content of solids and underlying permafrost from the McMurdo Sound region, Antarctica. *Antarctic Science, 6*(3), 307–314.

Chekroun, K. B., Sanchez, E., & Baghour, M. (2014). The role of algae in bioremediation of organic pollutants. *International Research Journal of Public and Environmental Health, 1*(2), 19–32.

Chuvilin, E. M., Naletova, N. S., Miklyaeva, E. C., Kozlova, E. V., & Instanes, A. (2001a). Factors affecting spreadibility and transportation of oil in regions of frozen ground. *Polar Record, 37*, 229–238.

Coulon, F., Pelletier, E., St. Louis, R., Gourhant, L., & Delille, D. (2004). Degradation of petroleum hydrocarbons in two sub-antarctic soils: Influence of an oleophilic fertilizer. *Environmental Toxicology and Chemistry, 23*(8), 1893–1901.

Curtosi A., Pelletier E., Vodopivez C. L., & Mac Cormack, W. P. (2007). Polycyclic aromatic hydrocarbons in soil and surface marine sediment near Jubany Station (Antarctica) – Role of permafrost as a low-permeability barrier. *Science of the Total Environment, 383*, 193–204.

Daryabeigi Zand, A., & Hoveidi, H. (2016). Plant-aid remediation of hydrocarbon-contaminated sites. *Pollution,* 2(3), 233–246.

Das, N., & Chandran, P. (2011). Microbial degradation of petroleum hydrocarbon contaminants: An overview. *Biotechnology Research International, 2011*(2011), 1-13. doi: 10.4061/2011/941810.

De Domenico, M., Lo Giudice, A., Michaud, L., Saitta, M., & Bruni, V. (2004). Diesel oil and PCB-degrading psychrotrophic bacteria isolated from Antarctic seawaters (Terra Nova Bay, Ross Sea). *Polar Research, 23*(2), 141–146.

Delille, D., Pelletier, E., Delille, B., & Coulon, F. (2003). Effect of nutrient enrichments on the bacterial assemblage of Antarctic soils contaminated by diesel or crude oil. *Polar Research, 39*, 1–10.

Det Norske Veritas. (2003). Russia pipeline oil spill study. ESMAP Technical Paper, vol. 034-03. *Joint UNDP/World Bank Energy Sector Management Assistance Programme* (p. 82). Norway.

Dias, R. L., Ruberto, L., Hernández, E., Vázquez, S. C., Balbo, A. L., Del Panno, M. T., & Mac Cormack, W. P. (2012). Bioremediation of an aged diesel oil-contaminated Antarctic soil: Evaluation of the ''on site'' biostimulation strategy using different nutrient sources. *International Biodeterioration and Biodegradation, 75*, 96–103.

Dutta, S. K., Hollowell, G. P., Hashem, F. M., & Kuykendall, L. D. (2003). Enhanced bioremediation of soil containing 2, 4-dinitrotoluene by a genetically modified *Sinorhizobium meliloti. Soil Biology and Biochemistry, 35*, 667–675.

Ferrari, B. C., Zhang, C., & Van Dorst, J. (2011). Recovering greater fungal diversity from pristine and diesel fuel contaminated sub-Antarctic soil through cultivation using both and high and low nutrient media approach. *Frontiers in Microbiology, 217*(2), 1-14. doi: 10.3389/fmicb.2011.00217.

Gesheva, V., Stackebrandt, E., & Vasileva-Tonkova, E. (2010). Biosurfactant production by halotolerant *Rhodococcus fascians* from Casey Station, Wilkes Land, Antarctica. *Current Microbiology, 61*, 112–117.

Goordial, J., Raymond-Bouchard, I., Ronholm, J., Shapiro, N., Woyke, T., Whyte, L., & Bakermans, C. (2015). Improved-high-quality draft gene sequence of *Rhodococcus* sp. JG-3, a eurypsychrophilic Actinobacteria from Antarctic Dry Valley permafrost. *Standard in Genomic Sciences, 10*(1), 61. doi: 10.1186/s40793-015-0043-8.

Greipsson, S. (2011). Phytoremediation. *Nature Education Knowledge, 3*(10), 7.

Harayama, S., Kishira, H., Kasai, Y., & Shutsubo, K. (1999). Petroleum biodegradation in marine environments. *Journal of Molecular Microbiology and Biotechnology, 1*(1), 63–70.

Harayama, S., Kok, M., & Neidle, E. L. (1992). Functional and evolutionary relationships aming diverse oxygenases. *Annual Review of Microbiology, 46*, 565–601.

Hughes, K. A., Bridge, P., & Clark, M. S. (2007). Tolerance of Antarctic soil fungi to hydrocarbons. *Science of the Total Environment, 372*, 539–548.

IARC. (2012). *Diesel engine exhaust carcinogenic*. International Agency for Research on Cancer. World Health Organization (WHO).

Jacques, N. R., & McMartin, D. W. (2009). Evaluation of algal phytoremediation of light extractable petroleum hydrocarbons in subarctic climates. *Remediation, 20*(1), 119–132.

Jesus, H. E., Peixoto, R. S., & Rosado, A. S. (2015). Bioremediation in Antarctic Soils. *Petroleum and Environmental Biotechnology, 6*(6), 1–11.

Johnson, G. R., & Olsen, R. H. (1997). Multiple pathways for toluene degradation in *Burkholderia* sp. strain JS150. *Applied and Environmental Microbiology, 63*, 4047–4052.

Jones, O. A. H., Spurgeon, D. J., Svendsen, C., & Griffin, J. L. (2008). A metabolomics based approach to assessing the toxicity of the polyaromatic hydrocarbon pyrene to the earthworm *Lumbricus rubellus*. *Chemosphere, 71*(3), 601–609.

Jones, R. K., Sun, W. H., Tang, C. S., & Robert, F. M. (2004). Phytoremediation of petroleum hydrocarbons in tropical coastal soils. II. Microbial response to plant roots and containment. *Environmental Science and Pollution Research, 11*(5), 340–346.

Kalman, B., & Johnson, R. (2007). *Endangered penguins*. New York: Crabtree Publishing and Cooperation.

Kennicutt, M. C., McDonald, T. J., Denoux, G. J., & McDonald, S. J. (1992). Hydrocarbon contamination on the Antarctic Peninsula: I. Arthur harbor –Subtidal sediments. *Marine Pollution Bulletin, 24*(10), 499–506.

Kim, K. H., Jahan, S. A., Kabir, E., & Brown, R. J. C. (2013). A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. *Environmental International, 60*, 71–80.

Kinawy, A. A. (2009). Impact of gasoline inhalation on some neurobehavioural characteristics of male rats. *BMC Physiology, 9*(21), 1–10.

Kuo, C. Y., Hsu, Y. W., & Lee, H. S. (2003). Study of human exposure to particulate PAHs using personal air samplers. *Archive of Environmental Contamination and Toxicology, 44*, 454–459.

Liu, H., Meng, F., Tong, Y., & Chi, J. (2014a). Effect of plant density on phytoremediation of polycyclic aromatic hydrocarbons contaminated sediments with *Vallisneria spiralis*. *Ecological Engineering, 73*, 380–385.

Lo Giudice, A., Bruni, V., De Domenico, M., & Michaud, L. (2010). Psychrophiles – Cold-adapted hydrocarbon-degrading microorganisms. In K. N. Timmis (Ed.), *Handbook of hydrocarbon and lipid microbiology* (pp. 1897-1921). doi: 10.1007/978-3-540-77587-4

Ma, Y., Wang, L., & Shao, Z. (2006). *Pseudomonas*, the dominant polycyclic aromatic hydrocarbon-degrading bacteria isolated from Antarctic soils and the role of large plasmids in horizontal gene transfer. *Environmental Microbiology, 8*(3), 455–465.

Macoustra, G. K., King, C. K., Wasley, J., Robinson, S. A., & Jolley, D. F. (2015). Impact of hydrocarbons from a diesel fuel on the germination and early growth of subantarctic plants. *Environmental Science Processes and Impacts, 17*(7), 1238–1248.

Malavenda, R., Rizzo, C., Michaud, L., Gerçe, B., Bruni, V., Syldatk, C., ... & Giudice, A. L. (2015). Biosurfactant production by Arctic and Antarctic bacteria growing on hydrocarbons. *Polar Biology, 38*(10), 1565–1574.

Margesin, R., & Schinner, F. (1997). Effect of temperature on oil degradation by a psychrotrophic yeast in liquid culture and in soil. *FEMS Microbiology and Ecology, 24*, 243–249.

Margesin, R., & Schinner, F. (1999). Biological decontamination of oil spills in cold environments. *Journal of Chemical Technology and Biotechnology, 74*, 381–389.

Massa, V., Infantino, A., Radice, F., Orlandi, V., Tavecchio, F., & Giudiuci, R. (2009). Efficiency of natural and engineered bacterial strains in the degradation of 4-chlorobenzoic acid in soil slurry. *International Biodeterioration and Biodegradation, 63*, 112–125.

McWatters, R. S., Wilkins, D., Spedding, T., Hince, G., Raymond, B., Lagerewskij, G., ... & Snape, I. (2016). On site remediation of a fuel spill and soil reuse in Antarctica. *Science of the Total Environment, 57*1, 963–973.

Merkl, N., Schultze-Kraft, R., & Infante, C. (2005). Phytoremediation in the tropics-influence of heavy crude oil on root morphological characteristics of graminoids. *Environmental Pollution, 138*, 86–91.

Mishra, K. P., Yadav, A. P., Sharma, Y. K., Ganju, L., & Singh, S. B. (2014). Effect of extreme conditions of Antarctica on human leukocyte antigen-G in Indian expeditioners. *Indian Journal of Medical Research, 140*(4), 520–523.

Mitton, F. M., Gonzalez, M., Monserrat, J. M., & Miglioranza, K. S. (2016). Potential use of edible crops in the phytoremediation of endosulfan residues in soil. *Chemosphere, 148*, 300–306.

Mohn, W. W., & Stewart, G. R. (2000). Limiting factors for hydrocarbon degradation at low temperature in Arctic soils. *Soil Biology and Biochemistry, 32*, 1161–1172.

Montagnolli, R. N., Lopes, P. R. M., & Bidoia, E. D. (2015). Screening the toxicity and biodegradability of petroleum hydrocarbons by a rapid colorimetric method. *Archives of Environmental Contamination and Toxicology, 68*, 342–353.

Morgan, P., & Watkinson, R. J. (1994). Biodegradation of components of petroleum. In C. Rathledge (Ed.), *Biochemistry of microbial degradation* (pp. 1–31). Dordrecht, Netherlands: Kluwer Academic Publishers.

Morita, R. Y. (1975). Psychrophilic bacteria. *Bacteriological Reviews, 39*, 144–167.

Muangchinda, C., Chavanich, S., Viyakarn, V., Watanabe, K., Imura, S., Vangnai, A. S., & Pinyakong, O. (2015). Abundance and diversity of functional genes involved in the degradation of aromatic hydrocarbons in Antarctic soils and sediments around Syowa Station. *Environmental Science and Pollution Research, 22*(6), 4725–4735.

Muthusamy, K., Gopalakrishnan, S., Ravi, T. K., & Sivachidambaram, P. (2008). Biosurfactants: Properties, commercial production and application. *Current Science, 94*(6), 736–747.

National Research Council. (2003). *Oil in the Sea III: Inputs, Fates and Effects*. Washington: The National Academic Press.

Nie, Y., Chi, C. Q., Fang, H., Liang, J. L., Lu, S. L., Lai, G. L., ... & Wu, X. L. (2014). Diverse alkane hydroxylase genes in microorganisms and environments. *Scientific Reports, 4*, 4968.

Nydahl, A. C., King, C. K., Wasley, J., Jolley, D. F., & Robinson, S. A. (2015). Toxicity of fuel-contaminated soil to Antarctic moss and terrestrial algae. *Environmental Toxicology and Chemistry, 34*(9), 2004–2012.

Obidike, I. R., Maduabuchi, I. U., & Olumuyiwa, S. S. O. (2007). Testicular morphology and cauda epididymal sperm reserves of male rats exposed to Nigerian Qua Iboe Brent crude oil. *Journal of Veterinary Science, 8*(1), 1–5.

Oboh, B. O., Adeyinka, Y. Awonuga, S., & Akinola, M. O. (2007). Impact of soil types and petroleum effluents on the earthworm *Eudrillus eugeniae*. *Journal of Environmental Biology, 28*(2), 209–212.

OEHHA (Office of Environmental Health Hazard Assessment). (2001). *Health effects of diesel exhaust.* Sacramento, California: California Environmental Protection Agency.

Pacwa-Plociniczak, M., Plaza, G. A., Poliwoda, A., & Piotrowska-Seget, Z. (2014). Characterization of hydrocarbon-degrading and biosurfactant-producing *Pseudomonas* sp. P-1 strain as a potential tool for bioremediation of petroleum-contaminated soil. *Environmental Science and Pollution Research, 21*, 9385–9395.

Paixão, J. F., Nascimento, I. A., Pereira, S. A., Leite, M. B. L., Carvalho, G. C. D., Silveira, J. S. C., ... & Rodrigues, I. L. P. (2007). Estimating the gasoline component and formulations toxicity to microalgae (*Tetraselmis chuii*) and oyster (*Crassostrea rhizophorae*) embryos: An approach to minimize environmental pollution risk. *Environmental Research, 103*(3), 365–374.

Perelo, L. W. (2010). Review: In situ and bioremediation of organic pollutants in aquatic sediments. *Journal of Hazardous Materials, 177*, 81–89.

Powell, S. M., Ferguson, S. H., Snape, I., & Siciliano, S. D. (2006). Fertilization stimulates anaerobic fuel degradation of Antarctic soils by denitrifying microorganisms. *Environmental Science and Technology, 40*(15), 2011–2017.

Rodrigues, J. L. M., Maltseva, O. V., Tsoi, T. V., Helton, R. R., Quensen, J. F., & Fakuda, M. (2001). Development of a *Rhodococcus* recombinant strain for degradation of products from anaerobic dechlorination of PCBs. *Environmental Science and Technology, 35*, 663–668.

Rodríguez-Trigo, G., Zock, J. P., Pozo-Rodríguez, F., Gómez, F. P., Monyarch, G., Bouso, L., ... & Barberà, J. A. (2010). Health changes in fishermen 2 years after clean-up of the *Prestige* oil spill. *Annals of Internal Medicine, 153*(8), 489–498.

Rojo, F. (2009). Degradation of alkanes by bacteria. *Environmental Microbiology, 11*(10), 2477–2490.

Rothwell, D. R., & Davis, R. (1997). *Antarctic environment protection: A collection of Australian and international instruments*. New South Wales: The Federation Press.

Ruberto, L., Dias, R., Lo Balbo, A., Vazquez, S. C., Hernandez, E. A., & Mac Cormack, W. P. (2009). Influence of nutrients addition and bioaugmentation on the hydrocarbon biodegradation of a chronically contaminated Antarctic soil. *Journal of Applied Microbiology, 106*(4), 1101–1110.

Ruberto, L., Vazquez, S. C., & Mac Cormack, W. P. (2003). Effectiveness of the natural bacterial flora, biostimulation and bioaugmentation on the bioremediation of the hydrocarbon contaminated Antarctic soil. *International Biodeterioration and Biodegradation, 52*(2), 115–125.

Ruberto, L. A. M., Vazquez, S., Lobaldo, A., & Mac Cormack, W. P. (2005). Psychrotolerant hydrocarbon-degrading *Rhodococcus* strains isolated from polluted Antarctic soils. *Antarctic Science, 17*(1), 47–56.

Saito, A., Iwabuchi, T., & Harayama, S. (1999). Characterisation of genes for enzymes involved in the phenanthrene degradation in *Nocardioides sp*. KP7, *Chemosphere, 38*, 1331–1337.

Santa Anna, L. M., Sebastian, G. V., Menezes, E. P., Alves, T. L. M., Santos, A. S., Pereira Jr, N., & Freire, D. M. G. (2002). Production of biosurfactants from *Pseudomonas aeruginosa* PA1 isolated in oil environments. Brazilian Journal of Chemical Engineering, 19(2), 159–166.

Saul, D. J., Aislabie, J. M., Brown, C. E., Harris, L., & Foght, J. M. (2005). Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica. *FEMS Microbiology Ecology, 53*, 141–155.

Sayler, G. S., & Ripp, S. (2000). Field applications of genetically engineered microorganisms for bioremediation processes. *Current Opinion in Biotechnology, 11*(3), 286–289.

Sekine, M., Tanikawa, S., Omata, S., Saito, M., Fujisawa, T., Tsukatani, N., ... & Nishiko, R. (2006). Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. *Environmental Microbiology, 8*(2), 334–346.

Singh, P., DeMarini, D. M., Dick, C. A., Tabor, D. G., Ryan, J. V., Linak, W. P., ... & Gilmour, M. I. (2004). Sample characterisation of automobile and forklift diesel exhaust particles and comparative pulmonary toxicity in mice. *Environmental Health Perspectives, 112*(8), 820–825.

Shukor, M. Y., Hassan, N. A. A., Jusoh, A. Z., Perumal, N., Shamaan, N. A., MacCormack, W. P., & Syed, M. A. (2009). Isolation and characterisation of a *Pseudomonas* diesel-degrading strain from Antarctica. *Journal of Environmental Biology, 30*(1), 1–6.

Snape, I., Morris, C. E., & Cole, C. M. (2001). The use of permeable reactive barriers to control contaminants dispersal during site remediation in Antarctica. *Cold Regions Science and Technology, 32,* 157–174.

Stallwood, B., Shears, J., Williams, P. A., & Hughes, K. A. (2005). Low temperature bioremediation of oil-contaminated soil using biostimulation and bioaugmentation with a *Pseudomonas* sp. from maritime Antarctica. *Journal of Applied Microbiology, 99*(4), 794–802.

Sun, W. H., Lo, J. B., Robert, F. M., Ray, C., & Tang, C. S. (2004). Phytoremediation of petroleum hydrocarbons in tropical coastal soils. I. Selection of promising woody plants. *Environmental Science and Pollution Research, 11*(4), 260–266.

Tang, J., Wang, M., Wang, F., Sun, Q., & Zhou, Q. (2011). Eco-toxicity of petroleum hydrocarbon contaminated soil. *Journal of Environmental Sciences, 23*(5), 845–851.

Tyagi, M., Da Fonseca, M. M. R., & De Carvalho, C. C. C. R. (2011). Bioaugmentation and biostimulation strategies to improve the effectiveness of bioremediation processes. *Biodegradation, 22*(2), 231–241.

Van Dyke, M. I., Lee, H., & Trevors, J. T. (1996). Survival of luxAB-marked *Alcaligenes eutrophus* H850 in PCB-contaminated soil and sediment. J*ournal of Chemical Technology and Biotechnology, 64*(7), 2730–2735.

Wang, X., & Bartha, R. (1990). Effects of bioremediation on residues, activity and toxicity in soil contaminated by fuel spills. *Soil Biology and Biochemistry, 22*(4), 501–505.

Webster, J., Webster, K., Nelson, P., & Waterhouse, E. (2003). The behaviour of residual contaminants at a fomer site, Antarctica. *Environmental Pollution, 123*(2), 163–179.

Wentzel, A., Ellingsen, T. E., Kotlar, H. K., Zotchev, S. B., & Throne-Holst, M. (2007). Bacterial metabolism of long-chain *n*-alkanes. *Applied Microbiology and Biotechnology, 76*, 1209–1221.

Whyte, L. G., Smits, T. H., Labbè, D., Witholt, B., Greer, C. W., & Van Beilen, J. B. (2002). Gene cloning and characterisation of multiple alkane hydroxylase systems in *Rhodococcus* strains Q15 and NRRL B-16531. *Applied and Environmental Microbiology, 68*(12), 5933–5942.

Xiao, N., Liu, R., Jin, C., & Dai, Y. (2015). Efficiency of five ornamental plant species in the phytoremediation of polycyclic aromatic hydrocarbon (PAH)-contaminated soil. *Ecological Engineering, 75*, 384–391.

Xu, X., & Zhu, X. (2004). Treatment of refectory oily wastewater by electro-coagulation process. *Chemosphere, 56*, 889–894.

Zock, J. P., Rodríguez-Trigo, G., Pozo-Rodríguez, F., Barberà, J. A., Bouso, L., Torralba, Y., ... & SEPAR-Prestige Study Group*. (2007). Prolonged respiratoory symptoms in clean-ip workers of the *Prestige* oil spill. *American Journal of Respiratory and Critical Care Medicine, 176*(6), 610–616.

Zock, J. P., Rodríguez-Trigo, G., Rodríguez-Rodríguez, E., Espinosa, A., Pozo-Rodríguez, F., Gómez, F., ... & Barberà, J. A. (2012). Persistent respiratory symptoms in clean-up workers 5 years after the *Prestige* oil spill. *Occupational and Environmental Medicine, 69*(7), 508–513.

Zock, J. P., Rodríguez-Trigo, G., Rodríguez-Rodríguez, E., Pozo-Rodríguez, F., Gómez, F. P., Fuster, C., ... & Barberà, J. A. (2009). Long-term health effects of the *Prestige* oil spill (Galicia, Spain). *Epidemiology, 20*(6), 242–243.

Pertanika J. Sci. & Technol. 26 (1): 21 - 36 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Improving Cost and Time Control in Construction Using Building Information Model (BIM): A Review

M. M. Tahir[1], Haron, N. A.[1]*, Alias, A. H.[1], Harun, A. N.[2], I. B. Muhammad[3] and D. L. Baba[4]

[1]Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Department of Management of Technology, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 UTM, Kuala Lumpur, Malaysia
[3]Department of Architecture, School of Environmental Technology, Federal University of Technology, Minna, Nigeria
[4]Department of Project Management Technology, School of Entrepreneurship and Management Technology, Federal University of Technology, Minna, Nigeria

ABSTRACT

The control of cost and time in construction projects is one of the most important issues in construction since the emergence of the construction industry. A successful project should meet not only quality output standards, but also time and budget objectives. The management and control of cost and time in construction is fundamental in every project. An effective cost and time management and control technique for construction projects is important in managing risk of cost overrun and delay in completion of projects. Construction projects are becoming more complex as they now involve many stakeholders from different disciplines. The emergence of Building Information Model (BIM), an alternative technology is believed to solve issues related to project cost and time control as it efficiently increases collaboration between stakeholders. The aim of this paper is to review and summarise the causes of delay and cost overrun in construction industries, which are the main causes of disputes and abandonment of projects in the industry. It was found that delays and cost overrun eat deep into the industry and leave the construction industry with a bad image for decades even with rapid advancement in technology. The review of the applications of BIM showed that most of the applications are geared towards minimising construction cost and time spent on projects. This means that the use of BIM in the management of construction projects has great impact on project cost and time.

Keywords: BIM, cost control, delay and cost overrun, time control

ARTICLE INFO

Article history:
Received: 25 November 2016
Accepted: 04 July 2017

E-mail addresses:
mtahir1129@gmail.com (M. M. Tahir),
nuzul@upm.edu.my (Haron, N. A.),
aidihizami@upm.edu.my (Alias, A. H.),
aizulnahar.kl@utm.my (Harun, A. N.),
mibdoko@gmail.com (I. B. Muhammad),
dladi28@yahoo.com (D. L. Baba)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Control of cost and time in a construction project is one of the most important issues in construction since the emergence of the industry (Minchin et al., 2013). In this light, a successful project should not only meet quality output standards, but also time and budget objectives. Time and cost performance is a fundamental criterion for the success of any project. However, delay of project completion is very common in the construction industry due to ineffective cost and time control (Forbes & Ahmed, 2010). The essence of cost and time control is to ensure that projects are finished on time, and this is attainable through constant measurement of progress, evaluation of plans and taking appropriate action on the project (Kerzner, 2013). Buttressing this also is the Project Management Body of Knowledge (PMBOK), which suggests that in order to achieve the baseline objectives of any project, there has to be effective monitoring of project cost and time control (PMI, 2013).

Inherently, cost and time are two major concerns in managing construction projects (Rasdorf & Abudayyeh, 1991). However, construction projects are becoming more complex as they now involve many stakeholders from different disciplines. Most of the features of projects that give rise to delay and cost overruns do vary alongside with the project type, location, size and scope. Most of the time, construction projects that are large in nature and scope are characterised by their complexity and capital demands (Torp et al., 2016).

In addition, the construction industry has many branches, and as such, it encompasses a lot of information about any one construction project. This is information that is very important to a project, and can be the basic foundation for decision-making, procurement and collaboration. The success of a project requires cost management among other factors to be considered before the commencement of the project (Masrom et al., 2015). Cost management starts with quantification, which takes a lot of time and is tedious in nature. Traditionally, the process is manually completed most of the time with high likelihood of human error, which tends to be higher when preparing estimates for complex projects. The use of computer-related applications allow for the making of more reliable decisions (Martínez-Rojas et al., 2016).

Globally, the construction industry is replete with high-profile projects that are faced with significant delays and cost overrun (Smith, 2014). For example, in Saudi Arabia, it was found out that only 30% of construction projects were completed within the scheduled completion dates and that the average time overrun was between 10% and 30% (Assaf & Al-Hejji, 2006). Malaysia, a fast-developing country in Southeast Asia, is no exception from this global phenomenon. In 2005, about 17.3% of government contract projects in Malaysia were considered sick due to three months of delay and therefore, abandoned.

However, the construction industry in Malaysia plays a vital role in the nation's economic growth. It brings job opportunity and increment to the people's quality of life by providing essential socioeconomic infrastructure, such as offices, roads, houses and schools. Additionally, Malaysia is progressively marching towards industrialisation and the role of the construction industry is to enhance and at the same time, realise the needs and aspiration of its population (Alaghbari et al., 2007). Unfortunately, construction project delays and cost overrun are the norm in the Malaysian construction industry, leading to additional project costs (Enshassi et al., 2009).

The Malaysian construction industry is regarded as an industry facing poor performance leading to failure in achieving effective and efficient cost and time management (Ismail et al., 2013). In addition, the Chartered Institute of Building (CIOB) stated that in 2008 the quality of time management of construction projects was generally poor (Purnus & Bodea, 2013). Research conducted by CIOB in 2008 indicated that growth in training, education and skill levels within the construction industry in the use of time management techniques has not kept pace with the technology available (Zhang & Gao, 2013). As such, the emergence of alternative techniques is believed to minimise the issues relating to project cost and time control. Furthermore, it is also believed that the emergence of the Building Information Model (BIM) can lead to greater efficiency by means of increased collaboration (Zhang & Gao, 2013).

The National BIM Standard (NBIMS) Project Committee of the Building SMART alliance (2010) referred to BIM as

> A digital representation of physical and functional characteristics of a facility, as such, it serves as a shared knowledge resource for information about a facility, forming a reliable basis for decisions during its life-cycle from inception onward. The BIM is a shared digital representation founded on open standards for interoperability. (Succar, 2009)

This suggests the Building Information Model (BIM) as an alternative approach to construction design; it does not only make digital representation of designs easier, it also provides all the necessary information for any project before construction (Xiao & Noble, 2014). Thus, the information represented by BIM models is very useful and can be analysed to optimise the design, planning and construction processes (Azhar, 2011). Moreover, according to Bryde et al. (2013), BIM is an appropriate tool for project managers and should be considered by project managers as a way to help manage construction projects. However, BIM as an alternative technology in Malaysia needs to be studied to provide proof that it can satisfy the industry's need in improving cost and time control. It is also important to further investigate how cost and time control can be improved upon using BIM technology.

Project Time Control

Project time control is the process of monitoring the status of project activities, which include updating the project progress and managing changes to the schedule baseline to achieve the plan (PMI, 2013). This process is beneficial as it provides the means to recognise deviation from the actual plan in order to take corrective and preventive measures to reduce risk. The duration of time it takes for the execution of a project is, most of the time, very important to the parties involved (Dalu, 2012). However, project delays are common globally, especially in developing and underdeveloped countries (Nassar et al., 2005).

Project time control process. The process of project time control involves three main parameters, which are input, tools and techniques and output (PMI, 2013). The input are project management plans, the project schedule, work performance data, the project calendar and organisation process assets. The tools and techniques used in project time control are

performance reviews, project management software, resource optimisation techniques, modelling techniques, leads and lags, schedule compression and use of scheduling tools (Masrom et al., 2015), while the output that is expected from these processes are work performance information, schedule forecasts, change request, project management plan updates, project document updates and the organisational process assets updates (Yun et al., 2016).

Project Cost Control

According to Dalu (2012), project cost control is a vital ingredient for a successful project. Project cost control is the process of monitoring the status of the project by updating the project costs and managing changes to the cost baseline. This provides means for the recognition of variance from the plan so as to take corrective actions and reduce risk (PMI, 2013). This, therefore, means that corrective action needs to be taken by personnel who incur cost in all companies irrespective of their size.

Project cost control process. Similar to time control, the process for project cost control involves three main parameters, which are input, tools and techniques and output. The process and its parameters for input are project management plan, project funding requirements, work performance data and organisational process assets. The tools and techniques comprise earned value management (EVM), forecasting, to-complete performance index (TCPI), performance reviews, project management software and reserve analysis, while the output process and its parameters include work performance information, cost forecast, change request, project management plan updates, project document updates and organisational process updates (PMI, 2013).

Causes of Delay and Cost Overrun in Construction Projects

However, even with the vast integrated knowledge areas in project management, delay and cost overrun are the most common issues in construction projects (Francois, 2016). Thus, it is significant to note that the degree of success of a project is defined within the triangle of scope, time and cost. As such, it is important to look at the causes of these delays and cost overrun in projects. In this regard, delay is generally acknowledged as the most common, costly, complex and risky problem encountered in construction projects.

Previous research conducted has shown that project delays are common and costly, making it an important study to know the causes of these problems for effective project management (Frimpong et al., 2003; Koushki & Kartam, 2004; Koushki et al., 2005; Abdullah & Tawie, 2006; Abdul-Muhid, 2006; Ramanathan et al., 2012). As a result of the overriding importance of time for both the owner (in terms of performance) and the contractor (in terms of money), delays are undoubtedly the source of frequent disputes and claims leading to lawsuits. Delays occur in most construction projects; the magnitude of these delays varies considerably from project to project (Alaghbari et al., 2007; Enshassi et al., 2009).

Chan et al. (1997) stated that poor site management and supervision, unforeseen ground conditions, delays in making decisions by the project team and changes in scope caused by initiated and necessary variations of work by the project sponsors are the main causes of cost and time overrun in Hong Kong. In Jordan, Al-Momani (2000) examined 130 government projects and stated that changes that were initiated by the designers, weather, client requirement, late deliveries, site condition and economic conditions were the main causes of delay in construction projects.

According to Frimpong et al. (2003), in Ghana, material procurement, escalation of material prices, poor contractor management, poor technical performance and difficulties arising from agencies' monthly payment are the major causes of delay and cost overrun in the construction of groundwater projects. In Hong Kong, the major cause of delays and cost overrun were identified to be the poor management of sites, unforeseen ground conditions, change orders and poor decision making (Chan & Kumaraswamy, 1997, 2002). Furthermore, according to Odeh and Battaineh (2002), from the viewpoint of contractors and consultants in the construction industry, owner interference, inadequate contractor experience, finance and payment, labour productivity, slow decision making, improper planning and subcontractors are the major causes of delay in construction projects.

Hsieh et al. (2004) conducted a study on 90 metropolitan projects in Taiwan, and identified planning and design as the major cause of change orders in the projects leading to delay and cost overrun, while Sambasivan et al. (2007) stated that predominantly, delays are caused as a result of changes in design, poor planning and labour productivity. Moreover, in a study conducted by Kaliba et al. (2009), other prime causes of delay and cost overrun are client organisations delaying in making payment, modification of contracts, economic hardship, procurement materials, design changes, staffing issues, lack or unavailability of working equipment, poor supervision, mistakes during construction, poor site coordination, specification changes and labour.

Abd El-Razek et al. (2008) conducted a survey on the causes of delay and cost overrun and found that financial problems experienced during construction by contractors, owners delaying payment to contractors, changes in design by clients or their agents during construction and the lack of use of professional construction and contractual management were the main causes of delay and cost overrun. In addition, research by Le-Hoai et al. (2008) on the causes of delay in Vietnam compared with those of other countries showed that in Vietnam, lack of experience, loose deadlines, poor cost estimates, design inefficiencies, labour incompetence, government-related issues and financial problems were the main causes of delay, whereas in Thailand, the main causes of delay were poor design, resource and labour shortages, poor project planning, inefficient contractor management, financial difficulties and change orders (Toor & Ogunlana, 2008).

Delay and cost overrun do occur internally and are generated by the sponsors, design team, contractors and consultants. At the same time, they could be externally caused through late material supply, government policies or weather conditions (Ahmed et al., 2003). Olawale and Sun (2010) reported change in design as the most significant cause of poor time and cost

control from the practitioner's point of view. The five main causes of delay in construction projects according to Alinaitwe et al. (2013) comprise scope change, payment delays, high cost of capital, poor monitoring and control and political insecurity and instability. Inadequate experience by the contractor, poor project planning, site management and change orders were among the 15 main causes of delay in Turkey according to Gündüz et al. (2012).

Samarghandi et al. (2016) conducted a study on the reasons for delay and cost overrun in construction projects in Iran and developed a statistical model that categorised delay factors into four main categories. The categories include owners' defect, contractors' defect, consultants' defect and government laws, regulations and other general defects. Delay factors such as shortage of materials, change orders, delay in payment of suppliers, poor management of site and late submission of drawings are the main causes of delay (Abdul Kadir et al., 2005; Ramanathan et al., 2012). Alaghbari et al. (2007) conducted a study on the causes of delay and ranked them. The researcher found that financial problems and coordination problems were the two most important factors causing delay in construction projects in Malaysia.

Similarly, Sambasivan and Soon (2007) conducted research using a questionnaire to describe the 10 main causes of delay in Malaysian construction projects and found that they included poor site management, late payment, labour supply, improper planning, lack of experience, problems with subcontractors and shortage of materials. Al-Tmeemy et al. (2012) listed several causes of delay in Malaysia that included labour productivity, slow decision making, inflation, material delivery and insufficient equipment. Shehu et al. (2014) also stated that contract delays were predominantly caused by the contractors, while the other factors were associated with finance. Ahmed et al. (2003) and Al-aghbari (2005) in their separate research classified the factors causing delay and cost overrun in Malaysian construction projects into four categories, which are contractor's responsibility, consultant's responsibility, owner's responsibility and external factors.

The causes of delay and cost overrun are summarised in Table 1 below.

Table 1
Causes of delay and cost overrun

Causes of Delay and Cost Overrun	Sources
Scope change, payment delays, high cost of capital, poor monitoring and control and political insecurity and instability	Alinaitwe et al. (2013)
Client organisations delay in payment, modification of contracts, economic hardship, procurement materials, design changes, staffing issues, lack or unavailability of working equipment, poor supervision, mistakes during construction, poor site coordination, specification changes and labour disputes	Kaliba et al. (2009); Samarghandi et al. (2016); Olawale and Sun (2010)
Inadequate experience of the contractor, poor project planning, site management and change orders	Gündüz et al. (2012)
Designers, weather, client requirement, late deliveries, site condition and economic conditions	Al-Momani (2000)
Material procurement, escalation of material prices, poor contract management, poor technical performance, difficulties arising from agencies' monthly payment	Frimpong et al. (2003)

Table 1 *(continue)*

Owner interference, inadequate contractor experience, finance and payment, labour productivity, slow decision making, improper planning and subcontractors	Odeh and Battaineh (2002)
Poor planning and design	Hsieh et al. (2004)
Changes in design, poor planning and labour productivity.	Sambasivan et al. (2007)
Poor site management and supervision, unforeseen ground conditions, slow rate of making decisions by the project team, changes in scope of work by the project sponsors	Chan et al. (1997); Chan and Kumaraswamy (1997, 2002)
Financial difficulty during construction by contractor, owners delaying payment to contractors, changes in design by clients or their agents during construction, lack of use of professional construction and contractual management	Abd El-Razek et al. (2008)
Lack of experience, loose deadlines, poor cost estimates, design inefficiencies, labour incompetence, government-related issues and financial issues	Le-Hoai et al. (2008)
Poor design, resource and labour shortages, poor project planning, inefficient contractor management, financial difficulties and change orders	Toor and Ogunlana (2008)
Late material supplies, government policies, weather conditions	Ahmed et al. (2003)
Shortage of materials, change orders, delay in payment to suppliers, poor management of site and late submission of drawings	Abdul Kadir et al., 2005; Ramanathan et al., 2012
Financial issues and coordination	Alaghbari et al. (2007); Shehu et al. (2014)
Poor site management, late payment, labour supply, improper planning, lack of experience, problems with subcontractors and shortage of materials	Sambasivan and Soon (2007)
Low labour productivity, slow decision making, inflation, material delivery and insufficient equipment	Al-Tmeemy et al. (2012)
Adherence to outdated/old construction methods, lack of knowledge about different defined execution models, poor contract management by consultants, governmental inefficiencies, mistakes in technical documents, delays in producing design documents, delays in reviewing and approval of design documents by consultant and client	Odeh and Battaineh, (2002); Faridi and El-Sayegh, (2006); Assaf and Al-Hejji, (2006).

From the review above, one can conclude that the cause of delay and cost overrun have eaten deep into the industry and have given the construction industry a bad image for decades. Unfortunately, such causes keep recurring even with the advancement in technology and rigorous research done to solve or minimise their recurrence. It is believed that the application of BIM will solve these issues as it integrates all the major stakeholders from different disciplines in a project when making and taking decisions.

Concept of Building Information Model (BIM)

The contractor's guide to Building Information Modelling (BIM) describes it as a process of developing and implementing computer-generated models to combine the design, planning, construction and operation of a facility. In this light, Masood et al. (2014) defined BIM as a 3D digital representation of a facility with essential components and characteristics made up

of intelligent building components. It is a means and practice of virtual design construction throughout the facility life-cycle that serves as a platform for knowledge and data sharing for communication between stakeholders (Eadie et al., 2015). While Olatunji et al. (2010) defined it as a representation of the combination of fairly revolutionary ideals for design technology, it portrays the geometry, geographic information spatial relationships, quantities and characteristics of building elements, material inventories, cost estimates and schedule of performance.

BIM is basically a 3D digital representation of a facility. The model can be used to express the entire facility life-cycle. The model is data rich because of the quantity of material involved, its properties can be easily obtained and the scope of work required can easily be defined and isolated from the model (Smith & Edgar, 2008). Contract documents, drawings, procurement details, specifications and other construction documents can easily be interrelated using the model (Bazjanac, 2006; Khemlani, 2007). Thus, a series of techniques that enable the practice and processes of construction and virtual designs through the project's life-cycle is the main concept of BIM (Zhang, 2012).

The BIM industry working group (BIWG) (2011) defined the levels of BIM from 0-3. Level 0 refers to the unmanaged CAD, Level 1 is the managed CAD, while Level 2 is the managed 3D of the different disciplines. The third level is a fully-open process and data integration system enabled by web services compliant with emerging Industry Foundation Class/International Framework for Dictionaries (IFC/IFD) standards managed by a collaborative model server.

BIM tools. A lot of tools have been developed as a result of the spread of the concept of BIM in achieving its perspectives. These tools are used to manage construction projects, most of which are designed for specific purposes to meet the need of users, while a few are designed for multiple functions and information collection. The type of tool to be used depends on the purpose, user and stage in which it will be used. BIM tools enable 3D modelling and the management of information. The use of these tools makes BIM a unified system that interacts with all its parts. The table below shows some known BIM tools, their manufacturers and their functions.

Table 2
Building information model tool

Manufacturer	Tool	Function
AutoDesk	Navisworks	To manage 3D model-based design and clash detection
Bentley	Bentley Navigator	Dynamic coordination between models and disciplines
Vico Software	Vico Office	Analysis of various 3D models for coordination, scheduling and estimating
Gehry Technologies	Digital Project Suit	Full-featured suite: For design, review and information management
Tekla	Tekla Structures	3D structural modelling and detailing
Solibri	Solibri Model Checker	For quality assurance/quality control (QA/QC)
Synchro Ltd.	Synchro Professional	Scheduling of systems and planning simulations

Applications of BIM in Project Management

The potential of the application of BIM in the management of construction projects is similar to the Project Management Body of Knowledge (PMBOK); hence, it is an important tool for effective and efficient project management as it integrates stakeholders (Rokooei, 2015). BIM as a promising technology facilitates project management and the possibility of integrating building models and products, giving it high potential for management of project life-cycles (Gourlis & Kovacic, 2016) as it can be used throughout a project's life-cycle. It helps in understanding the project needs by the owner. It is also used for analysis, design and development of the project by the design team. The contractor also makes use of it in managing the construction phase, as well as in decommissioning, maintenance and operation carried out by the facility manager (Grilo & Jardim-Goncalves, 2010).

BIM is regarded as a great visualisation tool as it provides 3D virtual representation of the facility (Zhang, 2012). This is the reason why in recent times, researchers and practitioners in the industry have been using it as an alternative means of interdisciplinary information sharing; it enables users to have an idea of the functional and physical characteristics of the facility in 3D visualisation. Walk-through, rendering and sequence of the model can be provided by the project manager during project bidding for ease of communication with interested contractors. An outlook of the project, when completed, is provided through visualisation of the model using BIM techniques, which solve the issues of having to combine the different 2D views of the proposed project to create a 3D view (Mohandes et al., 2015). The virtual models enhance collaboration and communication as it can be shown to the owner and designers during meetings. The sequence of the construction work and planning can be based upon the utility of the model component. The virtual models, which are cost effective, allow the contractor and the design team to work on the constructability analysis of the building, thereby reducing risk and potential design errors and saving time (Azhar et al., 2008).

Value for money can be improved through the use of BIM, as was shown through research conducted by Li et al. (2014). The evidence provided to justify their finding was from the Shanghai Disaster Recovery Centre project, which showed a potential benefit of optimised construction activities through cost-orientated activities. Moreover, BIM reduces waste and also optimises efficiency throughout the project life-cycle as it supports integrated project delivery through a collaborative process (Glick & Guggemos, 2009).

Lavy et al. (2014) conducted research using a case study of the Solibri Model Checker (SMC), and it was found out that maintainability of the facility could be checked during the design phase (i.e. using BIM), which will reduce the cost of maintenance during a project's life-cycle. This is possible as the 3D model offers facility managers the opportunity to anticipate maintenance accessibility problems and ways to resolve them.

Lu et al. (2014) conducted a study on the cost benefit of implementing BIM by comparing two projects i.e. one using BIM and the other, conventional methods. It was found that the implementation of BIM saved cost by about 7% per square metre for the project. At the construction stage, it decreased the cost per square metre of the gross floor area by 8.61% when compared to the conventional designed and built project. According to Lu et al. (2016), cash flow analysis can be automated and simplified through 3D modelling design by linking

cost and schedule information (i.e. 5D BIM); this saves more time and cost compared with using the traditional method, which is time consuming. On the contrary, Masood et al. (2014) stated that BIM has very low impact on cost reduction, time and human resource in Pakistan's construction industry.

According to Underwood (2009), BIM presents a non-redundant model of the project's life-cycle information to streamline its processes, which solves the problem of redundancy when using conventional methods. Improvement in quality of design, construction and minimising rework during construction are the three most important merits of using BIM. This was the perception of AEC professionals in Pakistan's construction industry (Masood et al., 2014).

The applications of BIM in project management are summarised in Table 3.

Table 3
Building information model applications

Application of BIM	Source
Quality compliance management	Chen and Luo (2014)
Reduction in Information Exchange (IE) waste	Dubler et al. (2010)
Improvement in the accessibility of facility management data	Kassem et al. (2015); Liu (2010); Meadati et al. (2010)
Sustainability of project design and building performance	Wong and Fan (2013)
Clash detection and coordination	Azhar et al. (2008); Foster (2008); Young et al. (2009); Arayici et al. (2011); Lahdou and Zetterman (2011)
Automated safety checking platform	Zhang and Gao (2013)
Constructability analysis	Foster (2008)
Visualisation and sequencing of activities	Tulke and Hanff (2007); Wilson and Koehn (2000); Ding et al. (2014)
Cost estimation of material and quantities	Azhar et al. (2008); Hergunsel (2011); Sabol (2008)
Integration of key stakeholders	Foster (2008)
Optimisation of prefabricated construction components	Hergunsel (2011); Winberg and Dahlqvist (2010)
Risk assessment of design component of facility for prevention through design	Kamardeen (2010)
Scope clarification	Bryde et al. (2013)

Table 3 shows that all the applications and the benefits from the applications of BIM are geared towards minimising construction cost and time spent on projects. This has shown that the use of BIM in the management of construction projects has great impact on project time and cost. Initial cost for the implementation of BIM might cost much, but it will increase the profitability of the company in the long run. The applications of BIM such as clash detection, quantity take-off, design and visualisation play an important role in the management of construction projects. The most common causes of delay and cost overrun arise from poor designs, inaccurate

estimates, mistakes and errors during construction due to clashes in designs. These issues can be minimised as there are reported benefits in the application of BIM that solve these issues by researchers.

CONCLUSION

The review highlighted the main causes of delay and cost overrun in the construction industry. It is generally acknowledged that the end result of delays and cost overrun is costly and risky, leading to many disputes and claims that end in lawsuits. Most of the causes of delay and cost overrun as stated by previous researchers can be classified into four categories. These are contractor's responsibility, consultant's responsibility, owner's responsibility and external factors. It is believed that the evolution of the Building Information Model (BIM) will be able to increase efficiency and quality of output in the construction industry by eliminating these causes of delay and cost overrun. The application of BIM and the potential benefits of its application were also reviewed. The benefits from using these applications are expected to minimise delay and cost overrun as they provide solutions for the main causes of delay and cost overrun such as estimation, clash detection, integration and many more.

REFERENCES

Abd El-Razek, M., Bassioni, H., & Mobarak, A. (2008). Causes of delay in building construction projects in Egypt. *Journal of Construction Engineering and Management, 134*(11), 831–841.

Abdul Kadir, M., Lee, W., Jaafar, M., Sapuan, S., & Ali, A. (2005). Factors affecting construction labour productivity for Malaysian residential projects. *Structural Survey, 23*(1), 42–54.

Abdullah, F., & Tawie, S. (2006). Abdullah peeved over delayed prisons projects. *New Straits Times*. Retrieved from http://rakan.jkr.gov.my/ups/krtnAkhbar/pdfAkhbar/NST% 2018%20APR.pdf

Abdul-Muhid, N. (2006). Lagi sekolah gagal disiapkan. *Utusan Malaysia*. Retrieved from www.utusan.com.my/utusan_Malaysia&sec¼Muka_Hadapan&pg¼mh_02.htm.

Ahmed, S. M., Azhar. S., Kappagntula, P., & Gollapudil, D. (2003). *Delays in construction: A brief study of the Florida construction industry.* Paper presented at the Proceedings of the 39th Annual ASC Conference, Clenson University, Clenson, SC.

Al-Aghbari, M. (2005). *Factors affecting construction speed of industrialized building system in Malaysia*. Serdang, Malaysia: Universiti Putra Malaysia.

Alaghbari, W. E., Razali A., Kadir, M., Salim, A., & Ernawati. (2007). The significant factors causing delay of building construction projects in Malaysia. *Engineering, Construction and Architectural Management, 14*(2), 192–206.

Alinaitwe, H., Apolot, R., & Tindiwensi, D. (2013). Investigation into the causes of delays and cost overruns in Uganda's public sector construction projects. *Journal of Construction in Developing Countries, 18*(2), 33.

Al-Momani, A. H. (2000). Construction delay: A quantitative analysis. *International Journal of Project Management, 18*(1), 51–59.

Al-Tmeemy, S. M. H., Abdul-Rahman, H., & Harun, Z. (2012). Contractors' perception of the use of costs of quality system in Malaysian building construction projects. *International Journal of Project Management, 3*0(7), 827–838.

Arayici, Y., Coates, P., Koskela, L., Kagioglou, M., Usher, C., & O'reilly, K. (2011). Technology adoption in the BIM implementation for lean architectural practice. *Automation in Construction, 20*(2), 189–195.

Assaf, S. A., & Al-Hejji, S. (2006). Causes of delay in large construction projects. *International Journal of Project Management, 24*(4), 349–357.

Azhar, S. (2011). Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry. *Leadership and Management in Engineering, 11*(3), 241–252.

Azhar, S., Hein, M., & Sketo, B. (2008). *Building Information Modeling (BIM): Benefits, risks and challenges*. Auburn, Alabama: McWhorter School of Building Science, Auburn University.

Bazjanac, V. (2006). Virtual building environments (VBE)-applying information modeling to buildings. In A. Dikbas & R. Scherer (Eds.), *eWork and eBusiness in Architecture Engineering and Construction* (pp. 58-72). London: Taylor & Francis Group.

BIM Industry Working Group (BIWG). (2011). A report for the government construction client group building information modelling (BIM). *Working party strategy paper*. Communications. London, UK.

Bryde, D., Broquetas, M., & Volm, J. M. (2013). The project benefits of building information modelling (BIM). *International Journal of Project Management, 31*(7), 971–980.

Chan, D. W., & Kumaraswamy, M. M. (1997). A comparative study of causes of time overruns in Hong Kong construction projects. *International Journal of Project Management, 15*(1), 55–63.

Chan, D. W., & Kumaraswamy, M. M. (2002). Compressing construction durations: Lessons learned from Hong Kong building projects. *International Journal of Project Management, 20*(1), 23–35.

Chen, L., & Luo, H. (2014). A BIM-based construction quality management model and its applications. *Automation in Construction, 46*, 64–73.

Ding, L., Zhou, Y., & Akinci, B. (2014). Building information modeling (BIM) application framework: The process of expanding from 3D to computable nD. *Automation in Construction, 46*, 82–93.

Dubler, C., Kreider, R., & Messner, J. (2010). Determining the frequency and impact of applying BIM for different purposes on projects. In *Proceedings 6th International Conference on Innovation in Architecture, Engineering and Construction (AEC)*. University Park, PA.

Eadie, R., Browne, M., Odeyinka, H., McKeown, C., & McNiff, S. (2015). A survey of current status of and perceived changes required for BIM adoption in the UK. *Built Environment Project and Asset Management, 5*(1), 4–21.

Enshassi, A., Al-Najjar, J., & Kumaraswamy, M. (2009). Delays and cost overruns in the construction projects in the Gaza strip. *Journal of Financial Management of Property and Construction, 14*(2), 126–151.

Faridi, A. S., & El-Sayegh, S. M. (2006). Significant factors causing delay in the UAE construction industry. *Construction Management and Economics, 24*(11), 1167–1176.

Forbes, L. H., & Ahmed, S. M. (2010). *Modern construction: Lean project delivery and integrated practices.* Boca Raton: CRC Press.

Foster, L. L. (2008). *Legal issues and risks associated with Building Information Modeling Technology.* (Doctoral dissertation). University of Kansas.

Francois Buys| LinkedIn. (n. d.). Retrieved August 23, 2016, from https://www.linkedin.com/in/francois-buys-77592116?trk=pulse-det-athr_prof-art_ftr

Frimpong, Y., Oluwoye, J., & Crawford, L. (2003). Causes of delay and cost overruns in construction of groundwater projects in developing countries: Ghana as a case study. *International Journal of Project Management, 21*(5), 321–326.

Glick, S., & Guggemos, A. (2009). IPD and BIM: Benefits and opportunities for regulatory agencies. In *Proceedings of the 45th Associated Schools of Construction National Conference* (pp. 2-4). Gainesville, Florida.

Gourlis, G., & Kovacic, I. (2016). Building Information Modelling for analysis of energy efficient industrial buildings – A case study. *Renewable and Sustainable Energy Reviews, 68*, 953-963.

Grilo, A., & Jardim-Goncalves, R. (2010). Value proposition on interoperability of BIM and collaborative working environments. *Automation in Construction, 19*(5), 522–530.

Gündüz, M., Nielsen, Y., & Özdemir, M. (2012). Quantification of delay factors using the relative importance index method for construction projects in Turkey. *Journal of Management in Engineering, 29*(2), 133–139.

Hergunsel, M. F. (2011). *Benefits of building information modeling for construction managers and BIM based scheduling*. (Master Thesis). Worcester Polytechnic Institute.

Hsieh, T. Y., Lu, S. T., & Wu, C. H. (2004). Statistical analysis of causes for change orders in metropolitan public works. *International Journal of Project Management, 22*(8), 679–686.

Ismail, I., Abdul Rahman, I., Memon, A. H., Karim, A., & Tarmizi, A. (2013). Comparative study on time management practices in construction industry between Kedah and Kelantan. In *Proceedings the 2nd International Conference on Global Optimization and Its Applications 2013 (ICoGOIA2013)*. Melaka, Malaysia.

Kaliba, C., Muya, M., & Mumba, K. (2009). Cost escalation and schedule delays in road construction projects in Zambia. *International Journal of Project Management, 27*(5), 522–531.

Kamardeen, I. (2010). 8D BIM modelling tool for accident prevention through design. In *26th Annual ARCOM Conference, Association of Researchers in Construction Management* (vol. 1, pp. 281-289). Leeds, United Kingdom.

Kassem, M., Kelly, G., Dawood, N., Serginson, M., & Lockley, S. (2015). BIM in facilities management applications: A case study of a large university complex. *Built Environment Project and Asset Management, 5*(3), 261–277.

Kerzner, H. R. (2013). *Project management: A systems approach to planning, scheduling, and controlling.* New Jersey: John Wiley & Sons.

Khemlani, L. (2007). Top criteria for BIM solutions. *A survey conducted by AECbytes*.

Koushki, P., Al-Rashid, K., & Kartam, N. (2005). Delays and cost increases in the construction of private residential projects in Kuwait. *Construction Management and Economics, 23*(3), 285–294.

Koushki, P. A., & Kartam, N. (2004). Impact of construction materials on project time and cost in Kuwait. *Engineering, Construction and Architectural Management, 11*(2), 126–132.

Lahdou, R., & Zetterman, D. (2011). BIM for project managers – How project managers can utilize BIM in construction projects. (Master Thesis). Chalmers University of Technology, Sweden.

Lavy, S., Liu, R., & RA Issa, R. (2014). *Design for maintenance accessibility using BIM tools*. Facilities, 32(3/4), 153–159.

Le-Hoai, L., Dai Lee, Y., & Lee, J. Y. (2008). Delay and cost overruns in Vietnam large construction projects: A comparison with other selected countries. *KSCE Journal of Civil Engineering, 12*(6), 367–377.

Li, J., Hou, L., Wang, X., Wang, J., Guo, J., Zhang, S., & Jiao, Y. (2014). A project-based quantification of BIM benefits. *International Journal of Advanced Robotic Systems, 11*(8), 123.

Liu, Z. (2010). *Feasibility Analysis of BIM based information system for facility management at WPI.* (Doctoral Dissertation). Worcester Polytechnic Institute.

Lu, Q., Won, J., & Cheng, J. C. (2016). A financial decision making framework for construction projects based on 5D Building information modeling (BIM). *International Journal of Project Management, 34*(1), 3–21.

Lu, W., Fung, A., Peng, Y., Liang, C., & Rowlinson, S. (2014). Cost-benefit analysis of building information modeling implementation in building projects through demystification of time-effort distribution curves. *Building and Environment, 82*, 317–327.

Martínez-Rojas, M., Marín, N., & Miranda, M. A. V. (2016). An intelligent system for the acquisition and management of information from bill of quantities in building projects. *Expert Systems with Applications, 63*, 284–294. doi: http://dx.doi.org/10.1016/j.eswa.2016.07.011

Masood, R., Kharal, M., & Nasir, A. (2014). Is BIM adoption advantageous for construction industry of Pakistan? *Procedia Engineering, 77*, 229–238.

Masrom, M. A. N., Rahim, M. H. I. A., Mohamed, S., Chen, G. K., & Yunus, R. (2015). Successful criteria for large infrastructure projects in Malaysia. *Procedia Engineering, 125*, 143–149. doi: http://dx.doi.org/10.1016/j.proeng.2015.11.021

Meadati, P., Irizarry, J., & Akhnoukh, A. K. (2010). *BIM and RFID integration: A pilot study*. In BIM and RFID integration: A pilot study (pp. 570–578). Cairo, Egypt.

Minchin, R. E., Li, X., Issa, R. R., & Vargas, G. G. (2013). Comparison of cost and time performance of design-build and design-bid-build delivery systems in Florida. *Journal of Construction Engineering and Management, 139*(10), 04013007.

Mohandes, S. R., Marsono, A. K., Omrany, H., Faghirinejadfard, A., & Mahdiyar, A. (2015). Comparison of building existing partitions through building information modeling (BIM). *Jurnal Teknologi, 75*(1), 287-298.

Nassar, K. M., Gunnarsson, H. G., & Hegab, M. Y. (2005). Using Weibull analysis for evaluation of cost and schedule performance. *Journal of Construction Engineering and Management, 131*(12), 1257–1262.

Odeh, A. M., & Battaineh, H. T. (2002). Causes of construction delay: Traditional contracts. *International Journal of Project Management, 20*(1), 67–73.

Olatunji, O., Sher, W., Gu, N., & Ogunsemi, D. (2010). Building information modelling processes: benefits for construction industry. *Paper presented at the W078-Special Track 18th CIB World Building Congress.* Salford, United Kingdom.

Olawale, Y. A., & Sun, M. (2010). Cost and time control of construction projects: Inhibiting factors and mitigating measures in practice. *Construction Management and Economics, 28*(5), 509–526.

Purnus, A., & Bodea, C. N. (2013). Considerations on project quantitative risk analysis. *Procedia-Social and Behavioral Sciences, 74*, 144–153.

Ramanathan, C., Potty, N. S., & Idrus, A. B. (2012). Analysis of time and cost overrun in Malaysian construction. In *Advanced Materials Research* (vol. 452, pp. 1002-1008). Trans Tech Publications.

Rasdorf, W. J., & Abudayyeh, O. Y. (1991). Cost-and schedule-control integration: Issues and needs. *Journal of Construction Engineering and Management, 117*(3), 486–502.

Rokooei, S. (2015). Building information modeling in project management: necessities, challenges and outcomes. *Procedia-Social and Behavioral Sciences, 210*, 87–95.

Sabol, L. (2008). Challenges in cost estimating with building information modeling. *IFMA World Workplace.* The power of process in the built environment.

Samarghandi, H., Tabatabaei, S. M. M., Taabayan, P., Mirhashemi, A., & Willoughby, K. (2016). Studying the reasons for delay and cost overrun in construction projects: The case of Iran. *Journal of Construction in Developing Countries* (in press).

Sambasivan, M., & Soon, Y. W. (2007). Causes and effects of delays in Malaysian construction industry. *International Journal of Project Management, 25*(5), 517–526.

Shehu, Z., Endut, I. R., & Akintoye, A. (2014). Factors contributing to project time and hence cost overrun in the Malaysian construction industry. *Journal of Financial Management of Property and Construction, 19*(1), 55–75.

Smith, D. K., & Edgar, A. (2008). *Building information modeling (BIM)*. Washington, U.S.A.: National Institute of Building Sciences.

Smith, P. (2014). Project cost management – Global issues and challenges. *Procedia-Social and Behavioral Sciences, 119*, 485–494.

Succar, B. (2009). Building information modelling framework: A research and delivery foundation for industry stakeholders. *Automation in Construction, 18*(3), 357–375.

Toor, S. U. R., & Ogunlana, S. O. (2008). Problems causing delays in major construction projects in Thailand. *Construction Management and Economics, 26*(4), 395–408.

Torp, O., Belay, A. M., Thodesen, C., & Klakegg, O. J. (2016). Cost development over-time at construction planning phase: Empirical evidence from Norwegian construction projects. *Procedia Engineering, 145*, 1177–1184. doi: http://dx.doi.org/10.1016/j.proeng.2016.04.152

Tulke, J., & Hanff, J. (2007). 4D construction sequence planning–new process and data model. In *Proceedings of CIB-W78 24th International Conference on Information Technology in Construction*. Maribor, Slovenia.

Underwood, J. (2009). *Handbook of research on building information modeling and construction informatics: Concepts and technologies.* United States of America, USA: IGI Global.

Wilson, J. M., & Koehn, E. E. (2000). Safety management: problems encountered and recommended solutions. *Journal of Construction Engineering and Management, 126*(1), 77-79.

Winberg, A., & Dahlqvist, E. (2010). BIM – *The next step in the construction of civil structures*. (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-36993

Wong, K.-D., & Fan, Q. (2013). Building information modelling (BIM) for sustainable building design. *Facilities, 31*(3/4), 138–57.

Xiao, H., & Noble, T. (2014). BIM's impact on the project manager. In *Proceedings of the 30th Annual ARCOM Conference* (pp. 693-702). Portsmouth, UK.

Young, N., Jones, S. A., Bernstein, H. M., & Gudgel, J. (2009). *The business value of BIM – Getting building information modeling to the bottom line* (p. 51). Bedford, Massachusetts: McGraw-Hill Construction.

Yun, S., Choi, J., Oliveira, D. P., Mulva, S. P., & Kang, Y. (2016). Measuring project management inputs throughout capital project delivery. *International Journal of Project Management, 34*(7), 1167–1182. doi: http://dx.doi.org/10.1016/j.ijproman.2016.06.004

Zhang, D. (2012). *Project time and cost control using building information modeling* (Master's Thesis). North Dakota State University of Agriculture and Applied Science, North Dakota, U.S.A. ProQuest LLC, (1550199).

Zhang, D., & Gao, Z. (2013). Project time and cost control using building information modeling. *ICCREM 2013*, 545–554.

Pertanika J. Sci. & Technol. 26 (1): 37 - 70 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Android Botnets: A Serious Threat to Android Devices

Shahid Anwar[1]*, Mohamad Fadli Zolkipli[1], Zakira Inayat[2,3], Julius Odili[1], Mushtaq Ali[1] and Jasni Mohamad Zain[4]

[1]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, 26300 UMP, Gambang, Malaysia
[2]Department of Computer Science, University of Engineering and Technology Peshawar, Peshawar 2500, Pakistan
[3]Center for Mobile Cloud Computing Research, University of Malaya, 50603 UM, Kuala Lumpur, Malaysia
[4]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

ABSTRACT

Android devices have gained a lot of attention in the last few decades due to several reasons including ease of use, effectiveness, availability and games, among others. To take advantage of Android devices, mobile users have begun installing an increasingly substantial number of Android applications on their devices. Rapid growth in many Android devices and applications has led to security and privacy issues. It has, for instance, opened the way for malicious applications to be installed on the Android devices while downloading different applications for different purposes. This has caused malicious applications to execute illegal operations on the devices that result in malfunction outputs. Android botnets are one of these malfunctions. This paper presents Android botnets in various aspects including their security, architecture, infection vectors and techniques. This paper also evaluates Android botnets by categorising them according to behaviour. Furthermore, it investigates the Android botnets with respect to Android device threats. Finally, we investigate different Android botnet detection techniques in depth with respect to the existing solutions deployed to mitigate Android botnets.

Keywords: Android botnets, malware, detection techniques, DDoS attacks, mobile security

ARTICLE INFO

Article history:
Received: 06 December 2016
Accepted: 27 September 2017

E-mail addresses:
shahidanwar.safi@gmail.com (Shahid Anwar),
fadli@ump.edu.my (Mohamad Fadli Zolkipli),
zakirainayat@uetpeshawar.edu.pk (Zakira Inayat),
odili_julest@yahoo.com (Julius Odili),
sunnygul1@gmail.com (Mushtaq Ali),
jasni@tmsk.uitm.edu.my (Jasni Mohamad Zain)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Mobile devices (mobile devices/android devices are used interchangeably in this article) are gaining popularity in the 21st century (Narudin, Feizollah, Anuar, & Gani,

2016). These devices offer a host of advanced capabilities and ample storage of large volumes of personal and confidential data (Peng, Yu, & Yang, 2014). Nowadays, most mobile devices offer more computing capabilities and memory storage than many personal computers did a few years back. According to former Android boss Andy Rubin, "There should be nothing that users can access on their desktop that they cannot access on their cellphone" (Rubin, 2008). Any mobile device has three core features, such as Applications, Storage and Connectivity. These key features make Android devices an attractive tool for malware writers to attack organsation/ individual devices. While protecting stored data on these devices is crucial against today's threats, most mobile devices use the Android operating system due to its open nature. Android provides a full set of software for Android devices including operating system, middleware and key Android applications (Sears, 2007). The open nature of Android devices make these devices an attractive source for cybercriminals and over the years there has been a number of threats faced by mobile devices such as spyware, botnet, vulnerable applications, privacy threats, drive-by-download, phishing scams, malware, network exploits, browser exploits and Wi-Fi sniffing (Fossi et al., 2011; Inayat, Gani, Anuar, Khan, & Anwar, 2016), Botnet is one of the most dangerous threats faced by mobile devices recently.

Malware is used to damage Internet-connected devices and gather sensitive information from individuals or it uses spyware for accessing the most private information on the infected device (Sharma, Chawla, & Gajrani, 2016). Spyware gathers all this information specifically for advertising purposes (Sheta, Zaki, El Salam, & Hadad, 2015). Privacy threats can be caused by those Android applications that may not be malicious by nature but use sensitive information obtained illegally from unsuspecting Android users. Vulnerable applications are those that contain deficiencies that may cause malicious attacks and malicious activities. Phishing scams are those that use the victim's device emails for sending the virus infected links to the Internet-connected devices (Naraine, 2012). In drive-by-download, the infected devices download an application when they access a website. While the browser-exploits benefits from the vulnerabilities in mobile device web browsers or applications launched by the browser such as flash player, PDF reader and much more, in network exploits, cybercriminals take advantage of Android operating system flaws for criminal activities (Naser, Zolkipli, Majid, & Anwar, 2014). When the data are transferred from one device to another connected by Wi-Fi as many applications do not use proper security rules, this results in data obstruction known as Wi-Fi sniffing. In this article, we focus on the Android botnet.

A botnet (Robot Network) is a type of malware that enables the infected devices to perform criminal activities according to the botmaster's instructions (Anwar, Zain, Inayat, Haq, Karim, & Jaber, 2016; Naser et al., 2014; Peng et al., 2014). A malicious Android application is installed in a susceptible host that is capable of carrying out a series of different harmful activities to the end user according to the botmaster's instructions. These applications can be downloaded to the victims' devices using different methods. The most common ways to infect a victim's device includes access to the infected websites, drive-by-download, spam emails, viral mechanism and much more (Anwar, Zain, Zolkipli, Inayat, Khan, Anthony, & Chang, 2017; Karim, Shah, Salleh, Arif, Md Noor, & Shamshirband, 2015). Once an end-user's device is infected with malicious software, it receives instructions from the cybercriminal (botmaster) through a command and control server using communication channels. Botmaster is the entity

that performs criminal activities from these bot devices, while a communication channel is the way through which a botmaster can communicate with the C&C server and bots. A bot can be a servant and a client as well at the same time. It can propagate themselves to infect vulnerable hosts (Silva, Silva, Pinto, & Salles, 2013).

To the best of our knowledge, this paper aims to present the Android botnet from first appearance. We aim to guide interested readers and researchers on Android botnets and detection techniques. This paper organises the Android botnet detection techniques with respect to their benefits and limitations; understanding this information can improve Android botnet detection techniques.

The key contributions of this survey paper are:

- It provides up-to-date information on mobile device threats: We provide comprehensive details of the possible threats to mobile devices. We have also categorised these threats in sub-groups according to their nature.
- It provides exhaustive information about Android botnets: We provide in-depth information about Android botnets, their background and timeline.
- We provide in-depth information about Android botnet detection techniques: This paper presents detailed information about Android botnet detection techniques. We explain these techniques regarding their benefits and limitations. These limitations are also explained in more detail in table form.
- We introduce future research challenges: We suggest potential research areas for Android botnet detection techniques and we highlight the challenges present in Android botnet detection techniques as well.

Classification of Mobile Device Threats

There are diverse types of threat to mobile devices that may badly affect mobile devices, such as viruses and spyware that can infect personal computers (PC). These threats can be divided into four broad categories: Application-Level, Web-Level, Network-Level and Physical-Level (see Figure 1).

Application-Level Threats

Application-level threats are based on the Applications, which are the core feature of every mobile device. These threats appear to be the most widely discussed threats in the literature, which presents application-level threats as the most widely discussed threat. Since the applications that run on these mobile devices are available from third-party markets, it is clear that they can be target vectors for mobile device security breaches (Faruki et al., 2015). Malware are Android applications that perform malicious activities can inject malicious codes into mobile device that send unsolicited messages and allow an adversary the ability to remotely control the device.

Malware. Malware is short for ‘malicious software’. This is specifically developed to damage machines on which they are executed or the network on which it communicates (Inayat, Gani,

Anuar, Anwar, & Khurram Khan, 2017; Preda, Christodorescu, Jha, & Debray, 2008). Malware is mostly installed on victims' devices to perform illegal activities without the knowledge of the owner. The range of malware varies; it can be as simple as pop-up advertising or so dangerous that it causes machine invasion or damage. Stealing owner-sensitive credentials and infecting new vulnerable devices are the main targets of malware. The most common malware is found as financial, crypto locker and advertisement malware (Anwar, Zain, Zolkipli, Inayat, Jabir, & Odili, 2015).

Financial malware is developed for scanning mobile devices to gather financial information, while crypto locker malware is used in cyber-criminal activities. According to the Symantec report published in 2013, ransomware evolves regularly in the Android operating system. Compared with other OS on mobile devices, Android is most frequently attacked because of its open nature (Narudin et al., 2016; Odili, Kahar, & Anwar, 2015; Teufl et al., 2013). Ransomware allows cybercriminals to hijack the victim's device, encrypt the victim's private files and then demand a ransom from the victim in order for the files to be released (Anwar et al., 2017). Malicious spyware is considered a significant threat to the confidentiality of mobile devices (Sheta, Zaki, El Salam, & Hadad, 2015). It covertly collects confidential data from the infected device and sends it to the cybercriminal (Botmaster) through the user's Internet connection without the owner's knowledge. These applications mostly contain freeware or shareware, which can be downloaded from third-party markets. Adware is another type of malware. It is a software package that automatically displays related advertisements to the mobile device users based on the user's pattern of web surfing. These advertisements may be present during the installing phase of any Android application, or they are present when an end-user is using these applications. This adware engages in collecting confidential information, frequently by user consent, while stealing this personal information for covert harmful activities.

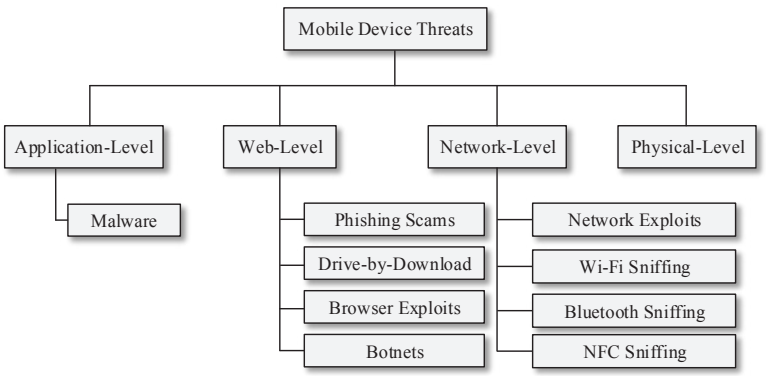


Figure 1. Threats to mobile device security

Web-Level Threats

The security and privacy threats to mobile devices from webs happens normally. The most dangerous web-level threats are phishing scams, drive-by-download and browser exploits. Phishing scams are the key web-level threat, which uses email or other social media apps

to send an unwitting user links to a phishing website designed to trick users into providing sensitive information such as user credentials. Phishing is one of the top seven security threats identified by the Kaspersky lab (Kaspersky, 2015). However, botware is the most dangerous threat to mobile devices nowadays. These are software programmes created to automatically perform specific operations.

Phishing scams. Phishing refers to the criminal action of generating a replica of web pages that exist to fool a mobile user entering private, extremely sensitive credentials, financial or online banking information and passwords (Alta, Loock, & Dabrowski, 2005). Phishing is a technique of attacking to obtain personal information from a mobile device user and is the main cause of various problems encountered by Internet users. This technique can cost the victim financially. Phishing is performed through instant messenger phishing, voice phishing, and flash phishing (Dunne, 2006; Milletary & Center, 2005). After launching a phishing attack on an individual or an organisation, the employees of the organisation handle the customer when he calls after losing his money.

Drive-by-Download. A drive-by-download refers to potentially harmful software code that is installed on a person's computer without the user's permission; the user may not even be aware that the software has been installed. Drive-by-downloads are a form of malware typically found on compromised web pages. By simply 'driving by', or visiting the web page, the drive-by-download begins to download and is then installed in the background on the computer or mobile device without alerting the user (Naraine, 2012).

Browser exploits. This is a malicious code that uses a piece of software or operating vulnerabilities to breach the security of the browser. Browser exploits perform these malicious activities without informing the owner of the device.

Botnets. Short for robot network, botnet, is the network of Internet-connected infected-devices (bots) under the control of a botmaster (cybercriminal) to perform cyber-criminal activities without the knowledge of the device owner (Anwar, Mohamad Zain, Zolkipli, & Inayat, 2014). There are two types of botnet: traditional botnets and mobile botnets. This paper focusses on mobile (Android) botnets. The purpose of Android botnets will most likely be similar to those of existing traditional botnets (e.g. providing means of DoS, DDoS and spam distribution); however, the targets are different (Enck, Ongtang, & McDaniel, 2009). In mobile botnets, the targets are mobile devices.

A common botnet having thousands of infected victims is called a bot (zombie). The botmaster sends instructions to all online bots to send queries to a particular system/server (Mirkovic & Reiher, 2004). By attacking a new victim from thousands of different bots in a botnet, the DoS (DDoS) is distributed. In a DDoS attack, the bot becomes harder to detect and it is difficult for cyber law enforcement to prevent DDoS attacks. Some DDoS attacks include UDP flood attacks, Zero-day DDoS attacks, Sync flood attacks, ICMP flood attacks, Slowloris and Ping of Death (LulzSec, 2011; Zang, Tangpong, Kesidis, & Miller, 2011). DDoS attacks are performed using diverse types of tool, such as agent- and IRC-based tools. These

attacks can be detected through screening of the time interval of requests and bandwidth size. Some DDoS attacks, such as Zero-day attacks, are unknown or new and thus, have no patch yet. The term DDoS is well-known among hackers as dealing with Zero-day vulnerabilities is a common activity.

Network-Level Threats

Any mobile device has three core features: applications, storage and connectivity. Network-level threats can occur due to mobile device connectivity with the cellular/mobile networks, local wireless networks or near field-communication (NFC). Network exploits, Wi-Fi sniffing, Bluetooth and NFC are the main types of network-level threats.

Network exploits. Network exploits take advantage of flaws in the mobile operating system or other software that operates on local or cellular networks, such as an International Mobile Subscriber Identity (IMSI) catcher. Once connected, they can intercept data connections and find a way to inject malicious software on users' phones without their knowledge.

Wi-Fi sniffing. Wi-Fi sniffing seizes data when they are traveling between the device and the Wi-Fi access point. Most Android applications do not use proper security measures while sending unencrypted data across the network. A cybercriminal can easily read the data as they travel. Public sites such as coffee shops, restaurants and bookstores may have WPA2, but it is likely that anyone with the password can decrypt your packets.

Bluetooth. People who leave BT on all the time leave themselves vulnerable to pairing from nefarious devices and the uploading of spyware. Blue jacking is an older-style attack in which a Bluetooth enabled device that is active is used by someone else. Blue jacking refers to the sending of unsolicited data (vCards etc.) to open Bluetooth listeners in the area. It has more recently been used for marketing, but many more modern smartphones are less vulnerable to Bluetooth stack exploits. This can lead to phishing attempts and the spread of malware or viruses.

Near field communication (NFC). Advanced mobile devices contain near field communication (NFC) as a medium for communication. NFC is a newly developed wireless technology that provides communication between two mobile devices, both of which must contain NFC tags using short-range radio waves. NFC enables the exchange of images, apps and other data between two devices without first pairing them. For this purpose, both devices use a feature that Google calls Android Beam (Sauter, 2013), while Beam is Android's trademark for NFC when the protocol is used for device-to-device communication. In the NFC communication, only two devices can communicate, such as the initiator and target. The initiator sends data, while the target receives them; both devices are active during the communication, consuming their own battery power. NFC provides extra opportunities to the attacker to compromise NFC-enabled devices, such as Wi-Fi and Bluetooth. So far, mobile threats are still mainly aimed at consumers rather than at enterprises.

Physical-Level Threats

Physical-level threats are more important than other mentioned threats. Since mobile devices are small, portable and valuable, it makes their physical security more important. Stealing and misplacing devices are the common issue among users of these devices. These devices are valuable not only because they are resold in the black market, but more importantly, because they contain sensitive organisational and personal data. Most mobile device users use their phone for banking, social communication and much more, while end-user are always connected to these accounts, which makes the stolen or misplaced phone more vulnerable for criminal activities. Furthermore, a lost or stolen device can be used to gain access to secret data stored on it.

Android Botnets as a Universal Threat for Mobile Device Users

The dramatic increment in the number of mobile device users has attracted cybercriminals to develop malicious applications (Narudin et al., 2016). In addition, mobile devices contain more sensitive information about the owners; this tends to be taken lightly by security organisations and individuals. A mobile device can be misused in many ways. The botnet is one of the most successful methods by which a mobile device can be misused for malicious activities against organisations or individuals.

Android Botnet as a Threat for Organisations

Android botnets are mostly used for organised economic fraud. Today, world economies must deal with a broad range of botnets that have caused a considerable amount of damage. About USD7.1 million was estimated lost due to click fraud performed by botnets using DDoS attacks in 2007 (Plohmann, Gerhards-Padilla, & Leder, 2011). It is very hard to detect click fraud, as these target legitimate users while they are surfing websites. According to a published report, on every USD3 million spent on digital advertisement, USD1million is spent on click fraud. Another statistics report showed that digital advertising hit the highest level of fraud in 2015, which was estimated at USD27.5 billion (Slefo, 2015). According to Kaspersky's monitoring results, 35,000 malicious mobile programmess were found at the end of 201. These malicious programmes steal sensitive data from the end-user devices, consume account balances while running digital advertisements, which is pushed by these malicious programmes (Christian & Maria, 2013). Kaspersky released security threat statistics for the year 2015, in which they blocked 0.8 billion attacks and used a list of 6.5 million unique host, to from web resources located in various countries around the world (LAB, 2015).

Android Botnet as a Threat for Individuals

Mobile devices offer advanced capabilities, with more storage capacity that can store organisational and confidential data of end users (Peng et al., 2014). Furthermore, advanced mobile devices offer more computing capabilities than many of personal computers offered a few years back. A mobile device has three main features, such as Android applications, storage, and connectivity with internet or cellular network (Rubin, 2008). Once a mobile device becomes

part of the Android botnet, botmasters attempt to take complete control of it to enable the botnet creator to perform illegal activities without the knowledge of the end user. After taking control of the mobile device, a botmaster can access everything from the compromised device. Furthermore, sending texts and making phone calls to premium numbers can be performed with these compromised devices. Botmaster can access contacts and messages on the compromised devices as well. These botnets take advantage of unpatched Android application updates.

Android Botnet Security

As smartphones are the largest category of Android devices, they have become an essential tool in how people communicate with one another. Each Android device has three key features: Applications, online and storage. Android applications are one of the core features of Android devices. They enable users to play games, read the news, connect with others, check weather conditions, perform online banking, read maps and navigators and perform many other functions. These applications are available from third parties like Google Play Store and Amazon (Silva et al., 2013). It may be a primary feature for many end users.

The core function of the Android device is to enable the user to make calls, take photographs, send text or picture messages and access personal data storage. It also allows the developer to develop richer applications. The developer may also access the user's address book, SMS content, GPS location data, movement data by G-sensor and accelerometer and even information in other applications. The Android does not differentiate between the phone's core applications and third-party applications. However, such applications from third parties can access personal/confidential information in the Android device very easily. These core features of Android devices make these devices an easy target for cybercriminals.

In this modern era, trojanised Android applications are a common infection method of Android devices. This is most often targeted by cybercriminals who use different types of malware. Botnet, a dangerous malware, compromises Android devices such as smartphones, smartware, tablets and notebooks, attempting to get full access to the device and provide control to the botmaster. The data found on Android devices include text messages (SMS/MMS), contacts, call logs, e-mail messages (Gmail, Yahoo), chats, location coordinates using the global positioning system (GPS), photographs, videos, web history, search history, driving directions, Facebook and Twitter information, music collections and other information. These third-party applications provide a simple and easy means of accessing content and services of Android devices. It is important to be aware of how to use these third-party applications safely and securely. Android botnets are able to spread themselves by sending copies to compromised devices.

The criminal activity the Trojan-Ransom.Android-OS.Small family is a multifunctional ransomware Trojan performed by an Android botnet. After connecting to the botnet army, it receives commands from the command and control channel and performs the activities received from them. Once run, it asks for the victim's device's admin rights and loads information about the victim's device to a malicious server. It can be an international mobile equipment identity (IMEI), international mobile subscriber identity (IMSI), device model, brand and phone number or other information.

In addition, the Trojan is registered in the Google cloud messaging (GCM) system. As such, the Trojan can receive commands from both the C&C server and via GCM. With this information, it can perform the commands shown in Table 1.

Table 1
Commands between C&C server and trojan

Command	Description
START	Start the main service of the Trojan
STOP	Stop the main service of the Trojan
RESTART	Restart the main service of the Trojan
URL	Change the C&C address
MESSAGE	Send an SMS to a specified number with a specified text
UPDATE_PATTERNS	Update the rules for processing incoming SMS
UNBLOCK	Disable the device administrator's rights
UPDATE	Download a file from the specified URL and instal it
CONTACTS	Send out a specified SMS to all contacts from the list of contacts
LOCKER_UPDATE	Update the text with the ransom demand
LOCKER_BLOCK	Block the device
LOCKER_UNBLOCK	Unblock the device
CHANGE_GCM_ID	Change the GCM id

The main idea behind botnets is to control interaction in Internet Relay Chat (IRC) chat rooms. They are able to interpret simple commands, provide administration support, offer simple games and other services to chat users and retrieve information about operating systems, logins, email addresses and aliases, in addition to other information (Silva et al., 2013). The first known iKee.B Mobile botnet was found in 2009. It was discovered to be using the Command and Control Server in the iPhone. This botnet is able to propagate itself and to instal third-party applications on the end-user's phone without user information (Peng et al., 2014).

Table 2 shows the timeline of Android botnets with respect to their first appearance in terms of year, platform, instruction, categories and C&C type in addition to other related information.

Table 2
Android Botnet timeline

Year	Name	C&C Type	Botnet Instructions	Criminal Activities by Default	Requires Permission
2010	SMSHowU.A	SMS	Leak location, GPS and maps through SMS	None	N/A
2011	Geinimi.A	HTTP	ON, OFF, ADD or Set or Rem Sender	IMEI, IMSI, SIM, SIM state, Build info, GPS, Board, Brand, CPU type, User, Software version, SIM country, SIM operator	N/A
	DroidKungFu.A	HTTP	Leak location, GPS and maps through SMS	Send sensitive data, execDelete, Exploit known vulnerabilities to gain root, Instal APK, execOpenUrl, execStartApp	N/A

Table 2 *(continue)*

2012	Fjcon.A	HTT Phone	ICCID;	Financial, Propagation of malware	N/A
	Rootsmart	HTTP	action.host start; action. boot; action.shutdown; action.install; action. installed; action.check live;action.download apk;	IMEI, IMSI, cell ID, location area code, mobile network code	N/A
	TigerBot.A	SMS	Change APN; Notify of SIM change; Kill running process	IMEI	N/A
2013	Stealer.B	HTTP and SMS	HTTP: time; sms; send; delete; smscf SMS: ServerKey +001; +002; anything	IMEI, IMSI, contacts	READ_SMS; INTERNET; RECEIVE_BOOT_ COMPLETED; READ_ PHONE_STATE; RECEIVE_SMS; READ_CONTACTS; SEND_SMS; WRITE_EXTERNAL_ STORAGE
	Tascudap.A	HTTP	time; sms; send; delete;smscfSMS: ServerKey + 001; 002; anything	Specify time when trojan should next contact C&C, send SMS, delete SMS from phone, selective SMS hiding, start application, forward received SMS, update	READ_SMS; ACCESS_NETWORK INTERNET; READ_PHONE_STATE; RECEIVE_SMS; READ_CONTACTS; SEND_SMS; WRITE_EXTERNAL_ STORAGA;
	BadNews.A	HTTP	news; showpage; install; showinstall; iconpage; coninstall; newdomen; seconddomen; stop; testpost	Propagation of possible malware; download and instal APK	RECEIVE_BOOT_ COMPLETED; SEND_ SMS; RECEIVE_SMS; INTERNET; ACCEESS_INTERNAL_ MEMORY; ACCESS_EXTENAL_ MEMORY;
	Spamsold.A	SMS	Display same icon on the menu, retain the image same but the name may change, instal APK once clicked	Sends SMS spam messages without the user's consent	INTERNET; CHANGE_ COMPONENT_ ENABLED; RECEIVE_SMS; READ_SMS; SEND_SMS WRITE_SMS; RECEIVE_SMS; RAISED_THREAD_ PRIORITY; READ_ CONTACTS; WRITE_EXTERNAL; RECEIVE_BOOT_ COMPLETED; WAKE_ LOCK;

Table 2 *(continue)*

2014	FrictSpy.E3	HTTP; SMS	Command and Control to execute malware activities such as call records, use camera for pictures and videos, use mic for recording voice	Incoming/Outgoing call; Incoming/Outgoing SMS, GPS location information, URLs that the device user accesses	ACCESS_NETWORK_ STATE; CALL_PHONE; GET_TASKS; INTERNET; READ_PHONE_STATE; READ_SMS; RECEIVE_BOOT_ COMPLETED; RECEIVE_SMS; SEND_SMS; SYSTEM_ALERT_ WINDOW; WAKE_LOCK; WRITE_SMS;
	Geinimi.A	HTTP	ON, OFF, ADD or Set or Rem Sender	User, Software version, IMEI, SIM State, CPU type, SIM country, IMSI, SIM, SIM operator, build info, GPS, Board, Brand	CALL_PHONE; GET_TASKS; INTERNET; READ_PHONE_STATE; READ_SMS; RECEIVE_BOOT_ COMPLETED; RECEIVE_SMS; SEND_SMS; SYSTEM_ALERT_ WINDOW; WAKE_LOCK; WRITE_SMS;
	SpyBubb.A	SMS	Leak location, GPS and maps through SMS; HTTP: time; sms; send; delete; smscf SMS: ServerKey +001; +002; anything	Collect SMS, Call, Fine location, Coarse location, GPS, Device info like IMEI, IMSI etc. Share phone information to vendor site	ACCESS_NETWORK_ STATE; ACCESS_WIFI_STATE; READ_PHONE_STATE; INTERNET; WAKE_LOCK;
2015	Leech.A	HTTP	action.host start; action. boot; action.shutdown; action.install; action. installed; action.check live;action.download apk	Instal itself persistently, run with full privileges, unwanted payment through SMS, spying activities, dynamically load command and control server	ACCESS_NETWORK_ STATE; ACCESS_WIFI_STATE; READ_PHONE_STATE; INTERNET; WAKE_LOCK;
	Tediss	SMS	N/A	Monitor calls, SMS and conversation applications	CALL_PHONE; GET_TASKS; INTERNET; READ_PHONE_STATE; READ_SMS; RECEIVE_BOOT_ COMPLETED; RECEIVE_SMS; SEND_SMS; SYSTEM_ALERT_ WINDOW; WAKE_LOCK; WRITE_SMS;
	WormHole.A	HTTP and SMS	WormHole.A HTTP and SMS	Instal applications without notification; Location information; Add contact items; Monitor list of applications	READ_EXTERNAL_ STORAGE; READ_PHONE_STATE; READ_NETWORK_ STATE; INTERNET; READ_INTERNAL_ STORAGE; WAKE_LOCK; READ_COARS_ LOCATION;

Table 2 *(continue)*

	SilverPush.A	HTTP and SMS	HTTP: time; sms; send; delete; smscf SMS: ServerKey +001; +002; anything; ON, OFF, ADD or Set or Rem Sender	IMEI number; Operating system version; Location; Potentially the identity of the owner; Behaviour of users using TVs; Web browsers; Radios	ACCESS_NETWORK_ STATE; CALL_PHONE; GET_TASKS INTERNET; READ_PHONE_STATE; READ_SMS; RECEIVE_SMS; SEND_SMS; WRITE_SMS;
2016	MazarBOT.A	SMS	N/A	Sends premium SMS, exfiltrate sensitive information and steal the received SMS messages by setting up a backdoor on device	ACCESS_NETWORK_ STATE; CALL_PHONE; GET_TASKS; INTERNET; READ_PHONE_STATE; READ_SMS; RECEIVE_BOOT_ COMPLETED; RECEIVE_SMS; SEND_SMS; SYSTEM_ALERT_ WINDOW; WAKE_LOCK; WRITE_SMS;
	Morder.A	HTTP and SMS	Command and Control to execute Malware activities such as calls record, use camera for pictures and videos, use mic for recording voice	Track location; Leak contacts to C&C Upload data from SD Card to C&C; Delete or download files in the infected device; Leak phone call history; Take pictures with the camera; Record audio and calls; Execute shell commands	ACCESS_NETWORK_ STATE; CALL_PHONE; GET_TASKS; ACCESS_FINE_ LOCATION; ACCESS_COARS_ LOCATION; INTERNET; READ_PHONE_STATE; READ_SMS; RECEIVE_BOOT_ COMPLETED; RECEIVE_SMS; SEND_SMS; SYSTEM_ALERT_ WINDOW; WAKE_LOCK; WRITE_SMS;
	Smishing.D	SMS	time; sms; send; delete; smscf SMS: ServerKey +001; +002; anything; ON, OFF, ADD or Set or Rem Sender	Detect text messages; Access fraudulent fake bank URL; Steal user's sensitive credential; Password stealing; Additional information stealing	ACCESS_NETWORK_ STATE; CALL_PHONE; GET_TASKS; INTERNET; READ_PHONE_STATE; READ_SMS; RECEIVE_SMS; SEND_SMS; WRITE_SMS;

NA=Not Available, HTTP=Hyper-Text Transfer Protocol, SMS=Short Message Service, PUP=Potential Unwanted Programmes, SD=Secure Digital, C&C=Command & Control Servers, IMEI=International Mobile Equipment Identity, IMSI=International Mobile Subscriber Identity, HTTP=Hyper-Text Transfer Protocol

Components of Android Botnet

A typical Android botnet has four elementary components as shown in Figure 2: bot, botmaster, command and control server and communication channel.

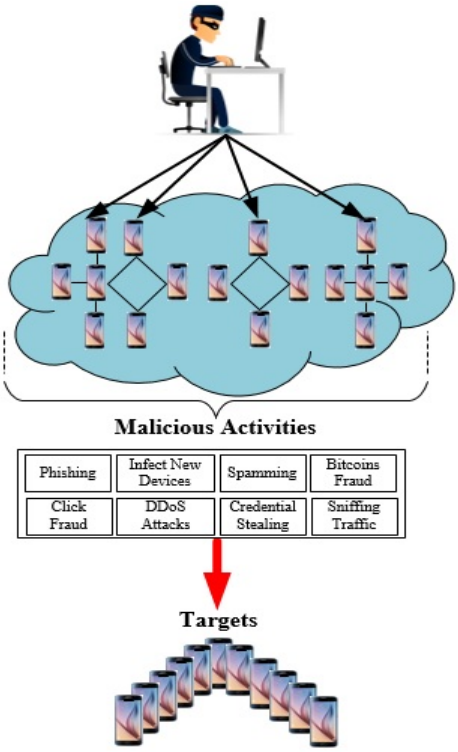


Figure 2. Typical android botnet structure

Bot. A bot is a malicious Android application that is installed in a susceptible host that can perform a series of different harmful actions upon the end user at a cybercriminal's command. This application can be installed to the victim devices in diverse ways. The most common ways include access to the infected websites, drive-by-downloads, spam emails, viral mechanisms and much more (Karim et al., 2015). Once an end-user device is infected with malicious software, it receives commands and controls from the botmaster through the command and control server using communication channels. A bot can be a servant and client at the same time.

Botmaster. The attacker is also known as the botmaster, who maintains and operates the command and control of botnets from remote areas. The botmaster, also known as the bot-herder, is responsible for a variety of malicious activities. Botmasters ensure that errors are fixed and that the bot does not break any of the rules of the channel or server it is logged into. Most botmasters hide their identity via proxies, the onion ring (TOR) and/or shells to disguise their ip address from detection by investigators and law enforcement agents.

Command control server. The term 'command and control' (C&C) is a military concept. Command and Control servers allow a bot entity to take new instructions and malicious capabilities as commanded by a remote individual (botmaster). These servers are used to control botnets, in particular. Command and control in the botnet's Fast-flux Domain Name System (DNS) can be used to make track the control server difficult to do, as the server may change from day to day. These servers may also hop from one DNS domain to another. The Domain Generation Algorithm (DGA) is used presently to create new DNS names for controller servers. A botnet may have different C&C server topologies like Star, Multi-Server, Hierarchical and Random topology.

Communication channels. The botnet communication channel refers to the protocol used by bots and the botmaster to communicate with each other. Bluetooth, Internet Relay Chat (IRC), Hypertext Transfer Protocol (HTTP), peer-to-peer (P2P) and voice over internet protocol (VoIP) servers are used to pass information between bots and the botmaster. The botmaster creates IRC channels on the C&C server, after which the compromised machines will wait for commands to perform malicious activities. An interesting feature of the IRC protocol is the possibility of multicast communication through groups. The IRC channel has some serious limitations like being easy to detect and interrupt. It is rarely used in corporate networks and is usually blocked (Silva et al., 2013).

Due to these limitations of the IRC channel, the HTTP has become the most usable mechanism for implementing command and control communication (Liu, Chen, Yan, & Zhang, 2008). The first Android botnet named SymbOS/Yxes, which appeared in 2009, (Suarez Tangil, Tapiador, Peris-Lopez, & Ribagorda, 2014) targeted the SYMBIAN OS platform using a rudimentary HTTP-based command and control (C&C) channel. Centralised botnets are not more secure as discussed above, so the trend shifted to decentralised botnets. Most of the decentralised botnets are based on a variety of P2P protocols (Jelasity & Bilicki, 2009). Similarly, VoIP is used as the communication channel in vishing (VoIP and phishing) instead of the more usual email technique.

Life Cycle of Android Botnet

Android botnets can come in different structures and sizes, but in general, they go through the same steps as computer botnets (Silva et al., 2013), as shown in Figure 3. An active botnet requires the bot device to complete its life cycle. A typical Android botnet can be developed and maintained in five phases: initial infection, secondary injection, connection, malicious command and control, update and maintenance. In the first phase, initial infection, the end-user device is infected and becomes an active member of the Android botnet. The second phase, secondary injecting, can be carried out by injecting the malicious code into the end-user devices through Bluetooth, drive-by-download, automatic scan, NFC and Wi-Fi (Faruki et al., 2015).

After injecting the code into the victim's device, the bot finds a way to connect to the command and control server. This happens in the connection phase, which is the only phase that may occur several times during the botnet's life cycle (Liu et al., 2008). Once an infected device connects to the Android botnet's command and control server, the botmaster will send

commands through the C&C server using communication channels, while the bots await commands from the botmaster.

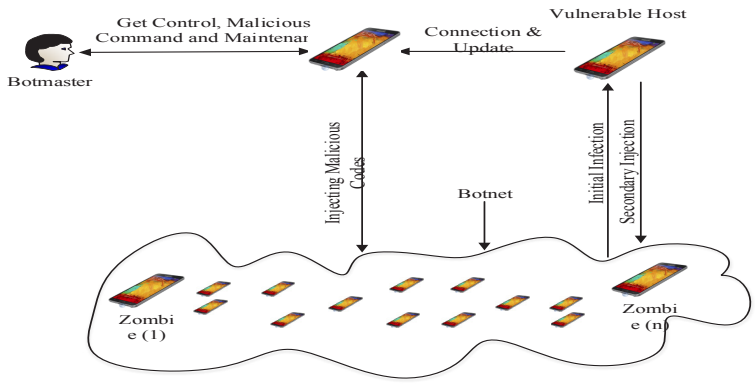


Figure 3. Android botnet life cycle

The last phase of the Android botnet's life cycle is updating and maintenance of the infected devices. Maintenance is important for a botnet to keep his army of infected devices active. The new updates are then sent to these bots many times for many reasons that propagate the different types of criminal activity such as spamming, identity theft, DDoS attacks and much more.

Mobile Device Infection Vectors

There are multiple infection vectors for delivering malicious content to mobile devices. In this survey, we classify infection vectors into four categories: SMS/MMS, Bluetooth, Internet access and file duplication with USB. Cellular services, such as short message service (SMS) and multimedia messaging service (MMS), can be used as attack vectors for smartphones, as shown in Figure 4. For example, SMS/MMS messages can be used to deliver malicious content and to maintain communication with an attacker. For example, ComWar is a worm that browses the host's phonebook and then spreads via SMS/MMS messages (Peng et al., 2014).

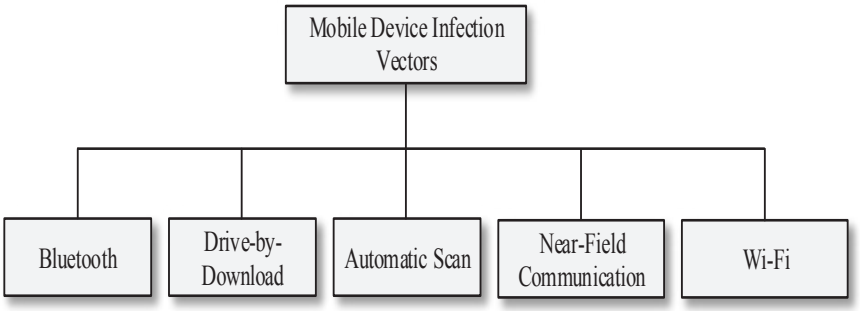


Figure 4. Mobile device infection vectors

Bluetooth

Bluetooth is short-range radio communication protocol used for exchanging data over a limited distance between Bluetooth-enabled devices. Device-to-Device (D2D) malware attacks are performed on the bases of Bluetooth. Once the cybercriminal infects a smartphone with the bot code, it can enable Bluetooth without the knowledge of the owner and target another Bluetooth-enabled device in its range. If the connection is established, the infected device sends the bot code to the targeted device using this Bluetooth vector. This limits the attack vector in some way, such as one-to-one connection, limited distance range and others.

Drive-by-Download

To secure Android devices from botnet attacks, users should only visit reputable websites for downloading application software and other video/audio materials. Botnet infection is possibly acquired by visiting a malicious website. When visited by a smartphone or tablet user, the malicious website forces the user to download the plugin software, which is in fact a malware. If a web page causes the automatic downloading and installation of software without the Android device user's consent, the page is considered malicious. This mechanism, which is also called drive-by-download, allows malware to control Android devices.

Automatic Scan

Automatic scan is performed to infect new victim's devices by compromising and influencing them to be a part of the botnet. In this technique, a new host inside the botnet must be recruited to establish a new botnet through vulnerability scanning (Ianelli & Hackworth, 2005). This goal can be achieved by infecting many hosts, which attempt to identify exploitable vulnerabilities in other new hosts. For example, FTP services suffer buffer overflow exploitation (Lashkari, Ghalebandi, & Moradhaseli, 2011).

Near-Field Communication

Near-Field Communication (NFC) is an advanced wireless technology that allows fast data transfer between two close devices with an enabled NFC setting. NFC is related to mobile payments, such that it has the personal banking information of a user. It has gained popularity among botmasters for spreading malicious commands to compromise other devices because of its fast data transfer capability. In addition, dependence on NFC has induced the C&C channel of botnets to be more challenging (Stevanovic, Revsbech, Pedersen, Sharp, & Jensen, 2012).

Wi-Fi

Wi-Fi has assured compensation over other communication media applicable to Android botnets. The use of open Wi-Fi networks for an Android botnet provides a higher level of stealth and fewer entry barriers than other communication media. Denial of service (DoS) attacks and distributed DoS attacks are threats that can simultaneously inflict devastation on many users. Apart from the aforementioned-infection vectors, smartphones could be compromised

using other methods e.g. the use of USB. If the files used to synchronise smartphones are compromised, malware can also infect smartphones. As a result, attackers can access the host's classified information and instal malicious applications on the smartphone.

Android Botnet Architecture

Personal-computer-based botnet are considered the most compromised platforms for botnet attacks compared to the recently evolved Android botnets due to some limitations, such as limited battery power, limited processing speed, limited internet access and limited memory storage. Android botnets have similar architecture as computer botnets, namely, centralised architecture, decentralised architecture and hybrid architecture. Table 3 shows the advantages and disadvantages of existing Android botnet architecture with respect to map complexity, detection, message latency and survivability.

Table 3
Command-and-Control architecture

Architecture	Centralised	Decentralised	Hybrid
Alias	Star	Peer-to-peer	Random
Map Complexity	Very low	Medium-High	Moderate
Detection	Medium	Low-Medium	High
Message Latency	Very low	Moderate	Moderate-High
Survivability	Low-Medium	Medium	High

Centralised botnet architecture. In centralised botnet architecture, all the bots relate to a central command-and-control server to establish a communication channel with central point as illustrated in Figure 5. In centralised architecture, the botmaster controls and supervises all bots in a botnet from a single C&C server. Botmasters are able to communicate with the bots continuously by sending instructions to them through these central servers (Anwar et al., 2014). As all bots receive commands and report to a C&C server, it is easy for botmasters to manage botnets using centralised architecture. Furthermore, centralised botnet architecture uses two types of topologies, star topology and hierarchical topology, and two types of protocols, Internet Relay Chat (IRC) and Hypertext Transfer Protocol (HTTP) (Khattak, Ramay, Khan, Syed, & Khayam, 2014; Li, Jiang, & Zou, 2009). The design of centralised architecture is less complex compared to other architecture, while message latency and survivability rate are low. This causes low reaction time, easy means of communication and direct feedback (Plohmann et al., 2011).

It also possesses some limitations. For instance, centralised architecture has more maximum failure chances compared with other architecture. If the C&C server fails, then all the botnets may stop working because of the central point of control. Detection of a botmaster is easier compared than if the decentralised and hybrid architecture were used (Bailey, Cooke, Jahanian, Xu, & Karir, 2009; Cooke, Jahanian, & McPherson, 2005; Zang et al., 2011).

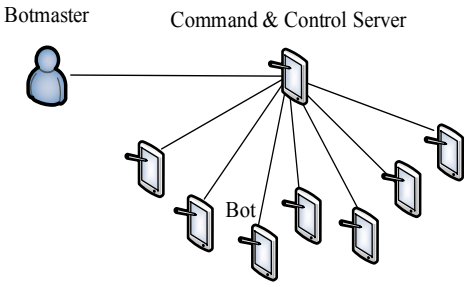


Figure 5. Centralised android botnet architecture

Decentralised botnet architecture. In decentralised botnet architecture, no single responsible entity controls different bots in a botnet. More than one C&C server communicate with various bots as described in Figure 3. Botnets using decentralised architecture are known as decentralised botnets. However, the term peer-to-peer botnet is also commonly used for this type of botnet. Decentralised botnets are more difficult to detect compared with centralised botnets. Figure 6 shows that no specific C&C server exists in decentralised architecture, and all bots act as the C&C server and the client at the same time (Dong, Wu, He, Huang, & Wu, 2008). Decentralised architecture is based on Peer-to-Peer (P2P) protocols. Compared with centralised architecture, the design of P2P architecture is more complex and detection of a botnet with the same architecture is more difficult than detection of other botnets. Message latency and survivability rate are higher than those of the centralised botnet architecture. Failure chances are lower in decentralised architecture than in centralised architecture because if a C&C server fails, then other C&C servers can manage and monitor the botnet (Cooke et al., 2005).

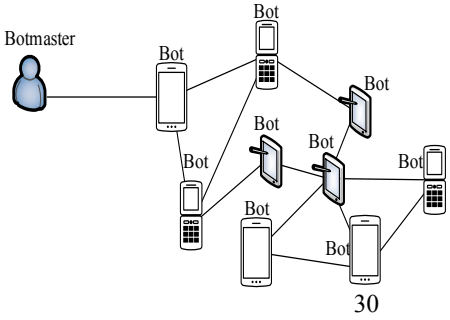


Figure 6. Decentralised architecture

Hybrid botnet architecture. Hybrid architecture is the combination of centralised and decentralised architecture as shown in Figure 7. Hybrid architecture comprises two types of bots, namely, the servant and the client. Bots are connected to the hybrid botnet as a client or a servant. Monitoring and detection of botnets with hybrid architecture is more difficult than detecting those with centralised and decentralised architecture. However, hybrid architecture is less complex in design.

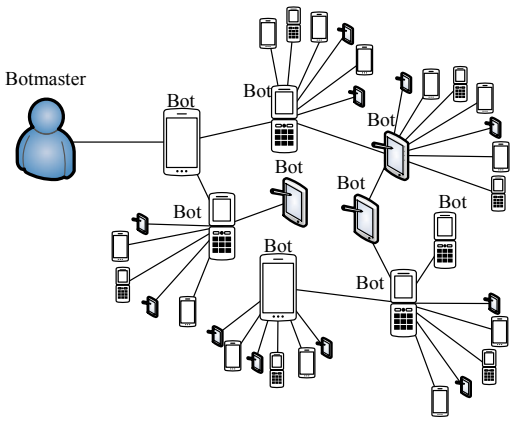


Figure 7. Hybrid architecture

Mobile Botnet Detection Techniques

The challenges faced in mobile-device security are quite similar to those faced in personal-computer security. To solve this problem, researchers have proposed and developed common desktop security solutions for smartphones. Some of the popular security solutions are listed below.

Kirin. Kirin security service is an OS-level protection service that provides enhanced security mechanisms for Android smartphone applications (Enck et al., 2009). This approach performs lightweight certification of applications to mitigate malware at installation time with modification of the Android applications installer. Kirin has different security rules; a well-known combination of permissions is the most important part in these rules. To define these security rules, a detailed understanding of malware and protection techniques is required; it is usually performed by security experts. Furthermore, it prevents access to sensitive information. However, once information enters the application, no additional mediation occurs.

Multi-Agent system. Szymczyk (2009) proposed the Multi-Agent Bot Detection System (MABDS) based on the hybrid approach. It is the combination of multiple agents such as administrative agent, user agent, a central knowledge database, system analysis, honeypots, agent collections and network analysis (Silva et al., 2013). In this technique, each agent observes traffic using different sensors by implementing the Markov chain model to perform dynamic risk assessment (Shameli, Cheriet, & Hamou-Lhadj, 2014). These systems in multifaceted, piercing, real-time domains involve autonomous agents that should act as a team to compete against malware (Castiglione, De Prisco, De Santis, Fiore, & Palmieri, 2014). The slow convergence of new signatures with the knowledge database is the key limitation of this technique. The new signature updates are another limitation of this system (Karim et al., 2014).

SAINT. SAINT (Ongtang, McLaughlin, Enck, & McDaniel, 2009) is a rule-based runtime approach for Android application security that defines application protection at runtime,

depending on the caller and permission constraints. It protects Android applications from one another by their policies during the installation time and runtime interaction. In this case, it allows an application to define which application can access its interfaces and how other applications use those interfaces. This technique has the same limitation as Kirin security services. To define these security rules, a detailed understanding of malware and protection techniques is required, and security experts are required to perform the job.

AASandbox. AASandbox was the first technique to perform both static and dynamic analysis of the Android applications and was proposed by Bläsing, Batyuk, Schmidt, Camtepe and Albayrak (2010). The static analysis scans the Android applications for malicious patterns without installation on the Android platform, while in the dynamic analysis, the Android application is executed in a fully isolated platform called sandbox. It also intervenes and logs low-level interaction with the system for further analysis during application execution. In contrast, both the detection algorithm and sandbox algorithm are implemented in the cloud. AASandbox uses a system called foot-printing approach for detecting suspicious Android applications. In its early days, when AASandbox was proposed, there were no known botnet malware samples available to evaluate this technique, although it seems to be unmaintained nowadays.

Paranoid. Considering various factors of smartphone technology including resources, storage, processing and memory, Paranoid Android malware detection technique was proposed for the first time in mobile technology (Portokalidis, Homburg, Anagnostakis, & Bos, 2010). Paranoid Android is a security model implemented on remote servers (cloud server) to observe the dynamic behaviour of Android applications and to detect zero-day attacks, system call anomaly and antivirus file scanning. Both Crowdroid and Paranoid Android incur a 15-30% overhead for smartphone devices. This particular technique records information that is necessary for application execution and transmits it to a cloud server over an encrypted channel. While a complete replica of the executing application is running parallel on the remote virtual machine, the server can detect the potential malware using this technique. Both the application and its replica are executing parallel to one another, which may cause a lot of space and time complexity. It also converts energy by using 'loose synchronisation', which may cause loss of battery power usage that specifically sends information when the mobile user is using the mobile device.

Crowdroid. Crowdroid is a dynamic approach based on the behaviour of Android applications and was proposed by Burguera, Zurutuza and Nadjm-Tehrani (2011). Crowdroid is a lightweight application available online on Google Play Store, which can be downloaded and installed on Android smartphone devices. It monitors and collects the API calls of apps that are running on mobile devices and sends them to a centralised server after preprocessing. With the application of cluster algorithms, Android applications can be evaluated with this approach. The given approach is also able to detect self-written malware.

DroidRanger. DroidRanger is the combination of two systems based on permissions' behaviour, foot-printing and heuristic-based filtering (Odili, Kahar, Anwar, & Ali, 2017). It was proposed by Zhou, Wang, Zhou and Jiang (2012). This technique applies both the static and dynamic approaches to detect malicious applications in existing Android markets. Permissions-based behaviour foot-printing is used for the detection of known malware, while heuristic-based filtering is used for unknown malware Android applications. Despite the advancements in the detection approaches applied by DroidRanger, the system has some limitations; it requires manual operation for analysing and collecting behaviour of Android applications (Babu Rajesh, Reddy, Himanshu, & Patil, 2015). Manual operation takes more time than other detection techniques (Odili & Kahar, 2015).

Bouncer. Bouncer was proposed in 2012 by Oberheide and Miller (2012). It provides static and dynamic scanning together with Android applications that are automatically performed on the server. Google Play Store uses this technique to scan an Android application before hitting the application market (Penning, Hoffman, Nikolai, & Wang, 2014). Bouncer has the potential to take newly-uploaded applications to the app market. If this application is able to send an SMS to the malicious sites or detect other criminal activities, it classifies that Android application as malware. If not, it classifies it as benign. However, in this advanced era, it seems that cyber attackers have found ways to bypass detection. This technique is better for those who download applications from Google Play Store, while those who download applications from third-party app stores are not protected by this technique.

RobotDroid. RobotDroid is an Android malware detection technique that is based on SVM machine learning classifier algorithm and was proposed by Zhao, Zhang, Ge and Yuan (2012). This technique focusses on the signature of the applications. It has the ability to detect unknown malware like Plankton, DroidDream and Gemini. This framework can be used only for these few types of malware; this is the main limitation of this framework.

DroidScope. DroidScope designed by Yan and Yin (2012) is a fine-grained dynamic binary instrumentation tool for Android. It rebuilds two levels of semantic information: OS and Java. It provides an instrumentation interface that can be used to write plug-ins. It implements API tracing, native instruction tracing, Dalvik instruction tracing and taint tracking plug-ins. DroidScope works entirely on the emulator level and requires no changes to the Android sources. It runs the analysis outside the smartphone software stack and can analyse kernel-level attacks. This system has a big drawback: not able to detect real-time attacks. The second drawback is that it does not cover the subtleties of real devices (Enck et al., 2014).

Table 4 shows the list of Android botnet detection techniques with respect to year, major contribution and limitations.

Table 4
List of Android Botnet detection techniques

Ref	Techniques	Year	Architecture	Key Concept	Major Contribution	Limitation
Enck et al., 2009	Kirin	2009	Static	Based on permissions	Used rules to detect malware in installation time	No access to sensitive information of application; Cannot detect new malware
Szymczy, 2009	Multi-Agent System	2009	Hybrid	Based on permissions and rules	Used agent for traffic analysis	Slow interaction with the knowledge database
Ongtang et al., 2009	SAINT	2009	Static/ Dynamic	Based on rules	Installation time detection of malware	No access to sensitive information of application; Cannot detect new malware
Bläsing et al., 2010	AASandbox	2010	Static/ Dynamic/ Hybrid	Base on signatures and behaviour of the logs	Before instal detection of malware	NA
Portokalidis et al., 2010	Paranoid Android	2010	Dynamic/ Hybrid	Based on behaviour	Dynamic analysis, memory scanners, system call anomaly detection	Consumption of more time, space and power
Burguera et al., 2011	Crowdroid	2011	Dynamic	Based on behaviour	Client APK, behavioural detection	More clients, dynamic analyser
Zhou et al., 2012	DroidRanger	2012	Static/ Dynamic	Permissions-based behaviour	Detection of known and unknown malware, 0-day malware detection	Only a few of all possible execution paths are negotiated within one analysis run
Oberheide & Miller, 2012	Bouncer	2012	Static/ Dynamic	Permissions-based	Detection of unknown malware	Can be easily evaded by cybercriminals
Zhao et al., 2012	Robotdroid	2012	Static	Signature-based	Detection of unknown malware such as Plankton, DroidDream, and Gemini	Detects only specific malware families such as Plankton, DroidDream and Gemini
Yan & Yin, 2012	DroidScope	2012	Dynamic		Dynamic binary instrumentation	Cannot detect real-time botnet attacks
Zhou et al., 2012	DroidMOSS	2012	Dynamic	Permissions-based,	Fuzzy Hashing Technique,	Identifies only repackaged official Android market applications
Alparslan, Karahoca, & Karahoca, 2012	Data Mining	2012	Static	Behaviour-based	utilise auditing programmes to extract and extend features	Cannot detect real-time botnet attacks
Faruki, Ganmoor, Laxmi, Gaur, & Bharmal, 2013	AndroSimilar	2013	Static/ Dynamic	Statistical features	Improbable signature generation, thwarts obfuscation and repackaging	Limited malware DB, more false positives; Cannot detect new malware
Spreitzenbarth, Freiling, Echtler, Schreck, & Hoffmann, 2013	Mobile-Sandbox	2013	Static	Smali, emulator	Both static and dynamic analysis, obfuscation resistance, native API call track, web accessibility	More detection time

Table 4 *(continue)*

Rastogi, Chen, & Enck, 2013	AppsPlayground	2013	Static/ Dynamic	N/A	he-based UI interaction based on contextual exploration	Cannot detect real-time botnet attacks
Reina, Fattori, & Cavallaro, 2013	CopperDroid	2013	Dynamic	Behaviour-based approach	Automatically performs dynamic analysis, reconstructs behaviour of Android	Identifies only repackaged official Android market applications
Gascon, Yamaguchi, Arp, & Rieck, 2013	Embedded call graph	2013	Dynamic	Function call graphs	Obfuscation resistance	Undecidability of static call graph construction; Cannot detect new malware
Abdelrahman, Gelenbe, Görbil, & Oklander, 2013	NEMESYS	2013	Static	Model-based approach	Generate background traffic of network for simulating smaller set of users, learn Random Neural Network,	Limited to a small number of users, space complexity
Roshandel, Arabshahi, & Poovendran, 2013	LIDAR	2013	Static/ Dynamic	Behaviour-based	automatically detects, analyses, protects, remediates	N/A
Moonsamy, Rong, & Liu, 2014	Mini Permission Pattern	2013	Static/ Dynamic	Based on permission	'Used' permission extraction, informative data from contrast permission patterns	Careful analysis of permissions, no repackaging resistance; Cannot detect new malware
Enck et al., 2014	TaintDroid	2014	Static	Behaviour-based	Different APIs, specifically SMS APIs	It does not track implicit control flows due to performance overhead
Suarez-Tangil, Tapiador, Pens-Lopez, & Blasco, 2014	Dendroid Approach	2014	Dynamic	Code Chunks	Unknown malware classification, fast and scalable, dendograms	No obfuscation resistance, large feature vectors; Cannot detect new malware
Dhaya & Poongodi, 2014	N-gram analysis	2014	Static	N-gram CVSS	Produced N-grams signatures	No obfuscation resistance; Cannot detect new malware
Lindorfer, Neugschwandtner, Weichselbaum, Fratantonio, Van Der Veen, & Platzer,, 2014	Andrubis	2014	Static/ Dynamic	Based on behaviour and rules	Static and dynamic analysis on both Dalvik VM and System Level	Dynamic analysis consumes more space; Cannot be used for latest Android applications
Andronio, Zanero, & Maggi, 2015	Heldroid	2015	Static/ Dynamic	Behaviour-based	Static and dynamic analysis	Portability, internationalisation and evasion

DroidMoss. Zhou et al. (2012) proposed DroidMoss in 2012, using the fuzzy hashing technique to effectively localise and detect repackaged and injected applications. This technique detects Android applications in the existing mobile app market that are injected with malicious codes using the repackaging technique. The main feature of the applications used in this technique is the Dalvik opcodes. DroidMOSS calculates fuzzy hashes on each N sequential opcode,

which then applies a measure function on the two applications to realise their similarity quantitatively. The use of DroidMOSS is limited to identifying repackaged official Android market applications.

Data mining. The hardest part of the detection of malicious traffic is to differentiate C&C data flow from normal data flow behaviour. To overcome this limitation, data mining techniques are used to recognise the pattern by extracting the unexpected network patterns (Alparslan et al., 2012). Data mining is the most used machine learning device method for classification, prediction, regression and inference. This technique is extensively used in anomaly detection, especially in establishing generic and heuristic methods (Odili, Kahar, & Noraziah, 2016; Schultz, Eskin, Zadok, & Stolfo, 2001). Data mining approaches detect structures in a wide range of data, such as byte code, and use these structures to detect upcoming malicious occurrences in related data. Researchers such as Gu, Perdisci, Zhang and Lee (2008), Gu, Porras, Yegneswaran, Fong and Lee (2007), Gu, Zhang and Lee (2008), Wang, Huang, Lin and Lin (2011) and Yu, Dong, Yu, Qin, Yue and Zhao (2010) proposed BotMiner, BotHunter, BotSniffer and behaviour-based botnet detection systems based on the data-mining approach. This technique is very effective though it has some limitation as well. In experiments, BotMiner and BotHunter have been able to achieve 99% success rate with 1% false alarm and 99.2% success rate with 0.8% false alarm, respectively (Zhao et al., 2013).

AndroSimilar. AndroSimilar (Faruki et al., 2013) detects Android malware regions of statistical similarity starting from the .dex file. This method employs the similarity digest hashing system on byte-stream-based robust statistical malicious features. Similarly, a digest hashing scheme uses this feature to generate a list of signatures for this app. Here, the feature values between 100 and 990 are selected and the rest are discarded using the Bloom filter. A set of malicious signatures are generated and thus, a database of signatures is created. For testing a sample app, its signature is created in the same way as described above and is matched against a signature database and is considered malware if the similarity score crosses 35% (Sharma et al., 2016). Authors obtain an accuracy of 72.27% using a dataset of 101 malicious applications. Androsimilar performs at file level as an alternative for codes in decompiling; therefore, control of shared library is not protected. Also, porting the approach to constrained memory and a strong database remains a concern.

Mobile-SandBox. Mobile-SandBox is a static and dynamic analysis system that is publicly available. It was proposed by Spreitzenbarth, Schreck, Echtler, Arp and Hoffmann (2015). In this technique, the comparison of applications occurs in different stages: first, it compares the hash value with the VirustTotal database of the running application; second, it extracts the manifest file for permissions, background services, broadcast receivers and intents. This technique also extracts API calls from the Dalvik bytecode; thiss happen frequently in botnets. Mobile-SandBox makes it very easy to submit applications for static and dynamic analysis because of its user interface. A user can easily upload an application for static and dynamic analysis to the Mobile-SandBox by using the user interface. However, in some aspects, Mobile-SandBox seems unable to cope with the submission load.

AppsPlayground. AppsPlayground, based on TaintDroid, is a scalable dynamic analysis system that is used for detection of possible data leaks (Skovoroda & Gamayunov, 2015). Proposed by Rastogi, Chen and Enck (2013), it employs a Javaapp that connects to an emulator running a modified version of the OS and governs app behaviour exploration logic. Simply, the aim of AppsPlayground is to improve the stimulation of apps during dynamic analysis because it also detects dangerous API calls.

This technique also helps to create a more realistic analysis environment. It tries to drive the app along paths that are likely to reveal interesting behaviour through targeted stimulation of UI elements. This approach can be seen as an intelligent enhancement of the Application Exerciser Monkey and the custom stimulation of activity screens. This technique is largely orthogonal, as it focusses on stimulating broadcast receivers, services and common events instead of UI elements. Its main contribution is a heuristic-based intelligent black box (Monkey Exerciser-like) execution approach to explore the app's GUI (Odili, Kahar, Anwar, & Azrag, 2015). This technique can be more useful if combined with the static analysis technique.

CopperDroid. CopperDroid is a dynamic detection system presented by Reina, Fattori and Cavallaro (2013) that is built on top of the quick emulator (QEMU). To the best of our knowledge, this is the first technique that performs system call monitors of the Android applications out-of-the-box through virtual machine introspection (VMI) by reconstructing Dalvik behaviour and monitoring Binder communication (Lindorfer et al., 2014). CopperDroid carried the binder analysis to perform the reconstruction of high level Android-specific behaviour. It is available to the public as a web application that users can use to submit samples.

Embedded call graphs. Embedded call graphs is a static approach proposed by Gascon Yamaguchi, Arp and Reick (2013) in 2013. This technique can be used to find similarities between samples: first,-it-extracts function call graphs and then employs explicit mapping through kernel graphs from map call graphs to-feature-space. Sharma, Chawla and Gajrani (2016) showed that time and space complexity are high and large, respectively. Its key concept is functions call graphs, while obfuscation resistance is the major contribution. Embedded call graphs specially observe assembly-level analysis and support vector-machine implementation. The main disadvantage of this technique is that it cannot decide the static call graph construction.

NEMESYS. NEMESYS is a network model-based security solution that combines learning and modelling for detection of anomalies and attacks in mobile network. It deals with every mobile connection during communication between devices in a network. The motivation behind this approach was the difference between the number of mobile users who are monitored and dealt with in real time. Furthermore, a clear and understandable approach was needed to deal with every unique call. The second consideration in constructing this approach was the computational tools that were developed for anomaly detection that were based on mathematical models. However, NEMESYS has some limitations. For instance, it is limited to a small number of users. Also, this approach is complex and it uses a huge amount of memory.

Layered intrusion detection and remediation. The Layered Intrusion Detection and Remediation (LIDAR) framework focusses on automatic detection, analysis, protection and remediation of security threats. This framework is specially designed to detect intrusion in a multi-dimensional mode that is of network dimension, application dimension and social dimension (Roshandel, Arabshahi, & Poovendran, 2013). This approach contains both local and remote analyses, which causes some drawbacks. For instance, a set of local analyses is performed in the mobile device to detect malware and intrusion, leading to the use of more battery power in addition to time consumption and space complexity, among other issues.

TaintDroid. TaintDroid is a system-wide dynamic taint tracking and analysis system for simultaneous tracking of multiple sources of sensitive data. This technique monitors methods, variables, files and messages during application execution according to data flow (Suarez-Tangil et al., 2014). TaintDroid uses tag chunks to keep track of data in order to find information leakage at runtime. Information flow tracking needs lots of memory. None of these schemes is energy-efficient; hence, they are not suitable for resource-constrained mobile platforms.

Dendroid approach. The dendroid approach is based on text mining and information retrieval techniques (Suarez-Tangil et al., 2014). This technique extracts code chunks (CC) to analyse and classify the code structures in Android malware families. The authors present a simple way to measure the similarity between malicious applications by formulating the modelling process. In the experiments performed, more than 33 families with 1249 malware applications (Sharma et al., 2016) were detected. This approach also provided automatic classification of zero-day malware samples, which is based on the applications-code structure. With respect to time and accuracy, this technique is very fast and accurate, with high scalability. However, this technique features vector growth and new families create issues, while the strategies of obfuscation are not implemented.

N-gram. The N-gram [12] analysis is a probabilistic approach to detect the presence of malware. Reverse engineering tools like DexToJar, Java Decompiler-Graphical User Interface (JD-GUI) and ApkTool are used in this technique to convert executable to source code (high level or low-level language), thereby creating the training dataset. After this, the source code is considered as N-gram signatures. This N-gram model is a popular machine-learning algorithm and it is a type of probabilistic language that predicts the next item in the sequence with given datasets of order (N−1) as in the Markov Model. These signatures are then stored in a Comma Separated Values (CSV) file for the reason that signatures occupy a lot of space.

After this, a Common Vulnerability Scoring System (CVSS) is used to assess the vulnerabilities' severity level in software applications. It is a freeware tool. By applying this tool on the APK file under test, the description of all the vulnerabilities and solutions to mitigate the same is appended to the CSV file. It makes intuitive use of the N-gram machine-learning algorithm to analyse the Android apps. Limitations include the obfuscation techniques not being implemented.

Andrubis. Andrubis is a cloud-based malware detection technique proposed by Lindorfer et al. (2014). This technique combines both static and dynamic analyses on both Dalvik VM and the system level. First, it performs the static analysis by extracting the information including broadcast receivers, requested permissions, activities, services, SDK version and package name from the application manifest and its bytecode. Andrubis uses the modified DroidBox output to generate XML files that contain the analysed results. In the dynamic stage, it executes the application in a complete Android environment; during the execution its action is monitored at both the Dalvik and the system level. Other than this, Andrubis provides a web interface for users to submit Android applications and, so far, it has collected a dataset of over one million Android applications, 40% of which are malware. The only disadvantage of this technique is that it cannot track native codes. API calls that frequently happen in botnet are extracted from the Dalvik code, while the Andrubis is limited to the application's API level 8.

HELDROID. HELDROID is a fully automated behaviour-based approach to recognise known and unknown ransomware and scareware (Andronio et al., 2015). This approach analyses the Android application statically and dynamically as well. By using the static taint analysis approach, it analyses the function calls flow. Its result is more accurate when compared to those of previous apps. Still, it has some limitations. Although they focus on the mobile case, the results shown are far from being accurate. Ransomware is a general problem, but in this approach, it is limited to mobile devices only. HELDROID is based on sentence structure; this needs internationalisation.

CHALLENGES AND FUTURE DIRECTIONS

In this section, we discuss the challenges and future directions based on the findings of our study. Research into Android botnets is still in its initial stages. Therefore, sufficient opportunity exists for the betterment of detection and prevention of botnet attacks. The following challenges will help the researchers, academics and industry players to enhance this field.

Hybrid Approach Towards Android Botnet Detection

There is no way to ignore the rising security threats to mobile devices at this time. Researchers and industry players have proposed different botnet detection techniques. Practically, most of the existing Android botnet detections are either static or dynamic, and can detect known and unknown Android botnets. As can be seen in Table 2, few of the existing detection techniques based on the hybrid approach have a low detection rate and maximum false-alarm rate. This is a disturbing discovery.

Limitation of Mobile Devices

Personal computers are considered a more suitable platform for botnet attacks compared with mobile devices. Mobile devices have certain limitations, such as limited power storage, limited memory storage, limited Internet access and resource constraints. They attract

cybercriminals because of the open environment they operate in as well as their availability, Internet connectivity and storage capacity.

Existing Android botnet detection techniques are based on the estatic and dynamic approaches. These require heavy battery consumption, which is the most crucial challenge in protecting mobile devices (Ali, Zain, Zolkipli, & Badshah, 2014). Furthermore, the dynamic approach of scanning and blocking malicious codes needs a runtime environment. This is a big issue for mobile devices due to their limited battery power.

Internet Service Providers Should Also Provide Security Measurements

One of the biggest challenges for Internet Service Providers (ISP) is to protect mobile device users from the botnet threats as they do not use static addresses. Mobile device users are continuously changing their location; this creates difficulties for ISPs.

Difficulties in Estimating Botnet Size

It is very difficult to estimate the botnet size. To the best of our knowledge, there is no technique that estimates the size of compromised bots in a botnet. With rapid developments in detecting botnet attacks in Android devices, researchers need to find a quantitative methodology to find the number of bots in a botnet (Odili & Kahar, 2016).

CONCLUSION

Android botnets are harmful to Android devices. The popularity of mobile devices has made it a soft target for potential attacks. This survey aimed to find the real threat behind Android botnets. We conducted a comprehensive survey of existing Android botnets and their detection techniques. We categorised the detection techniques according to their detection environment, such as static, dynamic and hybrid detection techniques. Limitations in the existing Android botnet detection techniques as well as their benefits are listed here in an organised way. To the best of our knowledge this is the most current organised survey on Android botnets and their detection techniques. This research will help both academics and industry players.

ACKNOWLEDGEMENT

The authors are grateful to the Faculty of Computer Systems & Software Engineering (FSKKP), Universiti Malaysia Pahang for funding this research study under the Grant GRS140392.

REFERENCES

Abdelrahman, O. H., Gelenbe, E., Görbil, G., & Oklander, B. (2013). Mobile network anomaly detection and mitigation: The NEMESYS approach. *Information Sciences and Systems 2013* (pp. 429–438). Switzerland: Springer.

Ali, M., Zain, J. M., Zolkipli, M. F., & Badshah, G. (2014). Mobile cloud computing & mobile battery augmentation techniques: A survey. In *Research and Development (SCOReD), 2014 IEEE Student Conference* (pp. 1-6). IEEE.

Alparslan, E., Karahoca, A., & Karahoca, D. (2012). BotNet detection: Enhancing analysis by using data mining techniques. In A. Karahoca (Ed.), *Advances in Data Mining Knowledge Discovery and Applications*. INTECH Open Access Publisher.

Alta, V. D. M., Loock, M., & Dabrowski, M. (2005). Characteristics and responsibilities involved in a phishing attack. In *Proceedings of the 4th International Symposium on Information and Communication Technologies* (pp. 249-254). Trinity College Dublin.

Andronio, N., Zanero, S., & Maggi, F. (2015). HelDroid: Dissecting and detecting mobile ransomware. In *International Workshop on Recent Advances in Intrusion Detection* (pp. 382-404). Springer International Publishing.

Anwar, S., Inayat, Z., Zolkipli, M. F., Zain, J. M., Gani, A., Anuar, N. B., . . . & Chang, V. (2017). Cross-VM cache-based side channel attacks and proposed prevention mechanisms: A survey. *Journal of Network and Computer Applications, 93*, 259–279.

Anwar, S., Mohamad Zain, J., Zolkipli, M. F., & Inayat, Z. (2014). A review paper on botnet and botnet detection techniques in cloud computing. In *ISCI 2014 – IEEE Symposium on Computers and Informatics* (pp. 28-29). IEEE.

Anwar, S., Zain, J. M., Inayat, Z., Haq, R. U., Karim, A., & Jaber, A. N. (2016, August, 11–12,). A static approach towards mobile botnet detection. In *2016 3rd International Conference on Electronic Design (ICED)* (pp. 563-567). IEEE.

Anwar, S., Zain, J. M., Zolkipli, M. F., Inayat, Z., Jabir, A. N., & Odili, B. (2015). Response option for attacks detected by intrusion detection system. In *4th International Conference on Software Engineering and Computer System* (pp. 195-200). IEEE.

Anwar, S., Zain, J. M., Zolkipli, M. F., Inayat, Z., Khan, S., Anthony, B., & Chang, V. (2017). From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions. *Algorithms, 10*(2), 1–23.

Babu Rajesh, V., Reddy, P., Himanshu, P., & Patil, M. U. (2015). Droidswan: Detecting malicious Android applications based on static feature analysis. *Computer Science and Information Technology* (pp. 163-178). Centre for Development of Advanced Computing, India.

Bailey, M., Cooke, E., Jahanian, F., Xu, Y., & Karir, M. (2009). A survey of botnet technology and defenses. In *Conference for Homeland Security, 2009. CATCH'09. Cybersecurity Applications and Technology* (pp. 299-304). IEEE.

Bläsing, T., Batyuk, L., Schmidt, A. D., Camtepe, S. A., & Albayrak, S. (2010). An android application sandbox system for suspicious software detection. In *Malicious and Unwanted Software (MALWARE), 2010 5th International Conference* (pp. 55-62). IEEE.

Burguera, I., Zurutuza, U., & Nadjm-Tehrani, S. (2011). Crowdroid: Behavior-Based malware detection system for Android. In *Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices* (pp. 15-26). ACM.

Castiglione, A., De Prisco, R., De Santis, A., Fiore, U., & Palmieri, F. (2014). A botnet-based command and control approach relying on swarm intelligence. *Journal of Network and Computer Applications, 38*, 22–33.

Christian, F., & Maria, G. (2013). *Kaspersky security bulletin 2013*. Retrieved from http://www.securelist.com/en/

Cooke, E., Jahanian, F., & McPherson, D. (2005). The zombie roundup: Understanding, detecting, and disrupting botnets. In *Proceedings of the USENIX SRUTI Workshop* (pp. 6-6). USENIX.

Dhaya, R., & Poongodi, M. (2014). Detecting software vulnerabilities in Android using static analysis. Paper presented at the *Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference* (pp. 915-918). IEEE.

Dong, D., Wu, Y., He, L., Huang, G., & Wu, G. (2008). Deep analysis of intending peer-to-peer botnet. *In 2008 Seventh International Conference on Grid and Cooperative Computing* (pp. 407-411). IEEE.

Doorey, A. M. (2016). *Contextualizing privacy concerns within mobile engagement: a comparative investigation of escalating risk among general, e-commerce and health-related use.* (Doctoral dissertation). The University of Texas, Austin.

Dunne, P. (2006). *The art of pishing: How to attract birds by mimicking their calls*. United States of A,erica, USA: Stackpole Books.

Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B.-G., Cox, L. P., ... & Sheth, A. N. (2014). TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS), 32*(2), 5.

Enck, W., Ongtang, M., & McDaniel, P. (2009). On lightweight mobile phone application certification. In *Proceedings of the 16th ACM Conference on Computer and Communications Security* (pp. 235-245). ACM.

Faruki, P., Bharmal, A., Laxmi, V., Ganmoor, V., Gaur, M. S., Conti, M., & Rajarajan, M. (2015). Android security: A survey of issues, malware penetration, and defenses. *IEEE Communications Surveys and Tutorials, 17*(2), 998–1022.

Faruki, P., Ganmoor, V., Laxmi, V., Gaur, M. S., & Bharmal, A. (2013). Androsimilar: Robust statistical feature signature for android malware detection. In *Proceedings of the 6th International Conference on Security of Information and Networks* (pp. 152-159). ACM.

Fossi, M., Egan, G., Haley, K., Johnson, E., Mack, T., Adams, T., ... & McKinney, D. (2011). Symantec internet security threat report trends for 2010. *Symantec Enterprise Security, 16*, 1-20.

Gascon, H., Yamaguchi, F., Arp, D., & Rieck, K. (2013). Structural detection of android malware using embedded call graphs. In *Proceedings of the 2013 ACM Workshop on Artificial Intelligence and Security* (pp. 45-54). ACM.

Gu, G., Perdisci, R., Zhang, J., & Lee, W. (2008). BotMiner: Clustering analysis of network traffic for protocol-and structure-independent botnet detection. In *USENIX Security Symposium* (pp. 139-154). USENIX.

Gu, G., Porras, P. A., Yegneswaran, V., Fong, M. W., & Lee, W. (2007). BotHunter: Detecting malware infection through IDS-driven dialog correlation. Paper presented at the *USENIX Security Symposium* (Vol. 7, pp. 1-16). USENIX.

Gu, G., Zhang, J., & Lee, W. (2008). BotSniffer: Detecting botnet command and control channels in network traffic. In *NDSS* (Vol. 8, pp. 1-18).

Ianelli, N., & Hackworth, A. (2005). Botnets as a vehicle for online crime. *CERT Coordination Center, 1*(1), 28.

Inayat, Z., Gani, A., Anuar, N. B., Anwar, S., & Khurram Khan, M. (2017). Cloud-Based intrusion detection and response system: Open research issues, and solutions. *Arabian Journal for Science and Engineering, 7*(65), 1–25.

Inayat, Z., Gani, A., Anuar, N. B., Khan, M. K., & Anwar, S. (2016). Intrusion response systems: Foundations, design, and challenges. *Journal of Network and Computer Applications, 62*, 53–74.

Jelasity, M., & Bilicki, V. (2009). Towards automated detection of peer-to-peer botnets: On the limits of local approaches. In *2nd USENIX Conference on Large-scale Exploits and Emergent Threats (LEET)* .

Karim, A., Bin Salleh, R., Shiraz, M., Shah, S. A. A., Awan, I., & Anuar, N. B. (2014). Botnet detection techniques: Review, future trends, and issues. *Journal of Zhejiang University-Science C-Computers and Electronics, 15*(11), 943–983. doi:DOI 10.1631/jzus.C1300242

Karim, A., Shah, S. A. A., Salleh, R. B., Arif, M., Md Noor, R., & Shamshirband, S. (2015). Mobile botnet attacks – An emerging threat: Classification, review and open issues. *KSII Transactions on Internet and Information Systems, 9*(4), 1471–1492. doi:10.3837/tiis.2015.04.012

Khattak, S., Ramay, N., Khan, K., Syed, A., & Khayam, S. (2014). A taxonomy of botnet behavior, detection, and defense. *IEEE Communications Surveys and Tutorials, 16*(2), 898–924.

LAB, K. (2015). *Kaspersky security bulletin 2015*. Retrieved February, 2016, from www.kaspersky.com

Lashkari, A. H., Ghalebandi, S. G., & Moradhaseli, M. R. (2011). A wide survey on botnet. *Digital information and communication technology and its applications* (pp. 445–454): Berlin, Heidelberg: Springer.

Li, C., Jiang, W., & Zou, X. (2009). Botnet: Survey and case study. In *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference* (pp. 1184-1187). IEEE.

Lindorfer, M., Neugschwandtner, M., Weichselbaum, L., Fratantonio, Y., Van Der Veen, V., & Platzer, C. (2014). Andrubis-1,000,000 apps later: A view on current android malware behaviors. In *2014 Third International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)* (pp. 3-17). IEEE.

Liu, L., Chen, S., Yan, G., & Zhang, Z. (2008). Bottracer: Execution-based bot-like malware detection. In T. C. Wu, C. L. Lei, V. Rijmen & D. T. Lee (Eds.), *Information Security* (pp. 97–113). Germany: Springer.

LulzSec. (2011). *Distributed denial of service attack (DDoS) definition*. Retrieved from http://www.incapsula.com/

Milletary, J., & Center, C. C. (2005). *Technical trends in phishing attacks*. *Retrieved December, 1*(2007), 3-3.

Mirkovic, J., & Reiher, P. (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Computer Communication Review, 34*(2), 39–53.

Moonsamy, V., Rong, J., & Liu, S. (2014). Mining permission patterns for contrasting clean and malicious android applications. *Future Generation Computer Systems, 36*, 122–132.

Naraine, R. (2012, February 27). Android drive-by download attack via phishing sms: ZDNet. Zero Day. Retrieved from http://www.zdnet.com/article/android-drive-by-download-attack-via-phishing-sms/

Narudin, F. A., Feizollah, A., Anuar, N. B., & Gani, A. (2016). Evaluation of machine learning classifiers for mobile malware detection. *Soft Computing, 20*(1), 343–357.

Naser, A., Zolkipli, M. F., Majid, M., & Anwar, S. (2014). Trusting cloud computing for personal files. In *Information and Communication Technology Convergence (ICTC), 2014 International Conference* (pp. 488-489). IEEE.

Oberheide, J., & Miller, C. (2012). *Dissecting the android bouncer* (pp. 1-62). SummerCon2012, New York.

Odili, J. B., Kahar, M. N. M., Anwar, S., & Ali, M. (2017). Tutorials on African buffalo optimization for solving the travelling salesman problem. *International Journal of Software Engineering and Computer Systems, 3*(3), 120–128.

Odili, J. B., & Kahar, M. N. M. (2015). African buffalo optimization (ABO): A new meta-heuristic algorithm. *Journal of Advanced and Applied Sciences, 3*(3), 101–106.

Odili, J. B., & Kahar, M. N. M. (2016). Solving the traveling salesman's problem using the African buffalo optimization. *Computational Intelligence and Neuroscience, 2016*(2916), 1–12.

Odili, J. B., Kahar, M. N. M., & Anwar, S. (2015). African buffalo optimization: A swarm-intelligence technique. *Procedia Computer Science, 76*, 443–448.

Odili, J. B., Kahar, M. N. M., & Noraziah, A.. (2016). Convergence analysis of the African buffalo optimization algorithm. *International Journal of Simulations: Systems, Science and Technology, 17*(44), 44.41–44.46.

Odili, J. B., Kahar, M. N. M., Anwar, S., & Azrag, M. A. K. (2015). A comparative study of African buffalo optimization and randomized insertion algorithm for asymmetric travelling salesman's problem. In *Software Engineering and Computer Systems (ICSECS), 2015 4th International Conference* (pp. 90-95). IEEE.

Ongtang, M., McLaughlin, S., Enck, W., & McDaniel, P. (2009). Semantically rich application-centric security in Android. *Security and Communication Networks, 5*(6), 658-673.

Peng, S., Yu, S., & Yang, A. (2014). Smartphone malware and its propagation modeling: A survey. *IEEE Communications Surveys and Tutorials, 16*(2), 925–941.

Penning, N., Hoffman, M., Nikolai, J., & Wang, Y. (2014). Mobile malware security challeges and cloud-based detection. In *Collaboration Technologies and Systems (CTS), 2014 International Conference* (pp. 181-188). IEEE.

Plohmann, D., Gerhards-Padilla, E., & Leder, F. (2011). Botnets: Detection, measurement, disinfection and defence. *European Network and Information Security Agency (ENISA), 1*(1), 1-153.

Portokalidis, G., Homburg, P., Anagnostakis, K., & Bos, H. (2010). Paranoid Android: Versatile protection for smartphones. In *Proceedings of the 26th Annual Computer Security Applications Conference* (pp. 347-356). ACM.

Preda, M. D., Christodorescu, M., Jha, S., & Debray, S. (2008). A semantics-based approach to malware detection. *ACM Transactions on Programming Languages and Systems (TOPLAS), 30*(5), 25.

Rastogi, V., Chen, Y., & Enck, W. (2013). AppsPlayground: Automatic security analysis of smartphone applications. In *Proceedings of the Third ACM Conference on Data and Application Security and Privacy* (pp. 209-220). ACM.

Reina, A., Fattori, A., & Cavallaro, L. (2013). *A system call-centric analysis and stimulation technique to automatically reconstruct android malware behaviors*. EuroSec.

Roshandel, R., Arabshahi, P., & Poovendran, R. (2013). LIDAR: A layered intrusion detection and remediationframework for smartphones. In *Proceedings of the 4th international ACM Sigsoft symposium on Architecting Critical Systems* (pp. 27-32). ACM.

Rubin, A. (2008). Google bets on Android future. Retrieved from http://news.bbc.co.uk/2/hi/technology/7266201.stm

Sauter, D. (2013). *Rapid android development: Build rich, sensor-based applications with processing*. Raleigh, NC: Pragmatic Bookshelf.

Schultz, M. G., Eskin, E., Zadok, E., & Stolfo, S. J. (2001). Data mining methods for detection of new malicious executables. In *Security and Privacy, 2001 Symposium, S and P 2001. IEEE Symposium* (pp. 38-49). IEEE.

Sears, N. (2007). *Android.* Retrieved from http://www.openhandsetalliance.com/index.html

Shameli, S. A., Cheriet, M., & Hamou-Lhadj, A. (2014). Taxonomy of intrusion risk assessment and response system. *Computers and Security, 45*, 1–16. doi:10.1016/j.cose.2014.04.009

Sharma, M., Chawla, M., & Gajrani, J. (2016). A survey of Android malware detection strategy and techniques. In *Proceedings of International Conference on ICT for Sustainable Development* (pp. 39-51). Springer Singapore.

Sheta, M. A., Zaki, M., El Salam, K. A., & Hadad, E. (2015). Design and implementation of anti spyware system using design patterns. *International Journal of Computer Applications, 123*(2), 9-13.

Silva, R. M., Pinto, R. C., & Salles, R. M. (2013). Botnets: A survey. *Computer Networks, 57*(2), 378–403.

Skovoroda, A., & Gamayunov, D. (2015). Review of the mobile malware detection approaches. In *Parallel, Distributed and Network-Based Processing (PDP) Conference, 2015 23rd Euromicro International Conference* (pp. 600-603). IEEE.

Spreitzenbarth, M., Freiling, F., Echtler, F., Schreck, T., & Hoffmann, J. (2013). Mobile-sandbox: Having a deeper look into android applications. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing* (pp. 1808-1815). ACM.

Spreitzenbarth, M., Schreck, T., Echtler, F., Arp, D., & Hoffmann, J. (2015). Mobile-Sandbox: combining static and dynamic analysis with machine-learning techniques. *International Journal of Information Security, 14*(2), 141–153.

Stevanovic, M., Revsbech, K., Pedersen, J. M., Sharp, R., & Jensen, C. D. (2012). A collaborative approach to botnet protection. In *CD-ARES* (pp. 624-638).

Suarez-Tangil, G., Tapiador, J. E., Pens-Lopez, P., & Blasco, J. (2014). DENDROID: A text mining approach to analyzing and classifying code structures in Android malware families. *Expert Systems with Applications, 41*(4), 1104–1117. doi:10.1016/j.eswa.2013.07.106

Suarez-Tangil, G., Tapiador, J. E., Peris-Lopez, P., & Ribagorda, A. (2014). Evolution, detection and analysis of malware for smart devices. *Communications Surveys and Tutorials, IEEE, 16*(2), 961–987.

Szymczyk, M. (2009). Detecting botnets in computer networks using multi-agent technology. In *Dependability of Computer Systems, 2009. DepCos-RELCOMEX'09 Fourth International Conference* (pp. 192-201). IEEE.

Teufl, P., Ferk, M., Fitzek, A., Hein, D., Kraxberger, S., & Orthacker, C. (2013). Malware detection by applying knowledge discovery processes to application metadata on the Android market (Google Play). *Security and Communication Networks, 9*(5), 389-419.

Thompson, B., Morris-King, J., & Cam, H. (2016, November). Controlling risk of data exfiltration in cyber networks due to stealthy propagating malware. *In Military Communications Conference, MILCOM 2016-2016 IEEE* (pp. 479-484). IEEE.

Wang, K., Huang, C. Y., Lin, S. J., & Lin, Y. D. (2011). A fuzzy pattern-based filtering algorithm for botnet detection. *Computer Networks, 55*(15), 3275–3286. doi:10.1016/j.comnet.2011.05.026

Yan, L. K., & Yin, H. (2012). Droidscope: Seamlessly reconstructing the OS and Dalvik semantic views for dynamic Android malware analysis. In *21st USENIX Security Symposium* (USENIX Security 12) (pp. 569-584). USENIX.

Yu, X., Dong, X., Yu, G., Qin, Y., Yue, D., & Zhao, Y. (2010). Online botnet detection based on incremental discrete Fourier transform. *Journal of Networks, 5*(5), 568–576.

Zang, X., Tangpong, A., Kesidis, G., & Miller, D. J. (2011). Botnet detection through fine flow classification. *Unpublished, Report No. CSE11-001*.

Zhao, D., Traore, I., Sayed, B., Lu, W., Saad, S., Ghorbani, A., & Garant, D. (2013). Botnet detection based on traffic behavior analysis and flow intervals. *Computers and Security, 39*, 2–16. doi:10.1016/j.cose.2013.04.007

Zhao, M., Zhang, T., Ge, F., & Yuan, Z. (2012). RobotDroid: A lightweight malware detection framework on smartphones. *Journal of Networks, 7*(4), 715–722.

Zhou, Y., Wang, Z., Zhou, W., & Jiang, X. (2012). *Hey, you, get off of my market: Detecting malicious apps in official and alternative Android markets.* In *NDSS* (Vol. 25, No. 4, pp. 50-52).

Pertanika J. Sci. & Technol. 26 (1): 71 - 84 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

An Update on Type 1 Diabetes Treatments: Insulin Treatment, Cell Therapy and Transplantation

Homayoun Hani[1], Mohd-Azmi Mohd-Lila[1]*, Rasedee Abdullah[1], Zeenathul Nazariah Allaudin[1], Kazhal Sarsaifi[2] and Faez Firdaus Jesse Abdullah[2]

[1]Department of Veterinary Pathology and Microbiology, Faculty of Veterinary Medicine, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Department of Veterinary Clinical Studies, Faculty of Veterinary Medicine, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Diabetes is one of the major life-threatening health problems worldwide today. It is one of the most fast-growing diseases that cause many health complications and a leading cause of decreasing life expectancy and high mortality rate. Many studies have suggested several different types of intervention to treat Type 1 diabetes such as insulin therapy, islet transplantation, islet xenotransplantation and stem cell therapy. However, issues regarding the efficacy, cost and safety of these treatments are not always well addressed. For decades, diabetes treatments with few side effects and long-lasting insulin independence has remained one of the most challenging tasks facing scientists. Among the treatments mentioned above, application of human islet transplantation in patients with type 1 diabetes has progressed rapidly with significant achievement. Again, the lack of appropriate donors for islet transplantation and its high cost have led researchers to look for other alternatives. In this review, we discuss very pertinent issues that are related to diabetes treatments, their availability, advantages, disadvantages and also cost.

Keywords: Cell therapy, diabetes economy, diabetes treatments, insulin, islet transplantation, stem cell therapy

ARTICLE INFO

Article history:
Received: 07 April 2017
Accepted: 05 December 2017

E-mail addresses:
h.hani1975@gmail.com (Homayoun Hani),
azmi@upm.edu.my (Mohd-Azmi Mohd-Lila),
rasedee@upm.edu.my (Rasedee Abdullah),
zeenathulnazariah@gmail.com (Zeenathul Nazariah Allaudin),
kazhalsarsaifi@gmail.com (Kazhal Sarsaifi),
jesse@upm.edu.my (Faez Firdaus Jesse Abdullah)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Diabetes mellitus is a life-threatening disease that might be complicated by cardiovascular and kidney diseases, ketoacidosis and skin conditions (Habener, 2004; Nathan et al.,

2009). Diabetes mellitus is a fast-growing metabolic disease (Kaul et al., 2013; IDF, 2015). In 2015, the International Diabetes Federation estimated that worldwide, more than 415 million people live with diabetes and 5% of that population are diagnosed with type 1 diabetes (IDF, 2015; ADA, 2017). The incidence of diabetes is dramatically increasing and predicted to more than double by the year 2040 (IDF, 2015). Patients with diabetes are costly to maintain and in 2015, it was shown to be an economic burden on the health maintenance schemes of undeveloped and developed countries, amounting to USD673 billion per annum, which is equivalent to 12% of total health expenditure (Guariguata, 2012). In Malaysia, prevalence of diabetes is growing and is expected to rise to 21.6% of the adult population by 2020. Statistics show that patients with type 1 diabetics are 0.6% of the whole population with diabetes in Malaysia (IDF 2015).

Diabetes is a chronic disease caused when the pancreas either does not produce sufficient insulin or the body fails to use insulin (IDF, 2017). Insulin facilitates cellular uptake of glucose and regulates carbohydrates and fat metabolism. In type 1 diabetes, insulin secreting cells are destroyed by autoimmune attack (Table 1). Although insulin injections or treatment using synthetic medication may temporarily control diabetes, these drugs cannot cure the disease.

Table 1
Autoantibodies against pancreatic antigens in type 1 diabetes

Auto-Ab	Abbreviation	Antigen Expression	Explanation
Anti-insulin Abs	IAA	Pancreas	Detected in 50% type 1 diabetic children
Anti-glutamate decarboxilase Abs	GADA	Pancreas and nervous system	Found in 70%-80% newly diagnosed diabetics
Anti-insulinoma associated 2 Abs	IA2-A	Pancreas	Found in about 60% of type 1 diabetics
Anti-islet cell Abs	ICA	Pancreas	Detected in about 70%-80% newly diagnosed type 1 diabeticd
Abs against the zinc channel ZnT8	SCL338A	Pancreas	-

(Akerblom et al., 2002; American Association for Clinical Chemistry, 2016).

Recently, cell therapy was deemed to hold great potential for developing a permanent cure for diabetes. One of the most recent successes in cell therapy is islet transplantation in the bile duct of the liver in type 1 diabetics (Malka et al., 2000; Margener & Baillie, 1997). However, the main problem with this treatment method is the lack of compatible human islet sources for transplantation. Stem cells such as embryo and adult stem cells that can potentially differentiate into insulin secreting islet-like clusters and xenogeneic islet have shown some promising results for treatment of diabetes (Habener, 2004). However, the success of cell therapy in treating diabetes is limited by the lack of human pancreatic β-cells to produce insulin. Alternatively, it was suggested that other animal species could be used as sources of pancreatic β-cells. It was demonstrated in diabetic mice that some β-cells, for example, embryo blastocysts and

pancreas, liver, bone marrow and islet cells, from various mammalian species have the potential to reverse symptoms of diabetes (Wedemeyer et al., 2016).

Insulin and Cell Therapies for Diabetes

In the early 1920s, the extract of bovine pancreas cells injected in a diabetic patient reduced diabetic signs such as glycosuria and hyperglycemia. In 1921, insulin was discovered by Frederick G. Banting (Karamitsos, 2011). Since the advent of genetic engineering, many useful immunoproteins and hormones, including biosynthetic human insulin (Sara et al., 1998), can be prepared by recombinant DNA techniques and produced in various efficient expression systems including bacteria, yeast and mammalian cells (Razis et al., 2006; Abdul Razis et al., 2008, Tam et al., 2012). Direct treatment with naked DNA containing the gene of interest, the insulin gene, into the host tissue or selected organs is possible but this method faces a lot of uncertainty in terms of potential integration in the host genome that may lead to unwanted cell transformation. Until now, biosynthetic insulin and its analogues serve as the mainstay in diabetic treatment. Insulin therapy reduces diabetic signs; however, diabetics require frequent monitoring of blood glucose while experiencing various side effects, including hypoglycemia (Cryer et al., 2009), unusual ocular disturbance (Lee & Traboulsi, 2008), lipohyperthrophy (Blanco et al., 2013), hypersensitivity (Wong et al., 2007), anaphylaxis (Ghazavi & Johnston, 2011), hypertension (Arima et al., 2002), weight gain (Russell-Johnes & Khan, 2007), myocardial infarction (Malmberg et al., 2005), renal dysfunction (Patrick et al., 1992), hemolytic anemia (Dhaliwal et al., 2004) and gastrointestinal distress (Drugs, 2017).

Edmonton Protocol

In 1972, Lacy et al. (1972) showed that transplantation of islet allografts in the portal vein of pancreactomised patients reduced insulin independence for long periods. In 2000, the Edmonton Protocol was introduced by a University of Alberta research group (Shapiro et al., 2000; McCall & Shapiro, 2012). The Edmonton Protocol is a method of implantation of pancreatic islets for treatment of type 1 diabetes mellitus. This protocol reduced the risk of islet graft rejection and inhibitors such as inadequate islet cluster, insufficient prophylaxis, diabetogenic consumption and drugs that prevented attainment of insulin independence (Shapiro et al., 2000). According to the protocol, in order to gain sufficient islet quantity and complete insulin independence, islets of two to four donors are required to be harvested and transplanted into recipients. The optimum islet number for transplantation in type 1 diabetics is approximately 11,000 IEQ/kg body weight. In this protocol, the period of cold ischemia and the need for exposure to xenoprotein such as fetal calf serum, were minimised and the islets could be transplanted immediately after harvest without need of prior long-term freezing or *in-vitro* maintenance (Gaglia, Shapiro, & Weir, 2005). The Edmonton Protocol success rate is 80% with insulin independence during the first year, although in the long term the results are quite varied (Ryan et al., 2002; Shapiro et al., 2006).

Stem Cell Therapy

Stem cells are unspecialised cells of multicellular organs with potential for differentiation into other cell types during the embryonic period or in adults. Pluripotent and multipotent stem cells can differentiate to other types of cell tissue to repair dysfunctional cells or maintain tissue or organ sustainability (Habener, 2004; Fuchs et al., 2004; Wagers & Weissman, 2004; Young et al., 2004). Stem cells have three main properties, which are, multipotency as in adult tissues, pluripotency as in the blastocyst of embryos and totipotency as in a fertilised egg. The most distinguishable characteristic of stem cells is their capacity for self-renewal, transmutability to other cell types, extreme motility and immune resistance in the host (Habener, 2004; Abraham et al., 2004; Yang, 2004).

Islet progenitor cells (IPC) are found not only in the islets but also in the ducts and acinar tissue of the pancreas (Tang et al., 2004; Pessina et al., 2004). The bone marrow, liver, umbilical cord blood cells and embryo stem cells are the origin of IPC. The IPC differentiate into an islet-like cluster (ILC) and is shown *in vitro* to have huge potential of proliferation and insulin secretion upon exposure to growth factors such as fetal serum, epidermal growth factor (EGF), nerve growth factor (NGF) and fibroblast growth factor (FGF). Treatment with proliferation and differentiation inhibitors provides the ILC with the ability to express and secrete insulin, glucagon, somatostatin, glucagon-like peptide-1 (GLP-1) and pancreatic polypeptide (PP) (Lechner & Habener, 2003). In addition, primary epithelial cells harvested from the skin, umbilical cord and intestine also have the capacity to differentiate into islets when maintained in the appropriate environment and culture medium. However, the insulin production and functionality of these epithelial cells are lower than that of native pancreatic islets (Lechner & Habener, 2003).

Islet precursor cells in the adult pancreas has been suggested to impact pancreas development during embryonic damage, such as those caused by surgery and drugs (Dor et al., 2004). Pancreatic β-cells can still maintain insulin secretion even in cases of insult to pancreatic tissues (Dor et al., 2004). Multipotent stem cells can regenerate new cells from damaged islets and can most probably differentiate into functional β-cells (Dor et al., 2004).

Islet Transplantation

Islet transplantation has been suggested to be the ultimate treatment for the recovery of glucose homeostasis in patients with type 1 or late type 2 diabetes (Figure 1). Functional transplanted islets alleviate hypoglycemia and reestablish glucose homeostasis through the restoration of insulin production. Type 1diabetes can be cured by pancreas replacement. This was shown in pancreactomised diabetic dogs transplanted with fragments of the pancreas beneath the skin. The grafts kept the dogs alive even though they were not in physical contact with the digestive organs. This procedure was used in human experimentation with the grafting of sheep pancreas in a patient with type 1 diabetes. Unfortunately, although the graft improved the glycosuria, the patient did not survive after falling into adiabetic coma (Gaglia et al., 2005).

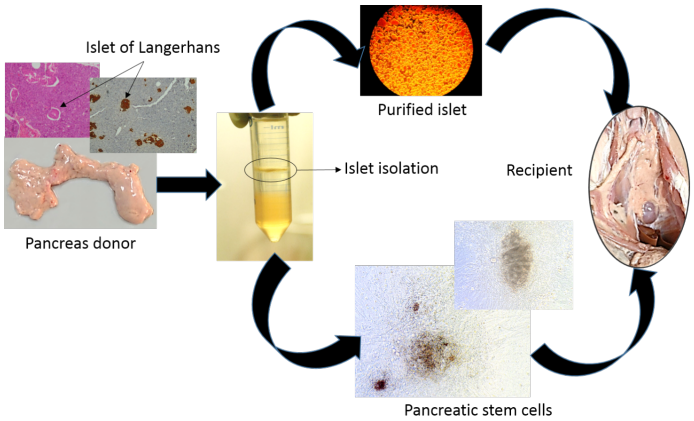


Figure 1. Cell therapy procedure in type 1 diabetes. Pancreata are collected from donors (human or animal), then islets or stem cells are harvested from pancreas. After islet purification and stem cell growth in culture, viable islets or cells are transplanted into the recipients' body

The use of improved surgical procedures and immunosuppression drugs have increased the success rate of vascularised intact pancreas transplantation. Between 1988 and 2016, there were 29,962 cases of successful vascularised pancreas transplantation worldwide (United Network for Organ Sharing, 2016). However, the use of islet transplantation to achieve insulin independence and alleviate complications of type 1 diabetes is suggested to be superior to whole pancreas transplantation or insulin therapy. Islet transplantation causes fewer complications and side effects than insulin therapy and it is easier to perform compared to whole pancreas transplantation. Insulin therapy is effective only if the patient fully subscribes to the consumption protocol and dosage and follow-up regimen and is supervised by a professional healthcare team. It should be noted that the aim of pancreas or islet transplantation is to achieve insulin independence, increase quality of life and diminish secondary diabetes complications (Robertson et al., 2010). At the early stages of islet transplantation experiments using induced diabetic rodents as models, the islets were shown to reverse diabetes. Transplantation of islet allografts at various sites on the body of diabetic rodents temporarily improved insulin requirement, but did not reverse diabetes for a period long enough to achieve complete insulin independence (Ballinger & Lacy, 1972; Najarian et al., 1977).

Islet Xenotransplantation in Diabetes

Currently, islet allotransplantation procedures are limited by lack of donor sources (Collaborative Islet Transplant Registry, 2009; Shapiro, 2011; Thompson et al., 2011; Hani et al., 2010; Hani et al., 2014). For that reason, over the last decade, fewer than 1000 islet transplantation procedures have been performed worldwide (Collaborative Islet Transplant Registry, 2009). To overcome the shortage, alternative sources of islets have been proposed including pigs, non-human primates, cattle, sheep, goats and fish (Hani et al., 2010; Hani et al.,

2014; Vakhshiteh et al., 2013; Hani e al., 2015; Hani et al., 2016; Hani et al., 2017; Kean et al., 2006). Although the best source of islets for human transplantation are non-human primates, due to issues like genetic homogeny, and ethics, safety and logistics, other sources have been investigated. Pig islets are one potential source for xenotransplantation for humans because of compatibility based on similarity to insulin molecules and glucose kinetics between these species (van der Windt et al., 2012). However, among Muslims, tissues from porcine sources are not generally acceptable for human transplantation.

As with all tissue transplantations, outbred diversity, heterologous immunity and MHC expressions are barriers to long-term xenograft survival; thus, islet transplantation needs to address these issues of immune response and rejection, for instance, through immunosuppression and encapsulation (Thompson et al., 2011).

Safety Issues in Islet Therapy for Diabetes Mellitus

Stem cell therapy. Stem cells such as mesenchymal, embryonic and hematopoietic stem cells, have the potential for differentiation into insulin-secreting or islet-like cluster cells. Some pancreatic islet cells or islet progenitor cells can differentiate into new insulin-secreting cells to replace injured and old β-cells that are undergoing apoptosis (Figure 1) (Kirk et al., 2014). In cell therapy for type 1 diabetes, adult stem cells are probably more suitable than embryonic stem cells. Adult pluripotential stem cells are teratogenic and their use give rise to fewer ethical issues. Adult pluripotential stem cells readily differentiate into the tissue cells of their origin and would be of greater application in cell replacement therapy for diabetes. However, diabetes treatment methods using these stem cells are still not fully developed for use without dire adverse consequences (Habener, 2004).

Safety of islet transplantation. Islets are fraught with the tendency to undergo apoptosis, diminished functionality and viability during purification (Lipsett et al., 2006). The isolation and purification of islets from the pancreas are governed by the presence of growth factors, supportive matrix and physical and chemical stresses, such as osmotic, hypoxia and mechanical stresses, that determine the survival of the cells. Thus, optimisation of the islet isolation procedure is imperative for obtaining viable cells for transplantation.

The success of islet transplantation is also determined by the site on the body, glucose and lipid concentrations (Lipsett et al., 2006) and the immunosuppressive drugs used. Graft sites that are naturally well-nourished and with an environment conducive for transplantation will increase the success of the treatment. In a recent study, the islets were first encapsulated with an immunprotector to avoid destruction by the immune system while retaining their ability to communicate with the environment (Lipsett et al., 2006).

Although anti-rejection regimens for both islet and pancreas transplantation recipients are the same, however, complications arising from pancreas transplantation are higher than from islet transplantation (Moassesfar et al., 2016). Whole pancreas transplantation is most common

for patients with pancreatitis and pancreas dysfunction or failure, while islet cell transplantation is usually recommended for those patients with islet mass loss or destruction, mostly in type 1 diabetes. Therefore, in type 1 diabetes, it is not necessary to transplant the whole pancreas even if it is less costly compared to islet transplantation. As the pancreas is a multifunction organ, if the graft is rejected by the immune system of recipients, then the whole organ, rather than just the islet cells, must be replaced.

Islet xenotransplantation. Safety protocols for islet transplantation are also applicable for islet allotransplantation and xenotransplantation. Among issues associated with transplantation is the transmission of infections. To ensure success of transplantation, precautions have to be taken to avoid transmission of infectious agents from the donor and the environment to the recipient (U.S. Food and Drug Administration, 2003; Mueller et al., 2011). One threat of infection concerns the use of porcine islet cells. Porcine endogenous retrovirus infection in islets is often a threat in xenotransplantation. Other infectious agents such as the cytomegalo virus, herpes virus and lymphotropic herpes virus as well as bacteria that are resident in porcine islets can pose a threat to recipients. Thus, to avoid infections, the animals serving as sources of islets must first be screened to ensure they are free from zoonotic organisms. However, it is often difficult to avoid contamination by infectious agents because the porcine retrovirus genome, for instance, is integrated in the animal genome and can go undetected and in this way, be transmitted to recipients during transplantation (van der Windt et al., 2012; Zhu et al., 2014).

Economics

To ensure the safety of recipients, animals like pigs that are the source of islets must be bred in expensive sterile and clean facilities. It is estimated that a facility complete with the equipment for cleaning breeding of 100 animals can cost more than USD10 million with a maintenance cost of between USD1 to 2 million per year. However, these breeding facilities are necessary to ensure supply of islets that are clean and safe for human use (van der Windt et al., 2012).

Based on reports, the cost of insulin therapy over the long term is higher than islet cell transplantation (Berwick, 2016). In 2016, the cost of insulin therapy was estimated at $71,000 per quality-adjusted life year (QALY) (Berwick, 2016), while for islet transplantation it was estimated at $50,000/QALY. The initial cost of islet transplantation is higher than that of insulin therapy, but over the years, due to continuous and long-term application, the total cost of insulin therapy eventually becomes much higher than that of islet transplantation (Beckwith et al., 2012).

Pancreas transplantation is marginally cheaper at a total cost of USD135,000 than islet cell transplantation at USD139,000. These expenses were calculated based on the cost of transplantation procedures and hospitalisation after surgery (Table 2).

Table 2
Comparison between treatments for type 1 diabetes

Therapy	Type	Global practice rate	Cost	Advantages	Disadvantages
Insulin Injection	Recombinant Animal	23.9 million (2000) 60,000 injections throughout life-time[a]	$663,000/20years $71,000 QALYs[b]	Easy access and personally practicable; Temporarily capable of controlling blood glucose level[c]	Hypoglycemia, unusual ocular disturbance, dermatologic reaction (lipohyperthrophy), hypersensitivity, immunologic response (anaphylaxis), hypertension, weight gain, myocardial infarction, renal dysfunction, hemolytic anemia and gastrointestinal distress[c]
Pancreas Transplantation	Intact Partial	2328/year[d]	$ ~134,750[e] (Mean total cost incorporating complications)	One proper donor may be sufficient[e]	Proper donor shortage; Need surgery; High rejection risk[e]
Islet Transplantation	Auto Allo	<100/year[f]	$659,000/20year $61,000 QALYs $ ~139,000 (Mean total cost incorporating second islet transplants (ITA)[b]	No surgery procedures necessary; Less post-transplantation complication compared to pancreas transplantation; Consistent islet yield[c]	Lack of donors; More than one donor may be needed to achieve insulin independency; Need isolation processing; Transplanted β-islet cells may be rejected within several years and must be repeated every couple of years at a cost of $120,000 per transplant[b,e]
Islet Xenotransplantation	Fatal	Not routinely practised	$60,700 QALYs[b]	No isolation procedure necessary; Proliferation and maturation *in vivo*	Not fully functional until >4 months after transplantation; Nonsurvival C-section in sow; Need for many foetuses
	Neonatal			No need for harmful purification process; Proliferation *in vitro* and *in vivo* after transplantation; Resistance against hypoxia; Preferable breeding logistics	Not fully functional until >4 weeks
	Young			More preferable breeding logistics vs. adult >2 years	Fragility of islets; Difficult to obtain consistent yields; Less preferable breeding logistics vs. neonatal[57]
	Adult			Consistent islet yields	Non-preferable breeding logistics; High costg
Stem cells	Polypotent Multipotent	N/A	N/A	Can be harvested from adult tissues (Habener, 2004); Unlimited sources of β-cell[h]	Teratogenic behaviour; Auto-immune attack by recipient body against autograft stem cells[h]

[a)] Al-Tabakha et al. (2008); [b)] Beckwith et al. (2012); [c)] Insulin side effects (2017); [d)] Estimated number of organ transplantations worldwide in 2014 (2014); e) Moassesfar et al. (2016); [f)] Islet Transplantation Technology (2010); [g)] van der Windt et al. (2012); [h)] Habener (2004).

CONCLUSION

Cell therapy, whether with islets or stem cells, is a promising treatment for type 1 diabetes that could provide long-lasting insulin independence. However, these methods of therapy for diabetics are limited by cell quality, donor availability and financial constraints. In the final analysis, the choice of therapeutic means is governed by reliability of the technique to provide a long-lasting cure for diabetes. Currently, cell therapy is only affordable by the affluent and is not within the means of low-income patients. Given more time and with greater advances in technology, cell therapy for diabetics may be affordable for all.

REFERENCES

Abdul Razis, A. F., Ismail, E. N., Hambali, Z., Abdullah, M. N. H., Ali, A. M., & Mohd Lila, M. A. (2008). Expression of recombinant human epidermal growth factor in Escherichia coli and characterization of its biological activity. *Applied Biochemistry and Biotechnology, 144*, 249–26.

Abraham, E. J., Kodama, S., Lin, J. C., Ubeda, M., Faustman, D. L., & Habener, J. F. (2004). Human pancreatic islet-derived progenitor cell engraftment in immunocompetent mice. *American Journal of Pathology, 164*, 817–830.

Akerblom, H. K., Vaarala, O., Hyoty, H., Ilonen, J., & Knip, M. (2002). Environmental factors in the etiology of type 1 diabetes. *American Journal of Medical Genetics, 115*, 18–29.

Al-Tabakha, M. M., & Arida, A. I. (2008). Recent challenges in insulin delivery systems: A review. *Indian Journal of Pharmaceutical Sciences, 70*, 278–286.

American Association for Clinical Chemistry. (2016). Diabetes-related Autoantibodies. Retrieved from https://labtestsonline.org/understanding/analytes/ diabetes-auto/tab/test/

American Diabetes Association. (2017). Type 1 Diabetes. Retrieved from http://www.diabetes.org/diabetes-basics/type-1/?loc=util-header_type1.

Arima, H., Kiyohara, Y., Kato, I., Tanizaki, Y., Kubo, M., Iwamoto, H. … Fujishima, M. (2002). Alcohol reduces insulin-hypertension relationship in the general population: The Hisayama study. *Journal of Clinical Epidemiology, 55*, 863–869.

Ballinger, W. F., & Lacy, P. E. (1972). Transplantation of intact pancreatic islets in rats. *Surgery, 72*, 175–186.

Beckwith, J., Nyman, J. A., Flanagan, B., Schrover, R., & Schuurman, H. J. (2012). A health economic analysis of clinical islet transplantation. *Clinical Transplantation, 26*, 23–33.

Berwick, D. M. (2016). Era 3 for medicine and health care. *Journal of the American Medical Association, 315,* 13.

Blanco, M., Hernandez, M. T., Strauss, K. W., & Amaya, M. (2013) Prevalence and risk factors of lipohypertrophy in insulin-injecting patients with diabetes. *Diabetes Metabolism, 39*, 445–453.

Collaborative Islet Transplant Registry. (2009, November 1). Sixth annual report. Retrieved from http://www.citregistry.com/. Accessed 2011, July 29.

Cryer, P. E., Axelrod, L., Grossman, A. B., Heller, R. S., Montori, V. M., Seaquist, E. R., & Service, F. J. (2009). Endocrine society. Evaluation and management of adult hypoglycemic disorders: An endocrine society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism, 94*, 709–728.

Dhaliwal, G., Cornett, P. A., & Tierney, L. M. (2004). *Journal of Hemolyticanemia. American Family Physician, 69*, 2599–2606.

Dor, Y., Brown, J., Martinez, O. I., & Melton, D. A. (2004). Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature, 429*, 41–46.

Drugs. (2017). Insulin side effects. Retrieved from https://www.drugs.com/sfx/ insulin-side-effects.

Estimated number of organ transplantations worldwide in 2014. (2014). *Statista*. Retrieved from https://www.statista.com/statistics/398645/global-estimation-of-organ-transplantations/.

Feisul, M. I., & Azmi, S. (2013). National diabetes registry report, volume 1, 2009-2012. Kuala Lumpur, Ministry of Health Malaysia. Retrieved from https://www.scribd.com/document/332019548/National-Diabetes-Registry-Report -Vol-1-2009-2012.

Fuchs, E., Tumbar, T., & Guasch, G. (2004). Socializing with the neighbors: Stem cells and their niche. *Cell, 116*, 769–778.

Gaglia, J. L., Shapiro, A. M., & Weir, G. C. (2005). Islet transplantation: Progress and challenge. *Archive of Medical Research, 36*, 273–280.

Ghazavi, M. K., & Johnston, G. A. (2011). Insulin allergy. *Clinical Dermatology, 3*, 300–305.

Guariguata, L. (2012). New estimates from the id diabetes atlas update for 2. *The Global Campaign, 57*.

Habener, J. F. (2004). A perspective on pancreatic stem/progenitor cells. *Pediatric Diabetes, 5*, 29–37.

Hani, H., Allaudin, Z. N., Ibrahim, T. A. T., Mohd-Lila, M. A., Sarsaifi, K., Camalxaman, S. N., & Othman, A. M. (2015). Morphological changes of post-isolation of caprine pancreatic islet. *Vitro Cellular and Developmental Biology-Animal, 51*, 113–120.

Hani, H., Allaudin, Z. N., Mohd-Lila, M. A., Ibrahim, T. A. T., & Othman, A. M. (2014). Caprine pancreatic islet xenotransplantation into diabetic immunosuppressed BALB/c mice. *Xenotransplantation, 21*, 174–182.

Hani, H., Allaudin, Z. N., Mohd-Lila, M. A., Sarsaifi, K., Rasouli, M., Tam, Y. J. … Othman, A. M. (2017). Improvement of isolated caprine islet survival and functionality in vitro by enhancing of PDX1 gene expression. *Xenotransplantation, 24*, DOI: 10.1111/xen.12302.

Hani, H., Nazariah, Allaudin, Z., Mohd-Lila, M. A., Sarsaifi, K., Tengku-Ibrahim, T. A., & Mazni Othman, A. (2016). Evaluation of caprine pancreatic islets cytoarchitecture by laser scanning confocal microscopy and flow cytometry. *Xenotransplantation, 23*, 128–136.

Hani, H., Tengku-Azmi, T. I., Abas, M. O., Mohd-Azmi, M. L., & Zeenathul, N. A. (2010). Isolation, density purification, and in vitro culture maintenance of functional caprine islets of Langerhans as an alternative islet source for diabetes study. *Xenotransplantation, 17*, 469–480.

IDF Annual Report. (2015). *Diabetes mellitus*. Retrieved from http://www.idf.org/sites/default/files/IDF_AnnualReport_2015_WEB.pdf.

IDF (2017). *About diabetes (What is diabetes?)*. Retrieved from http://www.idf.org/about-diabetes/what-is-diabetes.html.

Insulin side effects. (2017). Drugs.com. Retrieved from https://www.drugs.com/ cons/insulin-parenteral.html.

Islet Transplantation Technology. (2010). Revivicor. Retrieved from http://www.revivicor.com/islettech.htm.

Ismail, R., Allaudin, Z. N. & Lila, M. A. M. (2012). Scaling-up recombinant plasmid DNA for clinical trial: Current concern, solution and status. *Vaccine, 30*, 5914–5920.

Karamitsos, D. T. (2011). The story of insulin discovery. *Diabetes Research and Clinical Practice, 93*, 2–8.

Kaul, K., Tarr, J. M., Ahmad, S. I., Kohner, E. M., & Chibber, R. (2013). *Introduction to diabetes mellitus.* In S. I. Ahmad (Ed), *Diabetes* (pp 1–11). Austin, Texas, USA. Springer.

Kean, L., Gangappa, S., Pearson, T., & Larsen, C. (2006). Transplant tolerance in non-human primates: Progress, current challenges and unmet needs. *American Journal of Transplantation, 6*, 884–893.

Kirk, K., Hao, E., Lahmy, R., & Itkin-Ansari, P. (2014). Human embryonic stem cell derived islet progenitors mature inside an encapsulation device without evidence of increased biomass or cell escape. *Stem Cell Research, 12*, 807–814.

Lacy, P. E., Walker, M. M., & Fink, C. J. (1972). Perifusion of isolated rat islets in vitro: Participation of the microtubular system in the biphasic release of insulin. *Diabetes, 21*, 987–998.

Lechner, A., & Habener, J. F. (2003). Stem/progenitor cells derived from adult tissues: Potential for the treatment of diabetes mellitus. *American Journal of Physiology and Endocrinology Metabolism, 284*, 259–266.

Lee, B. J., & Traboulsi, E. I. (2008). Update on the morning glory disc anomaly. *Ophthalmic Genetics, 29*, 47–52.

Lipsett, M., Aikin, R., Castellarin, M., Hanley, S., Jamal, A. M., Laganiere, S., & Rosenberg, L. (2006). Islet neogenesis: A potential therapeutic tool in type 1 diabetes. *The International Journal of Biochemistry and Cell Biology, 38*, 715–720.

Malka, D., Hammel, P., & Sauvanet, A. (2000). Risk factors for diabetes mellitus in chronic pancreatitis. *Gastroenterology, 119*, 1324–1332.

Malmberg, K., Ryden, L., Wedel, H., Birkeland, K., Bootsma, A., Dickstein, K. … Waldenström, A. (2005). Intense metabolic control by means of insulin in patients with diabetes mellitus and acute myocardial infarction (DIGAMI 2): Effects on mortality and morbidity. *European Heart Journal, 26*, 650–661.

Margener, K., & Baillie, J. (1997). Chronic pancreatitis. *Lancet. 340*, 1379–1385.

McCall, M., & Shapiro, A. J. (2012). Update on islet transplantation. Cold Spring Harbor *Perspectives in Medicine, 2*, a007823.

Moassesfar, S., Masharani, U., Frassetto, L. A., Szot, G. L., Tavakol, M., Stock, P. G., & Posselt, A. M. (2016). A comparative analysis of the safety, efficacy, and cost of islet versus pancreas transplantation in nonuremic patients with type 1 diabetes. *American Journal of Transplantation, 16*, 518–526.

Mueller, N. J., Takeuchi, Y., Mattiuzzo, G., & Scobie, L. (2011). Microbial safety in xenotransplantation. *Current Opinion in Organ Transplantation, 16*, 201–206.

Najarian, J. S., Sutherland, D. E., Matas, A. J., Steffes, M. W., Simmons, R. L., & Goetz, F. C. (1977). Human islet transplantation: A preliminary report. *Transplantation Proceeding, 9*, 233–236.

Nathan, D. M., Zinman, B., Cleary, P. A., Backlund, J. Y. C., Genuth, S., Miller, R. … Diabetes Control and Complications Trial. (2009). Modern-day clinical course of type 1 diabetes mellitus after 30 years' duration: The diabetes control and complications trial/epidemiology of diabetes interventions and complications and Pittsburgh epidemiology of diabetes complications experience (1983-2005). *Archives of Internal Medicine, 169*, 1307–1316.

Patrick, A. W., Hepburn, D. A., Swainson, C. P., & Frier, B. M. (1992). Changes in renal function during acute insulin-induced hypoglycaemia in patients with type 1 diabetes. *Diabetic Medicine, 9*, 150–155.

Pessina, A., Eletti, B., Croera, C., Savalli, N., Diodovich, C., & Gribaldo, L. (2004). Pancreas developing markers expressed on human mononucleated umbilical cord blood cells. *Biochemical and Biophysics Research and Communication, 323*, 315–322.

Razis, A. F. A., Ismail, E. N., Hambali, Z., Abdullah, M. N. H., Ali, A. M., & Lila, M. A. M. (2006). The periplasmic expression of recombinant human epidermal growth factor (hEGF) in Escherichia coli. *Asia-Pacific Journal of Molecular Biology and Biotechnology, 14*, 41–45.

Robertson, R. P., Nathan, D. M., & Muylder, J. E. (2010). *Pancreas and islet transplantation in diabetes mellitus*. Retrieved from http://www.uptodate.com.

Russell-Jones, D., & Khan, R. (2007). Insulin-associated weight gain in diabetes – Causes, effects and coping strategies. *Diabetes Obesity and Metabolism, 9*, 799–812.

Ryan, E. A., Lakey, J. R., Paty, B. W., Imes, S., Korbutt, G. S., Kneteman, N. M. … & Rajotte, R. V. (2002). Clinical outcomes and insulin secretion after islet transplantation with the Edmonton protocol. *Diabetes, 50*, 710–719.

Sahu, S., Tosh, D., & Hardikar, A. A. (2009). New sources of β-cells for treating diabetes. *Journal of Endocrinology, 202*, 13–16.

Sara L. Noble, Pharm, D., Elizabeth Johnson, M. S. ED., & Bill Walton, D. O. (1998). Insulin lispro: A fast-acting insulin analog. *American Family Physician, 57*, 279–286.

Shapiro, A. M., Lakey, J. R., Ryan, E. A., Korbutt, G. S., Toth, E., Warnock, G. L. … & Rajotte, R. V. (2000). Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *New England Journal of Medicine, 343*, 230–238.

Shapiro, A. M. (2011). State of the art of clinical islet transplantation and novel protocols of immunosuppression. *Current Diabetes Reports, 11*, 345–354.

Shapiro, A. M., Ricordi, C., Hering, B. J., Auchincloss, H., & Lindblad, R. (2006). International trial of the Edmonton protocol for islet transplantation. *New England Journal of Medicine, 355*, 1318–1330.

Sutherland, D. E., Matas, J. A., Goetz, F. C., & Najarian, J. S. (1980). Transplantation of dispersed islet tissue in humans: Autografts and allografts. *Diabetes, 29*, 31–44.

Tam, Y. J., Allaudin, Z. N., Lila, M. A. M., Bahaman, A. R., Tan, J. S., & Rezaei, M. A. (2012). Enhanced cell disruption strategy in the release of recombinant hepatitis B surface antigen from Pichia pastoris using response surface methodology. *BMC Biotechnology, 12*, art. no. 70, DOI: 10.1186/1472-6750-12-70.

Tang, D. Q., Cao, L. Z., Burkhardt, B. R., Xia, C. Q., Litherland, S. A., Atkinson, M. A., & Yang, L. J. (2004). In vivo and in vitro characterization of insulin-producing cells obtained from murine bone marrow. *Diabetes, 53*, 1721–1732.

Thompson, D. M., Meloche, M., Ao, Z., Paty, B., Keown, P., Shapiro, R. J. … & Begg, I. (2011). Reduced progression of diabetic microvascular complications with islet cell transplantation compared with intensive medical therapy. *Transplantation, 91*, 373–378.

U. S. Food and Drug Administration. (2003). Guidance for industry: Source animal, product, preclinical, and clinical issues concerning the use of xenotransplantation products in humans. Retrieved from http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Xenotransplantation/ucm074354.htm.

United Network for Organ Sharing. (2016). Retrieved from https://www.unos.org/data.

Vakhshiteh, F., Allaudin, Z. N., Lila, M., & Hani, H. (2013). Size-related assessment on viability and insulin secretion of caprine islets in vitro. *Xenotransplantation, 20*, 82–88.

van der Windt, D. J., Bottino, R., Kumar, G., Wijkstrom, M., Hara, H., Ezzelarab, M. … & Ayares, D. (2012). Clinical islet xenotransplantation: How close are we? *Diabetes, 61*, 3046–3055.

Wagers, A. J., & Weissman, I. L. (2004). Plasticity of adult stem cells. *Cell, 116*, 639–648.

Wedemeyer, M., Bederman, S., & Steward, O. (2016). Stem cells and management of healthcare costs: Stem cell-based treatments and the societal balance sheet. *Journal of Regenerative Medicine, 5*. doi:10.4172/2325-9620.1000125.

Wong, J. T., Kim, P. T., Peacock, J. W., Yau, T. Y., Mui, A. L., Chung, S. W., … & Ong, C. J. (2007). Pten (phosphatase and tensin homologue gene) haploinsufficiency promotes insulin hypersensitivity. *Diabetologia, 50*, 395–403.

Yang, Y. G. (2004). Application of xenogeneic stem cells for induction of transplantation tolerance: Present state and future directions. *Springer Seminars Immunopathology, 26*, 187–200.

Young, H. E., Duplaa, C., Romero-Ramos, M., Chesselet, M. F., Vourc'h, P., Yost, M. J. … & Tamura-Ninomiya, S. (2004). Adult reserve stem cells and their potential for tissue engineering. *Cell Biochemistry and Biophysics, 40*, 1–80.

Zhu, H., Wang, W., Yu, L., & Wang, B. (2014). Pig-islet xenotransplantation: Recent progress and current perspectives. *Frontiers in Surgery, 1*, 1–8.

Pertanika J. Sci. & Technol. 26 (1): 85 - 94 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

The Simulation of a Force in Micro-scale Sensing Employing an Optical Double Ring Resonator System

Youplao, P.[1]*, Tasakorn, M.[1] and Phattaraworamet, T.[2]

[1]Department of Electrical Engineering, Faculty of Industry and Technology, Rajamangala University of Technology Isan Sakon Nakhon Campus, Sakon Nakhon 47160, Thailand
[2]Department of Electrical and Computer Engineering, Faculty of Science and Engineering, Kasetsart University, Chalermphrakiat Sakon Nakhon Province Campus, Sakon Nakhon 47000, Thailand

ABSTRACT

This paper presents an optical double-ring resonator system of which the design and analytical model are demonstrated to be useful as a novel force in micro-Newton measurement-sensing devices based on optical sensors. The sensing application can be accomplished by changing the optical filtering characteristic of an optical resonance structure such as the ring resonator system. Together with the concept of stress/strain and the elastic modulus of the waveguide material, the relationship between a slightly different value in the exerted force acting on the sensing unit and a difference in the waveguide length can be evaluated. Indeed, changing the optical path length (the waveguide length) causes the difference in peak spectrum of the filtering signals obtained from a ring resonator system. Hence, by measuring the spacing shift between the sensing and setting peak signal in the considered channel, the measurement of a slightly different value in the exerted forces on the sensing unit can be achieved. From the simulation results, an exerted force in small-scale ranges from 10 μN to 50 μN have been evaluated by measuring a spacing shift between the peak signals ranging from 35 pm to 225 pm. In this study, the potential of using such a double-ring resonator device for a force in micro-Newton sensing application is studied and discussed.

Keywords: Force sensing, optical sensors, optical resonance, optical filter, ring resonator

INTRODUCTION

Force sensing is crucial especially in micro-manipulation for monitoring mechanical behaviour or generating feedback signals for robotic systems in order to execute reliable operations and avoid damaging fragile objects such as the micro-force sensor used in minimally invasive surgery (Peirs et al., 2004; Baki et al., 2012; Abushagur et al., 2014). In

ARTICLE INFO

Article history:
Received: 02 June 2016
Accepted: 07 November 2017

E-mail addresses:
phichai.yo@rmuti.ac.th (Youplao, P.),
mtasakorn@yahoo.fr (Tasakorn, M.),
thanawat.pha@ku.th (Phattaraworamet, T.)
*Corresponding Author

ISSN: 0128-7680

addition, when manipulating biological cells, encountered forces are typically found in micro-Newton measurement, which the sensing devices use through microfabrication techniques. Various research methods in this field have been used such as the design and fabrication of a force sensor based on silicon micro-machining technology (Jin & Mote, 1998; Cappelleri et al., 2011), magnetic resonance imaging as a fibre-optic force sensor (Tokuno et al., 2008; Marchi et al., 2016), a microfiber-coupler-based reflective sensor, fabricated by fusing two twisted optical fibres and then connecting the two ends to form a Sagnac loop (Chen et al., 2014) and a parallel plate structure (Tan et al., 2011). Some methods used indirect measurement techniques as capacitance-based force measurement (Enikov & Nelson, 2000), with many micromachining steps of fabrication work (Zadeh et al., 2017). The piezo-resistive cantilever force measurement technique has also been used through electrical conversion of small gauges (Beyeler et al., 2007; Li et al., 2009; Alcheikh et al., 2013; Caseiro et al., 2014) and by using a composite material consisting of a conductive ink and silicon elastomer (Cho & Ryuh, 2016). However, the implementation of these measurement methods are difficult to conduct in micro-electromechanical systems due to complications in using the devices and the fabrication costs.

Ring resonators have been receiving extensive attention as useful micro-components, without any unnecessary electronic parts, in optical systems. Many of the potential applications have been investigated and proposed, such as optical logic gate operators (Bao et al., 2014; Rakshit & Roy, 2014), optical switch (Yan et al., 2010), nonlinear signal processing (Dumeige et al., 2005; Yupapin & Suwancharoen, 2007) and optical filters (Yang et al., 2003; Liua et al., 2007; Ma & Ogusu, 2011). In this paper, the design and simulation results of the proposed system consisting of two micro-ring resonators to be used as a force-sensing device are presented. The potential of using such a proposed system for measuring force in micro-scale levels is also examined and discussed.

THE PROPOSED RING RESONATOR SYSTEM

The proposed ring resonator system, of which the schematic diagram is shown in Figure 1, is employed as a micro-force sensing device. The system consists of two ring resonators, R_1 with an optical input port, and R_2 as a sensing unit with an optical drop port. In operation, a Gaussian light pulse from a monochromatic light source is used as the optical input signal that launches the system. The input signal is considered a function that consists of a constant light field amplitude, E_0, and random phase modulation, which is the combination of an attenuation, α, and a phase constant, ϕ_0, resulting in temporal coherence degradation. In the equation where L is the propagation distance, the time-dependent input optical field, $E_{in}(t)$, can be expressed as:

$$E_{in}(t) = E_0 \exp\left[-\alpha L + j\phi_0(t)\right] \quad [1]$$

When the Gaussian pulse is propagating through the ring resonator system, in the direction shown in Figure 1, the resonant optical fields are formed and then appear separately at the

throughput port and drop port of the system, which are represented by E_{th} and E_{drop}, respectively. The relations between each optical field output, E_{th} or E_{drop}, and the optical field input, E_{in}, in each round trip calculation can be expressed as:

$$\left|\frac{E_{th}}{E_{in}}\right|^2 = \frac{(1-\kappa_1)y_1^2 - 2x_1x_2y_1\sqrt{1-\kappa_1}e^{-\frac{\alpha}{2}L_1}\cos(knL_1) + y_1y_2e^{-\alpha L_1}}{y_1^2 - 2x_1x_2y_1\sqrt{1-\kappa_1}e^{-\frac{\alpha}{2}L_1}\cos(knL_1) + (1-\kappa_1)y_1y_2e^{-\alpha L_1}} \quad [2]$$

$$\left|\frac{E_{drop}}{E_{in}}\right|^2 = \frac{y_1\kappa_1\kappa_2\kappa_3e^{-\frac{\alpha}{2}(L_1+L_2)}}{y_1^2 - 2x_1x_2y_1\sqrt{1-\kappa_1}e^{-\frac{\alpha}{2}L_1}\cos(knL_1) + (1-\kappa_1)y_1y_2e^{-\alpha L_1}} \quad [3]$$

where, the optical field, E_i, propagates within each part, i, of the ring resonator system, as well described by Rabus (2002). The specified constant quantities of C_n, x_n and y_n are given in Table 1.

Table 1
Expression of optical field, $E_{(i)}$ and constant quantities of C_n, x_n and y_n

Optical Fields, E_i	Constant Quantities, C_n, x_n, y_n
$E_1 = E_{in}j\sqrt{\kappa_1} + E_4\sqrt{1-\kappa_1}e^{-\frac{\alpha L_1}{2}-jkn\frac{L_1}{2}}$	$C_1 = 1 - \sqrt{1-\kappa_2}\sqrt{1-\kappa_3}e^{-\frac{\alpha}{2}L_2-jknL_2}$
$E_2 = E_1j\sqrt{\kappa_2}e^{-\frac{\alpha L_1}{2}-jkn\frac{L_1}{2}} + E_3\sqrt{1-\kappa_2}e^{-\frac{\alpha L_2}{2}-jkn\frac{L_2}{2}}$	$C_2 = \sqrt{1-\kappa_2} - \sqrt{1-\kappa_3}e^{-\frac{\alpha}{2}L_2-jknL_2}$
$E_3 = E_2\sqrt{1-\kappa_3}e^{-\frac{\alpha L_2}{2}-jkn\frac{L_2}{2}}$	$x_n = \lvert C_n \rvert$
$E_4 = E_3j\sqrt{\kappa_2}e^{-\frac{\alpha L_2}{2}-jkn\frac{L_2}{2}} + E_1\sqrt{1-\kappa_2}e^{-\frac{\alpha L_1}{2}-jkn\frac{L_1}{2}}$	$y_n = \lvert C_n \rvert^2$

Equations [2] and [3] indicate that the ring resonator system in this particular case is similar to a Fabry-Perot cavity, which consists of two parallel highly reflecting mirrors that function as a fully reflecting mirror, κ, and a field reflectivity, (1-κ), where κ is the coupling coefficient. The details of the specific variables are as follows: the linear absorption coefficient is α, a roundtrip loss coefficient is $e^{-\alpha Li/2}$, the linear phase shift is $\phi_i = kn_i L_i$, the waveguide length of the ring radius R_i is $L_i = 2\pi R_i$ and n_i is the refractive index of the waveguide material of R_i.

In operation, the input optical field with a specified wavelength, E_{in}, i.e. a Gaussian pulse from a monochromatic light source, is launched into the proposed ring resonator system, of which the appropriate parameters, such as the radius of the rings (waveguide length) and the coupling coefficients, must be assigned so that the transmitted output signals can be controlled.

Figure 1 shows, in principle, while propagating within the ring system, the input light pulse is divided and sliced as the discrete signal by the first ring, R_1, and then spread to the another ring, R_2, with the direction as shown in the figure. Finally, the required signals are obtained separately via the throughput port and the drop port of the ring system.

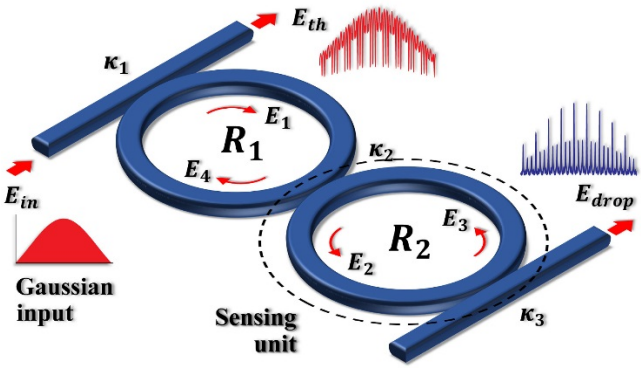


Figure 1. Schematic diagram of the proposed ring resonator system

SIMULATION RESULTS

Normally, the optical Vernier effects, which have a potential for use in high-resolution distance measurement, can be generated by using an optically resonant structure such as a micro-ring resonator (Boeck et al., 2010). A Gaussian input with a specified wavelength ranging from 1.545–1.555 μm (C-band), 10 nm bandwidth and a peak power of 0.2 W, can be propagated through the proposed system. Thereafter, the resonant outputs each at the throughput port and the drop port are obtained with the optical Vernier effects as shown in Figure 2. In this work, the ring resonator system is proposed to be useful as a force in a micro-scale sensing devices. Micro-sensing can be performed by measuring the spacing shift among the setting and sensing peak spectrum signals that are obtained from the drop port of the system. In order to associate the system with a practical device (Takara et al., 2002; Xu et al., 2004), the system parameters are specified as follows: the ring radius $R_1 = 300$ μm and the sensing unit ring $R_2 = 60$ μm. The refractive index of the waveguide material is fixed at n = 3.47 (Si–Crystalline silicon). The coupling coefficients, κ, of the system ranges from 0.10-0.20. The waveguide (ring system) loss is considered to be α= 0.5 dB/mm. For simplification, the optical field relation does not take into account the coupling losses (Rabus et al., 2002). In operation, the signal wavelength shift resolution in a nano-scale region within the interferometer ring R_2 can be achieved by changing its optical path length. For instance, the force exertion can be performed as a diffused force acting on the thin film, which is coated on the sensing unit; hence, the optical path length of the sensing unit within the thin film is also changed concurrently during the force exertion. The change in optical path length is related to the change in exerted force on the sensing unit, which is according to the stress/strain of the waveguide material. Finally, the setting and sensing signal are observed and compared to evaluate an amount of the exerted force on the sensing unit.

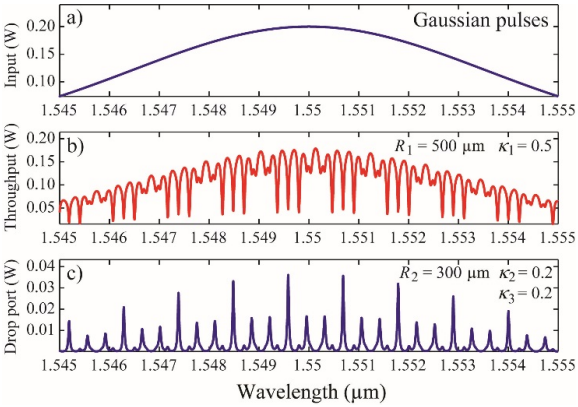


Figure 2. Simulation results with the obtained optical Vernier effects for a Gaussian pulse input of 10 nm bandwidth as a function of wavelength from 1.545 to 1.555 μm

The concept of stress and strain on the sensing unit by the elastic modulus of the waveguide material, which causes the difference in peak spectrum of the output signals, and the relationship between the exerted force and the change in waveguide length, can be expressed by equation [4] and [5], respectively.

$$Y_0 = \frac{Stress}{Strain} = \left(\frac{F}{A_0} \times \frac{L_0}{\Delta L}\right) \quad [4]$$

$$F = \left(\frac{Y_0 A_0}{L_0}\right) \times \Delta L \quad [5]$$

where, F is the exerted force in a small-scale ranging from 10 μN to 50 μN, Y_0 is Young's modulus of the waveguide material, which is approximately 160 kN/mm^2 for the Si–Crystalline silicon (Hopcroft et al., 2010). The initial cross-sectional area of the sensing unit waveguide is represented by A_0, which is specified as 0.054 μm^2, while L_0 is the initial waveguide length, which equals to $2\pi R_2$ μm, and ΔL is the change in waveguide length.

Figure 3 shows the simulation results of the proposed system as the setting spectrum signal (no force exertion) compared with each sensing spectrum signal, with the exerted force given as 10 μN, 20 μN, 30 μN, 40 μN and 50 μN, resulting in the changes in a spacing shift among the sensing and setting peak signal; this can be seen for instance in Channel 1 as approximately 35 pm, 72 pm, 110 pm, 149 pm and 189 pm, respectively.

Figure 4 shows the comparison between the setting and sensing spectrum signal in the other channels, Channel 2, Channel 3 and Channel 4, with the same mentioned conditions. The changes in spacing shift among the setting and sensing signal in each condition and channel are given by the figure.

By the well-known linear regression, with more than 99% of R^2, the linearity relations between the exerted forces and the free spectral range shifts of each considered channel, such as ch.1 to ch.4, have been illustrated as in Figure 5.

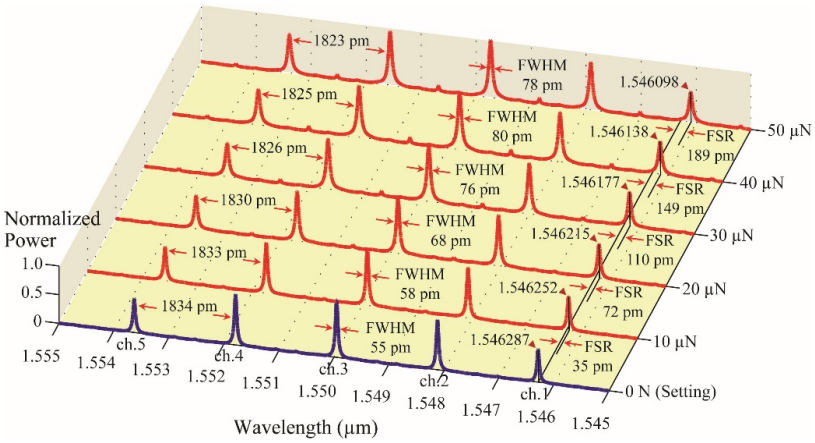


Figure 3. Simulation results of the comparing spectrum between the setting and sensing spectrum that can be obtained by the ring resonator system while exerting a force of 10 μN, 20 μN, 30 μN, 40 μN and 50 μN on the sensing unit

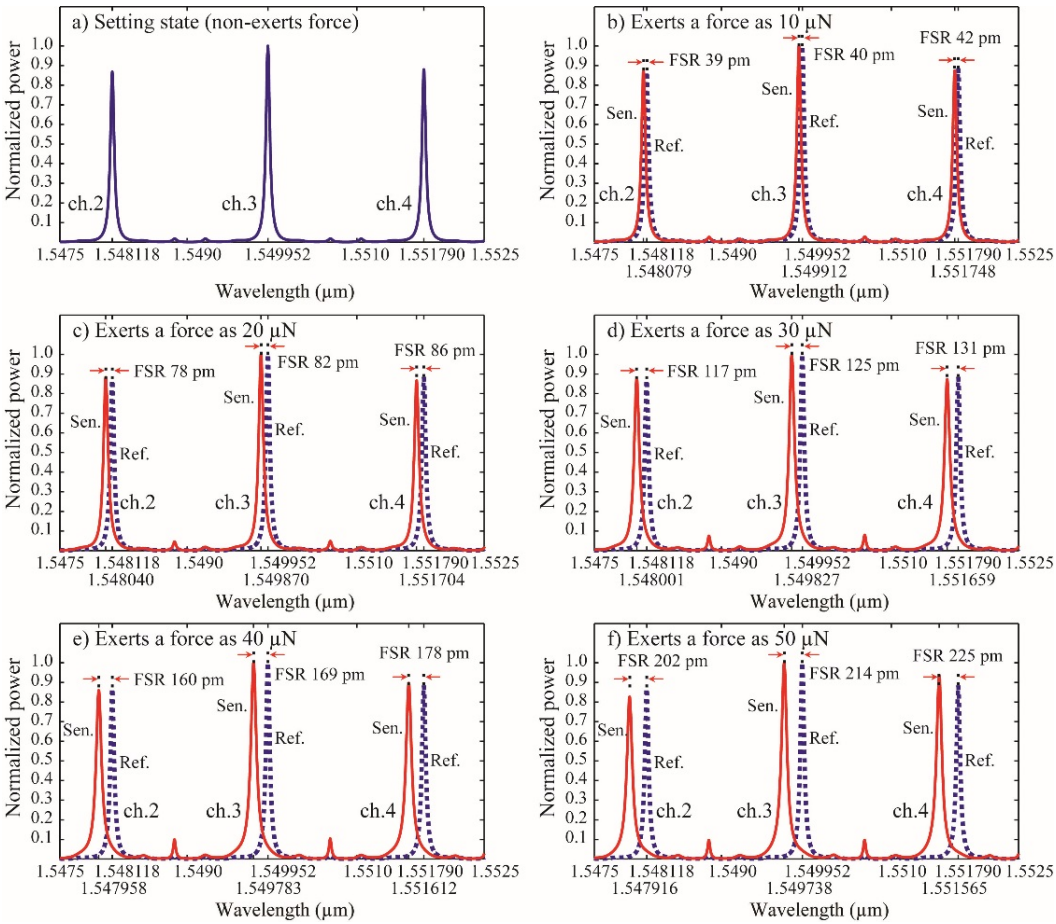


Figure 4. Simulation results of the comparing spectrum between the setting and sensing spectrum in channel 2, 3 and 4, a) for the setting state, b), c), d), e) and f), for a force exertion of 10 μN, 20 μN, 30 μN, 40 μN and 50 μN, respectively

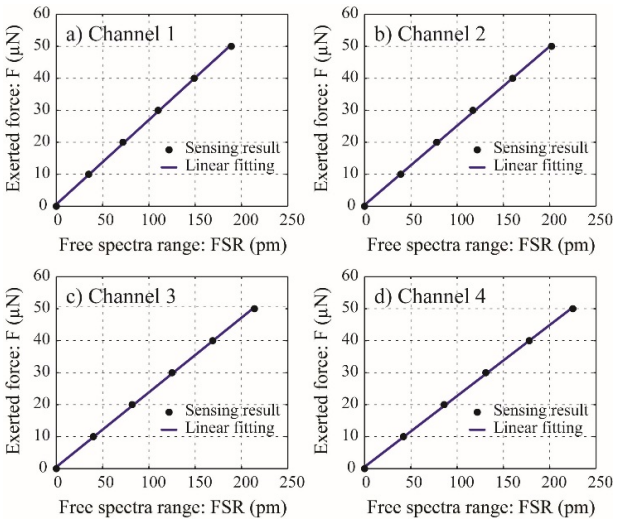


Figure 5. Linearity relation of the exerted forces as a function of free spectra range shifts in each considered channel

DISCUSSION

Although the proposed ring resonator system has the potential to be useful as a micro-force sensing device, there are significant uncertainties to be considered for its proper function. These uncertainties are listed below.

S1. Errors in the fabrication process: The proposed ring resonator system is designed to function within specific parameters, such as the ring radius, the coupling coefficients and the refractive index of the waveguide material. In practice, the resonant response of the ring resonator system needs to be customised for a particular device since the fabrication process may have some unavoidable errors in the required system parameters that would result in errors in the resonant outputs of the actual device. The most typical and straightforward approach for tuning the resonant response of a micro-ring resonator is to change the refractive index of the waveguide material, using perhaps the thermo-optic or electro-optic effect method, which applies the heat or electrical field to the waveguide (Prasad et al., 2016; Kim, 2015). The carrier injection method aims to change loss parameters and a refractive index of the material (Bulutay et al., 2010). The chemical method uses the chemical potential of graphene to tune the refractive index of the waveguides (Perez et al., 2016; Xing & Jian, 2017). Therefore, in order to function properly, the fabrication work must take into account the waveguide tuning techniques, which were not considered in this investigation.

S2. Elasticity of doped silicon: Generally, silicon wafers are not made of pure silicon, but usually have a certain amount of chemical impurities that are added to control the wafers' properties; this is called doping. The volume of the doping atoms on crystalline silicon is normally negligible but it may have an effect on the elastic behaviour of the waveguide.

However, the changes typically decrease by 1%-3% in heavy doping (Hopcroft et al., 2010); this percentage of doping can be neglected.

S3. Elasticity of thin film: As the force exertion is performed by a diffused force acting on the thin film, which is coated on the sensing unit, in practice, the change in the optical path length of the waveguide may also be influenced by the elasticity of the thin film. It is difficult to identify by how much the sensing results are exactly affected by the elasticity of the thin film since it can be made up of a variety of materials. However, the thin film could be made up of Low-Density Polyethylene (LDPE), which has a variety of applications in industry, of which Young's modulus is 150 N/mm^2 (Molaei, 2016), corresponding to 0.094% with respect to Young's modulus of the waveguide material (Si-Crystalline silicon), which can be neglected.

CONCLUSION

In this paper, an optical ring resonator system of which the resonant outputs were observed and compared to be useful as a force in a micro-scale sensing device was presented. The sensing application can be accomplished by evaluating the changes in the filtering characteristics together with the concept of stress/strain and the elastic modulus of the waveguide material. In operation, the force exertion can be performed as a diffused force acting within the thin film that is coated on the sensing unit, of which the optical path length can also be changed proportionally to the amount of the exerted force. From the simulation results, the exerted force ranged from 10 μN to 50 μN upon the sensing unit; this can be evaluated as the spacing shift between the setting and sensing peak signal of 35 pm to 225 pm. The sensing relation was obtained as good linearity, with more than 99% of R^2.

ACKNOWLEDGEMENT

We would like to express our thanks to Sakon Nakhon Rajabhat University International Conference 2015 (SNRU-IC 2015) for help in preparing this paper.

REFERENCES

Abushagur, A. A. G., Arsad, N., Reaz, M. I., & Bakar, A. A. A. (2014). Advances in bio-tactile sensors for minimally invasive surgery using the fibre Bragg grating force sensor technique: A survey. *Sensors, 14*(4), 6633-6665.

Alcheikh, N., Coutier, C., Giroud, S., Poulain, C., & Rey, P. (2013). Characterization and modeling of a piezoresistive three-axial force micro sensor. *Sensors and Actuators A: Physical, 201*, 188-192.

Baki, P., Szekely, G., & Kosa, G. (2012). Miniature tri-axial force sensor for feedback in minimally invasive surgery. In E. Guglielmelli (Ed.), *Proceedings of the 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics* (pp. 805-810). Rome, Italy: Universita Campus Bio-Medico di Roma.

Bao, J., Xiao, J., Fan, L., Li, X., Hai, Y., Zhang, T., & Yang, C. (2014). All-optical NOR and NAND gates based on photonic crystal ring resonator. *Optics Communications, 329*, 109-112.

Beyeler, F., Neild, A., Oberti, S., Bell, D. J., Sun, Y., Dual, J., & Nelson, B. J. (2007). Monolithically fabricated micro-gripper with integrated force sensor for manipulating microobjects and biological cells aligned in an ultrasonic field. *IEEE/ASME Journal of Microelectromechanical Systems, 16*(1), 7-15.

Boeck, R., Jaeger, N. A. F., Rouger, N., & Chrostowski, L. (2010). Series-coupled silicon racetrack resonators and the Vernier effect: Theory and measurement. *Optics Express, 18*(24), 25151-25157.

Bulutay, C., Turgut, C. M., & Zakhleniuk, N. A. (2010). Carrier-induced refractive index change and optical absorption in wurtzite InN and GaN: Full-band approach. *Physical Review B, 81*(15), 155206.

Cappelleri, D. J., Piazza, G., & Kumar, V. (2011). A two dimensional vision-based force sensor for microrobotic applications. *Sensors and Actuators A: Physical, 171*(2), 340-351.

Caseiro, D., Santos, S., Ferreira, C., & Neves, C. (2014). High sensitivity micro-machined piezoresistive strain sensor. *Procedia Engineering, 87*, 1362-1365.

Chen, Y., Yan, S. C., Zheng, X., Xu, F., & Lu, Y. Q. (2014). A miniature reflective micro-force sensor based on a microfiber coupler. *Optics Express, 22*(3), 2443-2450.

Cho, C., & Ryuh, Y. (2016). Fabrication of flexible tactile force sensor using conductive ink and silicon elastomer. *Sensors and Actuators A: Physical, 237*, 72-80.

Dumeige, Y., Arnaud, C., & Fe´ron, P. (2005). Combining FDTD with coupled mode theories for bistability in micro-ring resonators. *Optics Communications, 250*(4-6), 376-383.

Enikov, E., & Nelson, B. (2000). Three-dimensional microfabrication for a multi-degree-of-freedom capacitive force sensor using fibre-chip coupling. *Journal of Micromechanics and Microengineering, 10*(4), 492-497.

Hopcroft, M. A., Nix, W. D., & Kenny, T. W. (2010). What is the Young's modulus of silicon? *Journal of Microelectro-Mechanical Systems, 19*(2), 229-238.

Jin, W. L., & Mote, C. D. (1998). A six-component silicon micro force sensor. *Sensors and Actuators A: Physical, 65*(2), 109-115.

Kim, J. T. (2015). Silicon optical modulators based on tunable plasmonic directional couplers. *IEEE Journal of Selected Topics in Quantum Electronics, 21*(4), 3300108.

Li, X., Su, D., & Zhang, Z. (2009). A novel technique of microforce sensing and loading. *Sensors and Actuators A: Physical, 153*(1), 13-23.

Liua, C. P., Leea, K. J., Koa, C. H., & Dong, B. Z. (2007). Characteristics of optical multiple channelled filters made of aperiodically patterned phase elements. *Optics and Laser Technology, 39*(2), 415-420.

Ma, Z., & Ogusu, K. (2011). Channel drop filters using photonic crystal Fabry-Perot resonators. *Optics Communications, 284*(5), 1192-1196.

Marchi, G., Stephan, V., Dutz, F. J., Hopf, B., Polz, L., Huber, H. P., & Roths, J. (2016). Femtosecond laser machined micro-structured fiber Bragg grating for simultaneous temperature and force measurements. *Journal of Lightwave Technology, 34*(19), 4557-4563.

Molaei, S. (2016). The measurement of Young's modulus of thin films using secondary laser speckle patterns. *Measurement, 92*, 28-33.

Peirs, J., Clijnen, J., Reynaerts, D., Van Brussel, H., Herijgers, P., Corteville, B., & Boone, S. (2004). A micro optical force sensor for force feedback during minimally invasive robotic surgery. *Sensors and Actuators A: Physical, 115*(2-3), 447-455.

Perez, D., Domenech, D., Munoz, P., & Capmany, J. (2016). Electro-Refraction modulation predictions for silicon graphene waveguides in the 1540-1560 nm region. *IEEE Photonics Journal, 8*(5), 4501613.

Prasad, P. R., Selvaraja, S. K., & Varma, M. M. (2016). Full-Range detection in cascaded microring sensors using thermooptical tuning. *Journal of Lightwave Technology, 34*(22), 5157-5163.

Rabus, D. G. (2002). *Realization of optical filters using ring resonators with integrated semiconductor optical amplifiers in GaInAsP/InP*. (Doctoral thesis). Retrieved from Deposit Once Research Rata and Publications database. (ISBN 3-89959-022-8).

Rabus, D. G., Hamacher, M., Troppenz, U., & Heidrich, H. (2002). Optical filters based on ring resonators with integrated semiconductor optical amplifiers in GaInAsP-InP. *IEEE Journal of Selected Topics in Quantum Electronics, 8*(6), 1405-1411.

Rakshit, J. K., & Roy, J. N. (2014). Micro-ring resonator based all-optical reconfigurable logic operations. *Optics Communications, 321*, 38-46.

Takara, H., Ohara, T., Yamamoto, T., Masuda, H., Abe, M., Takahashi, H., & Morioka, T. (2002). Field demonstration of over 1000-channel DWDM transmission with supercontinuum multi-carrier source. *Electronics Letters, 41*(5), 270-271.

Tan, U. X., Bo, Y., Gullapalli, R., & Desai, J. P. (2011). Triaxial MRI-compatible fiber-optic force sensor. *IEEE Transactions on Robotics, 27*(1), 65-74.

Tokuno, T., Tada, M., & Umeda, K. (2008). High-precision MRI-compatible force sensor with parallel plate structure. In *Proceedings of the 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics* (pp. 33-38). Scottsdale, AZ, USA.

Xing, R., & Jian, S. (2017). Numerical analysis on tunable multilayer nanoring waveguide. *IEEE Photonics Technology Letters, 29*(12), 967-970.

Xu, Q., Almeida, V. R., Panepucci, R. R., & Lipson, M. (2004). Experimental demonstration of guiding and confining light in nanometer-size low-refractive-index material. *Optics Letters, 29*(14), 1626-1628.

Yan, X., Ma, C. S., Zheng, C. T., Wang, X. Y., & Zhang, D. M. (2010). Analysis of polymer electro-optic microring resonator switches. *Optics and Laser Technology, 42*(3), 526-530.

Yang, J., Zhou, Q., Zhao, F., Jiang, X., Howley, B., Wang, M., & Chen, R. T. (2003). Characteristics of optical bandpass filters employing series-cascaded double-ring resonators. *Optics Communications, 228*(1-3), 91v98.

Yupapin, P. P., & Suwancharoen, W. (2007). Chaotic signal generation and cancellation using a micro ring resonator incorporating an optical add/drop multiplexer. *Optics Communications, 280*(2), 343v350.

Zadeh, E. G., Charca, G. A., Gholamzadeh, B., Ghasemi, S., & Magierowski, S. (2017). Towards scalable capacitive cantilever arrays for emerging biomedical applications. *Sensors and Actuators A: Physical, 260*, 90-98.

Pertanika J. Sci. & Technol. 26 (1): 95 - 120 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Distributions, Composition Patterns, Sources and Potential Toxicity of Polycyclic Aromatic Hydrocarbons (PAHs) Pollution in Surface Sediments from the Kim Kim River and Segget River, Peninsula Malaysia

Mehrzad Keshavarzifard[1,2]*, Mohamad Pauzi Zakaria[3], Shahin Keshavarzifard[4] and Reza Sharifi[2]

[1]*Environmental Forensics Research Center, Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[2]*Department of Earth Sciences, Faculty of Sciences, Shiraz University, Shiraz 71454, Iran*
[3]*Institute of Ocean and Earth Sciences, University of Malaya, 50603 UM, Kuala Lumpur, Malaysia*
[4]*Department of Environmental Sciences, Faculty of Sciences, Zanjan University, Zanjan, Iran*

ABSTRACT

This study investigated distributions, composition patterns, sources and potential toxicity of polycyclic aromatic hydrocarbon (PAHs) pollution in surface sediments from the Kim Kim River and Segget River, Peninsular Malaysia. The samples were extracted using Soxhlet extraction, purified using two-step silica gel column chromatography and then analysed by gas chromatography mass spectrometry (GC-MS). The total PAH concentrations ranged from 95.17 to 361.24 ng g^{-1} dry weight (dw) and 330.09 to 552.76 ng g^{-1} dw in surface sediments from the Kim Kim and Segget Rivers, respectively. Source type identification using PAH molecular indices and hierarchical cluster analysis (HCA) indicated that PAHs were mostly of pyrogenic origin, while in some stations petrogenic sources had a significant portion. A PAH toxicity assessment using sediment quality guidelines (SQGs), mean effect range-median quotient (M-ERM-Q), benzo[*a*]pyrene (BaP) equivalent concentration and BaP toxicity equivalent quotient (TEQ^{carc}) indicated low probability of toxicity for both the Kim Kim and Segget Rivers. Moreover, the human health risk assessment applying Cancer $Risk_{ingestion}$ and Cancer $Risk_{dermal}$ indicated that probabilistic health risk to humans via ingestion and dermal pathways from sediments of the Kim Kim and Segget Rivers can be categorised as low-to-moderate risk.

Keywords: Kim Kim River, Segget River, Malaysia, pollution sources, Polycyclic Aromatic Hydrocarbons (PAHs), sediment, ecological risk assessment, human health risk assessment

ARTICLE INFO

Article history:
Received: 20 November 2016
Accepted: 09 November 2017

E-mail addresses:
mkf1759@yahoo.com (Mehrzad Keshavarzifard),
mpauzi57@um.edu.my (Mohamad Pauzi Zakaria),
shahinkf@yahoo.com (Shahin Keshavarzifard),
rezasharifi@shirazu.ac.ir (Reza Sharifi)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Johor State is one of the most developed states of Malaysia (Alkhadher et al., 2016), located in the southern part of Peninsular Malaysia. It is connected to Singapore by the Straits of Johor. The population size and density of Johor State have increased continuously in the last two decades. In 2012, the population of Johor State rose from 2.76 million in 2000 to 3.46 million (increase over 25%) (DOSM, 2014a, 2014b). Increase in the population size and density, quick development (industrialisation, urbanisation, agriculture and aquaculture) and marine transportation in Johor State and along the coastal areas of the Johor Straits can result in comprehensive physical and chemical changes in the aquatic environment. Industrial development in the region has also resulted in increase in the usage of petroleum and its derivatives as sources of energy. As a consequence, petroleum hydrocarbon contaminants released into this environment became significant. The Kim Kim River and Segget River run through Johor State. Pollutants, including polycyclic aromatic hydrocarbons (PAHs), originating from anthropogenic activities such as petrol combustion, industrialisation and multiple diffuse sources in Johor State and the neiguhboring area are released into the rivers via dry and wet deposit, urban runoff and direct input. It was reported that the Segget River is grossly polluted by domestic sewage (Lye & Eng, 1988; Tung et al., 2013).

Polycyclic aromatic hydrocarbons (PAHs) are categorised as one of the most important classes of contaminants. They have two or more benzene rings with variable physicochemical properties (Neff, 1979). PAHs are widely distributed in diverse ecosystems, including freshwater and marine, and they have drawn significant attention due to their persistence and toxicity and mutagenic, teratogenic and carcinogenic effects (Cao et al., 2009).

Many studies have reported that urban and riverine runoff (Alkhadher et al., 2015; Lipiatou et al., 1997; Magam et al., 2015; Motelay-Massei et al., 2006) and atmosphere (Liu et al., 2007b; Tsai et al., 2002) are the two most important transportation pathways of contaminants of the aquatic environment. Hoffman et al. (1984) demonstrated that urban runoff comprised 71% HMW PAHs and 36% total PAHs input into Narragansett Bay. This is consistent with Oros et al. (2007), who suggested that 51% of total PAH in the San Francisco Bay was transported by storm water runoff. The sources of PAHs in urban runoff are wear of asphalt surfaces, abrasion of vehicle tires, car exhaust and used crankcase oils (Takada et al., 1991). Incomplete combustion derived PAHs emitted into the atmosphere from different sources enter the water column directly via dry and wet deposition or are indirectly deposited to the terrestrial environment and urban runoff; they include street dust and storm water discharge, which are finally transported to the aquatic ecosystem.

PAHs can originate from natural processes, but the major source of PAH input is generally from anthropogenic activities (Keshavarzifard et al., 2017a; Keshavarzifard and Zakaria, 2015; Keshavarzifard et al., 2015; Masood et al., 2016). As for pollution due to anthropogenic activities, the origins are classified as pyrogenic PAHs (from high temperature incomplete combustion of organic matter) and petrogenic PAHs (from fresh petroleum input). Nevertheless, pyrogenic PAHs i.e. incomplete combustion derived PAHs can be adsorbed or absorbed to airborne organic particles and moved long distances by wind and other atmospheric phenomena.

Due to lipophilic characteristic, PAHs are mostly partitioned with organic matter in the environment and undergo sedimentation (Keshavarzifard et al., 2017b); however, occasionally these contaminants are not partitioned with organic matter, which indicates existence of soot particles in sediments (Gustafsson & Gschwend, 1997). PAH compounds are associated with soot particles and transported in long-range atmospheric pathway (Pereira et al. 1999). Petrogenic PAHs (associated with particles containing organic carbon) and pyrogenic PAHs (partitioned with airborne particles) reach the surface layers of river and marine water columns and pass through the water columns and then settle down in the bottom sediments. It was reported that the sorption of hydrocarbons in sediments is influenced largely by the content of organic carbons, particle size distribution and the octanol-water partition coefficients of the pollutant (Gustafsson et al., 1996). This statement is consistent with those reported by Karickhoff and Brown (1978) and Karickhoff et al. (1979). It was suggested that petroleum hydrocarbons deposited in sediments are more likely to partition into porewater than pyrogenic PAHs associated with soot particles (McGroddy & Farrington, 1995). Usually HMW PAHs i.e. 4-, 5- and 6-ring PAHs, which are heavier than LMW PAHs i.e. 2- and 3-ring PAHs, are placed in the centre of the particles and the LMW PAHs are placed around the particles and become heavier and then undergo sedimentation. Pyrogenic and petrogenic PAHs release into the river and marine ecosystems and then undergo dispersion, evaporation, settlement in the sediments, weathering, chemical changes, photooxidation and microbial degradation (bacteria and fungus) in short- and long-term periods (Aldarondo-Torres et al., 2010; Irwin, 1997; Mai et al., 2002; Neff, 1979; Ünlü et al., 2010).

The objectives of this study were: (a) to assess the concentration and distribution of 16 USEPA PAHs in surface sediment from the Kim Kim and Segget Rivers; (b) to identify the PAH sources using molecular indices; (c) to evaluate the toxicity potency and human health risk of PAHs. The results of this research in conjunction with long-term study of PAHs in the Malacca and Johor Straits enrich our knowledge of the distribution and pathways of PAHs in this region and introduce useful information about PAH pollution in order to improve regulations and consequently the environmental quality of the study area.

METHODOLOGY

Sampling

In May 2013, six surfacial sediment samples (0-4 cm) were collected from the Kim Kim and Segget Rivers using an Ekman Dredge sampler. The location of the sampling stations are shown in Figure 1 and Table 1. The top 4 cm of surface layer was carefully removed to identify modern input. After separating the organisms and shell debris from sediments at each sampling station, the samples were immediately placed in previously cleaned aluminum containers and Ziploc bags, and then stored in athe cooler box with dry ice. The samples were subsequently freeze-dried and stored at −20°C before analysis.

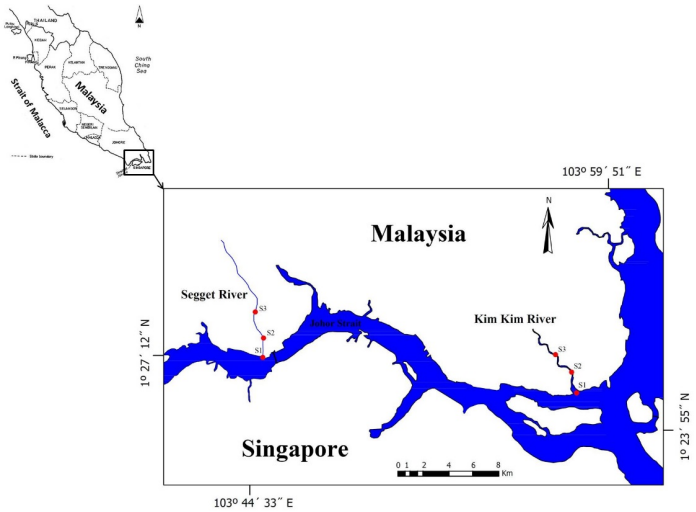


Figure 1. Location of sampling areas

Table 1
Details of water quality data of sampling stations

Sampling Location	Latitude (°N)	Longitude (°E)	DO (mg/l)	Turbidity (NTU)	Temperature (°C)	pH
Kim Kim River S1	01° 25'	103° 58'	5.9	15	29.8	7.55
Kim Kim River S2	01° 20"	103° 58'	4.7	7.32	29.9	7.47
Kim Kim River S3	01° 27'	103° 57'	6.54	8.6	20.3	7.27
Segget River S1	01° 27'	103° 45'	4.5	42.3	29	7.27
Segget River S2	01° 27'	103° 45'	2	54.8	28	7.63
Segget River S3	01° 26'	103° 45'	3.4	48	28.9	7.76

Extraction and Purification

Each sediment sample was extracted, purified and fractionated according to the method explained in Zakaria et al. (2002). In summary, hydrocarbons were extracted by Soxhlet Apparatus using 200 mL dichloromethane (DCM) for 10 h. A volume of 50 µL of 10 µg g^{-1} surrogate internal standards (SIS), namely naphthalene-d_8, anthracene-d_{10}, chrysene-d_{12}, and perylene-d_{12}, were added to the sample aliquot before extraction. The SIS compounds wereare used for recovery verification. The extraction was purified and fractionated through first- and second-step silica gel column chromatography. The extract was rotary evaporated and sulfur interference was removed using activated copper. The extract was allowed to pass through a silica gel column containng 5% H_2O-deactivated silica gel to purify non-polar compounds such as hydrocarbons. The hydrocarbon compounds (alkanes, hopanes, linear alkyl benzenes (LABs), PAHs and many more) were fractionated using a fully activated silica gel column (0.47 cm i.d × 18 cm height). A volume of 16 mL rinsing solution (hexane-DCM 3:1 v/v) was added. The

amunt of PAH fraction was further reduced by rotary evaporator and a gentle stream of pure N_2 gas. The final fraction was then adjusted to 200 µL with isooctane containing a 10-ppm internal injection standard (IIS), p-terphenyl-d_1Analysis of PAHs was performed by GC-MS Shimadzu QP5050A model and a BPX-5MS fused-silica capillary column (30 m by 0.25 mm i.d., 0.25-µm film thickness). The oven temperature was ramped from 60 to 260°C at 6° min^{-1}, and then held for 15 min. An aliquot 1 µL of purified sample was injected into the GC-MS injector. The oven temperature started at 70°C for 2 min with a ramp of 30°C.min-1 until 150°C. The temperature was further increased to 310°C with an increasing rate of 4°C.min^{-1} for 15 min. The 16 PAHs detected in sequence were as follows: naphthalene (Nap), acenaphthylene (Acy), acenaphthene (Ace), fluorene (Fl), phenanthrene (Phe), anthracene (Ant), fluoranthene (Fluo), pyrene (Py), benzo[*a*]anthracene (BaA), chrysene (Chr), benzo[*b*]fluoranthene (BbF), benzo[*k*]fluoranthene (BkF), benzo[*a*]pyrene (BaP), indeno[*1,2,3-cd*]pyrene (InP), dibenzo[*a, h*]anthracene (DBA), and benzo[*ghi*]perylene (BgP).

Total Organic Carbon (TOC) Analysis

TOC content of sediment sample was identified using a LECO CR-412 carbon analyser. Carbonates were eliminated by 2 mL of 1 M HCL at 100°C overnight before TOC identification (Nelson & Sommers, 1996).

Quality Control and Assurance

Quality control of sediment samples was applied to ensure the accuracy and precision of the analytical procedure. Surrogate internal standards were used to examine the recovery percentage of each sample. The recovery rates of 16 individual EPA PAHs ranged from 81% to 110%. Quantification of PAH compounds was processed based on external calibration curves using standard mixtures of PAH. Determination of target compounds was achieved based on matching target compound ionisation and retention times with the standard mixture of PAHs. Internal injection standard (p-terphenyl-d_{14}) was used for quantitation and identifying injection error. Procedural blanks were subjected for every batch of four samples; no contamination was detected in the analytical system and glassware.

RESULTS AND DISCUSSION

Distribution of PAHs in Surface Sediments

Analytical results of the 16 EPA priority PAHs measured at different sampling stations are listed in Table 2. The total PAH concentrations varied from 95.17 to 361.24 ng g^{-1} dry weight (dw) with a mean value of 218.1 ng g^{-1} dw in sediments from the Kim Kim River and 330.09 to 552.76 ng g^{-1} dw with a mean value of 437.5 ng g^{-1} dw in sediments from the Segget River. Therein, the highest concentrations were determined at Station 1 (estuary) of both the Kim Kim and Segget Rivers and and lowest concentrations recorded at Station 3 (upstream) of the Kim Kim and Segget Rivers. PAHs normally partition to lipophilic particles in the water column and as transported downstream, more particles clump together to flocculate. As the particles reach the salinity regime in estuaries, they settle to the bottom sediment, which usually

increases the concentration of PAHs in sediments closer to an estuary. Moreover, tidal current might also influence the PAH distribution pattern. Baumard et al. (1998) categorised the PAHs levels in the aquatic environment from 0-100 (low), 100-1000 (moderate), 1000-5000 (high) and >5000 ng g^{-1} (very high). According to this categorisation, the PAH levels can be classified as low-to-moderate in the Kim Kim River and moderate in the Segget River.

Table 2
PAHs in surface sediments from the Kim Kim and Segget rivers

	Kim Kim River			Segget River			[j]ERL	[k]ERM	[l]TEL	[m]PEL	[n]TEF
	Station 1	Station 2	Station 3	Station 1	Station 2	Station 3					
Nap	18.38	6.17	6.63	33.68	55.41	34.16	160	2100	34.6	391	0.001
Acy	19.64	7	9.38	6.33	51.1	4.89	44	640	5.87	128	0.001
Ace	39.72	2.74	4.23	5.72	47.77	2.4	16	500	6	88.9	0.001
Fl	18.56	10.45	8.81	7.49	75.55	11.26	19	540	21.2	144	0.001
Phe	16.22	8	19	56.92	10.41	51.98	240	1500	86.7	544	0.001
Ant	36.27	23.26	0.82	4.26	51.47	9.55	85	1100	46.9	245	0.001
Fluo	84.52	18	7.8	55.92	9.96	14.38	600	5100	113	1494	0.01
Py	90	17.31	12.43	52.1	16.19	10.72	665	2600	153	1398	0.001
BaA	14.54	30.8	2.73	47.3	3.22	0.61	261	1600	74.8	693	0.1
Chr	3.69	1.89	3.1	4.28	13.75	0.13	384	2800	108	846	0.01
BbF	3.59	15.63	7.7	44.51	2.15	0.4	-	-	-	-	0.1
BkF	2.1	7.19	1.27	10.99	25.28	46.1	-	-	-	-	0.1
BaP	4.55	24.23	2.31	84.39	28	53.88	430	1600	88.8	763	1
InP	2.25	11.46	1.28	54.91	5.9	15.5	-	-	-	-	0.1
DBA	2.11	6.9	0.8	21.48	4.68	53.64	63	260	6.2	135	1
BgP	5.1	6.75	6.88	62.48	28.86	20.49	-	-	-	-	0.01
[a]Total PAHs	361.24	197.78	95.17	552.76	429.7	330.09	4000	45000	1684	16770	-
[b]LMW PAHs	148.79	57.62	48.87	114.4	291.71	114.24	552	3160	-	-	-
[c]HMW PAHs	212.45	140.16	46.3	438.36	137.99	215.85	1700	9600	-	-	-
[d]LMW/ HMW	0.7	0.41	1.1	0.26	2.11	0.53	-	-	-	-	-
[e]Py/Fluo	1.06	0.96	1.59	0.93	1.63	0.75	-	-	-	-	-
fFluo/ (Fluo+Py)	0.48	0.51	0.39	0.52	0.38	0.57	-	-	-	-	-
[g]BaA/ (BaA+Chr)	0.79	0.94	0.47	0.92	0.19	0.82	-	-	-	-	-
[h]InP/ (Inp+BgP)	0.31	0.63	0.16	0.47	0.17	0.43	-	-	-	-	-
[i]TOC %	1.15	0.9	8.1	2.1	1.89	3.31	-	-	-	-	-

[a]Total PAHs = Sum of the concentrations of 16USEPA PAHs; [b]LMW PAHs = Sum of low molecular weight PAH concentrations from naphthalene to fluoranthene; [c]HMW PAHs = Sum of high molecular weight PAH concentrations from pyrene, benzo[a] anthracene to benzo[ghi]perylene
[d]LMW/HMW = The ratio of low molecular weight PAHs to high molecular weight PAHs; [e]Py/Fluo = The ratio of pyrene to fluoranthene; [f]Fluo/Fluo+Py = The ratio of fluoranthene to fluoranthene + pyrene; [g]BaA/(BaA+Chr) = The ratio of Benzo(a) anthracene to Benzo(a)anthracene+Chrysene; [h]InP/(Inp+BgP) = The ratio of indeno(1,2,3-cd)pyrene to indeno(1,2,3-cd) pyrene+benzo(ghi)perylene; [i]TOC = Total organic carbon
[j]ERL = Effects range-low value (Long et al. 1995); [k]ERM = Effects range-median value (Long et al. 1995); [l]TEL = Threshold effect levels (Macdonald et al. 1996); mPEL = Probable effect levels (Macdonald et al. 1996); [n]TEF = Toxic equivalency factor

The PAH concentrations in the Kim Kim River were compared with those of Zakaria et al. (2002), who found that PAH levels ranged from 37 to 41 ng g^{-1} (Table 3), confirming an increase in the levels of PAHs. However, more investigation is needed to support this statement. Among the PAH levels in the Malaysian aquatic ecosystems (Table 3), the concentrations of total PAH in sediments from the Kim Kim and Segget Rivers are close to those in sediments from the Selangor (Masood et al., 2014), the Merambong and Muar (Vaezzadeh et al., 2015) and the Perlis aquatic ecosystems (Keshavarzifard et al., 2015). However, they are lower than those of the Prai (Keshavarzifard et al., 2014b) and the Klang Rivers (Keshavarzifard et al., 2015). In comparison with other aquatic ecosystems in the world, sediment contaminations by PAHs in the Kim Kim and Segget Rivers are comparable with those of developing countries and much lower than those of developed and fast developing countries (Table 3).

Table 3
Comparison of PAH concentrations in surface sediments of different worldwide locations and this study

Location	Total PAHs (ng g^{-1})	[a]n	[b]Contamination Level	References
Boston Harbour, USA	7300-358000	16	Very high	Wang et al., 2001
Barcelona Harbour, Spain	1740-8420	14	High to very high	Baumard et al., 1998
Cape Cod, Massachusetts, USA	30000	38	Very high	Thorsen et al., 2004
Guanabara Bay, Brazil	91-8035	20	Low to very high	Meniconi et al., 2002
Patos Lagoon Estuary, Brazil	37-11779.9	23	Low to very high	Medeiros et al., 2005
Rybinsk Reservoir, Russia	178-5203	11	High to very high	Siddall et al., 1994
Lake iznik, Turkey	17-835	12	Low to moderate	Ünlü et al., 2010
Bohai Sea, China	0.14-10757	16	Low to very high	Zeng et al., 2013
Hyeongsan River, Korea	5.30-7680	16	Low to very high	Koh et al., 2004
Iranian coast of the Persian Gulf	93-4077	15	Low to high	Mirza et al., 2014
Gorgan Bay, Caspian Sea, Iran	107.9-516	16	Moderate	Araghi et al., 2014
Jakarta Bay, Indonesia	191-1252	18	Moderate to high	Rinawati et al., 2012
Rivers and Estuary of Thailand	6-8399	11	Low to very high	Boonyatumanond et al., 2006
Kim Kim River and Estuary, Malaysia	37-41		Low	Zakaria et al., 2002
Penang Estuary, Malaysia	924		Moderate	
Malacca River (1998), Malaysia	494	15	Moderate	
Klang Estuary, Malaysia	19-431		Low to moderate	
Johor Bahru Coast, Malaysia	239		Moderate	
Chini Lake, Malaysia	248-8098	16	Moderate to high	Bakhtiari et al., 2009
Prai River, Malaysia	1102-7938	16	High to very high	Keshavarzifard et al., 2014b

Table 3 *(continue)*

Malacca River, Malaysia	716-1210	16	Moderate to high	
Merambong River, Malaysia	38.6-122.8	16	Low to moderate	Vaezzadeh et al., 2015
Muar River, Malaysia	15.5-165.7		Low to moderate	
Selangor River, Malaysia	563-1037	16	Moderate to high	Masood et al., 2014
Perlis River, Malaysia	225.5-293.9		Moderate	Keshavarzifard et al., 2015
Kedah River, Malaysia	195.2-481.2		Moderate	
Merbok River, Malaysia	791.2-1995.4	16	Moderate to high	
Perak River, Malaysia	231.2-426.7		Moderate	
Klang River, Malaysia	3803.2-7442.7		High to very high	
Port Dickson, Malaysia	481.3-976.6		Moderate	Keshavarzifard & Zakaria, 2015
Johor Bahru Coast, Malaysia	650.5-1441.2	16	Moderate to high	Keshavarzifard et al., 2016
Malacca Strait, Malaysia	357.1-6257.1	16	Moderate to very high	Keshavarzifard et al., 2017d
Kim Kim River, Malaysia	95.17-361.24	16	Low to moderate	This study
Segget River, Malaysia	330.09-552.76		Moderate	This study

[a]Number of PAH compounds; [b]Scale of level of contamination: low: 0-100; moderate: 100-1000; high: 1000-5000; very high: >5000 (Baumard et al., 1998)

Determination of PAHs Sources Based on PAH Composition

Many investigations have confirmed that based on the characteristics of PAH composition and distribution pattern, the sources of anthropogenic PAHs can be distinguished (Keshavarzifard et al., 2017d; Liu et al., 2009; Vaezzadeh et al., 2014; Zakaria et al., 2002). Of anthropogenic PAHs, the incomplete combustion sources such as vehicle exhaust, domestic heating with coal and biomass burning contain predominantly HMW PAHs including 4-, 5- and 6-ring PAH compounds or pyrogenic PAHs, whereas the uncombusted sources (e.g. crude and fuel oil) contain predominantly LMW PAHs, including 2- and 3-ring PAH compounds or petrogenic PAHs (Bouloubassi et al., 2012; Irwin, 1997; Mai et al., 2002; Zakaria et al., 2002). In Stations 1 and 2 of the Kim Kim River, 4-ring PAHs had the highest portion of PAHs followed by 3-ring PAHs (Figure 2). In Station 3 of the Kim Kim River, 3-ring PAH had the highest portion, followed by 4-ring PAHs (Figure 2). In Stations 1 and 2 of the Segget River, PAHs were dominated by 4-and 3-ring PAHs, respectively. PAHs in Station 3 of the Segget River were dominated by 5-ring PAHs, followed by 6-ring ones (Figure 2). These results displayed that Stations 1 and 3 of the Segget River were significantly impacted by combustion-derived PAHs present in urban atmospheric particles, while Station 2 had received fresh petroleum

hydrocarbon. Based on the results, surface sediments from the Kim Kim and Segget Rivers showed that they contained both pyrogenic and petrogenic PAHs with pyrogenic PAHs dominating.

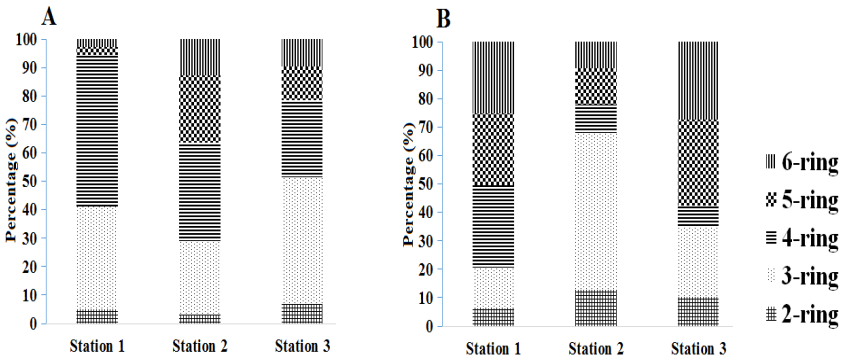


Figure 2. The composition pattern of PAHs by ring size in surface sediments from the Kim Kim River (A) and Segget River (B)

Determination of PAHs Sources Using Diagnostic Ratios

The diagnostic ratios of PAHs such as Py/Fluo, Fluo/(Fluo+Py), BaA/(BaA+Chr) and InP/(Inp+BgP) have been introduced as a useful tool to identify and assess the possible sources of PAHs (Keshavarzifard et al., 2014a; Qian et al., 2016; Yunker et al., 2002). For collected sediment samples of PAH isomer pair ratios, Py/Fluo was plotted against LMW/HMW and Fluo/(Fluo+Py) was plotted against BaA/(BaA+Chr) and InP/(Inp+BgP) to demonstrate the distribution of PAHs relative to their possible sources (Figure 3 and Figure 4). In general, a ratio of Py/Fluo > 1 indicates petroleum derivative PAHs, while a ratio < 1 suggests a dominance of combustion derivative PAHs (Sicre et al., 1987). Magi et al. (2002) reported that the ratio of LMW PAHs (2- and 3-ring PAHs) to HMW PAHs (4-, 5- and 6-ring PAHs) higher than 1 displays a pollution of petroleum contamination, while a ratio lower than 1 displays pyrogenic-origin PAHs. In this study, Py/Fluo ratios at Station 2 of the Kim Kim River and Stations 1 and 3 of the Segget River were lower than 1, indicating a dominance of pyrogenic PAHs. In the other stations, Py/Fluo ratios were lower than 1, suggesting a dominance of petroleum inputs (Table 2 and Figure 3). This is confirmed by the ratios of LMW/HMW, which were lower than 1 at all stations except Station 1 and 3 of the Kim Kim River and Station 2 of the Segget River (Table 2 and Figure 3).

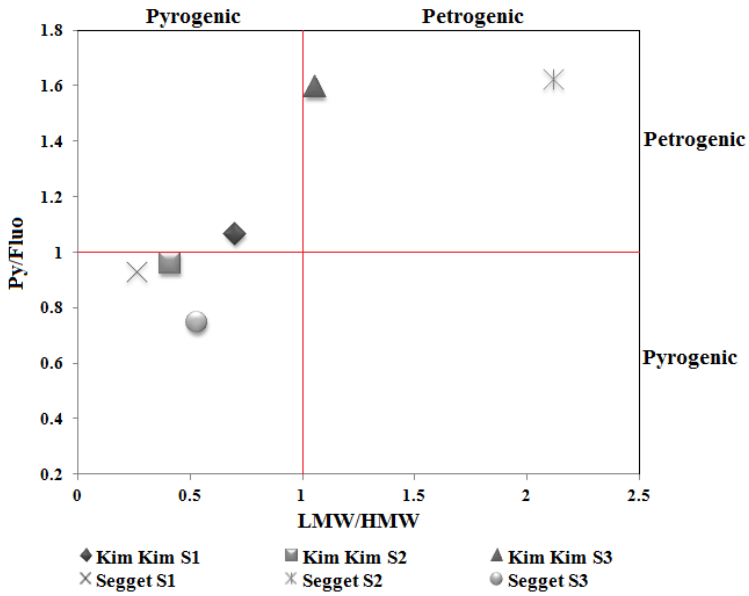


Figure 3. Plots of PAH isomer pair ratios for source identification: LMW/HMW versus Py/Fluo

Sediments with Fluo/(Fluo+Py) < 0.4 indicate petrogenic contamination, while Fluo/(Fluo+Py) > 0.5 suggests biomass and coal combustion origins and sediments with 0.4 < Fluo/(Fluo+Py) < 0.5 indicate petroleum combustion sources (Yunker et al., 2002). Those PAHs in sediments with the ratio of InP/(InP+BgP) < 0.2 were mainly from petrogenic sources, while those with InP/(InP+BgP) > 0.2 were typical of combustion-derivative PAHs (Budzinski et al., 1997; Yunker et al., 2002). Moreover, BaA/(BaA+Chr) ratios lower than 0.2, from 0.2 to 0.35 and higher than 0.35 are used as an indicator of petroleum, petroleum/combustion and combustion sources, respectively (Yunker et al., 2002). In this study, the ratios of Fluo/(Fluo+Py) at Station 3 of the Kim Kim River and Station 2 of the Segget River were lower than 0.4, indicating dominance of petroleum input, while at Station 1 of the Kim Kim River the ratio was 0.48, suggesting petroleum combustion processes make a more significant contribution than petroleum. Other stations had Fluo/(Fluo+Py) ratios more than 0.5, indicating biomass and coal combustion origins (Table 2 and Figure 4). This was confirmed by the results of InP/(InP+BgP) ratios (Table 2 and Figure 4). The ratio of BaA/(BaA+Chr) at Station 2 of the Segget River was 0.19, indicating petroleum input could be the main source. The BaA/(BaA+Chr) ratios in other stations were more than 0.35, implying combustion derivative PAHs played a major role.

Overall, sediments from the Kim Kim and Segget Rivers were affected by both petrogenic and pyrogenic sources, with a greater contribution from pyrogenic sources. This is consistent with Keshavarzifard et al. (2014b) and Keshavarzifard et al. (2015), who suggested dominance of combustion derivative PAHs in sediments from the major rivers of Peninsular Malaysia. Our findings are also consistent with those of industrialised countries (Notar et al., 2001;

Pereira et al., 1999; Stout et al., 2004), where pyrogenic PAHs were reported as mainly being of PAH origin in aquatic ecosystems. For example, recent widescale studies of background anthropogenic hydrocarbons in surficial sediments collected from the Arc River, the Berre Lagoon and an industrialised urban river (Huveaune) in France using PAH ratios to identify the sources of PAHs affirmed the pyrogenic-origin background of PAHs (Kanzari et al., 2014; Kanzari et al., 2012). On the other hand, Zakaria et al. (2002) reported petrogenic sources as a major source in Malaysian aquatic ecosystems incluing the Kim Kim River, which is in contrast with the results of this study. It can be suggested that the origin of PAHs in the Kim Kim River has shifted from petrogenic to pyrogenic sources; however, more studies are needed to confirm these findings.

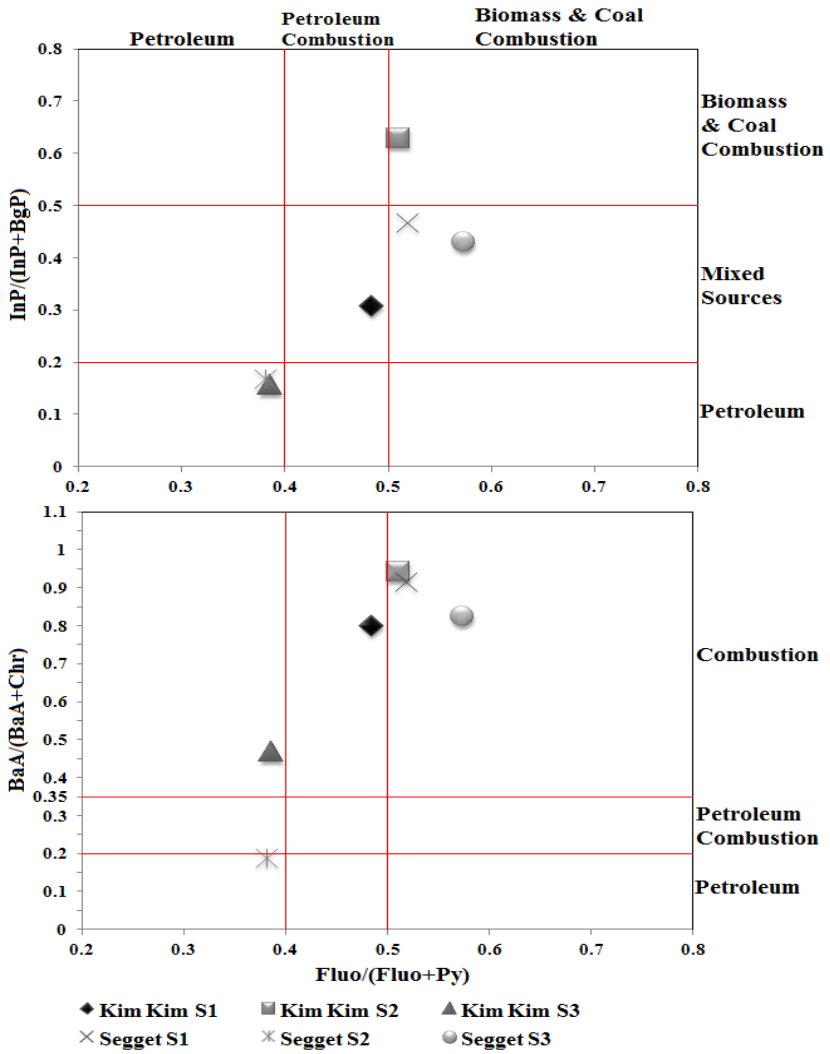


Figure 4. Plots of PAH isomer pair ratios for source identification: Fluo/(Fluo+Py) versus BaA/(BaA+Chr) and InP/(InP+BgP)

Determination of PAH Sources Using Hierarchical Cluster Analysis (HCA)

Hierarchical cluster analysis (Wards clustering method using average linkage between groups and Squared Euclidean distance as measure interval for PAH data normalised by Z score transformation) in package cluster (Maechler et al., 2012) was carried out to identify possible sources of individual PAHs in the surface sediments from the Kim Kim and Segget Rivers. It was assumed that PAHs with chemical configuration or similar ring number would be clustered among the samples. The hierarchical diagram clustered the 16 individual PAHs into major groups based on the similarity of PAH fingerprints (Figure 5 and Figure 6).

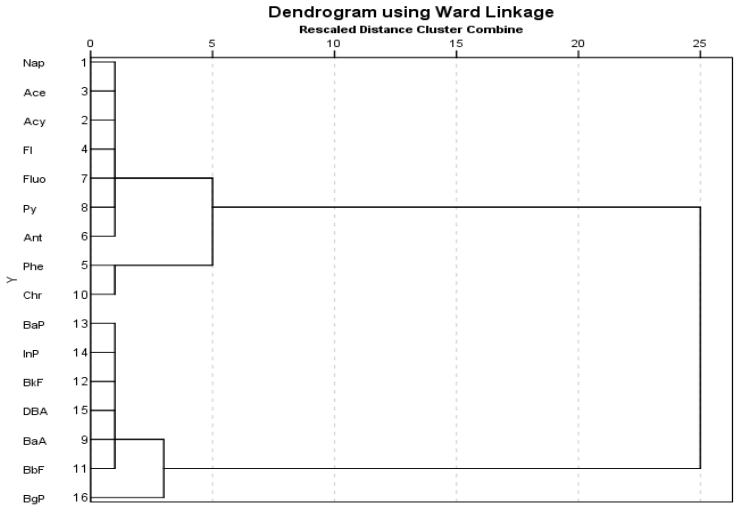


Figure 5. Hierarchical dendrogram for 16 individual PAHs in sediment from the Kim Kim River

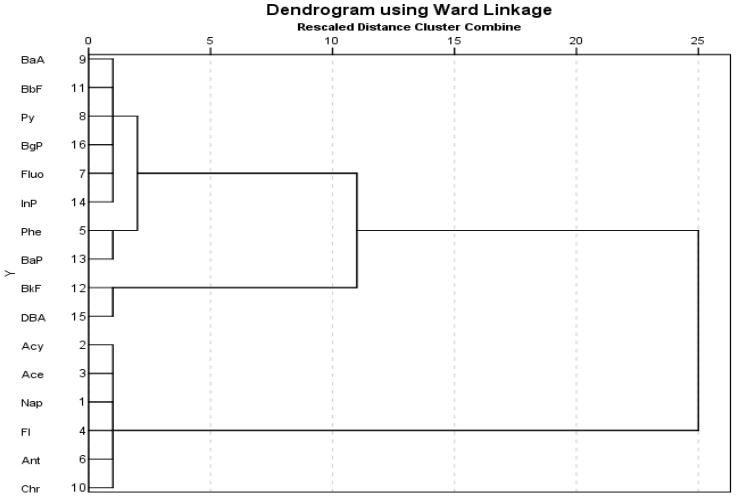


Figure 6. Hierarchical dendrogram for 16 individual PAHs in sediment from the Segget River

For PAHs in the Kim Kim River, cluster analysis led to four major groups. Group 1 included Nap, Ace, Acy, Fl, Fluo, Py and Ant, which belong mainly to lower molecular weight PAHs. Lower molecular PAHs including Nap, Acy, Ace, Fluo and Ant have been reported as markers of spills or crude oil leaks. Also, these compounds can be used as markers for low-temperature pyrogenic processes such as biomass combustion of straw and firewood (Jenkins et al., 1996; Yang et al., 2006; Zhang et al., 2008). As a result, Group 1 was believed to be mixed sources of PAHs derived from petroleum pollution and incomplete combustion of biomass. Group 2 included Phe and Chr, which can be identified as the markers of coal combustion source (Sofowote et al., 2008). The third group included BaP, InP, BkF, DBA, BaA and BbF. This group could be selected to represent emissions from vehicles because it aggregated mainly PAHs of higher molecular weight. Group 4 contained BgP, which identified as a tracer of gasoline emission (Harrison et al., 1996; Li & Kamens, 1993; Miguel & Pereira, 1989). The gasoline- and diesel-fuelled traffic on the road and boats in the study area may have contributed to these groups.

Similarly, cluster analysis represented four major groups for PAHs in the Segget River. The first group included BaA, BbF, Py, BgP, Fluo, InP, which are higher molecular weight PAHs with 4 to 5 rings. Five- and 6-ring PAHs were mostly found in road dust (Liu et al., 2007a; Ren et al., 2006) and vehicular emissions (Zuo et al., 2007). Group 2 was characterised by Phe and BaP. The research literature cited Phe as a tracer for coal combustion (Sofowote et al., 2008), while BaP is produced mostly during pyrolysis processes (Jung et al., 2010; Siddens et al., 2012). Based on field investigations, the Segget River passes through Johor Bahru; therefore, the road traffic in the city and residential and industrial activities near the Segget River may be the main contributors to PAH pollution in the Segget River. The third major group mainly consisted of BkF and DBA. Previous studies suggested that DBA indicates PAH compounds emitted from diesel combustion (Harrison et al., 1996; Larsen & Baker, 2003). Motelay-Massei et al. (2005) reported that BkF is a fossil-fuel combustion marker. Therefore, it is reasonable to attribute this group to vehicular emission. Group 4 was characterised by low molecular weight PAHs with 2 to 3 rings, including Acy, Ace, Nap, Fl and Ant, which are abundant in petrogenic sources and mainly are a result of petroleum spills (Dobbins et al., 2006; Liu et al., 2009). HCA suggested that both petrogenic and pyrogenic sources including traffic emissions, coal combustion, biomass burning and petroleum spills were the major sources of PAH contamination in sediments from the Kim Kim River and Segget River.

In the past two decades, population as well as urbanisation in Malaysia has had continuous growth, which caused an increase in the number of vehicles and road traffic. Possible sources for combustion-derivative PAHs in the sediments from the Kim Kim and Segget Rivers could be attributed to vehicular and industrial emissions, which are transported into the study areas via long-range atmospheric transport. This is consistent with Okuda et al. (2002) and Omar et al. (2002, 2006), who reported road-traffic emissions as the main origin of PAHs in atmospheric aerosols in Kuala Lumpur, Malaysia.

In addition, biomass burning from Sumatra, Indonesia has a portion of combustion-derivative PAHs in the study areas. He et al. (2010) reported higher level of PAHs in Singapore during the month of October 2006 compared to other months, especially two to three times at nighttime. Interestingly, satellite images showed hot spots in Sumatra Island in October, which

implied large biomass burning. The high concentration of retene in October in Singapore's atmosphere was attributed to long-range atmospheric transportation of carbonaceous material originated from biomass combustion in Indonesia (He et al., 2010). This is consistent with Okuda et al. (2002), who demonstrated that the large-scale biomass burning in Indonesia caused haze episodes in Malaysia, with high levels of carbonaceous material containing PAHs. As the Johor State is relatively close to Indonesia, the same influence can be suggested for the Johor State. However, local wood-burning can also be introduced as an important PAH source in the Kim Kim and Segget Rivers. For example, Okuda et al. (2002) reported a high portion (25%-35%) of local wood burning to Malaysian atmospheric PAHs. In other studies the measured PAH concentrations in Kula Lumpur during a hazy day were reported many folds higher than those on clear days (Abas et al., 2004; Omar et al., 2002), indicating the impact of biomass burning to PAH Levels.

Toxic Potency and Ecological Risk Assessment of PAHs

PAHs are widely distributed in diverse ecosystems, including freshwater and marine, and they have drawn significant attention due to their persistence and toxicity and mutagenic, teratogenic and carcinogenic effects. Sediment quality guidelines (SQGs) including effects range-low value (ERL), effects range-median value (ERM), probable effects level (PEL) and threshold effects level (TEL) (Long et al., 1995; Macdonald et al., 1996) are widely used for the purpose of ecological risk assessment of PAHs in aquatic ecosystems. The comparison between the measured concentrations of PAHs from sediments and the ERL, ERM, PEL and TEL values are listed in Table 2. The LMW PAHs, HMW PAHs and total PAH concentrations of sediment samples from the Kim Kim and Segget Rivers were below the ERL, ERM, PEL and TEL levels. It can be suggested that the ecological risk of PAHs was relatively low. However, PAH contamination of sediments from the study areas should be further evaluated for toxic potency assessment and ecosystem impairment due to rapid industrialisation, urbanisation and economy development in the region.

The ecological risk assessment technique of PAHs in marine and estuary surface sediments was also selected in our study (Long et al., 1995). The concentrations of effects range low (ERL) and effects range median (ERM) were applied to evaluate the ecological toxicity of individual PAH by some scholars (Mai et al., 2002). In addition, the quotient technique of average ERM was a complex manner of quantitative prediction for joint toxicity of pollutants in marine and estuary sediments. In the following equations,

$$\text{ERM-Q} = \frac{C_i}{ERM_i} \quad (1)$$

$$\text{M-ERM-Q} = \frac{\sum \text{ERM-Q}}{\text{n}} \quad (2)$$

where, C_i represents the measured concentration of given PAHs; ERM-Q is the risk quotient of ERM and M-ERM-Q stands for average risk quotients of ERM concentrations. When

M-ERM-Q is not 0.10, there rarely is observable toxicity to organisms (biological risk chance was less than 10%); when it is within 0.11-0.50, the possibility of ecological risk increases to 30% with low toxicity; when M-ERM-Q is between 0.51-1.50, the probability of ecological risk increases to 50% with high toxicity; when it is more than 1.50, the probability of ecological risk increases to 75% with higher toxicity (Long & MacDonald, 1998). The results indicated that values of M-ERM-Q in all sediment samples from the Kim Kim and Segget Rivers were lower than 0.1, which implied low probability (9%) of ecological effects.

Human Health Risk Assessment

The toxicity of a PAH compound is usually expressed by its BaP equivalent concentration (BaP_{eq}), and the toxicity equivalency factors (TEFs), proposed by Nisbet and LaGoy (1992). BaP is the most toxic PAH among all known potentially carcinogenic PAHs; hence, it is used as the reference chemical and assigned a value of 1 in the TEF system (Nisbet & LaGoy, 1992; Pufulete et al., 2004). The TEF values of other PAHs was calculated based on their carcinogenic level in comparison to that of BaP (Pufulete et al., 2004). The total BaPeq in sediment samples was calculated according to Eq. 3,

$$Total\ BaP_{eq} = \sum(C_i \times TEF_i) \quad (3)$$

where, BaP_{eq} is the equivalent BaP concentration of PAHs, Ci is the concentration of PAHi and TEFi is the toxic equivalency factor of individual PAHi (Table 2). Among 16 PAH compounds, seven PAHs, BaA, Chr, BaP, BbF, BkF, InP and DBA, have high toxic and carcinogenic effects. The potential toxicity of the sediment samples was then assessed using the toxic equivalent quotient (TEQ^{carc}) for seven carcinogenic PAHs calculated according to the following equation (Keshavarzifard et al., 2017c; Tian et al., 2013):

$$TEQ^{carc} = \sum(C_i \times TEF_i) \quad (4)$$

where, C_i is the concentration of an individual carcinogenic PAH (ng g^{-1}) and TEFi is the toxic factor of this carcinogenic PAH relative to BaP.

The BaP_{eq} of total PAHs in collected sediment samples from the Kim Kim River and the Segget River ranged from 4.65 to 37.98 ng g^{-1} and 37.17 to 123.03 ng g^{-1}, respectively (Table 4), indicating a relatively low level of BaP_{eq}. Total TEQ^{carc} in the Kim Kim River ranged from 4.44 to 37.66 ng g^{-1} -BaP_{eq}, with a mean concentration of 17.01 ng g^{-1} -BaP_{eq}. In addition, the sediment samples from the Segget River (TEQ^{carc} = 36.47 to 121.68 ng g^{-1} -BaP_{eq}, with a mean concentration of 90.64 ng g^{-1} -BaP_{eq}) contained slightly higher levels of total TEQ^{carc} than those from the Kim Kim River.

Table 4
Estimated ecological and human health risk caused by PAH contaminated sediment

Parameter	Kim Kim River			Segget River		
	S1	S2	S3	S1	S2	S3
M-ERM-Q	0.018	0.010	0.005	0.024	0.032	0.025
Total BaP_{eq}	10.08	37.98	4.65	123.03	37.17	114.26
TEQ^{car}	8.94	37.66	4.44	121.68	36.47	113.78
Cancer $risk_{ing}$	2.26E-04	1.24E-04	5.94E-05	3.45E-04	2.69E-04	2.06E-04
Cancer $risk_{der}$	1.17E-04	6.41E-05	3.09E-05	1.79E-04	1.39E-04	1.07E-04

In comparison, TEQ^{carc} values in sediments of the Kim Kim and Segget Rivers were lower than those of other areas reported in earlier studies (Table 5). The low BaPeq concentrations suggest potentially low carcinogenicity for both the Kim Kim and Segget Rivers. The contribution of each carcinogenic PAH to the toxic equivalent quotient of sediments in the study area varied according to the following order:

The Kim Kim River: BaP (60.91%) > DBA (19.22%) > BaA (9.42%) > BbF (5.27%) > InP (2.94%) > BkF (2.07%) > Chr (0.17%)

The Segget River: BaP (61.14%) > DBA (29.34%) > BkF (3.03%) > InP (2.81%) > BaA (1.88%) > BbF (1.73%) > Chr (0.07%).

Table 5
The toxic equivalent quotient (TEQ^{carc}) in sediment from different locations around the world

Location	TEQ^{carc} (ng g^{-1} -BaP_{eq})	References
Naples Harbour , Italy	2-4723	Sprovieri et al., 2007
Kaohsiung Harbour, Taiwan	3.9-1970	Chen et al., 2013
Meiliang Bay, China	94 to 856	Qiao et al., 2006
Barents Sea, Russia	18-300	Savinov et al., 2003
Langkawi Island, Malaysia	76.3 to174.6	Nasher et al., 2013

The results indicated that BaP and DBA had the highest portion of carcinogenic PAHs in sediment from both the Kim Kim and Segget Rivers.

Generally, there are pathways (i.e. ingestion and dermal contact) for humans to be exposed to PAHs via contaminated sediments. The assumption is that ingestion and dermal contact are more important ways. Therefore, the health risk assessments were determined based on the exposure factors and were calculated by assuming the daily intake, exposure time and skin contact (Keshavarzifard et al., 2017c). The health risk assessment through the ingestion and the dermal absorption pathways from sediments was calculated using the following equations:

$$Cancer\ Risk_{ingest} = \frac{C_s \times IngR \times EF \times ED}{BW \times At} \times CF \times SFO \quad (5)$$

$$Cancer\ Risk_{dermal} = \frac{C_s \times SA \times AF \times ABS \times EF \times ED}{BW \times At} \times CF \times SFO \quad (6)$$

where, Cancer $Risk_{ingest}$ is cancer risk through ingestion of sediment; Cs is the concentration of PAH compound; IngR is the ingestion rate: 100 mg day^{-1}; EF is the exposure frequency: 350 days $year^{-1}$; ED is the exposure duration: 70 years; BW is the body weight: 70 kg; At is the average day: 25,550 days; CF is the unit conversion factor: 10^{-6} kg mg^{-1}; SFO is the oral slope factor: 7.3 mg $(kg\ day)^{-1}$; Cancer $Risk_{dermal}$ is cancer risk through the dermal pathway of sediment; SA is the exposed skin surface area: 5700 cm^2; AF is the adherence factor from the sediment to skin: 0.07 mg cm^{-2}; and ABS is the dermal absorption from the sediment: 0.13. Due to different cancer risk ranges, the characteristics of cancer risk have been qualitatively categorised as $\leqq 10^{-6}$ (very low risk), from $10^{-6}<$ to $<10^{-4}$ (low risk), from $\leqq 10^{-4}$ to $<10^{-3}$ (moderate risk), from $10^{-3}\leqq$ to $<10^{-1}$ (high risk) and $\geqq 10^{-1}$ (very high risk) (Man et al., 2013).

The results showed that Cancer $Risk_{ingest}$ for sediment samples from Station 3 of the Kim Kim River was between 10^{-6} and 10^{-4}, indicating low risk from this station, while for Stations 1 and 2 of the Kim Kim River and all stations of the Segget River, it was from $\leqq 10^{-4}$ to $<10^{-3}$ (Table 4), implying moderate risk. The results demonstrated that Cancer $Risk_{dermal}$ for sediments from Stations 2 and 3 of the Kim Kim River was between 10^{-6} and 10^{-4}, indicating low risk level (Table 4). The Cancer $Risk_{dermal}$ for Station 1 of the Kim Kim River and all stations of the Segget River were between 10^{-4} and 10^{-3}, indicating moderate risk. Therefore, probabilistic health risk to humans via ingestion and dermal pathways from sediments of the Kim Kim and Segget Rivers can be categorised as low-to-moderate risk. The highest cancer risk occurred near Station 1 of both the Kim Kim River and the Segget River, which were near the coast. Therefore, people who live and work on the shore would likely have a higher risk of getting cancer.

CONCLUSION

The data from this study elucidated the 16 EPA PAHs contamination features in surface sediments from the Kim Kim River and Segget River, Peninsular Malaysia. The total PAH concentrations ranged from 95.17 to 361.24, with a mean value of 218.1 ng g^{-1} dw in sediments from the Kim Kim River and 330.09 to 552.76 with a mean value of 437.5 ng g^{-1} dw in sediments from the Segget River. The PAH concentrations can be classified as low to moderate in the Kim Kim River and moderate to the level of pollution in the Segget River. The higher concentrations of PAHs in the Segget River can be attributed to the heavy industrial activities and road traffic in Johor City and nearby areas. PAH molecular indices and HCA were applied to determine the possible sources of PAHs. The results indicated both petrogenic and pyrogenic input with significant dominance of pyrogenic PAHs. The concentrations of LMW, HMW and total PAHs were compared with SQGs to assess the potential ecological effects of the sedimentary PAHs in the study areas. The results indicated minimum adverse ecological effect from PAHs. The results of potential toxicity assessment including M-ERM-Q, total BaP_{eq}, and TEQ^{carc} implied low probability of ecological effects in the Kim Kim and Segget Rivers. The human health risk of PAH compounds in the Kim Kim and Segget Rivers through ingestion and dermal

contact of sediments was evaluated; the result was low-to-moderate risk from ingestion and dermal contact.

ACKNOWLEDGEMENT

This research was funded by the Inisiatif Putra Berkumpulan Grant (IPB) awarded by Universiti Putra Malaysia (Grant No. 9412401).

REFERENCES

Abas, M. R. B., Rahman, N. A., Omar, N. Y. M., Maah, M. J., Samah, A. A., Oros, D. R., ... & Simoneit, B. R. (2004). Organic composition of aerosol particulate matter during a haze episode in Kuala Lumpur, Malaysia. *Atmospheric Environment, 38*(25), 4223-4241. doi:10.1016/j.atmosenv.2004.01.048

Abdo Alkhadher, S. A., Zakaria, M. P., Yusoff, F. M., Kannan, N., Suratman, S., Magam, S. M., ... & Sani, M. S. A. (2016). Distribution and sources of linear alkyl benzenes (LABs) in surface sediments from Johor Bahru Coast and the Kim Kim River, Malaysia. *Environmental Forensics, 17*(1), 36-47. doi:10.1080/15275922.2015.1091405

Aldarondo-Torres, J. X., Samara, F., Mansilla-Rivera, I., Aga, D. S., & Rodríguez-Sierra, C. J. (2010). Trace metals, PAHs, and PCBs in sediments from the Jobos Bay area in Puerto Rico. *Marine Pollution Bulletin, 60*(8), 1350-1358. doi:10.1016/j.marpolbul.2010.06.006

Alkhadher, S. A. A., Zakaria, M. P., Yusoff, F. M., Kannan, N., Suratman, S., Keshavarzifard, M., ... & Sani, M. S. A. (2015). Baseline distribution and sources of linear alkyl benzenes (LABs) in surface sediments from Brunei Bay, Brunei. *Marine Pollution Bulletin, 101*(1), 397-403. doi:10.1016/j.marpolbul.2015.10.011

Araghi, P. E., Bastami, K. D., & Rahmanpoor, S. (2014). Distribution and sources of polycyclic aromatic hydrocarbons in the surface sediments of Gorgan Bay, Caspian Sea. *Marine Pollution Bulletin, 89*(1), 494-498. doi:10.1016/j.marpolbul.2013.12.001

Bakhtiari, A. R., Zakaria, M. P., Yaziz, M. I., Lajis, M. N. H., Bi, X., & Rahim, M. C. A. (2009). Vertical distribution and source identification of polycyclic aromatic hydrocarbons in anoxic sediment cores of Chini Lake, Malaysia: Perylene as indicator of land plant-derived hydrocarbons. *Applied Geochemistry, 24*(9), 1777-1787. doi:10.1016/j.apgeochem.2009.05.008

Baumard, P., Budzinski, H., Garrigues, P., Sorbe, J., Burgeot, T., & Bellocq, J. (1998). Concentrations of PAHs (polycyclic aromatic hydrocarbons) in various marine organisms in relation to those in sediments and to trophic level. *Marine Pollution Bulletin, 36*(12), 951-960. doi:10.1016/S0025-326X(98)00088-5

Boonyatumanond, R., Wattayakorn, G., Togo, A., & Takada, H. (2006). Distribution and origins of polycyclic aromatic hydrocarbons (PAHs) in riverine, estuarine, and marine sediments in Thailand. *Marine Pollution Bulletin, 52*(8), 942-956. doi:10.1016/j.marpolbul.2005.12.015

Bouloubassi, I., Roussiez, V., Azzoug, M., & Lorre, A. (2012). Sources, dispersal pathways and mass budget of sedimentary polycyclic aromatic hydrocarbons (PAH) in the NW Mediterranean margin, Gulf of Lions. *Marine Chemistry 142*, 18-28. doi:10.1016/j.marchem.2012.07.003

Budzinski, H., Jones, I., Bellocq, J., Pierard, C., & Garrigues, P. (1997). Evaluation of sediment contamination by polycyclic aromatic hydrocarbons in the Gironde estuary. *Marine Chemistry 58*(1-2), 85-97. doi:10.1016/S0304-4203(97)00028-5

Cao, B., Nagarajan, K., & Loh, K. C. (2009). Biodegradation of aromatic compounds: Current status and opportunities for biomolecular approaches. *Applied Microbiology and Biotechnology 85*(2), 207-228. doi:10.1007/s00253-009-2192-4

Chen, C. F., Chen, C. W., Dong, C. D., & Kao, C. M. (2013). Assessment of toxicity of polycyclic aromatic hydrocarbons in sediments of Kaohsiung Harbor, Taiwan. *Science of the Total Environment 463*, 1174-1181.

Dobbins, R. A., Fletcher, R. A., Benner, B. A., & Hoef, T. S. (2006). Polycyclic aromatic hydrocarbons in flames, in diesel fuels, and in diesel emissions. *Combustion and Flame 144*(4), 773-781. doi:10.1016/j.combustflame.2005.09.008

DOSM. (2014a). *Population and demography, report no. 4: Demographic indicators, Malaysia.* Malaysia: Department of Statistics Malaysia.

DOSM. (2014b). *Population and demography, report No. 1: Intercensal mid-year population estimates.* Malaysia: Department of Statistics Malaysia.

Guadalupe Meniconi, M. D. F., Gabardo, I. T., Carneiro, M. E. R., Barbanti, S. M., da Silva, G. C., & Massone, C. G. (2002). Brazilian Oil Spills Chemical Characterization--Case Studies. *Environmental Forensics, 3*(3-4), 303-321. doi:10.1080/713848377

Gustafsson, O., & Gschwend, P. M. (1997). Soot as a strong partition medium for polycyclic aromatic hydrocarbons in aquatic systems. In *ACS Symposium Series, 1997* (pp. 365-381). ACS Publications. doi:10.1021/bk-1997-0671.ch024

Gustafsson, Ö., Haghseta, F., Chan, C., MacFarlane, J., & Gschwend, P. M. (1996). Quantification of the dilute sedimentary soot phase: Implications for PAH speciation and bioavailability. *Environmental Science and Technology, 31*(1), 203-209. doi:10.1021/es960317s

Harrison, R. M., Smith, D., & Luhana, L. (1996). Source apportionment of atmospheric polycyclic aromatic hydrocarbons collected from an urban location in Birmingham, UK. *Environmental Science and Technology, 30*(3), 825-832. doi:10.1021/es950252d

He, J., Zielinska, B., & Balasubramanian, R. (2010). Composition of semi-volatile organic compounds in the urban atmosphere of Singapore: Influence of biomass burning. *Atmospheric Chemistry and Physics, 10*(23), 11401-11413 doi:10.5194/acp-10-11401-2010.

Hoffman, E. J., Mills, G. L., Latimer, J. S., & Quinn, J. G. (1984). Urban runoff as a source of polycyclic aromatic hydrocarbons to coastal waters. *Environmental Science and Technology, 18*(8), 580-587. doi:10.1021/es00126a003

Irwin, R. (1997). *National park service environmental contaminants encyclopedia PAHs entry.* Colorado: National Park Service.

Jenkins, B. M., Jones, A. D., Turn, S. Q., & Williams, R. B. (1996). Particle concentrations, gas-particle partitioning, and species intercorrelations for polycyclic aromatic hydrocarbons (PAH) emitted during biomass burning. *Atmospheric Environment, 30*(22), 3825-3835.

Jung, K. H., Yan, B., Chillrud, S. N., Perera, F. P., Whyatt, R., Camann, D., ... & Miller, R. L. (2010). Assessment of benzo (a) pyrene-equivalent carcinogenicity and mutagenicity of residential indoor versus outdoor polycyclic aromatic hydrocarbons exposing young children in New York City. *International Journal of Environmental Research and Public Health, 7*(5), 1889-1900.

Kanzari, F., Syakti, A., Asia, L., Malleret, L., Piram, A., Mille, G., & Doumenq, P. (2014). Distributions and sources of persistent organic pollutants (aliphatic hydrocarbons, PAHs, PCBs and pesticides) in surface sediments of an industrialized urban river (Huveaune), France. *Science of the Total Environment, 478*, 141-151. doi:10.1016/j.scitotenv.2014.01.065

Kanzari, F., Syakti, A. D., Asia, L., Malleret, L., Mille, G., Jamoussi, B., ... & Doumenq, P. (2012). Aliphatic hydrocarbons, polycyclic aromatic hydrocarbons, polychlorinated biphenyls, organochlorine, and organophosphorous pesticides in surface sediments from the Arc river and the Berre lagoon, France. *Environmental Science and Pollution Research, 19*(2), 559-576 doi:10.1007/s11356-011-0582-5

Karickhoff, S. W., & Brown, D. S. (1978). Paraquat sorption as a function of particle size in natural sediments. *Journal of Environmental Quality, 7*(2), 246-252. doi:10.2134/jeq1978.00472425000700020019x–

Karickhoff, S. W., Brown, D. S., & Scott, T. A. (1979). Sorption of hydrophobic pollutants on natural sediments. *Water Research, 13*(3), 241-248 doi:10.1016/0043-1354(79)90201-X

Keshavarzifard, M., & Zakaria, M. P. (2015). Polycyclic aromatic hydrocarbons (PAHs) contamination of surface sediments from Port Dickson, Malaysia: Distribution, sources and ecological risk assessment. *Environmental Forensics, 16*(4), 322-332. doi:10.1080/15275922.2015.1059392

Keshavarzifard, M., Moore, F., Keshavarzi, B., & Sharifi, R. (2017a). Distribution, source apportionment and health risk assessment of polycyclic aromatic hydrocarbons (PAHs) in intertidal sediment of Asaluyeh, Persian Gulf. *Environmental Geochemistry and Health*, 1-15. doi:https://doi.org/10.1007/s10653-017-0019-2

Keshavarzifard, M., Moore, F., Keshavarzi, B., & Sharifi, R. (2017b). Polycyclic aromatic hydrocarbons (PAHs) in sediment and sea urchin (Echinometra mathaei) from the intertidal ecosystem of the northern Persian Gulf: Distribution, sources, and bioavailability. *Marine Pollution Bulletin, 123*(1-2), 373-380. doi:https://doi.org/10.1016/j.marpolbul.2017.09.008

Keshavarzifard, M., Zakari, M. P., & Sharifi, R. (2017c). Ecotoxicological and health risk assessment of polycyclic aromatic hydrocarbons (PAHs) from consumption of short-neck clam (Paphia undulata), and contaminated sediment exposure, Malacca Strait, Malaysia. *Archives of Environmental Contamination and Toxicology*, 1-14. doi:10.1007/s00244-017-0410-0

Keshavarzifard, M., Zakaria, M. P., & Hwai, T. S. (2017d). Bioavailability of polycyclic aromatic hydrocarbons (PAHs) to short-neck clam (Paphia undulata) from sediment matrices in mudflat areas of West coast of Peninsular Malaysia. *Environmental Geochemistry and Health, 39*(3), 591-610. doi:10.1007/s10653-016-9835-z

Keshavarzifard, M., Zakaria, M. P., & Keshavarzifard, S. (2016). Evaluation of polycyclic aromatic hydrocarbons contamination in the sediments of the Johor Straits, Peninsular Malaysia. Polycyclic Aromatic Compounds, 1-16. doi:10.1080/10406638.2016.1257997

Keshavarzifard, M., Zakaria, M. P., Hwai, T. S., Halimoon, N., Mustafa, S., Vaezzadeh, V., ... & Weiyun, C. (2014a). Polycyclic aromatic hydrocarbons (PAHs) in sediments from Prai and Malacca Rivers, Peninsular Malaysia. In *From sources to solution* (pp. 415-420). Springer. doi:10.1007/978-981-4560-70-2_75

Keshavarzifard, M., Zakaria, M. P., Hwai, T. S., Yusuff, F. M., & Mustafa, S. (2015). Distributions and sour.ce apportionment of sediment-associated polycyclic aromatic hydrocarbons (PAHs) and hopanes in rivers and estuaries of Peninsular Malaysia. *Environment Science and Pollution Research, 22*(12), 9424-9437. doi:10.1007/s11356-015-4093-7

Keshavarzifard, M., Zakaria, M. P., Hwai, T. S., Mustafa, S., Vaezzadeh, V., Magam, S. M., ... & Abootalebi-Jahromi, F. (2014b). Baseline distributions and sources of polycyclic aromatic hydrocarbons (PAHs) in the surface sediments from the Prai and Malacca Rivers, Peninsular Malaysia. *Marine Pollution Bulletin, 88*(1), 366-372. doi:10.1016/j.marpolbul.2014.08.014

Koh, C. H., Khim, J., Kannan, K., Villeneuve, D., Senthilkumar, K., & Giesy, J. (2004). Polychlorinated dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs), biphenyls (PCBs), and polycyclic aromatic hydrocarbons (PAHs) and 2, 3, 7, 8–TCDD equivalents (TEQs) in sediment from the Hyeongsan River, Korea. *Environmental Pollution, 132*(3), 489-501. doi:10.1016/j.envpol.2004.05.001

Koike, T., Koike, H., Kurumisawa, R., Ito, M., Sakurai, S., Togo, A., ... & Takada, H. (2012). Distribution, source identification, and historical trends of organic micropollutants in coastal sediment in Jakarta Bay, Indonesia. *Journal of Hazardous Materials, 217*, 208-216. doi:10.1016/j.jhazmat.2012.03.023

Larsen, R. K., & Baker, J. E. (2003). Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. *Environmental Science and Technology, 37*(9), 1873-1881. doi: 10.1021/es0206184

Li, C. K., & Kamens R. M. (1993). The use of polycyclic aromatic hydrocarbons as source signatures in receptor modeling. *Atmospheric Environment Part A General Topics, 27*(4), 523-532,

Lipiatou, E., Tolosa, I., Simo, R., Bouloubassi, I., Dachs, J., Marti, S., ... & Albaiges, J. (1997). Mass budget and dynamics of polycyclic aromatic hydrocarbons in the Mediterranean Sea deep sea research Part II. *Topical Studies in Oceanography, 44*(3-4), 881-905. doi:10.1016/S0967-0645(96)00093-8

Liu, M., Cheng, S. B., Ou, D. N., Hou, L. J., Gao, L., Wang, L. L., ... & Xu, S. Y. (2007a). Characterization, identification of road dust PAHs in central Shanghai areas, China. *Atmospheric Environment, 41*(38), 8785-8795 doi:10.1016/j.atmosenv.2007.07.059

Liu, S., Tao, S., Liu, W., Liu, Y., Dou, H., Zhao, J., ... & Gao, Y. (2007b). Atmospheric polycyclic aromatic hydrocarbons in North China: A winter-time study. *Environmental Science and Technology, 41*(24), 8256-8261 doi:10.1021/es0716249

Liu, Y., Chen, L., Huang, Q. H., Li, W. Y., Tang, Y. J., & Zhao, J. F. (2009). Source apportionment of polycyclic aromatic hydrocarbons (PAHs) in surface sediments of the Huangpu River, Shanghai, China. *Science of the Total Environment, 407*(8), 2931-2938. doi:10.1016/j.scitotenv.2008.12.046

Long, E., & MacDonald, D. (1998). Recommended uses of empirically derived, sediment quality guidelines for marine and estuarine ecosystems. *Human and Ecological Risk Assessment, 4*(5), 1019-1039.

Long, E. R., MacDonald, D. D., Smith, S. L., & Calder, F. D. (1995). Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environmental Management, 19*(1), 81-97. doi:10.1007/BF02472006

Lye, K. H., & Eng, L. (1988). Environmental impact due to pollutant discharge in Johore Strait Towardsan: Integrated Management of Tropical Coastal Resources. In *ICLARM conference proceedings*. Manila.

Macdonald, D. D., Carr, R. S., Calder, F. D., Long, E. R., & Ingersoll, C. G. (1996). Development and evaluation of sediment quality guidelines for Florida coastal waters. *Ecotoxicology, 5*(4), 253-278. doi:10.1007/BF00118995

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2012). Cluster: cluster analysis basics and extensions. *R package version, 1*(2), 56.

Magam, S. M., Zakaria, M. P., Halimoon, N., Aris, A. Z., Kannan, N., Masood, N., ... & Sani, M. S. (2015). Evaluation of distribution and sources of sewage molecular marker (LABs) in selected rivers and estuaries of Peninsular Malaysia. *Environmental Science and Pollution Research, 23*(6), 5693-5704. doi:10.1007/s11356-015-5804-9

Magi, E., Bianco, R., Ianni, C., & Di Carro, M. (2002). Distribution of polycyclic aromatic hydrocarbons in the sediments of the Adriatic Sea. *Environmental Pollution, 119*(1), 91-98. doi:10.1016/S0269-7491(01)00321-9

Mai, B. X., Fu, J. M., Sheng, G. Y., Kang, Y. H., Lin, Z., Zhang, G., ... & Zeng, E. Y. (2002). Chlorinated and polycyclic aromatic hydrocarbons in riverine and estuarine sediments from Pearl River Delta, China. *Environmental Pollution, 117*(3), 457-474.

Man, Y. B., Kang, Y., Wang, H. S., Lau, W., Li, H., Sun, X. L., ... & Wong, M. H. (2013). Cancer risk assessments of Hong Kong soils contaminated by polycyclic aromatic hydrocarbons. *Journal of Hazardous Materials, 261*, 770-776.

Masood, N., Zakaria, M. P., Ali, M. M., Magam, S. M., Alkhadher, S., Keshavarzifard, M., ... & Hussein, M. A. (2014). Distribution of petroleum hydrocarbons in surface sediments from selected locations in Kuala Selangor River, Malaysia. In *From Sources to Solution* (pp. 351-356). Springer Singapore, Kuala Lumpur, Malaysia. doi:10.1007/978-981-4560-70-2_64, 351–356

Masood, N., Zakaria, M. P., Halimoon, N., Aris, A. Z., Magam, S. M., Kannan, N., ... & Alkhadher, S. A. A. (2016). Anthropogenic waste indicators (AWI) particularly PAHs and LABs in Malaysian sediments: Application of aquatic environment for identifying anthropogenic pollution. *Marine Pollution Bulletin, 102*(1), 160-175. doi:10.1016/j.marpolbul.2015.11.032

McGroddy, S. E., & Farrington, J. W. (1995). Sediment porewater partitioning of polycyclic aromatic hydrocarbons in three cores from Boston Harbor, Massachusetts. *Environmental Science and Technology, 29*(6), 1542-1550. doi:10.1021/es00006a016

Medeiros, P. M., Bícego, M. C., Castelao, R. M., Del Rosso, C., Fillmann, G., & Zamboni, A. J. (2005). Natural and anthropogenic hydrocarbon inputs to sediments of Patos Lagoon Estuary, Brazil. *Environment International, 3*1(1), 77-87. doi:10.1016/j.envint.2004.07.001

Miguel, A. H., & Pereira, P. A. (1989). Benzo (k) fluoranthene, benzo (ghi) perylene, and indeno (1, 2, 3-cd) pyrene: New tracers of automotive emissions in receptor modeling. *Aerosol Science and Technology, 10*(2), 292–295.

Mirza, R., Mohammadi, M., Faghiri, I., Abedi, E., Fakhri, A., Azimi, A., & Zahed, M. A. (2014). Source identification of polycyclic aromatic hydrocarbons (PAHs) in sediment samples from the northern part of the Persian Gulf, Iran. *Environmental Monitoring and Assessment, 186*(11), 7387-7398. doi:10.1007/s10661-014-3935-y

Motelay-Massei, A., Garban, B., Tiphagne-Larcher, K., Chevreuil, M., & Ollivon, D. (2006). Mass balance for polycyclic aromatic hydrocarbons in the urban watershed of Le Havre (France): Transport and fate of PAHs from the atmosphere to the outlet. *Water Research, 40*(10), 1995-2006. doi:10.1016/j.watres.2006.03.015

Motelay-Massei, A., Harner, T., Shoeib, M., Diamond, M., Stern, G., & Rosenberg, B. (2005). Using passive air samplers to assess urban − Rural trends for persistent organic pollutants and polycyclic aromatic hydrocarbons. 2. Seasonal trends for PAHs, PCBs, and organochlorine pesticides. *Environmental Science and Technology, 39*(15), 5763-5773.

Nasher, E., Heng, L. Y., Zakaria, Z., & Surif, S. (2013). Assessing the ecological risk of polycyclic aromatic hydrocarbons in sediments at Langkawi Island, Malaysia. *The Scientific World Journal 2013*.

Neff, J. M. (1979). *Polycyclic aromatic hydrocarbons in the aquatic environment. Sources, fates and biological effects.* London: Applied Science Publishers Ltd. (15 A NEF):274

Nelson, D., & Sommers, L. (1996). Total carbon, organic carbon and organic matter. In D. L. Sparks (Ed.), *Methods of soil analysis. Part 3: Chemical methods* (pp. 961-1010). Madison, WI: Soil Science Society of America.

Nisbet, I. C., & LaGoy, P. K. (1992). Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). *Regulatory Toxicology and Pharmacology, 16*(3), 290-300.

Notar, M., Leskovšek, H., & Faganeli, J. (2001). Composition, distribution and sources of polycyclic aromatic hydrocarbons in sediments of the Gulf of Trieste, Northern Adriatic Sea. *Marine Pollution Bulletin, 42*(1), 36-44. doi:10.1016/S0025-326X(00)00092-8

Okuda, T., Kumata, H., Zakaria, M. P., Naraoka, H., Ishiwatari, R., & Takada, H. (2002). Source identification of Malaysian atmospheric polycyclic aromatic hydrocarbons nearby forest fires using molecular and isotopic compositions. *Atmospheric Environment, 36*(4), 611-618. doi:10.1016/S1352-2310(01)00506-4

Omar, N. Y. M., Abas, M., Ketuly, K. A., & Tahir, N. M. (2002). Concentrations of PAHs in atmospheric particles (PM-10) and roadside soil particles collected in Kuala Lumpur, Malaysia. *Atmospheric Environment, 36*(2), 247-254. doi:10.1016/S1352-2310(01)00425-3

Omar, N. Y. M., Mon, T. C., Rahman, N. A., & Abas, M. (2006). Distributions and health risks of polycyclic aromatic hydrocarbons (PAHs) in atmospheric aerosols of Kuala Lumpur, Malaysia. *Science of the Total Environment, 369*(1), 76-81. doi:10.1016/j.scitotenv.2006.04.032

Oros, D. R., Ross, J. R., Spies, R. B., & Mumley T. (2007). Polycyclic aromatic hydrocarbon (PAH) contamination in San Francisco Bay: A 10-year retrospective of monitoring in an urbanized estuary. *Environmental Research 105*(1), 101-118. doi:10.1016/j.envres.2006.10.007

Pereira, W. E., Hostettler, F. D., Luoma, S. N., van Geen, A., Fuller, C. C., & Anima, R. J. (1999). Sedimentary record of anthropogenic and biogenic polycyclic aromatic hydrocarbons in San Francisco Bay, California. *Marine Chemistry 64*(1), 99-113 doi:10.1016/S0304-4203(98)00087-5

Pufulete, M., Battershill, J., Boobis, A., & Fielder, R. (2004). Approaches to carcinogenic risk assessment for polycyclic aromatic hydrocarbons: A UK perspective. *Regulatory Toxicology and Pharmacology, 40*(1), 54-66.

Qian, X., Liang, B., Fu, W., Liu, X., & Cui, B. (2016). Polycyclic aromatic hydrocarbons (PAHs) in surface sediments from the intertidal zone of Bohai Bay, Northeast China: Spatial distribution, composition, sources and ecological risk assessment. *Marine Pollution Bulletin 112*(1), 349-358.

Qiao, M., Wang, C., Huang, S., Wang, D., & Wang, Z. (2006). Composition, sources, and potential toxicological significance of PAHs in the surface sediments of the Meiliang Bay, Taihu Lake, China. *Environment International, 32*(1), 28-33.

Ren, Y., Zhang, Q., & Chen, J. (2006). Distribution and source of polycyclic aromatic hydrocarbons (PAHs) on dust collected in Shanghai, People's Republic of China. *Bulletin of Environmental Contamination and Toxicology, 76*(3), 442-449. doi:10.1007/s00128-006-0941-y

Savinov, V. M., Savinova, T. N., Matishov, G. G., Dahle, S., & Næs, K. (2003). Polycyclic aromatic hydrocarbons (PAHs) and organochlorines (OCs) in bottom sediments of the Guba Pechenga, Barents Sea, Russia. *Science of the Total Environment, 306*(1), 39-56.

Sicre, M., Marty, J., Saliot, A., Aparicio, X., Grimalt, J., & Albaiges, J. (1987). Aliphatic and aromatic hydrocarbons in different sized aerosols over the Mediterranean Sea: occurrence and origin. *Atmospheric Environment (1967), 21*(10), 2247-2259. doi:10.1016/0004-6981(87)90356-8

Siddall, R., Robotham, P., Gill, R., Pavlov, D., & Chuiko, G. (1994). Relationship between polycyclic aromatic hydrocarbon (PAH) concentrations in bottom sediments and liver tissue of bream (Abramis brama) in Rybinsk Reservoir, Russia. *Chemosphere, 29*(7), 1467-1476. doi:10.1016/0045-6535(94)90278-X

Siddens, L. K., Larkin, A., Krueger, S. K., Bradfield, C. A., Waters, K. M., Tilton, S. C., ... & Williams, D. E. (2012). Polycyclic aromatic hydrocarbons as skin carcinogens: comparison of benzo [a] pyrene, dibenzo [def, p] chrysene and three environmental mixtures in the FVB/N mouse. *Toxicology and Applied Pharmacology, 264*(3), 377-386.

Sofowote, U. M., McCarry, B. E., & Marvin, C. H. (2008). Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. *Environmental Science and Technology, 42*(16), 6007-6014.

Sprovieri, M., Feo, M. L., Prevedello, L., Manta, D. S., Sammartino, S., Tamburrino, S., & Marsella, E. (2007). Heavy metals, polycyclic aromatic hydrocarbons and polychlorinated biphenyls in surface sediments of the Naples harbour (Southern Italy). *Chemosphere, 67*(5), 998-1009.

Stout, S. A., Uhler, A. D., & Emsbo-Mattingly, S. D. (2004). Comparative evaluation of background anthropogenic hydrocarbons in surficial sediments from nine urban waterways. *Environmental Science and Technology, 38*(11), 2987-2994. doi:10.1021/es040327q

Takada, H., Onda, T., Harada, M., & Ogura, N. (1991). Distribution and sources of polycyclic aromatic hydrocarbons (PAHs) in street dust from the Tokyo Metropolitan area. *Science of the Total Environment, 107*, 45-69. doi:10.1016/0048-9697(91)90249-E

Thorsen, W. A., Cope, W. G., & Shea, D. (2004). Bioavailability of PAHs: Effects of soot carbon and PAH source. *Environmental Science and Technology, 38*(7), 2029-2037. doi:10.1021/es0306056

Tian, Y. Z., Li, W. H., Shi, G. L., Feng, Y. C., & Wang, Y. Q. (2013). Relationships between PAHs and PCBs, and quantitative source apportionment of PAHs toxicity in sediments from Fenhe reservoir and watershed. *Journal of Hazardous Materials, 248*, 89-96.

Tsai, P., Hoenicke, R., Yee, D., Bamford, H. A., & Baker, J. E. (2002). Atmospheric concentrations and fluxes of organic compounds in the northern San Francisco Estuary. *Environmental Science and Technology, 36*(22), 4741-4747. doi:10.1021/es011470b

Tung, C. Y., Ismail, R., & Jaafar, J. (2013). Passive sampling of polycyclic aromatic hydrocarbons in Johor Straits by Sol-gel Polydimethylsiloxane. *Jurnal Teknologi, 62*. doi:http://dx.doi.org/10.11113/jt.v62.1890

Ünlü, S., Alpar, B., Öztürk, K., & Vardar, D. (2010). Polycyclic aromatic hydrocarbons (PAHs) in the surficial sediments from Lake Iznik (Turkey): Spatial distributions and sources. *Bulletin of Environmental Contamination and Toxicology, 85*(6), 573-580. doi:10.1007/s00128-010-0134-6

Vaezzadeh, V., Pauzi Zakaria, M., Tan Shau-Hwai, A., Zaiton Ibrahim, Z., Mustafa, S., Keshavarzifard, M., ... & Abdullah Abdo Alkhadher, S. (2015). Source type evaluation of polycyclic aromatic hydrocarbons (PAHs) in surface sediments from the Muar River and Pulau Merambong, Peninsular Malaysia. *Environmental Forensics 16*(2), 135-142. doi:10.1080/15275922.2015.1022839

Vaezzadeh, V., Zakaria, M. P., Mustafa, S., Ibrahim, Z. Z., Shau-Hwai, A. T., Keshavarzifard, M., ... & Masood, N. (2014). Distribution of polycyclic aromatic hydrocarbons (PAHs) in sediment from Muar River and Pulau Merambong, Peninsular Malaysia. In *From sources to solution* (pp. 451-455). Springer Singapore, Kuala Lumpur, Malaysia. doi:10.1007/978-981-4560-70-2_81, 451-455

Wang, X. C., Zhang, Y. X., & Chen, R. F. (2001). Distribution and partitioning of polycyclic aromatic hydrocarbons (PAHs) in different size fractions in sediments from Boston Harbor, United States. *Marine Pollution Bulletin 42*(11), 1139-1149. doi:10.1016/S0025-326X(01)00129-1

Yang, H. H., Tsai, C. H., Chao, M. R., Su, Y. L., & Chien, S. M. (2006). Source identification and size distribution of atmospheric polycyclic aromatic hydrocarbons during rice straw burning period. *Atmospheric Environment, 40*(7), 1266-1274.

Yunker, M. B., Macdonald, R. W., Vingarzan, R., Mitchell, R. H., Goyette, D., & Sylvestre, S. (2002). PAHs in the Fraser River basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. *Organic Geochemistry 33*(4), 489-515. doi:10.1016/S0146-6380(02)00002-5

Zakaria, M. P., Takada, H., Tsutsumi, S., Ohno, K., Yamada, J., Kouno, E., & Kumata, H. (2002). Distribution of polycyclic aromatic hydrocarbons (PAHs) in rivers and estuaries in Malaysia: A widespread input of petrogenic PAHs. *Environmental Science and Technology, 36*(9), 1907-1918. doi:10.1021/es011278+

Zeng, S., Zeng, L., Dong, X., & Chen, J. (2013). Polycyclic aromatic hydrocarbons in river sediments from the western and southern catchments of the Bohai Sea, China: Toxicity assessment and source identification. *Environmental Monitoring and Assessment, 185*(5), 4291-4303. doi:10.1007/s10661-012-2869-5

Zhang, Y., Dou, H., Chang, B., Wei, Z., Qiu, W., Liu, S., ... & Tao, S. (2008). Emission of polycyclic aromatic hydrocarbons from indoor straw burning and emission inventory updating in China. *Annals of the New York Academy of Sciences, 1140*(1), 218-227.

Zuo, Q., Duan, Y., Yang, Y., Wang, X., & Tao, S. (2007). Source apportionment of polycyclic aromatic hydrocarbons in surface soil in Tianjin, China. *Environmental Pollution, 147*(2), 303-310. doi:10.1016/j.envpol.2006.05.029

Pertanika J. Sci. & Technol. 26 (1): 121 - 132 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Antiulcer Properties of Kelulut Honey against Ethanol-Induced Gastric Ulcer

Latifah Saiful Yazan[1,2]*, Nurul Amira Zainal[1], Razana Mohd Ali[2], Muhamad Firdaus Shyfiq Muhamad Zali[1], Ong Yong Sze[3], Tor Yin Sim[3], Banulata Gopalsamy[1], Voon Fui Ling[1], Sarah Sapuan[3], Nurulaidah Esa[3], Aminah Suhaila Haron[1], Fatin Hannani Zakarial Ansar[3], Ana Masara Ahmad Mokhtar[1] and Sharifah Sakinah Syed Alwi[1]

[1]*Department of Biomedical Science, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[2]*Deparment of Pathology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[3]*Laboratory of Molecular Biomedicine, Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

Ulcers in the gastrointestinal tract refer to any appreciable depth of break in the mucosa lining that may involve submucosa. Common types of ulcer include peptic, gastric and duodenal ulcer, which may lead to chronic inflammation. Ulcers may be caused by excessive alcohol intake or prolonged use of non-steroidal anti-inflammatory drugs (NSAID), in addition to several other factors. Conventional medication such as Omeprazole (proton pump inhibitor) and Ranitidine (H2 blockers) for management of ulcers may cause severe side effects such as myelosupression and abnormal heart rhythm. This has driven researchers to explore the potential of natural products for management of ulcers with reduced side effects. Kelulut honey (KH) is a type of honey that is produced by stingless bees from the *Trigona* species. It is believed to have a lot of medicinal properties such as being antimicrobial, antioxidant and antidiabetic. Yet, no scientific study has been carried out on its antiulcer properties. This study was carried out to determine the antiulcer properties of KH. Eighteen male *Sprague dawley* rats (5 to 6 weeks old, weighing between 200 and 300 g) were divided into three groups (n=6). The groups were 1) normal control group (without ulcer, without KH), 2) positive control group (with ulcer, without KH) and 3) treatment group (with

ARTICLE INFO

Article history:
Received: 24 November 2016
Accepted: 17 August 2017

E-mail addresses:
latifahsy@upm.edu.my (Latifah Saiful Yazan),
ieramyra_92@yahoo.com (Nurul Amira Zainal),
razanamohdali71@gmail.com (Razana Mohd Ali),
firdaus_shyfiq@yahoo.com
(Muhamad Firdaus Shyfiq Muhamad Zali),
ysooong_@hotmail.com (Ong Yong Sze),
yinsimtor@gmail.com (Tor Yin Sim),
banulatagopalsamy@yahoo.com (Banulata Gopalsamy),
voonfuiling@gmail.com (Voon Fui Ling),
sarajhr@yahoo.com (Sarah Sapuan),
nurulaidahesa@yahoo.com.my (Nurulaidah Esa),
aminahsuhaila91@gmail.com (Aminah Suhaila Haron),
fatinhannaniZA@yahoo.com (Fatin Hannani Zakarial Ansar),
anamasara@ymail.com (Ana Masara Ahmad Mokhtar),
sh_sakinah@upm.edu.my (Sharifah Sakinah Syed Alwi)
*Corresponding Author

ISSN: 0128-7680

ulcer, treated with KH). The treatment, KH (1183 mg/kg), was given twice daily for 30 consecutive days by oral administration. On Day 31, the rats were induced with absolute ethanol (5 mL/kg) via oral administration after being fasted for 24 h and were sacrificed 15 min after the induction. The stomach was collected for macroscopic and histopathological evaluation. Pretreatment with KH significantly reduced ($p<0.05$) both the total area of ulcer and the ulcer index compared to the positive control group. The percentage of ulcer inhibition in the KH pre-treated group was 65.56% compared with the positive control group. The treatment, KH, exhibited antiulcer properties against ethanol-induced gastric ulcer.

Keywords: Kelulut honey (KH), antiulcer, gastric ulcer

INTRODUCTION

The term 'ulcer' was first introduced in 1882 by Quike (Clinch, 1989). According to the 'Global Burden of Disease, 2004', 270,000 cases worldwide were reported on peptic ulcer (Mathers et al., 2008) and 360 cases at a ratio of 1:1 of duodenal ulcer to gastric ulcer reported in Malaysia by the Profile of Peptic Ulcer Disease (Kudva, 1988). Peptic ulcer is the most common ulcer presenting as a benign lesion that occurs mainly due to excessive acid and pepsin bathing the surface of the mucosa (Pillai et al., 2010). It can develop in different parts of the gastrointestinal tract such as the stomach as gastric ulcers and the duodenum as duodenal ulcers. Primarily, there are two major etiologies of peptic ulcer, infection by *Helicobacter pylori* and excessive intake of non-steroidal anti-inflammation agents (NSAIDs). Other contributory factors are smoking and nutritional deficiencies (Rang et al., 2014; Naesdal & Brown, 2006). Ulcers are usually detected upon symptoms such as burning epigastric pain that may occur before or after meals, bloating, fullness and nausea. In severe condition, this may lead to complications such as perforation and bleeding (Malfertheiner et al., 2009).

There are various classes of medication available for ulcer management including the Histamine-2 receptor antagonist such as cimetidine and ranitidine, proton-pump inhibitors such as omeprazole and esomeprazole and prostaglandin E analogues such as misoprostol (Chan & Lau, 2010). However, all these drugs may lead to several side effects such as nausea (Dharmani, 2006), headache, abnormal heart rhythm and myelosupression (Sandhya et al., 2013). Thus, it is important to continue searching for safer and more effective antiulcer drugs among natural products.

One of the potential candidates is Kelulut honey (KH), which is produced from bees of the *Trigona* sp. more commonly known as Kelulut bees in Malaysia. A characteristic that differs *Trigona* sp. from common bees of other species is that it is stingless. Honey contains roughly 80% carbohydrates (35% glucose, 40% fructose and 5% sucrose) and 20% water, serving as a great source of energy. Moreover, it also contains more than 180 substances, including amino acids, vitamins, minerals, organic acid phenol compounds and enzymes. The pH of honey is around 4.0 (Ouchemoukh et al., 2007). Traditionally, KH is used in treatments for anti-ageing, fast healing of internal injuries and cough and cold (Barakhbah, 2007). KH was also reported to have various pharmacological effects such as anti-diabetes, antioxidant and antibacterial (Siok

et al., 2014; Zainol et al., 2013). Antioxidation is considered to be associated with anti-ulcer properties (Bashkaran et al., 2011). The phenolic compound that is one of the phytochemicals found in propolis is believed to contribute to the antiulcer activity (Pillai et al., 2010) as it has strong correlation with antioxidant activity (Bertoncelj et al., 2007; Beretta et al., 2005; Meda et al., 2005). This study was carried out to determine the antiulcer properties of KH.

METHODOLOGY

Chemicals and Reagents

Formalin was purchased from Fisher Scientific, UK. Absolute alcohol was purchased from Sigma Aldrich, USA. Hematoxylin and eosin were purchased from CellPath, UK. Xylene was purchased from Jt Baker, USA. Kelulut honey (KH) was provided by Marbawi Food Processing and Trading, Kuala Kangsar, Perak.

Experiment Animals

Eighteen male *Sprague dawley* rats with initial weight approximately 200-300 g at the age of 5-6 weeks old were acclimatised with free access to standard chow pellets and water for about one week before the study. For animal husbandry, three rats were been placed in a cage. This study was conducted after obtaining ethical approval (FYP.2015/FPSK.001) from the Institutional Animal Care and Used Committee (IACUC) of Universiti Putra Malaysia.

Experiment Procedure

Following acclimatisation, the rats were divided randomly into three groups (n=6) as detailed below.

Group 1: Negative control group (without ulcer, without KH)

Group 2: Positive control group (with ulcer, without KH)

Group 3: Treatment group (with ulcer, with KH)

The animals were pre-treated with KH twice daily via oral administration at the dosage of 1183.33 mg/kg body weight of rat for 30 days. This dosage was calculated based on human consumption.

The body weight of the rats was recorded every 3 days until the day of termination. On Day 30, the rats were fasted for 24 hours. On the following day, the rats in Groups 2 and 3 were induced with absolute ethanol (5 mL/kg body weight of rat) via oral administration for an hour after the last treatment of KH. Fifteen minutes later, all the rats were anesthetised with ketamine-xylazine (75 mg/kg body weight of rat: 5 mg/kg body weight of rat) by intra-peritoneal injection. Approximately 5 mL of blood was collected via cardiac puncture using a 26 G, ½" needle (Terumo®, Belgium, Europe) into a non-heparinised and EDTA-containing tube for biochemical and hematological analyses, respectively. The blood that was collected in the non-heparinised tube was centrifuged at 780 x g for 10 min to separate the serum. For the

histopathological analysis, the stomach was removed, cut open through its greater curvature and placed on graph paper to be captured by camera. Finally, the stomach was fixed in 10% formalin for further analysis.

Determination of the Ulcer Index and Inhibition

The ulcer index (UI) was calculated according to Goyal (2002) as follows:

$$\text{Ulcer index (UI)} = \frac{10}{X}$$

$$\text{where, } X = \frac{\text{Total area of the stomach}}{\text{Area of ulcer on the stomach}}$$

$$\text{Percentage of ulcer inhibition} = \frac{\text{UI of positive control} - \text{UI of treated group}}{\text{UI of positive control group}} \times 100$$

Histopathological Analysis

Initially, the samples of stomach tissue were grossed using a grossing blade (FEATHER Microtome Blade High Profile) by cutting the area of the ulcer from the opened stomach to get all the four layers. The grossed samples were placed in a cassette and processed using an automated tissue processor (Leica TP 1020, Germany) for 14 h during which they were dehydrated, cleared and infiltrated with paraffin wax.

Next, the processed tissue was embedded in appropriate orientation onto a block with molten paraffin wax using an embedder machine (Leica EG 1160, Germany). The block containing the tissue was then trimmed to thickness of 10 µm and sectioned to thickness of 0.4 µm using a microtome machine (Leica RM2135, Germany). The thin section of the tissue was placed in a water bath at 42°C. Next, the sample was fished onto slides and dewaxed before being stained with hematoxylin and eosin using an automated staining machine (Tissue-Tek® Prisma™, Japan) for 1 h and 18 min. The stained slides were mounted with P-xylene-bis-pyridinium bromide (DPX) and left to dry before being observed under a light microscope (Leica, Germany).

Statistical Analysis

All the data were analysed using Graph Pad Prism version 5.0. Analysis of the ulcer area and percentage of inhibition between the treated and untreated group was performed using the unpaired t-test. The differences in body weight were analysed by One Way ANOVA. The values are considered significant at $p<0.05$.

RESULTS

Effect of Kelulut Honey on the Body Weight of the Rats

Figure 1 shows the body weight of the rats in the negative control, positive control and KH-treated groups throughout the experiment. The body weight of the rats in the three groups increased from Day 0 to Day 30. There was no significant difference ($p>0.05$) in the body weight of the rats among the groups throughout the experiment.

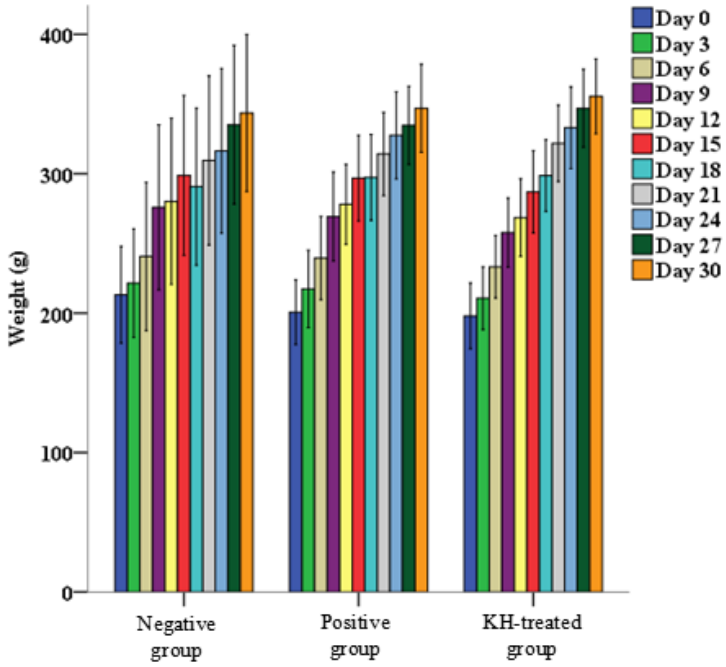


Figure 1. Effect of pretreatment with Kelulut honey for 30 days on the body weight of the rats. The data were analysed using One Way ANOVA and the values were expressed as Mean ± S.E.M. The value of $p<0.05$ was considered significant

Hematology Profile of Ethanol-Induced Gastric Ulcer in Rats

As shown in Figure 2, hematological analysis on the red blood cell (RBC), white blood cell (WBC), hemoglobin (Hb), thrombocytes, packed cell volume (PCV), mean corpuscular volume (MCV) and mean corpuscular hemoglobin concentration (MGHC) exhibited no significant difference between the positive control, negative control and pretreated group ($p>0.05$).

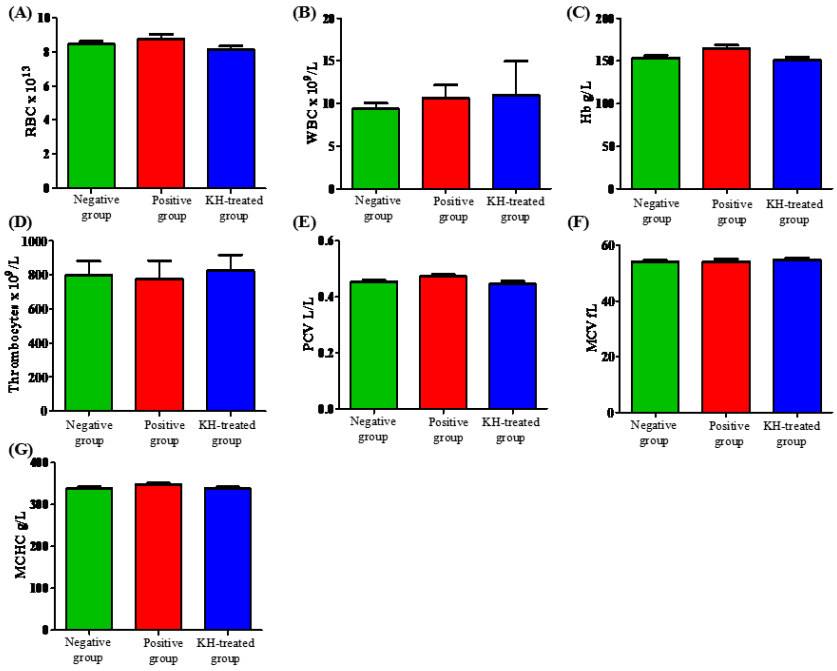


Figure 2. Effect of pretreatment with Kelulut honey for 30 days on the level of (A) RBC ($x10^{12}$/L), (B) WBC ($x10^{9}$/L), (C) Hb (g/L), (D) thrombocytes ($x10^{9}$/L), (E) PCV (L/L), (F) MCV (f/L), (G) MCHC (g/L) in the ethanol-induced gastric ulcer rats. The negative control group (without ulcer, without pretreatment) and positive control group (with ulcer, without pretreatment) were also included. The data were analysed using One Way ANOVA and the values were expressed as Mean ± S.E.M. The value of $p<0.05$ was considered significant

Effect of Pretreatment with Kelulut Honey on the Level of Liver Enzymes and Kidney Functions in the Ethanol-Induced Gastric Ulcer Rats

Figure 3 shows the level of urea, creatinine, aspartate transaminase (AST) alanine transaminase (ALT) and alkaline phosphatase (ALP) in the ethanol-induced gastric ulcer rats that were untreated (positive control group) and pretreated with KH for 30 days. The negative control was also included. The level of AST in the positive control and KH-treated groups was significantly higher ($p<0.05$) compared to the negative control group. Both levels of ALT and ALP showed no significant difference ($p<0.05$) when compared among the groups.

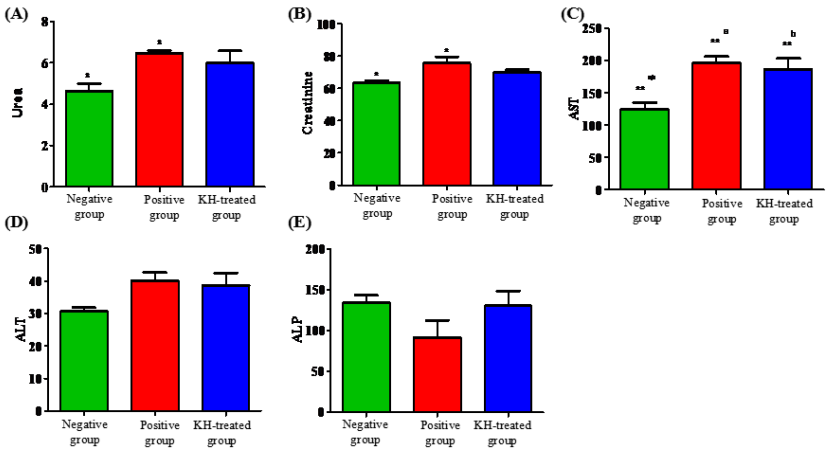


Figure 3. Effect of pretreatment with Kelulut honey for 30 days on the level of (A) urea (mmol/L), (B) creatinine (µmol/L), (C) AST (µ/L), (D) ALT (µ/L) and (E) ALP (µ/L) in the ethanol-induced gastric ulcer rats. The negative control group (without ulcer, without pretreatment) and positive control group (with ulcer, without pretreatment) were also included. The data were analysed using One Way ANOVA and the values were expressed as Mean ± S.E.M. The value of $p<0.05$ was considered significant

Effect of Pretreatment with Kelulut Honey on the Ulcer Area and Ulcer Index in the Ethanol-Induced Gastric Ulcer Rats

Figure 4 and Table 1 show the means of the ulcer area, ulcer index and percentage of ulcer inhibition in the stomach of the ethanol-induced gastric ulcer rats that were untreated (positive control group) and pretreated with KH for 30 days. There was a significant reduction ($p<0.05$) in the mean ulcer area at a value of 89.00 ± 38.87 and in the ulcer index at a value of 1.346 ± 0.5293 in the group pretreated with KH compared to the positive control group. The percentage of ulcer inhibition in the KH-treated group was 68.56%.

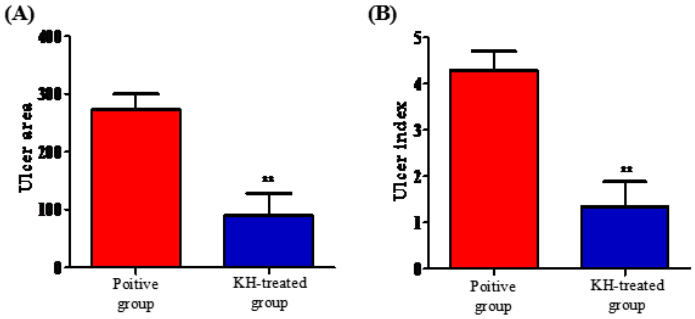


Figure 4. Effect of pretreatment with Kelulut honey for 30 days on the (A) ulcer area and (B) ulcer index in the ethanol-induced gastric ulcer rats. The positive control group (with ulcer, without pretreatment) was also included. The data were analysed using the unpaired t-test and the values were expressed as Mean ± S.E.M. The value of $p<0.05$ was considered significant

Table 1
Effect of pretreatment with kelulut honey on the ulcer area and ulcer index in ethanol-induced gastric ulcer rats between groups

Parameter	Positive Control Group (With Ulcer, Without KH Pretreatment)	Treatment Group (With Ulcer, with KH Pretreatment)	Percentage of Ulcer Inhibition (%)
Total Ulcer Area	272.4 ± 27.39	89.00 ± 38.87**	
Ulcer Index	4.281 ± 0.4157	1.346 ± 0.5293**	68.56

Note: All values were expressed as Mean ± S.E.M. The value of $p<0.05$ was considered significant

Effect of Pretreatment with Kelulut Honey on the Gross Pathology of the Stomach of the Ethanol-Induced Gastric Ulcer Rats

Figure 5 shows the macroscopic images of the stomach from the negative control, positive control and KH-treated groups opened along the greater curvature. Diffuse severe hemorrhage and multifocal area of hemorrhage were noted on the surface area of the stomach from the positive control and KH-treated groups, respectively.

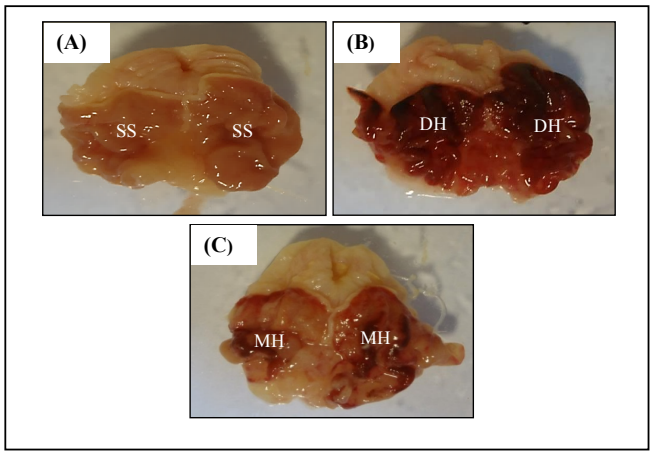


Figure 5. Macroscopic images of the stomach from (A) the negative control group with a smooth surface (SS), (B) positive control group with diffuse severe hemorrhage (DH) and (C) KH-treated group with the multifocal area of hemorrhage (MH) opened along the greater curvature

Effect of Pretreatment with Kelulut Honey on the Histological Analysis of the Stomach of the Ethanol-Induced Gastric Ulcer Rats

Figure 6 and 7 show the microscopic images of the stomach from the negative control, positive control and KH-treated groups after 30 days.

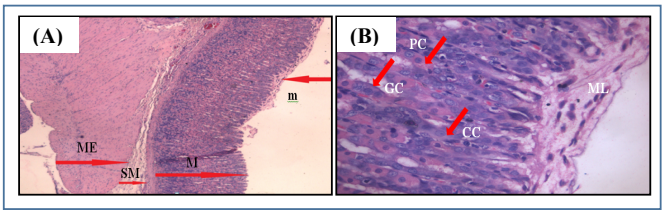


Figure 6. The microscopic images of the stomach from the negative control group; (A) histological analysis of a normal stomach with intact layer of mucosa (M), submucosa (SM), muscularis externa (ME) and the presence of a mucus layer on the epithelium surface (m); (B) histology of the upper portion of the normal stomach/gastric mucosa with a mucus layer (ML) covering its surface
Note: CC = chief cells; PC = parietal cells; GC = glandular cells. Image A was at 40× magnification, whereas image B was at 400× magnification

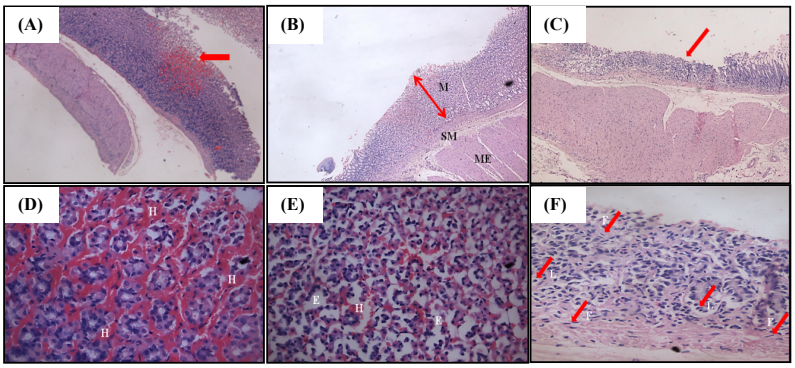


Figure 7. The microscopic images of the stomach from the positive and KH-treated groups; (A) histology of the untreated stomach with focal hemorrhage in the mucosa (arrow); (B) histology of the KH-treated stomach with intact mucosa (M), submucosa (SM) and muscularis external (ME) layer; (C) histology of the KH-treated stomach with a very thin layer of mucosa resulting from the erosion of the superficial part of the mucosa (arrow); (D) histology of the mucosa layer of the untreated stomach with hemorrhage (H) in the lamina propia. (E) histology of the mucosa layer of the untreated stomach with hemorrhage (H) and edema (E). (F) histology of the remaining part of the mucosa layer of the KH-treated stomach with fibroblast cells (F) and leukocyte infiltration (L)
Note: Images A-C were at 40× magnification, whereas images D-F were at 400× magnification

DISCUSSION

This study was done to evaluate the antiulcer properties of Kelulut honey (KH) in a model of ethanol-induced gastric ulcer in rats. Ethanol has been widely used to induce gastric lesions in experiment animals (Abdulla et al., 2010). The ability of absolute ethanol to induce gastric lesions is related to the formation of free radicals and reactive oxygen species (Nordmann, 1994; Cho et al., 1991). Ethanol has also been reported able to increase lipid peroxidation. Free radicals are able to attack cells, causing damage to the gastric cells of the gastrointestinal tract and leading to erosive lesions and ulceration (Hirokawa et al., 1998).

The results of this study demonstrated the potency of KH in increasing the rats' body weight in all three groups throughout the 30 days of experiment before the rats were induced with ethanol. It was also observed that KH was able to exert antiulcer effects in the ethanol-

induced gastric ulcer model via macroscopic and microscopic evaluation. Diffuse severe hemorrhage was first observed in the positive control group compared to only few areas of mild hemorrhage in the other treatment groups. This was due to the vascular injury caused by the absolute ethanol (Szabo et al., 1995). However, upon treatment with KH, the total ulcer area and ulcer index were reduced compared to the positive control group.

Further histological analysis demonstrated the disruption of the upper mucosal layer, accompanied by severe focal hemorrhage in the negative control group. This was caused by the ability of the ethanol to penetrate the protective mucus layer by dissolving the constituents of the stomach mucus (Szabo, 1987) and subsequently coming into contact with the epithelium surface, causing damage to the cells. It is also reported that severe focal hemorrhage and edema in the lamina propia of the mucosal layer were due to microvascular injury and an increase in vascular permeability as a consequence of microvascular endothelium disruption caused by absolute ethanol (Szabo et al., 1995).

Interestingly, upon pretreatment with KH, no severe hemorrhage or edema was observed in the mucosa. Although vascular congestion was detected, this could have been due to increased blood flow in the gastric mucosa caused by ethanol, leading to hyperemia (Endoh et al., 1993). Nevertheless, hyperemia is not as serious a condition as hemorrhage. The presence of tissue granulation and leukocyte infiltration such as by monocytes and neutrophils in this group could indicate the occurrence of the rapid healing process. The healing process may be divided into three overlapping phases, which are the inflammatory, proliferative and maturation phases (Regan & Barbul, 1994). Leukocyte infiltration such as by circulating monocytes (macrophage) and neutrophils as well as lymphocytes are involved in the inflammatory phase, in which they are crucial for healing (Barbul et al., 1989; Leibovich & Ross, 1975). This is due to the function of the macrophage and neutrophils in removing tissue debris caused by injury. In addition, the presence of fibroblasts in the mucosa cells of the KH pretreated group was an indicator of the proliferative phase of the healing process. Fibroblasts are important in repair processes as they are responsible for the production of structural proteins such as collagen, fibronectin, glycosaminoglycan and hyaluronic acid for the reconstruction of the affected tissue (Regan & Barbul, 1994).

The antiulcer properties of KH suggest that it is high in antioxidant properties. KH, which is from the *Trigona* sp., is intense in colour and high in phenolic content (Kek et al., 2014). Its colour intensity indicates the presence of carotenoids and flavonoids, which are renowned as natural antioxidants (Moniruzzaman et al., 2013b; Saxena et al., 2010), while total phenolic content is one of the reliable parameters for indicating antioxidative activities (Bertoncelj et al., 2007; Beretta et al., 2005; Meda et al., 2005) as well as effectively healing induced gastric ulcers (Bafna & Balaraman, 2005).

CONCLUSION

Kelulut honey was found to be able to increase rats' body weight as well as to possess antiulcer properties that were seen to reduce the total ulcer area and ulcer index in rats with ethanol-induced gastric.

AUTHORS' CONTRIBUTION

LSY supervised, evaluated the data and reviewed the manuscript for publication. RMA analysed the histological slides, while SSSA reviewed the manuscript for publication. NAZ performed the study as her undergraduate final-year project. OYZ, SS, NAE, MFSMZ, TYS, BG and VFL helped with preparation of the experiment studies. FHZA, ASH and AMAM helped with preparation of the manuscript. The authors declare that they have no competing interests.

ACKNOWLEDGEMENT

This work was supported by Universiti Putra Malaysia. The authors also wish to thank all staff of the Biomedical Science Department for help during the study.

REFERENCES

Abdulla, M. A., Al-Bayaty, F. H., Younis, L. T., & Abu Hassan, M. I. (2010). Anti-ulcer activity of Centella asiatica leaf extract against ethanol-induced gastric mucosal injury in rats. *Journal of Medicinal Plants Research, 4*(13), 1253-1259.

Bafna, P. A., & Balaraman, R. (2005). Anti-ulcer and anti-oxidant activity of pepticare, a herbomineral formulation. *Phytomedicine, 12*(4), 264-270.

Barakhbah, S. A. S. A. (2007). Honey in the Malay tradition. *Malaysia Journal of Medical Sciences, 14*(1), 106-110.

Barbul, A., Breslin, R. J., Woodyard, J. P., Wasserkrug, H. L., & Efron, G. (1989). The effect of *in vivo* T helper and T suppressor lymphocyte depletion on wound healing. *Annals of Surgery, 209*(4), 479-483.

Beretta, G., Granata, P., Ferrero, M., Orioli, M., & Facino, R. M. (2005). Standardization of antioxidant properties of honey by a combination of spectrophotometric/fluorimetric assays and chemometrics. *Analytica Chimica Acta, 533*(2), 185-191.

Bertoncelj, J., Doberšek, U., Jamnik, M., & Golob, T. (2007). Evaluation of the phenolic content, antioxidant activity and colour of Slovenian honey. *Food Chemistry, 105*(2), 822-828.

Chan, F. K. L., & Lau, J. Y. W. (2010). Treatment of peptic ulcer disease. In M. Feldman, L. S. Friedman, & L. J. Brandt (Eds.), *Sleingers and Fordtrans' gastrointestinal and liver Disease: Pathophysiology/ diagnosis/ management* (pp. 869-884). Philadelphia: Saunders Elsevier.

Cho, C. H., Pfeiffer, C. J., & Misra, H. P. (1991). Ulcerogenic mechanism of ethanol and the action of sulphanilyl fluoride on rat stomach *in vivo*. *Journal of Pharmacy and Pharmacology, 43*(7), 495-498.

Clinch, J. (1989). Perinatal mortality 1986 & 1987. *Irish Journal of Medical Sciences, 158*, 148-149.

Dharmani, P., & Palit, G. (2006). Exploring Indian medicinal plants for antiulcer activity. Indian Journal of Pharmacology, 38(2), 95-99.

Endoh, K., Kao, J., Domek, M. J., & Leung, F. W. (1993). Mechanism of gastric hyperemia induced by intragastric hypertonic saline in rats. *Gastroenterology, 104*, 114-121.

Goyal, R. K. (2002). *Practical in pharmacology* (pp. 7-10). Ahmedabad: Shah Prakahan.

Hirokawa, M., Miura, S., Yoshida, H., Kurose, I., Shigematsu, T., Hokari, R. … & Ishii, H. (1998). Oxidative stress and mitochondrial damage precedes gastric mucosal cell death induced by ethanol administration. *Alcoholism, Clinical and Experimental Research, 22* (Supplement), 111S-114S.

Kek, S. P., Chin, N. L., Yusof, Y. A., Tan, S. W., & Chua, L. S. (2014). Total phenolic content and colour intensity of Malaysian honeys from the *Apis spp*. and *Trigona spp*. bees. *Agriculture and Agricultural Science Procedia, 2*, 150-155.

Kudva, M. V. (1988). Profile of peptic ulcer disease in Malaysia. *Singapore Medical Journal, 29*(6), 544-547.

Leibovich, S. J., & Ross, R. (1975). The role of the macrophage in wound repair. A study with hydrocortisone and antimacrophage serum. *American Journal of Pathology, 78*(1), 71-100.

Malfertheiner, P., Chan, F. K., & McColl, K. E. (2009). Peptic ulcer disease. *Lancet, 374*(9699), 1449-1461.

Mathers, C., Fat, D. M., & Boerma, J. T. (2008). The global burden of disease: 2004 update. *World Health Organization* (pp. 1-133). Switzerland: World Health Organization.

Meda, A., Lamien, C. E., Romito, M., Millogo, J., & Nacoulma, O. G. (2005). Determination of the total phenolic, flavonoid and proline content in Burkina Fasan honey, as well as their radical scavenging activity. *Food Chemistry, 91*(3), 571-577.

Moniruzzaman, M., Sulaiman, S. A., Khalil, M. I., & Gan, S. H. (2013). Evaluation ofmphysicochemical and antioxidant properties of sourwood and other Malaysian,honeys: A comparison with Manuka honey. *Chemistry Central Journal, 7*, 1-12.

Naesdal, J., & Brown, K. (2006). NSAID-associated adverse effects and acid control aids to prevent them: A review of current treatment options. *Drug Safety, 29*, 119-132.

Nordmann, R. (1994). Alcohol and antioxidant systems. Alcohol and Alcoholism, 29(5), 513-522.

Ouchemoukh, S., Louaileche, H., & Schweitzer, P. (2007). Physicochemical characteristics and pollen spectrum of some Algerian honeys. *Food Control, 18*(1), 52-58.

Pillai, S. I., Kandaswamy, M., & Subramanian, S. (2010). Antiulcerogenic and ulcer healing effects of Indian propolis in experimental rat ulcer models. *Journal of ApiProduct and ApiMedical Science, 2*, 21-28.

Rang, H. P., Ritter, J. M., Flower, R. J., & Henderson, G. (2014). *Rang & Dale's pharmacology: With student consult online access.* Elsevier Health Sciences.

Regan, M. C., & Barbul, A. (1994). The cellular biology of wound healing. *Wound Healing* (pp. 3-17). New York: Springer-Verlag.

Sandhya S., Venkata, R. K., Vinod, K. R., Reddy, S., & Begum, A. (2013). Scope of medicinal flora as effective anti-ulcer agents. *African Journal of Plant Science, 7*(11), 504-512.

Saxena, S., Gautam, S., & Sharma, A. (2010). Physical, biochemical and antioxidant properties of some Indian honeys. *Food Chemistry, 118*(2), 391-397.

Szabo, S. (1987). Mechanisms of mucosal injury in the stomach and duodenum: Time sequence analysis of morphologic, functional, biochemical and histochemical studies. *Scandinavian Journal of Gastroenterology, 22*(S127), 21-28.

Zainol, M. I., Yusoff, K. M., & Yusof, M. Y. M. (2013). Antibacterial activity of selected Malaysian honey. *BMC Complementary and Alternative Medicine, 13*(1), 1-10.

Pertanika J. Sci. & Technol. 26 (1): 133 - 160 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Revocable and Non-Invertible Multibiometric Template Protection based on Matrix Transformation

Jegede, A.[1,2]*, Udzir, N. I.[1], Abdullah, A.[1] and Mahmod, R.[1]

[1]*Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

[2]*Department of Computer Science, University of Jos, 930001 Nigeria*

ABSTRACT

Biometric authentication refers to the use of measurable characteristics (or features) of the human body to provide secure, reliable and convenient access to a computer system or physical environment. These features (physiological or behavioural) are unique to individual subjects because they are usually obtained directly from their owner's body. Multibiometric authentication systems use a combination of two or more biometric modalities to provide improved performance accuracy without offering adequate protection against security and privacy attacks. This paper proposes a multibiometric matrix transformation based technique, which protects users of multibiometric systems from security and privacy attacks. The results of security and privacy analyses show that the approach provides high-level template security and user privacy compared to previous one-way transformation techniques.

Keywords: Multibiometric, matrix transformation, performance, privacy, security

INTRODUCTION

Multibiometric authentication systems use a combination of two or more biometric modalities to verify the identity of users. Multibiometric systems can be implemented using image and voice data (Nishino et al., 2012), face and iris (Karmakar & Murphy, 2014), palm print and iris (Hariprasath & Prabakar, 2012), ear and finger knuckle images (Tharwat et al., 2012), as well as hand vein, iris and fingerprint (Xiuyan et al., 2011). Performance evaluation results from these studies show that multibiometric systems generally have much better recognition accuracy than unibiometric approaches. This is because multibiometric systems use a richer set of inputs that helps to improve recognition performance. However, multibiometric

ARTICLE INFO

Article history:
Received: 08 August 2016
Accepted: 15 August 2017

E-mail addresses:
abayomi.jegede@ymail.com (Jegede, A.),
izura@upm.edu.my (Udzir, N. I.),
azizol@upm.edu.my (Abdullah, A.),
ramlan@gmail.com (Mahmod, R.)
*Corresponding Author

ISSN: 0128-7680

approaches face challenges in terms of template security and fusion complexity (Kaur, 3013). Multibiometric systems are not exempted from security challenges, such as replay attacks. Legitimate users of multibiometric systems also face privacy violations, such as template sharing and cross-matching. The risk of identity loss also occurs if an attacker who obtains digital versions of users' biometric data is able to reconstruct original biometric images from compromised biometric templates (Feng & Yuen, 2012; Li & Cot, 2011). Loss of identity is a severe risk because biometric modalities cannot be easily replaced unlike a password, PIN, or chip. This implies that victims will lose two or more biometric means of identity because multibiometric systems use at least two or more modalities. This paper addresses security, privacy and risk of loss of identity in multibiometric systems by proposing a one-way (or non-invertible) transformation approach known as matrix transformation. Matrix transformation is a simple and effective method that prevents security attacks on multibiometric systems and protects users against a privacy violation as well as identity loss. This approach guarantees the security of stored biometric data and the privacy of legitimate users, even if protected biometric data and security parameters are disclosed to an attacker. This approach also supports revocability or template renewability because a new biometric template can be created and used to replace the one suspected to have been stolen, corrupted or compromised.

Non-invertible transformation belongs to a general class of techniques known as cancellable biometrics, whose goal is to achieve 'an intentional, repeatable distortion of a biometric signal based on a chosen transform' (Ratha et al, 2001). This provides template security, revocability, unlinkability, and resistance to cross-matching attacks. Non-invertible transformations provide template security by preventing the recovery of original biometric data from a transformed template, even if an attacker knows the transformation parameters. Revocability allows administrators to remove a compromised template and reissue a new one based on the same biometric data. Template revocability is achieved by changing the transformation parameters used for the previous enrolment. Moreover, multiple transformed templates can be constructed from a single biometric input of the same subject, which provides unlinkability and prevents cross-matching among transformed templates that are stored in multiple databases.

Non-invertible transformation can be divided into two categories, namely image-level transforms and feature-level transforms. Image-level transforms, such as grid morphing, block permutation (Ratha et al., 2001), blind deconvolution (Campisi & Egiazarian, 2007; He et al., 2008), block re-mapping, texture warping (Farberbock et al., 2010; Hammerle-Uhl et al., 2009), GREY-COMBO (Zuo et al., 2008) and Log-Polar transform (Plesca & Morogan, 2013) are used to create an irreversible version of a given biometric image prior to feature extraction. Feature-level transformation uses approaches such as Cartesian; polar and functional transforms (Ratha et al., 2006; Ratha et al., 2007); revocable biotokens (Boult, 2006; Boult et al., 2007); pseudo-random permutations (Grassi & Faundez-Zanuy, 2009); Gaussian distribution (Jeong & Teoh, 2010); partial Hadamard matrix (Wang & Hu, 2013); pulse active transform (Safie et al., 2014); BIN-COMBO (Zuo et al., 2008); user-specific secret permutations (Rathgeb & Uhl, 2010); alignment-free adaptive bloom filter (Rathgeb et al., 2014); and the Delaunay triangle (Sandhya et al., 2016) to create non-invertible templates from extracted biometric features. A

number of techniques, such as user-dependent multi-state discretisation (Teoh et al., 2010), a combination of multi-dimensional iris codes, bit permutation and key binding (Ouda et al., 2011) and spiral cube (Moujahdi et al., 2012) focus on improving the security and recognition accuracy of some existing non-invertible transformation techniques.

Definitions

Galois field. A Galois (or finite) field contains a finite number of elements. A Galois field consisting of q elements is denoted as $gf(q)$. Formally, a Galois field, $gf(p^n)$ is defined in Equation 1 as:

$$\begin{aligned} GF(p^n) = (0, 1, 2, \dots, p-1) \cup (p, p+1, p+2, \dots, p+p-1) \\ \cup (p^2, p^2+1, p^2+2, \dots, p^2+p-1) \\ \cup \dots \cup p^{n-1}, p^{n-1}+1, p^{n-1}+2, \dots, p^{n-1}+p-1 \end{aligned} \quad (1)$$

where $p \in \mathbb{P}$ and $n \in \mathbb{Z}^+$ p^n defines the order of the field or the number of elements in the field, p is the characteristic of the field and the degree of polynomial of each element is at most n-1 (Benvenuto, 2012). The binary (base-2 number) system represents each value as 0 or 1. The binary system provides an alternative way to represent the elements of a Galois field. Each decimal element, x of a Galois field, can be expressed in binary as $a_n 2^n$. That is,

$$x = \sum_{n \in \mathbb{N}} a_n 2^n \quad (2)$$

where a_n is the binary coefficient and n is the degree of the polynomial.

A Galois field of two elements (also known as binary field), (2) , contains values that are represented by 0 and 1. The concept of $gf(2)$ is applicable to digital systems (such as computers) which represent data and operations in binary (series of 0s and 1s).

Permutation matrices. A permutation matrix is a square matrix whose elements are all 0s and 1s, with each row and column containing exactly a 1 (Fuzhen, 2011). It is a square matrix obtained from an $n \times n$ identity matrix by a permutation of rows (Grinshpan, 2011). Formally, a permutation is defined in Equation 3 as:

$$\pi = \begin{pmatrix} 1 & 1 & \dots & m \\ \pi(1) & \pi(2) & \dots & \pi(m) \end{pmatrix} \quad (3)$$

where $\pi(1), \pi(2), \dots, \pi(m) \in (1, 2, \dots, m)$; thus, a permutation matrix can be expressed as

$$P_{ij} = \begin{cases} 1, & if\ i = \pi(j) \\ 0, & otherwise \end{cases}$$ (Berisha et al., 2012).

Non-Invertible matrices. Given any two square ($n \times n$) matrices A and B, matrix A is said to be invertible (non-singular or non-degenerate) if the following condition holds:

$$A * B = B * A = I_n \tag{4}$$

where I_n is an $n \times n$ identity matrix and * denotes ordinary matrix multiplication. The matrix B referred to as the inverse of A (denoted by A^{-1}) is uniquely determined by A. A square matrix which does not satisfy the condition above is said to be non-invertible. Non-invertible matrices are also called singular or degenerate matrices. A square matrix is singular if and only if its determinant is 0. A square matrix that is not invertible is called singular or degenerate. A square matrix is singular if and only if its determinant is 0. A singular matrix is obtained by performing a random selection based on a continuous uniform distribution of its entries.

Non-invertible matrices in a Galois field of two elements (or $gf(2)$) are obtained by performing an *xor* operation on a pair of permutation matrices in $gf(2)$. If $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ are two elementary permutation matrices, a non-invertible matrix, is computed by the operation $C=A$ *xor* B. That is, $C = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

The matrix is non-invertible because the values for its determinant and inverse are undefined. That is, if $y=Cx$, then $x=yC^{-1}$. C^{-1} is undefined. Hence x cannot be computed given the value of y.

METHODOLOGY

This section discusses the methods used for feature extraction, implementation, performance evaluation and security analysis.

Feature Extraction

Binary feature vectors were extracted directly from pre-processed face images using the rotation invariant neighbour-based invariant local binary pattern (RINLBP) technique. RINLBP is a hybrid approach which integrates the generic local binary pattern (Ojala et al., 2002) and neighbour-based local binary pattern (Hamouchene & Aouat, 2014). The approach improves on the generic local binary pattern (LBP) by addressing poor recognition performance due to image rotation. The LBP is a texture classification method that combines a set of local texture descriptors to provide a global textural representation of an image. The LBP descriptor of a local circular region is computed by comparing the value of the central pixel with each of its

neighbours. The result of the comparison is 1 if the value of the pixel is greater that the central pixel, otherwise the result is 0. LBP is defined in Equation 5 as:

$$LBP_{R,P} = \sum_{p=0}^{P-1} s(g_p - g_c).2^p \tag{5}$$

such that

$$s(x) = \begin{Bmatrix} 1 & x \geq 0 \\ 0 & x < 0 \end{Bmatrix} \tag{6}$$

where
g_p is greyscale value of the neighbour pixel,
g_c is the value of the central pixel,
p is the index of the neighbour
R is the radius of the circular region
P is the number of sample points in the neighbourhood of the central pixel
(Ojala et al., 2002).

Figure 1 illustrates the operation of the generic LBP.

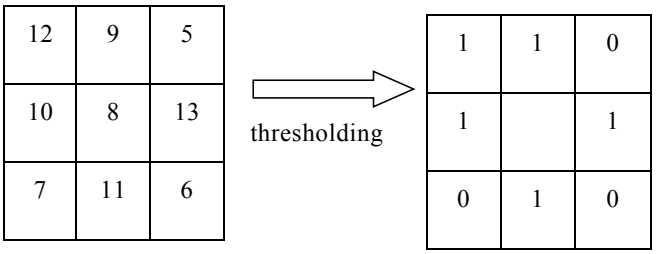


Figure 1. Generic LBP

The LBP is popular because it is simple to calculate and shows good performance. It is also robust against changes in illumination, which leads to changes in the values of pixel intensities. This is because features are not represented using the actual pixel values. Rather, they are computed by comparing the intensity values of a central pixel and its neighbours. A change in intensity value of a central pixel will lead to a corresponding change in the values of the neighbour pixels. Neighbour-based LBP (NLBP) compares the pixel value of each neighbour

of the central pixel with its next neighbour along the circular region. Neighbour-based LBP is defined in Equation 7 as:

$$NLBP_{R,P} = \sum_{p=0}^{P-1} s(g_p - g_{p+1}).2^p \quad (7)$$

such that

$$s(g_p - g_{p+1}) = \begin{cases} 1 & g_p \geq g_{p+1} \\ 0 & g_p < g_{p+1} \end{cases}$$

where
g_p is greyscale value of a neighbour pixel,
g_{p+1} is the value of the next pixel along the circular region,
p is the index of the neighbour
R is the radius of the circular region
P is the number of sample points in the neighbourhood of the central pixel (Hamouchene & Aouat, 2014).

The encoding begins with the topmost left neighbour and follows a clockwise direction (see Figure 2). This is unlike the generic LBP, which compares each neighbour with the central pixel. The generic LBP and the neighbour-based LBP generate different binary patterns from the same pixels.

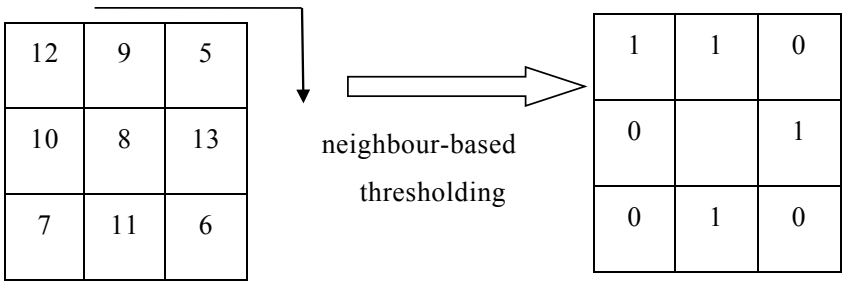


Figure 2. Neighbour-based LBP

The Rotation Invariant Neighbour-based LBP, $RINLBP_{R,P}$, is defined in Equation 8 as:

$$RINLBP_{R,P} = \sum_{p=0}^{P-1} s(g_p - g_{p+1}) . 2^{mod(p-d,P)} \quad (8)$$

such that

$$s(g_p - g_{p+1}) = \begin{cases} 1 & g_p \geq g_{p+1} \\ 0 & g_p < g_{p+1} \end{cases} \quad (9)$$

$$d = max|g_p - g_c| \quad (10)$$

$$p \in (0,1 \ldots P-1) \quad (11)$$

where
g_c, g_p, g_{p+1}, p, R and P are as previously defined,

d is the index of the neighbour pixel with the highest value, which defines the dominant direction in a neighbourhood. RINLBP provides rotation invariance by starting the encoding process with the neighbour pixel that has the highest value. This ensures that there is a corresponding rotation of the extracted binary pattern whenever the image is rotated.

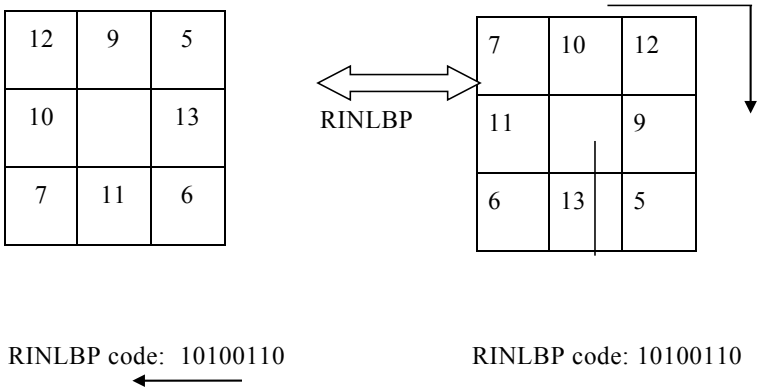


Figure 3. Rotation invariant neighbour-based LBP

The image in Figure 3 is rotated through an angle of 90° before applying the NLBP technique. The figure shows that RINLBP computes the same binary pattern from the original and the rotated images. This shows that image rotation does not affect the value of the binary pattern encoded by the RINLBP operator. We resized each face image to 16×8 before applying the RINLBP. This enabled us to obtain a 1,024-bit binary representation of the face image.

Segmentation isolates the iris from other structures that can affect the accuracy of the recognition process. This process involves detecting the inner and outer boundaries of the iris as well as the eyelids and eyelashes, which can interrupt the circular contour of the limbus boundary. The circular Hough transform is used for detecting the iris and pupil boundaries. Hough transform is defined as $x^2+y^2=r^2$, where (x,y) are the coordinates of the centre of the iris and pupil and r is the radius of the circular iris/pupillary boundaries. Figure 4 illustrates a segmented iris image.

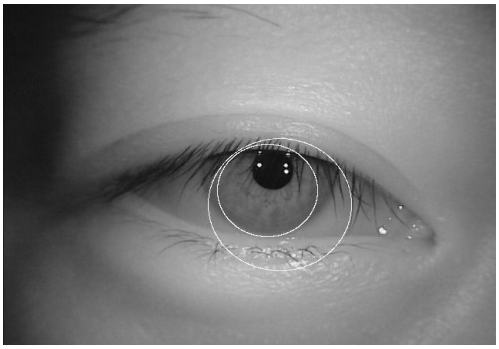


Figure 4. Segmented iris

Normalisation addresses variations in pupil size and provides for translation and scale invariance in order to ensure that irises of different individuals are mapped onto a common domain, since pupil size can vary across subjects. Normalisation is defined in Equation 12 as:

$$I(x(r,\theta),\ y(r,\theta)) \rightarrow I(r,\theta) \tag{12}$$

such that

$$x(r,\theta) = (1-r)x_p(\theta) + rx_l(\theta) \tag{13}$$

$$y(r,\theta) = (1-r)y_p(\theta) + ry_l(\theta) \tag{14}$$

where $I(x,y)$ is the iris image, (x,y) denotes the original Cartesian coordinates, (r,θ) are the corresponding normalised polar coordinates, (x_p,y_p) and $(x_l,y_l$ are the coordinates of the pupil and iris boundaries along the θ direction (Masek, 2003). Normalisation is usually carried out using the rubber sheet (Daugman, 2002) model. The rubber sheet model is illustrated in Figure 5.

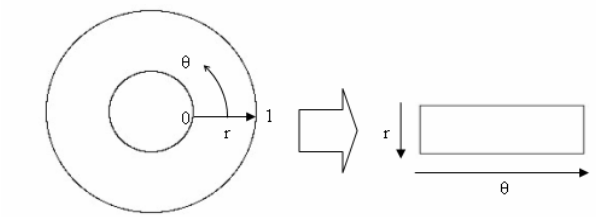


Figure 5. Rubber sheet model

Each point within the Cartesian coordinate is translated to a pair of polar coordinates (r, θ), where r lies within the range [0, 1] and θ is an angle in the range [0,2π]. Figure 6 illustrates a normalised iris image.

Figure 6. Polar iris image without noise

The rubber sheet model produces normalised irises of fixed dimension by taking both pupillary dilation and variations in pupil size into account. This ensures the extraction of iris codes of same dimension even if the size of the pupil varies across different subjects.

Feature extraction involves the application of a convolution operation to the 1D signals (obtained by breaking the 2D normalised iris image) using 1D Gabor wavelets. A Log-Gabor filter is defined in Equation 15 as:

$$G(f) = exp\left(\frac{-(log(f/f_0))^2}{2(log(\sigma/f_0))^2}\right) \quad (15)$$

where f_0 is the frequency and σ is the bandwidth of the filter (Field, 1987). The frequency response determines whether a given frequency value is quantised as 0 or 1. Feature extraction produces a binary template containing a number of bits of information that represent the iris image. The total number of bits in the template is two times the product of the angular resolution, the radial resolution and the number of filters used. A 1024-bit iris code is obtained by setting the values of angular resolution, radial resolution and filter to 8, 128 and 1, respectively.

The feature level fusion technique was used to create multibiometric templates from binary face and iris features. The process was carried out by transforming the templates into row vectors and appending one at the end of the other. That is,

$$MultiBio = FeatFace * FeatIris \quad (16)$$

where *MultiBio, FeatFace* and *FeatIris* are the multibiometric, face and iris feature vectors, respectively.

Implementation

The algorithm in Figure 7 describes the procedure for creating the transformation matrix.

Algorithm 1: *Computation of transformation matrix*

Input *P*

For *k = 1 to 2*

Read *i, j*

$row(i) \leftrightarrow row(i \pm j)$

Output m_1, m_2

$M_T = m_1 \oplus m_2$

End

Figure 7. Algorithm for computation of transformation matrix

Each elementary permutation matrix is computed by randomly selecting and interchanging a pair of rows of a general permutation matrix. Generally, a total of *n!* elementary permutation matrices can be computed from an $n \times n$ general permutation matrix. The combination of two elementary permutation matrices (from a given set of n! elementary permutation matrices) to obtain a non-invertible matrix produces a total of $\binom{n!}{2} = n!_{C_2} = \frac{(n!)!}{(n!-2)2!} = \frac{n!(n!-1) \ldots (n!-2+1)}{2!}$ non-invertible matrices. The non-invertible matrix (or transformation key), M_T is created by XORing two elementary permutation matrices, m_1 and m_2. That is,

$$M_T = m_1 \, XOR \, m_2 \tag{17}$$

The binary string, X (that is, the reference biometric template) is first converted into a one-dimensional column vector. A non-invertible template, X_T , is created by multiplying the transformation matrix, M_T , with the column vector representing X. That is,

$$X_T = M_T * X \tag{18}$$

The transformed template, X_T , is stored in the database instead of the original feature vector, X. The algorithm for enrolment is shown in Figure 8.

Algorithm 2: *Enrolment*

Input M_T, N

For *all* $i \in N$

Read X^i

$X_T^i = M_T * X^i$

End

Figure 8. Algorithm for enrolment

The algorithm in Figure 9 describes the procedure for authentication.

Algorithm 3: *Authentication*

Input M_T, N

For *all* $i \in N$

Read Y^i, X_T^i

$Y_T^i = M_T * Y^i$

Z=*hamming* (X_T^i, Y_T^i)

If $Z < \rho$ ***then***

accept

Else

reject

End

Figure 9. Algorithm for authentication

Performance Evaluation

The performance of the scheme was evaluated using 756 face images and 756 iris images of 108 subjects (or users) from CASIA Near Infrared face database (Li et al., 2007) and CASIA iris image database version 1 (biometrics.idealtest.org/dbDetailForUser.do?id=2), respectively. An additional dataset consisting of 196 face images of 100 subjects obtained from Labeled Faces in the Wild (Huang et al., 2008) was also used to evaluate the recognition accuracy of the proposed approach. The recognition accuracy was reported in terms of false acceptance

rate (FAR) and false rejection rate (FRR). False rejection results from a situation where the Hamming Distance between a pair of same-user templates is greater than the threshold. False acceptance occurs when the hamming distance between a pair of transformed templates belonging to different users is less than the threshold. The threshold used for the experiments was 0.30. The justification for the choice of this value is explained as follows. The range of intra-class variation between iris textures of the same person is 10-20% whereas irises of different subjects differ by 40-60% (Hao et al., 2006). This means that a threshold of at most 20% or 0.20 would effectively discriminate same-person iris images from those of different subjects. The same threshold was applied to face and multibiometric data in order to provide a fair comparison of the performance of the matrix transformation technique on different biometric modalities.

Table 1 presents the performance evaluation results for the application of matrix transformation on face images obtained from CASIA Near Infra Red (NIR) database.

Table 1
Performance results – Matrix Transformation (Face – CASIA NIR database)

Hamming distance	Recognition accuracy (%)	
	FRR	FAR
0.20	4.629	78
0.25	0.231	99
0.30	0.0	100
0.35	0.0	100

Results in the table show that FRR decreases as the hamming distance increases while FAR increases with a corresponding increase in threshold value. This is because an increase in hamming distance lowers the rate at which legitimate users are treated as impostors but with an increase in the number of impostors who are accepted as valid users. Conversely, a lower hamming distance results in an increase in FRR and a reduction in FAR. That is, there is more likelihood for legitimate users to be regarded as impostors and less likelihood for impostors to be accepted as valid users. An increase in hamming distance results in low intra-class variation and low inter-class distance. That is, there is higher correlation among the biometric data of same subjects and less variation among the biometric data of different subjects.

Table 2 shows the performance evaluation results for the application of matrix transformation on face images obtained from Labeled Faces in the Wild (LFW) dataset.

Table 2
Performance results – Matrix Transformation (Face – LFW database)

Hamming distance	Recognition accuracy (%)	
	FRR	FAR
0.20	0.0	14.4
0.25	0.0	33.1
0.30	0.0	88.9
0.35	0.0	100

The table shows that the FRR for all the Hamming distance values was 0%. This implies that a high correlation exists among face images of same users. The FAR increases with a corresponding increase in Hamming distance. A high Hamming distance value increases the collision among biometric data of different subjects. This leads to an increase in the rate at which impostors are accepted as valid users. A comparison of the results in Tables 1 and 2 shows that the proposed approach has better recognition accuracy for the LFW dataset than it does for the CASIA NIR database. This is because the LFW dataset consists of colour images while the CASIA NIR database contains greyscale images. The LBP and its variants have better recognition accuracy on colour images than greyscale images (Choi et al., 2012). However, both tables show that matrix transformation has low FRR (between 0% and ~4.63) but an intolerable FAR (between 14.4% and 100%). Hence, it is difficult to plot a suitable ROC curve based on the performance results obtained. This implies that the application of matrix transformation on face modality has poor recognition accuracy based on the chosen hamming distance values. Finding optimal hamming distance values required for good recognition accuracy for face (or any biometric) is outside the scope of this work. The focus of the work was to provide a fair comparison of the recognition performance of matrix transformation on face, iris and multibiometric data.

The result of the application of matrix transformation on iris data is presented in Table 3.

Table 3
Performance results – Matrix Transformation (Iris)

Hamming distance	Recognition accuracy (%)	
	FRR	FAR
0.20	64.35	0
0.25	32.02	0
0.30	8.769	0.007
0.35	1.466	0.139

The table shows that FRR decreases from 64.35% to 1.466% as the hamming distance increases from 0.20 to 0.35. On the other hand, the FAR increases from 0% to 0.139% for the chosen hamming distance values. Increasing the hamming distance results in lower intra-class variations among the biometric data of same subjects and higher correlation among the data of different subjects. Low intra-class variation leads to the rejection of fewer valid users (lower false rejection) and acceptance of more impostors (higher false acceptance). Moreover, increasing the hamming distance allows more impostors to be accepted as valid users and fewer valid users to be treated as impostors.

Figure 10 is the ROC curve for the application of matrix transformation technique on iris templates. The graph illustrates the relationship between FAR and FRR for different values of Hamming distance.

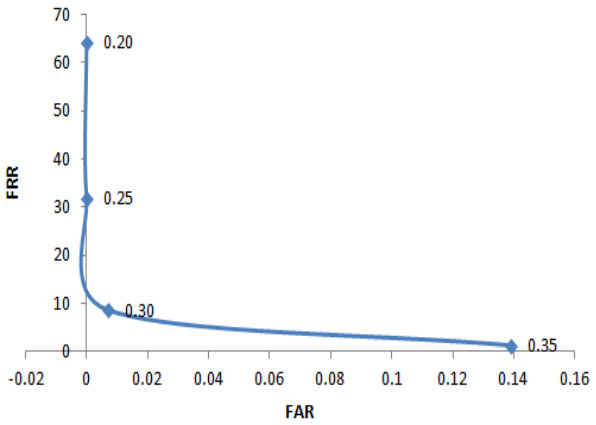


Figure 10. ROC curve for matrix transformation (iris)

The maximum value of FRR is approximately 64.4% and the minimum value of FRR is slightly below 1.47%. FAR has a maximum value of about 0.14% and a minimum value of approximately 0%. The curve shows that FRR reduces as FAR increases and vice versa. The value of FAR for each hamming distance is lower than the corresponding value of FRR. This implies that the approach sacrifices recognition accuracy and user convenience for security.

Table 4 shows the performance evaluation results for the application of matrix transformation on multibiometric data.

Table 4
Performance results – Matrix Transformation (Multibiometric)

Hamming distance	Recognition accuracy (%)	
	FRR	FAR
0.20	29.78	0.007
0.25	19.75	0.46
0.30	10.42	5.17
0.35	4.93	20.67

The FRR for the lowest hamming distance is ~29.8% while the highest hamming distance value has a false rejection rate of ~4.9%. Increasing the hamming distance leads to a reduction in both intra-class variation and inter-class distance. In other words, the rate at which genuine users are treated as impostors reduces with a corresponding increase in hamming distance. On the other hand, an increase in hamming distance results in the acceptance of more impostors as valid users. A lower hamming distance results in the rejection of more genuine users (higher FRR) and the acceptance of fewer impostors (lower FAR).

The relationship between the false rejection rate and false acceptance rate based on the application of matrix transformation on multibiometric data is illustrated by the ROC curve in Figure 11.

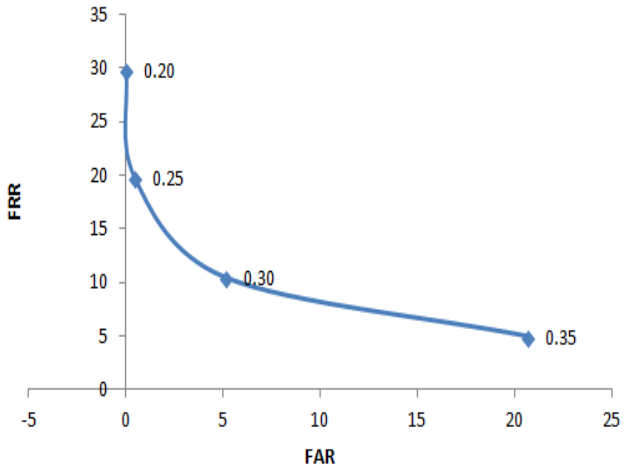


Figure 11. ROC curve for matrix transformation (multibiometric)

The maximum value of FAR is ~20.7% and the minimum value is ~0.007%. FRR has maximum and minimum values of approximately 29.8% and 4.9%, respectively. The curve shows that the FRR for each threshold is higher than the corresponding FAR. This implies that a higher premium is placed on security than on user convenience.

Security Analysis

The security of the proposed scheme was analysed using selected parameters, such as key length, key space, entropy and probability of correct guess in order to determine its resistance to security attacks (such as guessing and key exhaustion), privacy attacks (such as record multiplicity and cross matching) and template reconstruction attack.
Key length is the number of bits in the transformation key. It is expressed as the square of the dimension of the transformation matrix. Key length, K_l , is defined in Equation 19 as:

$$K_l = n^2 \tag{19}$$

where n is the dimension of the transformation matrix.

$\therefore$ key length 2^{20}= 1,048,576 bits. These keys are long enough to prevent guessing attack.

Key space, Ks_{pace} , is computed using the formula:

$$K_{space} = n!_{C_2} = \frac{(n!)!}{(n!-2)2!} = \frac{n!(n!-1)\ \ldots\ (n!-2+1)}{2!} \tag{20}$$

where n is the dimension of the transformation key.

The value of *n* used for the experiments is 1024.

$$\Rightarrow K_s = \frac{(1024!)!}{(1024!-2)2!} = \frac{1024!(1024!-1)\ \ldots\ (1024!-2+1)}{2!}$$

$\gg 1024!$ ($\gg$ is the symbol for much greater than)

The large key space prevents exhaustive search and cross-matching attacks.

Entropy, is expressed in Equation 21 as:

$$H = \log_2 N^K \tag{21}$$

where *N* is the symbol count and K is the key length (Shannon, 1948).

The value of key length is 2^{20} (or 1,048,576) bits

$\therefore H = \log_2 2^{1048576} = 1{,}048{,}576$ bits.

This is prohibitively large enough to prevent an attacker from carrying out a random guessing attack against the authentication system.

The probability of correct guess, $P_r\,(guess)$, measures the possibility that an attacker will guess a transformation key correctly. It is expressed as the inverse of the key space, K_{space} . That is,

$$P_r\,(guess) = {}^1\!/_{K_{space}} \tag{22}$$

$\therefore P_{guess} = {}^1\!/_{1024!} \ll 0$. The probability of guessing a transformation key is very low (much less than 0).

Theoretical analysis of irreversibility is used to determine the complexity of recovering an original biometric data from a transformed template and the transformation parameter. The analysis is presented as follows.

Consider a 4×4 general permutation matrix,

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

An elementary permutation matrix, p_a , is obtained by interchanging the first and second rows of *P*. That is,

$$p_a = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

The second elementary permutation matrix, p_b , is obtained by interchanging the third and fourth rows of P. That is,

$$p_b = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

The transformation matrix or key, is computed by the operation p_a *xor* p_b. Hence,

$$R = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Now consider a one-dimensional column vector, $x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. The transformed template Y is computed by the multiplication operation, $y = Rx$. That is,

$$y = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The non-invertible analysis is carried out based on two scenarios, namely, "with respect to *R* scenario" (that is, when an attacker knows the transformation key) and "with respect to without *R* scenario" (that is, when an attacker does not have access to the transformation key).

(a) Non-invertible analysis with respect to R scenario

$x = {}^{y}/_{R}$, where R is the transformation matrix and y is the transformed template.

$=R^{-1}y$

Recall that $R = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$,

hence $R^{-1} = \begin{bmatrix} inf & inf & inf & inf \\ inf & inf & inf & inf \\ inf & inf & inf & inf \\ inf & inf & inf & inf \end{bmatrix}$

The inverse of R is undefined because R is a singular matrix. That is, the determinant of R, $det(R) = 0$.

An alternative approach is to attempt to recover x from R and y using a system of simultaneous equations constructed from R and y (Li & Hu, 2013). Based on the example above, we attempt to retrieve the original biometric vector.

$x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ from the matrix $R = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$ and transformed template $y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

using a system of linear equations

$$0x_1 + 0x_2 + x_3 + x_4 = 1 \quad (23)$$

$$0x_1 + 0x_2 + x_3 + x_4 = 1 \quad (24)$$

$$x_1 + x_2 + 0x_3 + 0x_4 = 1 \quad (25)$$

$$x_1 + x_2 + 0x_3 + 0x_4 = 1 \quad (26)$$

The solutions for the equation will be the original biometric vector $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$

By eliminating the variables containing 0, we have

$$x_3 + x_4 = 1 \quad (27)$$

$$x_3 + x_4 = 1 \quad (28)$$

$$x_1 + x_2 = 1 \quad (29)$$

$$x_1 + x_2 = 1 \quad (30)$$

Using Equation 27 and 28,

$$x_3 = 1 - x_4 \quad \text{(from Equation 27)}$$

Substituting for x_3 in Equation 28, we have

$$1 - x_4 + x_4 = 1 \quad \text{or } 1 - 0 = 1$$

Moreover, $x_4 = 1 - x_3$ (from Equation 28)

Substituting for x_4 in Equation 27, we have

$1 - x_3 + x_3 = 1$ or $1 - 0 = 1$.

Using Equation 29 and 30,

$x_1 = 1 - x_2$ (from Equation 29)

Substituting for x_1 in equ. 30, we have

$1 - x_2 + x_2 = 1$ or $1 - 0 = 1$

Moreover, $x_2 = 1 - x_1$ (from Equation 30)

Substituting for x_2 in Equation 29, we have

$1 - x_1 + x1 = 1$ or $1 - 0 = 1$

The result in all cases is 1. But we do not know which variable has the value of 1. This makes it difficult to assign values to each x_i in $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$

Hence, it is impossible to retrieve *x* from *y* even if *R* is known.

(b) Non-invertible analysis with respect to without R scenario

In this case, only the transformed template $y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ is available.

A compromise of the authentication scheme will require two steps:

(i) An attacker will have to generate all possible values of *x*. Each *x* is an *n*×1 dimensional vector, where n is the number of bits in the compromised transformed template. This requires an effort of *n!*

(ii) Each *x* is transformed using each of the possible values of *R*.
This requires an effort of $n!_{C_2} = \frac{(n!)!}{(n!-2)2!} = \frac{n!(n!-1) \ldots (n!-2+1)}{2!}$
For *n!* vectors, the total effort required = $n!\left(\frac{(n!)!}{(n!-2)2!}\right)$

(iii) Each transformed *x* is compared with the compromised *y* resulting in an effort of $n!\left(\frac{(n!)!}{(n!-2)2!}\right)$.

(iv) A match between a transformed *x* and a compromised *y* indicates that both of them are generated from the same feature vector.

The complexity of this analysis is defined in terms of the total effort required to obtain all possible transformed values of *x* (exhaustive search) and to match each transformed *x* with a compromised *y*

$$= n!\left(\frac{(n!)!}{(n!-2)2!}\right) + n!\left(\frac{(n!)!}{(n!-2)2!}\right) = 2\left(n!\left(\frac{(n!)!}{(n!-2)2!}\right)\right)$$

Using this approach for a 4-bit template, we substitute $n = 4$.

The total effort required $2\left(4!\left(\frac{(4!)!}{(4!-2)2!}\right)\right) = 6.768\times10^{23}$ iterations.

Table 5 illustrates the relationship between the dimension of biometric data and complexity of an exhaustive search attack.

Table 5
Relationship between dimension of biometric data and complexity of an exhaustive search attack

Length of biometric template (n)	Complexity of operation (iterations)
4	6.768×10^{23}
5	6.80×10^{198}
6	overflow
⋮	⋮
1024	overflow

The complexity of irreversibility with respect to without *R* scenario increases as the dimension of biometric data increases. Matrix transformation uses a 1024-bit template, which makes it computationally infeasible to retrieve an original template from only its transformed version.

Privacy Analysis

The privacy capability of matrix transformation is analysed using the requirements proposed in the ISO/IEC JTC 1/SC 27 (BioKeySIII Final Report, 2011) standard. These include irreversibility, unlinkability, confidentiality and data minimisation.

Irreversibility analysis. With respect to matrix transformation, irreversibility is defined in terms of the difficulty an attacker faces in an attempt to recover original biometric data, from a transformed template, *Y*. Detailed analysis of irreversibility has been using both "with respect to *R* and without *R*" scenarios. The analysis shows that an attacker faces a computationally difficult task in an attempt to recover original biometric data from a transformed template.

Unlinkability analysis. This is used to determine the effort (computational complexity) required to match biometric references across multiple databases or applications. A pair of biometric data is 'linkable' if it is possible to establish that they are obtained from the same user. The analysis of unlinkability is carried out by considering both the average case and worst-case scenarios.

Average case. The average case involves matching a compromised template with only half of the possible instances of transformed templates. The effort required, E_{CM}

$$= {}^{n!_{C_2}}/_2 = \frac{(n!)!}{2(n!-2)2!} = \frac{n!\,(n!-1)\ \ldots\ (n!-2+1)}{2(2!)}$$

$$= \frac{n!\,(n!-1)\ \ldots\ (n!-2+1)}{4}$$

By substituting n = 1024, we have

$\frac{1024!(1024!-1)\ \ldots\ (1024!-2+1)}{4} \gg$ 1024! iterations

Worst case. This involves matching all possible instances of transformed templates. The effort required, E_{CM}

$$= n!_{C_2} = \frac{(n!)!}{(n!-2)2!} = \frac{n!\,(n!-1)\ \ldots\ (n!-2+1)}{2!}$$

$= \frac{1024!(1024!-1)\ \ldots\ (1024!-2+1)}{2!} \gg$ 1024! iterations.

The effort required when the entire transformed templates is matched is twice that which is required for matching only half of the transformed templates. However, the effort in both cases is in excess of 1024! iterations. This is prohibitively large enough to prevent a cross-matching attack in matrix transformation.

Confidentiality analysis. A secured template, y is created by applying a transformation matrix, on a biometric data, x. That is,

$$y = Rx$$

The transformed template is stored in the database while the original biometric data, , is discarded. The example above shows that the transformed template, $y = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$ is distinguishable from the original biometric data, $x = \begin{bmatrix}1\\0\\0\\1\end{bmatrix}$. In other words, the transformation operation uses R to conceal x in y.

Data minimisation. Matrix transformation stores only transformed templates, , and possibly the transformation key, R. Sensitive biometric data that may violate users' privacy are not retained in the authentication system during enrolment and authentication.

DISCUSSION OF RESULTS

Results in Tables 1, 2, 3 and 4 show that matrix transformation has low FRRs (between 0% and ~4.63%) when it is applied to face biometric than it does when applied to iris or multibiometric modality. However, this is achieved at the expense of intolerably high FARs (between 14.4% and 100%). This is due to the low inter-class distance among face images of different subjects (Wu & Yuan, 2010). High FARs also imply a high correlation among transformed templates of different users for the chosen hamming distance values. Low FRR and high FAR imply high usability (or user convenience) and low security. The application of matrix transformation on iris results in high FRRs of 64.35% and 32.02% for hamming distance values of 0.20 and 0.25, respectively. However, the FARs recorded for both hamming distance values is 0%. Increasing the hamming distance values to 0.30 and 0.35 results in improvements in recognition accuracy as the FRRs reduce to 8.796% and 1.466%, respectively. The FAR also increases from 0% to 0.007% and 0.139% for the respective hamming distance values. This implies that hamming distance values of 0.30 and 0.35 provides a good balance between security and user convenience in an iris-based matrix transformation scheme. Applying matrix transformation on multibiometric data results in high FRRs of 29.78%, 19.75% and ~10.42% for hamming distances of 0.20, 0.25 and 0.30, respectively. Increasing the hamming distance to 0.35 results in a lower FRR (~4.9%) and high FAR (~20.67%). Results from experiments show that the matrix transformation has better recognition accuracy when applied to iris biometric (especially at hamming distances of 0.30 and 0.35) than it does when it is applied to face or multibiometric data. The iris is a more reliable biometric modality because it exhibits low intra-class variation and high inter-class distance (Bowler et al., 2007). Matrix transformation also achieves improved performance results when it is applied to multibiometric modality than it does when face modality is used. This is because the presence of iris bits in multibiometric templates minimises the impact of the pervasiveness of the face on the recognition accuracy of the system. Security analysis shows that matrix transformation has long key length, large key space, high key entropy and low probability of correct guess of the key. Thus, matrix transformation is resistant to random guessing and exhaustive search attacks. Privacy analysis shows that matrix transformation provides irreversibility, confidentiality and data minimisation. The complexity of unlinkability analysis in matrix transformation is very high. This enables the approach to provide sufficient resistance against cross-matching attacks. Irreversibility prevents the recovery of original biometric data from transformed templates. Legitimate users are protected against loss of identity as attackers cannot reconstruct actual biometric images from compromised biometric templates.

A comparison of the recognition accuracy of the proposed matrix transformation and related works is presented in Table 6.

Table 6
Comparison between our approach and previous studies (recognition accuracy)

Author	Technique	Dataset	Performance (%)			
			FRR	GAR	FAR	EER
Moujahdi et al., 2012	Spiral cube	Yale face	100		0	
Zuo et al., 2008	BIN combo			100	0.1	
	GREY combo	MMU1 iris data set		99.5	0.01	
Our approach	Matrix transformation		10		5.5	
Hammerle-Uhl et al., 2009	Block re-mapping					1.2
	Image warping					1.3
Rathgeb et al., 2010	User-specific permutation	CASIA iris V3	3.821		0	
Rathgeb et al., 2014	Alignment-free adaptive Bloom filter		2.05		0.01	
Our approach	Matrix transformation		7.889		2.74	

Results from previous works show that the proposed approach has lower recognition accuracy than the previous works. The alteration of the bits in the original template makes matching difficult in the transformation domain. Hamming distance provides limited error correction capability because it corrects only bit errors and does not address burst errors. The long dimension of biometric data (1024 bits) used by matrix transformation imposes a high overhead on the error correcting capability of hamming distance and this results in performance degradation. However, it is pertinent to mention that the main goal of matrix transformation is to provide high template security and user privacy.

Table 7 presents a comparison of the security of the proposed matrix transformation and related works.

Table 7
Comparison between our approach and previous studies (security)

Author	Modality	Technique	Security			
			Key length (bits)	Key Space	Entropy (bits)	Pr (correct guess)
Zuo et al, 2008	Iris image	Grey-combo	260	2^{260}	260	5.39×10^{-79}
	Binary iris code	Bin-combo	560	2^{560}	560	0
Rathgeb & Uhl, 2010	Binary iris code	User-specific permutation		38!		1.91×10^{-45}
Moujahdi et al., 2012	Face	Spiral cube		100^{200}		$\ll 0$
Sandhya et al., 2016	Fingerprint	Delaunay		2.5 billion/		
		triangle		8.35 billion		
Our approach	Face, iris and multibiometric	Matrix transformation	2^{20} or 1,048,576	1024!	2^{20}	$\ll 0$

The proposed scheme has much higher key length, larger key space and higher entropy than the previous approaches. Its keys also have a much lower probability of correct guess compared to the previous schemes. Security and privacy analysis in Sections 5.7 and 5.8 show that the level of template security and user privacy provided by a biometric cryptosystem depends largely on its key length and key space. The proposed approach has a much higher key length and larger key space compared to the previous schemes. Hence, it is more robust to security and privacy attacks than existing approaches.

CONCLUSION

This paper proposed and applied the matrix transformation technique to three different biometric modalities (face, iris and multibiometric) unlike previous approaches that used only one biometric modality. The study highlights the effect of the nature of biometric data on the performance accuracy of the proposed approach. Matrix transformation provides high-level template security and user privacy compared to other approaches. The low recognition accuracy of the proposed multibiometric template protection scheme can be addressed by integrating a key binding technique that uses better error correction techniques; This will result in a hybrid multibiometric template protection scheme that provides improved template security and user privacy without compromising recognition accuracy.

Results from previous works show that the proposed approach has lower recognition accuracy than the previous works. The alteration of the bits in the original template makes matching difficult in the transformation domain. Moreover, Hamming distance provides limited error correction capability. That is, it corrects only bit errors and does not address burst errors. The long dimension of biometric data (1024 bits) used by matrix transformation imposes a high overhead on the error correcting capability of Hamming distance and this results in performance degradation. However, it is pertinent to mention that the main goal of matrix transformation is to provide high template security and user privacy.

ACKNOWLEDGEMENT

This material is based upon work supported by the Ministry of Higher Education Malaysia under Grant No. FRGS 08-01-15-1721FR.

REFERENCES

Boult, T. (2006). Robust distance measure for face recognition supporting revocable biometric tokens. In *Proceedings of 7th International Conference on Automatic Face and Gesture Recognition* (pp. 560–566). New York, NY, USA: IEEE.

Boult, T., Scheirer, W., & Woodworth, R. (2007). Revocable fingerprint biotokens: Accuracy and security analysis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. New York, NY, USA: IEEE. Doi: 10.1109/CVPR.2007.303110.

Bowler, K., Hollingsworth, K., & Flynn P. (2007). Image watermaking for iris: A survey. *Computer Vision and Image Understanding, 110*(2), 281–307.

BIT. (2010). *CASIA Iris Image Database Version 2.0*. Biometrics Ideal Test. Retrieved from http://biometrics.idealtest.org/

Campisi, P., & Egiazarian, K. (2007). *Blind image deconvolution: Theory and applications*. Raton, Florida, USA: Boca CRC Press.

Choi, J. Y., Ro, Y. M., & Plataniotis, K. N. (2012). Color local texture features for color face recognition. *IEEE Transactions on Image Processing, 21*(3), 1366–1380.

Farberbock, P., Hammerle-Uhl, J., Kaaser, E., Pschernig, D., & Uhl, A. (2010). Transforming rectangular and polar iris images to enable cancelable biometrics. In A. Camilho & M. Kamel (Eds.), *Image analysis and recognition – Lecture notes in computer science* (Vol. 6112, pp. 276–286). Berlin Heidelberg: Springer.

Feng, Y. C., & Yuen, P. C. (2012). Vulnerabilities in binary face template. *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop* (pp. 105–110). New York, NY, USA: IEEE.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America 4*(12), 2379–2394.

Grassi, M., & Faundez-Zanuy, M. (2009). Protecting DCT templates for a face verification system by means of pseudo-random permutations. In J. Cabestany, F. Sandoval, A. Prieto & J. M. Corchado (Eds.), *Bio-Inspired systems: Computational and ambient intelligence – Lecture notes in computer science* (Vol. 5517, pp. 1216–1223). Heildelberg, Berlin: Springer.

Hammerle-Uhl J., Pschernig E., & Uhl, A. (2009). Cancelable iris biometrics using block re-mapping and image warping. In P. Samarati (Ed.), *Information security – Lecture notes in computer science* (Vol. 5735, pp. 135–142). Heidelberg, Berlin: Springer.

Hamouchene, I., & Aouat, S. (2014, August). A cognitive approach for texture analysis using neighbors-based binary patterns. In *Cognitive Informatics and Cognitive Computing (CICC), 2014 IEEE 13th International Conference on* (pp. 94-99). IEEE.

Hao F., Anderson, R., & Daugman J. (2006). Combining cryptography with biometrics effectively. IEEE *Transactions on Computers, 55*(9), 1081–1088.

Hariprasath, S., & Prabakar, T. N. (2012). Multimodal biometric recognition using iris feature extraction and palmprint features. In *IEEE International Conference on Advances in Engineering, Science and Management* (pp. 174–179). New York, NY, USA: IEEE.

He, Y., Yap, K. H., Chen, L., & Chau, L. P. (2008). A novel hybrid model framework to blind color image deconvolution. *IEEE Transactions on System, Man and Cybernetics – Part A: Systems and Humans, 38*(4), 867–880.

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A survey. In M. Kawulok, M. Emre Celebi, & B. Smolka (Eds.), *Advances in face detection and facial image analysis* (pp. 189–298). New York, NY: Springer International Publishing.

Jeong, M. Y., & Teoh, A. B. J. (2010). Cancellable face biometrics system by combining independent component analysis coefficients. In H. Sako, K. Franke, & S. Saitoh (Eds.), *Computaional forensics – Lecture notes in computer science* (Vol. 6540, pp.78–87). Berlin, Heidelberg: Springer.

Karmakar, D., & Murphy, C. A. (2014). Generation of new points for training set and feature – Level fusion in multimodal biometric identification. *Machine Vision and Applications, 25*, 477–487.

Kaur, G. M. (2013). Multimodal-Based fuzzy vault using iris, retina and fingervein. In *Fourth International Conference on Computing, Communications and Networking Technologies*. New York, NY, USA: IEEE. doi: 10.1109/ICCCNT.2013.6726786.

Li, C., & Hu, J. (2013). Attacks via record multiplicity on cancelable biometric templates. *Concurrency and Computation: Practice and Experience, 26*(8), 1593-1605. doi: 10.1002/cpe.3042.

Li, S., & Cot, A. C. (2011). Attacks using reconstructed fingerprint. In *2011 IEEE International Workshop on Information Forensics and Security*. New York, NY, USA: IEEE. Doi: 10.1109/WIFS.2011.6123151.

Li, S. Z., Chu, R. F., Liao, S. C., & Zhang, L. (2007). Illumination invariant face recognition using near-infrared images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (Special issue on Biometrics: Progress and Directions), 29*(4), 627–639.

Masek, L. (2003). *Recognition of human iris for biometric identification*. (Upublished MSc Thesis). University of Western Australia, Australia.

Moujahdi, C., Ghouzali, S., Mikram, M., Rziza, M., & Bebis, G. (2012, June). Spiral cube for biometric template protection. In *International Conference on Image and Signal Processing* (pp. 235-244). Heildelberg, Berlin: Springer.

Nishino, T., Kajikawa, Y., & Muneyasu, M. (2012). Multimodal person authentication system using features of utterance. In *2012 IEEE International Symposium on Intelligent signal Processing and Communication Systems* (pp. 43–47). New York, NY, USA: IEEE.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(7), 971–987.

Ouda, O., Tusamura, N., & Nakaguchi, T. (2011). Securing bioencoded iriscodes against correlation attacks. In *2011 IEEE International Conference on Communications*. New York, NY, USA: IEEE. doi: 10.1109/icc2011.5963247.

Plesca, C., & Morogan, L. (2014). Efficient and robust perceptual hashing using log-polar image representation. In *2014 10th International Conference on Communications*. New York, NY, USA: IEEE. doi: 10.1109/ICComm.2014.6866755.

Ratha, N. K., Connell, J. H., & Bolle, R. M. (2001). Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal, 40*(3), 614–634.

Ratha, N. K., Connell, J. H., Bolle, R. M., & Chikkerur, S. (2006). Cancellable biometrics: A case study in fingerprints. In *Proceedings of 18th International Conference on Pattern Recognition* (pp. 370–373). New York, NY, USA: IEEE.

Ratha, N. K., Connell, J. H., Bolle, R. M., & Chikkerur, S. (2007). Generating cancellable fingerprints templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(4): 561–572.

Rathgeb, C., & Uhl, A. (2010). Secure iris recognition based on local intensity variations. In A. Campilho & M. Kamel (Eds.), *Image analysis and recognition – Lecture notes in computer science* (Vol. 6112, pp. 266–275). Heildelberg, Berlin: Springer.

Rathgeb, C., Breitinger, F., Busch, C., & Baier, H. (2014). On the application of bloom filters to iris biometrics. *IET Biometrics, 3*(4), 207–218.

Safie, S. I., Nurfazira, H., Azavitra, Z., Soraghan, J. J., & Petropoulakis, L. (2014). Pulse active transformation: A non-invertible transformation with application to ECG biometric Authentication. In *2014 IEEE Region 10 Symposium* (pp. 667–671). New York, NY, USA: IEEE.

Sandhya, M., Prasad, M. V. N. K., & Chillarige, R. R. (2016). Generating cancellable fingerprint templates based on Delaunay triangle feature set construction. *IET Biometrics, 5*(2), 131–139.

Shannon, C. E. (1948) A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.

Study of the Privacy and Accuracy of the Fuzzy Commitment Scheme. (2011). BioKeyS III-final report. *Bundesamt für Sicherheit in der Informationstechnik 2011*. Retrieved from http://www.bsi.bund.de

Teoh, A. B. J., Yip, W. K., & Toh, K. A. (2010). Cancellable biometrics and user-dependent discretization in biohash. *Pattern Analysis and Applications, 13*(3), 301–307.

Tharwat, A., Ibrahim, A. F., & Ali, H. A. (2012). Multimodal biometric authentication algorithm using ear and finger knuckle images. In *2012 7th International Conference on Computer Engineering and Systems,* Cairo (pp. 176–179). New York, NY, USA: IEEE.

Wang, S., & Hu, J. (2013). A Hadamard transformed-based method for the design of cancellable fingerprint templates. In *2013 6th International Congress on Image and Signal Processing* (pp. 1682–1687). New York, NY, USA: IEEE.

Wu, L., & Yuan, S. (2010). A face based fuzzy vault scheme for secure online authentication. In *2010 Second International Symposium on Data, Privacy and E-Commerce* (pp. 45–49). New York, NY, USA: IEEE.

Xiuyan, L., Changyun, M., Tiegen, L., & Chenhu, Y. (2011). Research on personal identity verification based on hand vein, iris and fingerprint. In *2011 International Symposium on Computer Science and Society* (pp. 16–19). New York, NY, USA: IEEE.

Zuo, J., Ratha, K., & Connell, J. H. (2008). Cancelable iris biometric. In *Proceedings of 19th International Conference on Pattern Recognition*. New York, NY, USA: IEEE. Doi: 10.1109/ICPR.2008.4761886.

Pertanika J. Sci. & Technol. 26 (1): 161 - 176 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

The Effects of Alkali Treatment on the Mechanical and Chemical Properties of Banana Fibre and Adhesion to Epoxy Resin

Zin, M. H.[1,2]*, Abdan, K.[3], Norizan, M. N.[3] and Mazlan, N.[1,4]

[1]*Institute of Tropical Forestry and Forest Products, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[2]*Aerospace Malaysia Innovation Centre, German Malaysian Institute, Jalan Ilmiah, Taman Universiti, 43000 Bangi, Selangor, Malaysia*
[3]*Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[4]*Department of Aerospace Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

The main focus of this study was to obtain the optimum alkaline treatment for banana fibre and the its effect on the mechanical and chemical properties of banana fibre, its surface topography, its heat resistivity, as well as its interfacial bonding with epoxy matrix. Banana fibre was treated with sodium hydroxide (NaOH) under various treatment conditions. The treated fibres were characterised using FTIR spectroscopy. The morphology of a single fibre observed under a Digital Image Analyser indicated slight reduction in fibre diameter with increasing NaOH concentration. The Scanning Electron Microscope (SEM) results showed the deteriorating effect of alkali, which can be seen from the removal of impurities and increment in surface roughness. The mechanical analysis indicates that 6% NaOH treatment with a two-hour immersion time gave the highest tensile strength. The adhesion between single fibre and epoxy resin was analysed through the micro-droplet test. It was found that 6% NaOH treatment with a two-hour immersion yielded the highest interfacial shear stress of 3.96 MPa. The TGA analysis implies that alkaline treatment improved the thermal and heat resistivity of the fibre.

Keywords: Alkaline treatment, banana fibre, biocomposites, interfacial shear stress, natural fibre

ARTICLE INFO

Article history:
Received: 24 August 2016
Accepted: 10 August 2017

E-mail addresses:
hanafeezin@gmail.com (Zin, M. H.),
khalina.upm@gmail.com (Abdan, K.),
zaidharithah66@yahoo.com (Norizan, M. N.),
norkhairunnisa.mazlan@gmail.com (Mazlan, N.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

The usage of composite material has been found to increase significantly in the manufacture of aircraft parts, from 5% in the 90s to 50% in 2010 (Gent et al., 2010). The composite parts are mainly made of petroleum derivative products, such as carbon fibre.

However, most of the synthetic fibres are costly and non-biodegradable and they are facing depletion. The rise of global awareness of environmental issues has attracted researchers in various areas to develop renewable materials based on sustainability principles (Mahjoub et al., 2014). A huge shift in the usage of natural fibres to produce composites worldwide has been reported (John et al., 2008). Natural fibre reinforced composites (NFRC) have been used in construction and automotive applications for a while, and are now becoming more substantial as secondary structures for next-generation aircraft (Koronis et al., 2013). The advantages of using natural fibre composites are environmental gain, reduced energy consumption and reduced dependency on petroleum-based materials; in addition the material is light weight and provides improved insulation and sound absorption properties (Joseph et al., 2002). Natural fibres such as jute, sisal, hemp, kenaf, banana and pineapple leaf (PALF) are renewable and non-abrasive and they can be incinerated for energy recovery. They possess good calorific values and present low safety risk during handling. They also exhibit excellent mechanical properties, have low density and are inexpensive (Boopalan et al., 2013).

Numerous research and studies have been carried out on natural fibre reinforced composites (NFRC) in recent years (Alavudeen et al., 2015; John et al., 2008; Sanjay et al., 2015). To develop NFRC, it is vital to understand the chemical composition and the surface adhesive bonding properties of natural fibre. The components of natural fibre include cellulose, hemicellulose, lignin, pectin, waxes and water-soluble substances (Li et al., 2007). Since natural fibre is hydrophilic in nature, chemical modification of the fibre is required in order to improve the interfacial properties between the fibre and any polymer matrix (Asim et al., 2015). Various chemical treatments for natural fibre, namely alkaline, silane, acetylation, benzoylation, acrylation, maleated coupling agent, permanganate and peroxide treatment have been studied (Li et al., 2007). Alkaline treatment is one of the most commonly used chemical treatments for reinforcing thermoplastics or thermosets. Alkaline treatment increases surface roughness and increases the amount of exposed cellulose on the fibre surface by removing a certain amount of the cellulosic content that covers the external surface of the fibre cell wall (Gurunathan et al., 2015) to produce better mechanical interlocking. Atiqah et al. (2014) treated kenaf fibre with 6% sodium hydroxide (NaOH) solution for 3 hours and recorded optimum results for flexural, tensile and impact strengths. Merlini et al. (2011) attempted alkaline treatment of short banana fibre with 10% NaOH solution for 1 hour and Panyasart et al. (2014) treated pineapple leaf fibre (PALF) with 5% NaOH solution and a 5-hour immersion period at room temperature. The results obtained from these alkaline treatments showed superior behaviour in mechanical properties compared to the untreated fibre. The focus of this study was to obtain the optimum alkaline treatment for banana fibre and its effect on the mechanical and chemical properties, surface topography, heat resistivity, as well as interfacial bonding with epoxy matrix.

EXPERIMENTAL DETAILS

Material

For this research investigation, loose pseudo stem banana fibre was obtained from local farmers in Kuala Langat, Malaysia. The common mechanical properties of banana fibre are shown in Table 1.

Table 1
Mechanical properties of banana fibre (Li et al., 2007; Merlini et al., 2011; Venkateshwaran et al., 2012)

Properties	Value
Density (g/cm^3)	1.35
Tensile Strength (MPa)	161.8 ± 11.8
Young's Modulus (GPa)	8.5 ± 0.9
Elongation at break (%)	2.0 ± 0.4
Diameter (μm)	120 ± 5.8

For the resin system, epoxy resin DM15 made of Bisphenol A was chosen as it is widely used in the automotive and aerospace industries and has moderate viscosity, low molecular weight and high mechanical strength. The resin consists of part A (base) and part B (hardener). The formulation of the resin is as shown below:

$$\text{DM15 (Part A)}: \text{DM15 (Part B)} = 5:1 \qquad \text{Equation 1}$$

Chemical Composition of Banana Fibre

Figure 1 shows strands of untreated banana fibre in loose form. The strands of fibre were cut into 400 mm length and separated into single filaments. The type of banana used cannot be revealed due to the confidential nature of the information.

Figure 1. Untreated banana fibre in loose form

However, the common chemical composition of banana fibre is cellulose, hemi-cellulose, lignin and other components as shown in Table 2.

Table 2
Chemical composition of banana fibre (Arthanarieswaran et al., 2014)

Fibre Type	Cellulose %	Hemi Cellulose %	Lignin %	Moisture Content %
Banana	62	18	5	11

Chemical Treatment

The strands of banana fibre were treated with sodium hydroxide (NaOH) of different combinations of concentration and duration, as presented in Table 3. In order to fill the gap in the findings of a previous study by Venkateshwaran et al. (2013) and to determine the optimum treatment from a smaller range that meets with industrial requirement, 4% (w/v) to 8% NaOH (w/v) concentration with a 2- and 4-hour immersion time was selected. In addition, this study used purely single strand pseudo stem banana fibre unlike previous research, which used banana/epoxy fibre. The NaOH concentration and immersion period are shown in Table 3.

Table 3
Banana fibre treatment using NaOH solution

Sample	NaOH %	Duration (Hours)
Untreated	0	0
4% 2h	4	2
4% 4h	4	4
6% 2h	6	2
6% 4h	6	4
8% 2h	8	2
8% 4h	8	4

The alkaline treatment method was adopted from a previous study by Mahjoub et al. (2014). In short, sodium hydroxide pellets were weighed according to the designated concentration; 40g NaOH was dissolved in 1 litre of distilled water to make 4% NaOH solution. Strands of banana fibre were cut into approximately 400-mm length and immersed in alkaline solution according to the period specified for each designated concentration. Each series was labelled with the group name and the type of treatment condition. Once the immersion time ended, the fibres were washed thoroughly with distilled water and pH paper was used to check the alkalinity of the treated strands of fibre. The pH paper was immersed in the solution in which the treated fibre strands were soaked and rinsed with distilled water. The fibres were washed until the neutral value of pH 7 was obtained. Finally, the strands were dried in an oven at 60°C for 24 hours.

Physical Analysis

The diameter of a single strand of banana fibre was measured using an Olympus SZX 12-CCD Digital Image Analyser at 20-times magnification. For each fibre strand, the diameter was

measured at five different points, and the average value was taken. For each combination of NaOH treatment, an average of 10 single fibre samples were obtained for recording.

Morphology Analysis

The surface microstructure of untreated and treated fibre strands was observed using a Hitachi S-3400N Scanning Electron Microscope (SEM) set at 5.0 kV and a magnification between 300 and 800 times.

Functional Chemical Group Analysis

Fourier Transformed Infrared Spectroscopy (FTIR) was performed on a Perkin Elmer Spectrum 100 instrument with a resolution setting of 4 cm^{-1} to identify the functional groups in untreated and treated fibres.

Mechanical Analysis

A single-fibre tensile test was conducted using an Instron 3365 Dual Column Table Top Universal Testing Systems with a 5 kN maximum load at a rate of 2 mm/min per ASTM D3039. A single strand of banana fibre was attached to a paper holder, as shown in Figure 2.

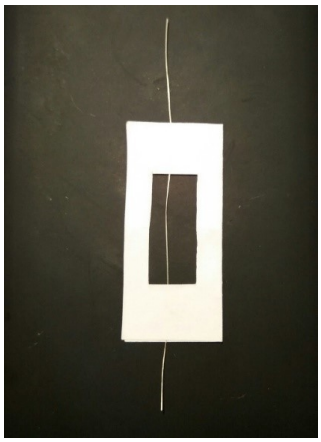


Figure 2. Banana fibre filament attached to a paper holder

The sample was then attached to the tensile machine gripper for testing. For each combination of NaOH treatment, the average based on 10 single fibre samples was recorded.

Adhesion Analysis

The micro-droplet test was conducted to evaluate the interfacial shear strength between the epoxy resin and the banana fibre using a technique adopted from a previous researcher, Dai et al. (2011). The resin was mixed according to a 5:1 ratio and then stirred for a few minutes. The resin was wetted onto the single banana fibre strand that had been fixed to a paper holder

to form a micro-droplet surrounding the fibre diameter due to surface tension. The embedded length of the micro-droplet was measured using a digital image analyser. The force required to de-bond the solid resin droplet from the fibre while the loading blade held the droplet was recorded. The schematic of the micro-droplet test is illustrated in Figure 3.

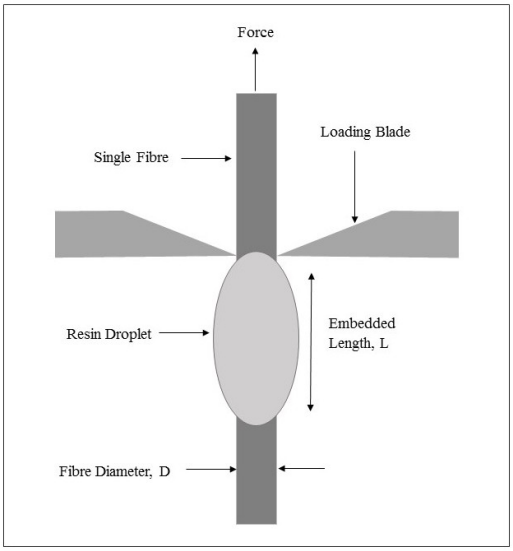


Figure 3. Schematic of the micro-droplet test

The figure shows vertical arrangement of the micro-droplet test. The top part of the fibre was clamped and pulled upwards while the bottom part of fibre strand was left unclamped. While the fibre was pulled upwards, the resin droplet was held by the loading blade, resisting the clamping force direction. The resisting force was recorded until the fibre breakage point. The interfacial shear strength (IFSS) was calculated based on:

$$\tau = \frac{F}{\pi DL} \qquad \text{Equation 2}$$

where F is the maximum load, D is the fibre diameter and L is the embedded length. The method was adopted from previous work by Dai et al. (2011). For each treatment condition, 10 samples were tested and recorded. The tensile test was carried out on the droplet samples using the Instron 3365 Dual Column Table Top Universal Testing Systems with 5 kN maximum load at a rate of 0.1 mm/min.

Thermal Stability Analysis

A thermogravimetric analysis (TGA) was performed to evaluate the thermal properties of banana fibre using the TA Instrument TGA Q500 with a temperature setting between 30°C

and 600°C at a rate of 10°C/min. Thermogravimetric analysis measures the amount and the changing rate of material weight as a function of time or temperature in a controlled atmosphere (Boopalan et al., 2013). TGA can also be used to characterise the effect of decomposition, oxidation and dehydration on the material's weight loss or gain.

RESULTS AND DISCUSSION

Physical Analysis

The average diameter (10 samples for each combination) for all treatment conditions is shown in Table 4.

Table 4
Average diameter of untreated and NaOH treated banana fibre strands

Sample	Average Diameter (mm)
Untreated	0.11 ± 0.01
4% 2h	0.10 ± 0.01
4% 4h	0.10 ± 0.006
6% 2h	0.11 ± 0.007
6% 4h	0.11 ± 0.008
8% 2h	0.09 ± 0.006
8% 4h	0.08 ± 0.005

In general, the average diameter showed slight variation in value across different concentrations of NaOH. The diameter of the untreated single fibre was 0.11 mm, which was close to the finding reported by another researcher, Idicula et al. (2005). The fibre diameter started showing reduction beyond 6% NaOH concentration. At the highest NaOH concentration of 8% and with a 4-hour immersion period, the diameter became reduced by 33% compared to the untreated fibre diameter. This could be due to the fact that higher alkaline concentration increases the delignification of banana fibre, thus reducing the diameter of the fibre, as reported by Gurunathan et al. (2015). The fibre delignification is illustrated in the FTIR Spectrography shown in this paper.

Morphology Analysis

The effect of alkaline treatment can be explained by considering SEM micrographs presented in Figures 4a, 4b and 4c.

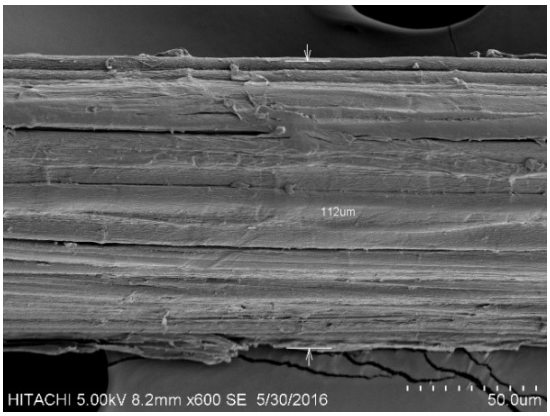


Figure 4a. SEM image of untreated banana fibre

Figure 4b. SEM image of banana fibre treated with 6% NaOH for 2 hours

Figure 4a shows considerable deposits of impurities and natural components on the surface of the untreated fibre. The deposits led to poor fibre-matrix interfacial adhesion by reducing the interfacial area of contact between the matrix and the fibre. This result was in agreement with previous work by Asumani et al. (2012). Figure 4b shows a rougher surface of treated banana fibre, as well as the removed impurities. Damage to the fibre is noticeable at a few locations. Removed impurities, as well as increased surface contact area, contribute to better surface interlocking between the fibre and epoxy resin, as also found by Mahjoub et al. (2014).

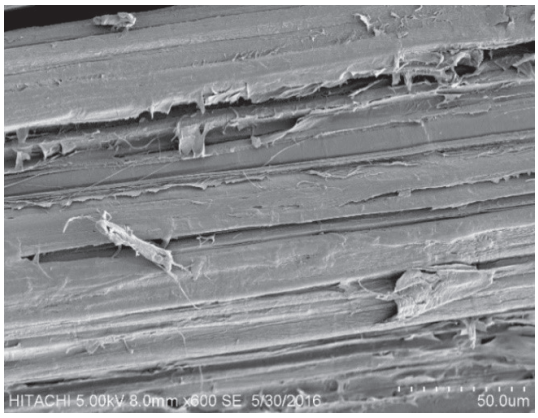


Figure 4c. SEM image of banana fibre treated with 8% NaOH 4 hours

However, as shown in Figure 4c, as the NaOH concentration reached 8%, the fibre damage was more severe, which in turn, reduced the strength of the banana fibre. This was in line with the tensile testing results as explained earlier.

Functional Chemical Group Analysis

The FTIR spectrum of untreated and treated banana fibres, which reflects the lignocellulosic components of the fibres, is illustrated in Figure 5.

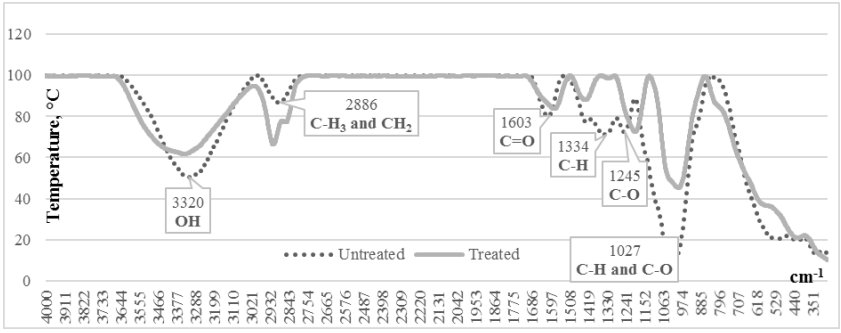


Figure 5. FTIR elements of treated and untreated fibres

The wide band at 3320 cm^{-1} indicates vibrational stretching of the hydroxyl group (OH) for natural fibre, as stated by Merlini et al. (2011). The treated fibre showed reduced hydroxyl stretching, which is responsible for the hydrophilic character of the fibre as also mentioned by Benítez et al. (2013) in their study. The sharp band at 2886 cm^{-1} demonstrates C-H stretching of methyl and methylene (alkene) groups. A similar pattern was reported by Corrales et al. (2007). The stretch at band 1603 cm^{-1} implies the presence of associated carbonyls (C=O) in the lignin, which was apparent in untreated banana fibre. This trend was also found by Asim et al. (2015).

The absorption reduced with alkaline treatment. The presence of C-H group with the frequency of vibration interval between 1535 cm^{-1} and 1330 cm^{-1} indicates the characteristic of methoxy radical in lignin, as also reported by Benítez et al. (2013). The reduced stretching within the region demonstrates that alkali treatment removed hemicelluloses and lignin from the surface of the natural fibre, as also reported for other vegetable fibre, as confirmed by Sgriccia et al. (2008). Meanwhile, the absorption at band 1245 cm^{-1} and 1027 cm^{-1} indicated C-O stretching of the acetyl group in hemicellulose, which reduced with alkaline treatment of NaOH. This result confirmed the finding by a previous team of researchers, Guimares et al. (2009).

Mechanical Analysis

Figure 4a presents the average ultimate tensile strength (UTS) and the tensile strain of the alkaline treated and untreated single banana fibre strands.

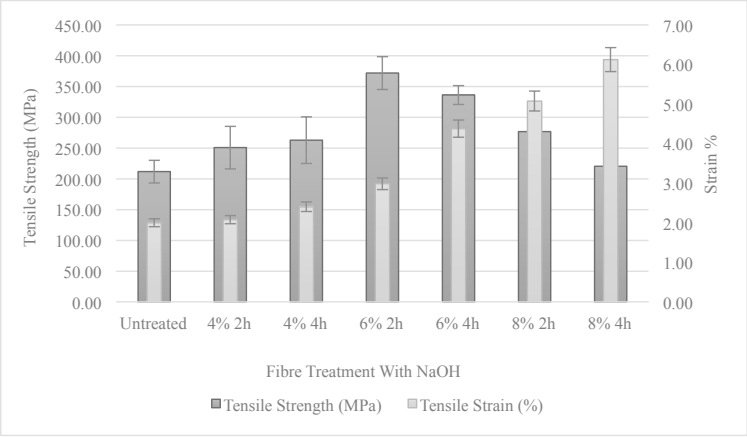


Figure 6a. Tensile strength and strain of untreated and treated single banana fibres

Based on the histogram, the tensile strength of the untreated fibre was 212 MPa, which is 30% higher compared to that reported in a previous study by Merlini et al. (2011). This could be due to various factors such as species, variety, type of soil used, plant age and weather factors, as mentioned by Rowell et al. (2000). The trend shows improved tensile properties for alkaline-treated banana fibre up to a certain point, followed by deteriorating tensile properties with increasing NaOH concentration and immersion time. The improvement in tensile properties was most noticeable for the fibre treated with 6% NaOH for 2 and 4 hours. The highest tensile strength recorded for a single strand of banana fibre was 371 MPa when treated with 6% NaOH with a 2-hour immersion time, which is equivalent to a 75% increment compared to that of untreated fibre. From the result, the tensile strength decreased with higher NaOH concentrations and longer immersion periods. This is because excessive delignification of natural fibre occurs in higher alkaline concentrations, resulting in weaker or damaged fibre. The result is in agreement with previous work by Asumani et al. (2012). This can be proven from the morphology as illustrated in Figure 4c.

Increased tensile strain was noticed for alkaline-treated samples. The increasing trend implies that alkaline treatment improved the fibre's ductility due to the removal of impurities, such as lignin and pectin, as found by a previous researcher, Gu (2009), hence resulting in more flexible fibre.

Figure 6b presents the tensile modulus of a single strand of banana fibre subjected to various alkaline treatments.

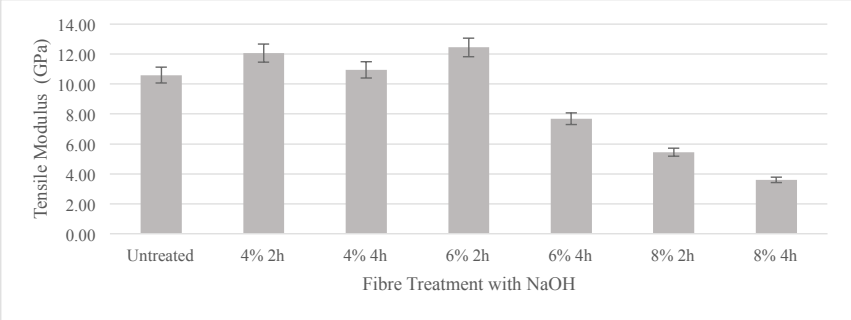


Figure 6b. Tensile modulus of untreated and treated single banana fibres

The variation in tensile modulus with alkaline concentration is similar to that of tensile strength, as presented in Figure 6a. In general, it can be observed that alkaline treatment improved the tensile modulus property of a single strand of banana fibre compared to that of untreated fibre. This was most noticeable for the fibre treated with 4% and 6% NaOH. The highest tensile modulus was 12.49 GPa, which was treated with 6% NaOH and given a 2-hour immersion time. The value was 53% higher than that of untreated fibre, with 8.09 GPa. Alkaline treatment may enhance fibre stiffness and expedite rearrangement of fibrils along the direction of the tension force after eliminating binding materials, as reported by Lopattananon et al. (2008). However, at 8% NaOH treatment, the tensile modulus reduced significantly, indicating decreased tensile strength due to the degradation of banana fibre, as also mentioned by Meon et al. (2012).

Adhesion Analysis

The interfacial shear strength is shown in Table 5.

Table 5
Interfacial shear strength (IFSS) for treated and untreated fibre

Sample	IFSS (MPa)
Untreated	1.12 ± 0.03
4% 2h	2.87 ± 0.02
4% 4h	3.74 ± 0.03
6% 2h	3.96 ± 0.03
6% 4h	3.85 ± 0.02
8% 2h	3.24 ± 0.02
8% 4h	2.51 ± 0.03

Looking at the trend, the interfacial shear stress increased with alkaline treatment to a certain value, then decreased with higher NaOH concentration. The interfacial shear strength of untreated fibre was 1.12 MPa, while 6% NaOH and a 2-hour immersion period treatment yielded the highest interfacial shear strength for treated fibre at 3.96 MPa. The same treatment condition also recorded the highest tensile strength for the treated single fibre strand, as illustrated in Figure 6a, which could be due to increased surface roughness, leading to better interlocking adhesion and greater amount of exposed cellulose on the fibre surface. The results obtained are in agreement with those obtained by previous researchers, Mostafa et al. (2015). This condition allowed better fibre wetting through increased number of possible reaction sites. Alkaline treatment also improved the surface adhesion properties of the fibre through the removal of natural and artificial impurities, and also created rougher surface topography. A similar trend was seen in Anuar et al. (2008). In this study, this was shown in the surface morphology result in Figure 4b and the FTIR spectroscopy results in Figure 5. Further treatment containing higher concentration of NaOH reduced the tensile strength due to excessive delignification, causing damaged and weaker fibre strands. The outcome was in accordance with that reported by Mishra et al. (2003).

Thermal Stability Analysis

Figure 7a illustrates the thermogravimetric analysis (TGA) of combustion residue for treated and untreated single banana fibre strands.

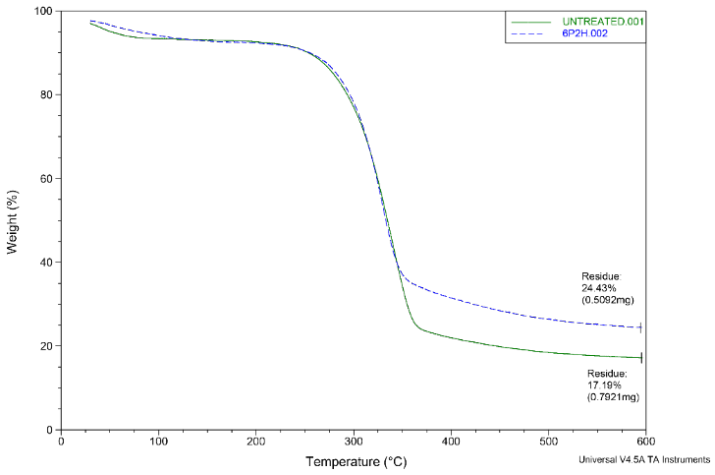


Figure 7a. TGA combustion residue for treated and untreated fibres

An initial weight loss of about 9.5% was recorded at a low temperature of 130°C due to the removal of moisture from untreated fibre. A similar pattern was found by Shih et al. (2014). The greatest weight loss occurred after 290°C with 75.61% weight reduction in untreated banana fibre, corresponding to component degradation and decomposition such as cellulose,

hemicellulose and lignin content in the fibre. The result is in line with the work of Zainudin et al. (2009). The starting temperature of weight loss for treated fibre shifted to a temperature beyond 290°C. This indicates that alkaline treatment resulted in higher thermal stability for the fibre, as also reported by Nopparut et al. (2016). For treated banana fibre, moisture was released from the fibre between 40°C and 175°C, and further degradation of cellulosic substances occurred between 175°C and 375°C. Nevertheless, subsequent decomposition requires elevated temperature as residues form and accumulate during degradation. The final residue degradation was recorded at 600°C. The higher percentage of residue for treated fibre (24.43%) showed that less fibre was burnt due to higher heat resistivity. This result is in line with the findings by Benítez et al. (2013).

The combustion temperature for treated and untreated fibre is shown in Figure 7b.

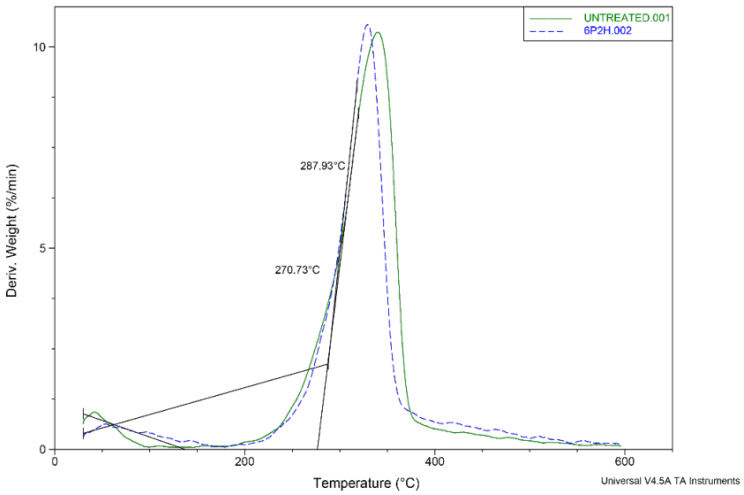


Figure 7b. TGA combustion temperature for treated and untreated fibres

The result showed that the decomposition temperature for untreated fibre at different weight-loss levels was lower than those of the alkaline-treated fibre. The burning of treated fibre occurred at 287.93°C, which was about 17°C higher than that required for untreated fibre. This indicated that NaOH treatment had slightly increased the fibre's heat resistivity due to the presence of inorganic materials such as sodium (Na) in the fibre, which requires a higher decomposition temperature, as also reported by Parker (2000).

CONCLUSION

This study showed that the optimum alkaline treatment for banana fibre is 6% NaOH concentration with a 2-hour immersion period, which resulted in 371 MPa tensile strength, 12.45 GPa tensile modulus and 3.96 MPa interfacial shear strength. The tensile strain increases with higher NaOH concentration. As the concentration increases beyond 6%, the mechanical properties of banana fibre deteriorate significantly. The removal of impurities and natural

deposits was evident through FTIR spectroscopy. The SEM micrographs showed that alkali treatment increased the surface roughness of the fibre, while the TGA results showed that the treatment enhanced the heat resistivity of the fibre.

ACKNOWLEDGEMENT

This study was conducted in a collaborative laboratory known as the Sustainable Aviation, a joint lab shared by Aerospace Malaysia Innovation Centre, Universiti Putra Malaysia and University of Nottingham Malaysia Campus, under Project no. BF-P06.

REFERENCES

Alavudeen, A., Rajini, N., Karthikeyan, S., Thiruchitrambalam, M., & Venkateshwaren, N. (2015). Mechanical properties of banana/kenaf fiber-reinforced hybrid polyester composites: Effect of woven fabric and random orientation. *Materials and Design, 66*, 246–257. https://doi.org/10.1016/j.matdes.2014.10.067

Anuar, H., Ahmad, S. H., Rasid, R., Ahmad, A., & Wan Busu, W. N. (2008). Mechanical properties and dynamic mechanical analysis of thermoplastic-natural-rubber-reinforced short carbon fiber and kenaf fiber hybrid composites. *Journal of Applied Polymer Science, 107*(6), 4043–4052. https://doi.org/10.1002/app.27441

Arthanarieswaran, V. P., Kumaravel, A., & Kathirselvam, M. (2014). Evaluation of mechanical properties of banana and sisal fiber reinforced epoxy composites: Influence of glass fiber hybridization. *Materials and Design, 64*, 194–202. https://doi.org/10.1016/j.matdes.2014.07.058

Asim, M., Abdan, K., Jawaid, M., Nasir, M., Dashtizadeh, Z., Ishak, M. R., & Hoque, M. E. (2015). A review on pineapple leaves fibre and its composites. *International Journal of Polymer Science, 2015*, 1–17. https://doi.org/10.1155/2015/950567

Asumani, O. M. L., Reid, R. G., & Paskaramoorthy, R. (2012). The effects of alkali-silane treatment on the tensile and flexural properties of short fibre non-woven kenaf reinforced polypropylene composites. *Composites Part A: Applied Science and Manufacturing, 43*(9), 1431–1440. https://doi.org/10.1016/j.compositesa.2012.04.007

Atiqah, A., Maleque, M. A., Jawaid, M., & Iqbal, M. (2014). Development of kenaf-glass reinforced unsaturated polyester hybrid composite for structural applications. *Composites Part B: Engineering, 56*, 68–73. https://doi.org/10.1016/j.compositesb.2013.08.019

Benítez, A. N., Monzón, M. D., Angulo, I., Ortega, Z., Hernández, P. M., & Marrero, M. D. (2013). Treatment of banana fiber for use in the reinforcement of polymeric matrices. *Measurement: Journal of the International Measurement Confederation, 46*(3), 1065–1073. https://doi.org/10.1016/j.measurement.2012.11.021

Boopalan, M., Niranjanaa, M., & Umapathy, M. J. (2013). Study on the mechanical properties and thermal properties of jute and banana fiber reinforced epoxy hybrid composites. *Composites Part B: Engineering, 51*, 54–57. https://doi.org/10.1016/j.compositesb.2013.02.033

Corrales, F., Vilaseca, F., Llop, M., Girones, J., Mendez, J. A., & Mutje, P. (2007). Chemical modification of jute fibers for the production of green-composites. *Journal of Hazardous Materials, 144*(3), 730–735. https://doi.org/10.1016/j.jhazmat.2007.01.103

Dai, Z., Zhang, B., Shi, F., Li, M., Zhang, Z., & Gu, Y. (2011). Effect of heat treatment on carbon fiber surface properties and fibers/epoxy interfacial adhesion. *Applied Surface Science, 257*(20), 8457–8461. https://doi.org/10.1016/j.apsusc.2011.04.129

Gent, B. C., Leuven, H., Rue, S. P., & Saint-Jean, B. (2010). *Composites: Materials of the future Part 2: Market and market developments* (pp. 1–7). SIRRIS: Driving Industry by Technology.

Gu, H. (2009). Tensile behaviours of the coir fibre and related composites after NaOH treatment. *Materials and Design, 30*(9), 3931–3934. https://doi.org/10.1016/j.matdes.2009.01.035

Guimarães, J. L., Frollini, E., Da Silva, C. G., Wypych, F., & Satyanarayana, K. G. (2009). Characterization of banana, sugarcane bagasse and sponge gourd fibers of Brazil. *Industrial Crops and Products, 30*(3), 407–415. https://doi.org/10.1016/j.indcrop.2009.07.013

Gurunathan, T., Mohanty, S., & Nayak, S. K. (2015). A review of the recent developments in biocomposites based on natural fibres and their application perspectives. *Composites Part A: Applied Science and Manufacturing, 77*, 1–25. https://doi.org/10.1016/j.compositesa.2015.06.007

Idicula, M., Malhotra, S. K., Joseph, K., & Thomas, S. (2005). Dynamic mechanical analysis of randomly oriented intimately mixed short banana/sisal hybrid fibre reinforced polyester composites. *Composites Science and Technology, 65*(7–8), 1077–1087. https://doi.org/10.1016/j.compscitech.2004.10.023

John, M., & Thomas, S. (2008). Biofibres and biocomposites. *Carbohydrate Polymers, 71*(3), 343–364. https://doi.org/10.1016/j.carbpol.2007.05.040

Joseph, S., Sreekala, M. S., Oommen, Z., Koshy, P., & Thomas, S. (2002). A comparison of the mechanical properties of phenol formaldehyde composites reinforced with banana fibres and glass fibres. *Composites Science and Technology, 62*(14), 1857-1868.

Koronis, G., Silva, A., & Fontul, M. (2013). Green composites: A review of adequate materials for automotive applications. *Composites Part B: Engineering, 44*(1), 120–127. https://doi.org/10.1016/j.compositesb.2012.07.004

Li, X., Tabil, L. G., & Panigrahi, S. (2007). Chemical treatments of natural fiber for use in natural fiber-reinforced composites: A review. *Journal of Polymers and the Environment, 15*(1), 25–33. https://doi.org/10.1007/s10924-006-0042-3

Lopattananon, N., Payae, Y., & Seadan, M. (2008). Influence of fiber modification on interfacial adhesion and mechanical properties of pineapple leaf fiber-epoxy composites. *Journal of Applied Polymer Science, 110*(1), 433–443. https://doi.org/10.1002/app.28496

Mahjoub, R., Yatim, J. M., Mohd Sam, A. R., & Hashemi, S. H. (2014). Tensile properties of kenaf fiber due to various conditions of chemical fiber surface modifications. *Construction and Building Materials, 55*, 103–113. https://doi.org/10.1016/j.conbuildmat.2014.01.036

Meon, M. S., Othman, M. F., Husain, H., Remeli, M. F., & Syawal, M. S. M. (2012). Improving Tensile properties of kenaf fibers treated with sodium hydroxide. *Procedia Engineering, 41*(Iris), 1587–1592. https://doi.org/10.1016/j.proeng.2012.07.354

Merlini, C., Soldi, V., & Barra, G. M. O. (2011). Influence of fiber surface treatment and length on physico-chemical properties of short random banana fiber-reinforced castor oil polyurethane composites. *Polymer Testing, 30*(8), 833–840. https://doi.org/10.1016/j.polymertesting.2011.08.008

Mishra, S., Mohanty, A. K., Drzal, L. T., Misra, M., & Parija, S. (2003). Studies on mechanical performance of biofibre/glass reinforced polyester hybrid composites, *63*, 1377–1385. https://doi.org/10.1016/S0266-3538(03)00084-8

Mostafa, M., & Uddin, N. (2015). Effect of banana fibers on the compressive and flexural strength of compressed earth blocks. *Buildings, 5*(1), 282–296. https://doi.org/10.3390/buildings5010282

Nopparut, A., & Amornsakchai, T. (2016). Influence of pineapple leaf fiber and its surface treatment on molecular orientation in, and mechanical properties of, injection molded nylon composites. *Polymer Testing, 52*, 141–149. https://doi.org/10.1016/j.polymertesting.2016.04.012

Panyasart, K., Chaiyut, N., Amornsakchai, T., & Santawitee, O. (2014). Effect of surface treatment on the properties of pineapple leaf fibers reinforced polyamide 6 composites. *Energy Procedia, 56*(C), 406–413. https://doi.org/10.1016/j.egypro.2014.07.173

Rowell, R. M., Han, J. S., & Rowell, J. S. (2000). Characterization and factors effecting fiber properties. *Natural Polymers an Agrofibers Composites*, 115–134.

Sanjay, M. R., Arpitha, G. R., & Yogesha, B. (2015). Study on mechanical properties of natural – Glass fibre reinforced polymer hybrid composites: A review. *Materials Today: Proceedings, 2*(4–5), 2959–2967. https://doi.org/10.1016/j.matpr.2015.07.264

Sgriccia, N., Hawley, M. C., & Misra, M. (2008). Characterization of natural fiber surfaces and natural fiber composites. *Composites Part A: Applied Science and Manufacturing, 39*(10), 1632–1637. https://doi.org/10.1016/j.compositesa.2008.07.007

Shih, Y. F., Chang, W. C., Liu, W. C., Lee, C. C., Kuan, C. S., & Yu, Y. H. (2014). Pineapple leaf/recycled disposable chopstick hybrid fiber-reinforced biodegradable composites. *Journal of the Taiwan Institute of Chemical Engineers, 45*(4), 2039–2046. https://doi.org/10.1016/j.jtice.2014.02.015

Sybil P. Parker, R. A. C. (2000). *McGraw-Hill encyclopedia of environmental science and engineering* (8th Ed.). New York: McGraw-Hill.

Venkateshwaran, N., Elaya Perumal, A., & Arunsundaranayagam, D. (2013). Fiber surface treatment and its effect on mechanical and visco-elastic behaviour of banana/epoxy composite. *Materials and Design, 47*, 151–159. https://doi.org/10.1016/j.matdes.2012.12.001

Venkateshwaran, N., & ElayaPerumal, A. (2012). Mechanical and water absorption properties of woven jute/banana hybrid composites. *Fibers and Polymers, 13*(7), 907–914. https://doi.org/10.1007/s12221-012-0907-0

Zainudin, E. S., Sapuan, S. M., Abdan, K., & Mohamad, M. T. M. (2009). Thermal degradation of banana pseudo-stem filled unplasticized polyvinyl chloride (UPVC) composites. *Materials and Design, 30*(3), 557–562. https://doi.org/10.1016/j.matdes.2008.05.060

Pertanika J. Sci. & Technol. 26 (1): 177 - 192 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

The Use of Environmentally Friendly Bio-Oil in the Production of Phenol Formaldehyde (PF) Resin

Naja Nadiera Omar[1]*, Iskandar Shahrim Mustafa[1], Nurhayati Abdullah[1] and Rokiah Hashim[2]

[1]School of Physics, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[2]School of Industrial Technology, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

ABSTRACT

Phenol Formaldehyde (PF) resin has been extensively used in the manufacturing industry as a binding agent, especially in the production of wood-based panels because of its ability to provide good moisture resistance, exterior strength and durability as well as excellent temperature stability. However, due to the use of limited petroleum-based phenol in its formulation, there is a strong interest in exploring renewable biomass material to partially substitute the petroleum-based phenol. In this study, the slow pyrolysis of biomass decomposition process was used to convert two types of biomass, namely, oil palm frond and *Rhizophora* hardwood, into bio-oil. The phenol-rich fraction of the bio-oil was separated and added into the formulation of PF resin to produce an environmentally-friendly type of PF resin, known as bio-oil-phenol-formaldehyde (BPF) resin. This BPF resin was observed to have comparable viscosity, better alkalinity, improved non-volatile content and faster curing temperature than conventional PF resin. Moreover, the particleboard bonded with this BPF resin was observed to have just as excellent bonding strength as the one bonded using conventional PF resin. However, the BPF resin exhibited an increased level of free formaldehyde and less thermal stability than the conventional PF resin, probably due to the addition of the less reactive bio-oil.

Keywords: Phenol formaldehyde resin, bio-oil, oil palm frond, *Rhizophora* hardwood

ARTICLE INFO

Article history:
Received: 23 November 2016
Accepted: 04 July 2017

E-mail addresses:
naja.nadiera@gmail.com (Naja Nadiera Omar),
iskandarshah@usm.my (Iskandar Shahrim Mustafa),
nurhaya@usm.my (Nurhayati Abdullah),
hrokiah@usm.my (Rokiah Hashim)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

PF resin, also known as phenolic resin, is a reddish-brown resin mainly produced from condensation reaction of petroleum-based phenol and excess formaldehyde. The reaction is also catalysed by an alkali to produce resole-type resin (Athanassiadou et al., 2002). The resole, frequently in liquid form, can be cured to a solid thermosetting network polymer using heat.

PF resin is extensively used in the manufacturing industry for production of moulding forms, thermo-insulation materials, rubbers, laminates as well as abrasive and frictional materials in which the resin acts as a binding agent. PF resin is also used as a binding agent in composite wood panels or wood-based panels such as particleboards, fibreboards, plywood and orientated strand boards (OSB). The latter application represents the major application segment of PF resin, with over 34% of the total volume production. In 2014, approximately 4.7 million tons of PF resin was produced globally, especially in the Asia-Pacific region, which consumed around 35% of the global consumption. In terms of revenue, the production of PF resin was valued at approximately USD10 billion in 2014 and is expected to reach approximately USD14 billion by 2020 (Zion Research, 2015).

As a matter of health concern, there are two potential sources of formaldehyde emission from composite wood panels or wood-based panels, namely, the unreacted free formaldehyde from synthesis of PF resin and the formaldehyde release resulting from breakdown of cured resin. The release of unreacted free formaldehyde from the panels can be greatly limited by controlling the formaldehyde to phenol molar ratio during resin production. In the case of PF resin, the amount of free formaldehyde has been reported as undetectable (Chaouch et al., 2014; Cheng et al., 2011; Shakhreet et al., 2013).

A potentially greater source of formaldehyde emission comes from the chemical breakdown of cured resin, known as hydrolysis, which usually occurs in the presence of water. However, it was reported that panels bonded with PF resin emit nearly no formaldehyde (Shakhreet et al., 2013). This is because PF resin tends to be more chemically stable than other formaldehyde-based resin such as urea formaldehyde (UF) resin or melamine-urea formaldehyde (MUF) resin. Furthermore, the chemical structure of cured PF resin is less hydrophilic towards water, which contributes to the fact that PF resin has good moisture resistance. The two characteristics make cured PF resin much less susceptible to hydrolysis that will induce formaldehyde emission. In both the United States and Europe, panels bonded with PF resin are classified as non-emitting and are exempted from formaldehyde emission regulations (Böhm et al., 2012).

However, the main drawback of conventional PF resin is that it can be very expensive due to the price of phenol. Hence, the use of more natural products such as bio-oil has been utilised. Bio-oil is a very suitable option because it is rich in phenols, which are mainly found within the bio-oil in the form of pyrolytic lignin (Kim et al., 2010).

Several attempts have been made to utilise bio-oil as a phenol substitute to produce BPF resin. These attempts include the incorporation of bio-oil obtained from fast pyrolysis of pine wood (Sukhbaatar et al., 2009), direct liquefaction of white pine sawdust (Cheng et al., 2011) as well as fast pyrolysis of white spruce and trembling aspen (Chaouch et al., 2014). From there, it was noticed that all of the available literature reported on the use of bio-oil originating from softwood biomass, probably because softwood contains a relatively higher content of lignin than any other types of biomass, and in turn, promotes the production of bio-oil with a higher amount of phenols (Demirbas, 2010).

This study was conducted to describe the area still unexplored in research related to the synthesis of BPF resin originating from non-softwood biomass, oil palm frond and *Rhizophora* hardwood. The rationale behind the selection of oil palm frond was due to its abundant availability across the world. Oil palm frond is usually harvested annually at about

10.9 tons per hectare from more than 13.5 million hectares of oil palm plantation around the world. Since oil palm frond has a very limited utility, an initiative was taken by this study to optimally exploit the enormous amount of oil palm frond (Kelly-Yong et al., 2007). Meanwhile, bonding strength of the BPF resin was determined by blending the resin with *Rhizophora* wood particles to produce a *Rhizophora* particleboard, which would be most beneficial in the diagnostic radiography industry (Abuarra et al., 2014). Therefore, the similar origin of resin and wood particles was chosen to investigate whether or not it would enhance the resulting bonding strength.

During this study, the expensive petroleum-based phenol usually used in conventional PF resin was substituted with phenol-rich bio-oil obtained from slow pyrolysis. It was believed that with the use of the environmentally-friendly bio-oil, BPF resin with comparable or better capability than the conventional PF resin might be produced.

MATERIALS AND METHODS

Materials

Oil palm frond was harvested in August 2014 in a plantation of Universiti Sains Malaysia (USM), Nibong Tebal, Pulau Pinang, Malaysia (5°08'48.2'N 100°29'32.0'E). Oil palm frond that was abandoned from the harvesting process was collected and the leaves attached to them were removed using a machete. *Rhizophora* hardwood was also collected in August 2014 from a charcoal factory in Kuala Sepetang, Perak, Malaysia (4°50'12.1"N 100°38'13.9"E). During the collection, bark of the *Rhizophora* was removed using a bark spud and a total of two units of bark-free hardwood were randomly chosen.

Immediately after retrieval, the oil palm fronds and the *Rhizophora* hardwood were dried in a Venticell oven at 105°C until their moisture content wasreduced to less than 10 mf wt% to avoid growth of fungus or microorganism (Abdullah et al., 2014). Then, a Hitachi band saw machine was used to cross cut the sample to an appropriate length. This was important to ensure that the samples could be milled by a Riken grinder with screen size of 1.5 mm. The properties of the two samples were later determined and are presented in Table 1. The reagents used during the synthesis of BPF resin were liquefied phenol (99%), formaldehyde (37%), sodium hydroxide (50%) and anhydrous ethanol (99%). All the reagents were of analytical or pharmaceutical grade.

Preparation of Bio-Oil: Slow Pyrolysis

In brief, slow pyrolysis of oil palm frond and *Rhizophora* hardwood was carried out in a fixed bed reactor (modified Thermolyne F62700 muffle furnace) equipped with a liquid collection system. In a typical run, approximately 200 g of biomass sample was introduced into a stainless steel pyrolyzer (diameter=7 cm, length=15 cm) and placed inside the muffle furnace. All of the related apparatus was then assembled according to the experimental setup shown in Figure 1. From the start, nitrogen gas was purged into the pyrolyser at 100 ml/min to facilitate the removal of pyrolysis vapour into the liquid collection system (Ertaş & Alma, 2010).

Table 1
Properties of oil palm frond and Rhizophora hardwood

Properties	Oil palm frond	*Rhizophora* hardwood	Method
Structural analysis			
Cellulose (wt%)	47.34 ± 1.07	49.42 ± 0.94	D1103 (ASTM, 1978a)
Hemicellulose (wt%)	27.25 ± 1.92	26.63 ± 1.19	D1104 (ASTM, 1978b)
Lignin (wt%)	20.08 ± 2.42	18.65 ± 1.84	D1106 (ASTM, 2001a)
Extractive (wt%)	3.60 ± 0.06	1.28 ± 0.04	D1107 (ASTM, 2007)
Elemental analysis			
C (wt%)	42.91 ± 1.21	44.15 ± 1.94	Perkin Elmer Series II CHNS/O Analyzer
H (wt%)	6.88 ± 1.36	7.13 ± 1.05	
N (wt%)	0.47 ± 0.09	0.34 ± 0.05	
S (wt%)	0.03 ± 0.02	0.41 ± 0.04	
O (wt%)	46.27 ± 2.41	47.38 ± 2.10	
Proximate analysis			
Moisture content (wt%)	7.47 ± 0.25	7.61 ± 0.15	A&D MX-50 moisture analyzer
Volatile matter (wt%)	82.42 ± 1.39	84.63 ± 0.52	E872 (ASTM, 2006)
Ash (wt%)	3.44 ± 0.30	0.59 ± 0.24	D1102 (ASTM, 2001b)
Fixed carbon (wt%)	6.67 ± 1.42	7.16 ± 0.64	By difference

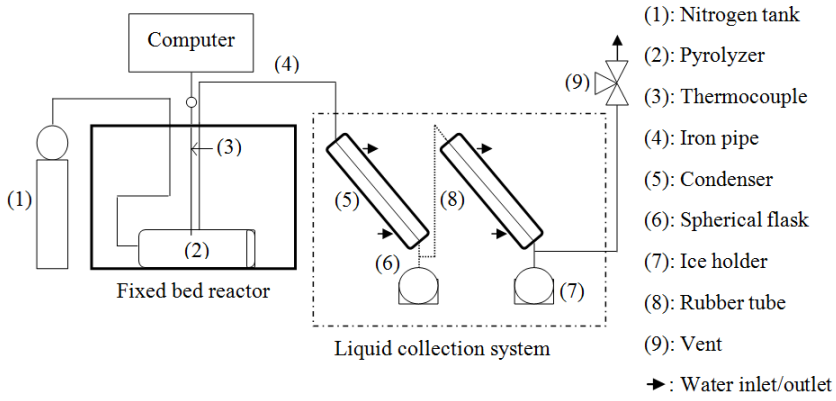


Figure 1. Experimental setup of pyrolysis

The furnace was then heated to 375°C at a steady rate of 100°C/min. These parameters correspond to the optimised yield of phenols obtained in the bio-oil (unpublished data). After one hour of holding time, the muffle furnace was turned off and allowed to stabilise to room temperature. The bio-oil, as obtained from the condensation of pyrolysis vapours, was collected, separated using dichloromethane and analysed as in Table 2 before it was used in the synthesis of BPF resin.

Table 2
Properties of Bio-oil

Properties	Oil palm frond	*Rhizophora* hardwood	Method
Element			
C (wt%)	66.23 ± 1.14	67.66 ± 1.59	Perkin Elmer Series II CHNS/O Analyzer
H (wt%)	9.35 ± 1.16	8.12 ± 1.44	
N (wt%)	0.05 ± 0.02	0.02 ± 0.01	
S (wt%)	0.00 ± 0.00	0.05 ± 0.03	
O (wt%)	23.77 ± 2.30	23.93 ± 2.58	
Ash (wt%)	0.61 ± 0.03	0.22 ± 0.06	D1102 (ASTM, 2001b)
Char (wt%)	0.32 ± 0.13	0.14 ± 0.08	Filtration method
pH	3.01 ± 0.03	2.63 ± 0.03	Accumet AB15 pH meter
Total peak area of phenols (%)	74.4	71.1	
Distribution of phenols (%)	Phenol (40.6)	Phenol (7.3)	Agilent Technologies 7890A/5975C GCMS
	Phenol, 2-methyl- (1.7)	Phenol, 2-methyl- (2.4)	
	Phenol, 3-methyl- (2.3)	Phenol, 3-methyl- (2.9)	
	Phenol, 2-methoxy- (8.5)	Phenol, 2-methoxy- (16.5)	
	4-Mercaptophenol (1.4)	Phenol, 2,4-dimethyl- (1.8)	
	Phenol, 2,5-dimethyl- (1.4)	Phenol, 2-methoxy-5-methyl- (5.8)	
	Phenol, 4-ethyl-2-methoxy- (4.4)	Phenol, 2-methoxy- (1.6)	
	Phenol, 2,6-dimethoxy- (14.2)	Phenol, 4-ethyl-2-methoxy- (3.9)	
		Phenol, 2,6-dimethoxy- (29.0)	

Synthesis of BPF Resin

The synthesis procedure was maintained in a closed system. In brief, the required volume of reagents was measured using a Fisherbrand pipette. Then, liquefied phenol, ethanol and sodium hydroxide were loaded into a three-necked flask connected to a pressure-equalising addition funnel, thermometer and cooling condenser and placed on a ceramic heating plate equipped with magnetic stirrer. Bio-oil, after the separation procedure, was also used to substitute the liquefied phenol at 25% and 75% of the total amount required. The mixture was heated to 65°C and maintained at that temperature for 30 minutes to ensure a homogenous alkaline medium was produced (Chaouch et al., 2014). After homogenisation had occurred, the temperature was raised to 80°C and formaldehyde solution was added step-wise over a period of 10 minutes. Finally, the reaction mixture was heated to 95°C and kept at that temperature for a period of time to allow condensation reaction and polymerisation to occur. After the required viscosity had seemingly been reached, the reaction was stopped and allowed to stabilise to room temperature before being dissembled. The BPF resin was then refrigerated in a sealed glass

bottle to prolong its pot life by minimising any additional slow polymerisation of phenolic rings (Sukhbaatar et al., 2009).

Characterization of BPF Resin

Several analyses such as viscosity, pH, non-volatile measurements, free formaldehyde level, pot life evaluation, DSC, TGA and bonding strength determination were conducted to study the properties of BPF resin. These properties were then compared with the properties of conventional PF resin.

Viscosity measurement was conducted using the Visco Basic Plus viscometer with L1 spindle according to D1084 (ASTM, 2016a). The pH value of the resin was determined using an Accumet AB15 pH meter. The non-volatile content of the BPF resin was evaluated at 105°C in reference to D4426 (ASTM, 2013). Free formaldehyde level was measured according to 11402 (ISO, 2012). In addition, pot life was denoted as the time taken for the initial viscosity of the resin to become twice its original, according to D1337 (ASTM, 2016b). Therefore, the viscosity of BPF resin was determined regularly, at a fixed time of interval from the first week of synthesis until the fifth week of storage.

On the other hand, DSC analysis was conducted to evaluate thermal curing properties of BPF resin using the Perkin Elmer DSC Pyris 6. Approximately 10 mg of resin was sealed in the given DSC aluminium pan, placed onto the sample holder and heated from 30°C to 250°C at 10°C/min. A flow of nitrogen gas at 20 ml/min was maintained over the sample to create a dry and reproducible atmosphere. Meanwhile, TGA was carried out to analyse thermal behaviour of the resin using the TGA/DSC-1 Mettler Toledo. The analysis started by heating the resin sample from room temperature to 700°C at a heating rate of 10°C/min. Together with the heating, nitrogen gas was purged into the system at a flow rate of 50 ml/min. To pronounce the decomposition pattern of each resin, cured conventional PF and BPF resins were used. These cured resins were obtained by heating the respective resins to their curing temperature as obtained from the DSC analysis.

Moreover, bonding strength was determined by fabricating a *Rhizophora* particleboard with 10% of the resin as binding agent. The fabrication procedure of the particleboard was conducted according to previous literature (Ngu et al., 2015). Following that, bonding strength was measured according to A5908 (JIS, 2015).

RESULTS AND DISCUSSION

Formulation

The formulation used to synthesise BPF resins was tabulated in Table 3. For easy reference, abbreviation of BPF(OP) and BPF(RS) was used denoting BPF resin originated from bio-oil of oil palm frond and bio-oil of *Rhizophora* hardwood, respectively.

Table 3
Formulation of BPF resin

Type of Resin	Bio-oil Substituted	F/P	Molar Ratio NaOH/P	EtOH/P	Time to Reach 200 cP (h)
BPF(OP25)	25	1.3	0.3	0.4	5.0
BPF(OP75)	75	1.3	0.5	0.4	3.5
BPF(RS25)	25	1.3	0.3	0.4	5.0
BPF(RS75)	75	1.3	0.5	0.4	3.5

From the preliminary study, molar ratio of formaldehyde to phenol at 1.3 was observed to produce free formaldehyde level less than the standard permissible value for formaldehyde-based resin, which was set at 0.4 wt% (Cetin & Özmen, 2002). Meanwhile, 0.3 to 0.5 molar ratio of sodium hydroxide to phenol was engaged to conform to the typical pH value of conventional PF resin, which is usually around 11 (Ayrilmis et al., 2008; Chaouch et al., 2014; Zhao et al., 2010). Different substitution levels of sodium hydroxide were initiated to counteract the different amount of acidic bio-oil within the formulation. In the case of ethanol, a lower ratio of ethanol to phenol reduced the solubility of bio-oil in the resin solution. From the preliminary study, it was observed that when the molar ratio of ethanol to phenol was less than 0.4, the BPF resin separated into two different fractions.

In general, for the same amount of bio-oil substituted during BPF resin synthesis, a similar amount of reagents and time was required to achieve the desired viscosity, regardless of the origin of the bio-oil. This was expected since the properties of the two bio-oils were comparable, as found from previous characterisation.

Meanwhile, when comparing different substitution levels of the same bio-oil, BPF resin with higher percentage of bio-oil reached the desired viscosity in a shorter time. The occurrence was most likely because a much larger and complex molecular structure existed in the bio-oil as compared to the petroleum-based phenol, increasing its degree of condensation and polymerisation (Cheng et al., 2011).

Furthermore, it was noticed that the condensation time required to achieve the desired viscosity was different in this study. Other studies reported that the condensation time taken was from 1 to 3 hours, whereas in this study, 3 to 5 hours was needed (Chaouch et al., 2014; Zhao et al., 2010). This could be explained by the inclusion of the low molar ratio of formaldehyde to phenol and the high molar ratio of ethanol to phenol during the synthesis. It is known that a low amount of formaldehyde would reduce the rate of reaction between formaldehyde and phenol, while a high amount of ethanol would slow down the condensation process as it introduced a side reaction between the ethanol and phenol (Mo et al., 2015; Yan et al., 2008). Hence, to ensure that maximum reaction and condensation occurred, a longer time was needed. Despite having an unfavourable effect on the synthesis procedure, the low amount of formaldehyde and the high amount of ethanol were still chosen to ensure that the final properties of the BPF resin were approximately similar to the conventional PF resin.

Properties of BPF Resin

The BPF resins obtained were homogenous and dark brown in color, slightly different from the reddish-brown conventional PF resins. This was probably due to the addition of dark-brown coloured bio-oil during the synthesis. The properties of the BPF resin were determined and are presented in Table 4. It was observed that all of the BPF resins successfully reached the desired viscosity of around 200 cP, as in conventional PF resin.

Table 4
Properties of resins

Type of Resin	Viscosity (cP)	pH	Non-volatile content (wt%)	Free Formaldehyde Level (wt%)
Conventional PF	250-500, 200, 233	11.75, 11.16, 10.5	47, 64, 49	Undetectable
	(Ayrilmis et al., 2008; Chaouch et al., 2014; Zhao et al., 2010)	(Ayrilmis et al., 2008; Chaouch et al., 2014; Zhao et al., 2010)	(Ayrilmis et al., 2008; Chaouch et al., 2014; Zhao et al., 2010)	(Chaouch et al., 2014)
PF	209 ± 3	9.36 ± 0.03	42.40 ± 0.11	0.00 ± NA
BPF(OP25)	220 ± 4	11.60 ± 0.03	56.96 ± 0.86	0.17 ± 0.05
BPF(OP75)	204 ± 2	11.83 ± 0.05	42.60 ± 0.41	0.41 ± 0.04
BPF(RS25)	227 ± 4	11.37 ± 0.02	59.21 ± 0.24	0.11 ± 0.02
BPF(RS75)	215 ± 2	11.60 ± 0.06	45.66 ± 0.20	0.35 ± 0.02

When comparing the origin of the resins, it was found that BPF(OP) resin had lower viscosity than BPF(RS) resin for both substitution levels. The lower viscosity of resin achieved after having similar formulation and time of condensation indicated that the BPF(OP) resin had a lower molecular weight than the BPF(RS) resin, probably due to the nature of the bio-oil (Cheng et al., 2011; Zhao et al., 2010).

It was also shown that the conventional PF resin had significantly lower pH than the other BPF resins. In most studies, it was reported that the ideal pH value for phenolic resin was between 10 and 12 (Ayrilmis et al., 2008; Chaouch et al., 2014; Zhao et al., 2010). Those with a lower pH may exhibit poor curing properties as polymerisation and curing of phenolic resin accelerated at a very alkaline pH. However, due to the complicated mechanism of the alkaline catalyst, the explanation behind the occurrence was reported as unclear (Pizzi, 1994). In this study, the formulation was prepared according to the typical pH value of conventional PF resin instead of the purchased conventional PF resin, hoping that with a higher pH, the curing properties of the BPF resin would improve.

In addition, for the similar amount of bio-oil substituted, BPF(OP) resin had a higher pH than did the BPF(RS) resin, even though a similar amount of sodium hydroxide was added. This finding was in accordance with the pH value of the respective bio-oil.

Furthermore, it was noticed that all of the BPF resins had higher non-volatile content than the conventional PF resin. The high amount of non-volatile content is favourable as it reduces the required amount of resin to be added during the production of panels; hence, decreasing the consumption of the formaldehyde-based resin makes it more economical and user-friendly.

Non-volatile content of phenolic resin is closely dependent on the volatile components such as free phenol and free formaldehyde. High amount of volatile components in the resin causes low value of non-volatile content as in the case of conventional PF resin. It was observed that the conventional PF resin had low free formaldehyde level, which raised the possibility that it might use a low level of formaldehyde. It is worthy to note that the addition of a low level of formaldehyde caused an imbalance reaction between the phenol and formaldehyde, leaving excess unreacted phenol at the end of the synthesis procedure. Upon heating, this free phenol vapourised from the resin and as phenol has a higher molecular weight than formaldehyde, the vaporisation of free phenol from the resin caused a more significant reduction in terms of non-volatile content than the vaporisation of formaldehyde.

On the other hand, when comparing a different substitution level of the same bio-oil, those with a higher substitution level had a lower value of non-volatile content. The most possible explanation was that together with the increase of bio-oil, the amount of unreacted free formaldehyde might have increased as well. Therefore, when this type of resin was heated, a large amount of volatile was evaporated and finally, the non-volatile content was reduced significantly. Similarly, when comparing the origin of the resins, BPF(OP) resin had lower non-volatile content. Hence, it was expected that this type of resin had a higher amount of free formaldehyde than the BPF(RS) resin.

From Table 4, it was found that all the resins exhibited free formaldehyde levels around 0.4 wt%. As for the purchased conventional PF resin, the free formaldehyde level was reported as being not detectable, cited from the Safety Data Sheet (SDS) provided by the supplier. Meanwhile, for all other BPF resins, the free formaldehyde level was observed to increase with an increase of the substitution level of phenol with bio-oil. This was mostly due to the lower reactivity of bio-oil over phenol or the lower number of active sites especially in *–otho* and *–para* position to the phenolic hydroxyl group of the bio-oil that might cause poor interaction between phenol and formaldehyde, leaving an amount of free formaldehyde at the end of the synthesis. When comparing the origin of the BPF resins, it was found that BPF(OP) resin had a higher free formaldehyde level than BPF(RS) resin, which suggested lower reactivity of bio-oil produced from oil palm frond than the one produced from *Rhizophora* hardwood, possibly due to lesser active sites in the phenolic rings (Chaouch et al., 2014; Cheng et al., 2011).

Pot life of the BPF resin was also investigated by constantly measuring of the viscosity for five consecutive weeks, as shown in Figure 2. In the earlier weeks, the viscosity increased at a slower rate. However, since resole-type phenolic resin is capable of curing without any addition of heat or curing agent, slow polymerisation occurred during storage time. The polymerisation process is an exothermic reaction due to the formation of new polymer chains that enhance the release of heat energy. This heat energy catalyses the rate of polymerization, causing an exponential increase in viscosity over time.

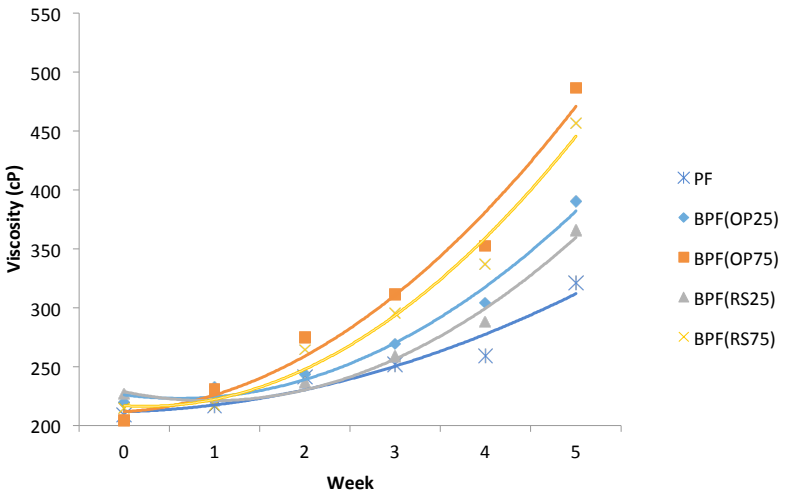


Figure 2. Pot life of resins

From Figure 2, it was seen that conventional PF resin had the greatest stability over time, with only 35% of increase from the initial. Since it had a lower pH value than the other BPF resins, its rate of polymerisation is lower, improving its stability over time. It was also observed that the viscosity of BPF resin with 25% substitution level increased at a slower rate compared to those with a 75% substitution level, with a total change of 38-44% and 53-58%, respectively. As previously characterised, the added bio-oil had a small amount of ash/char that affected the stability of bio-oil negatively (Pollard et al., 2012). Therefore, a higher substitution level of bio-oil in place of phenol increased the amount of ash/char in the BPF resin, which consequently reduced its stability compared to those with a lower substitution level. Similarly, the more pronounced increase in the viscosity of BPF(OP) resin than in that of BPF(RS) resin can be explained by the higher amount of ash/char content in the corresponding bio-oil.

Although the pot life of resins was observed to be around the fifth week (the week at which the viscosity of the resin doubled), it was important to note that with proper storage in a low temperature environment, the pot life can be prolonged.

Thermal curing properties of the resins were analysed using DSC results. In an attempt to enhance the exothermic peak of each resin, the DSC curves were plotted from 100°C to 230°C, shown in Figure 3. The temperature at which curing reaction occurred was represented by this exothermic peak. During curing reaction, further condensation and polymerisation occurred to produce a more stable cross-linked resin structure. Curing patterns obtained for all resins were found to be consistent with those typically observed for conventional PF resin, with a single exothermic peak only (Mo et al., 2015).

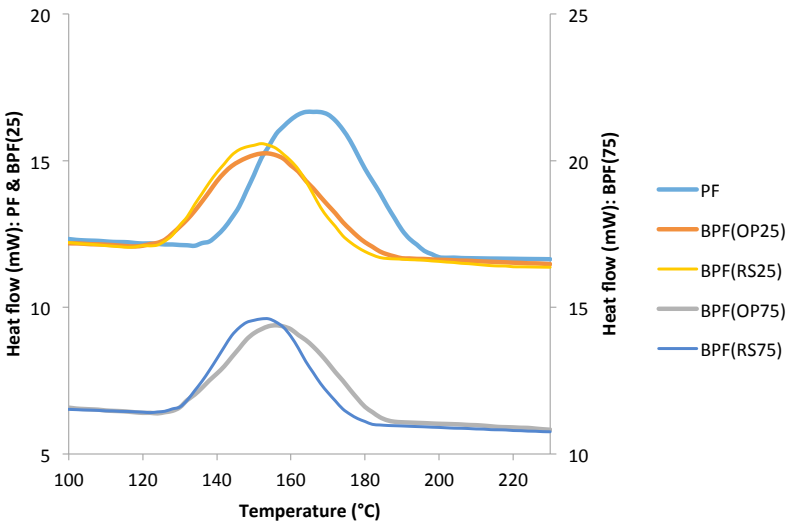


Figure 3. DSC of resins

However, it has been previously reported that the exothermic peak temperature of conventional PF resin was usually at 150°C (Cheng et al., 2011; Zhao et al., 2010). In this study, it was observed that the conventional PF resin had a higher exothermic peak temperature, with a value of 166°C. Since the detailed composition of the conventional PF resin used in this study and other studies was not known, it was difficult to speculate on the mechanism associated with the behaviour of both. However, the high peak temperature of the conventional PF resin used in this study might be attributed to its low pH value, causing it to less likely favour the condensation reaction and, therefore, require higher heat energy to allow complete condensation and polymerisation.

Interestingly, the curing temperature of BPF resins obtained was lower than that of the conventional PF resin used in this study and comparable with the conventional PF resin used in the other studies. The peak temperature of BPF(OP25), BPF(OP75), BPF(RS25) and BPF(RS75) resins were obtained at 154°C, 156°C, 152°C and 154°C, respectively. The occurrence was probably due to the enhanced alkaline environment of the BPF resins, which in turn helped to promote condensation and accelerate curing reaction (Pizzi, 1994).

One previous study suggested that the presence of bio-oil actually improved curing properties. A small amount of bio-oil was believed to favour thermal curing reaction governed by phenol, while a high amount of bio-oil would retard the thermal curing reaction due to the low reactivity of bio-oil (Cheng et al., 2011). In this study, the curing reaction between the BPF resins indeed supported the previous deduction.

In addition, the thermal stability of cured PF and BPF resins was evaluated according to the TG analysis. The result showed that all of the resins had an approximately similar thermal decomposition pattern. It has been previously reported that the decomposition of phenolic resins occurred in three major events known as post-curing, thermal decomposition and ring stripping (Cheng et al., 2011). Post-curing reaction, which could be attributed to further cross-linking and polymerisation reactions, usually occurs when cured resin is subjected to elevated

temperature. The purpose of post-curing reaction is to improve the mechanical properties of resin. However, during this post-curing reaction, low molecular weight compounds such as unreacted phenol or formaldehyde and short-chain polymers tend to vapourise (Ko & Ma, 1998). Meanwhile, thermal decomposition of resin could have been ascribed due to the breakdown of the previously formed methylene bridges linkage into aldehydes and phenols (Papadopoulou & Chrissafis, 2011, Chaouch et al., 2014). Ring stripping was associated with the breakdown of the phenolic ring network (Chaouch et al., 2014). All of these events consequently contribute to the weight loss of resin during TG analysis.

From Figure 4.16, it was observed that the first thermal event occurred from 50°C to 250°C with a total weight loss of 11% for conventional PF resin and 17% to 20% for BPF resins. The second thermal event was obtained from 250°C to 500°C with a total weight loss of around 10% for conventional PF, BPF(OP25) and BPF(RS25) resins. BPF(OP75) and BPF(RS75) resins experienced a significantly higher percentage of weight loss at 23% and 24%, respectively. The third thermal event was observed to occur within 500°C to 700°C with a comparable percentage of weight loss. Both conventional PF and BPF resins showed total weight loss between 9% and 12%.

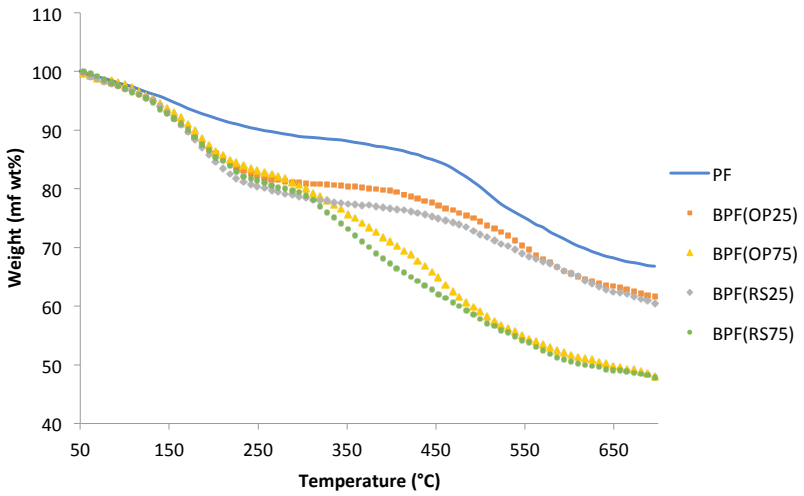


Figure 4. TGA of resins

It can be concluded that thermal stability of PF resin was superior to any other BPF resins, with the lowest percentage of weight loss happening in all three thermal events and the highest amount of final weight residue. In the case of BPF resin, it was seen that when a higher substitution level of bio-oil was incorporated, more pronounced weight loss occurred throughout the heating, resulting in lower weight residue. The occurrence was probably due to the addition of the less reactive and complex bio-oil to the synthesis of resin, causing an increase in the formation of any side chain present within the bio-oil molecules, such as the Tollens reaction, where the lignin side chains are substituted by aliphatic methylol groups

(Kalia & Avérous, 2016). This undesired side chain was easily decomposed with an increased temperature. Furthermore, the presence of an undesired chain might result in a rather weak cross-linked network of polymers and, therefore, affect thermal stability negatively (Cheng et al., 2011, Chaouch et al., 2014).

From Table 5, it was noticed that when different types of resin were used, no significant change to bonding strength was observed, except for a slight increase or decrease in accordance to the viscosity of resin. In cases of resin with lower viscosity such as PF resin and BPF resin with 75% substitution level, the ability of the resin to readily flow and spread itself within the particleboard increased. When this occurred, the distribution of the resin across the particleboard improved, causing more interaction between the resin and *Rhizophora* wood particles. During particleboard formation, the distributed resin cured to a stable polymer and became chemically attached to the wood particles, giving the particleboard higher bonding strength (Pizzi, 1994).

Table 5
Bonding strength of particleboard bonded with phenolic resins

Type of Resin	Bonding Strength (MPa)
PF	1.61 ± 0.13
BPF(OP25)	1.46 ± 0.03
BPF(OP75)	1.65 ± 0.10
BPF(RS25)	1.37 ± 0.13
BPF(RS75)	1.51 ± 0.02

CONCLUSION

The bio-oil produced from slow pyrolysis of oil palm frond and *Rhizophora* hardwood was successfully used to partially substitute petroleum-based phenol in the formulation of resole-type BPF resin. In this study, all of the synthesised resins were seen to successfully imitate the properties of conventional PF resin. These BPF resins were more than capable of reaching the typical viscosity and pH value of conventional PF resin. The BPF resins also had an improved value of non-volatile and better curing properties than the conventional PF resin. Bonding strength of particleboard bonded with the BPF resins was also found to be comparable with those bonded with conventional PF resin. The only downsides to these BPF resins were the increase in free formaldehyde level and the reduction of their thermal stability. These outcomes were produced probably due to the addition of less reactive bio-oil, preventing maximum interaction between phenol and formaldehyde during the synthesis procedure as well as inducing various undesired side reactions that would easily decompose in high temperature.

ACKNOWLEDGEMENT

The authors would like to acknowledge financial support from Universiti Sains Malaysia under the RUI grant (1001/PFIZIK/814250 and 1001/PFIZIK/811301) as well as from Institute of Postgraduate Studies (IPS) of Universiti Sains Malaysia under USM Fellowship 1/14.

REFERENCES

Abdullah, N., Sulaiman, F., Miskam, M. A., & Taib, R. M. (2014). Characterization of banana (Musa spp.) Pseudo-Stem and fruit-bunch-stem as a potential renewable energy resource. I*nternational Journal of Biological, Veterinary, Agricultural and Food Engineering, 8*(8), 712–716.

Abuarra, A., Hashim, R., Bauk, S., Kandaiya, S., & Tousi, E. T. (2014). Fabrication and characterization of gum Arabic bonded Rhizophora spp. particleboards. *Materials and Design, 60*, 108–115.

ASTM, D. (1978a). *1103, Standard test method for alpha-cellulose*. ASTM International.

ASTM, D. (1978b). *1104, Standard test method for holocellulose in wood*. ASTM International.

ASTM, D. (2001a). *1106, Standard test method for acid-insoluble lignin in wood*. ASTM International.

ASTM, D. (2001b). *1102, Standard test method for ash in wood*. ASTM International.

ASTM, D. (2007). *1107, Standard test method for ethanol-toluene solubility of wood.* ASTM International.

ASTM, D. (2013). *4426, Standard test method for determination of percent non-volatile content of liquid phenolic resins used for wood laminating*. ASTM International.

ASTM, D. (2016a). *1084, Standard test methods for viscosity of adhesives*. ASTM International.

ASTM, D. (2016b). *1337, Standard practice for storage life of adhesives by viscosity and bond strength.* ASTM International.

ASTM, E. (2006). *872, Standard test method for volatile matter in the analysis of particulate wood fuels.* ASTM International.

Athanassiadou, E., Tsiantzi, S., & Nakos, P. (2002). Wood adhesives made with pyrolysis oils. In *Proceedings of Euro-Sustain 2002*. Aristotle University of Thessaloniki, Laboratory of Heat Transfer and Environmental Engineering, Rhodes, Greece.

Ayrilmis, N., Candan, Z., & Hiziroglu, S. (2008). Physical and mechanical properties of cardboard panels made from used beverage carton with veneer overlay. *Materials and Design, 29*(10), 1897–1903.

Böhm, M., Salem, M. Z., & Srba, J. (2012). Formaldehyde emission monitoring from a variety of solid wood, plywood, blockboard and flooring products manufactured for building and furnishing materials. *Journal of Hazardous Materials, 221*, 68–79.

Cetin, N. S., & Özmen, N. (2002). Use of organosolv lignin in phenol-formaldehyde resins for particleboard production: I. Organosolv lignin modified resins. *International Journal of Adhesion and Adhesives, 22*(6), 477–480.

Chaouch, M., Diouf, P. N., Laghdir, A., & Yin, S. (2014). Bio-oil from whole-tree feedstock in resol-type phenolic resins. *Journal of Applied Polymer Science, 131*(6), 40014.

Cheng, S., D'Cruz, I., Yuan, Z., Wang, M., Anderson, M., Leitch, M., & Xu, C. (2011). Use of biocrude derived from woody biomass to substitute phenol at a high-substitution level for the production of biobased phenolic resol resins. *Journal of Applied Polymer Science, 121*(5), 2743–2751.

Demirbas, A. (2010). Fuels from biomass. In A. Demirbas (Ed.), *Biorefineries: For Biomass Upgrading Facilities* (pp. 33–73.). London: Springer.

Ertaş, M., & Alma, M. H. (2010). Pyrolysis of laurel (Laurus nobilis L.) extraction residues in a fixed-bed reactor: Characterization of bio-oil and bio-char. *Journal of Analytical and Applied Pyrolysis, 88*(1), 22–29.

Kalia, S., & Avérous, L. (2016). *Biodegradable and biobased polymers for environmental and biomedical applications*. New Jersey: John Wiley & Sons.

Kelly-Yong, T. L., Lee, K. T., Mohamed, A. R., & Bhatia, S. (2007). Potential of hydrogen from oil palm biomass as a source of renewable energy worldwide. *Energy Policy, 35*(11), 5692–5701.

Kim, S. J., Jung, S. H., & Kim, J. S. (2010). Fast pyrolysis of palm kernel shells: Influence of operation parameters on the bio-oil yield and the yield of phenol and phenolic compounds. *Bioresource Technology, 101*(23), 9294–9300.

Ko, T. H., & Ma, T. S. (1998). Effect of post-curing on the mechanical properties of carbonized phenolic resins. *Polymer Composites, 19*, 456–462.

Mo, X., Fan, D., Qin, T., & Chu, F. (2015). Curing characteristics and adhesion performance of phenol-formaldehyde resins with composite additives. *Journal of Tropical Forest Science, 27*, 248–254.

Ngu, K. T., Bauk, S., Hashim, R., Tajuddin, A. A., & Shukri, A. (2015). Fabrication of formaldehyde-based *Rhizophora* spp. particleboards and their mass attenuation coefficients at 15.77, 17.48, 21.18 and 25.27 keV photon energies. *Journal of Physical Science, 26*(1), 27–33.

Papadopoulou, E., & Chrissafis, K. (2011). Thermal study of phenol-formaldehyde resin modified with cashew nut shell liquid. *Thermochimica Acta, 512*(1), 105–109.

Pizzi, A., & Stephanou, A. (1994). Phenol-formaldehyde wood adhesives under very alkaline conditions. Part I: Behaviour and proposed mechanism. *Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood, 48*(1), 35–40.

Pollard, A., Rover, M., & Brown, R. (2012). Characterization of bio-oil recovered as stage fractions with unique chemical and physical properties. *Journal of Analytical and Applied Pyrolysis, 93*, 129–138.

Shakhreet, B., Bauk, S., & Shukri, A. (2013). Mass attenuation coefficients of fabricated *Rhizophora* spp. particleboard for the 15.77 to 25.27 keV range. *American Journal of Scientific and Industrial Research, 4*, 89–94.

Sukhbaatar, B., Steele, P. H., Ingram, L. I., & Kim, M. G. (2009). Use of lignin separated from bio-oil in oriented strand board binder phenol-formaldehyde resins. *BioResources, 4*(2), 789–804.

Yan, Z., Yujian, L., Qi, H., & Zhewen, H. (2008). Effect of solvent on the chain conformation and cure behavior of phenolic resin. *Journal of Applied Polymer Science, 108*(5), 3009–3015.

Zhao, Y., Yan, N., & Feng, M. (2010). Characterization of phenol-formaldehyde resins derived from liquefied lodgepole pine barks. *International Journal of Adhesion and Adhesives, 30*(8), 689–695.

Zion Research. (2015). *Phenolic resins (resols, novolac and others) market for wood-adhesives, molding compounds, laminates, insulation and other applications: Global industry perspective, comprehensive analysis, and forecast, 2014-2020*. Retrieved from http://www.marketresearchstore.com/report/phenolic-resins-market-for-z37563

Pertanika J. Sci. & Technol. 26 (1): 193 - 214 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

High Intrinsic Biosorption Efficiency of Cattle Manure on Cr(VI): A Potential Low-cost Fibre-rich Biosorbent

Yap, K. L.[1], Lee, C. M.[1], Gan, Y. L.[2], Tang, T. K.[1], Lee, Y. Y.[1], Tee, T. P.[3] and Lai, O. M.[1,2]*

[1]Bioprocess Laboratory, Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Department of Bioprocess Technology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[3]Department of Animal Science, Faculty of Agriculture, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Fibre-rich manure derived from grass-fed cattle showed significantly higher intrinsic sorption efficiency on Cr(VI) solution as compared to corncob, sawdust and cogon grass. This observation could be attributed to the ligneous nature and rough surface morphology of the cattle manure. Four-factor, three-level, face-centred composite design (FCCD) suggested the process was greatly affected by initial pH of the solution, contact time and sorbent dosage ($p<0.0001$), while stirring rate had negligible effect. Highest percentage removal (≥70%) happened at pH 1-1.56, 0.79-1 g sorbent and 57-300 min contact time in 200 mg/L Cr(VI) solution. The process is spontaneous, endothermic and best described by pseudo-second-order and Langmuir model. It was found that adsorbed Cr(VI) could be recovered and CM could be reused at least three times with >50% adsorption efficiency. It is predicted that both physisorption and chemisorption are involved in the sorption process.

Keywords: Biosorption, cattle manure, chromium (VI), heavy metals, response surface methodology

ARTICLE INFO

Article history:
Received: 16 November 2016
Accepted: 02 June 2017

E-mail addresses:
ykling1990@gmail.com (Yap, K. L.),
choonminlee2010@gmail.com (Lee, C. M.),
yeelingan@gmail.com (Gan, Y. L.),
teckkim1@gmail.com (Tang, T. K.),
leeying911@gmail.com (Lee, Y. Y.),
ttpoy@yahoo.com (Tee, T. P.),
omlai@upm.edu.my; omlai.biotech@gmail.com (Lai, O. M.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Heavy metal (HM) pollution has long been a serious environmental issue due to its perniciousness even at trace concentration. Non-biodegradability allows HMs to bio-accumulate and bio-magnify along the food chain; this worsens an already alarming situation. Discharge of untreated industrial effluent is the root cause, especially in

developing countries where more than 70% of industrial effluent is discharged without prior treatment (Water, 2009). Hexavalent chromium, Cr(VI), is a harmful HM categorised as a group 1 human carcinogen (IARC, 2012) as it is known to cause various cancers. Cr(VI) is commonly found in wastewater due to its high water solubility, which facilitates the release of toxicity into the environment.

Conventional methods such as chemical precipitation, membrane filtration, electrodialysis, ion exchange and adsorption perform at a comparable level of efficiency and thus, cost plays a decisive role in removing heavy metals from the surrounding environment. The sorption method (ion exchange and adsorption) is preferred due to its relatively cheaper cost (up to 200 USD/million litres) compared to other methods, which can cost up to 500 USD/million litres (Barakat, 2011; Fu & Wang, 2011; Gupta et al., 2012). Although the sorption method is more economical, the synthetic resin and activated carbon used in this method are still a heavy burden for industry to bear.

Agricultural waste (AW) with high lignocellulosic content possesses potential as a cost-effective alternative to replace commercial products. Various surface functional groups contributed by lignocellulose and some proteins of AW make AW suitable for removing divalent HMs (Bailey et al., 1999; Garg et al., 2008). Furthermore, the threatening environmental and disposal issues emerging from the overproduction of AW further intensify the value-added initiative, which is an ideal solution to achieve zero-waste.

Numerous studies on various types of AW have been carried out as stated in the review paper of Ngah and Hanafiah (2008). Miretzky and Cirelli (2010) focussed on the list of AW that were tested to remove Cr. Sawdust, corn waste, leaves, weeds, barks, bagasse and husk were some types of AW that have been studied. Animal waste, a subgroup of AW, is seldom used to remove Cr although it has been shown to have high sorption efficiency for other divalent HMs. Fresh manure excreted from different animals such as turkey (Lima & Marshall, 2005b), sheep (Al-Rub et al., 2002), poultry (Lima & Marshall, 2005a), broiler (Uchimiya et al., 2010), and swine (Meng et al., 2014) showed positive results. Similar studies have also been carried out on cattle-manure compost and cattle-manure vermicompost (Jordão et al., 2010; Jordão et al., 2011; Zaini et al., 2009). Fresh cattle manure (CM), however, has received less attention. It is estimated that there are 1.47 billion head of cattle throughout the world (FAOSTAT, 2013). According to Hofmann and Beaulieu (2006), cattle with an average weight of 635 kg can produce 14 tons of CM annually. Since CM is in abundance and it has higher fibre content over poultry and swine manure due to its feed intake (Chen et al., 2003), it has greater potential to be a cost-effective biosorbent.

The aim of this study was to investigate the feasibility of using native CM excreted from grass-fed cattle to remove Cr(VI). The efficiency was compared with other previously studied lignocellulose materials, namely corncob (C), sawdust (S) and cogon grass (CG) in their native form without structural modification through physical or chemical treatment. The response surface methodology (RSM) was employed to optimise and investigate the single and interactive effect of the important process variables, namely, initial solution pH, contact time (CT), stirring rate (SR) and sorbent dosage (SD) on the sorption process. In addition, the sorption process was examined through kinetic, isotherm and thermodynamic studies.

METHODOLOGY

Materials

Potassium dichromate ($K_2Cr_2O_7$), sodium chloride (NaCl) and nitric acid (HNO_3) were purchased from Fisher Scientific, USA. Sodium hydroxide (NaOH) and hydrochloric acid (HCl) were purchased from Merck Millipore, USA. All the chemicals were of analytical reagent grade except for the HNO_3 which was classified as trace metal grade.

Biosorbent Preparation

Corncob (C) was collected from the local corn industry; hardwood S was purchased from Northeastern Products Corp., NY, USA; CG was collected from a wild grass lawn around Selangor, Malaysia. Fresh CM excreted from grass-fed cattle was collected from the animal farm in Universiti Putra Malaysia. All AW was cleaned and washed with distilled water until no visible surface particulates were observed. Then, they were dried in a hot air oven at 70°C for 1 hour to kill pathogens followed by 60°C for 24 h. The AW was ground using a blender and sieved to particle size of ≤250 µm and stored in an air-tight container for further use. Physicochemical properties of AW are shown in Table 1. Moisture content, total solid, volatile solid and ash content were determined using the EPA method (EPA, 2001). Total carbon and nitrogen were analysed using the CNS TruMac analyser (LECO, USA). Cellulose, hemicellulose and lignin content were determined upon neutral detergent fibre, acid detergent fibre and acid detergent lignin analysis. Total surface negative charge was determined using the modified Boehm (1966) method described in Lima and Marshall (2005b). Point of zero charge was determined using the pH drift method as described in Zaini et al. (2009). The surface morphology before and after sorption experiment (Figure 1[a]-[h]) was observed using scanning electronic microscope (JSM-6400, JOEL, Japan).

Metal Solution Preparation

A Cr(VI) stock solution of 1000 mg/L was prepared by dissolving 2.282 g of $K_2Cr_2O_7$ metal salt in 200 mL of ultrapure (Type 1) water (18.2 MΩ at 25°C) and brought up to 1000 mL with ultrapure (Type 1) water. Working Cr(VI) solutions with different initial metal concentrations (MC) were freshly prepared by diluting the stock accordingly before use.

Table 1
The characteristics of C, S, CG and CM

Parameters	Value			
	C	S	CG	CM
Moisture content (% wet weight)	11.76	9.62	64.20	76.79
Total solid content (TS, % wet weight)	88.24	90.38	35.80	23.21
Volatile solid content (VS, % dry weight)	98.38	99.06	93.79	80.89
Ash content (% dry weight)	1.62	0.94	6.21	19.11
Total negative charge (mmol H+/g)	1.50	1.50	2.70	3.70

Table 1 *(continue)*

pH_{pzc}	4.10	6.80	6.80	7.30
Bulk density (g/cm^3)	0.20	0.26	0.33	0.38
Total carbon (TC, % dry weight)	31.20	33.20	30.10	25.22
Total nitrogen (TN, % dry weight)	0.45	0.01	0.51	1.13
Neutral detergent fibre (NDF, % dry weight)	67.28	76.61	69.28	74.63
Acid detergent fibre (ADF, % dry weight)	31.36	57.54	38.55	54.40
Acid detergent lignin (ADL, % dry weight)	4.13	11.15	4.41	22.15
Cellulose content (% dry weight)	27.23	46.39	34.14	32.25
Hemicellulose content (% dry weight)	35.92	19.07	30.73	20.23

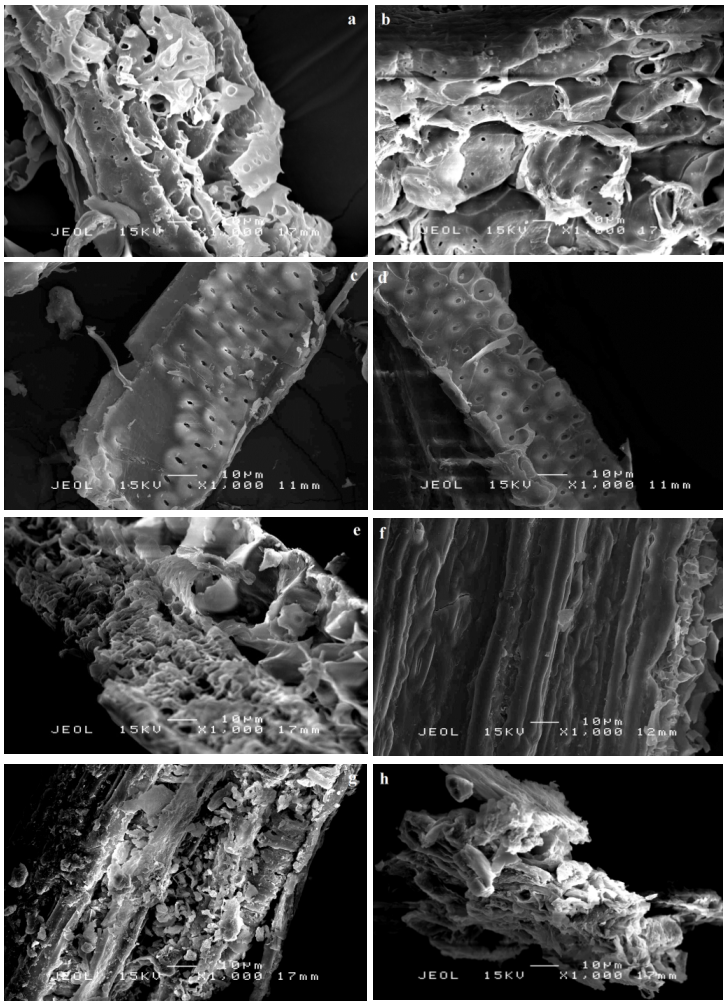


Figure 1. SEM image of (a) C before sorption, (b) C after sorption, (c) S before sorption, (d) S after sorption, (e) CG before sorption, (f) CG after sorption, (g) CM before sorption, (h) CM after sorption (× 1000)

Batch Biosorption Studies

Biosorption efficiency of C, S, CG and CM on Cr(VI) was carried out (i) in low (20 mg/L) and high (200 mg/L) concentrations of Cr(VI) solution at pH 4 (ii) at pH 2 and pH 4 in 200 mg/L Cr(VI) solution. The MC of 200 mg/L was employed because it is the upper limit of the majority of real-life conditions (Masood & Malik, 2011; Shazili et al., 2006). Biosorption was also carried out at 10 times lower MC (20 mg/L) to investigate the performance variation of biosorbents in low MC. Both pH 2 and pH 4 were selected as the initial pH of the solution because the pH of the effluents from the two major contributors of Cr(VI) i.e.: electroplating and tannery industries range from pH 2.2-pH 4.2 (Chowdhury et al., 2015; Vikramjit et al., 2016). Both MC and solution pH were set close to real-life conditions to ensure a practical and applicable biosorbent was identified from this study. Five hundred milligram of biosorbent in 50 mL of metal solution was allowed to equilibrate at 25±2°C for 24 h in both (i) and (ii). Control experiments were carried out without sorbents, and were taken as initial MC. The biosorption process of CM was further investigated in subsequent analysis. A kinetic experiment was performed in 200 mg/L Cr(VI) solution at 25±2°C with CT of 15-1440 min. Sorption isotherm and thermodynamic studies were carried out in MC range of 20-300 mg/L at 25±2°C, 35±2°C and 45±2°C with an equilibrium time of 480 min. The pH was adjusted using 0.1 M HNO3 and 0.5 M NaOH, and measured using a pH meter (Mettler Toledo Ross FE20, USA). The mixture was stirred at 350 r/min using a Teflon-coated magnetic stir bar on a stirring hotplate (Fisher Scientific, USA). It was then filtered through a 0.22 µm PVDF filter (JET BiofilFPV-213-000, China) with the aid of a vacuum pump (Rocker 300, Rocker Scientific Co. Ltd.). The final Cr(VI) concentration after sorption was analysed by ICP-OES (Perkin Elmer Optima 8300). The percentage removal and uptake of Cr(VI) was calculated using Equation [1] and Equation [2]. Graphs were drawn and analysis was done using GraphPad Prism Version 5.02.

$$Percentage\ removal\ (\%) = \frac{(C_0 - C_e)}{C_0} \times 100 \quad [1]$$

$$Uptake, q_e(mg/g) = \frac{(C_0 - C_e) \times V}{m} \quad [2]$$

where C_0 and C_e are the initial and final MC (mg/L), respectively, V is the volume of the metal solution (mL) and m is the amount of biosorbent used (g).

RSM Experimental Design

Biosorption of Cr(VI) by CM was optimised by the four-factor and three-level face-centred central composite design (FCCD) generated using Design Expert 7.0.0 software. The independent variables concerned were the initial pH of the metal solution (x_1, pH), CT (x_2, min), SR (x_3, r/min) and SD (x_4, g). A set of 30 experimental runs with 16 factorial points, eight axial points, and six replicates at the centre point was produced in the FCCD. Experimental runs were conducted randomly in batch mode at room temperature (RT) (25±2°C) in 200 mg/L Cr(VI) solution. The coded and actual values, as well as the range of the independent variables are shown in Table 2. The dependent variable i.e. the percentage removal (%) of Cr(VI) was related to independent parameters through a quadratic polynomial generated from regression analysis.

The generalised quadratic polynomial is shown in Equation [3]:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j \quad [3]$$

where Y is the dependent response (percentage removal of Cr(VI)), β_0 is the intercept, β_i, β_{ii}, β_{ij} are linear, quadratic and interaction coefficients, respectively; x_i and xj are independent variables in actual unit and k is the number of tested independent variables.

Table 2
The coded and actual values of independent variables

Independent Variables	Unit	Symbol	Range and Level (Coded)		
			-1	0	+1
pH	pH	x_1	1	3	5
Contact time	min	x_2	15	157.5	300
Stirring rate	r/min	x_3	60	205	350
Biosorbent dosage	g	x_4	0.1	0.55	1.0

Desorption and Regeneration

Cattle manure (CM) (0.5 g) was allowed to equilibrate in 50 mL of 200 mg/L metal solution at 25±2°C, pH 2 and 350 r/min for 24 h. The sorbent was then recovered and immersed in 50 mL desorbing agent (0.1 M NaOH) for 30 min, 5 h and 24 h at 25±2°C. The adsorption-desorption cycle was repeated three times to test the regenerability of the sorbent. Desorption efficiency was calculated using Equation 4.

$$\% \text{ desorption} = \frac{C_d}{C_0 - C_e} \times 100 \quad [4]$$

where C_0, C_e and C_d are the initial, equilibrium and desorbed MC (mg/L), respectively.

Quality Assurance

All glassware and plasticware were soaked in 2% HNO_3 overnight and rinsed five times with ultrapure water before use to prevent the contamination of metal ions present in the environment.

RESULTS AND DISCUSSION

Biosorption Efficiency of Biosorbents

The intrinsic biosorption efficiency of C, S, CG and CM in both low (20 mg/L) and high (200 mg/L) Cr(VI) solutions is shown in Figure 2. The sorbents did not undergo surface chemistry and structural modification by either thermochemical (pyrolysis) or chemical pre-treatment. The experimental conditions for the biosorption efficiency tests of the sorbents was kept the same to ensure a fair comparison of the results. At pH 4, higher biosorption was observed in 20 mg/L Cr(VI) solution for all sorbents tested (22.89% for C, 32.50% for S, 54.09% for

CG and 72.53% for CM). As the sorbents' binding sites to metal ions ratio is higher, lower competition increases the chance of Cr(VI) adhering to the binding sites. Biosorption of 200 mg/L Cr(VI) at pH 4 was extremely low for all the sorbents tested (9.49% for C, 14.08% for S, 18.25% for CG and 26.23% for CM), while increased sorption was observed when the pH decreased to pH 2 (35.94% for C, 35.64% for S, 38.37% for CG and 58.68% for CM). In acidic conditions, Cr(VI) exists as $Cr_2O_7^{2-}$ and $HCrO_4^-$ anionic form, and surface protonation occurs when pH of the solution is less than pH_{pzc}; thus, higher sorption was observed at pH 2. Biosorption of CM was significantly higher than C, S and CG, while C exhibited the lowest biosorption performance in both low and high MC and initial pH. Distinctive sorption efficiency may be related to physicochemical characteristics and surface morphology of the sorbents. Although the indigestible fibre content for all types of AW is high (NDF>67% dry weight), their lignocellulose components are different (Table 1). Hemicellulose is the major fibre component of C, while cellulose is the major component in S and CM. CG is composed of almost equal amounts of cellulose and hemicellulose. CM consists of the highest amount of lignin, double the amount of lignin in S and five times those in CG and C. High lignin content in CM could be due to the concentration of lignin after the digestion of cellulose and hemicellulose in the rumen of cattle (Harman et al., 2007). Lignin is an extensively-branched polyphenolic polymer composed of various functional groups responsible for the sorption of Cr(VI) such as hydroxyl, carbonyl and carboxyl groups. Dupont et al. (2004) demonstrated that Cr preferably adhered to microstructures rich in lignin.

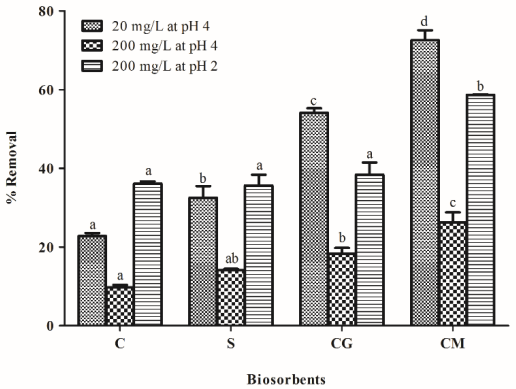


Figure 2. Biosorption efficiency of C, S, CG and CM in 20 mg/L Cr(VI) solution at pH 4 and 200 mg/L Cr(VI) solution at pH 2 and 4 (SD: 0.5 g, SR: 350 r/min, temperature: 25°C, CT: 24 h). Different letters indicate significant difference ($p<0.05$) between sorbents in the same MC and initial pH. Values are mean ± SD (n=3)

A similar finding was reported whereby lignin played a significant role in the biosorption of Cr, but no appreciable sorption was observed in pure cellulose (Dupont & Guillon, 2003). The study claimed that lignin acts as an electron donor and reduces Cr(VI) to Cr(III) in acidic conditions, while oxidation of lignin results in the increase of hydroxyl and carboxyl functional groups, which promotes the adhesion of Cr(III) to sorbents. This indicates that lignin plays the major role in the sorption of Cr(VI), and therefore explains the high biosorption efficiency

of CM. Besides lignocellulosic content, the surface morphology also influenced the sorption performance (Figure 1). Although S is composed of significantly higher lignin content than CG, it did not show greater sorption efficiency. This may have been due to the smooth surface of S although defined pores were observed, while CG possessed uneven surface and smaller pores, thus increasing the overall surface area and strengthening the adhesion between Cr ions and sorbent. Rough surface was also observed in CM, further explaining the high sorption performance. The surface of C consisted of observable pores, but the overall surface was sleek. In addition, protonation occured less significantly in C than in the other sorbents as the pH_{pzc} of C was 4.10, which resulted in low Cr(VI) sorption. The synergistic effect of high lignin content, high pHpzc and the rough surface morphology of CM was exemplified with high intrinsic sorption efficiency for Cr(VI).

Application of RSM

Determination of parameters range of operation. Before proceeding to RSM, a screening process was done to determine the region of interest and to select the suitable range of operation for the independent variables involved i.e. initial pH, CT, SR and SD. The effect of CT (15, 30, 60, 90, 180, 300, 480, 600, 840 and 1440 min) and SD (0.1, 0.25, 0.5, 0.75 and 1.0 g) were investigated in 50 mL of 200 mg/L Cr(VI) solution at pH 2 and RT. The respective % removal of Cr(VI) at different time intervals was 29.15%, 36.30%, 38.95%, 39.58%, 46.15%, 53.51%, 54.33%, 55.02%, 59.41% and 60.43%. The corresponding quantity SD range that was studied gave 30.58%, 50.30%, 62.62%, 79.89% and 84.96% of sorption. The ranges of 15-300 min and 0.1-1.0 g of SD were selected as rapid changes and the region of interest fall within this range. The range of initial pH was set at pH 1-5 because the sorption of Cr(VI) happens effectively at lower pH values (pH<3) (Saha et al., 2012). An SR with the range of 60-350 r/min was selected as higher SR increases the cost of operation, making it economically undesirable.

Model fitting. The 30 experimental runs of the four-factor three-level FCCD and their output responses (% removal) are shown in Table 3. The data were analysed through cubic, quadratic, two-factor interaction and linear models. A suitable model was selected based on a few important factors i) highest order polynomial where additional terms are significant, ii) not aliased, iii) insignificant lack-of-fit and iv) highest adjusted and predicted R-squared. A quadratic model was selected as it fulfilled all the aforementioned aspects (adjusted R^2=0.9635, predicted R^2=0.8819). The selected quadratic model was improved using backward elimination regression with alpha to exit equals to 0.1. Backward elimination bypassed terms with the highest p-value one by one until the next term satisfied the alpha criterion. SR was eliminated throughout the process as it showed insignificant effect and was not required to support the hierarchy of the model. Table 4 shows the analysis of variance (ANOVA) for the reduced quadratic model. The lowest and highest percentage removal was 15.90% and 74.63%, respectively. The biosorption of Cr(VI) was highly influenced by the initial pH, CT and SD ($p<0.0001$). Furthermore, the quadratic effects of pH and CT as well as the interactive effect of pH and SD also affected biosorption significantly ($p<0.0001$). The reduced quadratic model was statistically significant ($p<0.0001$) with a high R^2 (0.9788) and adjusted R^2 (0.9733). The predicted R^2 (0.9601) was in reasonable

agreement with the adjusted R^2 and had high adequate precision (48.649). This indicated that the model gave a good prediction on the response. Furthermore, the lack-of-fit test was also insignificant ($p>0.05$); this confirmed that the empirical relationship between the response and the variables involved can be sufficiently described through the following polynomial:

$$\% \text{ Removal} = 28.10 - 13.53x_1 + 0.16x_2 + 54.14x_4 - 8.80x_1 x_4 + 1.99x_1^2 - (3.75E\text{-}004)x_2^2$$

Table 3
Face-centred central composite design (FCCD) of four independent factors along with output responses

Standard	Run	Independent Variables				Percentage Removal (%)
		x_1	x_2	x_3	x_4	
14	1	5	15	350	1.00	20.65
5	2	1	15	350	0.10	21.65
6	3	5	15	350	0.10	17.20
9	4	1	15	60	1.00	65.13
13	5	1	15	350	1.00	66.46
2	6	5	15	60	0.10	15.90
10	7	5	15	60	1.00	20.25
19	8	3	15	205	0.55	23.95
1	9	1	15	60	0.10	20.35
25	10	3	157.5	205	0.55	34.80
17	11	1	157.5	205	0.55	62.70
30	12	3	157.5	205	0.55	37.55
28	13	3	157.5	205	0.55	33.40
23	14	3	157.5	205	0.10	21.75
21	15	3	157.5	60	0.55	38.00
26	16	3	157.5	205	0.55	38.55
18	17	5	157.5	205	0.55	30.60
29	18	3	157.5	205	0.55	37.95
24	19	3	157.5	205	1.00	50.96
22	20	3	157.5	305	0.55	38.60
27	21	3	157.5	205	0.55	36.55
8	22	5	300	350	0.10	25.40
16	23	5	300	350	1.00	37.95
4	24	5	300	60	0.10	23.90
15	25	1	300	350	1.00	74.63
11	26	1	300	60	1.00	73.07
3	27	1	300	60	0.10	36.35
20	28	3	300	205	0.55	38.15
12	29	5	300	60	1.00	37.85
7	30	1	300	350	0.10	39.85

x_1: initial pH; x_2: CT (min); x_3: SR (r/min); x_4: SD (g)

Table 4
Analysis of variance (ANOVA) for reduced quadratic model

Source	SS	df	MS	F Value	Prob > F	Significance
Model	7744.83	6	1290.80	177.28	<0.0001	*
x_1	2951.17	1	2951.17	405.32	<0.0001	*
x_2	742.67	1	742.67	102.00	<0.0001	*
x_4	2802.26	1	2802.26	384.87	<0.0001	*
x_1x_4	1004.57	1	1004.57	137.97	<0.0001	*
x_1^2	219.32	1	219.32	30.12	<0.0001	*
x_2^2	200.03	1	200.03	27.47	<0.0001	*
Residual	167.46	23	7.28			
Lack of Fit	147.56	18	8.20	2.06	0.2174	NS
Pure Error	19.90	5	3.98			
Cor Total	7912.29	29				
Standard Deviation	2.70		R-squared		0.9788	
Mean	37.34		Adj R-squared		0.9733	
C.V. %	7.23		Pred R-squared		0.9601	
PRESS	315.34		Adeq precision		48.649	

x_1: initial pH; x_2: CT (min); x_4: SD (g); SS: sum of squares; df: degree of freedom; MS: mean square; NS: not significant; * indicates significant

Effect of single factor. The perturbation graph shows and compares the effect of each factor on the percentage removal of Cr(VI) at the middle level in the design space (Figure 3a). Initial pH of the solution affected the biosorption most significantly, as the percentage removal increased from 30.60% to 62.70% when the pH decreased from 5 to 1 while the other factors remained constant. The initial pH of the solution had a high influence on biosorption because the pH not only affected the surface characteristic of the sorbent, but also changed the speciation of chromium ions. At pH<6, the dominant Cr(VI) species is $Cr_2O_7^{2-}$ and $HCrO_4^-$ anion (Rai et al., 1989). Meanwhile, the surface of the sorbent was protonated at pH<pH_{pzc} (pH_{pzc}=7.30), causing the surface to become positively charged. The dominant anionic species $HCrO_4$- and $Cr_2O_7^{2-}$ were therefore highly attracted to the positively-charged surface of the sorbent at low pH. As the pH increased, the degree of protonation decreased, making the surface less positive, which in turn reduced the percentage removal. Initial pH of the solution had little effect on biosorption at pH>4.

The SD affected percentage removal to the same extent as that of the initial pH of the solution. The percentage removal of Cr(VI) increased linear from 21.75% to 50.96% when the SD increased 10 times from 0.1 g, while the other factors remained constant. This is because there were more available sorption sites for the metal ions, creating a less competitive

environment as the SD increased; thus, more metal ions were successfully attached to the sorbent. The percentage removal had not reached saturation in the studied SD range.

Followed by initial pH of the solution and SD, the percentage removal of Cr(VI) was significantly influenced by CT. Biosorption happened rapidly from 15 min to around 240 min, where the percentage removal increased from 23.95% to 38.29% and reached a plateau around 240 min onwards and recorded 38.15% at 300 min. CT did not have a significant interactive effect with both solution pH and SD. Thus, the same trend was observed at different initial pH and SD ranges where biosorption was completed at around 240min.

The SR was not included in the reduced quadratic model as it did not have a significant effect on the biosorption. Although SR was expected to increase the kinetic energy of metal ions and enhance diffusion of metal ions through the sorbent, the percentage removal had a negligible effect when the SR increased from 60 to 350 r/min in the whole range of initial pH and SD studied. SR also did not accelerate the sorption process. This indicated that vigorous agitation is not required to enhance the biosorption, but minimum agitation which provides enough mixing of the mixture was adequate. This is good as increasing agitation speed would definitely increase the cost of operation.

Interaction between factors. The interaction between the solution's initial pH and SD was highly significant ($p<0001$). Interaction plots showed the interaction between these two variables at 157.5 min and 60 r/min (Figure 3b). The combined effect of low pH and high SD accounted for the highest percentage removal of Cr(VI) (78.31% removal at pH 1 and 1.0 g). However, the synergistic effect was greatly reduced as the pH increased from 1 to 5. At pH 1 and 3, the percentage removal increased from 37.50% to 78.31% and 21.75% to 50.96%, respectively when the SD was raised from 0.1 g to 1.0 g. The sorption efficiency was doubled at both pH 1 and 3 when SD was 10 times higher. However, the increment in sorption efficiency caused by higher SD dropped dramatically at pH=5. Only 27.74% removal was observed in 0.1 g sorbent and this increased slightly to 36.85% in 1.0 g of sorbent. This phenomenon occurred because protonation happened at low pH, causing the surface of the sorbent to be positively charged, aiding the sorption of anionic species of Cr(VI), i.e. $HCrO_4^-$ and $Cr_2O_7^{2-}$. Therefore, when the SD increased at low pH, all the active sites were adequately protonated; there were more functional positively-charged active sites available for sorption. However, when the pH shifted towards pH 5, the degree of protonation slowed down, hence, even when the SD was higher, the surface of the sorbent was not available for metal ions as they were not sufficiently protonated. The non-protonated active sites were unable to adsorb the anionic species due to the repulsive force between like charges.

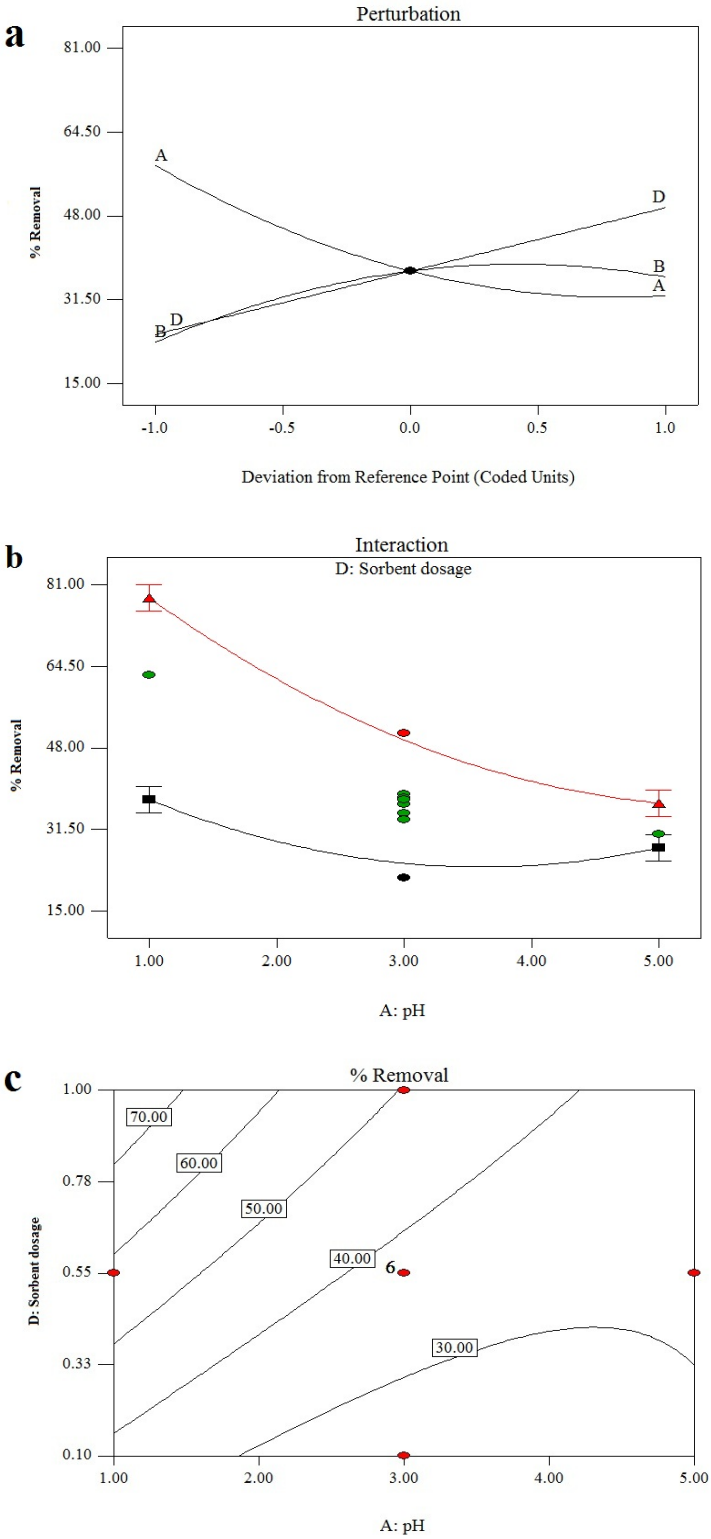


Figure 3. (a) Perturbation graph of the effect of single variables (b) Interactive effect of pH and SD and (c) Contour plot shows the optimum condition for the biosorption of Cr(VI) at the middle level in the design space

Response optimisation and validation experiment. The contour plot and numerical optimisation function in Design Expert 7.0.0 software can be used to predict the optimum condition. The optimum operational range was roughly estimated through the contour plot and the exact optimum range was determined through numerical optimisation function. The goal (maximise, minimise, target to or in range) was set for each variable and output response, and a list of solutions with desirability were generated. In the optimisation process, SR was minimised (60 r/min) since this variable had little effect on biosorption. Through numerical optimisation and the analysis of the contour plot generated as shown in Figure 3c, the maximum range of percentage removal (≥70%) occurred at the following optimum range: initial pH 1-1.56, 0.79-1 g of sorbent and 57-300 min CT. The specific optimum operating condition for maximum removal was then determined by setting the goal for initial pH, SD and CT within the optimum range as stated above. The predicted maximum percentage removal (78.43%) could be achieved at pH 1.04, 212.37 min, 60 r/min and 0.99 g of sorbent in 50 mL of 200 mg/L metal ion solution with desirability of 1. In order to validate the quadratic model, the solutions suggested by software were carried out experimentally (Table 5). In all three solutions, the goal for all the variables was set within the experimental range except when the SR was minimised. The output response for solution 1 and 2 was maximised, while it was set within the range of 50%-74.63% in solution 3. The experimental results were close to the predicted value with residual less than 4.65% in all the three solutions tested. A chi-square test (X^2) was carried out to determine the goodness-of-fit between the experimental and predicted values at the degree of freedom = 2. The results were accepted as significant when X^2 was more than the critical value at the 0.05 level of significance (5.991). The chi-square (0.47) obtained was smaller than the critical value (5.991), which indicated that there was no significant difference between the experimental and predicted values ($p>0.05$). In other words, the model was able to give a valid prediction.

Table 5
Response optimisation and model validation

Solution	x_1	x_2	x_3	x_4	D	Percentage Removal (%)		R	X^2
						Est.	Exp.		
1	1.04	157.97	60	0.99	1.000	77.21	79.00	1.79	0.47
2	1.15	212.11	60	0.96	1.000	75.25	75.07	0.18	
3	1.07	162.19	60	0.40	1.000	50.49	55.14	4.65	

x_1: initial pH; x_2: CT (min); x_3: SR (r/min); x_4: SD (g); D: desirability; Est.: estimated; Exp.: experimental; R: residual; X^2: Chi-square

Biosorption Kinetics

The pseudo-first-order equation is based on Lagergren's first-order equation (Lagergren, 1898), in which metal ions are assumed to bind to only one binding site and the adsorption rate is

proportional to the number of available sorption sites as expressed in Equation [5]. The integral and linear form is shown in Equation [6].

$$\frac{dq_t}{dt} = k_1(q_e - q_t) \quad [5]$$

$$log(q_e - q_t) = log\, q_e - \frac{k_1}{2.303}t \quad [6]$$

where t is the CT (min), q_e and q_t are the metal uptake of sorbent at equilibrium and at time t (mg/g), respectively, and k_1 is the pseudo-first-order rate constant (min^{-1}). Estimated q_e and k_1 were determined from the slope and intercept of the graph of log (q_e - q_t) against t (Figure 4a).

The pseudo-second-order reaction, in contrast, assumes metal ions are adsorbed to two binding sites (Ahmady-Asbchin et al., 2015). The relationship between the driving force, (q_e - q_t) and the adsorption rate is expressed in Equation [7], and its integral and linear form is shown in Equation [8].

$$\frac{dq_t}{dt} = k_2(q_e - q_t)^2 \quad [7]$$

$$\frac{t}{q_t} = \frac{1}{q_e{}^2 k_2} + \frac{1}{q_e}t \quad [8]$$

where k_2 is the pseudo-second-order rate constant (g mg^{-1} min^{-1}). The slope and intercept of the straight line from the plot of t/q_t against t (Figure 4b) were used to obtain k_2 and estimated q_e.

The parameters of both the pseudo-first-order and pseudo-second-order models are shown in Table 6. The biosorption data were best fitted in the pseudo-second-order with R^2=0.9964 rather than the pseudo-first-order, in which the R^2 was only 0.9427. Besides that, the good agreement of the experimental q_e (12.09 mg/g) to the estimated q_e (12.00 mg/g) from the pseudo-second-order further supported the conclusion that the sorption process is best described by pseudo-second-order model.

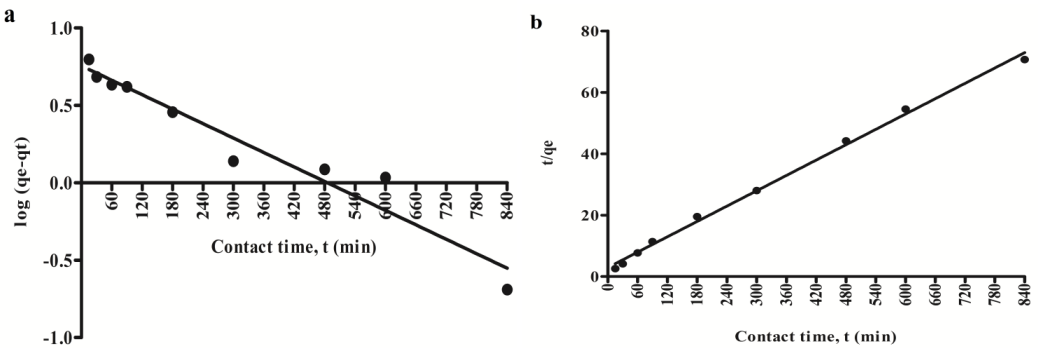


Figure 4. (a) Pseudo-first-order and (b) pseudo-second-order model for biosorption of Cr(VI) by CM (MC: 200 mg/L, pH: 2, SD: 0.5 g, SR: 350 r/min, temperature: 25°C)

Table 6
Kinetic parameters for the biosorption of Cr(VI) by CM

Kinetic models	Parameters	Value	R-squared
Pseudo-first-order	q_e (mg/g)	5.70	0.9427
	k_1 (min^{-1})	0.0037	
Pseudo-second-order	q_e (mg/g)	12.00	0.9964
	k_2(g mg^{-1} min^{-1})	0.0023	
	h (mg g^{-1} min^{-1})	0.3386	
Experimental	q_e (mg/g)	12.0860	NA

NA: not available

Biosorption Isotherm

Adsorption isotherm describes the equilibrium relationship between sorbate and sorbent. It also provides a better understanding of the sorption capacity and pathway involved. Langmuir and Freundlich isotherms are two widely used models in liquid-solid adsorption system.

Langmuir isotherm (Langmuir, 1916) presumes that the sorbent has a homogenous surface with a finite number of sorption sites that possess constant sorption affinity for the sorbate. It suggests monolayer adsorption, in which there is no interaction between adjacent adsorbed molecules. The Langmuir model and its linear form are expressed in Equation [9] and [10], respectively.

$$q_e = \frac{q_m b C_e}{1 + b C_e} \quad [9]$$

$$\frac{C_e}{q_e} = \frac{1}{b q_m} + \frac{1}{q_m} C_e \quad [10]$$

where q_e and C_e are the metal uptake of sorbent (mg/g) and final MC at equilibrium, respectively; q_m is the maximum monolayer sorption capacity (mg/g) and b is the Langmuir equilibrium constant related to sorption affinity (L/mg) obtained from the plot of C_e/q_e against C_e (Figure 5a).

The b value can be used to obtain the dimensionless separation factor through Equation [11]

$$R_L = \frac{1}{1 + b C_o} \quad [11]$$

where R_L is the dimensionless separation factor, b is the Langmuir constant and C_0 is the initial MC. R_L>1 indicates unfavourable sorption, 0<R_L<1 indicates favourable sorption, R_L=0 indicates irreversible sorption and R_L=1 is linear sorption.

The Freundlich isotherm (Freundlich, 1906), in contrast, is used to describe sorption on heterogeneous surfaces with varying sorption affinity. This model can be applied in a multilayer

sorption system, and it assumes that the sorption affinity decreases when a sorption site is being occupied (Vijayaraghavan et al., 2006). The relationship is expressed in Equation [12]

$$q_e = k_f C_e^{\frac{1}{n}} \quad [12]$$

The linear Freundlich model is shown in Equation [13]

$$log\ q_e = log\ k_f + \frac{1}{n}\ log\ C_e \quad [13]$$

where k_f is the Freundlich constant related to multilayer adsorption capacity and n is the Freundlich constant related to adsorption intensity. The slope and intercept of the graph of $log\ q_e$ against $log\ C_e$ (Figure 5b) were used to determine k_f and n, respectively.

The isothermal parameters of both models at different temperatures (298 K, 308 K and 318 K) are shown in Table 7. The q_m and k_f values increased with temperature. This may be due to the rise in kinetic energy, which improves the collision frequency between metal ions and the sorbent (Khan et al., 2012). In addition, the heat dilates the pores and slightly opens up the structure of the sorbents, which eases intraparticle diffusion within the pores (Meena et al., 2008). The n value obtained from the Freundlich model lay in the range of 1-10 (2.2753-2.6048), while the separation factor, R_L, was in the range of)-1 (0.03-0.45); both observations suggested favourable adsorption (Slejko, 1985; Treybal, 1980) in the whole range of the initial MC and temperature studied. Based on the correlated coefficient, the equilibrium relationship was better described by the Langmuir models (R^2>0.9882) compared to the Freundlich models (R^2<0.9523).

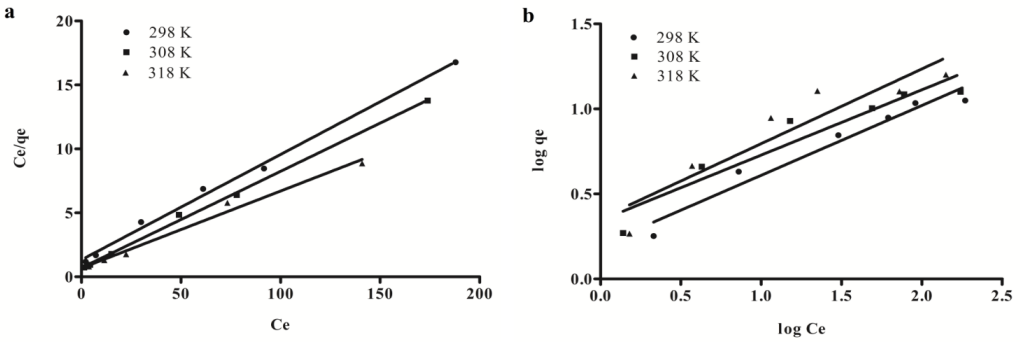


Figure 5. (a) Langmuir and (b) Freundlich isotherm for biosorption of Cr(VI) by CM at different temperatures (pH: 2, SD: 0.5 g, SR: 350 r/min, CT: 480 min)

Table 7
Isotherm parameters for the Biosorption of Cr(VI) by CM at different temperatures

Isotherm Models	Parameters	Value		
		298 K	308 K	318 K
Langmuir	q_m(mg/g)	12.14	13.26	16.60
	b (L/mg)	0.0621	0.1070	0.0888
	R^2	0.9950	0.9979	0.9882
	R_L	0.05-0.45	0.03-0.32	0.04-0.36
Freundlich	k_f[(mg/g)(L/mg)$^{1/n}$]	1.58	2.21	2.27
	n (g/L)	2.4349	2.6048	2.2753
	R^2	0.9523	0.9034	0.8622

Thermodynamic Studies

Thermodynamic parameters i.e. Gibb's free energy change (ΔG), enthalpy energy change (ΔH) and entropy energy change (ΔS) are related to each other in Equation [14] and ΔG is defined in Equation [15].

$$\Delta G = \Delta H - T\Delta S \quad [14]$$

$$\Delta G = -RT ln\, K \quad [15]$$

$$K = \left(\frac{C_s}{C_e}\right) \quad [16]$$

Equation [14] and [15] are combined and to yield the Van't Hoff equation as shown in Equation [17]

$$ln\, K = \left(\frac{\Delta H}{R}\right) - \left(\frac{\Delta S}{RT}\right) \quad [17]$$

where R is the universal gas constant (8.314 $J^{-1}mol\text{-}1K^{-1}$), T is the absolute temperature (K), C_s and C_e are the MC on the sorbent and the final MC remains in the solution at equilibrium (mg/g), respectively. ΔH and ΔS can be obtained from the Van't Hoff plot of $ln\, K$ against $1/T$.

The thermodynamic parameters are shown in Table 8. The negative value of ΔG indicated that the sorption process was spontaneous. ΔG was more negative in the lower initial MC and became less negative as the concentration increased, which revealed that the sorption process was more spontaneous in lower initial MC. This can be attributed to high availability of free sorption sites in relatively low MC solution. When the initial MC increased to 300 mg/L, the biosorption became unfavourable (ΔG became positive at 298 K and 308 K) as the binding sites of the sorbent had reached saturation. ΔG also reflected that the spontaneity was greater when the temperature increased. The positive value of ΔH indicated the biosorption process was endothermic. The positive value of ΔS indicated the increase in randomness of the sorbent-solution interface during the biosorption process (Ahmady-Asbchin et al., 2015).

Table 8
Thermodynamic parameters for the Biosorption of Cr(VI) by CM

Initial MC, C0 (mg/L)	Temperature (K)	ΔG (KJ mol^{-1})	ΔH (KJ mol^{-1})	ΔS (KJ mol^{-1} K^{-1})
20	298	-5.288	14.377	0.0668
	308	-6.670		
	318	-6.591		
50	298	-4.418	29.054	0.1129
	308	-6.072		
	318	-6.652		
100	298	-2.103	47.081	0.1657
	308	-4.417		
	318	-5.389		
150	298	-0.929	53.408	0.1814
	308	-1.856		
	318	-4.596		
200	298	-0.412	15.102	0.0523
	308	-1.146		
	318	-1.449		
300	298	1.280	24.966	0.0791
	308	0.823		
	318	-0.318		

Desorption and Regeneration

The amount of Cr(VI) desorbed after 30 min, 5 h and 24 h CT was 14.50 mg/L, 34.03 mg/L and 67.51 mg/L, respectively (Figure 6[a]). Since the concentration of NaOH could not be increased further as it would have destroyed the morphology of the sorbent (Wang et al., 2013), 24 h CT was chosen to desorb Cr(VI) in the regeneration study as it showed a higher desorption. Desorption efficiency in the first, second and third desorption cycle was 35.66%, 46.0% and 54.62%, respectively (Figure 6b). The adsorption efficiency decreased from 76.35% in the first cycle to 52.51% in the third cycle, suggesting that some Cr(VI) that had formed a stronger interaction with the binding sites were more difficult to desorb and they remained attached to the binding site; thus, fewer binding sites were available for subsequent sorption.

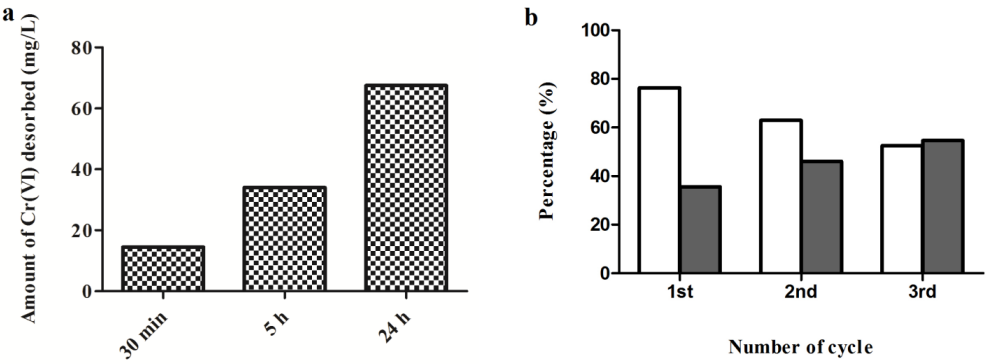


Figure 6. (a) Desorption of Cr(VI) using 50 mL of 0.1 M NaOH at different CT and (b) the adsorption (no filled column) and desorption (filled column) efficiency of CM loaded with Cr(VI)

Sorption Pathway

During rapid sorption in the first 30 min, ΔG in the range of -20 to 0 KJ mol^{-1} and $\Delta H<40$ KJ mol^{-1} in majority conditions indicated that physisorption was involved (Poulopoulos & Inglezakis, 2006; Weil, 1981). However, the endothermic nature observed wherein the q_m increased with temperature and a long equilibrium time (major sorption completed at 300 min) suggested that chemisorption also played a role in the sorption (Tran et al., 2016). Possible desorption confirmed that physisorption on the surface of the sorbents was involved. Nevertheless, low desorption efficiency and the significant drop in adsorption efficiency with the number of adsorption-desorption cycles confirmed that stronger chemical bonds were involved in addition to the physical bonding.

CONCLUSION

Biosorption of Cr(VI) using fresh CM without structural modification was examined in this study. Results showed that CM possessed significantly higher sorption efficiency compared to C, S and CG due to its fibrous, ligneous nature and rough surface morphology. The solution's initial pH, CT and SD significantly affected biosorption, while SR had negligible effect on biosorption. The combined effect of high SD and low pH was responsible for the highest sorption efficiency. The initial pH of the solution 1-1.56, 0.79-1 g of sorbent and 57-300 min of CT were able to remove ≥70% Cr(VI). Kinetic studies confirmed that the biosorption of Cr(VI) obeyed the pseudo-second-order model (R^2=0.9964). The equilibrium relationship between the sorbent and metal ions was well described by the Langmuir isotherm model (R^2>9882). The maximum monolayer sorption capacity of the CM was 12.14 mg/g at 298 K. Thermodynamic studies confirmed that the sorption process was spontaneous and endothermic. The recovery of Cr(VI) was possible and the CM was able to be reused at least three times with adsorption efficiency remaining >50%. Based on the combined observation of the kinetic, isotherm, thermodynamic and regeneration studies, it is suggested that the sorption of Cr(VI) involved both the physisorption and chemisorption processes. Intrinsically, the advantageous properties of CM (plentiful, low-cost, regenerable, fibrous, ligneous, rough surface morphology and

significantly high sorption for Cr[VI]) enabled it to be employed with minimum requirement of subsequent chemicals or thermochemical pre-treatment to enhance its sorption efficiency. Therefore, CM is a potential low-cost and an environmentally-friendly biosorbent for Cr(VI) removal.

REFERENCES

Ahmady-Asbchin, S., Safari, M., & Varposhti, M. (2015). Biosorption optimization of Cr(VI) using response surface methodology and thermodynamics modeling onto *Azolla filiculoides*. *Separation Science and Technology, 50*(4), 554–563.

Al-Rub, F., Kandah, M., & Al-Dabaybeh, N. (2002). Nickel removal from aqueous solutions using sheep manure wastes. *Engineering in Life Sciences, 2*(4), 111.

Bailey, S. E., Olin, T. J., Bricka, R. M., & Adrian, D. D. (1999). A review of potentially low-cost sorbents for heavy metals. *Water Research, 33*(11), 2469–2479.

Barakat, M. (2011). New trends in removing heavy metals from industrial wastewater. *Arabian Journal of Chemistry, 4*(4), 361–377.

Boehm, H. (1966). Chemical identification of surface groups. *Advances in catalysis* (pp. 179–274). New York: Academic Press.

Chen, S., Liao, W., Liu, C., Wen, Z., Kincaid, R., Harrison, J., Elliott, D., … & Stevens, D., (2003). *Value-added chemicals from animal manure-final technical report-contract DE-AC06-76RL0 1830*. US Department of Energy (USA).

Chowdhury, M., Mostafa, M., Biswas, T., Mandal, A., & Saha, A. (2015). Characterization of the effluents from leather processing industries. *Environmental Processes, 2*(1), 173–187. doi: 10.1007/s40710-015-0065-7

Dupont, L., Bouanda, J., Ghanbaja, J., Dumonceau, J., & Aplincourt, M. (2004). Use of analytical microscopy to analyze the speciation of copper and chromium ions onto a low-cost biomaterial. *Journal of Colloid and Interface Science, 279*(2), 418–424.

Dupont, L., & Guillon, E. (2003). Removal of hexavalent chromium with a lignocellulosic substrate extracted from wheat bran. *Environmental Science and Technology, 37*(18), 4235–4241.

EPA. (2001). *Method 1684 total, fixed, and volatile solids in water, solids, and biosolids* (pp. 1–13). US Environmental Protection Agency.

FAOSTAT. (2013). *Live animals' production*. Retrieved July 25, 2015, from http://faostat3.fao.org/download/Q/QA/E

Freundlich, H. (1906). Over the adsorption in solution. *Journal of Physical Chemistry, 57*(385), e470.

Fu, F., & Wang, Q. (2011). Removal of heavy metal ions from wastewaters: A review. *Journal of Environmental Management, 92*(3), 407–418.

Garg, U., Kaur, M., Garg, V., & Sud, D. (2008). Removal of nickel (II) from aqueous solution by adsorption on agricultural waste biomass using a response surface methodological approach. *Bioresource Technology, 99*(5), 1325–1331.

Gupta, V., Ali, I., Saleh, T., Nayak, A., & Agarwal, S. (2012). Chemical treatment technologies for waste-water recycling – An overview. *RSC Advances, 2*(16), 6380-6388.

Harman, G., Patrick, R., & Spittler, T. (2007). Removal of heavy metals from polluted waters using lignocellulosic agricultural waste products. *Industrial Biotechnology, 3*(4), 366–374.

Hofmann, N., & Beaulieu, M. (2006). *A geographical profile of manure production in Canada, 2001*. Statistics Canada, Agriculture Division.

IARC. (2012). Chromium (VI) compounds. IARC Monographs, 100 C, 147–167.

Jordão, C., Fernandes, R., de Lima Ribeiro, K., de Barros, P., Fontes, M., & de Paula Souza, F. (2010). A study on Al(III) and Fe(II) ions sorption by cattle manure vermicompost. *Water, Air and Soil Pollution, 210*(1–4), 51–61.

Jordão, C., Pereira, W., Carari, D., Fernandes, R., De Almeida, R., & Fontes, M. (2011). Adsorption from brazilian soils of Cu(II) and Cd(II) using cattle manure vermicompost. *International Journal of Environmental Studies, 68*(5), 719–736.

Khan, M., Ngabura, M., Choong, T., Masood, H., & Chuah, L. (2012). Biosorption and desorption of nickel on oil cake: Batch and column studies. *Bioresource Technology, 103*(1), 35–42.

Lagergren, S. (1898). About the theory of so-called adsorption of soluble substances. *Kungliga Svenska Vetenskapsakademies Handlingar, 24*(4), 1–39.

Langmuir, I. (1916). The constitution and fundamental properties of solids and liquids: Part I, solids. *Journal of the American Chemical Society, 38*(11), 2221–2295.

Lima, I., & Marshall, W. (2005a). Adsorption of selected environmentally important metals by poultry manure-based granular activated carbons. *Journal of Chemical Technology and Biotechnology, 80*(9), 1054–1061.

Lima, I., & Marshall, W. (2005b). Utilization of turkey manure as granular activated carbon: Physical, chemical and adsorptive properties. *Waste Management, 25*(7), 726–732.

Masood, F., & Malik, A. (2011). Biosorption of metal ions from aqueous solution and tannery effluent by Bacillus sp. FM1. *Journal of Environmental Science and Health, Part A, 46*(14), 1667–1674.

Meena, A., Kadirvelu, K., Mishra, G., Rajagopal, C., & Nagar, P. (2008). Adsorptive removal of heavy metals from aqueous solution by treated sawdust (*Acacia arabica*). *Journal of Hazardous Materials, 150*(3), 604–611.

Meng, J., Feng, X., Dai, Z., Liu, X., Wu, J., & Xu, J. (2014). Adsorption characteristics of Cu(II) from aqueous solution onto biochar derived from swine manure. *Environmental Science and Pollution Research, 21*(11), 7035–7046.

Miretzky, P., & Cirelli, A. (2010). Cr(VI) and Cr(III) removal from aqueous solution by raw and modified lignocellulosic materials: A review. *Journal of Hazardous Materials, 180*(1), 1–19.

Ngah, W., & Hanafiah, M. (2008). Removal of heavy metal ions from wastewater by chemically modified plant wastes as adsorbents: A review. *Bioresource Technology, 99*(10), 3935–3948.

Poulopoulos, S. G., & Inglezakis, V. J. (2006). *Adsorption, ion exchange and catalysis: Design of operations and environmental applications.* The Netherlands: Elsevier.

Rai, D., Eary, L., & Zachara, J. (1989). Environmental chemistry of chromium. *Science of the Total Environment, 86*(1–2), 15–23.

Saha, P., Dey, A., & Marik, P. (2012). Batch removal of chromium (VI) from aqueous solutions using wheat shell as adsorbent: Process optimization using response surface methodology. *Desalination and Water Treatment, 39*(1–3), 95–102.

Shazili, N., Yunus, K., Ahmad, A., Abdullah, N., & Rashid, M. (2006). Heavy metal pollution status in the Malaysian aquatic environment. *Aquatic Ecosystem Health and Management, 9*(2), 137–145.

Slejko, F. L. (1985). *Adsorption technology: A step-by-step approach to process evaluation and application*. Dekker New York: Basel.

Tran, H. N., You, S. J., & Chao, H. P. (2016). Thermodynamic parameters of cadmium adsorption onto orange peel calculated from various methods: A comparison study. J*ournal of Environmental Chemical Engineering, 4*(3), 2671–2682.

Treybal, R. (1980). *Mass transfer operations, 1980* (3rd Ed.). New York: Mc-Graw Hill Book Company.

Uchimiya, M., Lima, I., Thomas, K., Chang, S., Wartelle, L., & Rodgers, J. (2010). Immobilization of heavy metal ions (Cu(II), Cd(II), Ni(II), and Pb(II)) by broiler litter-derived biochars in water and soil. *Journal of Agricultural and Food Chemistry, 5*8(9), 5538–5544.

Vijayaraghavan, K., Padmesh, T., Palanivelu, K., & Velan, M. (2006). Biosorption of nickel (II) ions onto *Sargassum wightii*: Application of two-parameter and three-parameter isotherm models. *Journal of Hazardous Materials, 133*(1), 304–308.

Vikramjit, S., Chhotu, R., & Ashok, K. (2016). Physico-Chemical Characterization of Electroplating Industrial Effluents of Chandigarh and Haryana Region. *Journal of Civil and Environmental Engineering, 6*(4), 1-6.

Wang, J., Pan, K., He, Q., & Cao, B. (2013). Polyacrylonitrile/polypyrrole core/shell nanofiber mat for the removal of hexavalent chromium from aqueous solution. *Journal of Hazardous Materials, 244*, 121–129.

Water, U. N. (2009). *The United Nations world water development report 3 –Water in a changing world.* Paris: United Nations Educational Scientific and Cultural Organization.

Weil, K. G. (1981). MJ Jaycock, GD Parfitt: Chemistry of Interfaces. Ellis Horwood Limited Publishers, Chichester 1981. 279 Seiten, Preis:£ 27, 50. *Berichte der Bunsengesellschaft für physikalische Chemie, 85*(9), 718–718.

Zaini, M., Okayama, R., & Machida, M. (2009). Adsorption of aqueous metal ions on cattle-manure-compost based activated carbons. *Journal of Hazardous Materials, 170*(2), 1119–1124.

Pertanika J. Sci. & Technol. 26 (1): 215 - 224 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Building a Low Cost, Good Quality and Safe Infinity Mirror Room for Suroboyo Night Carnival Amusement Park

Rinda Hedwig[1]*, Andrew Nafarin[2], Ichwanhono Kosasih[2], Jimmy Linggarjati[1] and Wiedjaja Atmadja[1]

[1]Computer Engineering Department, Bina Nusantara University, Jl. K.H. Syahdan, Jakarta, Indonesia
[2]Srimulia Indo Pertiwi, Ltd., Jl. Kalisari, Surabaya, Indonesia

ABSTRACT

Amusement parks have been growing rapidly in Indonesia in the past five years but they seem affordable only for the middle and higher classes of society. Surabaya in East Java has several amusement parks that cater for low budgets and they have good-quality equipment and are safe to use. The purpose of this research is to show the results of building a low-cost amusement park. The amusement park was built by a player in the amusement park industry in collaboration with the Computer Engineering Department of Bina Nusantara University. An infinity mirror room (IMR) was built in one of the new amusement parks in Surabaya called Suroboyo Night Carnival. The full design of the IMR is discussed in this paper including the equipment used in the design. The IMR mainly uses fibre optics, LED and a mirror. The amusement park began operation on July 28, 2013, and the amount spent on one room was around US$8,500. No safety breaches have been reported. The facility has been able to attract 500 visitors on average from 2013 to 2016. The breakeven point of this facility was achieved in the first year of operation.

Keywords: Amusement park, infinity mirror room, fibre optics, low-cost, Surabaya

INTRODUCTION

According to Anand (1993), Clave (2007) and Milman (2010), entertainment is a human need. Amusement or theme parks provide pleasure for many (Nye, 1981; Lewi (2015), with many families flocking to visit them at weekends (Weinstein, 1992). This phenomenon is also true for Indonesia, a developing country. It has been observed that small amusement parks or trolley-parks that are less safe (Davis, 1996) have grown

ARTICLE INFO

Article history:
Received: 08 November 2016
Accepted: 11 September 2017

E-mail addresses:
rinda@binus.edu (Rinda Hedwig),
sip97@hotmail.com (Andrew Nafarin),
kosasih_ichwan@yahoo.co.nz (Ichwanhono Kosasih),
jimmyl@binus.edu (Jimmy Linggarjati),
steff@binus.edu (Wiedjaja Atmadja)
*Corresponding Author

ISSN: 0128-7680

in number in many places especially in urban areas, where most families come from the low-income segment of society.

Understanding the need for low cost, good quality and safe amusement parks that can cater for lower-income families, one player in the amusement park industry collaborated with the Computer Engineering Department of Bina Nusantara University to study, research and build an infinity mirror room (IMR) (Applin, 2012; Cutler, 2017; Kusama, 2013) in Surabaya, East Java. The IMR is an entertainment room that uses lights, mirrors and illusion to create special effects for the entertainment of visitors. The IMR in Surabaya was built similar to the one built by Ripley in Pattaya, Thailand (R. E. Inc., 2011) and to Infinity Gold Coast, Australia (I. Ltd., 2000), but at a lower cost while maintaining high quality. Before the project was undertaken, full research was carried out to ensure that the IMR would be built based on these requirements: safe design, low cost and good effects.

A scalable IMR was developed by the Computer Engineering Department, Bina Nusantara University. The laboratory work was funded by Srimulia Indo Pertiwi Ltd. We studied the effectiveness of light propagation and its reflectivity before finally building the IMR in Surabaya. The selection of light source, fibre optics and materials for the mirror walls are discussed in this paper with the objective that any developing country may also build low-budget yet safe and high quality IMRs if so desired. The IMR, as one feature of the amusement park, which offered a science exhibition, children's playground, racing arena and wax museum, among its features, was to be a room using illusion conjured through the interplay of light, mirrors and fibre optics to provide entertainment for visitors.

Since the IMR involved the use of mirrors and fibre optics, safety was a crucial issue. Therefore, it was important to use products that were strong and durable. In addition, visitors would be asked to use gloves and socks when entering the room. Closed circuit televisions would also be mounted in key areas including the emergency exit routes.

DESIGN AND IMPLEMENTATION

Our task was to build a dark chamber that could be used to create the illusion of infinity. We built a small mirror chamber measuring 40 cm × 40 cm × 40 cm that had holes drilled at random on one surface. Twenty strands of end-point fibre optics (Jeff, 2004) were attached in each hole and the total mirror surface covered with a vinyl sheet. The result can be seen in Figure 1. We were able to create an effect of stars sparkling in the deep space although lights were attached to only one side of the mirror as shown in Figure 2. The number of fibres to attach was a huge consideration in later implementations. The details can be seen in the block diagram of the experiment (see Figure 3).

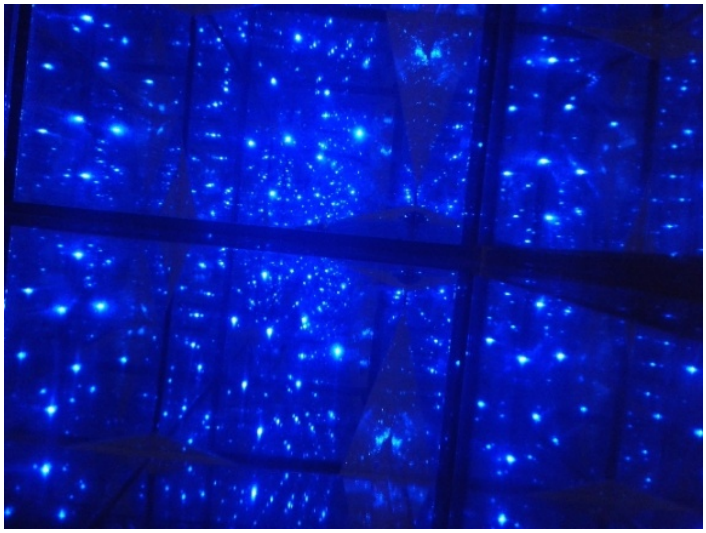


Figure 1. Infinity effect created in a mirror room

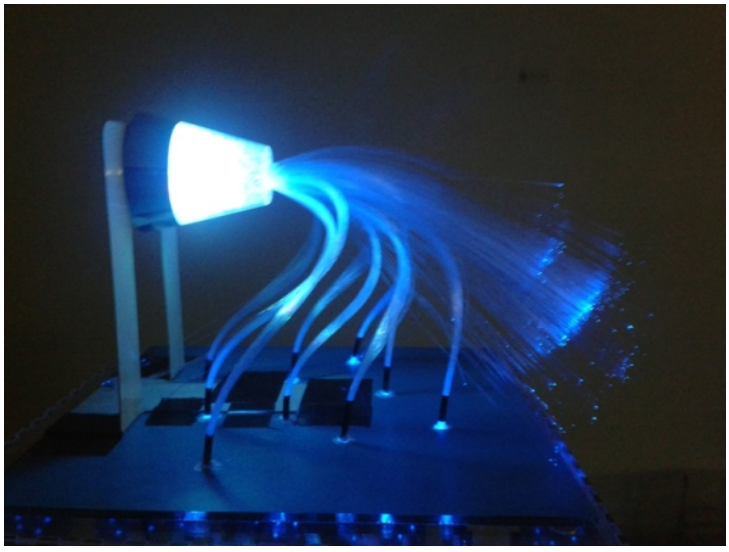


Figure 2. One side of the mirror chamber drilled and attached with 20 fibre optic strands per hole

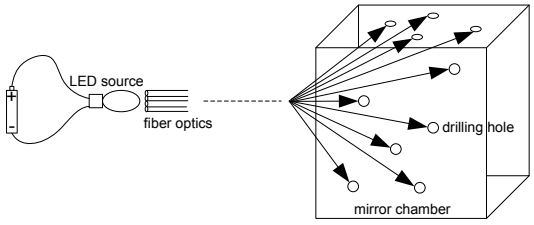


Figure 3. Block diagram of the experiment

In the early stage, we found that 5 mm hole was the smallest diameter that could be found in the Indonesian market. Thus, the number of fibre optic strands that should be filled in the hole had to be not less than 20 pieces. Due to this constraint, we chose a 2-mm acrylic mirror (Taylor, 1999) sheet so that we could drill holes that were 0.5 to 1 mm in diameter. The holes could be drilled in a specific pattern or in random design depending on the size of the mirror and how dense we wanted the light effect to be. Fibre optics imported from China, which came in small diameter sizes was used later in the actual chamber. The light source for the first experiment was a small LED powered by a 5-V battery, while the fibre optics were taken from children's toys and ornaments without detailed specification. Acrylic mirrors are safer than regular mirrors since they are not easily cracked in case of accidents. Its durability is due to the adhesive that holds the surface together.

Figure 4 is a map of the IMR. The infinity effect was focussed in the electron maze and star chamber room. The difference in the pattern of each room was determined by the name of the room. The electron maze room was designed to make people who entered the room feel like electrons were moving across the room, while the star chamber room was designed to show some of the constellations. However, the star chamber was only built later.

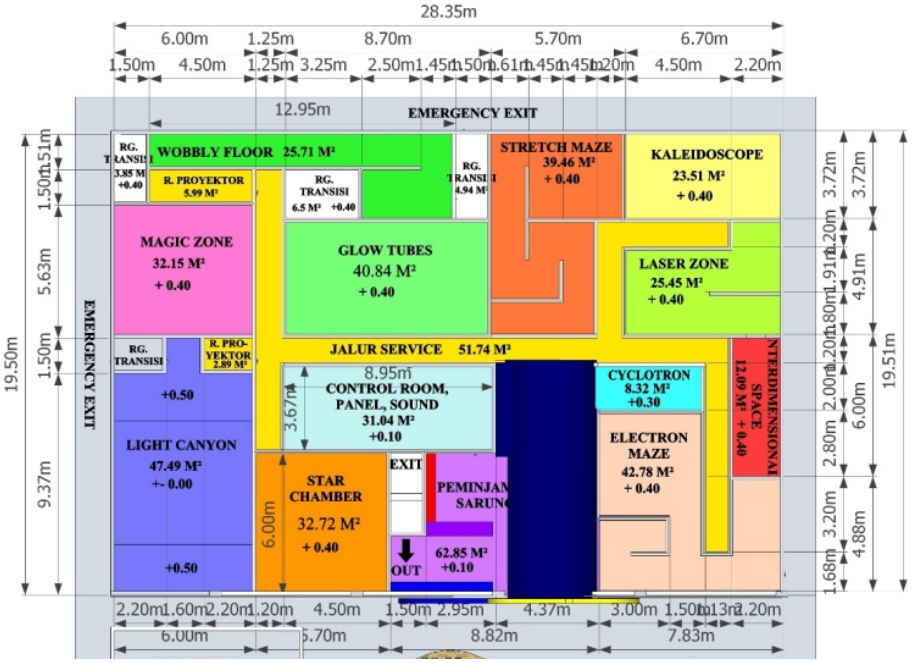


Figure 4. Infinity mirror room design

For additional safety, we applied tempered glass (USA Patent No. US 2177324 A, 1939) in front of the acrylic mirror so that visitors would not be in direct contact with the vinyl fibre optics or any other electric circuit exposed behind the mirror. We also used a tempered glass mirror for the floor so that a dramatic optic illusion of sparkling lights could be seen all around the room. As for the ceiling, we used an acrylic mirror sheet, which is lighter and more economical than using a tempered glass mirror (see Figure 5). These safety measures proved effective as no accidents were reported in the period 2013 to 2016 when the IMR was in operation. The mirror itself did not crack until the IMR was converted into a Kids' Play Station.

For the dramatic illusion of light, we used an LED fibre optic light engine, which produced a mix of red-green-blue colours and an end-side sparkling vinyl fibre optics. The items were supplied by a company in China; the quality of each item was good and the items were available at a lower cost. Building and instalment of the IMR took 3 months to complete.

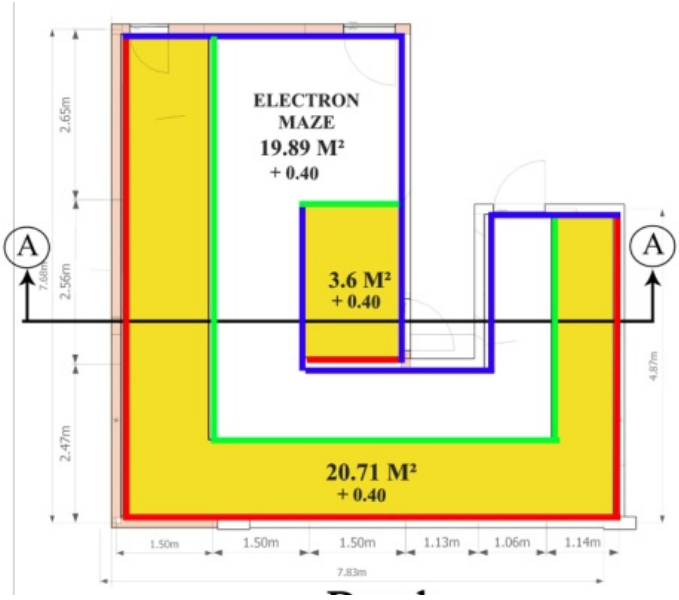


Figure 5. Detailed map of the electron maze chamber

DISCUSSION

Before finally building the IMR and installing the apparatus, we ran many tests to study and fine-tune the intended effect. We expected the light that was reflected from one side of the acrylic mirror to be refracted (Crowell, 2012) due to the fact that the surface of the acrylic mirror was not exactly flat. When we first designed the fibre attachment on the back side of the mirror, we expected that the laser ray would be reflected in the possible path shown in Figure 6. However, due to the fact that the surface was not flat, we could not get a good reflection of the laser light in the electron maze chamber. This interfered with the intended optical illusion

effect of infinity. To solve this problem, the number of fibre optics placed in the mirror was increased and the distance between the holes was made to be less than 1 cm. Each hole was fit only with a single strand of the fibre optics as shown in Figure 7.

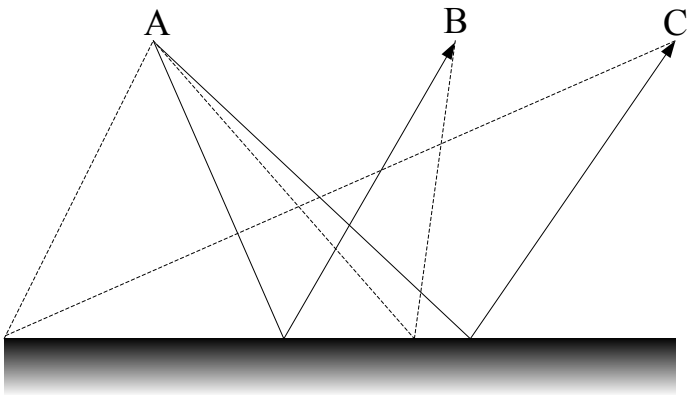
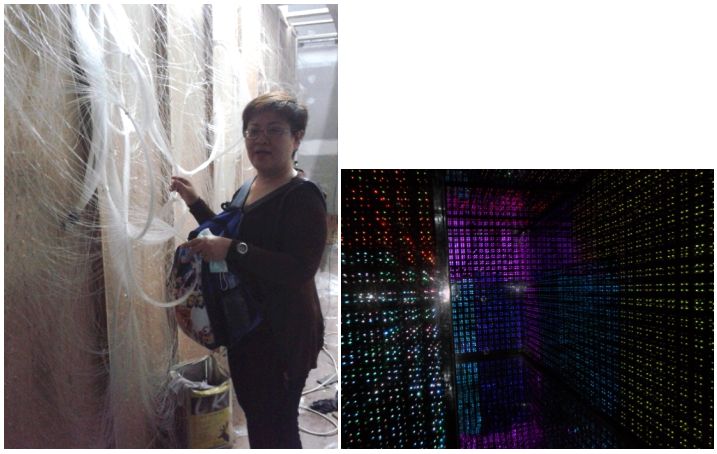


Figure 6. The solid lines are possible physical paths for light rays travelling from A to B or A to C (Crowell, 2012)

Although we had planted a great number of fibre optic strands in one mirror wall, numbering about 2,000 strands of fibre optics per each mirror wall of 2 m x 1 m, the reflection from the other side of the wall did not give as good an image as in our first experiment (see Figure 1). We used Side Sparking Fibre Optic, Mitsubishi fibre, 1250 pcs 4 m, 0.75 mm, 3 cm dots distance, connected by an engine connector. We also used an LED fibre optic light engine, six colours, with a remote controller, 220 V, UK plug as the light source. We found that the distance between the mirror planted with fibre optics and the two-way mirrors, which were placed parallel to one another, also played a crucial role, as shown in Table 1. Excellent reflection means that the reflection results were similar to those shown in Figure 1. The narrower the distance, the better the reflection we got. However, this condition could not be performed as the number of visitors entering the maze would have to be limited.

Table 1
Distance between Two-Way Mirror Walls

No.	Distance Between Two-Way Mirror Walls	Reflection Quality
1	2 metres	Poor/No reflection
2	1.5 metres	Poor/No reflection
3	1 metres	Fair
4	0.8 metres	Good
5	0.5 metres	Excellent
6	0.4 metres	Excellent
7	0.3 metres	Excellent



(a) (b)

Figure 7. (a) Fibre optics instalment behind the electron maze chamber (b) front view of the instalment

From this we understood that a good reflection could be achieved by planting 10 to 20 strands of fibre optics in one hole as shown in Figure 2 and using flat mirrors for the wall instead of acrylic mirrors. The infinity effect rested on the flatness of the mirrors' surface. The distance between two mirrors was also an important consideration in generating good continual reflectivity as expected.

Air circulation also played an important role in this design. A centralised air conditioner was chosen for use but having a ceiling duct would reduce the optical illusion. A completely dark room enhances light travelling across it, making it starkly visible to the eyes. Therefore, we used a split air conditioner that was placed behind the acrylic mirror in the service area. Good air circulation would ensure long life for both the LED engine and he fibre optics. Since there was also a small space between the ceiling and the acrylic mirror sheet, air could flow easily to fill the chamber for the comfort of the visitors. The light engine produced attractive patterns of different colours.

The first budget estimation for this project was US$12,500 (US$1 = Rp. 10000,-) for the imported products we used. The price was considered high due to the purchase tax that we had to pay. We managed to source for good products at a lower price, and were able to bring the cost down to US$8,500. This was our biggest expense, due mainly to the use of tempered glass for the mirrors.

We looked forward to constructing a model of the star chamber, which would be designed to show some of the constellations on the walls of the chamber. The number of visitors who visited the IMR can be seen in Figure 8; the average number fell to 500 visitors per month in 2014. The IMR itself reached breakeven point in the first year, but the number of visitors to the amusement park started to decrease in 2015. At the beginning of 2016, the new management of Suroboyo Night Carnival decided to remodel 30% of the amusement park's facilities, including the IMR. Most of the acrylic mirrors and fibre optics were still used but the chamber was transformed into a Kids' Play Station.

Cheaper entrance tickets did play a role in attracting visitors in the beginning but perhaps the attraction of the display faded after a while. The data revealed that the IMR had visitors did not revisit the IMR more than twice; this was one of the reasons why the new management demolished the IMR and some other facilities as well. Amusement parks need to provide new attractions every three years to draw public interest. It was considered easy and economical to demolish the facilities and construct new ones because the cost was low and the attractions could reach breakeven point quick quickly.

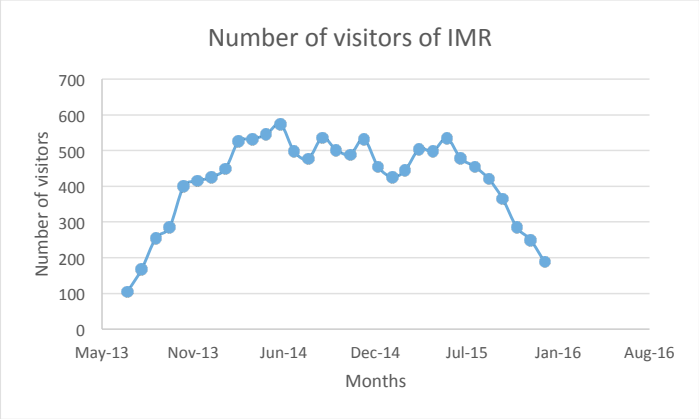


Figure 8. Number of visitors to the IMR

CONCLUSION

This research was simple but its results have wide reaching implications. The prototype can be used for students of physics to study reflection of light in a mirror chamber. The IMR also provides an experience of space and light for teachers and students. It is also a useful study aid for students of art, design and architecture. We also learned that the flatness of a mirror as well as the distance between two mirrors play important roles in generating continued reflectivity. The light source and the darkness of the room are also important considerations. This qualitative research can also benefit other developing countries that might desire to build an IMR for their amusement parks as it provides an affordable model using good quality products that are also safe.

The IMR built in the Suroboyo Night Carnival Amusement Park was inside the Rumah Kinclong building, which was visited by an average of 500 visitors per week. The ticket price was only US$1, which is affordable for low-income families. The IMR was safe; no accidents were reported among visitors or workers. However, the IMR lost its appeal among the public after about three years. Visitors were not interested to visit it more than twice. To retain public interest in its offerings, the new management of the amusement park closed the IMR in 2016, converting it into a Kids' Play Station.

ACKNOWLEDGEMENT

The authors would like to thank Srimulia Indo Pertiwi Ltd. for funding this research and building the IMR, 'Rumah Kinclong', as part of the Suroboyo Night Carnival Amusement Park.

REFERENCES

Anand, S., & Ravallion, M. (1993). Human development in poor countries: On the role of private incomes and public services. *The Journal of Economic Perspectives, 7*(1), 133–150.

Applin, J. (2012). *Yayoi Kusama: Infinity mirror room – Phalli's field*. London: Afterall Books, One Work Series.

Barsom, J. (1968). Fracture of tempered glass. *Journal of American Ceramic Society, 51*(2), 75–78.

Clave, S. A. (2007). *The global theme park industry*. Cambridge: CABI.

Crowell, B. (2012). *Light and matter, volume 1*. California: Benjamin Crowell.

Cutler, J. B. (2017). Narcissus, narcosis, neurosis: The visions of Yayoi Kusama. In J. Hirsh, *Contemporary art and classical myth*. Athens: Routledge.

Davis, S. (1996, July). The theme park: Global industry and cultural form. *Media, Culture and Society, 18*(3), 399–422.

I. Ltd. (2000). *Infinity*. Retrieved November 10, 2011, from http://www.infinitygc.com.au/

Jeff, H. (2004). *City of light: The story of fibre optics.* New York: Oxford University Press, Inc.

Kusama, Y. (2013). *Infinity net: The autobiography of Yayoi Kusama.* Tate Enterprises Ltd.

Lewi, H. (2015). Reviews: The architecture of pleasure: British amusement parks 1900–1939. *The Journal of the Society of Architectural Historians, Australia and New Zealand, 25*(2), 290–291.

Long, B. (1939). *USA Patent No. US 2177324 A*.

Milman, A. (2010). The global theme park industry. *Worldwide Hospitality and Tourism Themes* (pp. 220–237). United Kingdom, UK: CABI.

Nye, R. (1981). Eight ways of looking at an amusement park. *The Journal of Popular Culture, 15*(1), 63–75.

R. E. Inc. (2011). *Ripley's Pattaya Thailand*. Retrieved November 5, 2011, from http://www.ripleysthailand.com/index.php?p=maze

Taylor, C. (1999). *USA Patent No. US Patent 5864434A.*

Weinstein, R. (1992). Disneyland and Coney Island: Reflections on the evolution of the modern amusement park. *The Journal of Popular Culture, 26*(1), 131–164.

Pertanika J. Sci. & Technol. 26 (1): 225 - 234 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

The Value of 18F-Fluorodeoxyglucose PET (18F-FDG PET) and MRI Spectroscopy in Underpinning Suspicious Breast Cancer

Shazreen, S.[1]*, Shakher, R.[2], Shahrun Niza, A. S.[3] and Fathinul Fikri, A. S.[1]

[1]Departmentof Diagnostic and Nuclear Imaging, University Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Department of Nuclear Imaging, Monash University, Clayton VIC 3800, Australia
[3]Department of Surgical, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

ABSTRACT

The aim of this study was to evaluate the value of MRI spectroscopy and association with the altered glucose metabolism on 18-FDG PET/CT in patients with suspicious breast cancer. Eight selected breast cancer patients with BIRADS 4 or 5 on mammogram were recruited and patients underwent 18F-FDG PET/CT MRI (spectroscopy). The standardise uptake value (SUVmax) was analysed to determine the degree of the altered glucose metabolism on the PET. The metabolites of tumor lesions were measured using *in vivo* proton MR spectroscopy (MRS) of the breast. There were eight females with a mean age of 55.3±12.2 years with a biopsy result of invasive ductal carcinoma (2), lobular carcinoma (1) and benign lesion (5). There was a significant difference between the mean of the malignant tumour (SUVmax 4.28±3.74 g/ml) and the mean of the benign tumour (SUVmax: 2.33±0.9 g/ml). On the per-lesional basis of the MRS correlate with SUVmax, the suspicious breast tissue exhibited raised creatinine metabolites (mean: 3.39±0.54u) with significant correlation SUVmax mean 3.06±2.34 as compared to N-acetyl Aspartate (NAA), (mean: 2.84±0.99u) and choline (mean: 2.46± 0.70 u). This study showed that high SUVmax was associated with malignant cancer and the high creatinine metabolite that correlated with the SUVmax could potentially be utilised as a surrogate marker in detecting breast cancer.

Keywords: Magnetic Resonances Imaging Spectroscopy, Computed tomography, positron emission tomography, [18F]-fluorodeoxyglucose, multimodality

ARTICLE INFO

Article history:
Received: 16 December 2016
Accepted: 23 June 2017

E-mail addresses:
fir_reen@yahoo.com (Shazreen, S.),
sramdave@gmail.com (Shakher, R.),
shahrun72.sn@gmail.com (Shahrun Niza, A. S.),
ahmadsaadff@gmail.com (Fathinul Fikri, A. S.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Worldwide, breast cancer is becoming a leading cause of death. Mortality due to breast cancer increased from 20,100 deaths in 2008 to 21,700 deaths in 2012 worldwide, as reported by the International Agency for Research on Cancer (IACR) Globocan of the

World Health Organisation (WHO). The risk factors that contribute to breast cancer include hormone replacement therapy for a long period of time, being overweight especially after menopause and mutations in cancer-related genes (erbB2). Recent studies found that the development of cancer was due to choline intake (Zhang et al., 2013).

In breast cancer cells, the level of choline was found markedly higher compared to in normal human mammary epithelial cells (Aoyama et al., 2004). Both choline phosphorylation and transport were found to be augmented in human breast cancer cells (Cho et al., 2007). Overexpression of choline kinase (CK) will induce progression of human mammary epithelial cells from a normal to a malignant phenotype. These choline kinase will catalyse the phosphorylation of choline to form phosphocholine, followed by generation of phosphatidylcholine in tumor cell membranes (Molina et al., 2004). Recent studies proved that choline kinase, which transforms choline to phosphocholine (PCho), was increased in malignant tumours and caused progression of mammary epithelial cells *in vitro* (Ramirez et al., 2004).

In vivo proton MR spectroscopy (^{1}H-MRS) is a non-invasive technique that has great potential to generate tumor metabolism, which is a useful process in evaluation and diagnosis of the response to treatment of tumour progression (Meisamy et al., 2004). Recently, breast ^{1}H-MRS has been shown to improve cancer diagnosis based on elevated choline-containing compounds' (tCho) metabolite peak (Baek et al., 2008). Several researchers have found that 1.5T of *in vivo* ^{1}H-MRS can be used to detect and distinguish between malignant and benign cancer tissue (Baik et al., 2006). Nowadays, FDG ia used to detect tumour progression by providing useful information about tumor characterisation, staging, patient prognosis and monitor response of anticancer therapy to patients with suspected malignancies. Besides that, FDG uptake correlates with tumour growth and viability in recent studies (Fathinul Fikri et al., 2014). A combination of non-invasive imaging techniques of Positron Emission Tomography (PET) and MRI are beneficial in detecting the response to treatment to target tumours using the metabolics of choline. The aim of this study was to evaluate the value of metabolic MRI spectroscopy and its association with altered glucose metabolism on FDG-18 PET/CT in patients suspected of developing breast cancer tissue as evaluated by the BIRADS scoring system.

MATERIALS AND METHODS

Patient Selection

In the prospective group, eight selected breast cancer patients with mammogram results showed BIRADS 4 or 5, recurrent breast cancer and biopsy results with proven breast mass were recruited from an endocrinology clinic. Patients undergoing chemotherapy or who had undergone surgery were excluded from the studies. All the patients underwent 18F-FDG PET CT and MRI spectroscopy prospectively. All the patients who participated in this study gave written consent before commencement of the study.

Patient preparation. All the patients were required to fast from food consumption for at least 6 hr before examination. The fasting blood glucose level was measured in the morning and followed by injection of the tracer of 8-10mCi of ^{18}F-FDG intravenously. The patients

were required to rest with an uptake time approximately 60 min in a confinement room before examination. All the patients underwent MRI imaging first, followed by PET-CT imaging studies according to standard protocol.

Imaging Technique

^{18}F-FDG. The dedicated integrated PET-CT system (Siemens Biograph-64, Germany) was used to standardise the value of ^{18}F-FDG uptake. This dedicated PET scanner with optimum performance in 3-D Imaging using Lutetium Oxyyorthosilicate (LSO) scintillator crystal technology was used to provide efficient rejection of random events, fast coincidence timing, high count rates and high light output. The device was incorporated with a multislice CT scanner with capability for high spatial resolution and a 64-slice CT. For higher sensitivity, the acquisition time for PET using this system was calibrated using a 2-min-per-table position. The CT data were resized from a 512x512 matrix to a smaller matrix, 128x128, to match the PET data to allow the fusion image and CT transmission map to be generated. Ordered-subsets expected maximisation (OSEM) was used to reconstruct the PET images with segmented measurement of attenuation correction using CT data with four iterations and 16 subsets. Post reconstruction smoothing of images was performed using a 5-mm FWHM Gaussion filter. The PET and CT images were then fused, creating an image of distributed ^{18}F-FDG overlying the corresponding anatomy and physiology generated using a dedicated workstation.

MRI spectroscopy. MRI was performed using a 3.0-T system (Prisma, Siemens Heatlcare). All patients were examined in a prone position. A bosy coil was used for transmission and a double-breast coil (4-element Breast Matrix Coil, Siemens Helathcare) was used for the MR spectroscopy. All the MRI sequence steps were performed using a single-voxel ^{1}H MRS with a point-resolved spectroscopy sequence (PRESS). The parameter of MRS was 1,620/270;voxel size 10x10x10 mm^3; 256 acquisitions; spectral width 1000Hz; 1,024 data points; and the time of acquisition, 7 min. For voxel placement, coronal and sagittal positions were used as scout images, and a voxel of interest was placed to include the lesion. Shimming was performed automatically first, followed by manual shimming on the water resonance for optimisation of the homogeneity in each volume of interest. Water-peak line widths of 10-20Hz (full width at half-maximum [FWHMI]) were typically achieved. After the shimming procedure, spectra were acquired with water suppression by applying three chemical shift-selective excitation pulses. By spectral suppression using dual band-selective inversion with gradient dephasing, the transverse magnetisation was selectively dephased before and after the second spin-echo pulse.

Image Analysis

The PET/CT scans and MRI breast images were read and interpreted by an experienced radiologist with more than five years' clinical experience, who was blinded to the history and diagnosis of each patient and was unaware of any clinical findings or information.

^{18}F-FDG. The results were analysed according to the qualitative assessment on a per-lesion basis in each patient who had undergone the FDG-PET/CT examination. On per lesion basis,

the higher inherent metabolic activity was compared to the mediastinal blood pool on the FDG-PET/CT and the lesions were considered positive lesions. All of the reviews of the PET images and the determination of the PET parameter were performed via software using a setting that allowed maximum intensity projection (MIP) and three-dimensional displays (transaxial, coronal, sagittal) of PET, CT and fused PET/CT images. The SUVmax was derived from the mean of all the FDG-positive lesions. The CT scans were also analysed for lesion location, size and the pattern of regional tumour infiltration. For the quantitative assessment, a circular region of interest (diameter 1.5 cm) was placed over the tumour in the slice with the maximum SUV in the baseline scan (Figure 1).

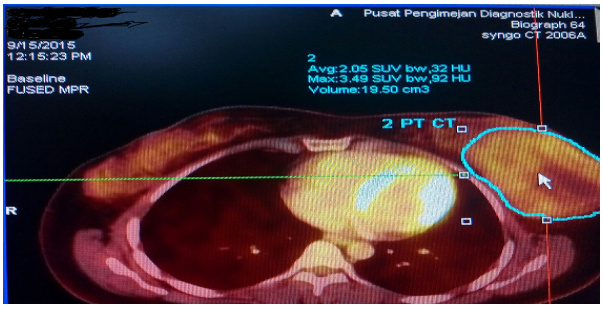
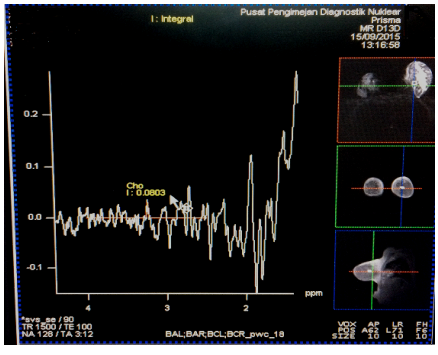
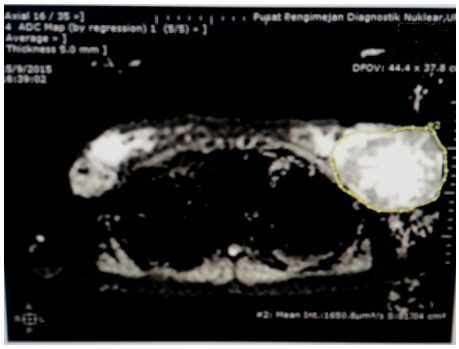


Figure 1. A circular region of interest (diameter 1.5 cm) was placed over the tumour in the slice with the maximum SUV of 3.49ug/dl

MRI spectroscopy. The obtained spectrum represents specific metabolites appearing in certain frequencies due to their specific chemical shifts. The resonance spectrum identifies metabolites by locating their peaks. Several peaks can characterise the same compound (i.e. doublet or triplet). Graphic representation of acquired data includes these metabolite peaks (represented on the horizontal axis of the graph) expressed as parts per million (ppm) and their relative signal amplitude in the vertical axis. The area or integral under each peak represents the relative concentration of the detected metabolite (Figure 2).



a



b

Figure 2. (a) The peak of tCho with 1.41 ppm, integral at 0.08 in benign tumour. (b) The circular region drawn at the breast lesion with *b* value of 50 showed non-restricted ADC value of 1.65mm^2/s, which was benign

Statistical Analysis

All the patients with biopsy-proven cancer were dichotomised into groups based on malignant and benign cancer. Univariate descriptive analysis was carried out using socio-demographic characteristics (age group and biopsy results). Numerical data described using means and standard deviation depending on the distribution of respective variables (Table 2). An independent t test was used to compare between the groups (Table 3). Data were analysed using the Pearson correlation coefficient. All statistics were analysed using the Statistical Package for Social Sciences programme for Windows 21.0 (SPSS 21) (IBM Corp, Somers, New York). P values <0.05 were considered statistically significant.

Table 1
Descriptive patient characteristics (n=8)

Characteristic	N (n=8)	Percentage
Age Group (Years)		
<30	1	12.5
31-40	5	62.5
41-50	2	25
Biopsy		
Fibroadenamatous	3	37.5
PASH	1	12.5
Fibroadenosis	1	12.5
IDC	2	25
ILC	1	12.5

RESULTS AND DISCUSSION

Patient Characteristics

The mean age of the patients in our study was 55.3±12.2 years, with middle age of 31-40 years being predominant (62.5%). Most of the subjects showed BIRADS 4 majority. The biopsy results showed five benign patients with three fibroadenoma, one Pseudoangiomatous Stromal Hyperplasia (PASH) and one fibroadenosis, while three patients showed malignant tumours with two Invasive Ductal Carcinoma (IDC) and one Invasive Lobular Carcinoma (ILC) (Table 1).

The relationship between biopsy with SUVmax and parameter of MRI spectroscopy. Table 2 shows there was significant relationship between SUVmax (p<0.05) and the biopsy results; the mean for malignant tumours was 4.28±3.74 and the mean for benign tumours was 2.33±0.90, while there was no significant correlation between the biopsy results and the MRI parameters.

Table 2
Relationship between biopsy with FDG uptake (SUVmax) and MRI spectrocopy using independent t test

	Biopsy	*n*	Mean	Std. Deviation	*p* value
FDG (SUVmax)	malignant	3	4.28	3.74	*0.01
	benign	5	2.33	0.90	
MRI spectroscopy choline	malignant	3	3.25	0.14	0.4
	benign	5	2.00	0.34	
MRI spectroscopy creatinine	malignant	3	3.45	0.70	0.44
	benign	5	3.39	0.52	
MRI spectroscopy N-acetyl Aspartate (NAA)	malignant	3	2.77	1.29	0.42
	benign	5	2.88	0.93	

*statistical significance p<0.05

Correlation between SUVmax and parameter of MRI spectroscopy. There was significant and direct correlation between FDG uptake with a mean of 3.06±2.34 and the creatinine mean of 3.39±0.54, where p=0.02. Besides that, there was also a significant and direct correlation between the FDG uptake with a mean of 3.06±2.34 and the N-acetyl Aspartate (NAA) mean of 2.84±0.99, where p=0.05 (Table 3).

Table 3
The correlation between FDG uptake (SUVmax) and parameter of MRI spectroscopy using Pearson correlation

SUVmax vs MRI spectroscopy	mean±S.D	r value	p value
-FDG (SUVmax)	3.06±2.34	0.16	0.32
-MRI spectroscopy choline	2.46±0.7		
-FDG (SUVmax)	3.06±2.34	0.63	*0.02
-MRI spectroscopy creatinine	3.39±0.54		
-FDG (SUVmax)	3.06±2.34	0.5	*0.05
-MRI spectroscopy NAA	2.84±0.99		

* statistical significance p<0.05

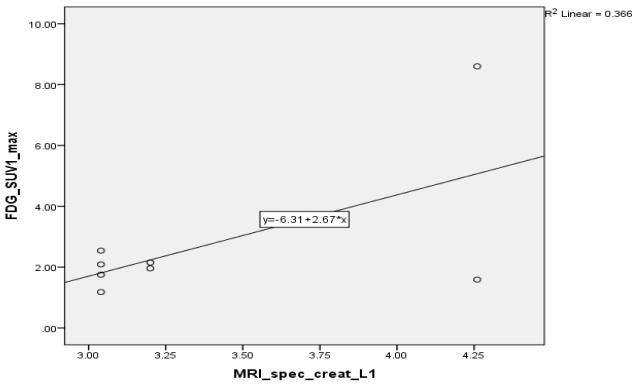


Figure 3. Scatter plot correlating the relationship between SUVmax and creatinine using Pearson correlation test

The level of choline in the human breast cancer cells was found to be markedly higher than in normal human mammary epithelial cells (Aoyama et al., 2004). Both choline transport and phosphorylation were found to be augmented in the human breast cancer cells. Progression of human mammary epithelial cells from normal to malignant phenotype was shown to be associated with an induced overexpression of choline kinase (CK) that catalysed the phosphorylation of choline to form phosphocholine followed by generation of phosphatidylcholine in the tumour cell membranes (Molina et al., 2004). In previous studies, the increased choline uptake in tumour cells was mainly explained by the upregulation of choline kinase due to an increased demand of membrane constituents (Molina et al., 2004). Recent studies found that the development of cancer was due to choline intake (Zhang et al., 2013). Besides that, other studies have proven that CK and phosphorylcholine (PCho) production increased in growth factor-induced mitogenic signaling of primary human breast epithelial cells in response to insulin and hydrocortisone (Ramirez et al., 2004).

Imaging of tumour cell metabolism has been remarkably successful in recent years. Numerous studies have demonstrated that malignant tumours can be detected with high sensitivity and specificity by imaging increased metabolic rates for glucose. FDG Positron Emission Tomography (PET) has been extensively used as an imaging biomarker for diagnosis (Kelloff et al., 2005) and for monitoring of treatment response in cancer patients. An accumulation of the radiotracer-marked glucose, in terms of 18f-fluorodeoxyglucose, indicates an increased metabolic activity of viable tumour cells that can be quantified by SUV measurements. Our studies showed that there was a significant and increased FDG uptake in malignant lesions with a mean of 4.28±3.74 compared to in benign lesions with a mean of 2.33±0.90. This study showed that a high SUVmax was associated with malignant cancer.

In the mammary gland area, total choline, tCho or Cho, is considered the most important metabolite in proton MR spectroscopy. It has been reported that the degree of elevated choline-containing compounds is related to the grade of the tumour, with higher levels in high-grade than in low-grade lesions (Baik et al., 2006). Malignant lesions are more likely to show high levels of choline-containing compounds compared to benign or normal breast tissues. In this preliminary study utilising the 3.0T MR revealed that there was insignificant relationship

between biopsy results with and without choline. The mean tCho for malignant lesions is 2.58±1.01u compared to benign lesions, with a mean of 2.39±0.82u ($p>0.05$). Confounding factors i.e. patient motion, contamination by hemorrhage and inclusion of lipids interspersed with stromal and inflammatory cells may affect a small preliminary sample and would not yield a significant statistical results. The 3.0 Tesla MRI can unveil new potential surrogate molecular markers for aggressive lesions.

The Positron Emission Tomography (PET) and MR methods are two clinically translatable, non-invasive imaging techniques that are increasingly being used to detect physiological changes and the response of tumors to target therapy. Based on our study, we may conclude that there is direct and significant correlation between SUVmax 3.06±2.34g/dl and the parameter of MRI spectroscopy i.e. creatinine 3.39±0.54u and N-acetyl Aspartate (NAA). The metabolite landscape pattern showed that the highest mean was creatinine followed by N-acetyl Aspartate (NAA), (mean: 2.84±0.99u) and choline (mean: 2.46±0.70u).

The limitation of this study was its small sample size and high number of benign cases that made it difficult to generalise the results to the general population. Besides that, small and non-palpable lesions bonded to a small volume of the samples, and these were below the PET resolutions. Future studies are recommended to used large sample sizes with various histologic prognostic parameters such as nuclear grade and estrogen receptor status that can correlate with PET/CT and MRI spectroscopy in detecting primary breast cancer.

CONCLUSION

This study showed that high SUVmax associated with malignant cancer and high creatinine metabolites that correlate with SUVmax could potentially be utilised as surrogate markers for suspicious lesions in detecting breast cancer lesion suspected during mammography. Thus, a combination PET/CT and MR spectroscopy can be used to classify breast lesion as indicating stability or progressive disease with potential therapeutic relevance in underpinning breast cancer.

ACKNOWLEDGEMENT

This work was supported by the Fundamental Research Grant Scheme (FRGS) from the Ministry of Higher Education, Malaysia, with ID No. 28612-61021.

REFERENCES

Aoyama, C., Liao, H., & Ishidate, K. (2004). Structure and function of choline kinase isoforms in mammalian cells. *Program Lipid in Research, 43*(3), 266–281.

Baek, H. M., Hon, J. Y., Chen, J. H., Nalcioglu, O., & Su, M. Y. (2008). Quantitative correlation between 1H MRS and dynamic contrast-enhanced MRI of human breast cancer. *Magnetic Resonance Imaging, 26*(4), 523–531.

Baik, H. M., Su, M. Y., Yu, H., Mehta, R., & Nalcioglu, O. (2006). Quantification of choline-containing compounds in malignant breast tumors by 1H MR spectroscopy using water as an internal reference at 1.5 T. *MAGMA, 19*(2), 96–104.

Burtscher, I. M., Skagerberg, G., Geijer, B., Englund, E., Stahlberg, F., & Holtas, S. (2000). Proton MR spectroscopy and preoperative diagnostic accuracy: An evaluation of intracranial mass lesions characterized by stereotactic biopsy findings. *American Journal of Neuroradiology, 21*(1), 84–93.

Chiuve, S. E., Giovannucci, E. L., Hankinson, S. E., Zeisel, S. H., Dougherty, L. W., Willett, W. C., & Rimm, E. B. (2007). The association between betaine and choline intakes and the plasma concentrations of homocysteine in women. *American Journal of Clinical Nutrition, 86*(4), 1073–1081.

Cho, E., Willett, W. C., Colditz, G. A., Fuchs, C. S., Wu, K., Chan, A. T. … Giovannucci, E. L. (2007). Dietary choline and betaine and the risk of distal colorectal adenoma in women. *Journal of the National Cancer Institute, 99*(16), 1224–1231.

Eberl, M. M., Fox, C. H., Edge, S. B., Carter, C. A. & Mahoney, M. C. (2006). BI-RADS classification for management of abnormal mammograms. *Journal of the American Board of Family Medicine, 19*(2), 161–164.

Fathinul Fikri, A. S., Abdul Jalil, N., & Lau, E. F. (2014). 18[F] FDG-PET/CT is a useful molecular marker in evaluating tumour aggressiveness; a revised understanding of an in-vivo FDGPET imaging that alludes the alteration of cancer biology. *Cell Biochemistry and Biophysic, 55*(5), 631–640.

Kelloff, G. J., Hoffman, J. M., Johnson, B., Scher, H. I., Siegel, B. A., Cheng, E. Y., … & Shankar, L. (2005). Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clinical Cancer Research, 11*(8), 2785–808.

Meisamy, S., Bolan, P. J., Baker, E. H., Robin, L., Bliss, M. S., Evin, G., … & Michael, G. (2004). Neoadjuvant chemotherapy of locally advanced breast cancer: predicting response with in vivo 1H MR spectroscopy – A pilot study at 4 T. *Radiology, 233*(2), 424–431.

Molina, R., Banez-Coronel, M., Gutierrez, R., Rodríguez-Gonzalez, A., Olmeda, D., Megías, D., & Lacal, J. C. (2004). Choline kinase activation is a critical requirement for the proliferation of primary human mammary epithelial cells and breast tumor progression. *Cancer Research, 64*(18), 6732–6739.

Moller-Hartmann, W., Herminghaus, S., Krings, T., Marquardt, G., Lanfermann, H., Pilatus, U., & Zanella, F. E. (2002). Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. *Neuroradiology, 44*(5), 371–381.

Zhang, C. X., Pan, M. X., Li, B., Wang, L., Mo, X. F., Chen, Y. M. … Ho, S. C. (2013). Choline and betaine intake is inversely associated with breast cancer risk: A two-stage case-control study in China. *Cancer Science, 104*(2), 250–258.

Pertanika J. Sci. & Technol. 26 (1): 235 - 246 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Vertically Integrated Moisture Flux Convergence over Southeast Asia and Its Relation to Rainfall over Thailand

Kosum Chansaengkrachang[1,2,3,4], Anirut Luadsong[1,5] and Nitima Aschariyaphotha[6]*

[1]Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
[2]The Petchra Pra Jom Klao Scholarship, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
[3]Centre of Excellence in Mathematics, Faculty of Science at Mahidol University, Phayathai Campus, Bangkok 10400, Thailand
[4]Faculty of Science and Technology, Surindra Rajabhat University, Surin 32000, Thailand
[5]Theoretical and Computational Science Center, Science Laboratory Building, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
[6]Ratchaburi Learning Park, King Mongkut's University of Technology Thonburi, Ratchaburi 70150, Thailand

ABSTRACT

The aims of this research are to study vertically integrated moisture flux convergence (VIMC) over Southeast Asia and to analyse its relationship to rainfall over Thailand during the period 1999 to 2013. Data reanalysed by the National Oceanic and Atmospheric Administration (NOAA) during the period 1999 to 2013 are used in this study. The monthly mean rainfall data are taken from the Global Precipitation Climatology Project (GPCP). Vertically integrated moisture transport (VIMT) is calculated by vertically integrating moisture fluxes of the u and v components. The finite difference method is applied to the vertically integrated moisture flux divergence (VIMD). The results show that VIMD over the Indian Ocean is strong, and the moisture is directed from the Indian Ocean to Thailand by southwest winds that cause strong moisture convergence over Thailand during the rainy season, while moisture in the summer season is a strong divergence. Moisture increases from the South China Sea to Thailand during October to December, causing more moisture convergence over northern and northeastern Thailand. That the relationship between rainfall and VIMC averaged over Thailand from the years 1999 to 2013 is confirmed by large positive correlations. The average from the years 1999 to 2013 over the study area is confirmed by Thailand's rainfall pattern.

Keywords: Moisture transport, moisture flux, rainfall, Southeast Asia, Thailand.

ARTICLE INFO

Article history:
Received: 01 December 2016
Accepted: 24 November 2017

E-mail addresses:
ck.kung02@gmail.com (Kosum Chansaengkrachang),
anirut.lua@kmutt.ac.th (Anirut Luadsong),
nitima.asc@kmutt.ac.th (Nitima Aschariyaphotha)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

The Southwest Monsoon and the Northeast Monsoon are related to rainfall over Thailand. The Southwest Monsoon, from May through September, brings a stream of warm moist air from the Indian Ocean towards Thailand, causing more rainfall over the area. The cold and dry air from the anticyclone on the China mainland migrates over Thailand during the Northeast Monsoon, starting in October (Climatological Group, 2015). Thailand had the highest average annual rainfall in 2011 and the lowest average annual rainfall in the year 2005, as presented by the Thai Meteorological Department (2014).

Agriculture in Thailand is highly competitive, diversified and specialised, and its exports are very successful internationally. Agricultural production as a whole accounted for an estimated 9% of Thai GDP and 40% of the population work in agriculture-related jobs (Luedi, 2016). Water is the main factor in agriculture, but the amount of rainfall is uncertain depending on whether it is a drought year or a wet year. This makes it difficult to manage the use of water in Thailand.

This research studied the factors that influence rainfall in Southeast Asia to attend to the predicted rainfall for sufficient water management especially in the drought years and wet years. The period 1999 to 2013 was also studied to determine the trend of rainfall.

Typical anomalous summer rainfall patterns of China were analysed using observational precipitation data (Zhou & Yu, 2005), the NCEP/NCAR and ERA40 reanalysis data related to the vertically integrated atmospheric water vapour transports derived from the Bay of Bengal and the South China Sea. These agreed with Wang and Chen (2012). They showed that mainly, atmospheric water vapour over south-eastern China during summer transported from the Indian Ocean and the tropical western Pacific occurred during the Indian and East Asian Monsoons. Ullah and Gao (2012) presented the convective centres of vertically integrated moisture transport flux over Pakistan and neighbouring regions from the divergent regions of the Arabian Sea and the Bay of Bengal, from which vertically integrated moisture convergence was deep in 1994. Ribeiro et al. (2014) found that winds in low level circulation from the ocean taking more moisture and the change of the wind vectors at high levels of the atmosphere supplied more rainfall in Belém. Trenberth, Fasullo and Mackaro (2011) stated that a better evaluation of the hydrological cycle components was provided by the atmospheric moisture budget, and the results also advised about E - P information. Guo et al. (2017) presented that the tropical cyclones associated with the moisture transport and rainfall over East Asia, and showed the comparison of the contribution of tropical cyclones to the area of water budget with other contributors. Jongaramrungruang, Seo and Ummenhofer (2017) found that convergence mainly due to low-level moisture was strong in the negative Indian Ocean dipole years. Further, He (2015) showed that anomalous northward moisture was taken from the western Pacific to central China, and summer precipitation anomalies were significantly positive over central China.

Other research has found one interesting factor affecting rainfall, namely, moisture transport. Therefore, the aim of this research was to study vertically integrated moisture flux convergence (VIMC) over Southeast Asia and to analyse its relationship to rainfall over Thailand during the period 1999 to 2013. We divided the seasons into three: summer (March-May), the rainy season (June-September) and winter (November-February).

THE STUDY AREA AND DATA USED

The Study Area

Southeast Asia was the designated study area. It is defined by latitude 0°S - 25°N and longitude 85°E - 120°E as shown in Figure 1. The map of Southeast Asia was plotted using the coastline extractor created by Signell (2014).

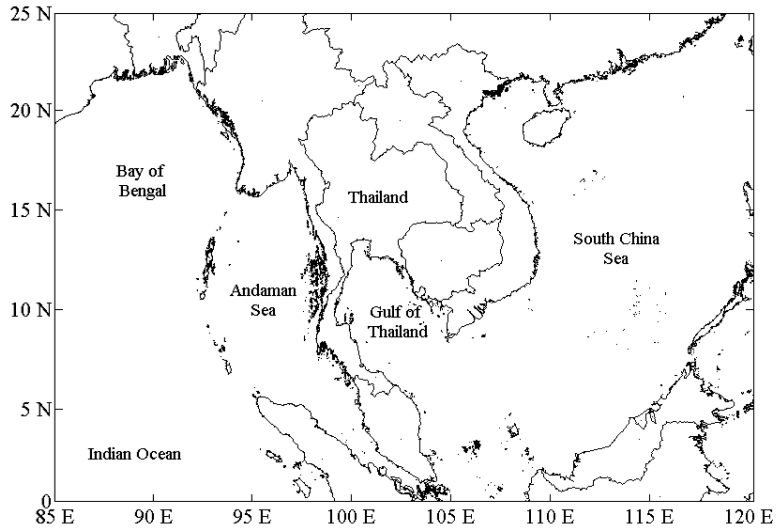


Figure 1. The study area (Signell, 2014)

Data Used

The data taken from the National Oceanic and Atmospheric Administration (NOAA) including the National Center for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) reanalysed data during the period 1999 to 2013, focussing on the u and v winds, specific humidity at the pressure levels 300 hPa, 400 hPa, 500 hPa, 600 hPa, 700 hPa, 850 hPa, 925 hPa and 1000 hPa and surface pressure at a resolution of 2.5° long × 2.5° lat (Kalnay et al., 1996; The NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, 2014). The monthly mean rainfall data were taken from the Global Precipitation Climatology Project (GPCP) at a resolution of 2.5° long × 2.5° lat during the period 1999 to 2013 (Adler et al., 2003; The NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, 2014). The evaporation minus precipitation (E-P) data were taken from the NCAR at a resolution of 0.7° long × 0.7° lat., which was the monthly mean during the period 1999 to 2013 (National Center for Atmospheric Research Staff, 2014; Trenberth & Fasullo, NCAR, 2014). All the data were interpolated with the computational grid at a resolution of 1° long × 1° lat.

The Moisture Conservation Equation

Water flux and transport information is important for understanding the global hydrological cycle, oceanic dynamics and global climate (Stewart, 2003). The moisture conservation equation in flux form of vertical integration suggested by Trenberth et al. (2011) is

$$\frac{\partial W}{\partial t} = -\nabla \cdot \frac{1}{g} \int_{300}^{p_s} q\mathbf{V} dp + E - P + R, \tag{1}$$

where, q is specific humidity; p is pressure; p_s is surface pressure; $\mathbf{V}$ is wind vector; g is gravitational acceleration; E is evaporation; P is precipitation and R is runoff. The left-hand side of Equation (1) is the change of total precipitable water (W) in column. The right-hand side of Equation (1) is the difference between evaporation and precipitation, vertically integrated moisture flux divergence (VIMD) and runoff (Ullah & Gao, 2012; Mo & Higgins, 1996; Trenberth & Guillemot, 1998).

The tendency term is small for long-term means; this is recommended by Mo and Higgins (1996). So, Equation (1) is now:

$$E - P = \nabla \cdot \frac{1}{g} \int_{300}^{p_s} q\mathbf{V} dp . \tag{2}$$

Equation (2) is the principal balance between E - P and vertically integrated moisture divergence (Trenberth et al., 2011). In this research, E - P is calculated from the NCAR data (National Centre for Atmospheric Research Staff, 2014; Trenberth & Fasullo, NCAR, 2014).

The wind vector, $\mathbf{V}$, is defined by $\mathbf{V} = (u, v)$, where u and v are the east-west and north-south components of wind. The vertically integrated moisture transport (VIMT) improved from Fasullo and Webster (2003) and Arias, Fu, Hoyos, Li and Zhou (2011) is calculated by $\int_{300}^{p_s} q\mathbf{V} dp \Big/ g$, and the vertically integrated moisture fluxes of u and v components are calculated by $\int_{300}^{p_s} qu dp \Big/ g$ and $\int_{300}^{p_s} qv dp \Big/ g$, respectively.

Specific humidity is small above the 300 hPa level and is not an effect to vertically integrated moisture transport (Fasullo & Webster, 2003). The vertically integrated moisture flux divergence (VIMD) is calculated by

$$\nabla \cdot \frac{1}{g} \int_{300}^{p_s} q\mathbf{V} dp = \frac{\partial \left(\int_{300}^{p_s} qu dp \Big/ g \right)}{\partial x} + \frac{\partial \left(\int_{300}^{p_s} qv dp \Big/ g \right)}{\partial y} . \tag{3}$$

The finite difference method using the central difference scheme is applied in Equation (3), which is a calculation of the monthly mean over Southeast Asia during the period 1999 to 2013. Thus, the VIMD can be approximated by

$$\left(\frac{\partial\left(\int_{300}^{p_s} qudp/g\right)}{\partial x}+\frac{\partial\left(\int_{300}^{p_s} qvdp/g\right)}{\partial y}\right)_{i,j} \approx \frac{\left(\int_{300}^{p_s} qudp/g\right)_{i+1,j}-\left(\int_{300}^{p_s} qudp/g\right)_{i-1,j}}{2\Delta x} + \frac{\left(\int_{300}^{p_s} qvdp/g\right)_{i,j+1}-\left(\int_{300}^{p_s} qvdp/g\right)_{i,j-1}}{2\Delta y}, \quad (4)$$

where, Δx and Δy are the grid spacing along east-west directions and north-south directions, respectively. The no-slip condition is used as the boundary condition in this paper. The specific humidity above pressure level 300 hPa is not an effect in studying vertically integrated moisture transport; this is recommended by Kalnay et al. (1996) and Fasullo and Webster (2003). Thus, the VIMT and VIMD perform from 300 hPa to the surface pressure level.

The vertically integrated moisture flux convergence (VIMC) averaged for Thailand during the period 1999 to 2013 was computed in this study. The VIMC is the negative of VIMD. The correlation between rainfall and VIMC averaged over Thailand was computed by:

$$r=\frac{n\sum_{i=1}^{n} X_i Y_i-\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{n\sum_{i=1}^{n} X_i^2-\left(\sum_{i=1}^{n} X_i\right)^2}\sqrt{n\sum_{i=1}^{n} Y_i^2-\left(\sum_{i=1}^{n} Y_i\right)^2}}, \quad (5)$$

where, X_i is rainfall averaged for Thailand; Y_i is the VIMC averaged for Thailand; n is the total number of months and i is the indices of the months. The formula for correlation coefficient r in Equation (5) was suggested by Hanke, Wichren and Reitsch (2001).

Absolute error was computed by $|\text{VIMD} - (E - P)|$, where $E - P$ is the difference between evaporation and precipitation and VIMD is the vertically integrated moisture flux divergence over the study area, as suggested by Olver (2008).

RESULTS

Moisture Transport

The VIMT and vertically integrated moisture fluxes of u and v components averaged from the years 1999 to 2013 are shown in Figure 2. In the colour bar, arctic relates to negative VIMT, and purple relates to positive VIMT. The results show that moisture transport appears from the southwest to the east of the study area towards Thailand from the Indian Ocean during the rainy season (June-September). Strong moisture with moisture components from the Indian Ocean

towards Thailand is carried by a southwest wind during the Southwest Monsoon associated with the rainy season in Thailand. The moisture transport is strongest (dark purple) in August and the vertically integrated moisture transport over Andaman Sea is strong at this time. This moisture transports from the Indian Ocean towards Thailand, and may bring more rainfall over Thailand. Moreover, strong moisture is in the South China Sea, and the moisture transports flow to the South China Sea during the rainy season. In the summer season (March-May), weak moisture flows from the east to west of the study area.

Moisture Convergence and Rainfall

The VIMD and winds averaged from the years 1999 to 2013 are shown in Figure 3. Blue relates to negative VIMD and purple relates to positive VIMD. The VIMC has the opposite sign of VIMD. During the rainy season, the VIMD over the Indian Ocean in September has high positive VIMD (dark purple), which is stronger than in June and July. Moisture direct from the Indian Ocean to Thailand by southwest winds causes strong moisture convergence over Thailand. Moisture convergence over Thailand is strongest in August (dark blue), inducing more rainfall during this period. Moisture convergence is strong over Thailand in May, which is the summer season. A strong divergence over the South China Sea occurs during October through December. Moisture is directed from the South China Sea through the Gulf of Thailand into Thailand during October to December; this causes more moisture convergence over the northern and northeastern parts of Thailand.

Figure 4 a) presents the monthly average values of rainfall and VIMC over Thailand from the years 1999 to 2013. The VIMC over Thailand during the rainy season is high, while during summer and winter it is low; rainfall during summer and winter is low. The relationship between rainfall and VIMC over Thailand from the years 1999 to 2013 has a high positive correlation of 0.8674. The monthly average values of rainfall and VIMC over Thailand in the years 2005 (a dry year) and 2011 (a wet year) are shown in Figure 4 b) and Figure 4 c), respectively. The correlation is highly positive i.e. 0.8938 and 0.8967, respectively. Obviously, VIMC in the year 2005 was smaller than the VIMC in the year 2011 during the rainy season, causing more rainfall in 2011. Figure 4 d) shows the difference of monthly averaged VIMC between the years 2011 and 2005. The difference in VIMC between the years 2011 and 2005 is visibly positive from July to September i.e. the rainy season. It can be inferred that moisture convergence from July to September in the year 2011 was greater than in 2005, and the rainfall in the year 2005 was also less than in 2011. There was a small difference in VIMC between the years 2011 and 2005 in the other months.

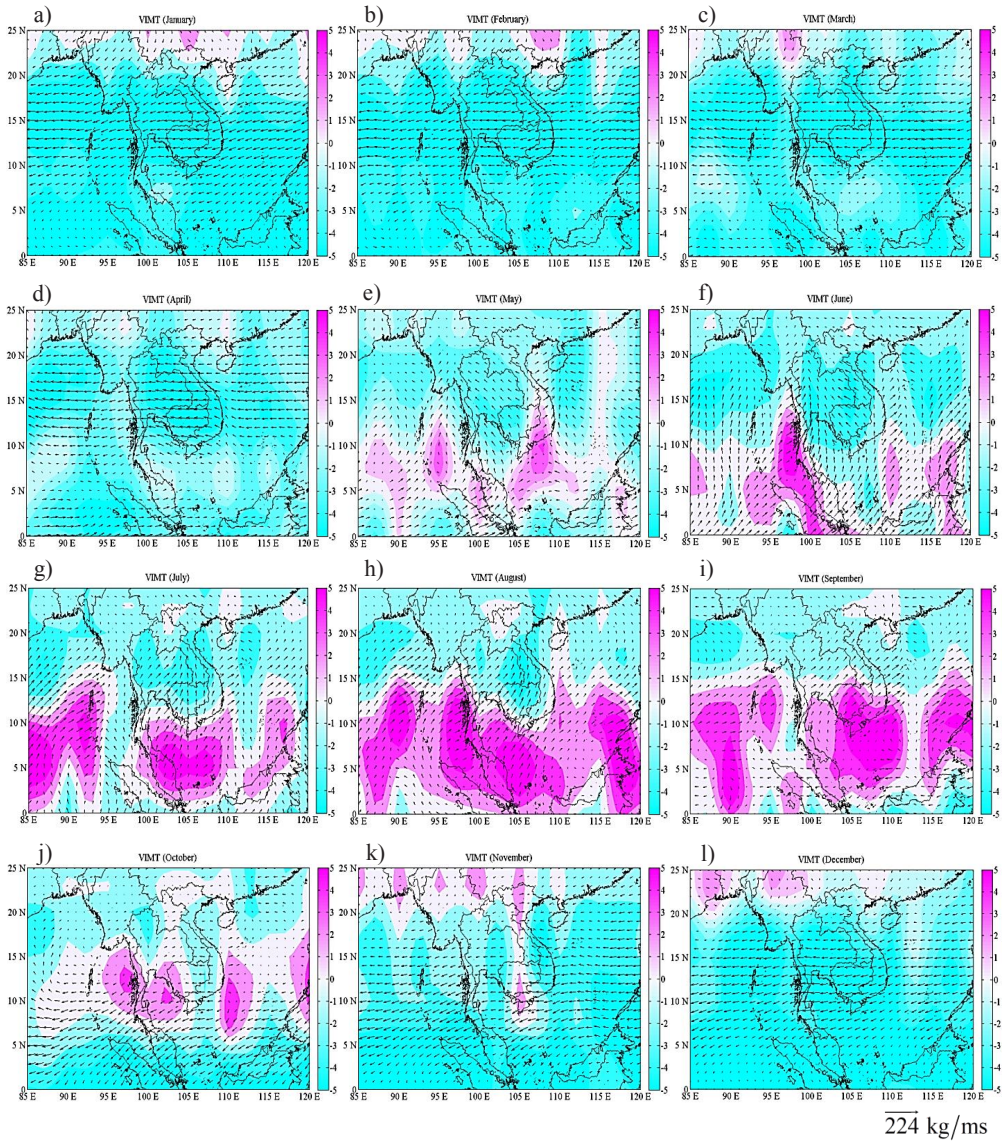


Figure 2. VIMT (kg/ms) shown as colour bars and vertically integrated moisture fluxes of u and v components averaged from the years 1999 to 2013 shown as vector during a) January, b) February, c) March, d) April, e) May, f) June, g) July, h) August, i) September, j) October, k) November and l) December

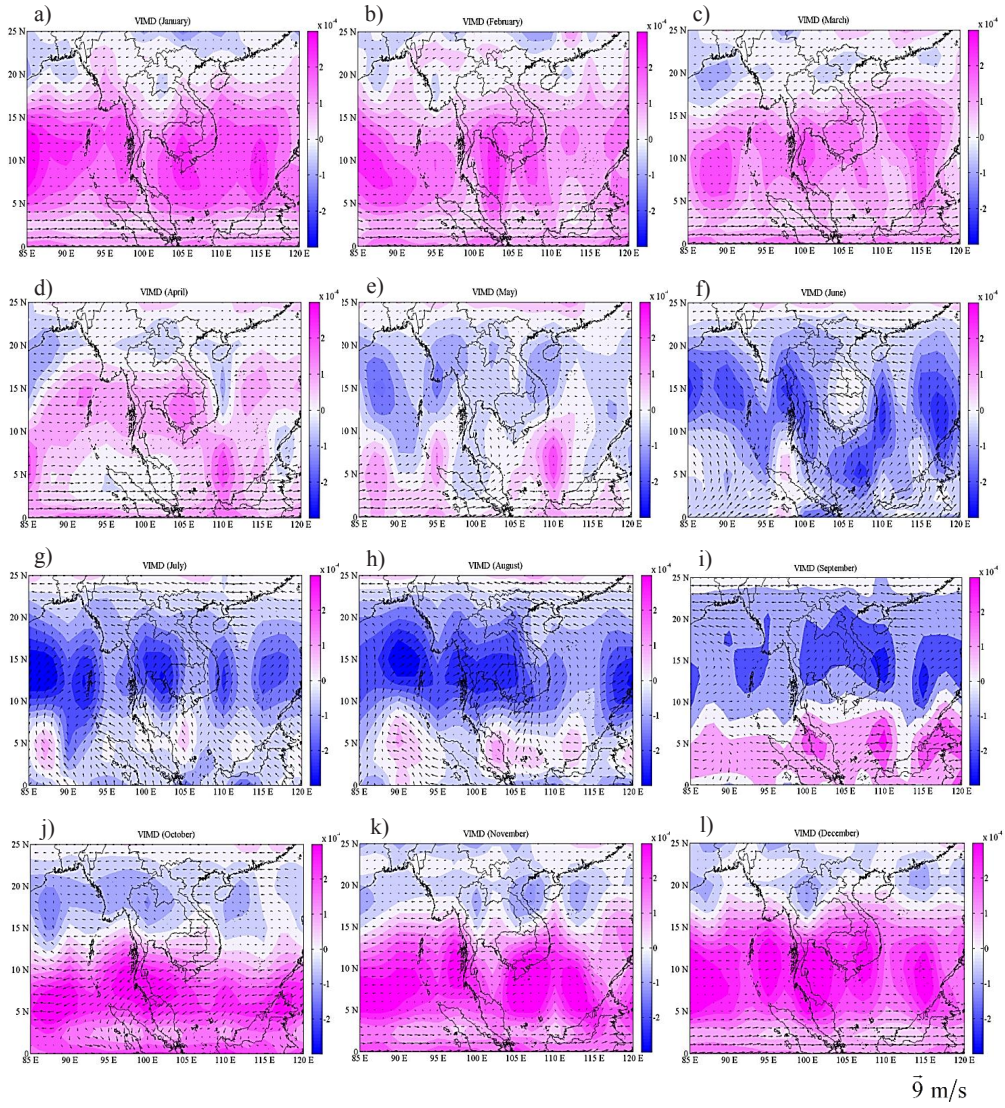


Figure 3. VIMD (kg/m^{2}s) shown as colour bars and winds averaged from the years 1999 to 2013 shown as vector during a) January, b) February, c) March, d) April, e) May, f) June, g) July, h) August, i) September, j) October, k) November and l) December

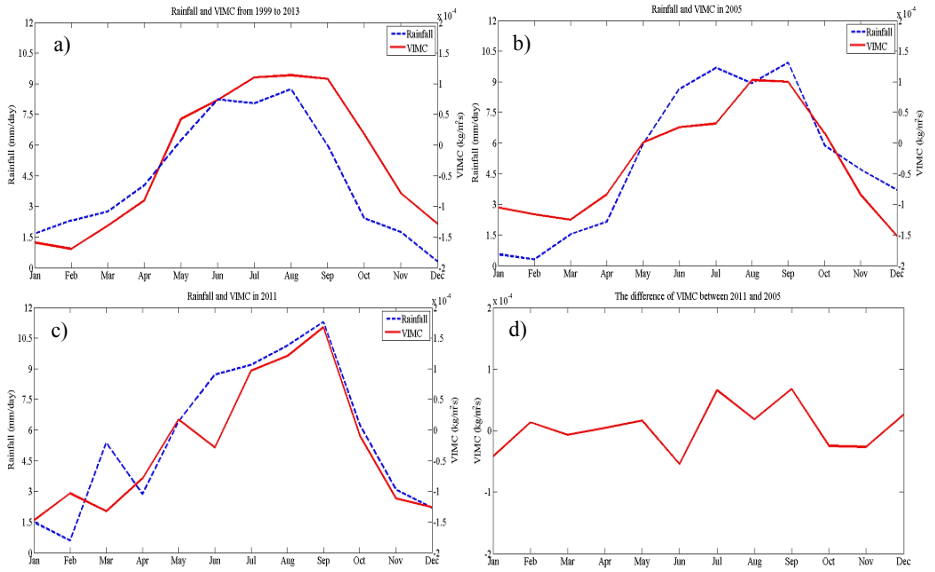


Figure 4. The monthly average values of rainfall and VIMC over Thailand a) from the years 1999 to 2013, b) in the year 2005 (dry year) and c) in the year 2011 (wet year); and d) the difference in monthly averaged VIMC between the years 2011 and 2005

Model Comparison with *E - P*

The *E - P* values averaged from the years 1999 to 2013 over the study area are shown in Figure 5. The *E - P* data were taken from the National Center for Atmospheric Research Staff (2014) and Trenberth and Fasullo, NCAR (2014). The *E - P* values were negative over the study area during May to October and highly negative over Thailand during the rainy season. It can be inferred that there was more precipitation than evaporation, causing more precipitation during the rainy season of Thailand. However, *E - P* values during January, February, March, April, November and December were positive over northern, northeastern, central and eastern parts of Thailand. The results show that there was more evaporation than precipitation; this corresponds to less precipitation during those periods over Thailand. Absolute error was employed in the evaluation study to find the absolute value of the difference between the calculated and actual values. If absolute error is very small, it can be said that *E - P* can be estimated by VIMD. In this study, the absolute error of *E - P* and VIMD are shown in Table 1. The small difference between *E - P* and VIMD infers that VIMD can be used to estimate *E - P*. The *E - P* value is the quantity that indicates the amount of water.

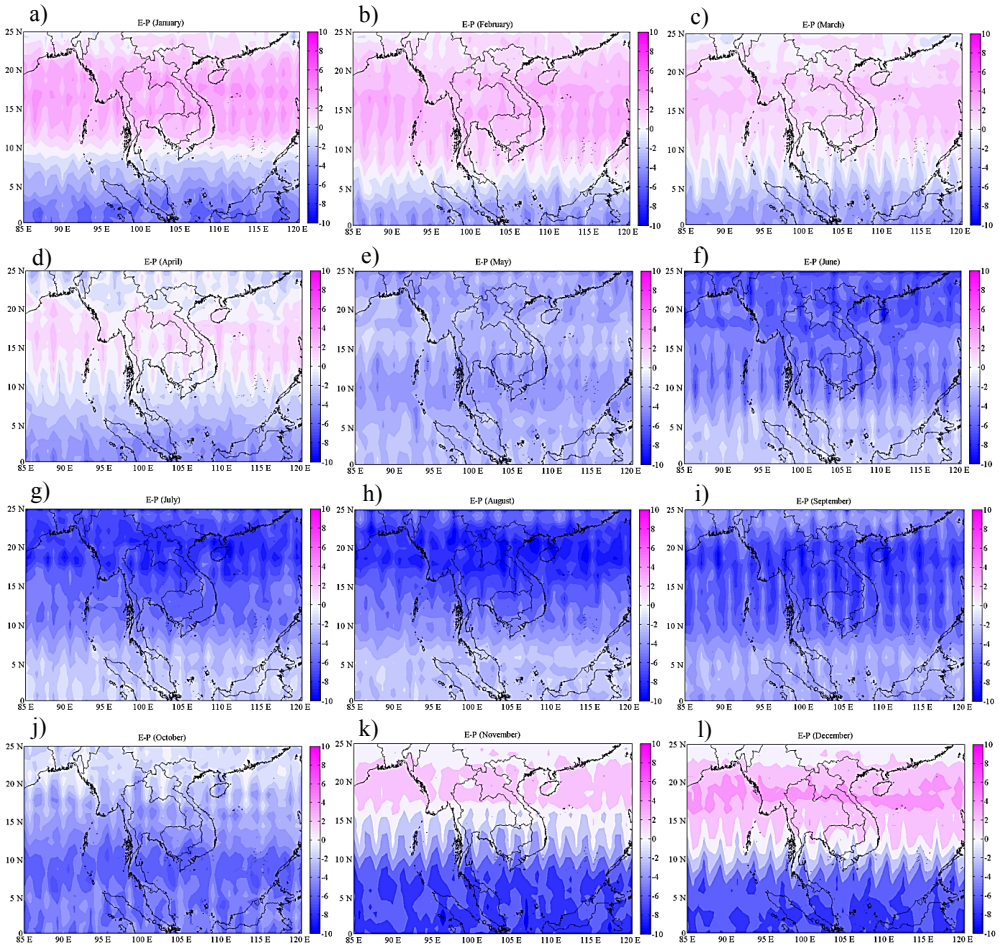


Figure 5. E - P taken from NCAR (National Center for Atmospheric Research Staff, 2014; Trenberth & Fasullo, NCAR, 2014) averaged from the years 1999 to 2013 (mm/day) during a) January, b) February, c) March, d) April, e) May, f) June, g) July, h) August, i) September, j) October, k) November and l) December

Table 1
Absolute error (kg/m²s) between and VIMD averaged from the years 1999 to 2013

Month	Absolute Error	Month	Absolute Error
January	0.000118	July	0.000010
February	0.000096	August	0.000009
March	0.000091	September	0.000065
April	0.000071	October	0.000141
May	0.000042	November	0.000147
June	0.000012	December	0.000146

CONCLUSION

In this analysis it can be summarised that moisture transport appears during the rainy season from southwest to east of the study area with southwest winds towards Thailand from the Indian Ocean. The strong moisture transports from the Indian Ocean towards Thailand may bring more rainfall over Thailand, and a weak moisture flow in the summer season occurs from west to east of the study area, causing less rainfall than during the rainy season. The VIMD over the Indian Ocean is strong. The moisture directs from the Indian Ocean to Thailand by a southwest wind, causing strong moisture convergence over Thailand during the rainy season. Moisture is directed from the South China Sea to Thailand during October to December, causing more moisture convergence over the northern and northeastern part of Thailand. The relationship between rainfall and VIMC averaged over Thailand from the years 1999 to 2013 is confirmed by large positive correlations. The difference in VIMC between the years 2011 and 2005 during the rainy season had positive values, and it can be inferred that the moisture convergence in the year 2011 was more than in 2005 during the rainy season, while rainfall in the year 2011 was greayer than in 2005. The small difference between E - P and VIMD inferred that VIMD can be used to estimate E - P.

ACKNOWLEDGEMENT

This research is supported by the Centre of Excellence in Mathematics (CEM) Scholarship, Faculty of Science, Mahidol University, Thailand and the Petchra Pra Jom Klao Scholarship, King Mongkut's University of Technology Thonburi (KMUTT), Thailand. The authors would like to thank the Department of Mathematics, King Mongkut's University of Technology Thonburi (KMUTT) for partial financial support and the use of their facilities during this research. We are grateful to all of the data sources used in the research.

REFERENCES

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., … & Nelkin, E. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology, 4*(6), 1147–1167.

Fasullo, J., & Webster, P. J. (2003). A hydrological definition of the Indian summer monsoon onset and withdrawal. *Journal of Climate, 16*(19), 3200–3211.

Guo, L., Klingaman, N. P., Vidale, P. L., Turner, A. G., Demory, M. E., & Cobb, A. (2017). Contribution of tropical cyclones to atmospheric moisture transport and rainfall over East Asia. *Journal of Climate, 30*(10). 3853–3865.

Hanke, J. E., Wichern, D. W., & Reitsch, A. G. (2001). *Business forecasting* (7th Ed.). NJ, USA: Prentice-Hall.

He, S. P. (2015). Potential connection between the Australian summer monsoon circulation and summer precipitation over central China. *Atmospheric and Oceanic Science Letters, 8*(3), 120–126.

Jongaramrungruang, S., Seo, H., & Ummenhofer, C. C. (2017). Intraseasonal rainfall variability in the Bay of Bengal during the Summer Monsoon: Coupling with the ocean and modulation by the Indian Ocean dipole. *Atmospheric Science Letters. 18*, 88–95.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … & Dennis, J. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society, 77*(3), 437–471.

Luedi, J. (2016). *Extreme drought threatens Thailand's political stability*. Retrieved from http://globalriskinsights.com/2016/01/extreme-drought-threatens-thailands-political-stability

MD. (2015). *The climate of Thailand.* Climatological Group, Meteorological Development Bureau, Meteorological Department. Retrieved from http://www.tmd.go.th/en/archive/thailand_climate.pdf

Mo, K. C., & Higgins, R. W. (1996). Large-scale atmospheric moisture transport as evaluated in the NCEP/NCAR and the NASA/DAO reanalyses. *Journal of Climate, 9*(7), 1531–1545.

NCARS. (2015). *The climate data guide: ERA-interim: Derived components*. National Center for Atmospheric Research Staff. Retrieved from https://climatedataguide.ucar.edu/climate-data/era-interim-derived-components

NOAA/OAR/ESRL PSD. (2014). *PSD Gridded Climate Datasets: All.* The, Boulder, Colorado, USA. Retrieved from http://www.esrl.noaa.gov/psd/

Olver, P. J. (2008). *Numerical analysis lecture notes*. Retrieved from http://www-users.math.umn.edu/~olver/num_/lna.pdf

Ribeiro, W. M. N., Souza, J. R. S., Lopes, M. N. G., Câmara, R. K. C., Rocha, E. J. P., & Almeida, A. C. (2014). Lightning and precipitation produced by severe weather systems over Belém, Brazil. *Revista Brasileira de Meteorologia, 29*, 41–59.

Signell, R. (2014). *The coastline extractor.* Retrieved from http://www.fr73.de/noaa/coast_panama.html

Stewart, R. H. (2003). *Introduction to physical oceanography*. Department of Oceanography, Texas. USA: Texas A & M University.

Thai Meteorological Department. (2014). *Mean annual rainfall in Thailand (mm)*. Retrieved from http://www.tmd.go.th/

Trenberth, K., & Fasullo, J. (2014). *ERA interim: Derived components, budget evaporation minus precipitation*. Retrieved from ftp://ftp.cgd.ucar.edu/archive/BUDGETS/ERAI

Trenberth, K. E., & Guillemot, C. J. (1998). Evaluation of the atmospheric moisture and hydrological cycle in the NCEP/NCAR reanalyses. *Climate Dynamics, 14*(3), 213–231.

Trenberth, K. E., Fasullo, J. T., & Mackaro, J. (2011). Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *Journal of Climate, 24*(18), 4907–4924.

Ullah, K., & Gao, S. T. (2012). Moisture transport over the Arabian Sea associated with summer rainfall over Pakistan in 1994 and 2002. *Advances in Atmospheric Sciences, 29*(3), 501–508.

Wang, H., & Chen, H. (2012). Climate control for southeastern China moisture and precipitation: Indian or East Asian monsoon? *Journal of Geophysical Research, 117*(D12), 1-9.

Zhou, T. J., & Yu, R. C. (2005). Atmospheric water vapor transport associated with typical anomalous summer rainfall patterns in China. *Journal of Geophysical Research, 110*, 1–10.

Pertanika J. Sci. & Technol. 26 (1): 247 - 260 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

New Smoothed Location Models Integrated with PCA and Two Types of MCA for Handling Large Number of Mixed Continuous and Binary Variables

Hamid, H.*, Ngu, P. A. H. and Alipiah, F. M.

School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM, Sintok, Kedah, Malaysia

ABSTRACT

The issue of classifying objects into groups when measured variables in an experiment are mixed has attracted the attention of statisticians. The Smoothed Location Model (SLM) appears to be a popular classification method to handle data containing both continuous and binary variables simultaneously. However, SLM is infeasible for a large number of binary variables due to the occurrence of numerous empty cells. Therefore, this study aims to construct new SLMs by integrating SLM with two variable extraction techniques, Principal Component Analysis (PCA) and two types of Multiple Correspondence Analysis (MCA) in order to reduce the large number of mixed variables, primarily the binary ones. The performance of the newly constructed models, namely the SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA are examined based on misclassification rate. Results from simulation studies for a sample size of n=60 show that the SLM+PCA+Indicator MCA model provides perfect classification when the sizes of binary variables (b) are 5 and 10. For b=20, the SLM+PCA+Indicator MCA model produces misclassification rates of 0.3833, 0.6667 and 0.3221 for n=60, n=120 and n=180, respectively. Meanwhile, the SLM+PCA+Burt MCA model provides a perfect classification when the sizes of the binary variables are 5, 10, 15 and 20 and yields a small misclassification rate as 0.0167 when b=25. Investigations into real dataset demonstrate that both of the newly constructed models yield low misclassification rates with 0.3066 and 0.2336 respectively, in which the SLM+PCA+Burt MCA model performed the best among all the classification methods compared. The findings reveal that the two new models of SLM integrated with two variable extraction techniques can be good alternative methods for classification purposes in handling mixed variable problems, mainly when dealing with large binary variables.

Keywords: Classification, large mixed variables, multiple correspondence analysis, Principal Component Analysis (PCA), Smoothed Location Model (SLM)

ARTICLE INFO

Article history:
Received: 06 February 2016
Accepted: 08 August 2017

E-mail addresses:
hashibah@uum.edu.my (Hamid, H.),
pennyngu90@hotmail.com (Ngu, P. A. H.),
fathilah@uum.edu.my (Alipiah, F. M.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Classification is a process of grouping objects into groups based on common attributes (Hunter, 2009). Classification tasks can be found in various fields ranging from medical, financial to education (Veer et al., 2002; Hauser & Booth, 2011). Many approaches such as quadratic discriminant analysis (Smith, 1947), logistic discrimination (Day & Kerridge, 1967), smoothed location model (Mahat et al., 2007; Hamid & Mahat, 2013) and k-nearest neighbour (Fix & Hodges, 1951) have been applied to solve classification problems. Compared to other approaches, the Smoothed Location Model (SLM) can be considered a good choice for handling mixtures of continuous and binary variables simultaneously (Vlachonikolis & Marriott, 1982). However, SLM is infeasible if dealing with a large number of binary variables.

In order to construct a classification model based on SLM, s cells of a multinomial table have to be generated from the b binary values for each group, where $s = 2^b$. Due to this structure, it was obvious that the number of multinomial cells in SLM increased exponentially with the size of binary variables considered in the study. This situation will increase the probability of the occurrence of empty cells if some multinomial cells are created. The occurrence of some empty cells will then cause the smoothed estimators of the location model to be biased and thus, affect the classification performance. Thus, it is very important to reduce the large number of binary variables in order to obtain an accurate classification model for the problem of mixed variables.

Multiple Correspondence Analysis (MCA) has been used to handle the problem of high dimensionality of categorical variables which has been proven to improve classification (Saporta & Niang, 2006; Nenadic & Greenacre, 2007). As expressed by Green et al. (1987) as well as Hoffman and Batra (1991), MCA has been widely applied in studies involving a large number of categorical variables. In fact, there are four different types of MCA i.e. Indicator MCA, Burt MCA, JCA and Adjusted MCA. A study by Hamid and Mahat (2013) focussed on high dimensional data, but only Burt MCA is used to reduce a large number of binary variables.

It is well known that Indicator MCA is a classic approach to MCA. It is to execute a simple correspondence analysis on the indicator matrix by performing singular value decomposition on the matrix of standardised residuals calculated on the indicator matrix. Burt matrix is actually the cross product of the indicator matrix. Due to the standard coordinates of the category points analysed by Burt MCA is similar to those analysed by the Indicator MCA, hence this study will observe the behaviour of the Indicator MCA and Burt MCA on the performance of the SLM. Our focus is on the performance of the newly constructed SLM, resulting from the integration of SLM with PCA and Indicator MCA as well as from the integration of SLM with PCA and Burt MCA in order to classify objects in some conditions, such as different sizes of binary and continuous variables and samples.

MATERIALS AND METHODS

Smoothed Location Model (SLM)

SLM is one of the methods that can handle both continuous and binary variables simultaneously. For classification tasks involving two groups, let Group 1 and Group 2 be denoted as π_1 and

π_2, respectively. A vector $\mathbf{z}^T = \left(\mathbf{x}^T, \mathbf{y}^T\right)$ is observed for each object in both groups, where the vector of b binary variables is represented by $\mathbf{x}^T = \{x_1, x_2, \ldots, x_b\}$ and the vector of, c, continuous variables is represented by $\mathbf{y}^T = \{y_1, y_2, \ldots, y_c\}$. To conduct a classification model, s cells of a multinomial table are generated from the b binary values for each group, where $s = 2^b$. The b binary variables will create some multinomial cells where the multinomial cell m can be defined by each different pattern of $\mathbf{x}$ uniquely with $\mathbf{x}$ falling in cell $m = 1 + \sum_{q=1}^{b} x_q 2^{q-1}$. The probability of obtaining an object in cell m of π_1 is denoted by p_{im}. We assume that c continuous variables have a multivariate normal distribution with mean $\boldsymbol{\mu}_{im}$ in cell m of π_1 and a common covariance matrix $\boldsymbol{\Sigma}$ across all cells and groups so that $Y_{im} \sim N(\boldsymbol{\mu}_{im}, \sum)$.

A future object $\mathbf{z}^T = \left(\mathbf{x}^T, \mathbf{y}^T\right)$ is allocated to π_1 if this object falls into the multinomial cell m and y satisfies

$$\left(\boldsymbol{\mu}_{1m} - \boldsymbol{\mu}_{2m}\right)^T \sum^{-1} \left\{ y - \frac{1}{2}\left(\boldsymbol{\mu}_{1m} - \boldsymbol{\mu}_{2m}\right) \right\} \geq \log\left(\frac{p_{2m}}{p_{1m}}\right) + \log(a) \tag{1}$$

otherwise, $\mathbf{z}^T$ will be allocated to π_2 (Krzanowski, 1980; 1993; 1995). For this classification model, we assume that a constant a, which is the misclassification costs, is equal to the prior probabilities in both groups, and hence $\log(a) = 0$.

However, the parameters of SLM in Equation (1) are commonly unknown and will be replaced with estimators obtained from the samples. By using non-parametric smoothing estimation, the mean $\boldsymbol{\mu}_{im}$ of each cell is fitted by a weighted average of all continuous variables from the data in the relevant group π_1. Thus, the vector mean of j^{th} continuous variables y of cell m in π_1 is estimated using:

$$\hat{\mu}_{imj} = \left\{ \sum_{k=1}^{s} n_{ik} w_{ij}(m,k) \right\}^{-1} \sum_{k=1}^{m} \left\{ w_{ij}(m,k) \sum_{r=1}^{n_{ik}} y_{rijk} \right\} \tag{2}$$

subject to

$$0 \leq w_{ij}(m,k) \leq 1 \text{ and } \left\{ \sum_{k=1}^{s} n_{ik} w_{ij}(m,k) \right\} > 0 \tag{3}$$

where $m, k = 1, 2, \ldots, s$; $i = 2$; n_{ik} is the number of objects falling in cell k of π_1; y_{rijk} is the j^{th} continuous variable of r^{th} object that falls in cell k of π_1 and w_{ij} (m, k) is a weight with respect to cell s of objects that fall in cell k.

In this study, the smoothing weight, w_{ij} (m, k), in the pattern of $w_{ij}(m,k) = \lambda_{ij}^{d(m,k)}$ is chosen where $0 < \lambda < 1$. This study chooses a method so that λ has the same value for all continuous variables in cells and groups and this could prevent the need to estimate many parameters. The d (m, k) explains the dissimilarity of the cell m and cell k of the binary vectors, which can be expressed as $d(\mathbf{x}_m, \mathbf{x}_k) = \left(\mathbf{x}_m - \mathbf{x}_k\right)^T \left(\mathbf{x}_m - \mathbf{x}_k\right)$.

The estimated means $\hat{\boldsymbol{\mu}}_{im}$ is then used to compute a smoothed pooled covariance matrix through

$$\hat{\Sigma} = \frac{1}{(n_1 + n_2 - g_1 - g_2)} \sum_{i=1}^{2} \sum_{m=1}^{s} \sum_{r=1}^{n_{im}} (y_{rim} - \hat{\mu}_{im})(y_{rim} - \hat{\mu}_{im})^T \tag{4}$$

where n_{im} is the number of objects falling in cell m of π_1; y_{im} is the vector of continuous variable of r^{th} object in cell m of π_1 and g_i is the number of non-empty cells in the training set of π_1.

Finally, the cell probabilities p_{im} can be obtained using the standardised exponential smoothing by

$$\hat{p}_{im(std)} = \hat{p}_{im} / \sum_{m=1}^{s} \hat{p}_{im} \tag{5}$$

where

$$\hat{p}_{im} = \frac{\sum_{k=1}^{s} w(m,k) n_{im}}{\sum_{m=1}^{s} \sum_{k=1}^{s} w(m,k) n_{im}} \tag{6}$$

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical tool used to reduce the dimensionality of the data while retaining important information of the original data as much as possible (Kemsley, 1996). PCA has been highlighted as an adequate variable extraction technique for continuous variables (Costa et al., 2013). PCA reduces the data dimension by choosing a few orthogonal linear combinations of the original variables which show the largest variance accounted for. The linear combination of variables with the largest variance is chosen and denoted as the first principal component (PC1). Meanwhile, the second principal component (PC2) will account for and explain the maximal variability, which is not included in the PC1, and this component is uncorrelated with PC1. The same process is continued to obtain PC3, PC4 and so on (Quinn & Keough, 2002; Rengner, 2008).

Consider a set of data that consists of p numeric variables with q principal components. The random vector is labelled $\mathbf{Y} = (y_1, y_2, ..., y_p)^T$ with a mean vector, $\boldsymbol{\mu} = E[y]$, while $\mathbf{C}_{jk} = E\left[(y - \mu)(y - \mu)^T\right]$ is a covariance matrix between y_j and y_k, where $j, k = 1, 2, \ldots, q$.

Next, the Eigenvectors (μ_j) and the respective Eigenvalues (λ_j) will be inserted into

$$\mathbf{C}u_j = \lambda_j u_j \tag{7}$$

Then, the eigenvalues are determined through

$$\left|\mathbf{C}_{jk} - \lambda\mathbf{I}\right| = 0 \tag{8}$$

where $\mathbf{I}$ is the identity matrix that has the same order as $\mathbf{C}_{jk}$ and $|\,.\,|$ is the determinant of a matrix.

Indicator MCA and Burt MCA

MCA has been demonstrated to show similar capability as PCA, but on the binary variables (Bar-Hen, 2002). MCA is good for detecting and representing the underlying structures of data (usually nominal categorical data) in a low dimensional space (Greenacre & Blasius, 2006). As such, it can be seen as a generalisation of PCA when the variables to be analysed are categorical instead of quantitative. The indicator **Z** is the matrix with cases (row) and category variables (column) where the category variables are coded in the form of dummy variables (binary matrix of indicator) with the value of 0 and 1. Meanwhile, Burt matrix is the cross product of the indicator matrix, which can be expressed in the form of **B** representing Burt matrix while **Z** is the indicator matrix (Greenacre, 2007; Nenadic & Greenacre, 2007).

Using the notation from Tenenhaus and Young (1985), suppose that a set of m categorical variables $X_1, X_2, ..., X_m$ with categorical size of $k_1, k_2, ..., k_m$, respectively is used to describe an original data matrix. Category l of variable j is defined as jl and coded into binary matrix **Z** where the general entries for **Z** are defined as

$$Z_{ijl} = \begin{cases} 1 & \text{if object } i \text{ from category } l \text{ of variable } j \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

A complete indicator $Z = \left[Z_1, Z_2, ...Z_d\right]$ with n rows and d columns $\left[d = \sum_{j=1}^{m} k_j\right]$ is obtained by merging the matrices **Z**. Then a (d, d) symmetric matrix of Burt **B** x $Z^T Z$ is built, where Z^T is a transpose matrix of **Z**. Let X be a (d, d) diagonal matrix that has the same diagonal elements as matrix **B**, then a new matrix **S** is constructed from Z and X by

$$\mathbf{S} = \frac{1}{d} Z^T Z X^{-1} = \frac{1}{d} B X^{-1} \tag{10}$$

In this study, the constructed SLM was evaluated and measured through a misclassification rate using leave-one-out (LOO) procedure. The misclassification rate can be obtained by taking the total number of misclassified objects and dividing it by the total number of objects in the group. The misclassification rate can be obtained by

$$LOO = \frac{\sum_{k=1}^{n} error}{n} \tag{11}$$

Algorithm 1 outlines the steps of the discrimination process with variable extraction techniques that are involved in this study for high dimensional data of mixed variables.

Data Generation

Thirty sets of data were generated for the purpose of this study using the R software package. The sample size (n) was set to have 60, 120 and 180 while the sizes of continuous variables (c) were set to 30, 60 and 90. The sizes of binary variables (b) are set to 5, 10, 15, 20 and 25. Observations were made on the behaviour of the Indicator MCA and Burt MCA towards the performance of the newly constructed SLM using these generated datasets. Thus, the focus of this study was on the performance of two newly constructed SLM models resulting from the

SLM integrated with PCA and Indicator MCA and SLM integrated with PCA and Burt MCA in order to classify objects when dealing with a large number of mixed variables, primarily the binary. Algorithm 1 describes the steps involved in constructing these new classification models, symbolised as SLM+PCA+Indicator MCA model and SLM+PCA+Burt MCA model, implemented in the leave-one-out fashion.

ALGORITHM 1

Step 1:	Omit an object k from the sample n, where the remaining objects are treated as a training set.
Step 2:	Perform PCA to extract and reduce the continuous variables using the training set.
Step 3:	Perform Indicator MCA to extract and reduce the binary variables using the training set.
Step 4:	Construct new SLM using the reduced sets of continuous and binary variables that have been extracted in Steps 2 and 3, which produces SLM+PCA+Indicator MCA model.
Step 5:	Predict the group of the omitted object k using the new model constructed in Step 4 and assign error if the prediction made is correct, otherwise assign.
Step 6:	Repeat all the steps from 1 to 5 for each object.
Step 7:	Compute the misclassification rate using the leave-one-out procedure for model evaluation.

Then, all the steps are repeated except for Step 3, where the Indicator MCA is replaced with the Burt MCA in order to construct another new SLM known as the SLM+PCA+Burt MCA model.

RESULTS AND DISCUSSION

Results from Simulation Study

Tables 1, 2 and 3 summarise the results of a simulation study for the two newly built SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for n=60, n=120 and n=180, respectively. The performances of these newly constructed models are assessed and compared based on the misclassification rates calculated through the LOO procedure.

Table 1 displays the performance of the SLM+PCA+Indicator MCA model as well as the SLM+PCA+Burt MCA model specifically for n=60. The highest misclassification rate was 0.5333 obtained by the SLM+PCA+Indicator MCA model when b=25, while the SLM+PCA+Burt MCA model only showed 0.0167 of the misclassification rate for the same binary size. From the results, it can be observed that the misclassification rate was strongly related with the number of binary components that were extracted. For the SLM+PCA+Indicator

MCA model, there was an increase of the misclassification rate when the number of binary sets extracted was increased. The same behaviour was observed for the SLM+PCA+Burt MCA model. The SLM+PCA+Indicator MCA model shows 0.3833 of the misclassification rate when nine binary components were extracted from the original b=20. In contrast, SLM+PCA+Burt MCA model classified all objects correctly with only six extracted components for the same original binary size. A similar pattern was obtained for b=25, which revealed that the SLM+PCA+Burt MCA model produces a much smaller misclassification rate compared to the SLM+PCA+Indicator MCA model due to the fact that the former model extracted much smaller binary components.

In SLM, sparseness of objects in multinomial cells may influence the estimation of parameters and the performance of classification models that is directly related to the misclassifying of objects into groups. Sparseness of objects is referred to as the occurrence of too many empty cells in the SLM. For example, in the case of n=60 and b=25 as can be seen in Table 1, the SLM+PCA+Indicator MCA model extracts 10 binary components. These 10 extracted components were considered very high in the context of SLM as they managed to produce up to 1,024 multinomial cells per group. Nevertheless, only 29 cells of π_1 and 28 cells of π_2 were not empty. This means that on average as high as 97.22% are empty cells and it is absolutely impractical to be used for the construction of the SLM. This is the main and most important reason why the SLM+PCA+Indicator MCA model showed the highest misclassification rate at 0.5333, which means that more than half of the objects had been misclassified due to the fact that the majority of the created cells were empty cells. In contrast, the SLM+PCA+Burt MCA model achieved a much smaller misclassification rate at 0.0167 due to only six binary components having been extracted, thus making the percentage of empty cells much lower i.e. 64.85% on average.

Table 1
Performance of SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for all binary sizes measured under n=60

	Size of Binary Variables				
SLM+PCA+Indicator MCA	5	10	15	20	25
Misclassification Rate	0	0	0.0333	0.3833	0.5333
Number of Binary Extracted (PC_B)	3	6	7	9	10
Number of Continuous Extracted (PC_C)	10	10	9	9	9
Number of Empty Cells (π_1, π_2)	(1,0)	(39,41)	(99,104)	(483,485)	(995,996)
KL Distance	397.94	16.51	3.77	0.77	0.40
SLM+PCA+Burt MCA	5	10	15	20	25
Misclassification Rate	0	0	0	0	0.0167
Number of Binary Extracted (PC_B)	2	4	5	6	6
Number of Continuous Extracted (PC_C)	10	10	9	9	9
Number of Empty Cells (π_1, π_2)	(0,0)	(2,2)	(11,11)	(40,39)	(42,41)
KL Distance	294.28	281.90	88.46	15.06	16.67

The misclassification rate in either model was found to be highly related to the number of binary components that were extracted. This binary extracted amount was then discovered to be closely associated with the Kullack-Leibler (KL) distance, which can indirectly affect the performance of the newly constructed models. For example, the misclassification rate of the SLM+PCA+Indicator MCA model for the case of n=120 as shown in Table 2 rose higher when the KL distance grew smaller. This model achieved 0.6721 of misclassification rate with 0.24 units of distance, while the misclassification rate for the SLM+PCA+Burt MCA model was only 0.0167 due to the distance being much greater, that is, 7.41 units. The performance of the newly constructed SLM+PCA+Indicator MCA model began to show a misclassification rate when the distance between the observed groups was less than 1.0 unit. This tells us that the smaller the distance between the groups under study, the higher the misclassification rate obtained.

In addition, the number of binary components extracted is also strongly related to the number of empty cells that occurred, which further gives impact to the operation of the constructed SLM. For example, the misclassification rate of the SLM+PCA+Indicator MCA model for the case of n=180 increased as the number of empty cells grew due to the fact that many binary components were extracted. As shown in Table 3, the number of empty cells rose when the number of binary components extracted increased. As a result, for the case of b=25, the SLM+PCA+Indicator MCA model achieved 0.5688 of the misclassification rate with 12 extracted binary components. This poor performance was due to the fact that nearly all the created cells were empty i.e. 97.95% of π_1 and 98.10% of π_2.

Besides that, sample size was another factor that affected the performance of the constructed models. The misclassification rate was smaller when the sample size was larger. For example, the misclassification rates of the SLM+PCA+Indicator MCA model decreased from 0.6667 and 0.6721 to 0.3221 and 0.5688 when the size of the sample was increased from n=120 to n=180, respectively for the cases b=20 and b=25. This demonstrated that the misclassification rate drops when the sample size is increased.

Table 2
Performance of SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for all Binary sizes measured under n=120

	Size of Binary Variables				
SLM+PCA+Indicator MCA	5	10	15	20	25
Misclassification Rate	0	0	0	0.6667	0.6721
Number of Binary Extracted (PC_B)	3	6	8	10	12
Number of Continuous Extracted (PC_C)	18	19	18	18	17
Number of Empty Cells (π_1, π_2)	(0,0)	(26,29)	(203,203)	(964,966)	(4012,4018)
KL Distance	684.04	48.29	7.72	0.27	0.24
SLM+PCA+Burt MCA	5	10	15	20	25
Misclassification Rate	0	0	0	0	0.0167
Number of Binary Extracted (PC_B)	3	5	6	7	8
Number of Continuous Extracted (PC_C)	18	19	18	18	17
Number of Empty Cells (π_1, π_2)	(2,2)	(4,3)	(26,24)	(82,78)	(203,202)
KL Distance	164.04	849.91	169.86	34.63	7.41

Table 3
Performance of SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for all binary sizes measured under n=180

	Size of Binary Variables				
SLM+PCA+Indicator MCA	5	10	15	20	25
Misclassification Rate	0	0	0.011	0.3221	0.5688
Number of Binary Extracted (PC_B)	4	6	9	10	12
Number of Continuous Extracted (PC_C)	26	26	26	26	28
Number of Empty Cells (π_1, π_2)	(0,0)	(16,16)	(428,432)	(964,966)	(4012,4018)
KL Distance	2592.71	775.56	6.46	3.45	1.97
SLM+PCA+Burt MCA	5	10	15	20	25
Misclassification Rate	0	0	0	0	0
Number of Binary Extracted (PC_B)	3	5	7	8	9
Number of Continuous Extracted (PC_C)	18	26	26	26	28
Number of Empty Cells (π_1, π_2)	(2,2)	(4,5)	(60,60)	(182,179)	(429,431)
KL Distance	205.56	1565.58	145.72	27.93	6.69

Tables 4 and 5 summarise the results of simulation studies for both SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for different data conditions under investigation. We compared the performance of the newly constructed models under the same binary size (i.e. b=20) for all samples tested as displayed in Table 4. The SLM+PCA+Indicator MCA model recorded the lowest misclassification rate at 0.3221 when n=180, while the highest was 0.6667 when n=120. On the other hand, the SLM+PCA+Burt MCA model achieved good performance as there were no objects that had been misclassified for all sizes of the samples examined.

Table 4
Performance of SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for b=20 based on all sample sizes examined

b=20	Misclassification Rate	
	SLM+PCA+Indicator MCA	SLM+PCA+Burt MCA
n=60	0.3833	0
n=120	0.6667	0
n=180	0.3221	0

Next, Table 5 shows the performance of the SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for n=120 with all sizes of the binary that were used for investigation purposes. The SLM+PCA+Indicator MCA model achieved zero misclassification rate for b=5, b=10 and b=15 while the SLM+PCA+Burt MCA model only showed a 0.0167 misclassification rate when b=25. The SLM+PCA+Indicator MCA model showed high misclassification rates when b=20

and *b*=25 due to a large number of binary components having been extracted by the indicator matrix, which were 10 and 12 for *b*=20 and *b*=25, compared to the second model, with 7 and 8 binary components that were extracted using Burt MCA.

Table 5
Performance of SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for n=120 based on all binary sizes measured

n=120	Number of Binary Extracted (Indicator, Burt)	Misclassification Rate	
		SLM+PCA+Indicator MCA	SLM+PCA+Burt MCA
b=5	(3, 3)	0	0
b=10	(6, 5)	0	0
b=15	(8, 6)	0	0
b=20	(10, 7)	0.6667	0
b=25	(12, 8)	0.6721	0.0167

Table 6 displays the average computational time for executing the whole process of simulation for all generated datasets. We discovered that computational time was strongly influenced by the number of binary components extracted and the size of the observed sample. Computational time increased with the number of binary components, where the number of multinomial cells grew as the binary variables grew. In addition, the amount of the sample used also greatly affected computational time as this study implemented double looping of the LOO procedure. Comparing the two constructed models, it can be observed that computational time for the SLM+PCA+Indicator MCA model was much higher than for the SLM+PCA+Burt MCA model, especially when the former had more binary variables and sample sizes. For example, for the case *n*=120 and *b*=15, the computational time for SLM+PCA+Burt MCA model was only 11 hours and 42 minutes, while the SLM+PCA+Indicator MCA model required 1 day and 14 hours to complete the simulation process. The time taken by the latter model was triple that of the former model. A similar pattern of computational time was observed for the other cases as well. Thus, it can be inferred that the SLM+PCA+Burt MCA model was more efficient in terms of computational time compared to the SLM+PCA+Indicator MCA model.

Table 6
Average computational time of SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA models for all simulated datasets

Sample Size	n = 60		Sample Size	n = 120		Sample Size	n = 180	
Mixed Variables Measured	Indicator	Burt	Mixed Variables Measured	Indicator	Burt	Mixed Variables Measured	Indicator	Burt
c=30, b=5	1 hour	8 minutes	c=60, b=5	2.62 hours	2.29 hours	c=90, b=5	14.24 hours	7.97 hours
c=30, b=10	1.73 hours	32 minutes	c=60, b=10	11.49 hours	6.66 hours	c=90, b=10	1.38 days	20.73 hours
c=30, b=15	3.3 hours	1 hour	c=60, b=15	38 hours	11.42 hours	c=90, b=15	10.08 days	2.75 days
c=30, b=20	17.1 hours	1.67 hours	c=60, b=20	9 days	19.27 hours	c=90, b=20	12.5 days	6.46 days
c=30, b=25	36 hours	3.15 hours	c=60, b=25	11.83 days	1.54 days	c=90, b=25	24.08 days	10.13 days

Results from Real Dataset

This study further utilised the new constructed models i.e. SLM+PCA+Burt MCA and SLM+PCA+Indicator MCA to interpret full breast cancer data associated with the influences of psychosocial behaviour among breast cancer patients conducted at King's College Hospital, London. The full breast cancer data were derived from 137 women with breast tumours who had been divided into two groups i.e. the benign tumour group (π_1) consisting of 78 women and the malignant tumour group (π_2) consisting of 59 women. The original dataset contained 15 variables comprising two continuous variables, six ordinal variables with 11 states each, four nominal variables with three states each and three binary variables. It has become the practice to treat ordinal variables as continuous and to convert nominal variables to binary variables (Krzanowski, 1975; Mahat et al., 2007; Hamid, 2014). Therefore, this study treats six ordinal variables as continuous variables and converts all four nominal variables to binary values. By treating ordinal variables as continuous and dichotomising nominal variables into binary variables, the dataset was given a new dimension consisting of eight continuous and 11 binary variables.

In order to assess the performance of the newly constructed models, we compared them with other classification methods including linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), logistic discrimination (logistic), linear regression model (regression), classification tree (tree), SLM with variable selections and SLM with double PCA (2PCA) using a real dataset i.e. for full breast cancer. Comparisons were made in terms of misclassification rate as shown in Table 7. Results from this data demonstrated that the SLM+PCA+Burt MCA and SLM+PCA+Indicator MCA models yielded a low misclassification

rate at 0.2336 and 0.3066, respectively, where the former model performed the best among all the classification methods compared. The second in the ranking was the SLM with double PCA. Meanwhile, the constructed SLM+PCA+Indicator MCA model was in seventh ranking, which means that it performed worse than LDA, regression and logistic discrimination. Nevertheless, the SLM with variable extractions performed better than the SLM with variable selections. The findings exhibited that SLM+PCA+Burt MCA was the most appropriate method to manage the extraction process of large continuous and binary variables before performing classification tasks.

Table 7
Comparison of eight classification methods for full breast cancer dataset

Classification Methods	Selection Strategy	Misclassification Rate	Performance Rating
LDA	Include all variables	0.2920	4
QDA	Include all variables	0.4453	12
Logistic	Include all variables	0.2847	3
Regression	Forward selection	0.3139	8
	Backward selection	0.2920	4
	Stepwise selection	0.2920	4
Tree	Auto termination	0.3139	8
Smoothed Location Model:			
(i) Smoothed LM with variable selections	Forward selection	0.3139	8
	Stepwise selection	0.3139	8
(ii) Smoothed LM with double PCA	PCA+PCA (2PCA)	0.2774	2
(iii) Smoothed LM with PCA and MCA	PCA + Indicator MCA	0.3066	7
	PCA + Burt MCA	0.2336	1

CONCLUSION

In this study, we investigate the performance of the newly constructed classification models i.e. SLM+PCA+Indicator MCA and SLM+PCA+Burt MCA, measured based on their misclassification rates using the location model as a basis for the construction. Both of these classification models showed good performance under b=5, b=10 and b=15 for all sizes of samples inspected. However, the overall results revealed that SLM+PCA+Burt MCA model performed better than the SLM+PCA+Indicator MCA model for all sample sizes and binary variables that were tested as well as in terms of computational time. This study also found that PCA and Burt MCA were superior in extracting and reducing the large number of continuous and binary variables. Findings from simulation and real datasets proved that the two newly constructed location models can be considered potential tools in discriminant analysis when practitioners are faced with a large number of mixed variables, mainly binary variables.

ACKNOWLEDGEMENT

We gratefully acknowledge financial support from Universiti Utara Malaysia (UUM) under its Postgraduate Scholarship Scheme.

REFERENCES

Bar-Hen, A. (2002). Generalized principal component analysis of continuous and discrete variables. *Journal of Applied Statistics, 8*(6), 11–26.

Costa, P. S., Santos, N. C., Cunha, P., Cotter, J., & Sousa, N. (2013). The use of multiple correspondence analysis to explore associations between categories of qualitative variables in healthy ageing. *Journal of Aging Research,* 1–12.

Day, N. E., & Kerridge, D. F. (1967). A general maximum likelihood discriminant. *Biometrics, 23,* 313–323.

Fix, E., & Hodges, J. L. (1951). *Discriminatory analysis, nonparametric discrimination: Consistency properties* (Report No. 4). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a800276.pdf

Green, P. E., Krieger, A. M., & Carroll, J. D. (1987). Multidimensional scaling: A complementary approach. *Journal of Advertisement Research, 27*, 21–27.

Greenacre, M. J. (2007). *Correspondence analysis in practice* (2nd Ed.). Boca Raton: Chapman and Hall/CRC.

Greenacre, M. J., & Blasius, J. (2006). *Multiple correspondence analysis and related methods.* London: Chapman and Hall/CRC.

Hamid, H. (2014). *Integrated smoothed location model and data reduction approaches for multi variables classification.* (Unpublished doctoral thesis). Universiti Utara Malaysia, Kedah, Malaysia.

Hamid, H., & Mahat, N. I. (2013). Using principal component analysis to extract mixed variables for smoothed location model. *Far East Journal of Mathematical Sciences (FJMS), 80*(1), 33–54.

Hauser, R. P., & Booth, D. (2011). Predicting bankruptcy with robust logistic regression. *Journal of Data Science, 9*, 565–584.

Hoffman, D. L., & Batra, R. (1991). Viewer response to programs: Dimensionality and concurrent behavior. *Journal of Advertising Research, 23*, 45–47.

Hunter, E. J. (2009). *Classification made simple: An introduction to knowledge organisation and informative retrieval* (3rd Ed.). England: Ashgate Publishing Company.

Kemsley, E. K. (1996). Discriminant analysis of high-dimensional data: A comparison of principal component analysis and partial least squares data reduction methods. *Chemometrics and Intelligent Systems, 33*, 47–61.

Krzanowski, W. J. (1975). Discrimination and classification using both binary and continuous variables. *Journal of the American Statistical Association, 70*(352), 782–790.

Krzanowski, W. J. (1980). Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics, 36*, 493–499.

Krzanowski, W. J. (1993). The location model for mixtures of categorical and continuous variables. *Journal of Classification, 10*, 25–49.

Krzanowski, W. J. (1995). Selection of variables and assessment of their performance in mixed variable discriminant analysis. *Computational Statistics and Data Analysis, 19*, 419–431.

Mahat, N. I., Krzanowski, W. J., & Hernandez, A. (2007). Variable selection in discriminant analysis based on the location model for mixed variables. *Advances in Data Analysis and Classification, 1*(2), 105–122.

Nenadic, O., & Greenacre, M. (2007). Correspondence analysis in R, with two and three dimensional graphics: The CA package. *Journal of Statistical Software, 20*(3), 1–13.

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. New York, NY: Cambridge University Press.

Rengner, M. (2008). What is principal component analysis? *Nature Biotechnology, 26*(3), 303–304.

Saporta, G., & Niang, N. (2006). Correspondence analysis and classification. In M. Greenacre & J. Blasius (Eds.), *Multiple correspondence analysis and related methods* (pp. 122–150). Boca Raton: Chapman and Hall/CRC.

Smith, C. A. B. (1947). Some examples of discrimination. *Annals of Eugenics, 13*, 272–283.

Tenenhaus, M., & Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika, 50*(1), 91–119.

Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., ... & Schreiber, G. J. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature, 415*(6871), 530–536.

Vlachonikolis, I. G., & Marriott, F. H. C. (1982). Discrimination with mixed binary and continuous data. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 31*(1), 23–31.

Pertanika J. Sci. & Technol. 26 (1): 261 - 282 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Effect of Distilled and Sea Water Absorption on Mechanical Behaviour of Short Coir Fibre Epoxy Composite/Sawdust Filler

Moorthy M. Nair[1], Nagaraja Shetty[2]*, Divakara Shetty S.[2] and Shambhavi Kamath M.[1]

[1]Department of Civil Engineering, Manipal Institute of Technology, Manipal University, Manipal-576 104, Karnataka, India

[2]Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal University, Manipal-576 104, Karnataka, India

ABSTRACT

Research on natural fibres has been carried out from past decades as a result of developing low cost, eco-friendly materials. The objective of this study is to fabricate composites utilising sawdust of various proportion and compare the mechanical properties like tensile, flexural and hardness within dry condition and with respect to specimens immersed in distilled and salt water. The composites with constant reinforcement 15%, different percentage of matrix (85, 80, 75, 70%) and filler (0, 5, 10, 15%), respectively by mass, are developed by hand layup method and is compared for their mechanical properties. Mechanical properties of composite fabricated from sawdust up to certain percentage showed an improvement when compared to composite with no filler; with further increase in filler, a drop in mechanical performance is noticed. An increase in tensile strength by 12.75%, flexural strength by 5.94%, hardness by 18.34%, tensile modulus by 100.3%, flexural modulus by 60.4% is observed in dry condition compared to the composite with no filler. Mechanical degradation in tensile strength, tensile modulus, hardness, flexural strength for the samples subjected to ageing in sea and distilled water is observed. Flexural modulus after ageing increased with filler addition up to a certain percentage and with further increase in filler, a decrease is noticed. Higher mechanical degradation (except flexural modulus) is observed for those specimens immersed in sea water.

Keywords: Ageing, composite, epoxy, mechanical properties, sawdust, short coir fibre

ARTICLE INFO

Article history:
Received: 18 May 2017
Accepted: 09 November 2017

E-mail addresses:
moorthymnair@yahoo.in (Moorthy M. Nair),
hosadunagaraj@gmail.com (Nagaraja Shetty),
shetty.divakar@manipal.edu (Divakara Shetty S.),
shambavi.kamath@manipal.edu (Shambhavi Kamath M.)
*Corresponding Author

INTRODUCTION

As a result of increased environmental awareness, natural fibres have gained importance due to their unique properties such as ease of availability, being light weight,

ISSN: 0128-7680

strength, renewability, eco-friendliness, being low cost, low density and biodegradability. Glass fibre composites are environmentally inferior when compared to natural fibre composites in most cases for the following reasons: (a) Production of natural fibre prevents environmental impressions compared to glass fibre; (b) Fibre content is high in natural fibre, resulting in reduced base polymer pollution content; (c) lightweight natural fibre composites improve productivity and diminish pollution in the use phase of the component, especially in auto applications and (d) Disposed natural fibres are incinerated to ebb energy and carbon credits (Joshi et al., 2004). Plant fibres are of interest due to their availability and reinforcement, which can be done at low cost (Ramesh et al., 2012). Most commonly used eco-friendly materials are plant fibres such as coir, jute, sisal, kenaf, banana, hemp and pineapple. Natural fibre composites are being more and more extensively used. Coir-fibre-reinforced polymer composites are being utilised for industrial and socioeconomic advantages such as automotive interior, panelling and roofing as building materials, storage tank, packing material, helmets, post boxes, mirror casing, paperweights, projector covers and voltage stabiliser (Ayrilmis et al., 2011; Yousif et al., 2012; Verma et al., 2013; Abilash & Sivapragash, 2013). Maximum application of natural fibre composites is noticed in the automobile sector (Emad et al., 2016).

Coir is extracted from coconut husk. Durability of coir is better than that of any other natural fibres due to its high lignin content (Verma et al., 2013; Khan et al., 2012). Total world coir fibre production is 250,000 tones/year (Coir Board Ministry of MSME, Govt. of India). Today India, mainly from the coastal region of Kerala State, produces 60% of the total world supply of coir fibre. Coir fibre is more productive and pompous in reinforcement performance compared to other reinforcement composites (Ticoalu et al., 2010). Coir exhibits low moduli due to a high microfibrilar angle. Toughness and percentage break elongation is higher in coir and it has good resistance to weather, fungal and bacterial infection due to its high content of lignin (Nam et al., 2011). Coir possesses 26-43% cellulose, 0.2% hemicellulose and 41-45% lignin content by mass (Nam et al., 2011). Due to high lignin and low cellulose content, unlike other natural fibres, coir needs treatment for effective interfacial bonding (Binu, 2012). The major drawbacks of natural fibres are weak adhesion with the matrix material as a result of incompatibility and a high water-absorption rate due to its hydrophilic nature. Incompatibility occurs when reactive groups of hydrophilic fibre are covered by dirt and polar groups, resulting in inept coupling with hydrophobic matrix (Chern et al., 2014).

Moisture diffusion in polymeric composites is governed by three different techniques (Papanicolaou et al., 2008; Zainab, 2009). The first involves diffusion of water molecules inside the micro gaps between the polymer chains. The second involves capillary transport into the gaps and flaws at the interfaces between the fibre and the matrix. This is a result of poor wetting and impregnation during the initial manufacturing stage. The third involves transport of micro cracks in the matrix arising from the swelling of fibres. To make fibres more compatible with hydrophobic matrix surfaces, treatments are performed to improve interfacial adhesion between them and other materials. Chemical treatment of fibres is a common method of cleaning and modifying the fibre surface in order to lower the surface tension and enhance the interfacial adhesion between a natural fibre and a polymeric matrix (Bledzki & Gassan, 1999). Hydrogen bonding from the fibre structure is removed with the help of alkali treatment, which substantially decreases moisture absorption and also alters surface morphology (Ramadevi et

al., 2015). Due to changes in surface morphology more reactive groups of hydrophilic fibres are exposed, promoting efficient integration with hydrophobic matrices (Chern et al., 2014).

In this study short coir fibres are reinforced with epoxy matrix and sawdust as filler. The main factors like fibres election, matrix selection, fibre orientation and the fabrication process should be pre-decided to determine the final mechanical properties (Pickering et al., 2016). Epoxy resins have an advantage that they reduce volume shrinkage (Sheng et al., 2012). The ring-opening polymerisation of the epoxy resin frees its constrained three-member ring structure, resulting in a much larger free volume (Odian, 2004). Water-immersed specimen behaviour is influenced by reinforcement, matrix and filler nature along with relative humidity and the type of fabrication method employed, which influences porosity and volume fraction of the fibres. Water intake for some natural fibres can also result in the water plasticing effect (Stambouliset al., 2000).

Sawdust is a major biological waste generated in wood polishing firms. Its storage in uncontrolled condition can be a factor responsible for environmental pollution (Deac et al., 2016). Variation in chemical composition for sawdust observed was 27.2-31.7% of lignin, 44.5-46.65% cellulose, 15.69-16.29% hemicellulose and 0.01-0.02% nitrogen content (Mercy et al., 2011; Campbell et al., 1966). From the environmental point of view, it is very important to prevent pollution resulting from sawdust. In this research sawdust waste was utilised to improve the mechanical performance of short coir fibre epoxy composite by replacing a proportion of epoxy with sawdust.

MATERIALS AND METHOD

Coir fibre procured from local resources was used for the preparation of the composites. Initially, the coir fibre was segregated, finely alkalised and minced to a length of 12-15 mm. An epoxy resin (LY 556) was used as the matrix binder. The mechanical and thermal properties of epoxy resins are extremely poor. To improve the properties, the resin should undergo a curing reaction in which the linear epoxy resin assembly is modified to form a three-dimensional cross-linked thermoset structure. A hardener, a curing agent in ratio to epoxy 1:10, is added to accelerate the curing action. The hardener-epoxy reaction releases an enormous amount of heat, resulting in the homopolymerisation of the resin. The curing agent or hardener used in this study was triethylenetetramine (HY-951). Sawdust was procured from local resources and tested for its physical properties.

Alkali Treatment

In alkali treatment in this study, pre-washed coir fibres were immersed in 6% (w/v) NaOH aqueous solution for 72 h. The fibres were removed from the alkali solution and rinsed with distilled water and dried to constant weight.

Composite Fabrication

In this study the fabrication of the composite was completed using an economical method called the hand lay-up method as this technique can be successfully employed in composite fabrication

(Srinivasababu et al., 2009). A mould of dimensions 300×300×6 mm^3 was used. A composite of 300 × 300 × 3.2 mm^3 was fabricated for different proportions of reinforcement, matrix and filler as shown in Table 1. Epoxy resin with hardener in a ratio of 10:1 was thoroughly mixed with a certain proportion of sawdust. Mould releasing wax was used on the mould for easy removal of the final product. Chopped fibre, agitated with epoxy, was gently poured on to the sheet inside the mould. The mixture was then allowed to settle within the mould for a whole day under pressure over the cast. The final specimen was machined as per ASTM standard dimension for mechanical testing.

Table 1
Proportion of coir fibre, epoxy and sawdust used for fabrication

Sample / Specimen Number	Type of Fibre Used	Reinforcement (Mass Percentage)	Matrix (Mass Percentage)	Filler (Mass Percentage)
1	Short Fibres	15	85	00
2	Short Fibres	15	80	05
3	Short Fibres	15	75	10
4	Short fibres	15	70	15

Calculation used for fabrication of composite. Table 2 shows the calculation to determine the quantity of short coir fibre, epoxy resin and sawdust utilised for fabrication of composites based on different proportions.

Table 2
Calculation to determine quantity of materials utilised for composite fabrication

Density of epoxy	'X' g/cm^3
Density of fibre	'Y' g/cm^3
Density of Sawdust	'Z' g/cm^3
Density of composite ($\rho_{composite}$)	Density of epoxy × (Mass Percentage of epoxy)+ Density of fibre× (Mass percentage of coir fibre) + Density of sawdust× (Mass percentage of sawdust)
$\rho_{composite}$	'G' g/cm^3
Mass of composite	$\rho_{composite}$ × volume of plate = G × (300 × 300 × 3.2) × 10^{-3}
Mass of composite	'M' g
Mass of Epoxy	Mass percentage of epoxy × M = 'E' g
Mass of Fibre	Mass percentage of fibre × M = 'F' g
Mass of Sawdust	Mass percentage of sawdust × M = 'S' g

Physical Properties of Sawdust

Specific gravity (IS:2720, Part 3, 1964). Specific gravity is determined using bottle density as per IS:2720 (Part 3), 1964. A clean, dry and cool density bottle with stopper was weighed initially (w_1), $^2/_3$ of the bottle was filled with sawdust sample for weighing (w_2). After removing the entrapped air, the sample was filled with distilled water and weighed (w_3). The empty bottle with water was weighed (w_4). Specific gravity was calculated using Eq. [1].

$$Specific\ gravity\ (G) = \frac{(w2-w1)}{(w2-w1)-(w3-w4)} \quad [1]$$

Particle size determination (IS:2720, Part 4, 1965). Particle size is determined using the hydrometric method as per IS:2720 (Part 4), 1965. A sample of 50 g (w) was accurately weighed and mixed with distilled water to form a smooth paste. An amount of 1 g of sodium silicate was added to the paste and the mixture was washed into the mixer. Mixing was carried out for 10 min and the specimen was graduated into a specimen jar. Enough water was added to bring the volume to 1000 cc (v). Water was added to the mixture. A hydrometer was inserted and a stopwatch was started. The hydrometer reading (Rh) was taken at different elapsed times (t) without removing the hydrometer. A graph was plotted with particle size readings in the x-axis and percentage finer in the y-axis. Particle size was determined using Eq. [2]. The percentage finer was calculated using Eq. [3].

$$\text{Particle size} = \sqrt{\frac{18\mu}{G-1}} \times \frac{He}{t} \quad [2]$$

where, He = Effective height (cm)
μ = viscosity of sample-water suspension (g-s/cm^3)

$$\text{Percentage finer} = \left[\frac{Rh}{1000} \times \frac{G}{G-1}\right] \times 1000 \quad [3]$$

Density of particle. Particle density of the sawdust was determined (Araki & Terazewa, 2004; Ruhlmannetal, 2006). It was calculated by adding the sawdust (100 cm^3) into a graduated volumetric cylinder to reach the marked 100 cm^3 volume (V_o), and its weight (g) could then be known by subtracting the combined weight of the sawdust and volumetric cylinder (W_b) with the weight of the empty volumetric cylinder (W_a) alone. Sawdust – particle density could then be calculated using Eq. [4].

$$\text{Particle density} = \frac{Wb-Wa}{Vo} \quad [4]$$

Porosity of sawdust particle. Sawdust porosity is measured by unoccupied space, which principally is composed of the inter-space among and intra-space within the particles (Agnew & Leonard, 2003; Bouma et al., 2003) or the percentage of sawdust volume occupied by air and water that fills the void (Baker et al., 1998). The percentage of porosity was determined (Horisawa et al., 1999). Sawdust with an apparent volume of 100 cm^3 and known weight (W_s in g) was at first placed in a volumetric cylinder. Tap water was then poured gently into it until

the surface of the water reached a marked line at the 100 cm3 level. A meshed top as a stopper was equipped at the 100 cm^3 level so that the sawdust, mostly floating on water, would not go beyond its surface. Porosity was expressed as the following Eq. [5].

$$\text{Porosity (\%)} = \frac{Va}{Vo} * 100 \quad [5]$$

where, V_a and V_o are consumptively the volume of poured water (cm^3) together with the water in the sawdust, and the volume of sawdust (100 cm^3), respectively. The volume of poured water, with the water in the sawdust could be calculated using Eq. [6].

$$V_a \text{ (cm}^3\text{)} = W_{comb} - W_s - W_{vs} \quad [6]$$

where, W_{comb} is the combined weight (g) of the volumetric cylinder, sawdust particles and poured water (g); W_s is the weight (g) of the sawdust particles (oven-dry weight equivalent) and Wvs is the weight (g) of the volumetric cylinder. Note that the density of tap water was assumed to be unity (1 gcm^3).

Properties of Distilled Water and Sea Water

Distilled water collected from the Environmental Engineering Lab, MIT Manipal, was tested for its pH value using a pH meter. Sea water collected from Panambur Beach, Mangalore, was checked for its pH using a pH meter. Salinity was determined using a salinity refractometer.

Mechanical Property Tests

Density test. A change in density will be observed in the actual composite laminate when compared to theoretical density due to the presence of void content, which accounts for the change in mechanical performance. Air void content was calculated using Eq. [7]

$$\text{Void content (\%)} = \frac{\text{Theoretical density} - \text{Actual density}}{\text{Theoretical Density}} \quad [7]$$

Tensile test. ASTM D3039 standard was employed for performing a tensile result using a Universal Testing Machine (UTM). A composite laminate with dimensions of 250 × 25 × 3.2 mm^3 was machined. The test was conducted at a constant strain rate of 2 mm/min tensile strength and the modulus was calculated using Eq. [8] and Eq. [9], respectively.

$$\text{Tensile strength (MPa) } (\sigma_t) = \frac{\text{Load in KN}}{\text{Cross-sectional area of specimen(mm*mm)}} \quad [8]$$

$$\text{Tensile modulus (GPa)} = \frac{\text{Tensile strength} * \text{Length (mm)}}{\text{Displacement (mm)}} \quad [9]$$

Flexural test. The flexural test on composites was performed as per ASTM D790-03 standards with the dimensions 12.5 × 125 × 3.2 mm^3. The three-point bend test was performed on the

composites using the same UTM (FIE UNITEK 9450) at a crosshead speed of 1 mm/min with a span length to depth ratio of 16:1. The flexural strength was determined using Eq. [10]. Flexural modulus was calculated using Eq. [11]

$$\text{Flexural strength (MPa) } (\sigma f) = \frac{3PL}{2bt*t} \quad [10]$$

where, L is the span length of the sample (mm); P is maximum load (N); b the width of specimen (mm); t the thickness of specimen (mm) and d is displacement(mm);

$$\text{Flexural modulus (GPa)} = \frac{P*L*L*L}{4*b*t*t*t*d} \quad [11]$$

Micro Vickers hardness. The Matzusawa micro-hardness tester was employed in measuring the micro-hardness of composite specimens as per ASTM E384-16. A diamond intender with an apical angle of 136° was indented over the surface of the specimen under a load of 100 gf under a dwell time of 15 s. After the removal of the load the two diagonals, D1 and D2, of the indentation were measured. The hardness value was noted.

Water absorption test. Water absorption specimen of size 75 × 25 × 3.2 mm^3 was machined as per ASTM D 570 and immersed in distilled water and sea water at room temperature (27ºC). The water absorption percentage was calculated using Eq. [12]. A graph of the moisture intake percentage vs time in hours was plotted and the diffusion coefficient (m^2/sec) was calculated using Eq. [13].

$$\text{Water absorption (\%)} = \frac{\text{Final weight after immersion} - \text{Initial weight before immersion}}{\text{Initial weight before immersion}} \quad [12]$$

$$\text{Diffusion coefficient } (m^2/sec) = \pi \text{ x } \left(\frac{B}{4Ms}\right)^2 \text{ x (slope)}^2 \quad [13]$$

RESULTS AND DISCUSSION

Physical Properties of Sawdust

Table 3 shows the physical properties of sawdust like density, specific gravity, porosity and particle size, all of which play a major role in mechanical performance.

Table 3
Physical properties of sawdust

Specific gravity	2.68
Density (g/m^3)	0.83
Porosity (%)	37
Particle size (μm)	< 65μ, Well graded

Density Test

Change in initial density and final density of composite samples shows the percentage void content present in the sample (Menard, 1999). Density test results for short coir fibre composites are shown in Table 4. Void content increased with increase in filler up to 10% mass and a decrease was noticed with 15% filler by mass. This may be due to improved fabrication skill. A higher void content percentage of 27.7% was observed for 10% filler by mass. This could have been the result of agglomeration of filler and matrix, resulting in improper wetting due to higher porosity at the agglomerated area. Surface voids and pores were minimal, so the chance of water diffusing through the sample was limited. Short coir fibre epoxy composites showed lower void content than long coir fibre epoxy composites i.e. three times more than that of the short coir fibre epoxy composite. This was mostly due to agglomeration of long coir fibre offering resistance to epoxy towards uniform distribution. Due to a pile up of fibre the resin disintegrated, causing the specimen to fracture (Venkatasubramanian & Raghuraman, 2015). The density result of the short coir fibre epoxy composite shows an acceptable void content in the composite, confirming the fabrication route as a noble one. The overall density test results showed that short coir fibre gave a better end product with less void than did the long coir fibre epoxy composite.

Table 4
Void content in composite sample

Sample No.	1	2	3	4
Void content (%)	23.6	23.8	27.7	25.6

Water Absorption Test

Water absorption specimens of different proportions were machined as per ASTM D 570 standard and were immersed in distilled water (pH = 7) and sea water (pH = 8.1, salinity = 29 ppt). Figure 1 and 2 shows that weight gain of the composite increased with immersion time at the initial stage and later kept stable, indicating that the specimens were apparently saturated with moisture. The distilled water absorption percentage decreased with the increase in filler loading up to 10% by mass. This was because well graded fillers clogged the pores and micro cracks within the surface of the composite, causing a lower water absorption percentage. It has been reported that water absorption can significantly be minimised if the filler is thoroughly encapsulated by the matrix (Sanadi et al., 1995). A further increase in the filler resulted in increased water absorption percentage as in Figure 1.

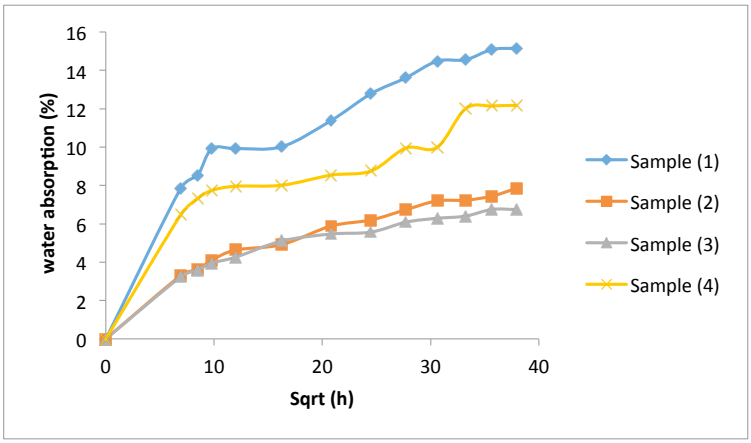


Figure 1. Percentage of distilled water absorbed

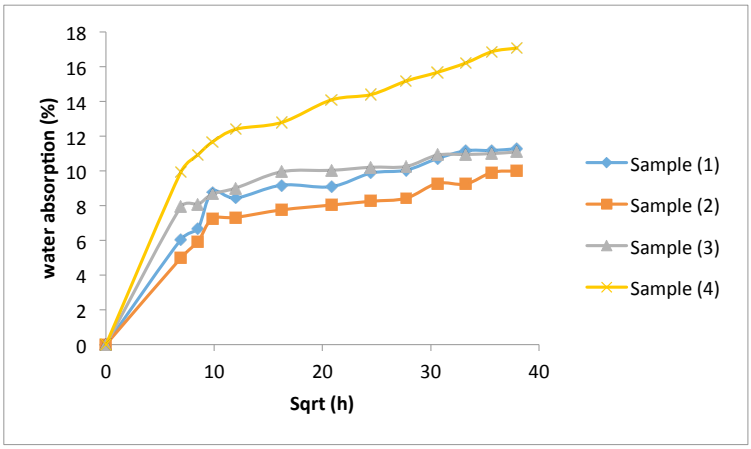


Figure 2. Percentage of sea water absorbed

From the density test it was clear that the third sample exhibited a higher void content. Figure 1 shows that lower water absorption was observed in the third sample; this was due to the fact that the sample machined for the distilled water absorption test had a lower void content than the overall composite laminate i.e. the voids were not uniformly distributed over the composite laminate.

The percentage moisture gain in samples immersed in sea water was higher than that of distilled water. This may be due to perforation of solution with greater pH, which induces more micro cracks leading to weight gain (Karbhari & Chu, 2005). Higher weight gain is due to the salts accumulated in the rough surface of the specimen; this was clear from the morphological analysis seen in Figure 9(e). When the filler concentration increased to 5% by mass, the water absorption percentage was reduced and with further loading, an increase in water absorption

was noticed. Increase in the sawdust filler elevated the agglomeration developed due to non-uniform dispersion of filler. Filler clouding in the composites escalated the water absorption percentage of the composites (Abdullah et al., 2011).

Table 5 shows the diffusion coefficient of the composite samples immersed in sea water and distilled water. The diffusion coefficient of the samples immersed in distilled water went on decreasing with the increase in filler content. This was because an increase in clouding of the filler load along with the matrix will withstand the passing of water molecules through the nano pores (Sunil et al., 2015). A sudden increase in absorption percentage in the fourth sample was noticed in Figure 1 after 10% water absorption. This was because when the composite was exposed to moisture, the hydrophilic nature of fibre causes it to swell. As a result, the brittle thermosetting resin developed microcracks, allowing a higher percentage of water to penetrate the composite. Further penetration of water into the interface through micro cracks was due to the high cellulose content of the fibre, leading to the composite failure (Bismarck et al., 2002). As the composite material undergoes cracks, capillarity mechanism and transport via micro cracks become more active (Dhakal et al., 2007).

Table 5
Diffusion coefficient of samples immersed in sea and distilled water

Sample No.	Diffusion coeff- Distilled water (m^2/sec) ($\times 10^{-13}$)	Diffusion coeff- Sea water (m^2/sec) ($\times 10^{-13}$)
1	4.62	3.92
2	4.56	3.36
3	4.06	2.52
4	3.80	3.58

The diffusion coefficient in sea water observed is less than that of distilled water. This clearly shows that most of the weight gain was due to salts accumulated in the composite sample surface. This was because the surface finish of the randomly orientated coir fibre composite was rough. Resin surface condition is influenced by rate of diffusion (Asmaashawky et al., 2013). As the surface of the samples had a minimum number of pores and voids, the overall diffusion rate of the short coir fibres was less than that of the long coir fibres.

Hardness Test

About seven sets of trials for hardness values in each sample were conducted, and the average of all the values exhibited the hardness value of each sample as shown in Figure 3. Due to agglomeration of filler at a higher percentage resulting in improper matrix wetting, the composite exhibited a larger variation in hardness value. This was seen in the third and fourth samples as seen in Table 6. The hardness value improved with the increase in filler percentage as increase in filler loading improves matrix surface resistance to indentation. A maximum value of 30.6 HV was observed with the fourth sample, which was 8.8% more than that of the long coir fibre epoxy composite.

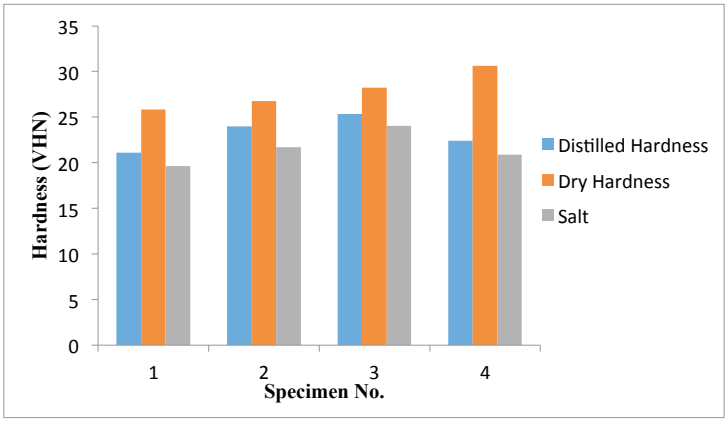


Figure 3. Hardness value of composite

Table 6
Change in hardness value (VHN) observed throughout the trails

Sample 1	Sample 2	Sample 3	Sample 4
26.2	24.91	29.37	36.14
24.62	29.36	38.16	35.13
29.08	24.91	25.92	25.20
29.37	29.37	24.62	29.08
25.34	26.92	25.77	28.80
25.63	25.05	31.24	31.30
22.46	22.66	29.80	28.50

An increase by 18.34% in hardness value was noticed compared to the composite with no filler. A drop in hardness value was noticed when subjected to distilled and sea water immersion. A higher drop by 26.7 and 31.69% for both distilled and sea water immersion was observed in the fourth sample due to the existence of more voids, sawdust agglomeration, improper wetting by matrix and higher water absorption. Figure 1 and 2 show that the fourth sample absorbed a higher moisture percentage than did the other composite with filler percentage. Increased water absorption resulted in bound water absorption and a drop in free water. In this situation, water diffused into the fibre cellulose network through capillaries and gaps between the low bound fibril area. Groups in the cellulose molecules were chemically linked by water. Water molecules in the cellulose network structure behave as a plasticiser and ruin the rigidity of the cellulose, permitting cellulose molecules to bulge independently. Simultaneously, the cellulose mass gets mushy, which in turn can easily alter the dimensions of the fibre with the execution of a load.

When absorption reaches the threshold level, bound water and free water continue as a reservoir, leading to softening of fibres and the weakening of fibre matrix adhesion, resulting in a reduced hardness value (Alomayria et al., 2014). The intra molecular hydrogen bond between polar groups of resin have to be broken under low temperature in order to develop bonding between the OH groups and water, forming hydrogen bonds between the water molecules. The

polar hydroxyl groups of the network will disrupt the inter change hydrogen bonding, thereby altering the structure of the molecule due to which the mechanical performance is affected (Nogueiraet al., 2001). Higher degradation is observed for those samples immersed in sea water due to NaCl, which exists as cations and anions, causing more damage to the matrix, fibre and their interface (Ashbee & Wyatt, 1969).

Tensile Test

Figure 4 shows the effect of filler percentage on tensile strength. It can be noticed that a maximum tensile strength of 20.8 MPa was exhibited by the sample with 5% filler by mass. With further increase in filler, a drop in tensile strength was observed due to the increase in interfacial area with worsening interfacial bonding between filler and matrix (Mosadeghzad et al., 2009). At a higher filler percentage, a drop in tensile strength occurred due to the agglomeration of the filler, which results in improper curing of the composites (Rahul et al., 2014). At a higher filler percentage, the filler starts engulfing the matrix isolating the hardener and resin, causing improper curing of the composite. The result exhibited an increase by 57.40% in tensile strength compared to that of the long coir fibre epoxy composite.

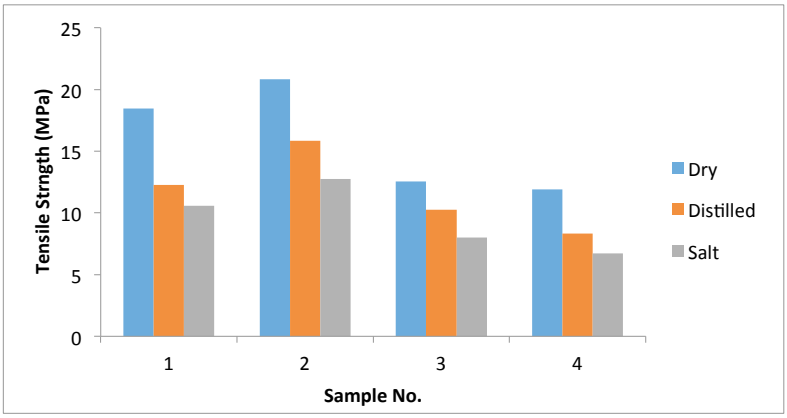


Figure 4. Tensile strength of composites

It is believed that fillers act as flaws at a higher filler mass fraction due to lack of resin, which wets fibre surfaces, resulting in inept stress transfer. An increase in the filler also increased the micropores between the filler and matrix, which in turn weakened the adhesion between the matrix and filler interface, affecting tensile strength and modulus (Hardinnawirda & Sitiribiatull, 2012).

A higher loss in tensile strength was noticed in the third sample i.e. the sample with 10% filler by mass. The density test revealed that the third sample had higher void content, resulting in a greater drop in tensile strength. Lack of resin and improper curing in the third and fourth samples resulted in fibre pull out.

A decrease in mechanical properties was noticed in those specimens immersed in distilled and sea water. The specimen immersed in sea water solution had the largest reduction, with a reduction of 42.79, 38.65, 36.21 and 43.55% for the first, second, third and fourth specimens. The specimen immersed in water had the smallest reduction, with a reduction of 33.61, 26.3, 28.01 and 29.9% for the first, second, third and fourth specimens.

Higher degradation was observed in the first specimen due to localised shrinkage of the epoxy occurring because of non-uniform distribution and in the fourth specimen due to epoxy deficiency, resulting in improper wetting leading to higher water absorption settled as both bound water and free water. The comparison in strength reduction indicated that the presence of NaCl in sea water solution led to faster degradation in tensile strength compared with in water that had no ions as was the case with distilled water. The possible reason is that the alkali metal oxides in sea water might accelerate transmission of water into the fibre/matrix interface, resulting in debonding (Ashbee & Wyatt, 1969).

Figure 5 shows the tensile modulus of the specimens. The tensile modulus increased with increase in filler loading up to 5% by mass. A higher tensile modulus of 5.55 GPa was observed in the second sample. This was due to the fact that filler material resulted in more uniform distribution of the epoxy matrix material, resulting in improved bonding between the matrix-fibre interphase. At higher filler loading, the tensile modulus decreased (Agomide et al., 2009). The possible reason is that at higher percentage of filler, the observed drop in tensile behaviour of the composites occurred due to the agglomeration of the filler molecules around the matrix, preventing the proper curing of the composite (Rahul et al., 2014).

Tensile modulus reduced by 40.9, 49, 55.1 and 65.6% in the first, second, third and fourth samples, respectively, that were immersed in distilled water. The percentage was 70.6, 64.7, 75.2 and 79.8% for the first, second, third and fourth samples immersed in salt water. A lower drop of tensile modulus in the second specimen immersed in sea water was due to the lower water absorption percentage as observed in Figure 5.

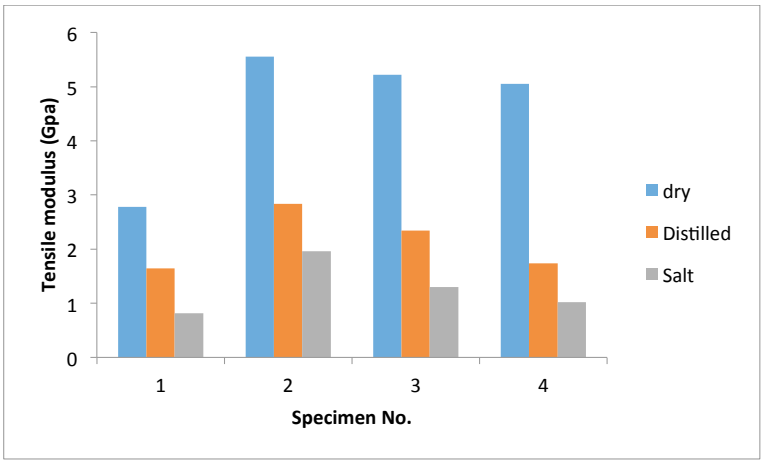


Figure 5. Tensile modulus of specimens

Flexural Test

The results of different filler loading are plotted in Figure 6. Initially, the flexural property increased with the increase in filler loading by mass up to 5%, and from 10% filler loading by mass, there was a decrease in flexural strength. The reduction in mean flexural strength of specimen above 5% filler by mass may have been due to the fact that the deficiency of epoxy resin and increased filler led to poor wetting of the fibre. This in turn might have led to matrix failure as shown in Figure 7. The same was reported by Binu (2012). Improved interfacial adhesion between the fibre and matrix resulted in higher flexural strength of the composites; this was mainly due to alkali treatment of the coir fibre. An increase in strength by 5.9% was observed in the filler compared with the short fibre composite without filler. An increase by 54.8% was observed in flexural strength compared with that of the long coir fibre epoxy composite.

From Figure 6 it is clear that lower degradation took place in the third sample due to less water absorption. A decrease by 14.7, 14.8, 6.25 and 18.38% when immersed in distilled water and 24.23, 21.3, 14.5 and 40.9% when immersed in salt water was noticed in the first, second, third and fourth specimens. In both the immersions, higher degradation was observed in the fourth specimen due to epoxy deficiency, resulting in improper wetting that led to higher water absorption that settled as both bound water and free water.

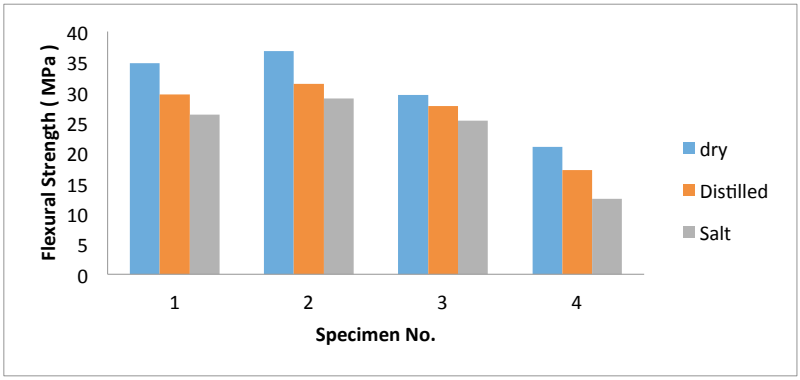


Figure 6. Flexural strength of composites

Figure 7. Matrix crack observed in composite specimen

Figure 8 shows the flexural modulus of short coir fibre composites. The flexural modulus increased with an increase in sawdust. Higher flexural modulus in dry condition was observed in the third sample, in which the molecules were tightly bound to each other (Agomide et al., 2009). With further increase in filler, a decrease in flexural modulus was noticed due to epoxy deficiency, resulting in improper wetting of the fibre and sawdust agglomeration.

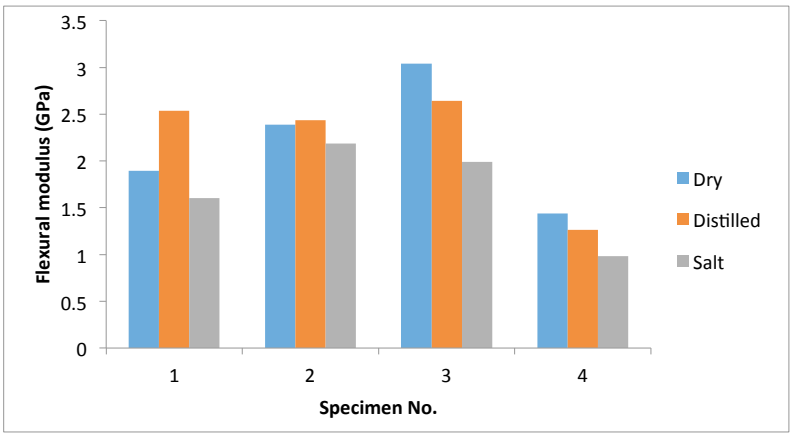


Figure 8. Flexural modulus of the composite

The strain observed in the third sample was lower even though the stress was high. If, however, the molecules were loosely bonded to each other, a relatively small amount of stress would have caused a large amount of strain (Agomide et al., 2009). The strain was higher for the first and second samples immersed in distilled water compared to the dry specimens of the same; this could have been due to the fact that fibre swelling caused by water absorption crowds the space between the fibre and polymer matrix and eventually leads to an increase in the mechanical properties of the composites (Karmakar & Hoffmann, 1994). Similar observations have been reported for jute fibre reinforced polymer composites (Ayensu, 2000).

In the third sample, due to the increased voids, water absorbed may have been present as free water, reducing the integrity of the matrix interface. When the fibre swelled, the weakened matrix had a higher chance to crack, failing to fill in the gaps. At higher filler loading resulting in water absorption led to higher elongation as the water molecules acted as a lubricant. The fibres could slip over one another during loading, resulting in extra extension and elongation (Azwa et al., 2013).

Mechanical properties like tensile, flexural and hardness degrade at a higher moisture rate due to generation of hydrogen bonding across water molecules and cellulose fibre. A high percentage of hydrogen bonding is formed between cellulose macromolecules and polymers due to hydroxyl groups (–OH) present in hydrophilic natural fibres (Sombastsompop & Chaochanchaikul, 2004).

Low moisture resistance is exhibited by natural fibres like coir due to the presence of a significant –OH group percentage. This alters the dimensional aspects of composites and causes

deficient interfacial bonding between the fibre and matrix, causing a decrease in the mechanical properties (Diamant et al., 1981). Water absorbed in polymers is generally divided into free water and bound water (Alomayria et al., 2014). Water with molecules that are free to move through available space is considered free water and water with molecules that are dispersed in epoxy matrix and adhered to polar groups is termed bound water (Maggana & Pissis, 1999).

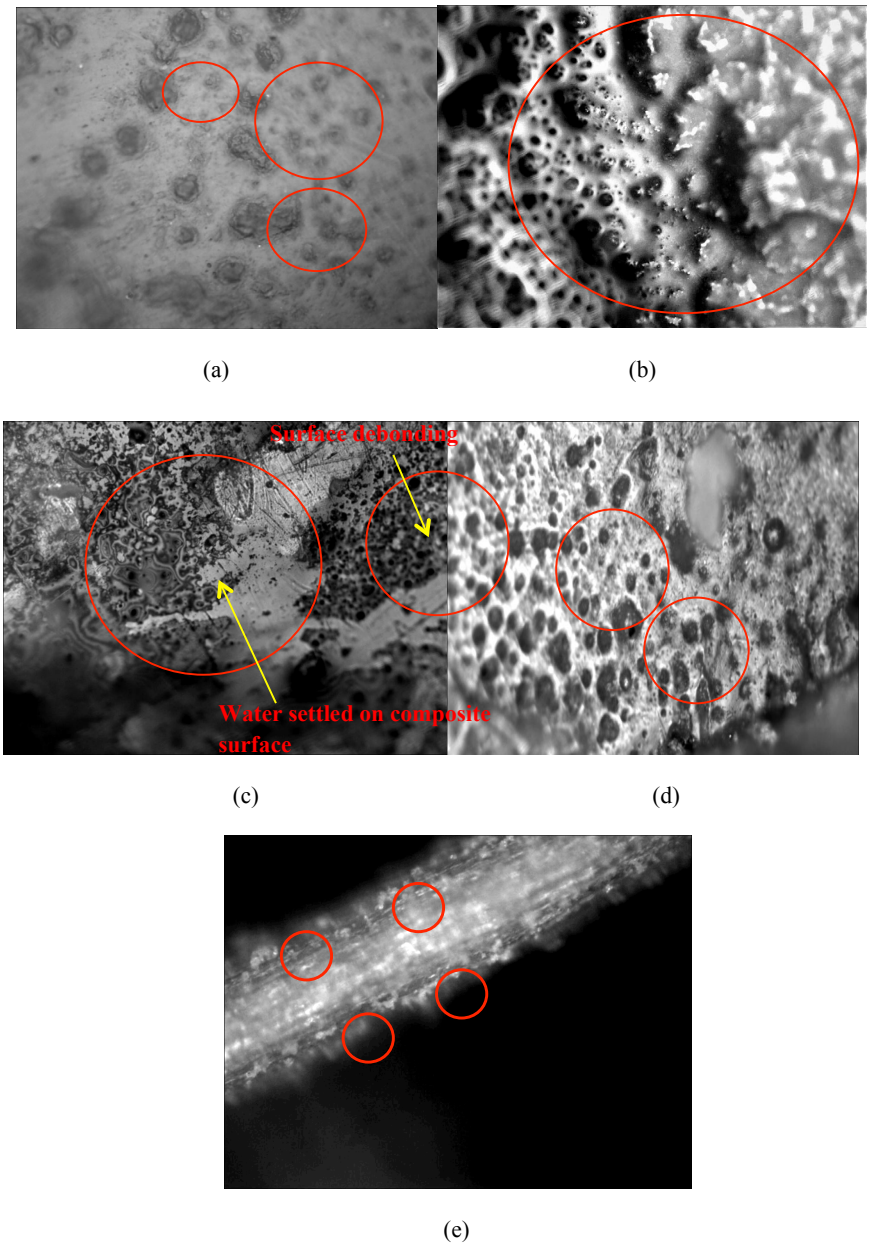


Figure 9. Microstructure analysis of composite samples (a) Sawdust agglomeration at higher filler percentage, (b) Non-uniform wetting at lower matrix percentage, (c) Water settled on uneven composite surface after ageing, resulting in surface debonding, (d) Pores formed on composite surfaces (e) Salt settled on fibre surface

Morphological Analysis

Microstructure analysis was carried out to study the surface morphology of the composite specimen. Alkalisation treatment of coir fibre resulted in fibre roughness, which exhibited advanced fibre and matrix, concluding substantially higher mechanical properties. As filler percentage increased, agglomeration was noticed in Figure 9(a), influencing improper curing of matrix, being a major rationale for varying hardness value and drop in tensile strength. Often, the composite fabricated by the hand lay-up process results in voids, cracks and flaws (Libo & Nawawi, 2015).

Figure 9(d) shows voids formed due to gas porosity and sawdust porosity, which have an effect on mechanical performance. Voids can also be formed as a result of improper matrix wetting and by implementation of the hand lay-up technique as it is person dependent. Figure 9(b) shows a web-like structure due to improper wetting of the epoxy matrix in the fourth composite sample, which is a major cause for reduced mechanical performance.

Figure 9(c) shows water accumulated on the uneven surface of the composite, resulting in weight gain. The water accumulated on the surface can soften the matrix surface, resulting in surface matrix debonding. This could be the effect of water absorption by the hydrophilic fibre due to the presence of -OH radicals. Incompatibility is higher when hydrophilic sawdust is adhered with hydrophobic epoxy matrix; this incapability was utilised by the water molecules, forming a hydrogen bond with free -OH radicals, causing sawdust swelling. This swelling resulted in matrix debonding and cracks that resulted in the mechanical degradation.

CONCLUSION

Composites fabricated using a short coir fibre epoxy showed variation in density. A higher diffusion rate was observed in the sample without a filler that was immersed in distilled water and sea water, respectively due to the higher percentage of water absorption. An increase by 18.34% in hardness value was noticed compared to composites with no filler. An increase in tensile strength by 12.75%, flexural strength by 5.94% and a higher tensile modulus of 5.55 GPa in dry condition was observed in the specimen with a 5% filler by mass. A higher flexural modulus of 2.3 GPa i.e. an increase by 60.4% in dry condition was observed in the specimen with a 10% filler by mass due to tightly bonded filler molecules. Reduction in tensile strength, hardness and flexural strength was observed in specimens immersed in distilled and sea water, respectively. Flexural modulus for the specimen without filler immersed in distilled water showed a remarkable increase, but with an increase in filler, higher degradation due to the hydrophilic nature of the fibre and filler was noticed, resulting in reduced flexural modulus. Flexural modulus for the specimen immersed in sea water increased with an increase in the filler up to 5% loading by mass and then reduced with a further increase in loading by mass. An improved fabrication method is suggested to fabricate high strength and durable short coir fibre epoxy composites with sawdust filler.

REFERENCES

Abdullah, A. H., Rusel, D. S., & Abdulwahab, A. S. (2011). Water absorption and mechanical properties of high-density polyethylene/egg shell composite. *Journal of Basrah Researches (Sciences), 37*(3A), 36–42.

Abilash, N., & Sivapragash, M. (2013). Environmental benefits of eco-friendly natural fibre reinforced polymeric composite materials. *International Journal of Application or Innovation in Engineering and Management, 2*(1), 53–59.

Agnew, J. M., & Leonard, J. J. (2003). The physical properties of compost (Literature review). *Composite Science and Utilization, 11*, 238–264.

Alomayria, T., Assaedi, H., Shaikh, F. U. A., & Low, I. M. (2014). Effect of water absorption on the mechanical properties of cotton fabric-reinforced geopolymer composites. *Journal of Asian Ceramic Societies, 2*, 223–230.

Araki, Y., & Terazawa, M. (2004). Physical properties of sawdust and soil (in Japanese). In *Proceedings of the Hokkaido Branch of Japan Wood Research Society, 36*, 67–70.

Aramide, F. O, Oladele, I. O., & Folorunso, D. O. (2009). Evaluation of the effect of fibre volume fraction on the mechanical properties of a polymer matrix composites. *Leonardo Electronic Journal of Practices and Technologies, 14*, 134–141.

Ashbee, K. H. G., & Wyatt, R. (1969). Water damage in glass fibre/resin composites. *Proceedings of the Royal Society: A Mathematical, Physical and Engineering Sciences, 53*, 312–553.

Asmaashawky, K. (2013). The effect of particles as additives on water absorption for epoxy resin. *International Journal of Application or Innovation in Engineering and Management, 2*(5), 131–136.

Ayensu, A. (2000). Interfacial debonding of natural fibre reinforced composites. *Quarterly Science Vision, 6*(1), 25–34.

Ayrilmis, N., Jarusombuti, S., Fueangvivat, V., Bauchongkol, P., & White, R. H. (2011). Coir fibre reinforced polypropylene composite panel for automotive interior applications. *Fibres and Polymer, 12(*7), 919–926.

Azwa, Z. N., Yousif, B. F., Manalo, A. C., & Karunasena, W. (2013). A review on the degradability of polymeric composites based on natural fibres. *Materials and Design, 47*, 424–442.

Badu, M., Twumasi, S. K., & Boadi, N. O. (2011). Effects of lignocellulosic in wood used as substrate on the quality and yield of mushrooms. *Food and Nurition Science, 2*, 780–784.

Baker, S. M., Richard, T., Zhang, Z., & Monteiro, S. (1998). Determining the free air space inside compost mixtures using a gas pycnometer. *American Society Agricultural Engineering* (paper no. 984094).

Binu, H. (2012). *A study on mechanical and thermal behaviour of coir fibre reinofrced epoxy composites*. Mechanical engineering thesis report, National Institute Rourkela, India.

Bismarck, A., Askargorta, I. A., Springer, J., Lampke, T., Wielage, B., & Stamboulis, A. (2002). Surface characterization of flax, hemp and cellulose fibres; Surface properties and the water uptake behaviour. *Polymer Composite, 23*, 872–94.

Bledzki, A. K., & Gassan, J. (1999). Composites reinforced with cellulose based fibres. *Progress in Polymer Science, 24*, 221–74.

Bouma, J., Rao, P. S., & Brown, R. B. (2003). *Soil as a porous medium: Basics of soil-water relationships-Part I.* USA: University of Florida, IFAS extension.

Campbell, W. B., Bamiord, K. F., & Simatupang, M. H. (1966). *Zur Chemic Und Biochemic Des Teakholzes* (Tectona Grandis L. Fil.). *Holz Als Roh-Und Werkstoff,* 24, 190–204.

Chern, C. E, Nor, A. I., Norhazlinv, Z., Hidayah, A., & Wan, M. (2014). Impact strength and flexural properties enhancement of methacrylate silane treated oil palm mesocarp fibre reinforced biodegradable hybrid composites. *Hindawi Publishing Corporation, 2014*(2014), 1-8.

Deac, T., Fechete-Tutunaru, L., & Gaspar, F. (2016). Environmental impact of sawdust briquettes use-experimental approach. *Energy Procedia, 85*, 178–183.

Dhakal, H. N., Zhang, Z. Y., & Richardson, M. O. W. (2007). Effect of water absorption on the mechanical properties of hempfibre reinforced unsaturated polyester composites. *Composites Science and Technology, 67*(7–8), 1674–1683.

Diamant, Y., Marom, G., & Broutman, L. J. (1981). The effect of network structure on moisture absorption of epoxy resins. *Journal of Applied Polymer Science, 26*, 3015–3025.

Hardinnawirda, K., & Sitirbiatull, A. (2012). Effect of rice husks as filler in polymer matrix composites. *Journal of Mechnical Engineering and Science, 2*, 818–186.

Horisawa, S., Sunagawa, M., Tamai, Y., Matsuoka, Y., Miura, T., & Terazawa, M. (1999). Biodegradation of nonlignocellulosic substances II: Physical and chemical properties of sawdust before and after use as artificial soil. *Journal of Wood Science, 45*, 492–497.

Joshi, S. V., Drzal, L. T., Mohanty, A. K., & Arora, S. (2004). Are natural fibre composites environmentally superior to glass fibre reinforced composites? *Composites Part A: Applied Science and Manufacturing, 35*, 371–376.

Karbhari, V. M., & Chu, W. (2005). Degradation kinetics of pultruded E-glass/vinylester in alkaline media. *ACI Materials Journal, 102*(1), 34–41.

Karmakar, A. C., & Hoffmann, A. G. H. (1994). Influence of water uptake on the mechanical properties of jute fibre reinforced polypropylene. *Journal of Applied Polymer Science, 54*, 1803–1807.

Khan N. M. D., Roy, J. K., Akter, N., Zaman, H. D., & Islam, T. (2012). Production and properties of short jute and short E-glass fibre reinforced polypropylene based composites. *Open Journal of Composite Materials, 2*, 40–47.

Libo, Y., & Nawawi, C. (2015). Effect of water, seawater and alkaline solution ageing on mechanical properties of flax fabric/epoxy composites used for civil engineering applications. *Construction and Building Materials, 99*, 118–127.

Maggana, C., & Pissis, P. (1999). Water absorption and diffusion studies in an epoxy resin system. *Journal of Polymer Science, 37*, 1165–82.

Menard, K. P. (1999). *Dynamic mechanical analysis: A practical introduction* (p. 208). Boca Raton: CRC Press.

Mosadeghzad, Z., Ahmad, I., Daik, R., Ramli, A., & Jalaludin, Z. (2009). Preparation and properties of acacia sawdust/UPR composite based on recycled PET. *Malaysian Polymer Journal, 4*(1), 30–41.

Nam, T. H., Ogihara, S., Tung, N. H., & Kobayashi, S. (2011). Effect of alkali treatment on interfacial and mechanical properties of coir fibre reinforced poly (butylene succinate) biodegradable composites. *Composites Part B: Engineering, 42*(6), 1648–1656.

Nogueira, P., RamiRez, C., Torres, A., Abad, M. J., Cano, J., LoPez, J., Lopez-Bueno, I., & Barra, L. (2001). Effect of water sorption on the structure and mechanical properties of an epoxy resin system. *Journal of Applied Polymer Science, 80*, 71–80.

Odian, G. (2004). *Principles of polymerization* (4th Ed.). New York: Wiley-Interscience.

Omrani, E., Menezes, P. L., & Rohatgi, P. K. (2016). State of the art on tribological behavior of polymer matrix composites reinforced with natural fibers in the green materials world. *Engineering Science and Technology, an International Journal, 19*(2), 717–736.

Papanicolaou, G. C., Xepapadaki, A. G., & Zarouchas, D. (2008). Effect of water uptake on the creep behavior of glass-epoxy composite. In *13th Eurepan Conference on Composite Materials*. Stockholm, Sweden.

Pickering, K. L., Arunan, M. G., & Lee, T. M. (2016). A review of recent developments in natural fibre composites and their mechanical performance. *Composite: Part A, 83*, 98–112.

Rahul, K., Kausik, K., Prasanta, S., & Sumit B. (2014). Study of mechanical properties of wood dust reinforced epoxy composite. 3rd International Conference on Materials Processing and Characterisation (ICMPC 2014). *Procedia Materials Science, 6*, 551–556.

Ramadevi, P., Dhanalakshmi, S., Basavaraju, B., Raghu, P. G., & Chikkol, V. S. (2015). Influence of fibre content and effect of chemical pre-treatments on mechanical characterizationof natural abaca epoxy composites. *Indian Journal of Science and Technology, 8*(11), 1–11.

Ramesh, M. B., Kumar, K. P., & Reddy, K. H. (2012). Mechanical property evaluation of sisal-jute-glass reinforced polyester composites. *Composite Part B: Engineering, 48*, 1–9.

Ruhlmann, J., Korschens, M., & Graefe, J. (2006). A new approach to calculate the particle density of soils considering properties of the soil organic matter and the mineral matrix. *Geoderma, 130*(3), 272–283.

Sanadi, A. R., Caulfield, D. F., Jacobson, R. E., & Rowell, R. M. (1995). Renewable agriculture fibres as reinforcing fillers in plastics: Mechanical properties of kenaf fibre polypropylene composites. *Industrial and Engineering Chemistry Research, 34*(5), 1889–1896.

Sheng, H. H., Rung, S. C., Yuan, L. C., Min, H. C., Kuo, C. C., & Wei, F. S. (2012). Biphenyl liquid crystalline epoxy resin as a low-shrinkage resin-based dental restorative nanocomposite. *Acta Biomaterialia, 8*, 4151–4161.

Sombastsompop, N., & Chaochanchaikul, K. (2004). Effect of moisture content on mechanical properties, thermal and structural stability and extruded texture of poly (vinyl chloride)/wood sawdust composites. *Polym International, 53*, 1210–1218.

Srinivasababu, N., Murali, M., & Suresh, J. (2009). Experimental determination of tensile properties of okra, sisal and banana fibre reinforced polyester composites. *Indian Journal of Science and Technology, 2*(7), 35–38.

Stamboulis, A., Baillie, C. A., Garkhail, S. K., Van, H. G. H., & Peijs, T. (2000). Environmental durability of flax and their composites based on polypropylene matrix. *Applied Composite Materials, 7*, 273–294.

Sunil, K. M., Gopalakrishna, M. G., Sujay, M., & Shivakumar, H. R. (2015). Water absorption and diffusion properties of areca fibre reinforced epoxy and vinyl ester composites. *International Journal of Advance Research in Science and Engineering, 4*(1), 147–1652.

Ticoalu, A., Aravinthan, T., & Cardona, F. (2010). A review of current development in natural fibre composites for structural and infrastructure applications. In *Southern Region Engineering Conference* (pp. 113-117). Engineers Australia.

Venkatasubramanian, H., & Raghuraman, S. (2015). Mechanical behaviour of abaca-glass-banana fibre reinforced hybrid composites. *Journal of Engineering Science and Technology, 10*(8), 958–971.

Verma, D., Gope, P. C., Shandilya, A., Gupta, A., & Maheshwari, M. K. (2013). Coir fibre reinforcement and application in polymer composites: A review. *Journal of Material Environment Science, 4,* 263–276.

Yousif, B. F., & Ku, H. (2012). Suitability of using coir fibre/polymeric composite for the design of liquid storage tanks. *Materials and Design, 36*, 847–853.

Zainab, R. (2009). Water absorption of denture base polymer reinforced with glass fibres in chopped and woven form. In *Proceedings of The College of Science 4th Scientific Conference* (pp. 68-73). College of Science, University of Babylon, Republic of Iraq.

Pertanika J. Sci. & Technol. 26 (1): 283 - 298 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Green Wall for Retention of Stormwater

J. T. Lau* and D. Y. S. Mah

Department of Civil Engineering, Faculty of Engineering, University Malaysia Sarawak, 94300 UNIMAS, Kota Samarahan, Sarawak, Malaysia

ABSTRACT

Urbanisation increases the level of imperviousness in a catchment, and more runoff is converted from rainfall in urban areas. To mitigate this adverse situation, dispensed green infrastructure presents the best solution for delivering results in reducing stormwater impact. Green roofs and rain gardens are extensively studied and widely available in the literature, but this is not the case for green walls, which more often than not, are treated as ornaments. Thus, this study developed a computer-aided stormwater model that incorporates a green wall to investigate its effectiveness as an urban drainage system. The effectiveness of employing a green wall as a stormwater component is tested using USEPA SWMM 5.1 and the embedded bioretention cell interface. Four simulation models according to different conditions and precipitation input are tested, compared and discussed. The conditions include investigation of different soil types, average recurrence interval (ARI) and storm duration with design and observed rainfall. The results reveal that synthesis precipitation data, used in Scenario 1, 2 and 3, decreased runoff by more than half, at 55% on condition of one-year ARI and 5 minutes of storm duration. Meanwhile, Scenario 4 also shows a repetition of runoff reduction by half after the integration of the green wall using the observed rainfall data. Thus, it is verified that a green wall can be effectively used as an urban drainage system in reducing surface runoff.

Keywords: Bioretention, green wall, runoff, SWMM, urban stormwater management

ARTICLE INFO

Article history:
Received: 26 June 2016
Accepted: 09 May 2017

E-mail addresses:
dante.lau91@hotmail.com (J. T. Lau),
ysmah@feng.unimas.my (D. Y. S. Mah)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

The process of urbanisation turns natural ground cover into urban infrastructure or utility developments. Impervious surfaces such as roofs, paved roads and parking lots have expanded significantly together with post-development progress (see Figure 1). Consequently, the infiltration of stormwater into the ground as depression storage is reduced with the gradual elimination of

vegetation as a natural filter. Thus, overland flows tend to travel faster and a huge quantity of runoff is discharged into urban stormwater conveying systems. Surface runoff is increased in urbanised watersheds, creating greater peak discharge. As the consequence of pre-development flow regime changes, natural disasters like flash floods, occur when the capacity of a drainage system fails to sustain the overwhelming quantity of runoff.

Hence, new, exhaustive and integrated stormwater management strategies are now required to underpin the Malaysian government's target of achieving sustainable urban drainage systems nationwide (DID, 2002). These new strategies incorporate various aspects of drainage, including runoff source control, management and delayed disposal of a catchment area on proactive and multifunction bases.

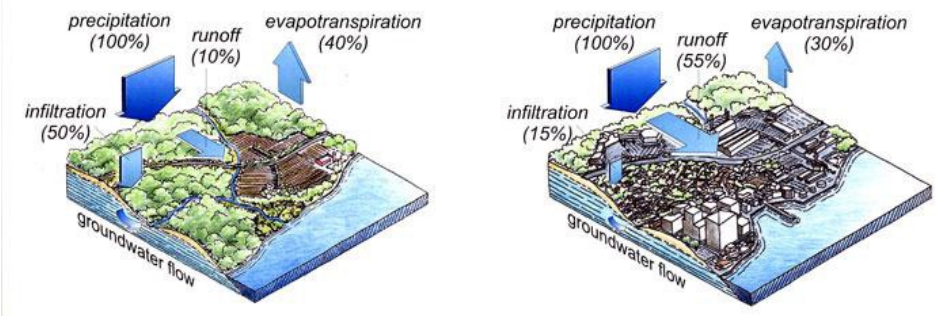


Figure 1. Typical degree of impervious areas that affect stormwater runoff, from (left) pre-development to (right) post-development (Commonwealth of Massachusetts, 2008)

Water Sensitive Urban Design

There are several well-known best practices of stormwater management used around the world that have been applied in urban developments of different countries, and one of them is Water Sensitive Urban Design (WSUD). WSUD is gaining popularity as an important element in sustainable supply planning in urban areas and has the added advantage of contributing to flood mitigation and maintaining safe water quality. Nevertheless, in order to resolve issues of high cost, land space utilisation and aesthetics of metropolitan areas, innovative stormwater management tools have emerged and been implemented. The Malaysian Urban Stormwater Management Manual (MSMA) has introduced a storage-orientated retention system that is water sensitive. The stormwater best management practices (BMPs), involving greenery and live plants, are designed to promote evapotranspiration and infiltration while minimising or delaying runoff from stormwater events (DID, 2002; 2012).

Green Infrastructure

Green infrastructure (GI) manages water and creates healthier urban environments by utilising vegetation, soil and natural processes. Considering that stormwater runoff is generated across distributed areas, the application of dispensed green infrastructure presents the best approach for delivering manifold ideal results in reducing stormwater impact. At the scale of a neighbourhood

or site with small linear features, GI is referred to as a stormwater management system that mimics the nature of soaking up and storing water, for instance bioretention (well-known feature that creates 'rain gardens'), green roofs or green walls (USEPA, 2014). Green roofs and rain gardens are extensively studied and widely available in the literature, but this is not the case for green walls, which more often than not, are treated as ornaments. Therefore, it was the intention of this project to look into the applicability of green walls in capturing runoff from roof tops. Computer modelling of the mentioned system provided initial results to guide the expected working of the actual model.

Green Wall

The terminology "green wall" refers to all forms of vegetated wall surfaces (GRHC, 2008). A green wall is basically a bioretention system, but it is structured in a vertical manner on a façade or wall, without the traditional requirements of space by sacrificing built-up areas. If the conventional bioretention system is a component of the urban runoff control, then theoretically, a green wall should have the same function.

Incorporated as part of a sustainable urban drainage system, green walls can mitigate water runoff and reduce stormwater flows (Green over Grey, 2009). Percolation of rainfall within modular green walls reduces the runoff rate (see Figure 2) and offers true benefits to urban stormwater management (Loh, 2008). Stormwater can be gathered for the purpose of irrigating a green wall, which in turn increases on-site infiltration and evapotranspiration. Several preliminary studies suggest that these systems retain as much as 45% to 75% of rainfall (Webb, 2010). In addition, green walls might become one of the effective stormwater management systems via vertical planting as wall area far exceeds roof area, especially in urban development areas (Kew et al., 2013). However, it is also reported that green walls hold less potential in producing much better results than green roofs (Higgs et al., 2011).

Figure 2. Modular green wall (GRHC, 2008)

METHODOLOGY

The study site was situated within Central City, which lies strategically between the Kuching and Samarahan link way (see Figure 3) in the state of Sarawak. Kota Samarahan has been on the Government's development radar screen for the past 10 years, and it seems that this will remain the case in the foreseeable future. The township experiences a high growth rate of economic development due to its function as a hub for higher education and technology. Due to rapid property and infrastructural development, flash floods often hit residential areas lying nearby. Therefore, Central City was chosen for this study to investigate ways for combating this problematic impact of urbanisation.

The main idea of this study was to propose green walls as a component of the local urban drainage system. The main goal was to devise an effective stormwater management system and to reduce the velocity of runoffs from rainfall events to downstream reaches. This gives a clearer picture of the objective of this study as it reduces the scope of the control design variables. Modular green walls were chosen as they have a sufficient volume of growing medium with retaining characteristics to control stormwater. Dependent variables in the design included the size of the modular cell, types of planting media and design rainfall.

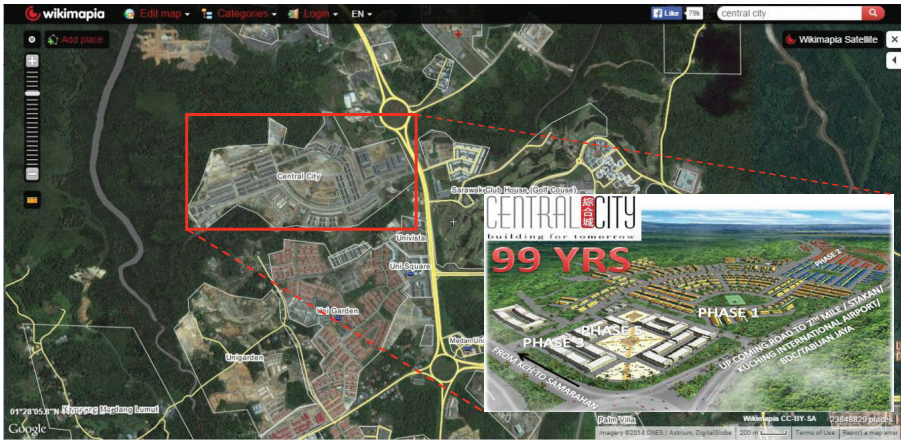


Figure 3. Aerial map of Central City via satellite image (http://wikimapia.org)

The study site was then narrowed down to a specific property in Central City. A shophouse in Phase 3 was chosen as the best option to implement the study of green walls. The major reason for this preference was that the three-storey shophouse had a relatively flat plain wall in front of the building and this available space could be put into good use. Apart from stormwater management, the green wall enhanced the aesthetic view of the commercial building and significantly reduced the urban heat island effect in Central City. A corner unit of a commercial building was chosen for a maximum roof catchment area of 139.08 m^2 for the design (see Figure 4).

Figure 4. Selected corner unit of three storey shophouse painted with orange (MD Kwang Tai, 2010)

The Elmich Green Wall was adapted for use in this project. The Elmich Green Wall is a modular system that consists of Elmich Vertical Greening Modules (VGMs); each module encases a VGM bag containing planting media, metal support frames and anchoring pilasters as shown in Figure 5 (Elmich, 2008). The green wall proposed in this study was assembled in front of a column of the commercial building. The top received runoff from the roof; then, the water slowly infiltrated the modules by gravitational force to the ground level, flowing finally into the culvert and roadside drain. The proposed size for a single green wall module was: height = 700 mm, width = 700 mm and depth = 200 mm. There were a total of 17 modules to be assembled in a straight upward manner and parallel to the column, which was approximately the height of the building.

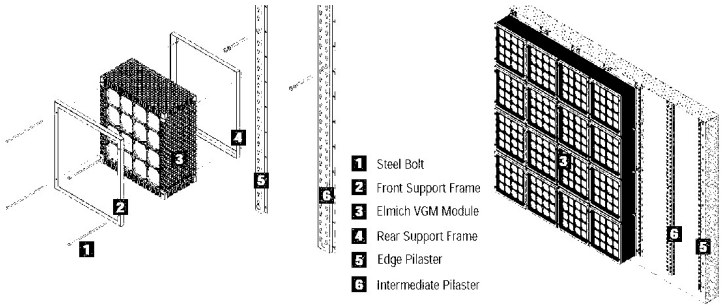


Figure 5. Elmich vertical greening modules (Elmich, 2008)

In this study, the soil for the green wall was required to allow water to pass through relatively fast. Thus, the recommended range of soil permeability was around 0.5 to 6.0 in/hr. Hence, four types of soil, namely sand, loamy sand, sandy loam and loam were chosen for analysis; the aim was to determine the best type among the four growing media for optimum performance of the

green wall as an urban drainage system. Hydraulic conductivity and other useful parameters for the four selected soils are shown in Appendix A.

The performance of the green wall system was assessed within a range of ARI with 1, 2, 5, 10, 20, 50 and 100 years and storm duration of 5, 10 and 15 minutes to determine the most suitable values for satisfactory performance of the system. Design rainfall intensity (mm/hr) depends on duration (minute) and ARI (year). In this study, the intensities of different ARI and storm duration in Kota Samarahan were estimated using the computerised intensity-duration-frequency (IDF) curves generated by the Department of Irrigation and Drainage (DID) Sarawak. Apart from that, the actual data of observed rainfall (24-hour precipitation data) were used to examine the effectiveness of the green wall system as an urban drainage system. Two sets of hourly rainfall data were used for analysis: data from January 2014 represented the highest accumulated rainfall depth of the year, while data from February 2014 represented normal rainfall in Kota Samarahan, showing only about half of January's rainfall.

Using the Rational Method, peak flow, Q_1, for a 15-minute storm was manually calculated; the value derived was 0.005022 cms. Roof runoff, Q_2, for a 15-minute storm, was generated from SWMM simulation. Both results, Q_1 and Q_2, were compared, and it was found that there were no significant differences, as shown in Table 1. A comparison of 10- and 5-minute storms are presented in Table 2 and Table 3. The runoff generated by SWMM was calibrated and the performance of the green wall system was further investigated using the bioretention interface in the SWMM simulation.

Hand calculation using Rational Method can be represented as:

$$Q_1 = \frac{CiA}{360} \quad [1]$$

where

Q_1 = Peak flow (cms);
C = Runoff coefficient;
i = Average rainfall intensity (mm/hr); and
A = Drainage area (ha)

Given that at the existing condition,

- ARI = 1 year
- Rainfall intensity, i, corresponding to a 15-minute storm = 130 mm/hr
- Roof runoff coefficient, C = 1.0
- Roof catchment area, A = 139.08 m^2 = 0.013908 ha

thus, peak flow, $Q_1 = \frac{CiA}{360} = \frac{1.0(130)(0.013908)}{360} = 0.005022$ cms

Table 1
SWMM calibrations for a 15-minute storm

ARI (year)	Rainfall Intensity, *i* (mm/hr) corresponding to 15-minute storm	Peak Flow, Q_1 (cms)	SWMM-generated Roof Runoff, Q_2 (cms) for 15-minute storm
1	130	0.005022	0.005017
2	160	0.006181	0.006176
5	170	0.006568	0.006562
10	180	0.006954	0.006954
20	190	0.007340	0.007334
50	210	0.008113	0.008106
100	230	0.008886	0.008879

Table 2
SWMM calibrations for a 10-minute storm

ARI (year)	Rainfall Intensity (mm/hr) corresponding to 10-minute storm	Peak Flow, Q_1 (cms)	SWMM-generated Roof Runoff, Q_2 (cms) for 10-minute storm
1	130	0.005022	0.005015
2	160	0.006181	0.006174
5	170	0.006568	0.006560
10	180	0.006954	0.006952
20	190	0.007340	0.007333
50	210	0.008113	0.008106
100	230	0.008886	0.008879

Table 3
SWMM calibrations for a 5-minute storm

ARI (year)	Rainfall Intensity (mm/hr) corresponding to 5-minute storm	Peak Flow, Q_1 (cms)	SWMM-generated Roof Runoff, Q_2 (cms) for 5-minute storm
1	130	0.005022	0.005006
2	160	0.006181	0.006167
5	170	0.006568	0.006554
10	180	0.006954	0.006941
20	190	0.007340	0.007328
50	210	0.008113	0.008102
100	230	0.008886	0.008875

However, SWMM used another equation based on nonlinear reservoir representation (see Figure 6). Each subcatchment surface was treated as a nonlinear reservoir. Inflow came from precipitation and the runoff from any designated upstream subcatchment areas. Outflow consisted of infiltration, evaporation and surface runoff. The capacity of this 'reservoir' was the maximum depression storage, which is the maximum surface storage provided by ponding, surface wetting and interception. Surface runoff, Q, occurred only when the depth of water, d, in the 'reservoir, exceeded the maximum depression storage, d_p, in which case the outflow was given by Manning's equation:

$$Q = W\frac{1.49}{n}\left(d - d_p\right)^{5/3} S^{1/2} \quad [2]$$

where
W is the subcatchment's characteristic width;
S is slope;
n is Manning roughness value; and
Depth of water, d_p, over the subcatchment was continuously updated with time by solving numerically a water balance equation over the subcatchment. Therefore, the hand calculation is not shown here.

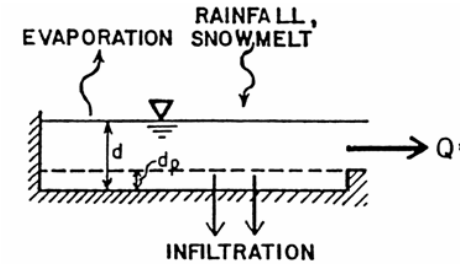


Figure 6. Nonlinear reservoir representation of a subcatchment (Huber & Dickinson, 1988)

RESULTS AND DISCUSSION

The process of employing a green wall as a stormwater component to reduce rainfall runoff was tested using Storm Water Management Model (SWMM) 5.1, which was carried out utilising an embedded bioretention cell interface. Four simulation models set to different conditions and precipitation input are shown in Table 4.

Table 4
Scenarios for modelling

Scenario	Precipitation Input	Condition
1	Design rainfall	Examine different soil media
2	Design rainfall	Examine average recurrence intervals (ARIs)
3	Design rainfall	Examine storm duration
4	Observed rainfall	Examine the effectiveness of green wall as an urban drainage system

After examining all the criteria in Scenario 1, 2 and 3 using design rainfalls, three soil types i.e loamy sand, sandy loam and loam showed equally good ability in surface runoff reduction, excluding sand, which acts as the control item. The following sizes were the design parameters that give the best result for optimum green wall performance: 12000 mm thickness, 1 year ARI and 5 minutes of storm duration (see Figures 7 to 9).

In Scenario 1, the thickness of the green wall was fixed at 12000mm and roof runoff was fixed at one-year ARI, while storm duration was 5 minutes for a single unit of corner shophouses. Hence, the types of growing media for a green wall are the dependent variables that determine the performance of the green wall. As shown in Figure 7, sand is expected to have the highest percentage of runoff reduction, followed by loamy sand, sandy loam and loam. The range is from 55.1% to 54.6%, which is a difference of only about 0.5%; thus all the soil types were considered equal in terms of runoff reduction.

All the soil types under study had higher composition of sand, which brought about a faster rate of storm water absorption. Although the permeability of clay and silt was low, their water-holding capability for retaining water was for a long period. The larger the soil particle size, the higher the conductivity. As a result, water infiltration rate increases in tandem with an increase in porosity or void between soil particles.

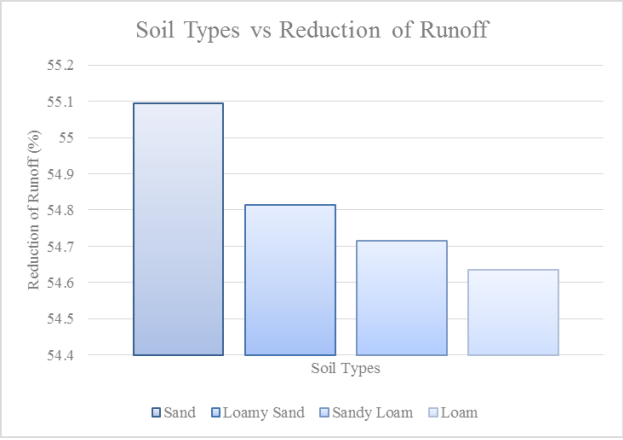


Figure 7. Reduction of runoff based on soil types for 5 minutes of one-year ARI event

In Scenario 2, the variables were given the following values: the thickness of the green wall was fixed at 12000 mm and storm duration was fixed at 5 minutes for a single corner shophouse. The ARIs were the dependent variables used to measure the performance of the green wall. Figure 8 shows the ARI traits for all the soil types, giving similar declivitous patterns from ARI year 1 until 100. The range of drop in runoff reduction was around 55% to 20%.

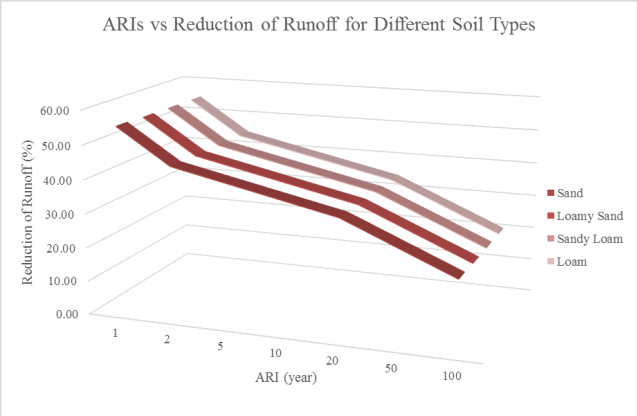


Figure 8. Reduction of runoff based on soil types and ARIs

According to DID (2010), a green wall (bioretention) is a minor system intended to collect, control and convey runoff from facilities in areas with relatively frequent storm events (recommended up to 10-year ARI) that minimises inconvenience and nuisance of flooding. The rationale of adopting a higher standard for minor systems in large commercial, business and industrial areas is that a minor system has a greater potential to cause damage and disruption in the event of flooding.

The trend of Scenario 1 was repeated for analysis, reiterating that the performance of the three soil types was the same. However, even at the extreme event of 100-year ARI, the three soil types were shown to be able to reduce about 20% of peak runoff.

In Scenario 3, the thickness of the green wall was fixed at 12000mm and the storm ARI was set at 1 year for a single corner shophouse. Hence, storm duration became the dependent variable that showcased the performance of the green wall. As shown in Figure 9, the storm duration traits for all the soil types showed similar declivitous patterns from 5 until 15 minutes. The drop in runoff reduction ranged from 55% to 22%.

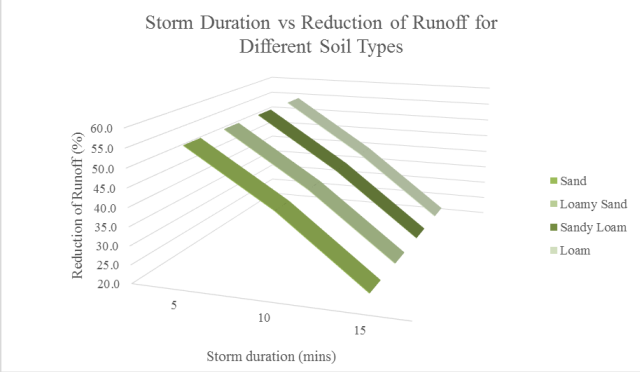


Figure 9. Reduction of runoff based on soil types and storm durations

Duration of storm is an important parameter that defines the intensity for a given ARI and thus, affects the resulting runoff peak. The storm duration that produces the maximum runoff peak traditionally is defined as the time of concentration – the sum of the travelling time to an inlet plus the time of travel in the stormwater conveyance system (DID, 2002). Although travel time from individual elements of a system may be short, the total nominal travel time of flow for all individual elements within any catchment to their points of entry into the stormwater drainage network should not be less than 5 minutes (DID, 2002).

The analysis of Scenario 3 showed that it was acceptable to use 5 minutes as the critical storm duration in order to enhance the performance of the green wall system; this was to prolong the serviceability period of the minor urban drainage system and to reduce the volume of storm runoff.

In Scenario 4, the observed rainfall data were used to test and determine the effectiveness of the green wall as an urban drainage system. Appendix B and Appendix C, respectively, show the rainfall-runoff simulations of the existing drainage system without the integration of the green wall for two months, namely January and February 2014, using two sets of actual observed rainfall data for Central City, Kota Samarahan. The simulation of roof runoffs with and without the green wall with respect to the three selected soil types (loamy sand, sandy loam and loam) were carried out and the results were compared and analysed. The performance of the green wall was not solely dependent on rainfall pattern, but also on the interval between storms for particular rainfall events. Therefore, four specific periods of storm duration with various intensity levels were extracted from January and February 2014 and analysed, as shown in Figure 10.

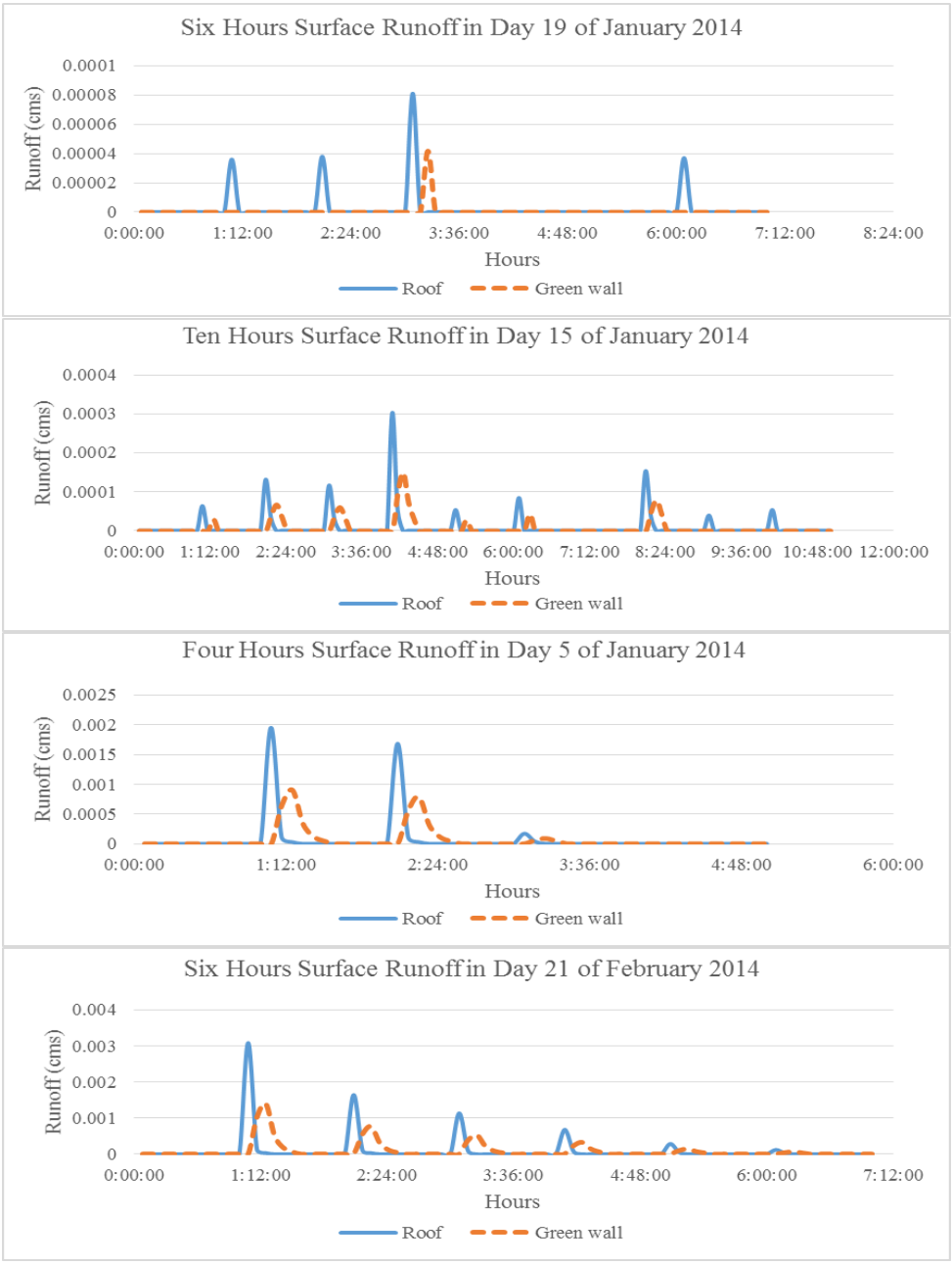


Figure 10. Four analyses of specific periods of storm duration with various intensity levels in January and February 2014

In this project, the effectiveness of a green wall as an urban drainage system was summarised by measuring hydrologic performance, which mainly focusses on deviation of runoff, with integration of the green wall system as shown in Table 5.

Table 5
Summary of hydrologic performance incorporating green wall

Scenario	Date	Continuous Storm Time Span	Intensity (mm/hr)		Average Runoff Reduction (%)
			Range	Average	
4	19 Jan	6 hrs	0.5-3.5	2.0	87
	15 Jan	10 hrs	1.5-9.5	5.5	55
	5 Jan	4 hrs	1.5-51.5	26.5	52
	21 Feb	6 hrs	4.5-80.5	42.5	52
1, 2 & 3	-	Single storm at 5 mins for one year of ARI	-	100.0	55

Scenario 4 showed the average runoff reduction gradually going down from 87% to 52% as the average rainfall intensity rose from 2.0 to 42.5 mm/hr. The increment of storm duration also degraded the performance of the green wall; changes in continuous storm time spanned from 6 to 10 hours, significantly reducing runoff reduction. With the same continuous storm time span of 6 hours, both events on 19 January and 21 February, with the intensity of 2.0 mm/hr and 42.5 mm/hr respectively, showed large gaps in runoff reduction at 35%. Thus, it is evident that when both storm duration and rainfall intensity increase, runoff reduction decreases and so does the effectiveness of the green wall as an urban drainage system facility.

CONCLUSION

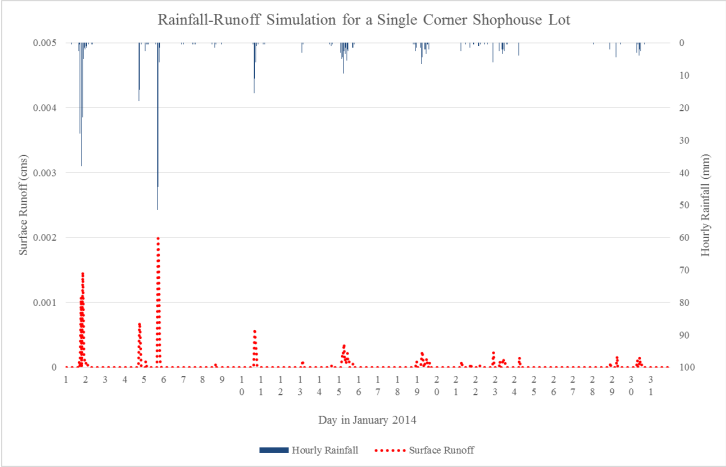
This project demonstrated green wall simulation for a commercial shophouse in Central City, Kota Samarahan. The experiment was carried out using synthesised or actual precipitation data and it tested four scenarios set up on different conditions. The simulation model on the hydrologic condition for the study area was developed to verify the effectiveness of using a green wall as an urban drainage system for reducing surface runoff using USEPA SWMM 5.1. Initially, when using synthesised precipitation data, the runoff decreased by half at 55% on the condition of one-year ARI and 5 minutes of storm duration. The results obtained in Scenario 4 showed repetition of runoff reduction by half after the integration of a green wall using the observed rainfall data.

The green wall model proposed in this study demonstrated the effectiveness of using a green wall as a component of the stormwater management system through four scenarios using SWMM simulation. The results and data displayed can be a guide for future practical design and the building of an actual model.

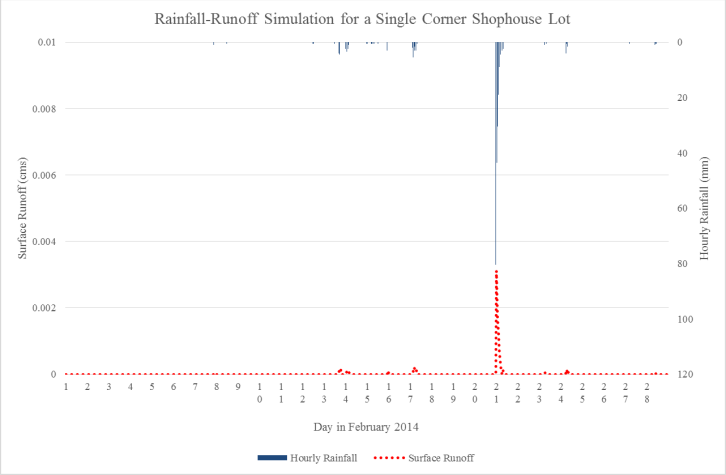
APPENDIX A

Soil Texture Class	Saturated Hydraulic Conductivity (in/hr)	Suction Head (in.)	Porosity (fraction)	Field Capacity (fraction)	Wilting Point (fraction)
Sand	4.74	1.93	0.437	0.062	0.024
Loamy Sand	1.18	2.40	0.437	0.105	0.047
Sandy Loam	0.43	4.33	0.453	0.190	0.085
Loam	0.13	3.50	0.463	0.232	0.116

APPENDIX B



APPENDIX C



ACKNOWLEDGEMENT

The authors thank Universiti Malaysia Sarawak (UNIMAS) and DID Sarawak for assistance and for providing the hydrological data. The authors are also grateful to UNIMAS for financial support through the Special Grant Scheme F02/SpGS/1405/16/6.

REFERENCES

Commonwealth of Massachusetts. (2008). *Smart Growth/Smart Energy Toolkit.* Low Impact Development (LID). Retrieved June 27, 2014, from http://www.mass.gov/envir/smart_growth_toolkit/pages/mod-lid.html

DID. (2002). *Urban stormwater management manual for Malaysia* (Manual Saliran Mesra Alam Malaysia, MSMA) (1st Ed.). Department of Irrigation and Drainage Malaysia. Retrieved June 24, 2014, from http://www.water.gov.my/index.php/pages/view/887

DID. (2012). *Urban stormwater management manual for Malaysia* (Manual Saliran Mesra Alam Malaysia, MSMA) (2nd Ed.). Department of Irrigation and Drainage Malaysia. Retrieved June 24, 2014, from http://www.water.gov.my/index.php/pages/view/887

Elmich Pte Ltd. (2008). *Elmich Green Wall*. Retrieved February 2, 2015, from http://www.elmich.com

Green over Grey. (2009). *Living walls and design.* Retrieved July 7, 2014, from http://www.greenovergrey.com/green-wall-benefits/leed-credits.php

GRHC. (2008). *Introduction to green walls: Technology, benefits and design*. Green Roofs for Healthy Cities. Retrieved July 5, 2014, from http://web.peralta.edu/das/files/2012/03/Green-Walls-Intro-908b_c2.pdf

Higgs, K., Patel, J., Sparks, A., & Young, C. (2011). Biologically engineered & irrigated sustainable structures: A means of addressing urban runoff. *Senior design capstone project report*. Retrieved July 11, 2014, from http://www.natuculture.org/docs/BIOE_502-_Green_Sockwell_Project_Report-_2011.pdf

Huber, W., & Dickinson, R. (1988). *Stormwater management model. version 4, part a. User´s manual.* U.S. Environmental Research Agency. Office of Research and Development. EPA/600/3-88/001a

Kew, B., Echols, S., & Pennypacker, E. (2013). Green Walls as a LID practice for stormwater mitigation: Can green walls provide similar attributes as green roofs? In *6th International Conference of Education, Research and Innovation* (pp. 2598–2607). Seville, Spain.

Loh, S. (2008). Living walls – A way to green the built environment. *BEDP Environment Design Guide, 1*(TEC 26), 1–7.

MD Kwang Tai Realty Sdn Bhd. (2010). *Project development: Central City.* Retrieved November 15, 2014, from http://mdkwangtai.com/

USEPA. (2014). *What is green infrastructure?* United States Environmental Protection Agency. Retrieved July 5, 2014, from http://water.epa.gov/infrastructure/greeninfrastructure/gi_what.cfm

Webb, V. (2010). Green walls: Utilizing & promoting green infrastructure to control stormwater in Mobile, Alabama. *Emerging Issues along Urban/Rural Interfaces III: Linking Science and Society, 256*, 130–136.

Pertanika J. Sci. & Technol. 26 (1): 299 - 308 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Investigating the Randomness and Duration of PM10 Pollution Using Functional Data Analysis

Shaadan, N.[1]*, Deni, S. M.[1] and Jemain, A. A.[2]

[1]Center for Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara, 40450 UiTM, Shah Alam, Selangor, Malaysia
[2]School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

ABSTRACT

Information on situation of air pollution is critically needed as input in four disciplines of research including risk management, risk evaluation, environmental epidemiology, as well as for status and trend analysis. Two normal practices were identified to evaluate daily air pollution situation; first, pollution magnitude has been treated as the common indicator, and second, the analysis was often conducted based on hourly average data. However, the information on the magnitude level alone to represent the pollution condition based on a rigid point data such as the average was seen as insufficient. Thus, to fill the gap, this study was conducted based on continuously measured data in the form of curves, which is also known as functional data, whereby pollution duration is emphasised. A statistical method based on curve ranking was used in the investigation. The application of the method at Klang, Petaling Jaya and Shah Alam air quality monitoring stations located in the Klang Valley, Malaysia, has shown that pollution duration decreases as the magnitude increases. Shah Alam has the longest pollution duration at low and medium magnitude levels. Meanwhile, all the three stations experienced quite a similar length of average pollution duration for the high magnitude level, that is, about 2.5 days. It was also shown that the occurrence of PM10 pollution at the area is significantly not random.

Keywords: Air pollution, functional data analysis, PM10, curve ranking, Malaysia

ARTICLE INFO

Article history:
Received: 03 March 2017
Accepted: 28 September 2017

E-mail addresses:
shahida@tmsk.uitm.edu.my (Shaadan, N.),
sayang@tmsk.uitm.edu.my (Deni, S. M.),
azizj@ukm.edu.my (Jemain, A. A.)
*Corresponding Author

Current Affiliation:
Shaadan, N. and Deni, S. M.
Advanced Analytics Engineering Center,
Faculty of Computer and Mathematical
Sciences, Universiti Teknologi Mara,
40450 UiTM, Shah Alam, Selangor, Malaysia

ISSN: 0128-7680

INTRODUCTION

Having exposed to air pollution has been shown to affect the life quality of human beings, reduce health and decrease life expectancy to a certain extent (Bae, Pan, Kim, Park, Kim, & Kim, 2010; Policheti, Cocco, Spinali, Trimarco, & Nunziata, 2009; Mahiyuddin, Sahani, Aripin, Latif, Thach, & Wong, 2013). Thus, understanding the situation of air pollution including the intensity or the magnitude, as well as their characteristics of occurrences such as their duration and arrival is necessary for preparing and choosing suitable adaptation and mitigation strategies. By any means, the information regarding the situation must be made available. Furthermore, the information is often of a great importance to be used as an input in four disciplines, namely, risk management, risk evaluation, environmental epidemiology, as well as for status and trend analysis (IPCS, 2008).

PM10, a particulate matter with a size of less than 10 micrometer, has been known as one of the important air pollutants other than Ozone (O_3), Nitrogen Oxide (NO_x), Sulphur Dioxide (SO_2) and Carbon Monoxide (CO). In Malaysia, PM10 pollution has become an important environmental problem of concern that needs higher attention as compared to other kinds of environmental pollution in the country; in fact, it has been identified as one of the dominant pollutants in the country other than Ozone (DOE, 2006; Field, Werf & Shen, 2009). The presence of high levels of PM10 in the atmosphere has been reported to significantly associate with haze, which has become a repeated typical problem in the country since 1980s. There have been several haze episodes occurring in Malaysia in the years of 1983, 1987, 1991, 1994, 1997, 2002, 2004, 2005 and 2006 (Afroz, Hassan, & Ibrahim, 2003; DOE, 2006; Field et al., 2009). Although those in 1997 and 2005 were the two extreme years of haze reported in the history of Malaysia, the incidence in 1997 was the most serious as it had caused a huge socioeconomic impact in various sectors including health, production, tourism, transportation and fisheries (Afroz et al., 2003; Abas, Oros, & Simoneit, 2004). For instance, about 83.2% of the population were exposed to health risk during the episode and there was a loss of RM802 million between August and October 1997 due to haze damages (Othman & Shawahid, 1999). The Klang valley region has been reported to be the area most frequently affected by PM10 pollution almost every year, which is regularly caused by transboundary PM10 from Sumatera Island and PM10 emitted from mobile sources, particularly transport, and both are known as the major sources of PM10 in Malaysia (Azmi, Latif, Ismail, Juneng, & Jemain, 2010).

In normal practices, daily air quality data are often recorded discretely by an hourly basis. In the majority of pollution investigation analysis conducted, average data have been popularly used in order to represent the pollutant concentration level for a day period. In a particular perspective, however, the nature of pollutant behaves continuously with time (Gao, 2007; Pudasainee, Sapkota, Shreshta, Kaga, Kondo, & Inoue, 2006). Therefore, it is more ideal to have the continuous evolution of PM10 concentration within the day process taken into consideration in the analysis part. Furthermore, using average data in the analysis is not sufficient enough due the fact of losing information or data reduction as a consequence of averaging or summarising process. To fill the gaps, this paper discusses and proposes the employment of functional data in the study analysis in order to study a situation of PM10 pollution at an area. Specifically, this study aims to investigate the randomness and duration of PM10 pollution. Therefore, to

achieve the study objective, functional data analysis (FDA), which is defined as the statistical methods to analyse functional data or curve data, is used.

METHODOLOGY

Several steps are needed to enable the analysis to be conducted using functional data. The analysis starts with data conversion process, whereby original observed data are transformed into a day curve. Next, the process is followed by the determination of the situation between polluted (abnormal) and normal day based on the concept of curve ranking. Further investigation on the randomness of the pollution occurrences is then conducted.

Data Conversion Process

Converting discrete original recorded data into functional data or curves is the first step that needs to be done. Various approaches can be used either by parametric or non-parametric. In this study, a famous and the most flexible method namely the basis expansion method is chosen (Ramsay & Silverman, 2006). Using the method, the discrete points of PM10 level (y_j) recorded at time point t_j for j=1, ..., 24 of day i are converted into a continuous function $x_i(t)$ using the following formula:

$$x_i(t) = \sum_{k=1}^{K} c_k \phi_k(t), \quad t \in [1,24] \tag{1}$$

The term $\phi(t)$ is the chosen basis system consisting of K number of appropriate basis functions, while c is the corresponding basis coefficient. In this study, K is determined using Bayesian Information Criteria (BIC) and the function's bases are the cubic B-spline. The coefficient c is determined using the least square methods by minimising the error terms based on the following equations:

$$SSE = \sum_{j=1}^{24} (y_j - x(t_j))^2 = \sum_{j=1}^{24} \left(y_j - \sum_{k=1}^{K} c_k \phi_k(t_j) \right)^2 \tag{2}$$

Determining the Status of PM10 Condition Based on Curves Ranking

Consider a set of functional data consisting of n daily PM10 curves, represented by functions $x_i(t)$ for i=1, ..., n. The next step is to determine which day is considered as polluted (abnormal) and which one is considered as normal. PM10 pollution refers to the concentration level in the atmosphere that exceeds a specific allowable level or a threshold level. For this study, a polluted (abnormal) day is defined as any day curve that is higher than a given threshold curve.

The three threshold curves that represent the three different degrees of pollution magnitude are the 50th, 75th and the 90th percentile curves. The 50th percentile curve is used to represent pollution with low magnitude; the 75th percentile is aimed for pollution with medium magnitude level, while the 90th percentile represents pollution with high magnitude. The identification of the percentile curves and determination of the pollution condition of any day is conducted based on curve ranking. In the ranking process, the curves position can be assessed using an

index named SNJF. The index is positive in the value and defined as the ratio of distances and constructed using the following formula:

$$\text{SNJF}(x_i) = \frac{A}{B} \tag{3}$$

Where:

$$A = \frac{d(x_i, x_{\min})}{d(x_i, x_{\min}) + d(x_i, x_{\max})},$$

$$B = \frac{d(x_{\text{median}}, x_{\min})}{d(x_{\text{median}}, x_{\min}) + d(x_{\text{median}}, x_{\max})}$$

$d(x_i, x_{\min})$ is the distance between a curve and the minimum curve
$d(x_i, x_{\max})$ is the distance between a curve and the maximum curve
$d(x_{\text{median}}, x_{\min})$ is the distance between median and minimum curve
$d(x_{\text{median}}, x_{\max})$ is the distance between median and the maximum curve

The distance between the two curves, namely, curve x_1 and x_2, which is obtained using the L_2 norm as in the following equation.

$$\begin{aligned} d(x_1, x_2) &= \|x_1 - x_2\|_2 \\ &= \left(\int_1^{24} |x_1(t) - x_2(t)|^2 \, dt \right)^{1/2} \end{aligned} \tag{4}$$

The minimum curve is defined as $\check{x}_j = \underset{i}{\text{minimum}}\,(x_i(t_j))$ and the maximum curve is $\hat{x}_j = \underset{i}{\text{maximum}}(x_i(t_j))$, while the median curve is obtained based on Fraiman and Muniz (2001). Any day with SNJF exceeding the SNJF of a given threshold curve indicates that the day is considered as a polluted or an abnormal day.

Obtaining the Duration of PM10 Pollution

A coding system is used to obtain a data set that contains the polluted (abnormal) and normal days such that; 0 for normal and 1 for polluted (abnormal) day. In order to investigate the duration of pollution, the data (0,1) are used. Examples of the data are given in Table 1. Based on Table 1, within the 15-day period, there are 2 runs of polluted days. The first run occurred in January 2001 within one-day period and the second run occurred within 2-day period on the 12th and 13th January 2001. In the context of this study, the pollution duration is defined as the number of consecutive days in the state of being polluted or abnormal (i.e., the length of run).

Table 1
Examples of data on the occurrences of polluted and normal day given the SNJF for a threshold curve (the 50th percentile) as 1.037 for a 15-day period

Year	Month	Day	SNJF	Occurrences	
2001	1	1	1.7717	1	} 1st run
2001	1	2	0.5736	0	
2001	1	3	0.7699	0	
2001	1	4	0.5590	0	
2001	1	5	0.7392	0	
2001	1	6	0.9931	0	
2001	1	7	0.5751	0	
2001	1	8	0.5699	0	
2001	1	9	0.6175	0	
2001	1	10	0.5450	0	
2001	1	11	0.9262	0	
2001	1	12	1.3545	1	} 2nd run
2001	1	13	1.1356	1	
2001	1	14	0.9694	0	
2001	1	15	0.8775	0	

Checking the Randomness of Pollution Occurrences

A non-parametric run test known as the Bradley run test was conducted to investigate the behaviour of PM10 pollution occurrences (Bradley, 1968). The aim is to determine whether the occurrences of PM10 pollution were random. Since the analysis involved a large size of data, and thus based on the occurrences and using normal approximation, the run test was conducted to examine the hypothesis whether the distribution of the pollution occurrences was random (null hypothesis) or otherwise (alternative hypothesis) using statistic z such that:

$$z_{calc} = \frac{\left| NR - \overline{NR} \right| - 0.5}{S_R} \tag{5}$$

Where, R is the number of runs, $\overline{NR}$ is the expected number of runs and S_R is the standard deviation of the number of runs, which are given by the following equation:

$$\overline{NR} = \frac{2n_a n_b}{n_a + n_b} + 1 \tag{6}$$

$$S_R = \sqrt{\frac{2n_a n_b (2n_a n_b - n_a - n_b)}{(n_a + n_b)^2 (n_a + n_b - 1)}} \tag{7}$$

where, n_b is the number of normal days and n_a is the number of polluted (abnormal) days. At 5% significance level ($\alpha = 0.05$), for large sample size, the null hypothesis is rejected if the z_{calc} statistic in equation (5) lies outside the interval $[Z_{\alpha/2}, Z_{1-\alpha/2}]$, whereby z_{calc} follows the standard normal distribution with mean 0 and variance 1 (Mendenhall, 1982). Alternatively,

using p-value, which is defined as the probability of $Z < -z_{calc}$ or $Z > z_{calc}$, the null hypothesis is also rejected when the p-value is less than the significance level, α. The run test can only be conducted when there is at least one run for normal and polluted days over the considered period of time such that $n_a, n_b \neq 0$.

APPLICATION OF THE METHOD

The methods explained in the previous section were applied to investigate the PM10 pollution situation at several locations in the Klang valley of Peninsular Malaysia. The analysis was conducted using daily curves data for a period of 10 years (i.e., from 2001 until 2010). The original observed data (daily by hourly discrete data) were obtained from the Department of Environment, Malaysia, involving three air quality monitoring stations, namely, Klang (S1), Petaling Jaya (S2), and Shah Alam (S3). Based on equation (4), the observed data were converted into daily curves using the pre-determined number of basis function K which equals 15, 19 and 17 at each station (S1, S2 and S3) respectively by means of BIC values.

RESULTS OF THE ANALYSIS

For illustration purposes, we can considered a set of five-day PM10 data that were recorded at an hourly basis, as shown in Table 2. The computed SNJF index for each curve is shown in Table 3, while the physical form of the five-day curves is represented in Figure 1. Using the SNJF index, the position of the curves in term of their magnitude can be determined. For this sample of the curves, based on SNJF index, day 4 was found to be the lowest curve, while day 5 is the highest curve.

Table 2
Examples of the 5-day recorded PM10 data

Day	Hour									
	1	2	3	4	5	…	21	22	23	24
1	38	29	23	44	49	…	31	43	37	48
2	51	28	26	29	19	…	29	30	35	44
3	50	51	27	46	39	…	57	44	45	43
4	29	24	16	16	17	…	48	39	38	34
5	19	28	26	30	26	…	59	69	66	80
$\bar{y}$	37.4	32.0	23.6	33.0	30.0	…	44.8	45	44.2	49.8

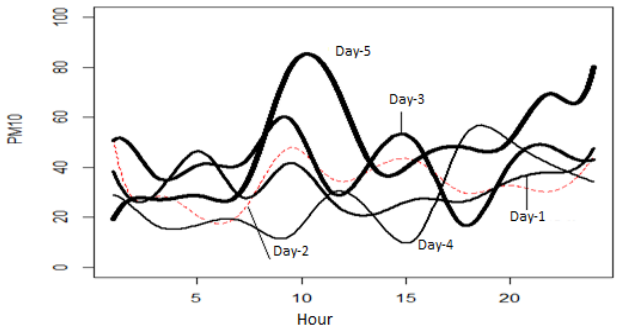


Figure 1. Data set for the five days curve

Table 3
Curve (day) ranking based on SNJF (1-lowest, 5- highest)

Curve (day)	1	2	3	4	5
SNJF Index	1.00	1.11	1.54	0.74	2.09
Position	2	3	4	1	5

Results of the analysis conducted for the data from the three air quality monitoring stations are recorded in Table 4. The results indicate that PM10 pollution duration decreased as the magnitude increased. Among the three stations, with respect to their own local environment, Shah Alam experienced the longest pollution duration at low and medium magnitude levels, while for the high magnitude level, the three stations experienced quite a similar length of pollution duration that was about 2.5 days.

Table 4
The mean duration of PM10 pollution (in days) at three different magnitude levels

Station Code	Name	Threshold Curve		
		P50	P75	P90
		Pollution duration (days)		
S1	Klang	3.92	2.94	2.54
S2	Petaling Jaya	5.20	3.36	2.56
S3	Shah Alam	4.76	3.71	2.47

At the 5% significance level, after the randomness (run) test was conducted, the analysis revealed that the p-value for the test at each level of magnitude; the low (50^{th} percentile), medium (75^{th} percentile) and high (90^{th} percentile) were near to zero (0.00). Therefore, we can conclude that the occurrence of PM10 pollution at the three stations is not random, thus the results suggest that PM10 pollution in the area is not generated by a random process.

CONCLUSION

This study has highlighted a statistical methodology on how to investigate the situation of air pollution using functional data as the input in the study analysis. In contrast to many previous research, functional data have been chosen to incorporate full evolution of the pollutant concentration level within a day period of time. Instead of preventing insufficient information contained in the day data, the application was found to be more appropriate due to the continuous nature of the pollutant process with respect to time.

The identification of polluted (abnormal) and normal day was made based on the simplest approach, that is, via curve ranking procedure. The magnitude level of each PM10 curve (curve) is represented by an index called the SNJF. The three situations of PM10 pollution were considered with respect to three categories of magnitude levels; low, medium and high. SNJF was used to determine the threshold curve (i.e. the 50^{th}, 75^{th} and 90^{th} percentile), as well as indicate whether a day curve is considered as polluted or normal. A day with the SNJF index higher than SNJF index of a threshold curve is detected as being in a polluted (abnormal) situation. Further investigations to study the duration and randomness behaviour of the PM10 pollution occurrences have also been conducted.

The application of the discussed method for the data set at the three air quality monitoring stations located in the Klang Valley revealed that with respect to their respective environment, Shah Alam experienced the longest pollution duration at low and medium magnitude levels, whereas for the high magnitude level, the three stations experienced quite a similar length of pollution duration of 2.5 days. The occurrence of PM10 pollution in the area is significantly not random.

In conclusion, as the application of functional data is becoming popular in many areas including environmental research, this paper would help researchers to conduct a study analysis using curves or functional data.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Environment Malaysia and Prof. Dr. Mohd Talib Latif from Universiti Kebangsaan Malaysia (UKM) for providing the information and data. This work is supported by the UKM's Research University Grant [UKM-AP-2011_19].

REFERENCES

Abas, M. R. B., Oros, D. R., & Simoneit, B. R. T. (2004). Biomass burning as the main source of organic aerosol particulate matter in Malaysia during haze episodes. *Chemosphere, 55*(8), 1089-1095.

Afroz, R., Hassan, M. N., & Ibrahim, N. A. (2003). Review of air pollution and health impacts in Malaysia. *Environmental Research, 92*(2), 71-77.

Azmi, S. Z., Latif, M. T., Ismail, A. S., Juneng, L., & Jemain, A. A. (2010). Trend and status of air quality at three different monitoring stations in Klang Valley, Malaysia. *Air Quality Atmospheric Health, 3*(1), 53-64.

Bae, S., Pan, X., Kim, S., Park, K., Kim, Y., & Kim, H. (2010). Exposures to particulate matter and polycyclic aromatic hydrocarbons and oxidative stress in schoolchildren. *Environmental Health Perspective, 118*(4), 579-583.

Bradley, J. V. (1968). *Distribution-free Statistical Tests*. Englewood Cliffs, New Jersey: Prentice-Hall.

DOE. (2006). Department of Environment. *Malaysia Environmental Quality Report, 2005*. Ministry of Environment and Natural Resources, Putrajaya.

Field, R. D., Werf, G. R. V., & Shen, S. S. P. (2009). Human implication of drought-induced biomass burning in Indonesia since 1960s. *Nature Geoscience, 2*(3), 185-188.

Fraiman, R., & Muniz, G. (2001). Trimmed means for functional data. *Test, 10*(2), 419-440.

Gao, H. O. (2007). Day of week effects on diurnal ozone, NOx cycles and transportation emissions in Southern California. *Transportation Research Part D, 12*(4), 292-305.

IPCS. (2008). *Human Exposure Assessment, Environmental Health Criteria 214 WHO, Geneva, 2000.* International Program on Chemical Safety (IPCS) and INCHEM (Chemical Safety Information from Intergovernmental Organisation).

Mahiyuddin, W. R. W., Sahani, M., Aripin, R., Latif, M. T., Thach, T. Q., & Wong, C. M. (2013). Short-term effects of daily air pollution on mortality. *Atmospheric Environment, 65*, 69-79.

Mendenhall, W. (1982). *Statistics for Management and Economics*. Boston: Duxbury Press.

Othman, J., & Shahwahid, O. H. M. (1999). Cost of Trans-boundary haze externalities. *Jurnal Ekonomi Malaysia, 33*, 3-19.

Policheti, G., Cocco, S., Spinali, A., Trimarco, V., & Nunziata, A. (2009). Effects of particulate matter (PM10, PM2.5 and PM1) on cardiovascular system. *Toxicology, 261*(1), 1-8.

Pudasainee, D., Sapkota, B., Shreshta, M. L., Kaga, A., Kondo, A., & Inoue, Y. (2006). Ground level ozone concentrations and its association with NOx and meteorological parameters in Kathmandu valley Nepal. *Atmospheric Environment, 40*(40), 8081-8087.

Ramsay, J. O., & Silverman, B. W. (2006). *Functional data analysis*. New York: Springer.

Pertanika J. Sci. & Technol. 26 (1): 309 - 316 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Application of Active Contours Driven by Local Gaussian Distribution Fitting Energy to the Computed Tomography Images

Nurwahidah, M.*, Wan, E. Z. W. A. R. and Shaharuddin, C. S.

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia

ABSTRACT

This paper presents the application of active contours region-based method of image segmentation to Computed Tomography (CT) images. Previous researchers applied this region based method on Magnetic Resonance Image (MRI), in vivo images and synthetic images which contain intensity inhomogeneities. In this paper, a different modality known as Computed Tomography (CT) scan was applied. CT scan also produces images containing intensity inhomogeneity, and it is predicted that this method provide good segmentation results. The main objective of applying this method is to check its applicability on CT images. The segmentation process begins by finding the area of interest (black region). Results from this experiment are then used in estimating time of death. Experimental results show that this method has successfully segmented the black region when some parameters changed, provided that the regions are closed to each other. If the black regions are located far from each other, then this method will only segment certain areas.

Keywords: Local Gaussian distribution, computed tomography images, segmentation

INTRODUCTION

Segmentation is one of the basic tasks involved in digital image processing. Digital image processing involves the processes of creating, processing, communicating and displaying images (Gonzalez & Woods, 2008). It can be used to convert signals from image sensor to digital images, increase clarity of images by removing noise and other artefacts, extract the size, scale or the number of objects in a scene, prepare images

ARTICLE INFO

Article history:
Received: 03 March 2017
Accepted: 28 September 2017

E-mail addresses:
nurwahidah5008@gmail.com (Nurwahidah, M.),
waneny@tmsk.uitm.edu.my (Wan, E. Z. W. A. R.),
shahar@tmsk.uitm.edu.my (Shaharuddin, C. S.)
*Corresponding Author

ISSN: 0128-7680

for display or printing, and compress images for communication across a network (Mathwork, 2013). There are many types of images that can be used in segmentation process such as medical images from different modalities, for example, Magnetic Resonance Images (MRI), Computed Tomography Images (CT) and Ultrasound.

In real world images, there are some difficulties in segmenting medical images due to intensity inhomogeneity. Intensity inhomogeneity is pondered to be the multiplicative low-frequency variation of intensities. It is caused by anomalies of the magnetic fields of the scanners (Rajapakse & Kruggel, 1998) such as the presence of intensity inhomogeneity in CT images due to the effects of beam hardening while MRI intensity inhomogeneity arises from radio-frequency coils or acquisition sequences.

This paper focuses on segmenting Computed Tomography (CT) images by using Active Contours driven by local Gaussian distribution fitting energy. The main objective of doing this segmentation is to check the applicability of this method on CT images as there is a future research related to this paper, which is finding the area of segmented region.

This paper is organised in several sections. Section 2 discusses the materials and method used in this paper. This section includes the procedure for CT image segmentation. Meanwhile, experimental result and discussion are presented in Section 3, and finally conclusions are illustrated in Section 4.

Active contour was introduced by Kass, Witkin, and Terzopoulos (1988). Currently, active contours model is widely used in image segmentation. This method produces good results because the model can be achieved by subpixel accuracy. Besides, this model also provides closed and smooth contours on surfaces (Wang, He, Mishra, & Li, 2009).

There are two categories of active contour which are parametric and geometric (Mumford, & Shah, 1989). Parametric active contour model is represented explicitly as parametric curve while geometric active contour model is based on the theory of curve evolution and geometric flows. Basically, active contours can be classified into two classes; edge-based models and region-based models. Edge-based models focus on employing local image gradients to attract contours towards object boundaries, while region-based model focuses on employing global image information in each region such as the distribution of intensities, colours, textures and motions to move the contours towards the boundary. Examples of region-based model are Mumford-Shah functional model (Mumford & Shah, 1989), Chan-Vese Model (Chan, & Vese, 2001), local Binary Fitting (Tsai, Yezzi, & Willsky, 2001) and local Gaussian distribution (Wang et al., 2009).

MATERIALS AND METHODS

Region-based Geometric Active Contour

The Chan-Vese algorithm is an example of the geometric active contour model and it was the first region-based model. This model uses the technique of mean curvature motion. The aim is to evolve the contour in such a way that it stops on the boundaries of the foreground region. Basically in the region based model, it is assumed that there are two regions in each image,

which are, foreground region and background region. This model is known as the Chan-Vese's Piecewise Constant model. It is defined by the energy functional, $F(c_1;c_2;C)$and that energy functional consists of the energy inside and outside the contour. There are also two additional parameters and that energy functional is defined as follows:

$$F(c_1,c_2,C)=\rho_1\int_{C_{in}}|f(x,y)-c_1|^2\,dxdy+\rho_2\int_{Cout}|f(x,y)-c_2|dxdy+\kappa.Length(C)+\mu.Area(C) \quad (1)$$

Where,

ρ_1,ρ_2,κ,μ	:	fixed positive integers
$f(x, y)$	:	pixel in the image
c_{in}	:	the region inside the contour
c_{out}	:	the region outside the contour
c_1	:	the mean intensity of region inside the curve of evolution
c_2	:	the mean intensity of region outside the curve of evolution

The main objective of this region-based GAC is to determine contour, $F(c_1;c_2;C)$ that minimise the energy functional. This energy functional can be minimised by transforming (1) to the level set function.

Level Set Equation

In level set function, there is a curve C known as the contour which divides the image into two regions. If the contour is defined as closed curves, it can be implicitly represented with higher-dimensional function. The first is the inside region denoted by C_{in} and the second is the outside region denoted by C_{out}. The interface of level set function is defined as $\phi(x)=x^2+y^2=1$, where $\phi(x)=0$. The example is as shown in Figure 1. The signed distance function can be defined as curve function, ϕ (Uppu, 2006). In Matlab, an image is defined in matrix form. Suppose, there is point x_1 for i=1,2,...,m and m represents the number of column for the image. This point can be determined whether it is placed inside or outside the contour C by checking the sign of ϕ.

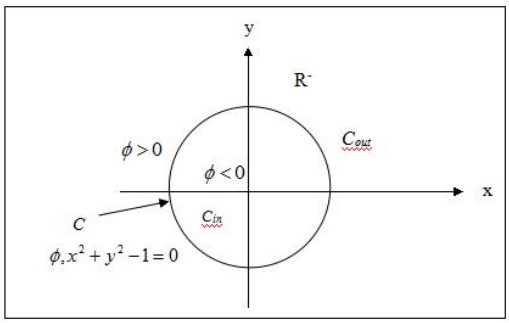


Figure 1. The Initial Contour

Local Gaussian Distribution

Local Gaussian Distribution Fitting (LGDF) is used to model the probability density functions, where the mean and variance are used as varying parameters. In order to effectively exploit information on local intensities, the distribution of local intensities is characterised by the partition of neighbourhood. In this research, local Gaussian distribution is used to obtain local mean and local variance.

In order to obtain the LGDF energy functional for all points in the entire image domain R, the integral of the following equation is minimised to give the following LGDF energy functional:

$$\begin{aligned} E^{LGDF} &= \int_R \sum_{i=1}^{M} \rho_1 \int_{Ri} -W(x-y)\log p_{i,x}(h(y))dy \\ &= \int_R E_x^{LGDF} dx \end{aligned} \tag{2}$$

In LGDF, there are level set functions which are contained inside and outside the contour region. Both the contour regions are named as foreground and background, respectively. Hence, it implies that the image domain where $R = R_1 \cap R_2$ R_1 denotes as the foreground region, while denotes as the background. Thus, (2) can be written as:

$$\begin{aligned} E(u_1(x),u_2(x),\sigma_1^2(x),\sigma_2^2(x),C) = {} & \rho_1 \int_R \int_{R_1} -W(x-y)\log p_{1,x}(h(y))G_\varepsilon(\phi(x,y))dydx \\ & + \rho_2 \int_R \int_{R2} -W(x-y)\log p_{2,x}(h(y))(1-G_\varepsilon(\phi(x,y)))\,dydx \\ & + \kappa \int_R |\vartheta(\phi(y)|\nabla(\phi(y))dxdy \end{aligned} \tag{3}$$

Then, adding the term *Length (C)* to (3) transforms the LGDF energy functional to the level set function. Therefore, the LGDF energy functional becomes:

$$\begin{aligned} E(u_1(x),u_2(x),\sigma_1^2(x),\sigma_2^2(x),C) = {} & \rho_1 \int_R \int_{R_1} -W(x-y)\log p_{1,x}(h(y))G_\varepsilon(\phi(x,y))dydx \\ & + \rho_2 \int_R \int_{R2} -W(x-y)\log p_{2,x}(h(y))(1-G_\varepsilon(\phi(x,y)))\,dydx \\ & + \kappa \int_R |\vartheta(\phi(y)|\nabla(\phi(y))dxdy + \mu \int_R \frac{1}{2}(|\nabla\phi(x)|-1)^2\,dxdy \end{aligned} \tag{4}$$

By differentiating energy functional in (4), with respect to $u_i(x)$ and $\phi_i(x)^2$, while considering $\phi(y)$ as a constant, both mean intensities and variance intensities can be defined as follows:

$$u_i(x) = \frac{W(y-x)f(y)M_{i,\varepsilon}(\phi(y))dy}{W(y-x)M_{i,\varepsilon}(\phi(y))dy} \tag{5}$$

$$\sigma_i(x)^2 = \frac{W(y-x)(u_i(x)-f(y))^2 M_{i,\varepsilon}(\phi(y))dy}{W(y-x)M_{i,\varepsilon}(\phi(y))dy} \tag{6}$$

Procedure for CT Image Segmentation

An image segmentation method, which is local Gaussion distribution fitting energy, was applied to the CT images. The CT images were taken from Hospital Kuala Lumpur for testing purposes. After undergoing the segmentation process, the segmented region was obtained. The aim of this method is to introduce a localised energy functional using a truncated Gaussian Kernel. Wang et al. (2009) used double integral for the energy minimisation. This method also assumes that the mean and variance of the local Gaussian distribution are spatially varying parameters. Generally, the procedure is as follows:

Step 1: View the image in Dicom Viewer.

Step 2: Reduce the image pixels to 256×256 using Adobe Photoshop.

Step 3: Read image in Matlab.

Step 4: Update local means and local variances in (5) and (6).

Step 5: Update the location of the region of interest to locate the initial contour.

Step 6: Repeat step 4 until the criteria is met.

Step 7: Display the result.

The difference between this paper and the previous research is that this paper focused on the black region area. Previously, Wang et al., (2009) focused on the white region and applied on synthetic images, MR images and in vivo images. Our focus is to segment the black region in CT images since the black region can be seen clearly in the CT images.

The method of Wang et al. (2009) has successfully segmented the black region when the region is near to each other and the initial contour is on the region itself. The initial contour needs to be set manually. This means that each image uses different initial contours and the position of the initial is not same. Then, the evolution of the contour changes at each iteration until it reaches the zero level set.

For image iteration, different images require different amounts of iteration. The iteration is set by running the LGDF coding several times with different numbers of iteration. This is done to determine the suitable values which the contour stop evolves. After all the images have been segmented, the results will be used for calculating the area of the black region, which is the next process of this research.

RESULTS AND DISCUSSION

The CT images data originated from Hospital Kuala Lumpur. Dicom Viewer is used to view the images. Then, the CT images are converted into bitmap image file (BMP) and the image pixels are reduced to since LGDF coding cannot read large image pixels. Then, the edited images are read in MATLAB until the results are obtained.

For all the images in this paper, there is a default setting of the parameters which are σ=3, μ=1, timestep Δt=0.1 and v=0.001×255×255. Other parameters will change according to the images. This means that different images will use different parameters.

Table 1
Results of the segmented images of the black region based on local Gaussian Distribution fitting energy

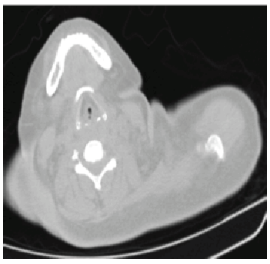
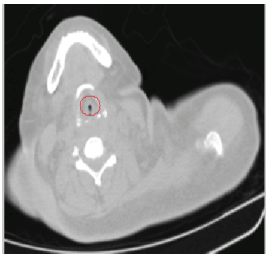
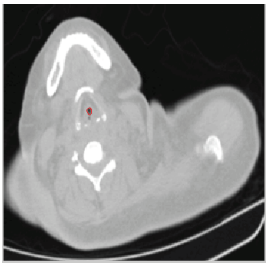
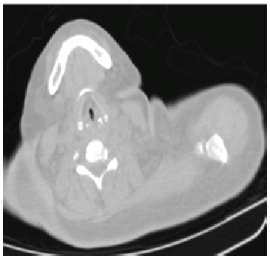
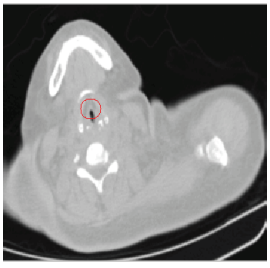
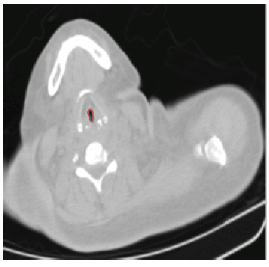
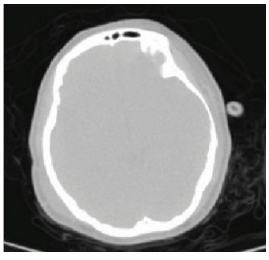
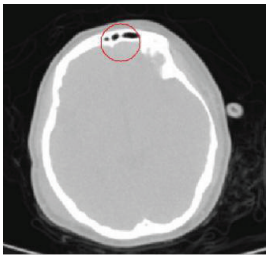
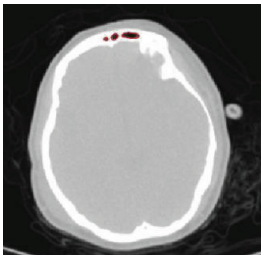
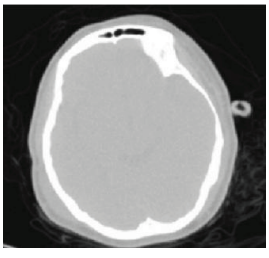
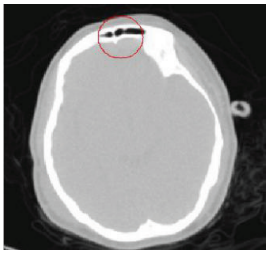
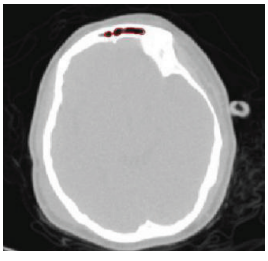
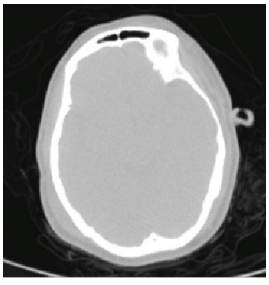
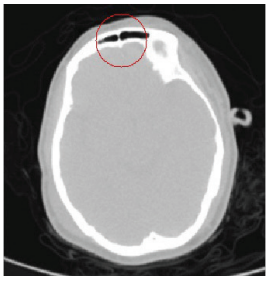
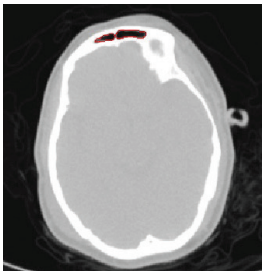
Images	Original images	Initial Contour	Segmented images
Image 1			
Image 2			
Image 3			
Image 4			
Image 5			

Table 1 shows the results of segmented images. The process of segmentation starts with initial contour and will iterate until the black region is fully segmented. The successful results are shown under segmented images and it is clearly seen that this method is able to segment the images when the region of interest is close to each other. The results for iteration and CPU time are recorded in Table 2. It was observed that the iteration took less than 10 seconds for each image to finish the segmentation process. It is a good sign since this method can abbreviate computation time.

Table 2
Number of iteration and CPU time (in s)

Image	Number of Iteration	CPU time
Image 1	40	1.431318
Image 2	40	1.300227
Image 3	100	3.315977
Image 4	140	8.279270
Image 5	140	3.961400

CONCLUSIONS

In this paper, we have presented the application of active contours driven by local Gaussian distribution fitting energy to the CT images. Based on the results given in the previous section, the method proposed by Wang et al. (2009) was found to have successfully segmented the black region provided that the regions are close to each other (as in Figure 2) and make some changes on several parameters. If the images have more than one black region and are located far from each other (as in Figure 3), this method only could segment certain areas of the CT images. Thus, it is a necessity to make an improvement to the model in order to make it applicable with CT images and will definitely give good results.

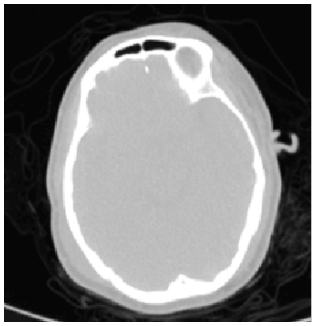


Figure 2. CT images with more than one black region (near to each other)

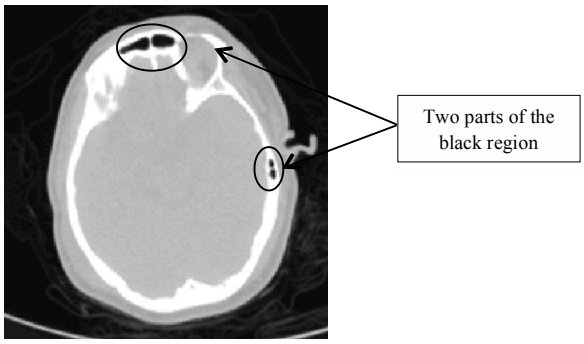


Figure 3. CT images with two black regions (far from each other)

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support received from the Tabung Amanah Pembangunan Akademik Pelajar. The authors also would like to thank Forensic Unit in Hospital Kuala Lumpur for of the permission given to do a research by providing the data. This work also received the ethics approval from Medical Research and Ethics Committee, Ministry of Health (NMMR-15-2001-28698).

REFERENCES

Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing, 10*(2), 266-277.

Gonzalez, R. C., & Woods, R. E. (2008). *Digital Image Processing.* Upper Saddle River, NJ: Pearson Prentice Hall.

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision, 1*(4), 321-331.

Mathwork. (2013). *Digital image processing - matlab and simulink.* Retrieved May 20, 2015, from http://www.mathworks.com/discovery/digital-imageprocessing.html.

Mumford, D., & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics, 42*(5), 577-685.

Rajapakse, J. C., & Kruggel, F. (1998). Segmentation of MR images with intensity inhomogeneities. *Image and Vision Computing, 16*(3), 165-180.

Tsai, A., Yezzi, A., & Willsky, A. S. (2001). Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE transactions on Image Processing, 10*(8), 1169-1186.

Uppu, S. K. (2006). *A Study of Parametric Active Contours for the Application of Mr Axial Abdominal Images.* (Doctoral dissertation). University of Cincinnati, Cincinnati, Ohio, United States.

Wang, L., He, L., Mishra, A., & Li, C. (2009). Active contours driven by local Gaussian distribution fitting energy. *Signal Processing, 89*(12), 2435-2447.

Pertanika J. Sci. & Technol. 26 (1): 317 - 328 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

General Formula for Calculating Accumulated Amount Based on Average Lowest Balance Concept

Wahab, Z. A.[1]*, Embong, R.[1], Azmi, A.[2] and Isa, N. B. M.[1]

[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia*

[2]*Technology Division, Petronas Carigali Sdn. Bhd., 50088 Kuala Lumpur, Malaysia*

ABSTRACT

One of the ways to calculate dividend for an investment is by using average lowest balance (ALB) concept. The existing calculation of dividend based on ALB concept can only be done yearly. This paper discusses on the development of a general formula to calculate the accumulated amount for any period of time, based on the ALB concept that considers different yearly dividend rates. The patterns for each variable and coefficient for the calculated yearly accumulated amount were analysed. The general forms of each variable and coefficient were then combined to form the general formula for calculating the accumulated amount. Validity of the general formula is confirmed by calculating the percentage errors and proven by using mathematical induction.

Keywords: Investment, average lowest balance, pattern analysing, general formula, accumulated amount

INTRODUCTION

In finance, the three most common formulas used to calculate dividend earned are simple interest, compound interest and annuity formulas. In annuity, equal amount of money is invested at equal interval of time over a certain period, where the rate of dividend is assumed to be the same throughout the period of investment. Annuity uses the concept of compound interest. In this paper, the assumption made is almost the same as in annuity, that is equal amount of money is invested at an equal interval of time. However, the yearly dividend rates are assumed to be different and the dividend calculated is based on the ALB concept.

In Malaysia, among the investment plans that use the ALB concept in calculating dividends are Tabung Haji (TH), Amanah

ARTICLE INFO

Article history:
Received: 03 March 2017
Accepted: 28 September 2017

E-mail addresses:
zubaidah@tmsk.uitm.edu.my (Wahab, Z. A.),
erokiah@tmsk.uitm.edu.my (Embong, R.),
afifah.azmi@yahoo.com (Azmi, A.),
norabaizura@tmsk.uitm.edu.my (Isa, N. B. M.)
*Corresponding Author

ISSN: 0128-7680

Saham Bumiputera (ASB) and Skim Simpanan Pendidikan Negara (SSPN-i) (Azlan, 2012; Zaaba, 2010). The current practice of finding the accumulated amount based on ALB concept is by adding up the lowest balance of each month in a year to obtain the average balance for the year. Then, the average balance will be used to find the dividend earned by using the simple interest formula. The accumulated amount obtained for the first year will be carried forward to the next year. The same calculation will be done for the second year, third year and so on. Thus, investment plans based on ALB cannot use annuity formula to find the accumulated amount. As of now, there is no general formula to calculate the accumulated amount over a period of time based on the ALB concept. This paper discusses the development of the general formula for calculating the accumulated amount based on ALB by observing the pattern of yearly dividend and accumulated amount earned.

Many researchers have developed mathematical models to solve problems in various fields such as in genetic study (Diehl & Görg, 2003), growing of cells (Finegood et al., 1995) and describing a real phenomenon (Nei & Li, 1979). While some other researchers from different fields such as music (Jones, 1987; Liu et al., 1999; Conklin, 2002; Meredith et al., 2002; Lu et al., 2004) and image processing (Jamil et al., 2004; Jamil & Bakar, 2006) analysed patterns to solve their problems. There are also some social studies done on investment of Tabung Haji (TH), Amanah Saham Bumiputera (ASB) and Skim Simpanan Pendidikan Negara (SSPN-i) (Zin, 1999; Ishak, 2011; Yusuf, 2011; Yahaya et al., 2009; Haron et al., 2013; Zainal et al., 2009; Hamzah et al., 2011; Musa et al., 2011). In this paper, a mathematical model is developed to solve a problem in the finance field. The method used is developing the model by analysing patterns.

In this study, the general formula is developed based on the existing calculations for calculating yearly accumulated amount. The calculation of dividend earned for the i^{th} year, D_i is as follows:

$$D_i = P_i r_i t_i, \tag{1}$$

where P_i is the average lowest balance of each month in i^{th} year, r_i is the dividend rate for the i_{th} year and t_i is the term (in year). The term P_i can be obtained by using (2):

$$P_i = \frac{\sum_{n=1}^{12} P_n}{12}, \tag{2}$$

where P_n is the monthly lowest balance in a year. After all the values of D_i have been calculated, the accumulated amount for m^{th} year, S_m can then be obtained using (3), as follows.

$$S_m = \sum_{i=1}^{m} D_i + 12Bm, \tag{3}$$

where D_i is the dividend earned until the m^{th} year, B is the monthly savings amount and m is the number of years.

The frequency of depositing money into the account, f, is assumed to be of monthly basis or equivalent to 12. Based on Azlan (2012), Zaaba (2010), and Diehl and Görg (2003), the equations for calculating yearly accumulated amount are as follows:

$$S_1 = B(12 + 6.5r_1) \tag{4}$$

$$S_2 = B(24 + 18.5r_2 + 6.5r_1(1 + r_2)) \tag{5}$$

$$S_3 = B(36 + 30.5r_3 + 18.5r_2(1 + r_3) + 6.5r_1(1 + r_2)(1 + r_3)) \tag{6}$$

$$S_4 = B(48 + 42.5r_4 + 30.5r_3(1 + r_4) + 18.5r_2(1 + r_3)(1 + r_4) + 6.5r_1(1 + r_2)(1 + r_3)(1 + r_4)) \tag{7}$$

Equations (4), (5), (6) and (7) are used as references in order to develop the general formula.

DEVELOPING THE GENERAL FORMULA

The General Form of the Variables

The development of the general formula begins with analysing the pattern of the variables. The variables for the second term until the last term of each equation (4), (5), (6) and (7) are put in a triangular coefficient, as shown in Figure 1.

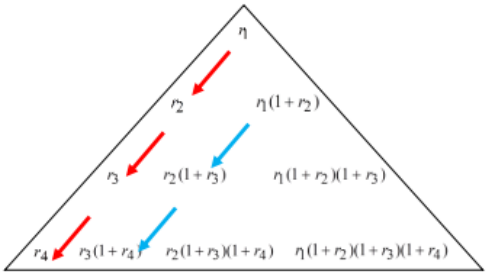


Figure 1. Triangular variable

The patterns of the variable for the second and the third terms are obvious, as shown by the red and blue arrows, respectively. Tables 1 and 2 show a summary of variables for the second and the third terms.

Table 1
A summary of the variables for the second term

Number of Year(s), m	S_m	Variable for Second Term
1	S_1	r_1
2	S_2	r_2
3	S_3	r_3
4	S_4	r_4

Table 2
Summary of variables for the third term

Number of Year(s), m	S_m	Variable for Second Term
2	S_2	$r_1(1+r_2)$
3	S_3	$r_2(1+r_3)$
4	S_4	$r_3(1+r_4)$

From Tables 1 and 2, the general form for the variable of the second term is r_m, where m=1,2,3,... while the general form for the variable of the third term is $r_{m-1}(1+r_m)$ where m=2,3,4,...

Rearranging equation (7) is done to obtain the general formula for the variable of the fourth term and above.

$$S_4 = B(48 + 42.5r_4 + 30.5r_3(1+r_4) + 6.5r_1(1+r_2)(1+r_3)(1+r_4) + 18.5r_2(1+r_3)(1+r_4)) \quad (8)$$

Equation (8) can be visualised as in Figure 2.

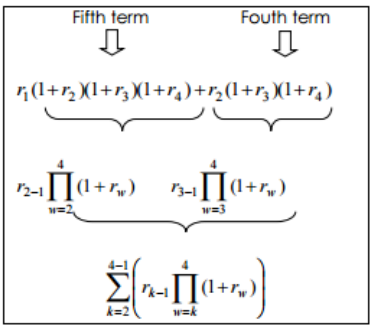


Figure 2. A summation of fourth and fifth terms in S_4

Figure 2 explains the formation of general formula for variables of the fourth term and above. Even though the equation for calculating the accumulated amount has more than five terms, it was found that the pattern of summation was still the same as in Figure 2. Hence, the general formula for the variables of the fourth term and above is;

$$\sum_{k=2}^{m-1}\left(r_{k-1}\prod_{w=k}^{m}(1+r_w)\right) \quad \text{where} \quad m = 3,4,5,...$$

The General Formula for the Coefficients

The coefficients in equations (4), (5), (6) and (7) are arranged and can be visualised as follows:

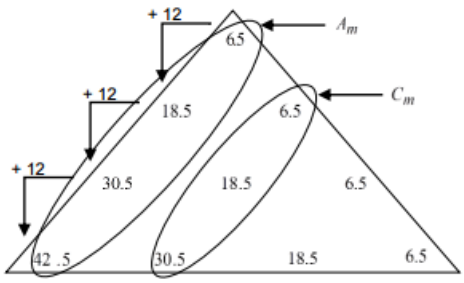


Figure 3. Triangular coefficient

In Figure 3, the value of the coefficient at the vertex is 6.5. This coefficient can be calculated based on the formula of average lowest balance for the first year, which is:

$$P_1 = \frac{\left(a^2 \sum_{y=1}^{f} y_n\right) - 12(a-1)}{12}, \tag{9}$$

where P_i is the average lowest balance for the first year, f is the frequency of depositing money into the account and $a = \frac{12}{f}$. In the example, f is 12 and thus, a=1. Equation (9) is only suitable for f equals to 1, 2, 3, 4, 6 and 12 only. Next, the general formula for A_m is developed (refer to Table 3).

Table 3
Summary of variables for the third term

Number of Year(s), m	S_m	Coefficient for Second Term	General Form of Coefficient
1	S_1	6.5	P_1+12(m-1)
2	S_2	18.5	P_1+12(m-1)
3	S_3	30.5	P_1+12(m-1)
4	S_4	42.5	P_1+12(m-1)

It can be seen that the coefficient for the second term is the addition of P_1 and 12(m-1) . This is shown in Figure 3, whereby the first coefficient of the following row is obtained by adding 12 to the coefficient of the previous row. Hence, the general formula for A_m is:

$$A_m = \frac{\left(a^2 \sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(m-1), \tag{10}$$

As with A_m, the value of P_1 is still maintained, but it must be added with $12(m\text{-}2)$. Hence, the general formula for C_m is as follows:

$$C_m = \frac{\left(a^2 \sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(m-2), \tag{11}$$

Previously, to form the general formula for the variables of the fourth term and above, the summation symbol was used. Since the coefficient of the fourth term and above must be added, the general form of these terms must be placed inside the summation, too. Hence, the general form of the fourth term and above must be written in terms of the variable k and not in terms of variable m. In the general formula of the variables for the fourth term and above, the index of summation started from k=2 until k=m-1. By relating the index k with the coefficient for the fourth and the fifth terms, the general form for the coefficients are summarised in Table 4.

Table 4
A summary of coefficient for fourth and the fifth terms

Sm	Value of k	Coefficient for Fourth Term	Coefficient for Fifth Term	General Form of Coefficient
S_3		6.5	-	P_1+12(k-2)
S_4			6.5	P_1+12(k-2)
		18.5		P_1+12(k-2)

Hence, from Table 4, the general form for the coefficients of the fourth terms and above is P_1+12(k-2) and it is written as follows:

$$D_k = \frac{\left(a^2 \sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(k-2), \tag{12}$$

Combining the General Formulas for Variables and Coefficients

The general formula for the first term, second term and the consecutive terms, as well as the coefficients for the terms can be combined to produce a general formula to calculate the accumulated amount. The general formula is as follows:

$$S_m = \begin{cases} B(12m + A_m r_m) \\ B(12m + A_m r_m + C_m r_{m-1}(1 + r_m)) \\ B\left(\begin{array}{l} 12m + A_m r_m + C_m r_{m-1}(1 + r_m) \\ + \sum_{k=2}^{m-1}\left(D_k r_{k-1} \prod_{w=k}^{m}(1 + r_w)\right) \end{array}\right) \end{cases} \tag{13}$$

Where,

$$A_m = \frac{\left(a^2 \sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(m-1)$$

$$C_m = \frac{\left(a^2 \sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(m-2)$$

$$D_k = \frac{\left(a^2 \sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(k-2)$$

Hence, the general formula for calculating the accumulated amount of money based on average monthly lowest balance that considers different yearly dividend rates has been obtained. Nonetheless, there are a few limitations to this formula. First, a person must practise discipline savings. Discipline saving refers to depositing the same amount of money at regular interval of time. In this study, a person may choose either to deposit money into the account every month, every two months, every three months, every four months, every six months or every twelve months. The frequency of depositing money into the account must be the same throughout the savings period amount. It is also assumed that no withdrawal is made during the savings period.

RESULTS AND DISCUSSION

The accuracy and reliability of the developed general formula is tested using percentage errors and mathematical induction.

Percentage Errors

The accumulated amount was calculated by using the general formulas that have been developed by using MAPLE. The results of calculations from MAPLE were then compared with the yearly calculations done using Microsoft Office Excel by finding the percentage error. The percentage errors were calculated for all the frequencies of depositing money into the account that was being considered in this study. Table 5 shows the percentage errors calculated in which the frequency is 12 (i.e., monthly deposits).

Table 5
The percentage error for monthly deposits

Year	Existing model (RM)	Developed model (RM)	Percentage error (%)
1	626.0000	626.0000	0
2	1306.8350	1306.8350	0
3	2029.2226	2029.2226	0
…			
7	5868.6116	5868.6117	1.70E-06
…			
15	17 526.4867	17 526.4867	0
16	19 420.7194	19 420.7195	5.15E-09
17	21 462.1570	21 462.1570	0
…			
23	37 695.8275	37 695.8275	0

From Table 5, most of the percentage errors are 0%. The largest error is 1.7×10^{-6}%, which is equivalent to 0.0000017% and it is considered to be a very small error that might be due to rounding off. Hence, the developed general formula can be considered as accurate and reliable.

Mathematical induction

According to Solow (2005), mathematical induction must be considered when a statement has the form "for every integer $m \geq 1$, something happens" The "something happens" is some statement $P(m)$, whereby it depends on the integer, m. Solow (2005) also mentioned that there are two words related to induction, which are, "integer" and also "≥1". Based on these statements, this study chose to prove the developed general formula by using the mathematical induction. Generally, there are two steps in induction, which are to verify that $P(1)$ is true and prove that $P(i+1)$ is true, assuming that $P(i)$ is true. Proving of the general formula (14) by using the mathematical induction is explained below:

For every integer $m \geq 3$,

$$S_m = B\left(12m + A_m r_m + C_m r_{m-1}(1+r_m) + \sum_{k=2}^{m-1}\left(D_k r_{k-1}\prod_{w=k}^{m}(1+r_w)\right)\right)$$

where

$$A_m = \frac{\left(a^2\sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(m-1) = P_1 + 12(m-1)$$

$$C_m = \frac{\left(a^2 \sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(m-2) = P_1 + 12(m-2)$$

$$D_k = \frac{\left(a^2 \sum_{y=1}^{f} y\right) - 12(a-1)}{12} + 12(k-2) = P_1 + 12(k-2)$$

The analysis of the proof is discussed below. Let assume that the statement is written as follows:

$$P(m): \quad S_m = B\left(12m + A_m r_m + C_m r_{m-1}(1+r_m) + \sum_{k=2}^{m-1}\left(D_k r_{k-1} \prod_{w=k}^{m}(1+r_w)\right)\right)$$

where

$$A_m = P_1 + 12(m-1)$$

$$C_m = P_1 + 12(m-2)$$

$$D_k = P_1 + 12(k-2)$$

The induction steps are as follows:

Step 1: Verify that $P(3)$ is true.

In this part, monthly deposits case was considered. Hence, f=12 and so a=1. All the values of m are replaced with 3. Therefore, by substituting these values into $P(m)$, $P(3)$ is obtained as follows:

$$P(3): S_3 = B(36 + 30.5r_3 + 18.5r_2(1+r_3) + 6.5r_1(1+r_2)(1+r_3))$$

Hence, the statement is clearly true for m=3.

Step 2: Assume $P(i)$ is true. By substituting m with i, $P(i)$ is as follows:

$$P(i): \quad S_i = B\left(12i + A_i r_i + C_i r_{i\ 1}(1+r_i) + \sum_{k=2}^{i\ 1}\left(D_k r_{k\ 1} \prod_{w=k}^{i}(1+r_w)\right)\right)$$

where

$$A_i = P_1 + 12(i-1)$$

$$C_i = P_1 + 12(i-2)$$

$$D_k = P_1 + 12(k-2)$$

Step 3: Show $P(i+1)$ is true. Then, $P(i+1)$ that needs to be obtained is as follows:

$$P(i+1): \quad S_{i+1} = B\left(12(i+1) + A_{i+1}r_{i+1} + C_{i+1}r_i(1+r_{i+1}) + \sum_{k=2}^{i}\left(D_k r_{k-1}\prod_{w=k}^{i}(1+r_w)\right)\right) \quad (15)$$

where

$$A_{i+1} = P_1 + 12i$$

$$C_{i+1} = P_1 + 12(i-1)$$

$$D_k = P_1 + 12(k-2)$$

In order to show that $P(i+1)$ is true, the left hand side (LHS) of (15) can be written as follows:

$$P(i+1): \quad S_{i+1} = S_i + [(S_i + BP_1)r_{i+1}] + 12B \quad (16)$$

The accumulated amount until the i^{th} year is added with the dividend $i+1^{th}$ of the year and the total monthly savings for the $i+1^{th}$ year. The right hand side (RHS) of (16) is expanded and hence, $P(i+1)$ can be obtained as follows:

$$P(i+1): \quad S_{i+1} = S_i + [(S_i + BP_1)r_{i+1}] + 12B$$

$$= B\left[12i + A_i r_i + C_i r_{i-1}(1+r_i) + \sum_{k=2}^{i-1}\left(D_k r_{k-1}\prod_{w=k}^{i}(1+r_w)\right)\right] + \left[B\left(\begin{matrix}12i + A_i r_i + C_i r_{i-1}(1+r_i) + \\ \sum_{k=2}^{i-1}\left(D_k r_{k-1}\prod_{w=k}^{i}(1+r_w)\right)\end{matrix}\right) + BP_1\right] r_{i+1} + 12B$$

$$= B\left[\begin{matrix}12i + A_i r_i + C_i r_{i-1}(1+r_i) + \sum_{k=2}^{i-1}\left(D_k r_{k-1}\prod_{w=k}^{i}(1+r_w)\right) + 12ir_{i+1} + A_i r_i r_{i+1} + C_i r_{i-1} r_{i+1}(1+r_i) \\ + (r_{i+1})\sum_{k=2}^{i-1}\left(D_k r_{k-1}\prod_{w=k}^{i}(1+r_w)\right) + P_1 r_{i+1} + 12\end{matrix}\right]$$

$$= B\left[12(i+1) + (12i+P_1)r_{i+1} + A_i r_i(1+r_{i+1}) + C_i r_{i-1}(1+r_i)(1+r_{i+1}) + \left(\left(\sum_{k=2}^{i-1}\left(D_k r_{k-1}\prod_{w=k}^{i}(1+r_w)\right)\right)(1+r_{i+1}\right.\right.$$

$$= \left[\begin{matrix}12(i+1) + (12i+\ _1)\ _{i+1} + (\ _1 + 12(i-1))\ _i(1+\ _{i+1}) + (\ _1 + 12(i-2))\ _{i-1}(1+\ _i)(1+\ _{i+1}) \\ + \left(\left(\sum_{k=2}^{i-1}\left(\ \ _k\ _{k-1}\prod_{w=k}^{i}(1+\ _w)\right)\right)(1+\ _{i+1})\right)\end{matrix}\right]$$

$$= B\left[12(i+1) + A_{i+1}r_{i+1} + C_{i+1}r_i(1+r_{i+1}) + D_i r_{i-1}(1+r_i)(1+r_{i+1}) + \sum_{k=2}^{i-1}\left(D_k r_{k-1}\prod_{w=k}^{i}(1+r_w)(1+r_{i+1})\right)\right]$$

$$= B\left[12(i+1) + A_{i+1}r_{i+1} + C_{i+1}r_i(1+r_{i+1}) + \sum_{k=2}^{i}\left(D_k r_{k-1}\prod_{w=k}^{i+1}(1+r_w)\right)\right]$$

Therefore, equation (16) is equal to the right hand side of (15). Thus, the statement is true for i+1 and hence, it is true for all positive integers m. Therefore, the proof is complete.

CONCLUSION

The general formula for calculating the accumulated amount of money based on average monthly lowest balance at any period of time with different yearly dividend rates has been developed. Nonetheless, the general formula can only be used to calculate the accumulated amount of money if a person deposits the same amount of money into the account every month, every two months, every three months, every four months, every six months or every twelve months only.

The formula has been tested by using the MAPLE software. The accumulated amounts obtained from the calculations done using the MAPLE software were then compared with the accumulated amounts calculated using the existing formula. Comparisons were made by looking at the percentage error and the result shows that most of the percentage errors are 0%. Therefore, it can be concluded that the developed general formula gives accurate accumulated amount. Then, the general formula was proven by using mathematical induction. This formula can be used as an alternative general formula for calculating the accumulated amount as it is simple and easy to use. Finally, this is beneficial to the financial institutions that choose to adopt this concept of calculating dividends.

REFERENCES

Azlan, N. H. M. (2012). Calculations of dividend and accumulated amount for savings in Tabung Haji using triangular coefficient diagram. *Final Year Project B.Sc. (Hons) Mathematics*. UiTM Shah Alam, Malaysia.

Conklin, D. (2002). Representation and discovery of vertical patterns in music. In *Music and Artificial Intelligence* (pp. 32-42). Springer Berlin Heidelberg.

Diehl, S., & Görg, C. (2003). Experiencing natural phenomena with virtual, constructed and mathematical models. *Simulation Conference, Proceedings of the 2003 Winter* (pp. 778-781). IEEE.

Finegood, D. T, Scaglia, L., & Bonner-Weir, S. (1995). Dynamics of β-cell mass in the growing rat pancreas: Estimation with a simple mathematical model. *Diabetes, 44*(3), 249-256.

Hamzah, H., Musa, R., Muda, M., & Mohamed, R. N. (2011). Analysing the predicting factors influencing parents to save in National Education Savings Scheme [SSPN]. *Humanities, Science and Engineering (CHUSER), IEEE Colloquium* (pp. 951-956). IEEE.

Haron, S. A, Sharpe, D. L., Abdel-Ghany, M., & Masud, J. (2013). Moving up the savings hierarchy: Examining savings motives of older Malay Muslim. *Journal of Family and Economic Issues, 34*(3), 314-328.

Ishak, M. S. B. H. (2011). Tabung Haji as an Islamic Financial Institution for sustainable economic development. In *2nd International Conference on Humanities, Historical and Social Sciences IPEDR*. IACSIT Press, Singapore.

Jamil, N., & Bakar, Z. A. (2006). Shape-based image retrieval of songket motifs. In *19th Annual Conference of the NACCQ* (pp. 213-219). Wellington, New Zealand.

Jamil, N., Bakar, Z. A., & Sembok, T. M. T. (2004). A comparison of noise removal techniques in songket motif images. *Computer Graphics, Imaging and Visualisation, Proceedings, International Conference* (pp. 139-143).IEEE.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research, *Perception and Psychophysics, 41*(6), 621-634.

Liu, C. C., Hsu, J. L., & Chen, A. L. (1999). Efficient theme and non-trivial repeating pattern discovering music in databases. *Data Engineering, Proceedings, 15th International Conference* (pp. 14-21). IEEE.

Lu, L., Wang, M., & Zhang, H. J. (2004). Repeating pattern discovery and structure analysis from acoustic music data. In *Proceedings of the 6thACM SIGMM International Workshop on Multimedia Information Retrieval* (pp. 275-282). ACM.

Meredith, D., Lemström, K., & Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research, 31*(4), 321-345.

Musa, R., Lennora, P., & Hamzah, H. (2011). *Precautionary saving intention in SSPN by Malaysian parents: Theory of planned behaviour approach*. Research Management Institute of UiTM Shah Alam.

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, 76*(10), 5269-5273.

Solow, D. (2005). *How to read and do proofs* (4th Ed.). Cleveland, OH: John Wiley & Sons Inc.

Yahaya, A., Yahaya, N., Ramli, J., Boon, Y., & Abd Ghaffar, M. (2009). Amanah Saham National Berhad's: Promotional strategies and its relationship with customer motivation. *Journal of Social Sciences, 5*(4), 283-290.

Yusuf, M. B. (2011). Effects of spiritual capital on Muslim economy: The case of Malaysia. *Research on Humanities and Social Sciences, 1*(2), 23-40.

Zaaba, F. S. (2010). The mathematical model for education saving plan in National Education Savings Scheme [SSPN]. *Final Year Project B.Sc. (Hons) Mathematics*. UiTM Shah Alam, Malaysia.

Zainal, N. R., Kamaruddin, R., & Nathan, S. B. S. (2009). Socio-Economic Status and Parental Savings for Higher Education among Malaysian Bumiputera Families. *International Journal of conomics and Finance, 1*(2), 170-173.

Zin, R. H. M. (1999). Malaysian reverse investments: Trends and strategies. *Asia Pacific Journal of Management, 16*(3), 469-496.

Pertanika J. Sci. & Technol. 26 (1): 329 - 340 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Feature Selection Methods: Case of Filter and Wrapper Approaches for Maximising Classification Accuracy

Yap Bee Wah[1]*, Nurain Ibrahim[1], Hamzah Abdul Hamid[1,2], Shuzlina Abdul-Rahman[3] and Simon Fong[4]

[1]Advanced Analytics Engineering Centre, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia
[2]Institute of Engineering Mathematics, Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 UMP, Arau, Perlis, Malaysia
[3]Centre of Information Systems Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia
[4]Faculty of Science and Technology, University of Macau, Avenida da Universidade, Taipa, Macau, China

ABSTRACT

Feature selection has been widely applied in many areas such as classification of spam emails, cancer cells, fraudulent claims, credit risk, text categorisation and DNA microarray analysis. Classification involves building predictive models to predict the target variable based on several input variables (features). This study compares filter and wrapper feature selection methods to maximise the classifier accuracy. The logistic regression was used as a classifier while the performance of the feature selection methods was based on the classification accuracy, Akaike information criteria (AIC), Bayesian information criteria (BIC), Area Under Receiver operator curve (AUC), as well as sensitivity and specificity of the classifier. The simulation study involves generating data for continuous features and one binary dependent variable for different sample sizes. The filter methods used are correlation based feature selection and information gain, while the wrapper methods are sequential forward and sequential backward elimination. The simulation was carried out using R, an open-source programming language. Simulation results showed that the wrapper method (sequential forward selection and sequential backward elimination) methods were better than the filter method in selecting the correct features.

Keywords: Feature selection methods, filter method, logistic regression, simulation, wrapper method

ARTICLE INFO

Article history:
Received: 03 March 2017
Accepted: 28 September 2017

E-mail addresses:
beewah@tmsk.uitm.edu.my (Yap Bee Wah),
nurain9270@salam.uitm.edu.my (Nurain Ibrahim),
amz_bst@yahoo.com (Hamzah Abdul Hamid),
shuzlina@tmsk.uitm.edu.my (Shuzlina Abdul-Rahman),
ccfong@umac.mo (Simon Fong)
*Corresponding Author

INTRODUCTION

Classification is one of the most important tasks in many diverse areas such as business, finance, marketing, engineering, medicine,

ISSN: 0128-7680

bio-informatics, and bio-medical engineering. Classification techniques are used to assign subjects to a specific class of a target variable. In classification problems, predictive models are developed to predict the target variable based on several input variables (features). Features, which are also referred to as attributes, are independent variables. Classification problems, such as classification of cancer tumour, images, handwriting, or spam emails, usually involve many features. Therefore, there is continuing research on finding an efficient method to select relevant features with minimal information loss. Classification of data often contains redundant, irrelevant, useless and misleading features. Hence, feature selection plays an important role in solving classification problems (Jirapech-Umpai & Aitken, 2005; Gheyas & Smith, 2010; Yongjun, Minghao, Kiejung, & Keun, 2012; Zhongyi, Yukun, Tao, & Raymond, 2015).

A complex classification problem involves a large number of features. The classifier will take a longer time to classify the observations when the number of features is very large. Several feature selection methods have been developed to solve classification problems. Feature selection methods deal with dimensionality reduction of a large number of features due to irrelevant and redundant features that may negatively affect the accuracy of classification. The main aim of feature selection is to minimise the dimensionality of the features, maximise the accuracy of classification and prevent overfitting.

Most studies have compared feature selection methods using several datasets. This study compares the selected features using a filter and wrapper methods via a simulation study. The selected filter methods are information gain and correlation based feature selection, while the wrapper methods are sequential forward selection and sequential backward elimination. The simulation procedure was carried out using R-an open source programming language. The feature selection methods were then applied to three datasets obtained from the UCI Machine Learning Laboratory.

In Section 2, we discussed some reviews on feature selection methods. The simulation procedure is given in Section 3. The results are presented in Section 4 and Section 5 concludes the paper.

FEATURE SELECTION METHODS

The feature selection methods have a lot of advantages such as reducing the cost of acquiring data and probably making the classification models much easier to understand (Cantú-Paz, 2004). In general, the feature selection methods can be categorised into the filter, wrapper and embedded methods (Ladha & Deepa, 2011; Naqvi, 2012). The advantages and disadvantages of filter, wrapper and embedded methods have been summarised by Ladha and Deepa (2011), Saeys, Inza, and Larranaga (2007), Bolón-Canedo, Sánchez-Maroño, and Alonso-Betanzos (2013), and Bolón-Canedo, Sánchez-Maroño, Alonso-Betanzos, Benítez, and Herrera (2014). Generally, the filter methods are faster and independent of the classifier. Meanwhile, the wrapper and embedded methods are classifier dependent, which means they interact with the classifier. The wrapper methods are simple methods but there may have a risk of overfitting the model. Some examples of the filter methods are chi-square, information gain, correlation based feature selection and relief. The wrapper methods apply searching techniques such as the sequential forward selection, sequential backward elimination, and plus-1-take-away-r with a classifier.

The embedded methods involve classifier such as decision trees, weighted naive Bayes and weight vector of Support Vector Machine (SVM) in Saeys, Inza, and Larranaga (2007), and SVM-RFE (Guyon, Barnhill, & Vapnik, 2002) and kernel-penalized SVM (Maldonado, Weber, & Basak, 2011). Meanwhile, Hui-Huang, Cheng-Wei, and Ming-Da (2011), and Uğuz (2012) and Naqvi (2012) proposed hybrid feature selection by combining the filter and wrapper methods.

The Filter Methods

The filter methods assess the relevance of features by using a ranking procedure that consequently removes low-scoring features. The filter methods are found to be fast, scalable, computationally simple and independent of the classifier. The methods are divided into two categories: the univariate filter method and multivariate filter method. The univariate methods evaluate the features independently, thereby ignoring feature dependencies and leading to poor feature subsets (Yongjun, Minghao, Kiejung, & Keun, 2012; Yusta, 2009). Unlike the univariate methods that ignore feature dependencies and interaction with the classification algorithm, the multivariate methods consider these two factors to a certain degree (Saeys, Inza, & Larranaga, 2007). The first two methods that will be explained in the subsequent section fall into the mutivariate category, while the last two methods fall into the univariate category.

- *Correlation based Feature Selection* - Correlation based feature selection deals with the features that have redundancy among the features. Correlation based feature selection finds features that are highly correlated with the target variable, but have low inter-correlation between the features by using the correlation coefficient (Yongjun, Minghao, Kiejung, & Keun, 2012; Hall, 1999). For correlation based feature selection, the correlation of each pair of features will be calculated. The highest correlation coefficient value will be the first feature to be selected. The equation of correlation based feature selection is (Huiqing, Jinyan, & Limsoon, 2002):

$$M_S = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}} \quad (1)$$

  where, M_S is the heuristic merit of a feature subset containing k features, $\bar{r}_{cf}$ is the average of the correlation between the features and the target variable, and $\bar{r}_{ff}$ is the average inter-correlation between the features.

- *Fast Correlation-based Filter* - Fast correlation-based filter (FCBF) starts with a full set of features. Fast correlation-based filter uses symmetrical uncertainty to calculate the dependency of features and removes redundant features by using the backward selection method (Zeng, Li, & Chen, 2010). This method has inside stopping criterion to stop it from eliminating the features. Fast correlation-based filter is faster than other feature selection methods. Lei and Huan (2004) provided the algorithm for FCBF method.

- *Information Gain* - A measure based on the information theory of entropy. Entropy is a measure of *disorderliness or noisiness*. Information gain measures the reduction in entropy before and after including the features (Uğuz, 2012; Lei & Huan, 2004). A feature with a high information gain value should be preferred over other features. Information gain does not remove redundant features. The information gain about X provided by Y is calculated as follows:

$$IG(X|Y) = H(X) - H(X|Y) \tag{2}$$

Where,

$$H(X) = -\sum_{i=1}^{k} P(x_i)\log_2 P(x_i) \tag{3}$$

is the entropy of the variable X, and

$$H(X|Y) = -\sum_{i} P(y_j)\sum_{i} P(x_i|y_j)\log_2\left(P(x_i|y_j)\right) \tag{4}$$

is the entropy of X after observing another variable Y. Continuous features need to be discretised when using entropy (Liu, Hussain, Tan, & Dash, 2002).

Each feature will be ranked based on their respective information gain value. Basically, the higher the value, the more informative the feature is.

- *Chi-squared Statistics* – The chi-squared statistics method evaluates association of two categorical variables. Thus, numeric variables need to be discretised into several intervals. The Chi-square statistic is obtained as follows (Huiqing, Jinyan, & Limsoon, 2002):

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\left(\frac{A_{ij} - E_{ij}}{E_{ij}}\right) \tag{5}$$

where m is the number of intervals, k is the number of classes, A_{ij} is the number of samples in the i^{th} interval j^{th} class, R_i is the number of samples in the i^{th} interval, C_j is the number of samples in the j^{th} class, N is the total number of samples, and is the expected frequency of $A_{ij}\left(E_{ij} = R_i * C_j / N\right)$.

Basically, the larger the calculated chi-squared value, the more important the feature is.

The Wrapper Methods

The wrapper methods function almost similar to the filter methods except that they make use of a predefined classification algorithm instead of an independent measure for the subset evaluation. The wrapper methods give a better result compared to the filter methods, but they tend to be more computationally expensive when the number of features becomes very large (Yongjun,

Minghao, Kiejung, & Keun, 2012; Kohavi & John, 1997; Inza, Sierra, Blanco, & Larrañaga, 2002). The first two searching techniques in the wrapper methods described in the subsequent section are the two most common greedy methods frequently employed for feature selection.

- *Sequential Forward Selection* - Sequential forward selection (SFS) starts from the empty set. It performs best when only a small number of features are involved. Nonetheless, the main disadvantage of sequential forward selection is that it is unable to remove features that become insignificant after the addition of other features.

- *Sequential Backward Elimination* - Sequential Backward Elimination (SBE), which is also known as Sequential Backward Selection (SBS), works in the opposite direction of sequential forward selection. Basically, sequential backward elimination starts with a full set of features. Sequential backward elimination works best with a large number of features in the dataset (Ladha & Deepa, 2011).

- *Plus-1-take-away-r* - This method attempts to overcome the nesting effect. In the case of SFS, the nesting effect is a situation whereby once the selected features are selected, they cannot be removed and similar to SBE, once the selected features are removed, they cannot be re-selected. This method allows SFS to use l times forward and then r back-tracking steps of SBS. The challenge with the "plus-l-take-away-r" method is predicting the best (l, r) values to obtain good results with moderate computation (Unler & Murat, 2010).

- *Sequential Floating Forward and Backward Selection (SFFS and SFBS)* - These two methods were introduced by Pudil, Novovičová and Kittler (1994). Backtracking is controlled without any parameter setting. These methods allow a more flexible method since the number of forward and backtracking steps is not predetermined, but instead, it is dynamically changed (Shuzlina, 2012). The SFFS and SFBS are probably the most effective FS methods (as cited in Yusta, 2009). These floating methods allow dynamic addition and deletion of the feature subsets until a suitable number of feature subsets are obtained. The benefit offered by the floating method over the plus-l-take-away-r is its ability to sweep through feature subsets to obtain good results.

SIMULATION PROCEDURES

In this simulation study, data X were simulated and assigned to group 1 or 0 using the following logistic regression model:

$$P(Y=1)=\frac{1}{1+e^{-z}} \tag{6}$$

where $z=\beta_0+\beta_1X_1+\beta_2X_2+\cdots\beta_kX_k$ and k is the number of features. For k=10, we set 6 significant features, while for k=50 we set 20 significant features. The predictors (or features) were set as significant features with odds-ratio greater than 1. For example, the odds-ratios for X5 and X8 are significant features with exp(0.4)=1.087 and exp(0.5)=1.359, respectively.

An odds-ratio close to 1 indicates the feature is not significantly related to Y. Ddata for the logistic regression model were generated using the same technique used by Hosmer and Hjort (2002). The simulation procedure for k=10 is as follows:

1. Generate continuous random features from X1, X2 … to X10 from a standard normal distribution.
2. Calculate z = (0.7 + 0.0000001*x1 + 0.0000001*x2 + 0.0000001*x3 + 0.0000001*x4 + 0.4*x5 + 0.4*x6 + 0.4*x7 + 0.5*x8 + 0.5*x9 + 0.5*x10) and $\pi(x)=\frac{1}{1+e^{-z}}$.
3. Generate the data u from a uniform distribution, U(0,1).
4. Generate outcomes for binary logistic regression by using the rule y=1 if $\mu \leq \pi(x)$ and y=0 otherwise.
5. Apply the feature selection method FSM(j).
6. Count the number of correctly selected features. Repeat 1-6, 1000 times, and obtain an average percentage of the correct features selected.

RESULTS

Results obtained from the simulation study and real datasets are discussed in this section.

Simulation Results

Based on the simulation results in Table 1 and Table 2, the percentages of correctly selected features increase as the sample size increases. The wrapper methods perform better than the filter methods when the model contains 10 features. Meanwhile, the information gain method did not perform well compared to the correlation-based and wrapper methods.

Table 1
Percentage of correctly selected features (k=10, 6 significant features

Feature Selection Methods	Sample Size				
	100	200	300	500	1000
Correlation-based feature selection	74.4%	84.7%	90.7%	96.2%	99.4%
Information Gain	33.3%	33.3%	33.3%	33.3%	35.0%
Sequential Forward Selection	80.3%	89.8%	94.9%	99.2%	100%
Sequential Backward Elimination	80.5%	89.8%	94.9%	99.2%	100%

Table 2
Percentage of correctly selected features (k=50, 20 significant features)

Feature Selection Methods	Sample Size				
	100	200	300	500	1000
Correlation-based feature selection	63.2%	80.5%	88.4%	94.7%	98.7%
Information Gain	0.4%	0.3%	0.4%	0.7%	4.2%
Sequential Forward Selection	58.5%	74.9%	84.4%	93.2%	99.2%
Sequential Backward Elimination	57.0%	75.0%	84.4%	93.2%	99.2%

Application to real datasets

Next, the feature selection methods were applied to three real datasets. The datasets used were Pima Indians Diabetes, Breast Cancer Wisconsin and Spambase obtained from UCI Machine Learning Repository. The sample size for the Pima Indians Diabetes is 768 with eight continuous features. Outcome variable is a binary variable which is denoted as 1 if a patient is tested positive and 0 if it is negative for diabetes. Table 3 summarises the results for the Pima Indians Diabetes dataset.

Table 3
Pima Indians diabetes dataset results

Method/ Performance	AIC	BIC	AUC	ACC	SEN	SPEC
No feature selection (x1*,x2*x3*,x4, x5,x6*, x7*,x8)	741.45	783.24	0.8394	78.2%	58%	89%
Correlation-based feature selection (3 features: x1*, x2*, x6*)	752.12	770.70	0.826	76.69%	57%	87%
Information Gain (2 features: x2*, x6*)	777.4	791.33	0.8109	76.43%	53%	89%
Sequential Forward Selection (6 features: x2*,x6*,x1*,x7*, x3*,x8)	739.46	771.97	0.8348	77.34%	58%	88%
Sequential Backward Elimination (6 features: x1*,x2*,x3*,x6*, x7*, x8)	739.46	771.97	0.8384	77.34%	58%	88%

*Significant feature
Source: https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

Based on the results in Table 3, the filter methods selected have fewer significant features compared to the wrapper methods. The correlation based feature selection selected three features (x1, x2, x6), while information gain selected only two features (x2, x6). Meanwhile, both the sequential forward selection and sequential backward elimination selected six significant features. The logistic regression has higher accuracy, sensitivity and specificity with the eight features selected by the wrapper methods.

The sample size for the Breast Cancer Wisconsin Dataset is 669 with nine continuous features. The outcome variable is a binary variable, which is denoted as 1 if the tumour is malignant and 0 if benign. Results for the Breast Cancer Wisconsin Dataset are shown in Table 4. The correlation-based feature selection only selects one variable (x9), while the information gain selected two features (x2 and x7). Meanwhile, both the sequential forward selection and sequential backward elimination selected the first eight features. These results show that wrapper methods, using the sequential forward selection and sequential backward elimination, managed to select more significant features compared to the filter methods. The application to two real datasets confirms the simulation result indicating that the wrapper methods are better than the filter selection methods. The logistic regression has lower accuracy, sensitivity and specificity, with the single feature selected by the correlation based method.

Table 4
Breast cancer Wisconsin dataset results

Method/ Performance	AIC	BIC	AUC	ACC	SEN	SPEC
No feature selection						
(x1*,x2,x3,x4*,x5,x6, x7*, x8, x9)	132.18	177.64	0.9959	97.0%	98.0%	95.0%
Correlation-based feature selection						
(x9*)	735.08	744.19	0.7101	79.0%	97.0%	44.0%
Information Gain						
(x7*, x2*)	232.67	246.32	0.984	93.5%	97.0%	88.0%
Sequential Forward Selection						
(x3,x6*,x1*,x8,x7*,x2, x5,x4*)	134.28	175.23	0.9955	97.0%	98.0%	95.0%
Sequential Backward Elimination						
(x1*,x2,x3, x4*,x5, x6*,x7*,x8)	134.28	175.23	0.9955	97.0%	98.0%	95.0%

*Significant feature
(*Source:* https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes)

The sample size for the Spambase dataset is 4601 with 57 continuous features. Results for the Spambase dataset are shown in Table 5. The outcome is a binary variable which denotes whether the email is considered a spam email (1), or otherwise (0). The result shows that the information gain method selects only two significant features (x52 and x53). Meanwhile, the correlation based method selected sixteen significant features. The sequential backward elimination selected 44 features (3 were not significant, x11, x18 and x22), while the sequential backward elimination selected only ten significant features. The logistic regression performs best with the 44 features selected by the sequential backward elimination with highest accuracy, sensitivity and specificity, lowest AIC and BIC values. This results of this study support the findings by Inza, Larrañaga, Blanco, and Cerrolaza (2002), whereby applications show that the wrapper methods perform better than the filter methods in gene selection in DNA microarray domains.

Table 5
Spambase dataset results

Method/ Performance	AIC	BIC	AUC	ACC	SEN	SPEC
No feature selection						
(x1*, x2*, x5*, x6*, x7*, x8*, x9*, x10*, x12*, x15*, x16*, x17*, x19*, x20*, x21*, x23*, x24*, x25*, x26*, x27*, x28*, x29*, x33*, x35*, x36*, x39*, x41*, x42*, x44*, x45*, x4x6*, x48*, x49*, x52*, x53*, x54*, x5x6*, x57*)	1931.8	2304.94	0.98	93.0%	89.0%	96.0%
Correlation-based feature selection						
(x3*, x5*, x6*, x7*, x8*, x16*, x17*, x18*, x19*, x20*, x21*, x22*, x23*, x24*, x52*, x53*, x57*)	2893.7	3009.54	0.94	89.0%	79.0%	95.0%
Information Gain						
(x52*, x53*)	4472.9	4492.21	0.88	82.0%	61.0%	95.0%
Sequential Forward Selection						
(x51*, x36*, x31*, x15*, x13*, x41*, x29*, x14*, x28*, x32*)	5035.3	5106.1	0.8	73.0%	46.0%	91.0%
Sequential Backward Elimination						
(x1 *, x2* , x3 ,x4 ,x5* ,x6*, x7*, x8*, x9*, x10* , x11 , x12* , x16* , x17* , x18 , x19* , x20* , x21* , x22 , x23* , x24* , x25* , x26* , x27* , x30 , x33* , x34 , x35* , x37 , x38 , x39* , x40 , x42* , x43 , x44* , x45 , x46* , x47 , x48* ,x49* , x50 , x52* , x53* , x54* , x56* , x57*)	1969.8	2272.19	0.98	93.0%	89.0%	96.0%

CONCLUSION

Feature selection methods depend on types of features. This simulation study with continuous features shows that the wrapper method selected more significant features compared to the filter methods. Nonetheless, the information gain did not perform well for continuous features. The application to three datasets confirms that the wrapper method using sequential backward elimination is the best selection method for data with continuous features. The feature selection methods can be easily applied using R, an open source programming language. The simulation study is being extended to compare the performance of the filter and wrapper methods for categorical type features. Future research can also look into feature selection using random forest (Genuer, Poggi, & Tuleau-Malot, 2010), the multivariate-based feature filter method called the kernel PLS-based filter method (Sun, Peng, & Shakoor, 2014) or the hybrid methods Hsu, Hsieh, & Lu, 2011; Uğuz, 2012; Naqvi, 2012; Shilaskar & Ghatol, 2013).

The R-syntax for the simulation study and application of feature selection using R can be obtained from the corresponding author.

REFERENCES

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems, 34*(3), 483-519.

Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benítez, J. M., & Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Information Sciences, 282,* 111-135.

Cantu-Paz, E. (2004). Feature subset selection, class separability, and genetic algorithms. In *Genetic and evolutionary computation conference* (pp. 959-970). Springer Berlin Heidelberg.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using Random Forests. *Pattern Recognition Letters, 31*(14), 2225-2236.

Gheyas, I. A., & Smith, L. S. (2010). Feature subset selection in large dimensionality domains. *Pattern Recognition, 43*(1), 5–13.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*(1-3), 389-422.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. (Doctoral Dissertation). The University of Waikato, New Zealand.

Hosmer, D. W., & Hjort, N. L. (2002). Goodness-of-fit processes for logistic regression: simulation results. *Statistics in Medicine, 21*(18), 2723-2738.

Hsu, H. H., Hsieh, C. W., & Lu, M. D. (2011). Hybrid feature selection by combining filters and wrappers. *Expert Systems with Applications, 38*(7), 8144-8150.

Huiqing, L., Jinyan, L., & Limsoon, W. (2002). A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. *Genome Informatics, 13*, 51-60.

Inza, I., Larrañaga, P., Blanco, R., & Cerrolaza, A. J. (2002). Filter versus wrapper gene selection approaches in DNA microarray domains. *Artificial Intelligence in Medicine, 31*(2), 91-103.

Inza, I., Sierra, B., Blanco, R., & Larrañaga, P. (2002). Gene selection by sequential search wrapper approaches in microarray cancer class prediction. *Journal of Intelligent and Fuzzy Systems, 12*(1), 25-33.

Jirapech-Umpai, T., & Aitken, S. (2005). Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes. *BMC Bioinformatics, 6*(1), 148-158.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*(1-2), 273-324.

Ladha, L., & Deepa, T. (2011). Feature Selection Methods and Algorithms. *International Journal on Computer Science and Engineering (IJCSE), 3*(5), 1787 – 1797.

Lei, Y., & Huan, L. (2004). Efficient Feature Selection via Analysis of Relevancy and Redundancy. *Journal of Machine Learning Research, 5*(Oct), 1205-1224.

Liu, H., Hussain, F., Tan, C. L., & Dash, M. (2002). Discretization: An enabling technique. *Data Mining and Knowledge Discovery, 6*(4), 393-423.

Maldonado, S., Weber, R., & Basak, J. (2011). Simultaneous feature selection and classification using kernel-penalized support vector machines. *Information Sciences, 181*(1), 115-128.

Naqvi, G. (2012). *A Hybrid Filter-Wrapper Approach for Feature Selection.* (Master's Thesis). Orebro University, Sweden.

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters, 15*(11), 1119-1125.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*(19), 2507-2517.

Shilaskar, S., & Ghatol, A. (2013). Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. *Expert Systems with Applications, 40*(10), 4146-4153.

Shuzlina, A. R. (2012). *Multivariate filter with particle swarm optimisation variants for feature selection.* (Doctoral dissertation). Universiti Kebangsaan Malaysia, Malaysia.

Sun, S., Peng, Q., & Shakoor, A. (2014). A kernel-based multivariate feature selection method for microarray data classification. *PloS ONE, 9*(7), e102541.

Uğuz, H. (2012). A hybrid system based on information gain and principal component analysis for the classification of transcranial Doppler signals. *Computer Methods and Programs in Biomedicine, 107*(3), 598-609.

Unler, A., & Murat, A. (2010). A discrete particle swarm optimization method for feature selection in binary classification problems. *European Journal of Operational Research, 206*(3), 528-539.

Yongjun, P., Minghao, P., Kiejung, P., & Keun, H. Y. (2012). An ensemble correlation-based gene selection algorithm for cancer classification with gene expression data. *Bioinformatics, 28*(24), 3306-3315.

Yusta, S. C. (2009). Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters, 30*(5), 525-534.

Zeng, X. Q., Li, G. Z., & Chen, S. F. (2010, December). Gene selection by using an improved Fast Correlation-Based Filter. In *Bioinformatics and Biomedicine Workshops (BIBMW), 2010 IEEE International Conference on* (pp. 625-630). IEEE.

Zhongyi, H., Yukun, B., Tao, X., & Raymond, C. (2015). Hybrid filter–wrapper feature selection for short-term load forecasting. *Engineering Applications of Artificial Intelligence, 40*, 17–27.

Pertanika J. Sci. & Technol. 26 (1): 341 - 352 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Neurocomputing Approach for Firearm Identification

Nor Azura Md Ghani[1]*, Choong-Yeun Liong[2] and Abdul Aziz Jemain[2]

[1]Center for Statistical and Decision Sciences Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia
[2]School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

ABSTRACT

This paper is an attempt to perceive and order guns using a two-layer neural system model taking into account a feedforward backpropagation calculation. Numerical properties from the joined pictures were utilised for enhanced gun characterisation execution. Inputs of the system model were 747 pictures blackmailed from the discharging pin impression of five differing guns model, Parabellum Vector SPI 9mm. Components created from the dataset were further grouped into preparation set (523 components), testing set (112 components) and acceptance set (112 components). Under managed learning, exact results exhibited that a two-layer BPNN of 11-11-5 arrangement, with tansig/purelin exchange capacities and a "trainlm" preparing calculation, had productively delivered 87% right aftereffect of grouping. The order result serves to be progressed and contrasted with the previous works. Finally, the joined picture districts can offer some accommodating data on the grouping of gun.

Keywords: Firearm classification, combined images, geometric moments, backpropagation neural network

INTRODUCTION

Acts, discernibled as 'wrongdoing', have been troubling humanity for quite a while. Wrongdoing insights, including handguns and different guns, have been consistently expanding subsequent to the nineteenth century, i.e. after the great advancement had been seen in weapon innovation and in the huge creation of handguns everywhere throughout the world. In this way, the measures received to substantiate and test into the utilisation of gun in wrongdoings as court proof, or criminological ballistics as its

ARTICLE INFO

Article history:
Received: 03 March 2017
Accepted: 28 September 2017

E-mail addresses:
azura@tmsk.uitm.edu.my,
azura158@salam.uitm.edu.my (Nor Azura Md Ghani),
lg@ukm.edu.my (Choong-Yeun Liong),
azizj@ukm.my (Abdul Aziz Jemain)
*Corresponding Author

ISSN: 0128-7680

normal term, are crucial to the measure of moderating the wrongdoing rates. In the legal unit of the Royal Malaysian Police, legal ballistics indicates an extensive investigation of the brunt of shots and the shot qualities where noteworthy confirmation found can issue authoritative verdicts. To include, weaponry, ammo and imprint printing investigation are included in the instrument imprint and gun assessments to confirm the utilisation of specific gun or barrier apparatus in the occasions of wrongdoing. These imprints and physical signs of the qualities on the cartridge packaging head have a considerable part to play as they can indicate to us the part where a shot is activated from (Widrow, Rumelhart, & Lehr, 1994; Smith, Cross & Variyan, 1995). Exceptional imprints will be unmistakably seen on the housings and the projectile undertaking after dispatching a slug. Through and through, more than thirty distinct qualities of these imprints can deliver traceable ballistics fingerprinting which takes after typical human fingerprints (Burrard, 1995). As gun markings, the sort, model, and characteristics of every weapon work well with each other. Along these lines, this legal strategy is major to legitimate confirmation for various gun-related violations to be lessened.

There have been a few programmed ballistics recognisable proof frameworks developed such as IBIS, CONDOR, ALIAS, FIREBALL and EVOFINDER, which encourage specialists in drawing so as to make associations of wrongdoing cases a correlation of each trademark on the projectiles and the cartridge case pictures that have been identified in the present wrongdoing to the prior confirmations recorded in the database (Smith et al., 1995, Smith & Li, 2008; Geradts, Bijhold, Hermsen, & Murtagh, 2001). Normally, such errand experience time limitation since it must be contrasted and an effective measure of accessible confirmations in the records beforehand settled, furthermore a considerable amount of guns to be coordinated. An issue frequently happens when the agents judge wrongly through their naked eyes, a rate regularly termed as 'human mistakes'. It is additionally challenging to these examiners as manual investigation requires fastidious arranging and undisputable ability.

Along these lines, the guns inventoring utilising numerical attributes of the joined shooting pin impression picture in the feedforward backpropagation neural system (BPNN) are the highlight of this study because of the proficiency and adequacy of the counterfeit neural systems (ANNs) in grouping and making arrangements (Widrow et al., 1994). This has been made apparent through the ANNs' versatility to different trains and tries. The present study is an attempt to identify the gun by taking into account the guided elements of consolidated discharging pin impression pictures using the two-layer feed forward BPNN.

In Section 2, past inquires about that have stayed into gun distinguishing proof have been re-investigated, and segment 3 tests into the information foundation. A succinct portrayal of the neural system approach executed is given in segment 4, while the outcomes, together with compact exchanges on the neural system models that have been adjusted and the different backpropagation preparing calculations, are elaborated in segment 5. Finally, the general finish of the discoveries and some achievable proposals for further work are offered in segment 6.

PAST RESEARCH

Bonfanti and Kinder (1999) called attention to that the begining of gun fingerprints is perceivable from the time the projectile is dispatched until it hits the objective, where each specific imprint is discovered either at the slug, cartridge packaging or discharging barrel. At that point, it is proposed by Xin et al. (2000) that a recognisable proof technique connected in gun official acknowledgment depends on magazine cases situated at the focal point of the shooting device of the cartridge cases. An intelligent framework and some proper methodologies that enable a prompt identification of the ID rates have been developed. Geradts, Bijhold, Hermsen, and Murtagh (1999, 2001) explained that the terminating pin is intense as it comprises of individual subtle elements for the cases portrayal. Meanwhile, Ghani, Liong, and Jemain (2009a, 2009b and 2010) stated that the vast majority of numerical components extricated from the terminating pin impression picture arrangement are runs of the mill highlights in gun ID. Some momentous essential measurements have effectively been developed to serve as the similar pointers in scientific ballistics examples. Leng and Huang (2012) presented another component termed as circle moment invariants for a more refined and discernable gun characterisation using some circle-brought together pictures.

The neural system used to distinguish guns was initially presented in 1994, where Kou, Tung, and Fu (1994) embraced the Self Organising Feature Map (SOFM) method in gun capabilities. Nonetheless, their work merely re-affirmed what past exploration had made on the subject, and effective occasions have yet to be accounted for. Geradts et al. (1999, 2001) have added to some picture coordinating polar calculations for gun recognisable proof using the shooting pin impression pictures and rupture face marks. Their work sets up that expel impact can be thought about all the more effectively utilising relationship examination. To acquire the most suitable lighting conditions required to elucidate the subject, diminish the blunder and streamline the yield, a handling step was initially forced through the histogram evening out strategy, where it was appeared to give a superior yield with a superior regenerative capacity. Xin, Zhou, and Rong (2000) utilised a combination of technique as a part of an intelligent way to achieve better results using cartridge cases with a groundwork situated at the focal point of the cases. Eventually, fruitful components were created as the work lands at its decision. Later, Kong, Li, and Watson (2003) and Li (2006) utilised neural system that weights on the SOFM neural system model, where the system is coordinated with a choice making procedure to group guns. Their findings indicated that the hybridised SOFM model has solid execution and quality. With reference to the past works (Kamaruddin, Ghani, Liong & Jemain, 2011), numerical elements of the whole terminating pin impression pictures were utilised to make order of guns. On the same note, ANNs were used to characterise execution by taking into account the cross-validation results.

DATA BACKGROUND

The numerical components used are the optional information set forth by Ghani et al. (2010). Illustrations of a cartridge packaging head, the whole terminating pin impression, full ring

picture of the terminating pin impression, and middle terminating pin impression are given in Figures 1(a), (b), (c) and (d), respectively. Initially, there were 6 best components of the entire terminating pin impression image, 8 best elements of the inside terminating pin impression picture, and 6 best elements of the ring terminating pin impression picture (Ghani et al., 2009a, 2009b and 2010). For the purpose of this study, however, focus was given on the eleven consolidated elements officially recognised in past studies. The specimen size of this study contained 747 numerical component vectors, and this implied that we have adequate example components serving as inputs in the neural system strategy, as proposed by Masters (1993), who stated that the minimal number of inputs in a neural system must be more than hundreds. The required variables are diverse sorts of guns utilised, marked as A, B, C, D and E. The autonomous variables are the best eleven geometric moment components of the consolidated picture. The neural system technique was executed using MATLAB R2010a.

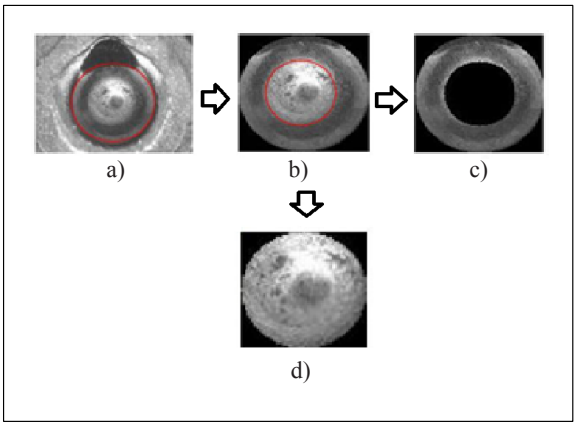


Figure 1. Extracted regions from the cartridge case image

The portrayals and documentations of the eleven components of the picture as the free variables are given in Table 1. The elements under centre were inclining toward the proposals built up in the past studies. For further elaboration on the components, consistent reference is made to Ghani et al. (2009a, 2009b and 2010).

Table 1
The eleven best features of the combined firing pin impression images

No.	Description		Notation
1.	M_{03}	whole image	MP_{03}
2.	M_{12}	whole image	MP_{12}
3.	M_{20}	whole image	MP_{20}
4.	M_{01}	center image	MT_{01}
5.	M_{02}	center image	MT_{02}
6.	M_{10}	center image	MT_{10}
7.	M_{11}	center image	MT_{11}
8.	M_{12}	center image	MT_{12}
9.	M_{21}	center image	MT_{21}
10.	M_{10}	ring image	MC_{10}
11.	M_{11}	ring image	MC_{11}

NEURAL NETWORK APPROACH

The component of normal human cerebrum capacities and the complicated way of learning and making judgments are repeated by the neural system, which includes a to a great extent parallel-circulated processor (neuro-PC) conveying the standard inclination for maintaining experiential information and programmed encompassing adjustment capacity. The learning procedure is likewise named the learning calculation. The solidified attestation on the utilisation of manufactured neural systems has been dictated by the understanding that individuals have over the confounded arrangement of the human mind, with the guide of the traditions of the computerised PC.

Upheld by adequate information depictions and evaluated results, the fake neural system is organised conveniently. Essentially, the Multilayer perceptrons (MLPs) are the feedforward systems with various circumspect neurons with no less than one careful layer. In this study, we utilised a double layer learning neural system.

A two-layer feedforward BPNN was utilised to separate the five distinct sorts of guns. The two layers in the system involve a computationally ascertained concealed layer and an objective yield layer. The neurons in the info layer are connected with the standardised inputs of the numerical elements of the shooting pin impression pictures, while the quantity of yield components is identified with the quantity of particular classes; this is the diverse sorts of guns explored. In the shrouded layer, the weight frameworks are associated with the inputs as they are data weight (IW).

At that point, they are associated from the concealed layer to the yield layer by layer weights (LW). Moreover, S1 and S2 speak to the first and second layers, separately. In the neural system, the yields of each middle person layer are likewise the approaching inputs to the resulting layer. For our situation, layer 2 is analysed as a layer system with LW2,1 weight network of S2 x S1 measurement, where S1 is the yield of the principal layer and S2 is the yield of the second layer. The inputs to layer 2 can be set apart as A1, although A2 is the system craved yield, Y. The inclination of the first and the second layers is connoted by B1 and B2, respectively. In

the interim, F1 and F2 are the actuation elements of layer one and layer two in the two layers of the system, respectively. In general, the mathematical statements of the double layer neural system proposed in this study can be made simpler, as communicated beneath:

The first layer,

$$A1=F1(IW^{1,1}[P]+B1) \tag{1}$$

and then the second layer,

$$A2=F2(LW^{2,1}[A1]+B2)=F2(LW^{2,1}[F1(IW^{1,1}[P]+B1)]+B2)$$

$$A2=Y \tag{2}$$

The learning calculation viewed as normal for a preparation set of the MLP is the backpropagation calculation (BPA). For the premise of our comprehension, BPA worries with two stages; feedforward and in reverse stages. At the feedforward stage, the inputs are proliferated forward through the info hubs keeping in mind the end goal to work out the yield units, and in the regressive stage, the association weights are recharged with the figured and the real units subtracted at the yield layer (Master, 1993; Eberhart & Dobbins, 1990).

In this study, we concentrated on directed preparing, where the inputs and yields are as recognised before by Panca, Moch, and Nanang (2010). In the example acknowledgment hypothesis, the information was isolated into three sets; the preparation (70%), testing (15%) and acceptance (15%) sets. There are some key inputs contained in the preparation set for acknowledgment, predictable with the yields wanted. The subset or tests in the preparation set had been chosen and concentrated on, in a steady progression mechanically in the system. moment elements of the consolidated terminating pin impression pictures as the information. For each specimen, the yield acquired in the system was made to fit the desired yield until all the subsets of the preparation tests were completely utilised, as the weight that associated the system neurons was reestablished. For this case, the blunders set apart at each layer of the system could be lessened, and ideal results were accomplished with respect to the craved yield.

Figure 2 represents the applied procedure, keeping in mind the end goal to hunt down ideal feedforward BPNN. As shown in the figure, five stages were adapted from Ghani et al. (2010), including the cartridge case inputs, picture handling and division, and numerical components determination using Geometric Moments. In this study, a change is made by adding a double layer feedforward BPNN, starting from the procedure of information standardisation to that of weight and inclination modification with various backpropagation preparing calculations accessible in MATLAB planning to get the most ideal arrangement results.

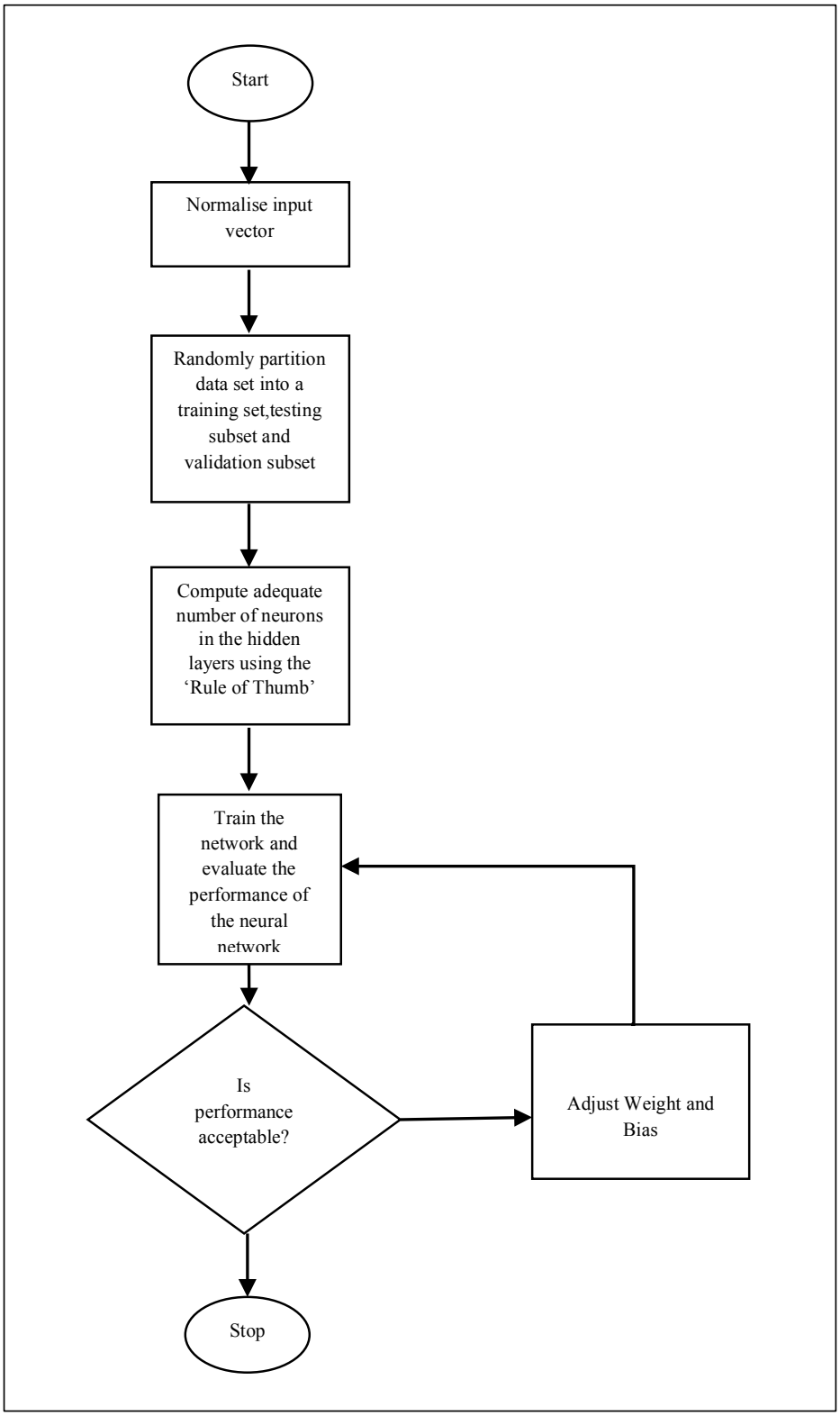


Figure 2. The steps in finding the optimal backpropagation neural network

The MATLAB Neural Network Toolbox (Demuth & Beale, 2001) can actualise speedier preparing procedures and not simply controlling the slope drop calculations, which can work up a hundred times quicker. The most workable calculations can be ordered into the heuristic methodologies (received in e.g. traingda, traingdx and trainrp) and numerical improvement approaches (embraced in e.g. traincgf, traincgp, trainscg, trainbfg, trainoss, and trainlm). The previous alludes to the progression from the standard inclination plummet calculation, while the last serves as the standard advancement application methods. The conspicuousness of this study is obvious in the utilisation of the arbitrary backpropagation preparing calculations since none of the calculations is best connected to each of the issues, as pointed out by Demuth and Beale (2001).

In the principal layer, a double layer system of the sigmoid initiation capacity and in the second layer, the straight enactment capacity, can be utilized effectively to register any useful relationship of the data and yield (2001). In the study, two distinct ways to deal with system structures (double layer tansig/tansig system and double layer tansig/purelin system) are embraced (see Figure 3).

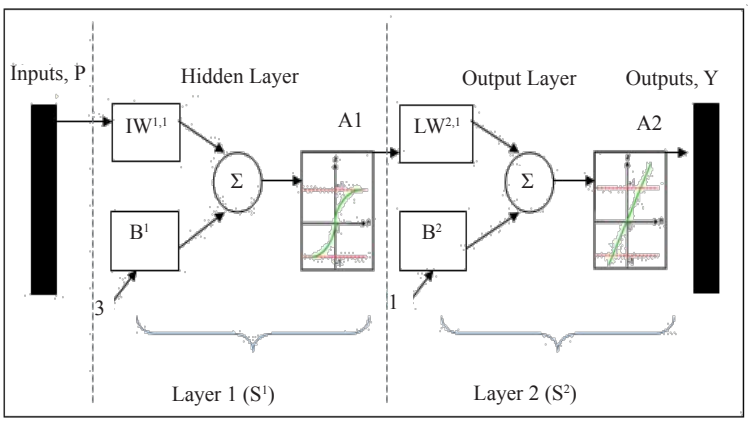


Figure 3. Two-layer tansig/purelin network

The focal point of our consideration is the utilisation of tansig enactment capacity in both the first and second layers since it is the most adaptable exchange capacities for comprehending the boolean values (0 to 1). The information and yield of the system were straightforwardly changed over into twofold values in the information standardisation stage, and later, some arrangement values were created by the system. The double layer tansig/purelin system can incorporate any utilitarian relationship between the inputs and yields with enough neurons.

RESULTS AND DISCUSSION

In this area, grouping exhibitions are quickly highlighted, alluding to two unmistakable neural system approaches with fluctuating sorts of BPA at a general right order rate, as to the eleven components of the consolidated impression pictures of the terminating pin.

Table 2 builds up the characterisation results from 11 × 523 inputs. The inputs were analysed and ordered by the two-layer tansig/tansig system, as per the gun sorts that had been utilised. It is clear that most guns were not grouped effectively with the outcome appeared to be under 70 percent. There were only three preparing calculations performing with over 70 percent accurate characterisation rates, namely, traingd (91%), traincgp (79%), traincgb (76%), and trainscg (79%). The sign is that the other preparing calculations were viewed as inadmissible for this situation. It is obvious that the numerical components of the consolidated impression pictures of the terminating pin are sufficient for ordering the gun. It is apparent that the traincgp and trainscg preparing calculations, in the direct two layer feedforward BPNN with tansig/purelin exchange capacities, have shown the most exact characterisation rate of the joined impression pictures of the terminating pin (79%).

Table 2
Classification rates of different backpropagation training algorithms based on the two-layer network of tansig/tansig transfer functions

Backpropagation Training Algorithms	Overall Correct Classification Rate	Number of neurons	
		L1	L2
traingd	65%	11	5
traingdm	67%	11	5
traindx	49%	11	5
trainrp	52%	11	5
traincgf	68%	11	5
traincgp	79%	11	5
traincgb	76%	11	5
trainscg	79%	11	5
trainbfg	56%	11	5
trainoss	64%	11	5
trainlm	43%	11	5
trainbr	48%	11	5

Table 3 outlines the arrangement consequences of fluctuating backpropagation preparing calculations taking into account the right grouping rate in respect of the components of the impression picture. The arrangement of the two-layer tansig/purelin of the 11-11-5 direct structure system is executed. Just like the past tansig/tansig system, the distinctive backpropagation preparing calculations were also received so that their suitability in the system could be inspected and the arrangement execution could be seen at the accurate grouping rate, which is more than 60%. As shown in Table 3, the most suitable backpropagation preparing calculations incorporate traingd (73%), traindx(75%), traincgf (73%), traincgp (78%), trainscg (84%), and trainbfg (81%). Thus, our perusing demonstrates that trainlm revealed the most elevated score of 87 percent, showing that the calculation relegates the most focused on set to the approved set precisely as to the gun sorts. As a rule, the system of two-layer tansig/purelin beat the two-layer tansig/tansig system in this study (see Figure 3).

Table 3
Classification rates of different backpropagation training algorithms based on the two-layer network of tansig/purelin transfer functions

Backpropagation Training Algorithms	Overall Correct Classification Rate	Number of neurons	
		L1	L2
traingd	73%	11	5
traingdm	64%	11	5
traindx	75%	11	5
trainrp	67%	11	5
traincgf	73%	11	5
traincgp	78%	11	5
traincgb	64%	11	5
trainscg	84%	11	5
trainbfg	81%	11	5
trainoss	66%	11	5
trainlm	87%	11	5
trainbr	65%	11	5

CONCLUSION

By and large, the most attributes must be given to the two layer direct feedforward BPNN with tansig/purelin exchange capacities inclining toward the "trainlm" preparing calculation with the system design 11-11-5. It is further affirmed by the entire accurate order rate of 87%indicating that the numeric geometric moment element in gun grouping is viable.

Thus, it can be concluded that the numerical elements of the consolidated terminating pin impression pictures have demonstrated to be better pointers for gun characterisation. The outcome is very low in comparison with our past works using the numerical elements of the entire terminating pin impression pictures as in Kamaruddin et al. (2011), and the ring terminating pin impression pictures in Kamaruddin, Ghani, Liong, and Jemain (2012), where the general order rates recorded 96% and 98%, respectively. The outcome confirms that the joined picture areas yield significant data on gun characterisation. The characterisation result perceptible in Kamaruddin et al. (2012) is as significant as that of Leng and Huang (2012).

Further investigation should utilise crucial measurable elements that had been created in the past works of Ghani et al. (2009a, 2009b and 2010), in which other neural system strategies such as the convolution neural system can be broken down. This is the point at which picture subtleties can be concentrated more effectively.

ACKNOWLEDGEMENT

The authors would like to express gratitude to the Royal Malaysian Police for its assistance on the laboratory procedures. Special thanks also go to Universiti Teknologi MARA for supporting this research under Research Grants 600-RMI/DANA 5/3/PSI (65/2013) and No. 600-RMI/DANA 5/3/CIFI (65/2013).

REFERENCES

Bonfanti, M. S., & Kinder, J. D. (1999). The Influence of Manufacturing Processes on the Identification of Bullets and Cartridge Cases- A Review of the Literature. *Journal of Science and Justice, 39*(1), 3-10.

Burrard, G. (1951). *Identification of Firearms and Forensic Ballistics.* London: Herbert Jenkins.

Demuth, H., & Beale, M. (2001). *Neural Network for Use with MATLAB*. Version Four, 3 Apple Hill Drive, Natick, MA: TheMathWorks.

Eberhart, R. C., & Dobbins, R. W. (1990). *Neural Network PC Tools: A Practical Guide*. San Diego, California: Academic Press.

Geradts, Z., Bijhold, J., Hermsen, R., & Murtagh, F. (1999). Pattern Recognition in a Database of Cartridge Cases. In *Proceedings of the International Society for Optical Engineering: Investigation and Forensic Science Technologies* (pp. 104-115). SPIE International Society For Optical.

Geradts, Z., Bijhold, J., Hermsen, R., & Murtagh, F. (2001). Image Matching Algorithms for Breech Face Marks and Firing Pin in a Database of Spent Cartridge Cases of Firearms. *Forensic Science International, 119*(1), 97-106.

Ghani, N. A. M., Liong, C. Y., & Jemain, A. A. (2009a). Extraction and Selection of Basic Statistical Features for Forensic Ballistic Specimen Identification. *Sains Malaysiana, 38*(2), 249-260.

Ghani, N. A. M., Liong, C. Y., & Jemain, A. A. (2009b). Analysis of Geometric Moments as Features for Identification of Forensic Ballistics Specimen. *Lecture Notes in Computer Science* (pp. 604-611). Berlin: Springer.

Ghani, N. A. M., Liong, C. Y., & Jemain, A. A. (2010). Analysis of Geometric Moments as Features for Firearm Identification. *Forensic Science International, 198*(1-3), 143-149.

Kamaruddin, S. A., Ghani, N. A. M., Liong C. Y., & Jemain, A. A. (2011). Firearm recognition based on whole firing pin impression image via back propagation neural network. In *Proceedings of IEEE: International Conference on Pattern Analysis and Intelligent Robotics* (pp. 177-182). IEEE.

Kamaruddin, S. A., Ghani, N. A. M., Liong, C. Y., & Jemain, A. A. (2012). Firearm Classification using Neural Networks on Ring of Firing Pin Impression Images. *Advances in Distributed Computing and Artificial Intelligence, 1*(3), 27-35.

Kong, J., Li, D. G., & Watson, A. C. (2003). A Firearm Identification System based on Neural Network. *Advances in Artificial Intelligence, Lecture Notes in Computer Science* (pp. 315-326). Springer, Berlin, Heidelberg.

Kou, C., Tung, C. T., & Fu, H. C. (1994). Firearms Identification based on SOFM Model of Neural Network. In *Proceedings of IEEE: International Carnahan Conference on Security Technology* (pp. 120-125). IEEE.

Leng, J., & Huang, Z. (2012). On Analysis of Circle Moments and Texture Features for Cartridge Images Recognition. *Expert Systems with Applications, 39*(2), 2902-2101.

Li, D. (2006). A New Approach for Firearm Identification with Hierarchical Neural networks based on Cartridge Case Images. In *Proceedings of IEEE: 5th IEEE International Conference on Cognitive Informatics* (pp. 923-928).IEEE.

Masters, T. (1993). *Practical Neural Network Recipes in C++*. San Diego: Academic Press.

Panca, M. R., Moch, R., & Nanang, S. (2010). The Implementation of Feedforward Backpropagation Algorithm for Digit Handwritten Recognition in a Xilinx Spartan-3. *Journal Electrics Electronics Communication Controls Informatics Systems, 4*(2), 17-21.

Smith, C. L., Cross, J. M., & Variyan, G. J. (1995). *FIREBALL: An Interactive Database for the Forensic Ballistic Identification of Firearms.* Research Report, Australian Institute of Security and Applied Technology, Edith Cowan University, Western Australia.

Smith, C. L., & Li, D. (2008). Intelligent imaging of forensic ballistics specimens for ID. In *Proceedings of IEEE: Congress on Image and Signal Processing* (pp. 37-41). IEEE.

Widrow, B., Rumelhart, D., & Lehr, M. A. (1994). Neural Networks: Applications in Industry, Business and Science. *Communications of the Association for Computing Machinery, 37*(3), 93-105.

Xin, L. P., Zhou, J., & Rong, G. (2000). A Cartridge Identification System for Firearms Authentication. In *Proceedings of IEEE: 5th International Conference on Signal Processing* (pp. 1405-1408). IEEE.

Pertanika J. Sci. & Technol. 26 (1): 353 - 366 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Consolidated Backpropagation Neural Network for Malaysian Construction Costs Indices Data with Outliers Problem

Saadi Ahmad Kamaruddin[1], Nor Azura Md Ghani[2]* and Norazan Mohamed Ramli[2]

[1]Computational and Theoretical Sciences Department, Kulliyyah of Science, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia
[2]Centre for Statistical and Decision Sciences Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor Malaysia

ABSTRACT

Neurocomputing has been adjusted effectively in time series forecasting activities, yet the vicinity of exceptions that frequently happens in time arrangement information might contaminate the system preparing information. This is because of its capacity to naturally realise any example without earlier suspicions and loss of sweeping statement. In principle, the most widely recognised calculation for preparing the system is the backpropagation (BP) calculation, which inclines toward minimisation of standard slightest squares (OLS) estimator, particularly the mean squared mistake (MSE). Regardless, this calculation is not by any stretch of the imagination strong when the exceptions are available, and it might prompt bogus expectation of future qualities. In this paper, we exhibit another calculation which controls the firefly algorithm of least median squares (FFA-LMedS) estimator for neural system nonlinear autoregressive moving average (ANN-NARMA) model enhancement to provide betterment for the peripheral issue in time arrangement information. Moreover, execution of the solidified model in correlation with another hearty ANN-NARMA models, utilising M-estimators, Iterative LMedS and Particle Swarm Optimisation on LMedS (PSO-LMedS) with root mean squared blunder (RMSE) qualities, is highlighted in this paper. In the interim, the actual monthly information of Malaysian Aggregate, Sand and Roof Materials value was taken from January 1980 to December 2012 (base year 1980=100) with various levels of anomaly issues. It was found that the robustified ANN-NARMA model utilising FFA-LMedS delivered the best results, with the RMSE values having almost no mistakes at all in all the preparation, testing and acceptance sets for every single distinctive variable. Findings of the studies are hoped to assist the regarded powers including the PFI development tasks to overcome cost overwhelms.

Keywords: ANN, time series, robust backpropagation, firefly algorithm, least median squares

ARTICLE INFO

Article history:
Received: 03 March 2017
Accepted: 28 September 2017

E-mail addresses:
adi8585@yahoo.com (Saadi Ahmad Kamaruddin),
azura@tmsk.uitm.edu.my;
azura158@salam.uitm.edu.my (Nor Azura Md Ghani),
norazan@tmsk.uitm.edu.my (Norazan Mohamed Ramli)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Private Financial Initiative (PFI) is currently a pattern in Malaysia as it is steady with the administration advancing more noteworthy private division's association in maintaining the notoriety of open administration. The most basic benefactor of PFI is value for money (VFM), where ideal nature of development tasks for customer's fulfilment and ventures in the long run are accomplished effectively. It is pivotal to figure on material costs that are brought about through PFI developments to guarantee that overspending will not happen. Since the development works and administration conveyance are the main motivations in the Malaysian PFI, endeavors have been made to foresee the current record of development material value files in Malaysia. It was settled that concrete's controlled cost has been wrecked by the Malaysian government, beginning on 5 June 2008 (Foad & Mulup, 2008). From that point on, there was a significant increment of the bond cost in June 2008, which was by 23.3% in Peninsula Malaysia, while this was 6.5% for Sabah and 5.2% in Sarawak (Foad & Mulup, 2008).

Malaysian government had executed Goods and Services Tax (GST) all through the country since 1 April 2015. Products and Services Tax (GST) is a multi-stage charge on local utilisation. GST is charged on every assessable supply of products and administrations in Malaysia, with the exception of those exempted. GST is likewise charged on importation of merchandise and administrations into Malaysia (Goh, 2015). Because of the GST implimention in Malaysia, engineers are principally hit by the expense of crude materials (Royal Malaysian Customs, 2014). The worst effect is, industry players and specialists expect the costs of private properties to rise 2% to 4% post-GST in spite of the way that such properties are not subject to the GST. In this way, with the execution with GST, combined with the harder working environment, property engineers are liable to methodologies to cradle any negative effect.

The value augmentation is additionally appropriate to the remaining development materials-steel, prepared blend concrete and a few others (Kamaruddin, Ghani, & Ramli, 2014). As development material costs in Malaysia have been met with vulnerability, the best strategy has been examined to give estimation of the development material costs as per the focal area of Malaysia. Next, the related writing is exhibited in the subsequent section, and the foundation of information utilised as a part of this study is depicted in the section that follows. Strategy review is additionally supplied, and the technique used to dissect the information clarified. Next, the concluded results and discourse on the best anticipating approach for evaluating the material value files, as indicated by Malaysian areas, are elaborated in result and discussion section. Finally, the conclusion of the study is elaborated at the end of the paper, together with recommendations for future work.

RELATED LITERATURE

The immediate thought of making the customary neural system learning calculation all the more effective towards remote information is by substituting the mean square errors (MSE) with an alternate symmetric and persistent cost capacity. This will bring about a nonlinear

impact capacity (Rusiecki, 2012) with the ability to provide food for the impacts of extensive mistakes. This must be performed by making the misfortune capacities hearty utilising the factual vigorous strategies to lessen the effects of anomalies issue (Rusiecki, 2012; El-Melegy, Essai, & Ali, 2009), where typical exceptions include event in routine information ranges up to 10% or significantly more (Rusiecki, 2012; El-Melegy et al., 2009; Zhang, 1997), which is the essential subject of this paper.

ANNs serves to be the object of enthusiasm of this exploration as they have turned out to be compelling in numerous exploratory zones (Sugunnasil, Somhom, Jumpamule, & Tongsiri, 2014). This is contemplated by the capacity of the mainstream feedforward neural systems as a general capacity approximator (El-Melegy et al., 2009). The greater part of past studies have attempted to enhance adaptation so as to learn calculation of feedforward neural systems, the M-estimators, overwhelmingly.

In 1996, Liano (1996) presented the LMLS (Least Mean Log Squares) strategy. He presented the logistic mistake capacity by shaping a presumption of the blunders produced utilising the Cauchy appropriation. This commitment has motivated different creators to make some more equipped capacities. The thought of M-estimators by Hampel (Hampel, Ronchetti, Rousseeuw & Stahel, 1986) had been proceeded by Chen and Jain (Chen & Jain, 1994) as they added to another mistake basis called Hampel's hyperbolic digression, where β estimator was utilised to characterise the extent of residuals thought to be anomalies.

Hector, Claudio, and Rodrigo (2002) found that a vigorous calculation for nonlinear autoregressive (NAR) models utilising the summed up most extreme probability (GM) sort estimators beat the minimum squares technique in dealing with the exceptions. In a study by Chuang, Su, and Hsiao (2000), the toughening plan was connected to decrease the estimation of β with the preparation progress. There were likewise approaches that additionally have execution capacities taking into account the tau-estimators (Pernia-Espinoza, Ordieres-Mere, Martinez-de-Pison, & Gonzalez-Marcos, 2005) and the LTS (Least Trimmed Squares) estimator, while the start-up information examination with the MCD (Minimum Covariance Determinant) estimator was recommended (Rusiecki, 2012). El-Melegy et al. (2009) have exhibited the Simulated Annealing for Least Median of Squares (SA-LMedS) calculation, as they connected the reproduced strengthening procedure to relieve the execution measured by the middle of squared residuals. A few endeavours to make the learning techniques for outspread premise capacity organises all the more effective, after the methodologies for the sigmoid systems, have additionally been practiced (Chuang, Jeng, & Lin, 2004; David, 1995). The most recent vigorous learning strategies to be specified are powerful co-preparing in view of the authoritative connection examination, as set forth by Sun and Jin (2011), and hearty versatile learning utilising direct grid imbalance methods (Jing, 2012).

In a paper composed by Rusiecki (2012), another hearty learning calculation in view of the iterated Least Median of Squares (LMedS) estimator was presented. This new approach is a great deal more compelling and strikingly speedier than the SA-LMedS technique (El-Melegy et al., 2009). It likewise accomplishes better imperviousness to imperfect preparing

information. To guarantee the power of the preparation process that the execution capacity is changed, information suspected to be exceptions is evacuated iteratively. A rough technique to minimise the LMedS blunder rule was proposed.

In any case, it is clear that each of these works has concentrated on the NAR model only. In other words, none of the works has considered utilising a strong methodology to enhance the NARMA model. The general execution of the NARMA model is superior to the NAR model (Bruna, 1994). It is the curiosity of the methodology that the current vigorous estimators are executed on BPNN of the NARMA models. Another new variable of the examination is interpreted in the augmentation of study towards the utilisation of molecule swarm advancement (PSO) to minimise the LMedS mistake standard, as started by Shinzawa, Jiang, Iwahashi and Ozaki (2007), with adjustment of the NARMA model.

PSO, created by Eberhart and Kennedy (1995), is a stochastic inquiry technique which takes motivation from the demonstration of winged animals rushing. Like the hereditary calculation (GA), PSO is a populace based enhancement apparatus that searches for optima by upgrading eras (Eberhart & Kennedy, 1995; Shi & Eberhart, 1998; Clerc & Shi, 2002; Eberhart & Shi, 2001; Yu, Wang, & Xi, 2008). Be that as it, unlike the GA, no development administrators were incorporated by the PSO (Goldberg, 1989). When contrasted with GA, a striking favourable position of PSO is that its calculation has a great basic idea, while calculation expenses are not high and only a couple of flexible parameters are required.

Also, in 2007, Xin-She Yang from Cambridge University added to another metaheuristic calculation known as the firefly (FA) calculation (Yang, 2008; Yang & Deb, 2009; Yang, 2009; Yang & He, 2013; Yang, 2010a; Yang, 2010b). The firefly calculation was found to perform better in comparison with molecule swarm advancement in taking care of the abnormal state of commotion (Pal, Rai & Singh, 2012). In this study, another methodology, with robustify the backpropagation learning calculation of nonlinear neural system time arrangement models, was used utilising FA-LMedS estimator. This paper includes the execution of LS, M-estimators, ILMedS, PSO-LMedS and FA-LMedS in the backpropagation calculation of both BPNN-NAR and BPNN-NARMA models.

DATA BACKGROUND

Information was incorporated from three unique sources of Unit Kerjasama Awam Swasta (UKAS) of Prime Minister's Department, Construction Industry Development Board (CIDB) and Malaysian Statistics Department which have supported the PFI development material value records for the Central area of the Peninsula comprising four states of Kuala Lumpur Federal Territory, Selangor, Negeri Sembilan and Melaka. The genuine modern month-to-month information of Malaysian Aggregate, Sand and Roof Materials value records from January 1980 to December 2012 (base year 1980=100) were adjusted, with various rates of anomalies issues, 3.9%, 0% and 8.1% individually.

Table 1 displays the synopsis measurements of the variables of hobby. The aggregate N=408 (12 months × 34 years) was from January 1980 to 2013 (base 1980=100). The mean of sand is the most astounding (198.6969), trailed by rooftop materials (131.6038) and total (113.7731). Definitively, the cost of sand is the most exorbitant compared to rooftop materials and in total.

Table 1
Summary statistics of the construction materials price indices data

Notation	N	Mean	Std. Dev.	Max	Min	Skewness	Kurtosis	J-B
Aggregate	408	113.7731	7.63405	140.63	99.2	1.409	2.803	0.873**
Sand	408	198.6969	68.4966	287.88	100	0.143	-1.730	0.828**
Roof Materials	408	131.6038	8.21297	150.04	100	-0.321	3.508	0.786**

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

Likewise, sand demonstrates the most elevated standard deviation (68.4966) compared to the total (7.63405) and rooftop materials (8.21297). Both total and sand are emphatically skewed which are 1.409 and 0.143, respectively. However, rooftop materials are contrarily skewed (- 0.321). Be that as it may, taking into account the Jarque-Bera test for ordinariness, each of the three variables are very critical at 99% certainty interim; total (J-B=0.873, p=0.000), sand (J-B=0.828,p=0.000), and rooftop materials (J-B=0.786,p=0.000). The variables of interest experienced the ill effects of anomalies issues, as illustrated in Figure 1, Figure 2 and Figure 3, respectively.

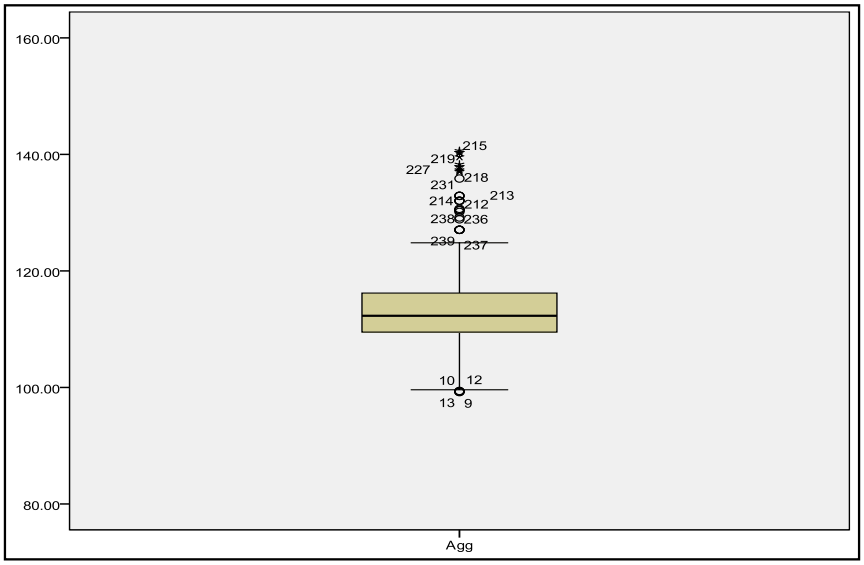


Figure 1. The boxplot of Malaysian Aggregate data

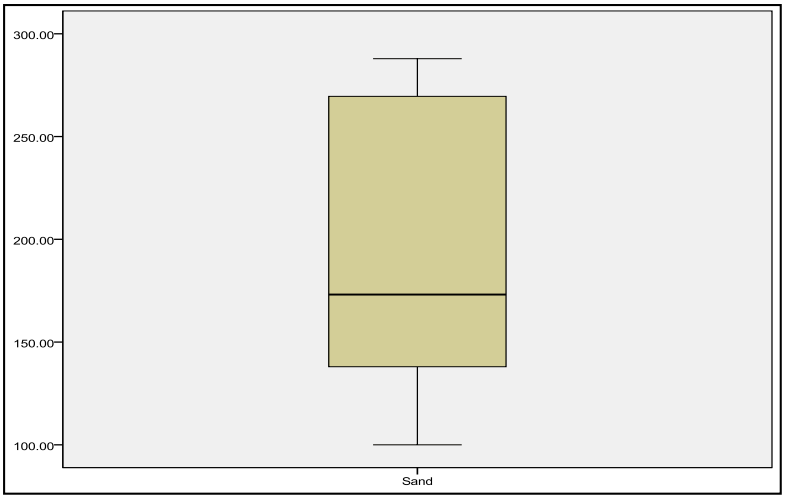


Figure 2. The boxplot of Malaysian Sand data

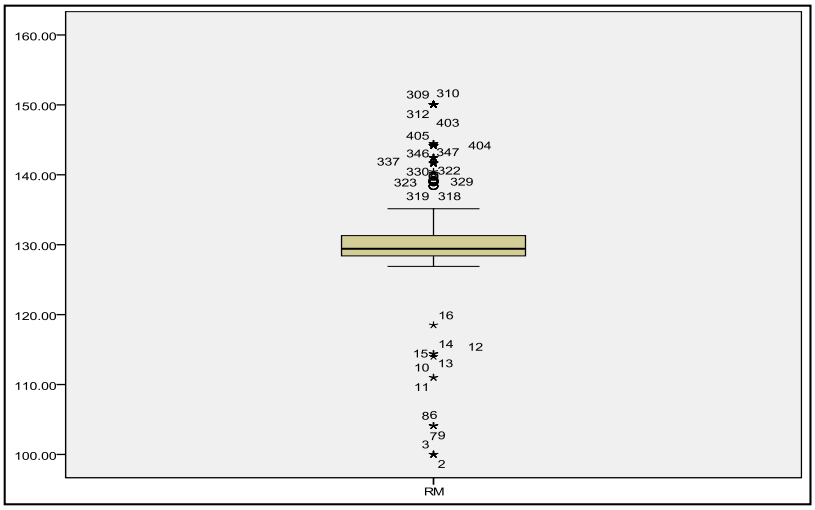


Figure 3. The boxplot of Malaysian Roof Materials data

Table 2
Stopping criteria

MATLAB Terms	Values	NN Terms
net.trainParam.epochs	1000	Maximum number of epochs to train
net.trainParam.goal	0	Performance goal
net.trainParam.max_fail	6	Maximum validation failures
net.trainParam.min_grad	1e^-7	Minimum performance gradient
net.trainParam.mu	0.001	Initial μ
net.trainParam.mu_dec	0.1	μ decrease factor
net.trainParam.mu_inc	10	μ increase factor
net.trainParam.mu_max	1e^10	Maximum μ

METHODOLOGY

A flowchart of the examination is given in Figure 4. In the figure, the current vigorous estimators on backpropagation neural system are actualised. In order to obtain the primary target of the

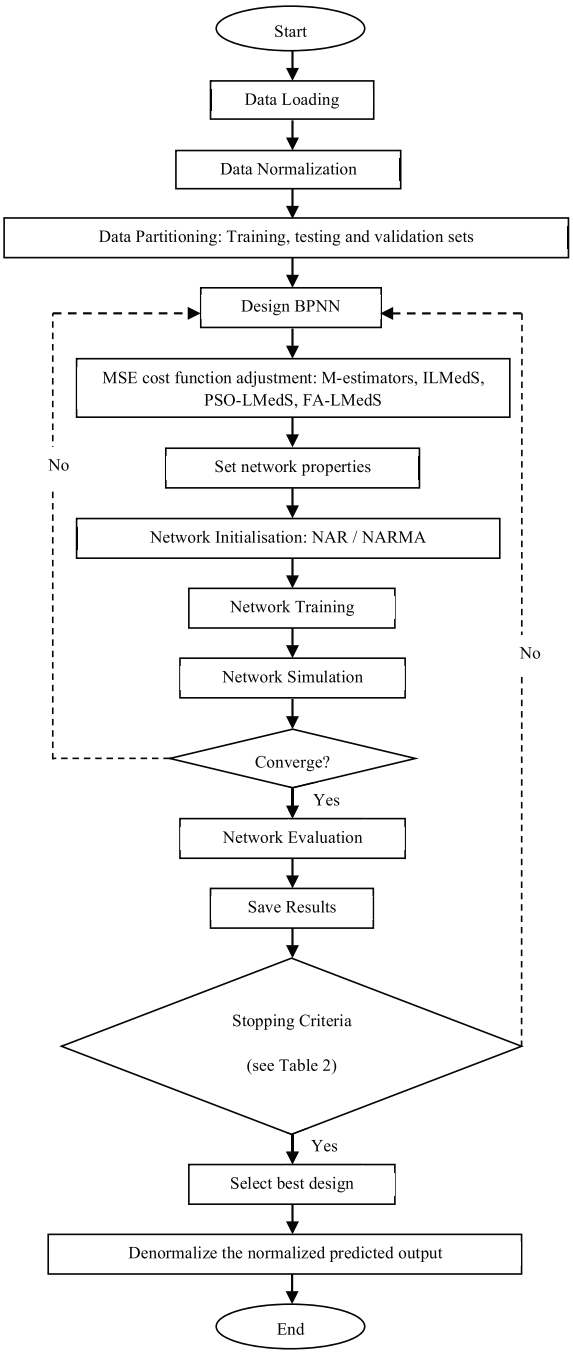


Figure 4. Flowchart of proposed robust backpropagationNN

study, the conceivable powerful estimators half breed in nonlinear autoregressive (NAR) and nonlinear autoregressive moving normal (NARMA) of neural system time arrangement were done using MATLAB R2012a. NARMA model was extended from NAR model by including the error terms as new inputs. At this stride, MATLAB scripts or codings were composed parallel to the numerical plan done prior to that. This was followed by execution of the proposed robustified neural system models, which were thought about utilising genuine information using standard execution measures (RMSE). The best relative results were drawn where the best model was picked. At the end of this examination, a programmed anticipating framework improvement was set up using MATLAB guide client interface (GUI) that had been effectively created. Finally, forecasts without bounds value lists of the Malaysian development material in the coming years before the best model were done at this stage. Details of the fundamental NAR-ANN are given below. The basic NAR-ANN formulation can be represented as follows:

$$H(x) = purelin\left[\sum_{j=1}^{m} w_{jk}\left[\tanh\left(\sum_{i=1}^{l} w_{ij}\left[x(t-1), x(t-2),..., x(t-n_y)\right] + \varepsilon(t)\right)\right]\right] \quad (1)$$

The finalised NARMA-ANN formulation can be represented as follows:

$$H(x) = purelin\left[\sum_{j=1}^{m} w_{jk}\left[\tanh\left(\sum_{i=1}^{l} w_{ij}\begin{bmatrix}[x(t-1), x(t-2),..., x(t-n_y),\\ \varepsilon(t-1), \varepsilon(t-2),..., \varepsilon(t-n_\varepsilon)] + \varepsilon(t)\end{bmatrix}\right)\right]\right] \quad (2)$$

where

H(x) is the estimated model,

x(t-1), x(t-2),…, x(t-ny) are lagged input terms,

ε(t-1), ε(t-2),…ε(t-nε) are lagged residual terms, and the lagged residual terms are obtained recursively after the initial model (based on the input and output terms) has been found.

Hence, ε(t) are the white noise residuals.

l is the input neurons with index i

m is the hidden neurons with index j

n is the output neurons with index k

Robust Backpropagation Algorithm

The most essential part of the study is the scientific definition change a portion of backpropagation neural system calculation utilising measurable vigorous estimators. In order to make the customary backpropagation calculation effective in light of the M-estimators idea for lessening anomaly impact, the squared residuals ε_i^2 in the network error by another capacity of the residuals

$$E = \frac{1}{N}\sum_{i}^{N} \varepsilon_i^2 \quad , \quad (3)$$

and this yields,

$$E = \frac{1}{N}\sum_{i}^{N}\rho(\varepsilon_i) \quad , \tag{4}$$

where N is the total number of samples available for network training. The updated network weights was obtained based on the gradient descent learning algorithm. To prevent generality loss, a feedforward neural network, with one hidden layer, was implemented in this study. The weights from the hidden neurons to output neurons, *Wj,i*, are expressed as

$$\Delta W_{j,i} = -\alpha\frac{\partial E}{\partial W_{j,i}} = -\frac{\alpha}{N}\sum_{i}^{N}\frac{\partial\rho(\varepsilon_i)}{\partial W_{j,i}} \tag{5}$$
$$= -\frac{\alpha}{N}\sum_{i}^{N}\varphi(r_i).\frac{\partial f_j}{\partial net_j}.O_i,$$

where α is a user-supplied learning constant, Oh is the output of the i^{th} hidden neuron, $O_j=f_j(net_j)$ is the output of the j^{th} output neuron, $net_j=\sum_i W_{ji}O_i$ is the induced local field produced at the input of the activation function associated with the output neuron (*j*), and f_j is the activation function of the neurons in the output layer. In this paper, a linear activation function (purelin) is used in the output layer's neurons. The weights from the input to hidden neurons $W_{j,i}$ are updated as:

$$\Delta W_{ji} = -\alpha\frac{\partial E}{\partial W_{ji}} = -\frac{\alpha}{N}\sum_{i}^{N}\frac{\partial\rho(\varepsilon_i)}{\partial W_{j,i}} \tag{6}$$
$$= -\frac{\alpha}{N}\sum_{i}^{N}\sum_{j}\ \varphi(r_i).\frac{\partial f_j}{\partial net_j}.W_{j,i}.\frac{\partial f_i}{\partial net_i}.I_i,$$

where I_i is the input to the i^{th} input neuron, $net_j=\sum W_{j,i}O_i$ is induced local field produced at the input of the activation function associated with the hidden neuron (i), and f_j is the activation function of the neurons in the hidden layer. We have the intention to use the tan-sigmoid function as the activation function for the hidden layer's neurons because of its flexibility.

The least-median-of-square (LMedS) method estimates the parameters by solving the nonlinear minimisation problem.

$$minmed_i\ \varepsilon_i^2 \tag{7}$$

That is, the estimator must create the least worth for the least median squares figured for the whole information set. It creates the impression that this strategy is extremely hearty to false matches, particularly to anomalies inferable from terrible limitation (El-Melegy at al., 2009). Unlike the M-estimators, the LMedS issue can not be lessened to a weighted slightest squares issue. It is probably impossible to write a clear equation for the subordinate of LMedS estimator. Subsequently, deterministic calculation will not have the capacity to minimise that estimator.

The Monte-Carlo method (Zhang, 1997; Aarts, Korst, & Michiels, 2005) has been honed to take care of this issue in some non-neural applications. Stochastic calculations are likewise distinguished as the enhancement calculations which utilise arbitrary hunt to achieve an answer. Stochastic calculations are generally moderate, yet there is a probability that it will locate the worldwide least. One very well known improvement calculation to minimise a LMedS-based system blunder is mimicked toughening (SA) calculation. SA is an eminent calculation as it is moderately broad and has the inclination not to get stuck in either the neighbourhood least or most extreme (El-Melegy at al., 2009). Nonetheless, Rusiecki (2012) found that iterated LMedS has a tendency to beat the SA-LMedS.

RESULTS AND DISCUSSION

Table 3, 4 and 5 demonstrate the correlations between execution aftereffects of robustified nonlinear autoregressive and nonlinear autoregressive moving normal of the artificial neural system time arrangement models on Malaysian Aggregate, Sand and Roof Materials value files information, respectively.

Table 3
Comparison of the best results of ordinary and modified backpropagation algorithms on Malaysian Aggregate Price Index Data

Malaysian Aggregate Price Index Data							
BP Learning Algorithm	Input Lags	Error Lags	Hidden Nodes		NAR		NARMA
MSE	10	10	20		592.877		253.726
M-estimators (L2)	15	15	25		0.123		0.013
M-estimators (L1)	15	15	25		0.123		0.013
M-estimators (L1-L2)	10	10	20		0.053		0.006
M-estimators (LP)	25	25	40		0.040		0.002
M-estimators (Fair)	15	15	25		0.074		0.006
M-estimators (Huber)	15	15	20		0.006		0.000
M-estimators (Cauchy)	25	25	40		0.094		0.094
M-estimators (Geman-McClaure)	15	15	25		0.072		0.006
M-estimators (Welsch)	15	15	20		0.015		0.000
M-estimators (Tukey)	20	20	35		0.070		0.002
ILMedS	15	15	20		0.053		0.003
BP Learning Algorithm	Input Lags	Error Lags	Hidden Nodes	Swarm Size	Iteration	NAR	NARMA
PSO-LMedS	15	15	40	40	20	0.005	0.005
FFA-LMedS	15	15	20	20	20	0.070	0.002

Table 4
A comparison of the best results of ordinary and modified backpropagation algorithms on Malaysian Sand Price Index Data

Malaysian Sand Price Index Data							
BP Learning Algorithm	Input Lags	Error Lags	Hidden Nodes		NAR		NARMA
MSE	10	10	10		0.004		0.006
M-estimators (L2)	20	20	20		0.120		0.013
M-estimators (L1)	20	20	20		0. 119		0.013
M-estimators (L1-L2)	25	25	35		0.049		0.011
M-estimators (LP)	25	25	35		0.878		0.002
M-estimators (Fair)	10	10	10		0.136		0.017
M-estimators (Huber)	25	25	25		0.235		0.000
M-estimators (Cauchy)	25	25	35		0.152		0.114
M-estimators (Geman-McClaure)	40	40	45		0.005		0.003
M-estimators (Welsch)	25	25	25		0.172		0.000
M-estimators (Tukey)	40	40	45		0.161		0.002
ILMedS	25	25	25		0.053		0.003
BP Learning Algorithm	Input Lags	Error Lags	Hidden Nodes	Swarm Size	Iteration	NAR	NARMA
PSO-LMedS	10	10	40	40	25	0.005	0.002
FFA-LMedS	10	10	20	20	25	0.063	0.000

Table 5
A comparison of the best results of ordinary and modified backpropagation algorithms on Malaysian Roof Materials Price Index Data

Malaysian Roof Materials Price Index Data							
BP Learning Algorithm	Input Lags	Error Lags	Hidden Nodes		NAR		NARMA
MSE	20	20	35		309.435		87.614
M-estimators (L2)	10	10	20		0.113		0.013
M-estimators (L1)	10	10	20		0.143		0.012
M-estimators (L1-L2)	10	10	20		0.088		0.007
M-estimators (LP)	15	15	30		0.878		0.002
M-estimators (Fair)	10	10	10		0.134		0.017
M-estimators (Huber)	15	15	40		0.235		0.000
M-estimators (Cauchy)	15	15	30		0.152		0.094
M-estimators (Geman-McClaure)	20	20	25		0.005		0.003
M-estimators (Welsch)	15	15	40		0.172		0.000
M-estimators (Tukey)	20	20	25		0.161		0.002
ILMedS	15	15	40		0.053		0.003
BP Learning Algorithm	Input Lags	Error Lags	Hidden Nodes	Swarm Size	Iteration	NAR	NARMA
PSO-LMedS	20	20	40	40	35	0.005	0.002
FFA-LMedS	20	20	20	20	35	0.070	0.000

The outcomes depend on the diverse parameter settings blends in both the ANN-NAR and ANN-NARMA models.

CONCLUSION

In this study, nonlinear time arrangement neural system models (NAR and NARMA) were utilised to adapt instability without bounds (Kilicman & Roslan, 2009). As it is difficult to get rid of the vicinity of anomalies in genuine information set, preparing feedforward neural systems by the prevalent backpropagation calculation might create wrong and offbase models on the ground that the first MSE learning calculation is not hearty, and accordingly, effectiveness is lost (Norazian, Shukri, Azam, & Bakri, 2008). In this manner, there is a need to supplant the MSE cost capacity with other strong cost capacities such as M-estimators, ILMedS, PSO-LMedS and FFA-LMedS.

In future work, FFA-LMedS should be investigated for true information which comprises 30% to half distant information. Finally, the proposed calculations for preparing neural systems might be adjusted to different fields of counterfeit consciousness, framework distinguishing proof, design acknowledgment, machine learning, quality control and streamlining and exploratory processing.

ACKNOWLEDGMENT

Authors would like to acknowledge Unit Kerjasama Awam Swasta (UKAS) of Prime Minister's Department, Construction Industry Development Board (CIDB) and Malaysian Statistics Department. The authors gratefully acknowledge the financial support from the Ministry of Higher Education, Malaysia, and Universiti Teknologi MARA for the Research Grant No. 600-RMI/DANA 5/3/CIFI (64/2013) and Fundamental Research Grant Scheme (FRGS) under the Research Grant No. 600-RMI/FRGS 5/3 (137/2014). The authors also wish to thank the International Islamic University, Malaysia, and MOHE for the research grant awarded to this project, RIGS 16-092-0256.

REFERENCES

Aarts, E., Korst, J., & Michiels, W. (2005). Simulated annealing. In E. K. Burke & G. Kendall (Eds.), *Search Methodologies* (pp. 187-210). United States of America, USA: Springer International Publishing AG.

Bruna, G. M. (1994). Short term load forecasting using non-linear models. The Netherlands: Measurement and Control Section ER. Electrical Engineering. (Master Thesis Report). Eindhoven University of Technology, Netherlands.

Chen, D. S., & Jain, R. C. (1994). A robust back propagation learning algorithm for function approximation. *IEEE Transactions on Neural Networks, 5*(3), 467-479.

Chuang, C. C., Jeng, J. T., & Lin, P. T. (2004). Annealing robust radial basis function networks for function approximation with outliers. *Neurocomputing, 56*, 123-139.

Chuang, C., Su, S., & Hsiao, C. (2000). The annealing robust backpropagation (ARBP) learning algorithm. *IEEE Transactions on Neural Networks, 11*(5), 1067-1077.

Clerc, M., & Kennedy, J. (2002). The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation, 6*(1), 58-73.

David, S. V. A. (1995). Robustization of a learning method for RBF networks. *Neurocomputing, 9*(1), 85-94.

Eberhart, R. C. & Shi, Y. (2001). Particle swarm optimization: developments, applications and resources. In *Proceedings IEEE: Congress on Evolutionary Computation* (pp. 81-86). IEEE.

El-Melegy, M. T., Essai, M. H., & Ali, A. A. (2009). Robust training of artificial feedforward neural networks. In A. E. Hassanien, A. Abraham, A. V. Vasilakos & W. Pedrycz (Eds.), *Foundations of Computational, Intelligence Volume 1* (pp. 217–242). Springer, Berlin, Heidelberg.

Foad, H. M., & Mulup, A. (2008, June 2). Harga siling simen dimansuh 5 Jun. *Utusan*. Retrieved from http://ww1.utusan.com.my/utusan/info.asp?y=2008&dt=0603&sec=Muka_Hadapan&pg=mh_02.htm

Goh, J. (2015, February 9). Developers strategizing to buffer GST impact. *The Edge Malaysia, MSN News.*

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading: Addison-Wesley.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust Statistics, The approach based on influence functions*. New York: Wiley.

Hector, A., Claudio M., & Rodrigo, S. (2002). Robust Estimator for the Learning Process in Neural Network Applied in Time Series. In *International Conference on Artificial Neural Networks: Lecture Notes in Computer Science Springer, Berlin, Heidelberg* (pp. 1080-1086).

Jing, X. (2012). *Robust adaptive learning of feedforward neural networks via LMI optimizations.* Neural Networks, 31, 33-45.

Kamaruddin, S. B. A., Ghani, N. A. M., & Ramli, N. M. (2014). Best Forecasting Models for Private Financial Initiative Unitary Charges Data of East Coast and Southern Regions in Peninsular Malaysia. *International Journal of Economics and Statistics, 2*, 119-127.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of IEEE: International Conference on Neural Networks, 2*(4), 1942-1948.

Kilicman, A., & Roslan, U. A. M. (2009). Tuberculosis in the Terengganu region: Forecast and data analysis. *ScienceAsia, 35*, 392-395.

Liano, K. (1996). Robust error measure for supervised neural network learning with outliers. IEEE Trans. *Neural Networks, 7*(1), 246-250.

Norazian, M. N., Shukri, Y. A., Azam, R. N., & Bakri, A. M. M. A. (2008). Estimation of Missing Values in Air Pollution Data using Single Imputation Techniques. *ScienceAsia, 34*(2), 341-345.

Pal, S. K., Rai, C. S., & Singh, A. P. (2012). Comparative Study of Firefly Algorithm and Particle Swarm Optimization for Noisy Non-Linear Optimization Problems. *International Journal of Intelligent Systems and Applications, 4*(10), 50-57.

Pernia-Espinoza, A. V., Ordieres-Mere, J. B., Martinez-de-Pison F. J., & Gonzalez-Marcos, A (2005). A TAO-robust backpropagation learning algorithm. *Neural Network, 18*(2), 191–204.

RMC. (2014, October 29). *Good and Sevices Tax: Guide on Construction Industry.* Royal Malaysian Customs.

Rusiecki, A. (2012). Robust Learning Algorithm Based on Iterative Least Median of Squares. *Neural Processing Letters, 36*(2), 145-160.

Shi, Y., & Eberhart, R. A. (1998). Modified particle swarm optimizer. In *Evolutionary, Proceedings of IEEE: World Congress on Computational Intelligence* (pp. 69-73). IEEE.

Shinzawa, H., Jiang, J. H., Iwahashi, M., & Ozaki, Y. (2007). Robust curve fitting method for optical spectra by least median squares (LMedS) estimator with particle swarm optimization (PSO). *Analytical Sciences, 23*(7), 781-785.

Sugunnasil, P., Somhom, S., Jumpamule, W., & Tongsiri, N. (2014). Modelling a neural network using an algebraic method. *ScienceAsia, 40*, 94-100.

Sun, S., & Jin, F. (2011). Robust co-training. *International Journal of Pattern Recognition Artificial Intelligence, 25*(7), 1113-1126.

Yang, X. S. (2008). *Nature-Inspired Metaheuristic Algorithms*. United Kingdom: Luniver Press.

Yang, X. S. (2009). Firefly Algorithms for Multimodal Optimization. *Proceeding of the 5th Symposium on Stochastic Algorithms, Foundations and Applications.* In O. Watanabe & T. Zeugmann (Eds.), Lecture Notes in Computer Science, 5792 (pp. 169-178).

Yang, X. S. (2010a). *Engineering Optimisation: An Introduction with Metaheuristic Applications.* USA: John Wiley and Sons.

Yang, X. S. (2010b). A New Metaheuristic Bat-inspired algorithm. In J. R. Gonzales, D. A. Pelta, C. Cruz, G. Terrazas & N. Krasnogor (Eds.), Nature Inspired Cooperative Strategies for Optimisation (pp, 65-74). Springer, SCI. 284.

Yang, X. S., & Cuckoo, D. S. (2009). Search via Levy Flights. In *Proceedings of IEEE: World Congress on Nature & Biologically Inspired Computing* (pp. 210-214). IEEE.

Yang, X. S., & He, X. (2013). Firefly Algorithm: Recent Advances and Applications. *International Journal Swarm Intelligence, 1*(1), 36-50.

Yu, J., Wang, S., & Xi, L. (2008). Evolving artificial neural networks using an improved PSO and DPSO. *Neurocomputing, 71*(4), 1054-1060.

Zhang, Z. (1997). Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing, 15*(1), 59-76.

Pertanika J. Sci. & Technol. 26 (1): 367 - 378 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Application of the First Order of Markov Chain Model in Describing the PM10 Occurrences in Shah Alam and Jerantut, Malaysia

Mohamad, N. S.[1]*, Deni, S. M.[1] and Ul-Saufie, A. Z.[2]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Malaysia

[2]Department of Computer and Mathematical Sciences, Universiti Teknologi MARA, 13500 UiTM, Permatang Pauh, Pulau Pinang, Malaysia

ABSTRACT

PM_{10} has been identified as being a common problem in Malaysia and many other countries all over the world. A Markov chain probability model is found to fit the average daily PM_{10} concentrations data of urban station (Shah Alam) and background area station (Jerantut) in Malaysia. This study aims to identify the occurrence of polluted and non-polluted days affected by PM_{10} concentrations based on data for 12 years' period (2002-2013). The first order transition probability matrix of a Markov chain model and a two-state Markov chain, which are polluted days (1) and non-polluted days (0), were used for this purpose. The threshold value used in this study is referring to WHO 2006 guidelines ($50\mu gm^{-3}$). Results of the analysis shows that there is a high probability that the next day event depends on what has happened on the previous day. The recurrence of the polluted day for Shah Alam is 4-5 days, while 2-3 days for Jerantut. By fitting the first order of Markov chain model, the results show that the higher order of Markov chain model is needed in order to get the best fitted distribution of polluted events at these two monitoring stations. Thus, the prediction of PM_{10} concentrations event can be made by considering the conditions of the previous day event.

Keywords: Markov chain model, PM_{10} concentrations, polluted days, non-polluted days, occurrence

ARTICLE INFO

Article history:
Received: 03 March 2017
Accepted: 28 September 2017

E-mail addresses:
norsalwani630217@gmail.com (Mohamad, N. S.),
sayang@tmsk.uitm.edu.my (Deni, S. M.),
ahmadzia101@ppinang.uitm.edu.my (Ul-Saufie, A. Z.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Particles or particulate matter is a type of air pollution. PM_{10}, which is a particulate matter with aerodynamic diameter less than 10 micrometre, has been identified as being a common problem in Malaysia and some other countries. Furthermore, PM_{10} is also mainly considered in Malaysia as this pollutant gives

greatest concern to public health, apart from being the main air pollutant that has exist since the haze event in 1980 (Awang et al., 2000). Cardiovascular diseases, respiratory problems and cancer are some examples of health problems caused by this air pollutant (EPA, 2001). PM_{10} is emitted by both anthropogenic and natural sources. There are five major PM_{10} emissions in Malaysia, which are motor vehicle exhausts, industrial sources, open burning, heat and power plants (DoS, 2013). However, the main causes of this pollutant in Malaysia are industrial activities and heavy traffic (Ul-Saufie et al., 2012). Many efforts have been done in order to monitor the effects of this pollutant; these include the regression model that is used to predict PM_{10} concentrations (Paschalidou et al., 2011; Chaloulakou et al., 2003; Pires et al., 2008). However, the study of the sequence of polluted and non-polluted days using Markov chain model still receives less attention.

Markov chain model was introduced at the beginning of the 20th century and it has been used in many disciplines including geography, meteorology, biology, physics, behaviour, social sciences, chemistry and others (Berchtold & Raftery, 2002). Many researchers have used this method to model the daily occurrences of rainfall and have successfully used it especially for the first order (Gabriel & Neumann, 1962; Mangaraj et al., 2013; Basu, 1971). The first order Markov chain is simple and it does not require a lot of computational efforts. The study of Berchtold and Raftery (2002) successfully used Markov chain model in describing the occurrence of daily rainfall in Tel Aviv for the mid-winter season. Besides that, other researchers such as Haan et al., (1976), Jimoh and Webster, (1996), Rahimi et al. (2011) and Chin (1977) also used Markov chain model in their fields of study. The order of the chain gives the number of time steps in the past influencing the probability distribution of the present state, which could be greater than one. Several authors have found that the order more than one gave the optimum result for certain conditions. For example, Deni et al. (2009) found that the first order of Markov chain model is the optimum order for the rainfall occurrence during the northwestern and eastern regions of peninsula, while a higher order is fit for the rainfall occurrence during the northeast monsoon season. However, this study only considers the first order of Markov chain model because of its simplicity.

The main purpose of this study is to determine the occurrence of polluted and non-polluted days affected by PM_{10} concentrations at an urban area (Shah Alam) and a background area (Jerantut) in Malaysia. This kind of study still receives less attention from other researchers. This study was conducted as follows. After a brief description of the PM_{10} concentrations at both stations in the data and methodology section, the explanation on how to determine the polluted day and non-polluted day and the transition probability are also discussed. Testing for the Markov chain property is also discussed in this section to determine whether or not this study possesses the first order of Markov property. This is followed by the persistency of the polluted (non-polluted) events. Results of the analysis and conclusion of this study are discussed in the results and discussion, and conclusion, respectively.

DATA AND METHODOLOGY

Study Area and Data Preparation

Shah Alam is the state capital of Selangor, Malaysia (Figure 1). It is located about 25 kilometres west of the country's capital, Kuala Lumpur. Shah Alam monitoring station is located at Sekolah Kebangsaan Raja Muda, Shah Alam. This station was chosen based on its land use background (urban area) as the absence of PM_{10} concentrations is higher at this area because of the daily activities and heavy traffic emissions. In Jerantut, the background station is located at the Meteorology Monitoring Station, Batu Embun, Jerantut. This monitoring station was selected as the background area because it is situated at the rural area and still has low potential of recording air pollution in Malaysia (Latif et al., 2014). The latitude and longitude for both stations are shown in Table 1.

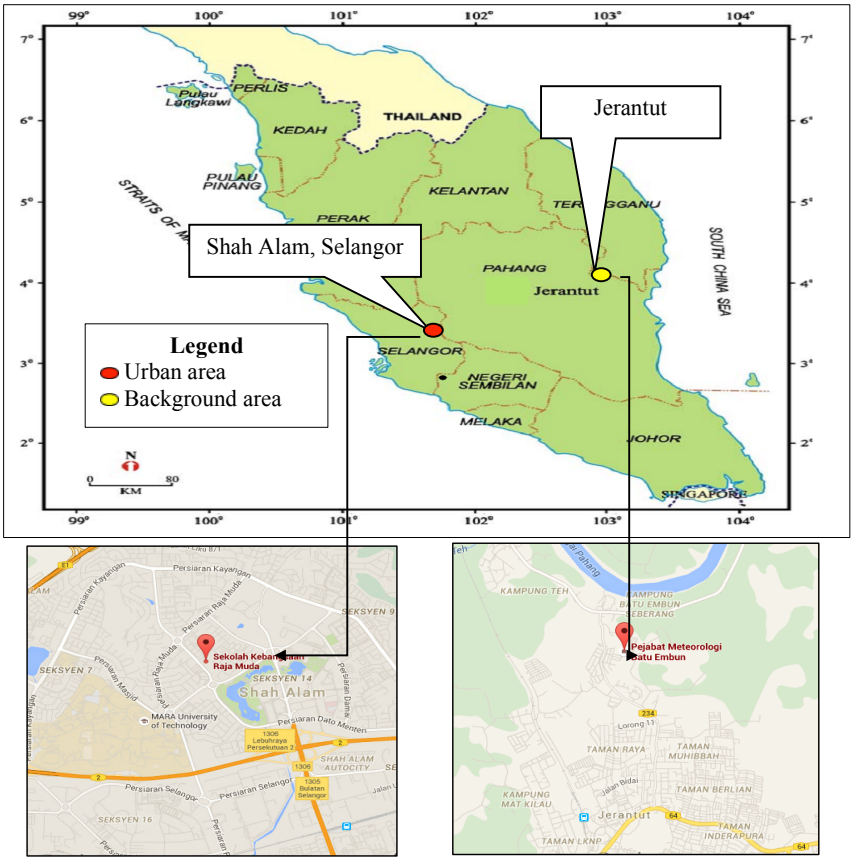


Figure 1. The location of the monitoring stations for Shah Alam and Jerantut, Malaysia

Data for twelve years (2002-2013) of the PM_{10} concentration monitoring, provided by the Department of Environment (DoE) Malaysia, are used in this study. In this study, the hourly data of PM_{10} concentrations were transformed into daily average data. However, the PM_{10} concentrations data were incomplete due to missing values as they were obtained using an automated machine. Thus, according to Junninen et al. (2004), the missing value for the air pollution monitoring record is in the category of missing at random and as the data used in this study contained small number percentage of missing value (which is < 1%) (Table I), the missing value is completely ignored. Little and Rubin (2002) stated that the low percentage of missing value can use the easiest and most common approach method to deal with these missing values, which is to completely ignore the missing values (delete all the missing values) and continue with the complete datasets.

Table 1
Basic descriptions of the monitoring stations and descriptive statistics

Station	Shah Alam	Jerantut
Latitude (°N)	3.08	3.95
Longitude (°E)	101.51	102.36
Valid Data	4369.00	4342.00
% missing	0.32	0.94
Mean (μg/m^3)	56.28	38.63
Median (μg/m^3)	51.57	36.22
Standard Deviation (μg/m^3)	26.15	15.27
Minimum value (μg/m^3)	14.00	10.00
Maximum value (μg/m^3)	587.00	211.00

Methodology

The PM_{10} concentrations were divided into two states: polluted day and non-polluted day. A polluted day can be defined as a day when the PM_{10} concentrations exceed or is equal to a certain threshold value, while the day is considered as a non-polluted day when the PM_{10} concentrations are lesser than the threshold value. According to DoE (2013), the current guidelines used in Malaysia for all the parameters, including PM_{10} concentration, are less relevant and outdated. Thus, this study considers the threshold value based on the World Health Organization (WHO), which is the PM_{10} concentrations should not exceed 50 μg/m^3 for 24 hours as one of the issues discussed and updated in Malaysia Ambient Air Quality Standard 2013 as the new standard in the line with WHO 2006 guidelines (see Table 2). Table 2 was adopted from ADB and CAI-Asia (2006).

Table 2
The Malaysia and WHO 2006 ambient air quality guidelines

Pollutant	Averaging Time	Malaysian Ambient Air Quality Guidelines	WHO (2006)
PM_{10}	24 hrs	150	50
	1 year	50	20

The Markov Chain is intended to be a simple model by requiring only two parameters. The assumptions in this model are that the probability of a polluted (or non-polluted) day depends only on whether it has polluted (or not) on the previous day; and the probability of polluted (or non-polluted) is assumed to be independent of the preceding days.

These probabilities are also known as transitional probabilities, denoted by P_{11} and P_{00} for the sequence of two-polluted days and two-consecutive non-polluted days, respectively. The estimation of P_{11} and P_{00} is directly counting methods from the available daily PM_{10} concentrations record. The formula for first order of Markov is given in equation 1 and 2.

$$P_{11} = P(X_t = 1 \quad X_{t-1} = 1) \tag{1}$$

$$P_{01} = P(X_t = 0 \quad X_{t-1} = 1) \tag{2}$$

The transition probability matrix P, which describes the 2-state Markov chain model, is as in Equation 3 below.

$$P = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix} \tag{3}$$

The polluted days are denoted by 1 and non-polluted days are denoted by 0. Each of the elements in this matrix is defined as:

P_{00}: Probability of a non-polluted day given that the previous day was also a non-polluted day

P_{01}: Probability of a polluted day given that the previous day was a non-polluted day

P_{10}: Probability of a non-polluted day given that the previous day was a polluted day

P_{11}: Probability of a polluted day given that the previous day was also a polluted day

The calculation of each of the elements in the matrix is as follows:

$$P_{01} = \frac{n_{01}}{n_{01} + n_{00}} \rightarrow P_{00} = 1 - P_{01} \tag{4}$$

$$P_{10} = \frac{n_{10}}{n_{10} + n_{11}} \rightarrow P_{11} = 1 - P_{10} \tag{5}$$

n denotes the total number of days for a given state.

The Markov chain properties of the daily occurrence can be tested statistically by checking whether or not the successive events are independent of each other. If the successive events are dependent, the events can form the first-order Markov chain model. The hypotheses are as follows:

H_0: The successive events are independent

H_1: The successive events are dependent

According to Moon et al. (1994), if the successive events are independent, then the statistics, α, is defined by

$$\alpha = 2\sum_{i,j}^{m} n_{ij} \ln(P_{ij} / P_j) \tag{6}$$

Where equation 6 is distributed asymptomatically as χ^2 with $(m - 1)^2$ degree of freedom. Meanwhile, m is the total number of state (in this case: m =2) and the marginal probabilities for j^{th} column of the transition probabilities matrix is presented by P_j.

Besides that, P_{11} and P_{00} can be used to determine the average lengths of the polluted and non-polluted spell, as follows:

The average length of polluted spells is

$$\frac{1}{1 - P_{11}} \tag{7}$$

whereas the average length of non-polluted spells is

$$\frac{1}{1 - P_{00}} \tag{8}$$

Bessons's coefficient of persistence is used to determine whether persistency is commonly used in meteorology. The formula has the following form:

$$R_B = \frac{1 - P_1}{1 - P_{11}} - 1 \tag{9}$$

where P_1 is the probability of a polluted day and P_{11} is the probability of a polluted day, given that the previous day was also a polluted day. The positive value of R_B indicates that the occurrence of a polluted or non-polluted event followed by an immediate preceding event. If the value is zero, then it indicates no persistence for a similar event. Many other researchers such as Deni et al. (2009) and Dahale et al. (1994) have used this particular method.

In fitting the first order of the Markov chain model, the expected distribution closer to the observed distribution of polluted events was analysed. The purpose of fitting this Markov chain model was to check whether the first order obtained was the appropriate fitted order to describe the observed distribution of polluted events for the data set at these two monitoring

stations. The first order of the Markov chain model considers only the condition of the state, either polluted or non-polluted, for one preceding day only. Equation 10 shows the probability of the duration of n days for the polluted events.

$$P_{(n)} = P_{11}^{n-1} \times P_{10} \tag{10}$$

The chi-square goodness of fit test was employed to compare the observed and expected distributions of polluted days. The calculation of the chi-square test statistics is as follows (Skuriat-Olechnowska, 2005):

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{11}$$

Where O_i = observed data

E_i = Expected data

This statistics follows the chi-square distribution, with a number of degree of freedom (d=v-1), where v is the length of polluted days. All the expected frequencies must be at least five days and the hypothesis is that the data follows the observed distribution when $\chi^2 < \chi^2_{1-\alpha,v-1}$ is at 5% significant level.

RESULTS AND DISCUSSION

Over the study period of 12 years, Shah Alam recorded the maximum value of PM_{10} concentration at 587 μg/m^3, while the maximum value of PM_{10} concentrations for Jerantut was 211 μg/m^3 (see Table 1). The absence of this pollutant at both stations is over the threshold value of PM_{10} concentrations for certain days. Besides that, the average daily PM_{10} concentrations for Shah Alam and Jerantut shows that for both stations, the PM_{10} concentrations are higher between day 177 (June) up to day 221 (August), as shown in Figure 2. This result indicates that the present of this pollutant is higher during the South west monsoon (June to September) or dry weather condition (Yusof et al., 2011). This study also shows that Shah Alam recorded the highest number of polluted days (2396 days), whereas Jerantut recorded the highest number of non-polluted days (3552 days).

The frequencies of the first order transitions by considering today's polluted (1) and non-polluted (0) day, followed by yesterday's events (polluted or non-polluted day) for both the stations, are shown in Table 3. Shah Alam recorded the highest frequency of polluted days given that the previous day was also a polluted day, while Jerantut recorded the highest frequency of a non-polluted day, given that the previous day was also a non-polluted day. These results show that Jerantut still has low potential of recording air pollution in Malaysia.

Table 3 also shows results of the transition probabilities for a first order Markov chain obtained directly by using the transition counts using the formula in equation 4 and equation 5 for both stations. Shah Alam and Jerantut recorded a high probability of the next day being polluted when the previous day was polluted (P_{11}) and also recorded a low probability for the next day being not polluted when the previous day was polluted (P_{10}). This finding can help

responsible authorities to take action by providing early warning to the respective population when the previous day is a polluted day and also made prediction of PM_{10} event by referring to the condition of the previous day event.

Table 3
Tally matrix and transition probabilities of the occurrence of polluted and non-polluted days for 12 years at both stations

Station	Tally Matrix		Transition probabilities	
Shah Alam	1472	501	0.7461	0.2539
	502	1894	0.2095	0.7905
Jerantut	3271	281	0.9209	0.0791
	281	513	0.3539	0.6461

The successive state of daily PM_{10} concentrations was checked by computing a statistic of independence, α, and the results are shown in Table 4. The value of α at both the stations are larger than the χ^2 value of 3.841 at the 5 per cent level, with one degree of freedom. Thus, H_0 is rejected and the successive transitions are dependent. Hence, this study concludes that the transitions of daily PM_{10} concentrations occurrence possesses the first-order Markov chain model property, where the successive events are dependent on each other. The analysis of the mean length of polluted and non-polluted days, based on equation 7 and equation 8 for Shah Alam, indicates that on average, 4-5 days for the recurrence of the polluted days, while this is about 3-4 days for the recurrence of the non-polluted day. On average, Jerantut showed 2-3 days for the recurrence of the polluted day and 12-13 days for the recurrence of non-polluted days. Thus, these reflect the fact that Jerantut is relatively a less polluted area compared to Shah Alam.

Table 4
The mean length of polluted and non-polluted days for both stations

Station	Mean		α	Besson's coefficient (R_B)
	Polluted days	Non-polluted days		
Shah Alam	4.77	3.94	660.22	1.61
Jerantut	2.83	12.64	567.92	1.31

Analysis of the persistency of the polluted (non-polluted) events was done by applying the Besson's coefficient of persistence (R_B) as in equation 9. Table 4 shows results of the coefficient and for both stations, the value of R_B was found to be positive, indicating that the occurrence of a polluted (or non-polluted) event was followed by an immediate preceding event. The result shows that the land used background has a strong influence on persistency since the persistent polluted events in Shah Alam were relatively much longer compared to Jerantut.

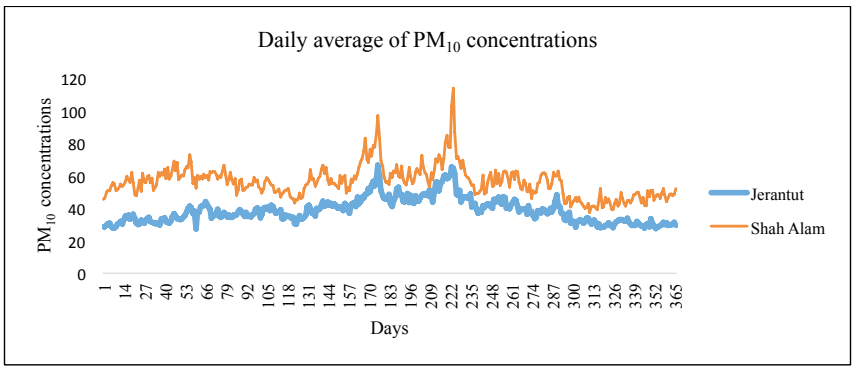


Figure 2. The average daily PM_{10} concentrations in Shah Alam and Jerantut, Malaysia, for a period of twelve years (2002-2013)

Results of the chi-square test at 5% level of significant shown in Table 5 show that the first order of Markov chain model does not a good fitting on the PM_{10} concentrations data at the two monitoring stations. The value of chi-square at both monitoring stations (χ^2) have bigger value than $\chi^2_{0.95,v-1}$ and the hypothesis that the data follow the observed distribution cannot be accepted at 5% significant level.

Table 5
The observed and expected frequency distributions computed with the first order of Markov chain model in Shah Alam and Jerantut monitoring stations

Length of polluted day	Shah Alam		Jerantut	
	Observed (day)	Expected (day)	Observed (day)	Expected(day)
1	209	105	131	99
2	90	83	52	64
3	49	66	44	41
4	24	52	15	27
5	24	41	8	17
6	16	32	7	11
7	10	26	3	7
8	13	20	4	8
9	12	16	16	6
10	8	13	-	-
11	2	10	-	-
12	5	8	-	-
13	1	6	-	-
14	6	9	-	-
15	6	6	-	-
16	13	8	-	-
Total	488	501	280	283
Degree of freedom χ^2 p-value	15.00		8.00	
	173.30		45.32	
	2.20E-16		3.20E-07	

Figure 3 shows the observed and expected frequencies of polluted events for the first order of Markov chain model for the Shah Alam and Jerantut monitoring stations. From the figures, it can be seen that the expected distribution of polluted events obtained does not closely describe the observed distribution. Thus, it can concluded that there is a need to use the higher order Markov chain model in assessing the best or appropriate fitting in describing the distribution of the polluted events.

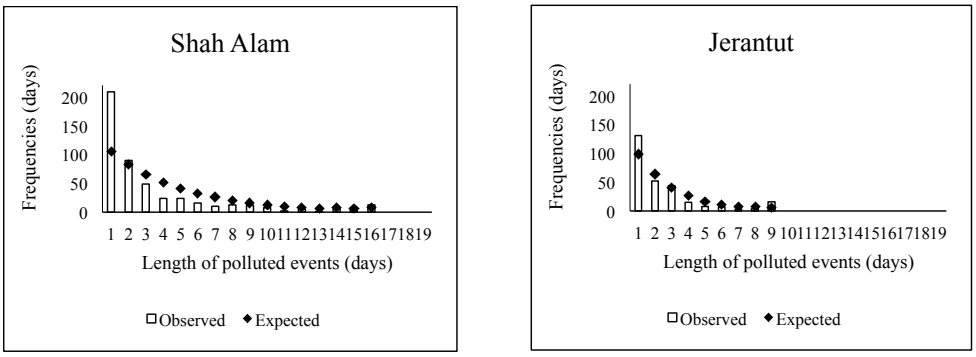


Figure 3. The observed (line) and expected (dot) frequencies of polluted event of the first order for both stations

CONCLUSION

The results obtained from the analysis can provide a better understanding of PM_{10} occurrences and the pattern at these two monitoring stations. In this study, the first order of Markov chain model was applied and the threshold value used was referring to WHO 2006 guidelines ($50\mu gm^{-3}$). The results of the analysis show the high probability that the next day event depends on what has happened on the previous day. Besides that, the analysis of persistency showed that for both stations, the occurrence of polluted (non-polluted) event, followed by an immediate preceding event and the recurrence of polluted day in Shah Alam (4-5 days), is longer compared to Jerantut (2-3 days). By fitting the first order of Markov chain model, the results showed that the higher order of Markov chain model is needed in order to get the best fitted distribution of polluted events at these two monitoring stations. This study can help the responsible authorities to provide early warning to the public so that they can avoid from being exposed to air pollutant when the previous day is a polluted day. Further analysis could be done by considering other monitoring stations in Malaysia and different threshold value of PM_{10} concentrations in order to get more valuable information for the purpose of prediction. Besides that, the higher order (order more than one) could be carried out to find the optimum order that best describes the occurrence of PM_{10} concentrations in Malaysia and thus, a better prediction can be made.

ACKNOWLEDGEMENT

This study was funded by Universiti Teknologi MARA (UiTM), under the Grant 600-RMI/IRAGS 5/3 (36/2015). Special recognition goes to the Department of Environment (DoE) and Alam Sekitar Malaysia Sdn. Bhd. (ASMA) for providing the air quality data for this study. Special appreciation also goes to Universiti Teknologi MARA (UiTM).

REFERENCES

ADB & CAI-Asia. (2006). *Country Synthesis Report on Urban Air Quality Management.* Malaysia: Asian Development Bank (ADB) and the Clean Air Initiative for Asian Cities (CAI-Asia) Centre.

Awang, M., Jaafar, A. B., Abdullah, A. M., Ismail, M., Hassan, M. N., Abdullah, R., & Noor, H. (2000). Air quality in Malaysia: impacts, management issues and future challenges. *Respirology (Carlton, Vic.), 5*(2000), 183-96.

Basu, A. (1971). Fitting of a Markov chain model for daily rainfall data at Calcutta. *Indian Journal of Meteorology and Geophysics, 22*, 67-74.

Berchtold, A., & Raftery, A. (2002). The Mixture Transition Distribution Model for High-Order Markov Chains and Non-Gaussian Time Series. *Statistical Science, 17*(3), 328-356.

Chaloulakou, A., Grivas, G., & Spyrellis, N. (2003). Neural network and multiple regression models for PM10 prediction in Athens: a comparative assessment. *Journal of the Air and Waste Management Association (1995), 53*, 1183-1190. http://doi.org/10.1080/10473289.2003.10466276

Chin, E. H. (1977). Modeling daily precipitation occurrence process with Markov Chain. *Water Resources Research, 13*(6), 949-956.

Dahale, S. D., Panchawagh, N., Singh, S. V., Ranatunge, E. R., & Brikshavana, M. (1994). Persistence in rainfall occurrence over Tropical south-east Asia and equatorial Pacific. *Theoretical and Applied Climatology, 49*(1), 27-39. http://doi.org/10.1007/BF00866286

Deni, S. M., Jemain, A. A., & Ibrahim, K. (2009). Fitting optimum order of Markov chain models for daily rainfall occurrences in Peninsular Malaysia. *Theoretical and Applied Climatology, 97*, 109-121.

DoE. (2013). *Malaysia Environmental Quality Report 2013*. Kuala Lumpur: Department of Environment. Ministry of Natural Resources and Environment, Malaysia.

DoS. (2013). *Compendium of Environment Statistics.* Malaysia: Department of Statistics, Malaysia.

EPA. (2001). *Latest Findings on National Air Quality: 2000 Status and Trends*. United State Environmental Protection Agency.

Gabriel, K. R., & Neumann, J. (1962). A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society, 88*(375), 90-95.

Haan, C. T., Allen, D. M., & Street, J. O. (1976). A Markov Chain Model of daily rainfall. *Water Resources Research, 12*(3), 443-449. http://doi.org/10.1029/WR012i003p00443

Jimoh, O. D., & Webster, P. (1996). The optimum order of a Markov chain model for daily rainfall in Nigeria. *Journal of Hydrology, 185*, 45–69.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., & Kolehmainen, M. (2004). Methods for Imputation of Missing Values in Air Quality Data Sets. *Atmospheric Environment, 38*(18), 2895-2907.

Latif, M. T., Dominick, D., Ahamad, F., Khan, M. F., Juneng, L., Hamzah, F. M., & Nadzir, M. S. M. (2014). Long Term Assessment of Air Quality from A Background Station on the Malaysian Peninsula. *Science of the Total Environment, 482-483*(2), 336-348.

Little, J., & Rubin, D. (2002). *Statistical Analysis with Missing Data*. New York: Wiley.

Mangaraj, A. K., Sahoo, L. N., & Sukla, M. K. (2013). A Markov Chain Analysis of Daily Rainfall. *Journal of Reliability and Statistical Studies, 6*(1), 77-86.

Moon, S. E., Ryoo, S. B., & Kwon, J. G. (1994). A Markov chain model for daily precipitation occurrence in South Korea. *International Journal of Climatology, 14*, 1009-1016.

Paschalidou, A. K., Karakitsios, S., Kleanthous, S., & Kassomenos, P. A. (2011). Forecasting hourly PM10 concentration in Cyprus through artificial neural networks and multiple regression models: Implications to local environmental management. *Environmental Science and Pollution Research, 18*(2), 316-327.

Pires, J. C. M., Martins, F. G., Sousa, S. I. V., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2008). Prediction of the Daily Mean PM_{10} Concentrations Using Linear Models. *American Journal of Environmental Sciences, 4*(5), 445-453.

Rahimi, J., Bazfarshan, J., & Rahimi, A. (2011). Study of Persistence of Days Infected Pollutant Particulate Matter (PM_{10}) in City of Tehran Using Markov Chain Model. In *13th Conference on Atmospheric Chemistry* (p. 231).

Skuriat-Olechnowska, M. (2005). *Statistical inference and hypothesis testing for Markov chains with Interval Censoring*. Delft University of Technology. Retrieved from http://www.ewi.tudelft.nl/fileadmin/Faculteit/EWI/Over_de_faculteit/Afdelingen/Applied_Mathematics/Risico_en_Beslissings_Analyse/Theses/MSkuriat_thesis.pdf

Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., & Hamid, H. A. (2012). Robust Regression Models for Predicting PM_{10} Concentration in an Industrial Area. *International Journal of Engineering and Technology, 2*(3), 364-370.

Yusof, N. F. F. M., Ramli, N. A. R., & Yahaya, A. S. (2011). Extreme Value Distribution for Prediction of Future PM_{10} Exceedences. *International Journal of Environmental Protection, 1*(4), 28-36.

Pertanika J. Sci. & Technol. 26 (1): 379 - 390 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Application of System Dynamic Approach for Family Takaful Product Analysis

Mohamad, A. I.[1]*, Tumin, M. H.[1], Noor, N. L. M.[2], Saman, F. M.[2] and Amin, M. N. M.[1]

[1]Centre for Actuarial Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40000 UiTM, Shah Alam, Selangor, Malaysia
[2]Centre of Information Technology Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40000 UiTM, Shah Alam, Selangor, Malaysia

ABSTRACT

Takaful, the Islamic alternative to conventional insurance, is based on the concept of social solidarity, cooperation and mutual indemnification of losses of members. The 'transparency' offered in the Takaful system will eliminates the elements of gharar (uncertainty), maisir (gambling) and riba (usury). Due to the dynamicity and complexity of cash flows in the Takaful system, the application of system dynamic approach is used in order to discover any possible internal and external impacts in the assumptions used in determining contributions rate from the participants. The traditional approach, which is the deterministic approach, has limitations where changes of the actual experience may cause operators to stop issuing the contract or product. Using system dynamic, these possible effects from the actual experience can be determined in terms of the amounts transferred to shareholder's fund and results obtained can assist the management to decide which assumptions to be used so that the operators will continue solvent and making profit at the same time. The results of System Dynamic simulation analysis in this paper represent the impacts of component changes in the Takaful model. The results can be used as decision tools for the Takaful operators to determine the best assumptions and strategies in order to maximise their profits.

Keywords: Actuarial Science, System Dynamic, Insurance, Takaful

ARTICLE INFO

Article history:
Received: 03 March 2017
Accepted: 28 September 2017

E-mail addresses:
arwin@tmsk.uitm.edu.my (Mohamad, A.),
hanafi@tmsk.uitm.edu.my (Tumin, M. H.),
norlaila@tmsk.uitm.edu.my (Noor, N. L. M.),
fauzi@tmsk.uitm.edu.my (Saman, F. M.),
nazrul@tmsk.uitm.edu.my (Amin, M. N. M.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Takaful, the Islamic alternative to conventional insurance, is based on the concepts of social solidarity, cooperation and mutual indemnification of losses of members. It is a pact among a group of persons who agree to jointly indemnify the loss or damage that may inflict upon any of them, out of the fund they donate collectively.

This fund is being managed by the Takaful operator. It is widely understood that the participants consider each other as a "family member" and that each and every one pools together in times of loss. In the event of a loss, the operator will pay out the specified amount to the respective participants. Should there be any surplus, it will be distributed to all its participants. This proves that the Takaful operator only works as a fund manager, or a protection and profit sharing venture between the operator and participants.

At the moment, there are fifteen Takaful operators in Malaysia, consisting of two retakaful operators, five foreign Takaful operators, and eight local Takaful operators. Since the endorsement of the Takaful Act in 1984, the Takaful industry has increased significantly, with an annual growth of 27% in assets and a net contribution increase of 19%. All these were evident from the years 2003 to 2007. This rapid change in the Takaful industry has made the Takaful operators to re-evaluate and study the product features to make it more competitive in the market.

One of the tools available to re-evaluate and study the Takaful product features and assumptions is the system dynamic approach. System dynamics is a computer-based model that specially designed to solve looping cases. A dynamical system is a function with an attitude that tends to do the same thing over and over again and it can estimate the next steps well. This system dynamic model tells us how the behaviour of the system results from the interactions of its component parts. This model combines both qualitative and quantitative aspects and aims to enhance the understanding of the complex systems to gain insights into the system behaviour (Srijariya, Riewpaiboon, & Chaikledkaew, 2008).

The research focuses on the applicability of the system dynamic in order to discover the possible impact in the assumptions used in determining contributions rate from the contract holder. The operation of the cash flows from the participants within the Takaful model or system is a dynamic and complex system due to many inter-related factors such as management expenses, investment earns, death claims, withdrawal claims, economic factors and also competitions among other Takaful operators. By using the traditional pricing methods in the Family Takaful product (premium for life insurance), decision makers for Takaful product development who may be the appointed actuary or members from the top management of the Takaful operator need to do more study on each variable or factor to gain a better understanding of the impacts of the pricing and operational strategy.

The scenario now is that the components in the Takaful model have been changing due to the impacts of economic condition of the country, as well as the policy intervention of Malaysian central bank, Bank Negara Malaysia (BNM). As a result, more Takaful licenses are issued to the industry and with the entrance of new competitors into the industry, the level of industry competitiveness has increased as well. The consequences of those changes could hurt the Takaful operators where the possibility of a large-scale loss to the Takaful operators may occur when issuing a new Takaful product.

The traditional pricing method approved by the appointed actuary of the Takaful operators may be sufficient to help the Takaful operator to remain solvent. However, any other changes that arise in the industry would give an impact to the product, perhaps benefiting the customers but may not be beneficial to the Takaful operator. Since the word 'Takaful' represents the

Islamic insurance system, it is crucial for the Takaful operators to able to produce a product that is fair and profitable for both parties, namely the Takaful customers and Takaful operators.

The perspective of the Takaful system has never been viewed from a dynamic system perspective. The study provides an opportunity to explore and exploit the dynamicity in the Takaful system using the system dynamic approach. A system dynamic model can be used to visualise any possible impact arising from the changes in economic condition, policy, behaviour of the policy holders and other changes that could give significant impacts to the profitability of the products of the Takaful operators. However, the traditional methods are not able to view this dynamicity and thus limits the capability of the Takaful operator to gain a good understanding of the bigger picture of the opportunity to perform Takaful product pricing. It seems that system dynamics can be used as an alternative approach to visualise all impacts arising from the changes of each component in the Takaful system and will act as decision support information for the pricing of Takaful products.

Some examples of possible changes that may give impacts to the Takaful model are increases in death claims, increases in withdrawal claims when the participants switch to another Takaful operator, unfavourable economic condition that will result in lower investment earnings from the expected and also high management expenses due to inflation. It is difficult to get a clear picture of the combined effects of the components due to the complexity and dynamicity of the Takaful system. The detailed Takaful model will be presented in the subsequent section. Each component of the model has a complex interconnection and relationship with other external factors and it can be further enhanced by using the system dynamic approach. In this study, system dynamic is used to model the components, factors and units in the Takaful model.

This study focuses on applying the system dynamic model in the Takaful model as an extension of the traditional approach of pricing. In general, this study provides an extra step to re-evaluate the pricing of family Takaful product and ensure sustainability and profitability of the product in the market. On a more specific term, this study contributes towards offering the decision makers of Takaful operators the opportunity to explore new methods in making better decisions within their scope of work. These decisions made may be crucial in order to meet the objectives of the Takaful operators. The simulation model developed from system dynamic approach can provide insight or visualisation into the impacts of any changes in the component of the model, which can be the translated into decisions that should be done by the Takaful operators.

Introduction to System Dynamic

System Dynamics is a computer-based model especially to solve looping cases. A dynamical system is a function with an attitude that tends to do the same thing over and over again and can estimate the next steps well. This system dynamic model tells us how the behaviour of the system results from the interactions of its component parts. This model combines both qualitative and quantitative aspects and aims to enhance the understanding of complex systems to gain insights into system behaviour (Srijariya, Riewpaiboon, & Chaikledkaew, 2008).

System Dynamic is applicable because the methodology can be used in solving real world crisis, which is depending upon reaction structure (Grosser, 2005). This methodology is also capable of capturing the effect of hypothetical assumption. It includes concepts, variable interdependency and dynamic systems with powerful shifting effects which can be used to analyse large and complex real-world situations that cannot be solved through conventional quantitative analysis. Furthermore, System Dynamic can be used to determine important variables for both long-term and short-term effects. Rational and logical future behaviour also can be drawn but it is not used as forecasting method. Other than that, its simulation-based theory can helps us to evaluate simulation-based scenarios, in order to find the best assumptions to price the contributions of a Takaful product.

In this paper, the System Dynamic model was modelled based on two main scopes; which are the internal and external factors of the component in the Takaful system.

System Dynamic Advantages

Sterman (2006) mentioned that System Dynamics is a method to enhance learning in complex systems. The ability to model diverse relationships between elements of a particular system and how these relationships influence the behaviour of the system over time has made it a better model compared to the traditional statistic approach. System Dynamics model educates us by identifying the inconsistency and allows one to see how the identified complex interactions work when all are active at one time (Taylor & Lane, 1998; Sapiri, Kamil, Mat Tahar, & Tumin, 2010).

Systems Dynamics provides a general communication instrument connecting many academic disciplines (Martin, 1997). The perspective of System Dynamics, with its emphasis on feedback, changes over time and the role of information delays actually notify the policy makers the projected and unintended alternative consequences. In addition, System Dynamics was used to design and test policies intended at altering a corporation's arrangement so that its behaviour would develop and become more (Radzicki, 2011). One feature that is common to all systems is that a system's structure determines its behaviour.

System Dynamics also can be used to analyse how structural changes in one part of a system affects the behaviour of the system as a whole. System Dynamics was initially based on the feedback control theory which consists of both quantitative (hard) and qualitative (soft) approaches in examining the dynamic behaviours of the development and changes of a system (Maani & Cavana, 2000). This model also seeks to predict dynamic implications of policy, not forecasting the outcome at a given time in future.

This theory is supported by Hirsch (1979). Grosser (2005) and Sterman (2006) stated that System Dynamics models can be used to draw reasonable future behaviour, depending upon the structure of the system. This will help policymakers and planners to have a better understanding of the relative impacts of alternative interventions they might be taking into account and anticipate unintended by-products of new programmes and policies. These models are not intended to serve as forecasting method. The policymakers will make decision based on the historical data and performance goals as the baselines to determine whether a particular policy generates behaviour of the main variables that is better compared to the baseline or even other policies.

The strength of System Dynamics model is the ability to justify own assumption of the system faced (Olaya, 2011). The wonderful advantage of the System Dynamics model is that its parameters can be accustomed by policy makers based on the changing situation or to construct visualised alternatives in planning (Udompanich, 1997). Furthermore, System Dynamics acknowledges the critical role of an individual and organisational core model, which means, it does not plainly model them. In other words, it only utilises factual data. System Dynamics is not about point prediction, the behaviour characteristics can evaluate whether or not the model can meet the conditions and behaviour leading to the event (Grosser, 2005).

System Dynamics is more effective than others in changing the thinking and actions of the audience. This is because it consists of stock and flow diagram and also the ability to track movements of people and things in a clear and systematic way. This methodology is also useful in providing qualitative system behaviour, whereas discrete event simulation is superior in revealing detailed features related to discrete queue dynamics. In terms of data requirements, System Dynamics is typically very much less as System Dynamics models are usually of higher-level and more aggregated. More useful modelling can be done with less data. Some healthcare settings are rich in useful data. Other than that, System Dynamics approach can include macroeconomics variable, behavioural variables and application variable which are not applicable in the traditional approach.

In this paper, System Dynamics approach will be tested for its applicability in the Takaful System. For this purpose, a few macroeconomic variables or economic conditions (e.g., inflation rate, favourable economic condition and Takaful operator's competitors) are tested. These variables are proven to be significant based on the secondary sources. Thus, it was included in this research, along with the basic assumptions of contribution calculation by using the traditional approach.

DESIGN APPROACH

Takaful Model

For this paper, the hybrid model of Takaful, that combines the model of Wakalah and Mudharabah, was selected because it appeared to be more well-accepted and favourable than other models, apart from being widely adopted by many newly-established Takaful operators (Islamic Financial Services Act 2013, n.d).

The main reason for using the system dynamic approach is to visualise the set of assumptions used in order to find the optimum contribution rate, which is fair to the participants and profitable to the Takaful operators. The main problem with the traditional approach is that it will take longer time to test each set of the assumptions. Moreover, the traditional approach will not consider any changes in the components after the product is launched. Using the system dynamic, the set of assumptions will be used as the initial values and it will be tested by changing the components in the system to see whether it will give positive impact towards profitability, or vice versa.

The model is detailed out in Figure 1 below:

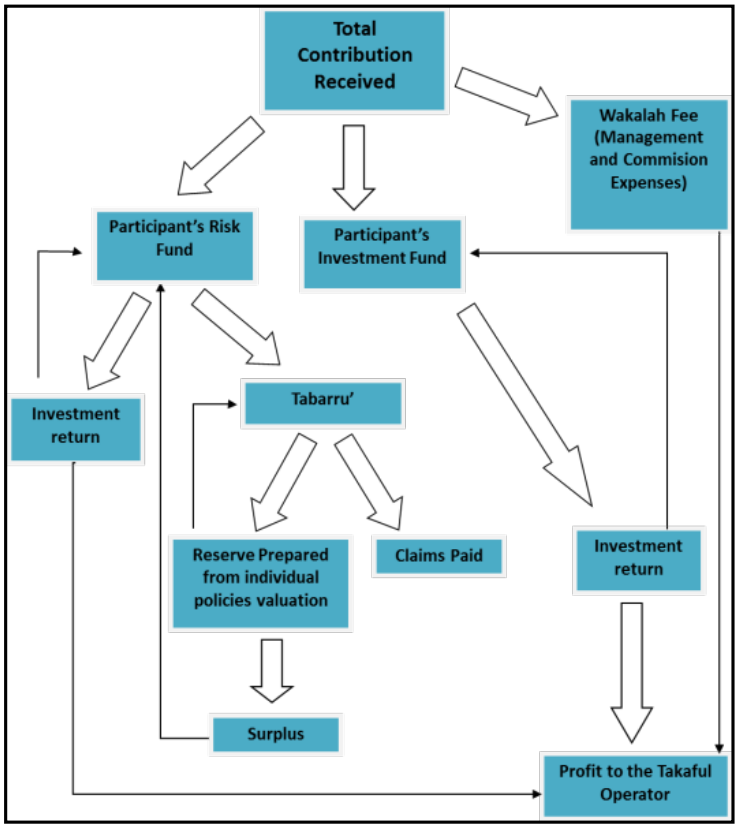


Figure 1. A detailed workflow of the Takaful System

All of these elements in the Takaful system must be properly determined in order for the Takaful operator to remain solvent and continue to make profit. Elements such as participants to Takaful operator's agreed ratio, contributions rate, commission and administration expenses, rate for investment return, reserve assumptions, as well as claims and withdrawal rate have to be properly assumed and calculated.

In this research, the system dynamic model is developed by first exploring the internal factors or components in the model, followed by combining them with the external components which involve uncertainties. The effects of each component on the others, including both internal and external, are examined separately.

Development of the Causal Loop Diagram

The involvement of investment activities, influence of the rapid growth of Takaful operators, and effects of inflation in the expenses may also cause significant effects on the Takaful operator's profitability. There is no use of issuing a contract or policy that will burden any parties involved, be it the operators or contract holders.

In this study, for the purpose of demonstrating the system dynamic approach, some of the economic variables and external components having relationship with internal components were used. Hypothetically, the relationship of the economic variables and the internal components can be best described as in Table 1 below:

Table 1
Hypothetical relationship between changes in economic variables and vhanges in internal components of the model

Changes in economic variables	Changes in internal components
Unfavourable market condition	Increase in actual expenses
(Increase in inflation rate)	Decrease in investment performance
(Decrease in overall national interest rates)	Decrease in number of participants
	Increase in number of withdrawals
Increase in number of claims and withdrawals	Increase in reserving assumptions
	Decrease in overall investment performance
Increase in competitiveness in the market	Increase in the number of participants
	Increase in total contribution received

The relationship of both external components (including economic variables) and internal components is best described in the causal loop diagram below. It is important to determine any real positive and negative relationships in order to have a system dynamic model that will represent the real situation in the simulation result.

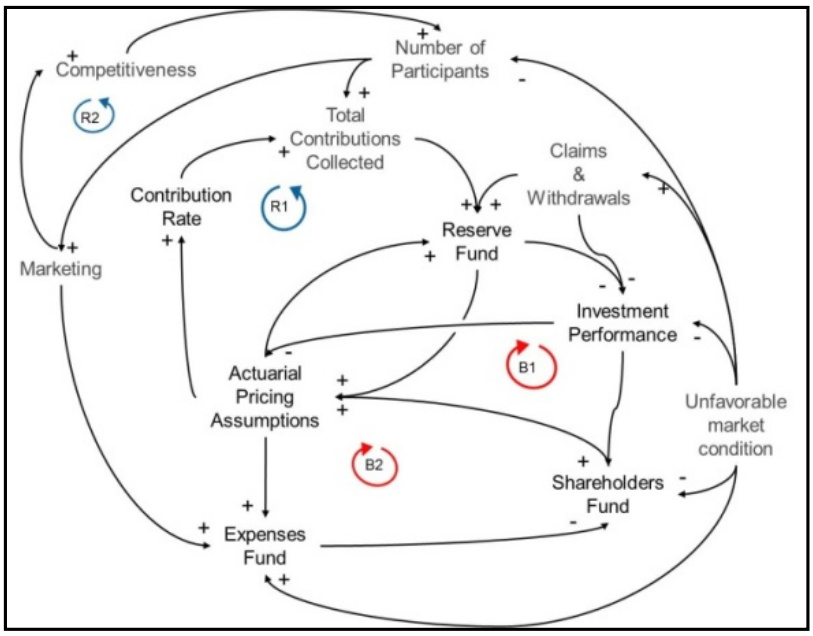


Figure 2. CLD for both the internal and external components

The system dynamic model proposed in this paper will assist the management to increase any factors that will lead to increases in cash inflow in terms of total contribution collected. The operators can make changes whether to have an improvement in marketing that also will lead to increases in the expenses by the operators. This improvement in marketing will also enhance the competitive element of the product compared to the other operators, and hence, will increase the number of participants for that particular product.

Figure 2 contains several balancing and reinforcement loops. The first balancing loop links between the actuarial pricing assumptions, reserve fund prepared, investment performance of the operators and shareholders' fund. Increase in Actuarial pricing assumptions will increase the reserve fund needs to be prepared by the operators. This will result in lesser fund provided for the investment where the investment for the reserve fund needs to be more prudent to meet the obligations towards the participants. Lesser fund for investment will lead to lesser amount to be transferred to the shareholders' fund as a profit sharing between the participants and the Takaful operators. The shareholders fund's income depends entirely on the investment performance of the operators, where higher provision required for the investment will result in the more amounts to be transferred to the shareholders' fund.

The second balancing loop links between the actuarial pricing assumptions, expenses fund prepared and shareholders' fund. Prudent assumptions in the actuarial pricing assumptions will lead the operators to prepare more funds in the expenses fund, and this will result in lesser amount to be transferred to the shareholders' fund. Motivated by maximising returns by the operators to continue operates, the shareholders of the operators will somehow force the management to reduce the pricing assumptions in order to provide more funds to be transferred to the shareholders' fund.

Proposed Stock and Flow Diagram for Takaful System

Two major characteristics, which are mainly provided by the system dynamic approach, are changes over time and feedback-transmission and receipt of information. In the system dynamic model, stock and flows are used to model the flow of work and resources through a project. Stocks represent the stored quantities and characterise the state of the system and generate the information upon which decisions are based. Flows are the rate of increase or decrease in stocks.

In this study, the unit used for time is in a yearly basis. This is because most of the components in the system happen on a yearly basis such as investment earned, total expenses for the year, total contribution received for the year, yearly claims rate and withdrawal rate, surplus, inflation rate, etc.

Based on the earlier causal loop diagram, the final simulation model was developed to enable more investigation on the changes of components by using simulation. Different decisions by the operators will be simulated by this model to evaluate the outcomes and visualise the impacts without having to conduct real experiments. The developed system dynamic model is shown in Figure 3.

In this model, reserve fund is prepared in the Participant Risk Fund (PRF) and the fund used for investment purposes is the Participant's Investment Fund (PIF), as suggested in the BNM Guidelines (2010) on Takaful Operator Framework. The investment portion (in terms

of percentage) from the Total contribution collected is a balance resulting from the deduction into Expenses Fund from the wakalah fee and also deduction into PRF from the tabarru rate.

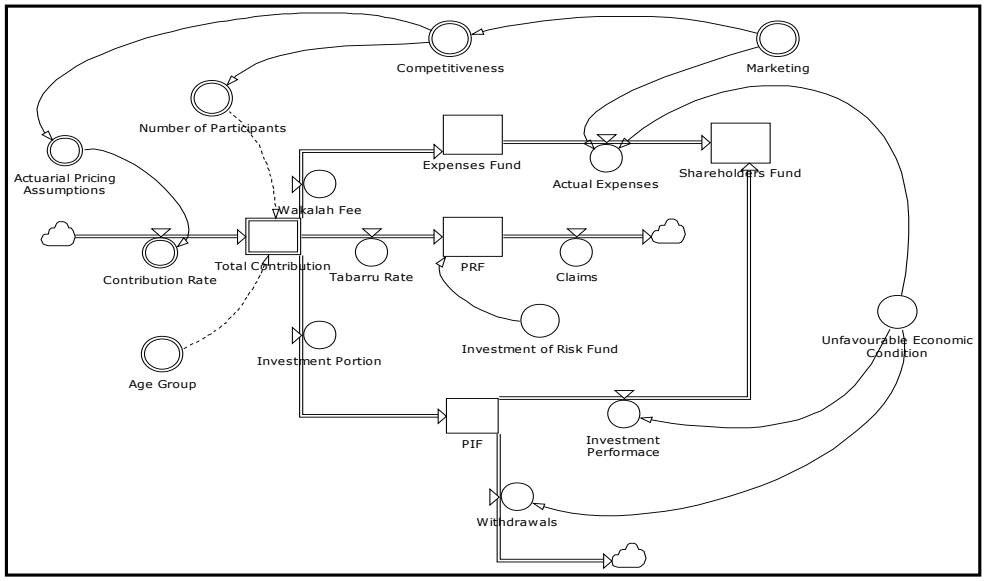


Figure 3. The proposed stock and flow diagram for the Takaful System

RESULTS AND DISCUSSION

Simulation Analysis of the Proposed System Dynamic Model

The simulation test combines two external variables which can lead to loss for the Takaful operator. These variables are the inflation rate and competition of the market. Increase in competition in the market, combined with the increase in inflation, will supposedly result in a major loss for the Takaful operator.

For example, through economic crisis, there would be increase in the inflation rate that would result in increases in the operational cost and expenses. Furthermore, during this time, there will be lesser participants who are willing to participate in buying the Takaful products. This analysis will test the point which will be the bench mark for the decision makers to be prepared on the worst case scenario.

Inflation rate, combined with the competitive market, are two variables that cannot be controlled by decision makers. However, this analysis can give a warning alarm at which point these combined variables would affect the profitability of the Takaful operator. Increase in the competitive market will result in fewer participants coming in to buy the product. On top of that, the inflation rate will increase the expenses occur each year.

The table below shows the input used in the simulation for this analysis:

Table 2
Input for the simulation model of unfavourable market condition

Changes in Component	Profit 1	Profit 2	Profit 3	Profit 4	Profit 5
Total Number of Policies	2500	2500	2500	2500	2500
% increase of number of policies every year	1%	-1%	-2%	-10%	-10%
Inflation rate	4%	4%	4%	0%	4%

The simulation result based on the inputs from Table 2 can be referred in Figure 4.

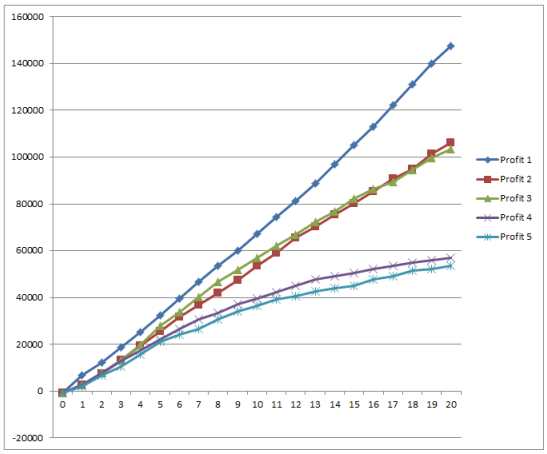


Figure 4. Inflation and competitive market vs. profit

The results above show that even when the number of policies dropped to 10% every year, and the inflation rate increased to 4% every year, the Takaful operator will still be making positive profit throughout the year. The expected outcome from this analysis was that the Takaful operator would result in loss when these two economic factors (inflation rate and competition) are combined in the market. This result of the analysis somewhat contradicts with the earlier expected outcome stated in the earlier chapter. However, it gives a very valuable insight into the current market situation which can lead to the Takaful operators' profitability. Based on this particular analysis, it is highly unlikely for a Takaful operator to result in loss or insolvency due to unfavourable market condition.

As stated earlier on, the Takaful operator will still be making profit throughout the year even though both the variables of the inflation rate and the competitive market are combined. Nonetheless, small changes to both the components will not give a big impact on profitability. However, the changes do lower any potential profit in the long run. In this Takaful system, it is clear that the decision maker can apply the system dynamic approach to help them make better decision to meet the objectives of the Takaful operator.

CONCLUSION

As a conclusion for the analysis, the results of profit differ for each variation in the Takaful model components. However, the Takaful operators will still continue to make a positive profit in the long run, regardless of the assumptions used in the pricing with the condition in Malaysia.

All of the visualisation results from the system dynamic simulation model can be used to help the Takaful operator plan for any future product to be launched and to be prepared for any changes that may happen around the industry. In this Takaful system, it is clear that the decision maker can apply the system dynamic approach to help them make better decision to meet the objectives of the Takaful operator.

This system dynamic model has also been proven to be used in visualising the impacts of changes to each component. Every small or big change can be tested using the system dynamic model so as to help determine the point at risk where this point can lead to insolvency, and also loss, from the product launched.

REFERENCES

BNM Guidelines. (2010). *Concept Paper – Guidelines on Takaful Operational Framework*. Bank Negara Malaysia, Malaysia.

Grosser, S. (2005). Modelling the Health Insurance System of Germany: A System Dynamics Approach. In *23rd International Conference of the System Dynamics Society*. Boston (MA), USA.

Hirsch, G. B. (1979). System Dynamics Modelling in Health Care. *ACM SIGSIM Simulation Digest, 10*(4), 38-42.

Maani, K. E., & Cavana, R. Y. (2000). *Systems Thinking and Modelling: Understanding Change and Complexity.* NZ: Pearson Education Press.

Martin, L. A. (July 1997). *The First Step.* Massachusetts Institute of Technology, USA.

Olaya, C. (2011). System Dynamics Philosophical Background and Underpinnings. In *Complex Systems in Finance and Econometrics* (pp. 812-832). Springer New York.

Radzicki, M. J. (2011). System Dynamics and Its Contribution to Economics and Economic Modelling. In *Complex Systems in Finance and Econometric*s (pp. 727-737). Springer New York.

Sapiri, H., Kamil, A. A., Mat Tahar, R., & Tumin, H. (2010). Dynamics Simulation Approach in Analyzing Pension Expenditure. *International Journal of Mathematical. Computational Science and Engineering, 4*(10), 1-7.

Srijariya, W., Riewpaiboon, A., & Chaikledkaew, U. (2008). System Dynamics modelling: An alternative method for budgeting. *Value in Health, 11*(1), S115-S123.

Sterman, J. D. (2006). Learning from evidence in a complex world. *American Journal of Public Health, 96*(3), 505-514.

Taylor, K., & Lane, D. (1998). Simulation applied to health services: opportunities for applying the system dynamics approach. *Journal of Health Services Research and Policy, 3*(4), 226-232.

Udompanich, S. (1997). System Dynamics Model in Estimating Manpower Needs in Dental Public Health. *Human Resources Development Journal, 1*(1), 35-47.

Pertanika J. Sci. & Technol. 26 (1): 391 - 408 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Monodispersed and Size-controllable Potassium Silicate Nanoparticles from Rice Straw Waste Produced Using a Flame-assisted Spray Pyrolysis

Asep Bayu Dani Nandiyanto[1]*, Rena Zaen[1], Rosi Oktiani[1], Ade Gafar Abdullah[2] and Ari Arifin Danuwijaya[3]

[1]Department of Chemistry, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia
[2]Department of Electro Technique, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia
[3]Department of English, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154, Indonesia

ABSTRACT

The purpose of this study was to demonstrate a facile method to produce monodispersed and size-controllable potassium silicate nanoparticles from rice straw waste. Different from other methods that use expensive raw chemicals, our method utilises rice straw waste as a source of silica that is cost-free and largely available. In the experimental procedure, rice straw waste was burned. Then, the burned rice straw waste was put into the alkaline extraction process (using potassium solution) and flame-assisted spray-pyrolysis apparatus system. To support the flame-assisted spray-pyrolysis, we utilised commercially available liquid petroleum gas as a fuel source for the flame combustion process. Experimental results showed that the prepared particles are monodispersed, spherical, and having sizes in the range of nanometer (from 20 to 80 nm). The fuel flow rate plays important role in controlling particle size in the final product. Increases in the fuel lead to the formation of larger particles. Since the present method can convert rice straw into useful and valuable potassium silica particles, further development of this study would give a positive impact for the reduction of rice straw waste emission.

Keywords: Aerosol process, controllable size,flame spray pyrolysis, growth particle, nucleation, potassium silicate, silica nanoparticles, spray drying

ARTICLE INFO

Article history:
Received: 25 April 2017
Accepted: 28 November 2017

E-mail addresses:
nandiyanto@upi.edu (Asep Bayu Dani Nandiyanto),
rena.zaen95@student.upi.edu (Rena Zaen),
rosioktiani@student.upi.edu (Rosi Oktiani),
ade_gaffar@upi.edu (Ade Gafar Abdullah),
aridanuwijaya@upi.edu (Ari Arifin Danuwijaya)
*Corresponding Author

INTRODUCTION

Potassium silicate is an attractive material used in a wide range of applications, specifically for fertiliser (Kikuchi, 1999). This material is also widely used due to its high stability, chemical

ISSN: 0128-7680

flexibility, and biocompatibility (Tokunaga, 1991). For example, as fertiliser, this material is effective for increasing the amounts of sugar and amino acid in plants, making them resistant to insects, pests, and diseases (Kikuchi, 1996). Potassium silicate is also used for improving quality of fruit in terms of taste, shape, and colour (Tokunaga, 1991).

To produce potassium silicate, many methods have been reported by several research teams (see Table 1) (Beneke & Lagaly, 1989; Rastsvetaeva, Aksenov, & Taroev, 2010; Novotny, Hoff, & Schuertz, 1993; Baghramyan, Sargsyan, & Harutyunyan, 2016; Tokunaga, 1991; Kikuchi, 1999). The method typically employed direct interaction and reaction between "silica source" and "potassium hydroxide". Potassium hydroxide acts as a potassium source and an agent for dissolving silica component from its originated source. Then, the dissolved silica component (typically as silicic acid ($Si(OH)_4$)) incorporates with potassium element to create crystal/ composite potassium silicate.

Table 1
Current reports on the production of potassium silicate material

Method	Materials	Result	References
Chemical treatment	Fly ash and KOH	Powder	Tokunaga (1991)
	SiO_2 and K_2O	Solution	Santmyers (1957)
	Waste bottle and KOH	Solution	Mori (2003)
	SiO_2 and KOH	Hydrated micropowder	Beneke & Lagaly (1989)
	Coal ash and KOH	Powder	Kikuchi (1999)
Hydrothermal reaction	Quartz sand and KOH	Solution	Novotny, Hoff, & Schuertz (1993)
	Ground diatomite and KOH (with microwave process)	Glass	Baghramyan, Sargsyan, & Harutyunyan (2016)

Beneke and Lagaly (1989) reported the production process using commercially available silica. Their process is effective to form potassium silicate particles with sizes in micrometer range. However, their particle sizes are broad and there is no information regarding control of particle size. Successful method using commercially available silica has been reported by Rastsvetaeva et al. (2010). Their method is effective to create unique crystal. However, their method is complicated as it utilises extreme process condition [e.g., high pressure (up to 1000 atm)] and time-consuming process. In addition, both the methods employed commercially available silica, which is impractical for practical uses since the price of these chemicals is typically expensive. As a consequence, the selling price of potassium silicate products is uncompetitive against other available products.

To against the use of above raw materials, Novotny et al. (1993) used silica quartz sand in the hydrothermal method. Mori (2003) reported the use of waste bottle, which is contacted with potassium hydroxide. Bagramyan et al. (2016) reported the production of potassium silicate from dissolving ground diatomite rock. They utilised microwave treatment to assist the heat treatment process. However, both methods are limited to the production of potassium silicate solution. There is no information on the possibility process for generating powder. Other papers

have been reported by Tokunaga (1991) and Kikuchi (1999), who used fly ash and coal ash as the silica sources, respectively. Although both the methods are effective for the production of potassium silicate particles, they did not report in detail about the physicochemical properties and characteristics of potassium silicate product.

Based on the above methods, several problems still persisted. Specifically, to the best of our knowledge, there is no information on how to create monodispersed and size-controllable potassium silicate nanoparticles in existence. Nanoparticles are required because these types of materials give better performance. The physicochemical properties of nanomaterial are different from that of bulk material. Control of particle size is also crucial since the size determines the material performance (Nandiyanto & Okuyama, 2011).

In our previous report, silica particles were successfully produced from rice straw waste (Nandiyanto et al., 2016b). In short, we used a combination of alkali and acid treatment for extracting and isolating silica from its originated source. Spherical and pure silica with sizes in the submicrometer range were produced. Although the combination of alkali and acid treatment is effective to create submicrometer silica particles, the production of potassium-embeded silica particles is typically difficult. Potassium and other inorganic components leach during the washing process (that is conducted in each step). This is the fundamental reason for why all papers related to such combination of alkali and acid treatment always created pure silica material (Kalapathy, Proctor, & Shultz, 2000; Zaky et al., 2008).

The purpose of this study was to demonstrate a facile technique to produce monodispersed and size-controllable potassium silicate nanoparticles from rice straw waste. Rice straw was selected because this material contains high amount of silica. Rice straw waste is also one of the largest agricultural wastes produced in large quantities but has not been used commercially (Fadhlulloh, Rahman, Nandiyanto, & Mudzakir, 2014; Permatasari, Sucahya, & Nandiyanto, 2016).

In the experimental procedure, rice straw waste was burned. Then, the burned rice straw waste was put into the alkaline extraction process (using potassium solution) and flame-assisted spray-pyrolysis apparatus system. While other methods utilise the acid precipitation method to precipitate silica-related material, our method uses flame-assisted spray-pyrolysis technique that is effective to generate particles with sizes in the nanometer range (Nandiyanto & Okuyama, 2011). To support the flame-assisted spray-pyrolysis method, we utilised commercially available liquid petroleum gas (LPG) as a fuel source for the flame burner as it is largely available. Therefore, the use of this LPG gives possibility for further scaling up process (Nandiyanto, Fadhlulloh, Rahman, & Mudzakir, 2016a).

Direct application of the flame-assisted spray-pyrolysis method to the alkaline-treated solution is also important for maintaining the existence of potassium in the final product. The flame-assisted spray-pyrolysis method is also prospective to produce agglomerate-free and pure particles in just a few of seconds (Karthikeyan et al., 1997). Further, this method has many advantages, including its low-cost production, easy-to-control particle size, simple processing, high-production yield, and ease of conversion to mass manufacturing (Nandiyanto & Okuyama, 2011).

Experimental results showed that the prepared particles are spherical with the sizes from 20 to 80 nm. We also varied the fuel flow rate in the pyrolysis process, which is effective to

control the particle size. In addition, since the product is generated from the rice straw waste which has been underutilised till the present day, further development of the present study would have a great positive impact for the reduction of rice straw waste emission and conversion of waste into useful and valuable products.

BASIC INFORMATION ON RICE STRAW AND FLAME-ASSISTED SPRAY-PYROLYSIS METHOD

What Happened with the Rice Straw?

Rice straw in the agricultural country raises environmental problems (van Soest, 2006). For example, Indonesian Statistic Bureau in 2015 recorded that the rice production was 75.36 million tons. The production of rice increases every year (see https://www.bps.go.id/brs/view/id/1271; accessed on 24 December 2016). This huge amount remains for waste problems that are about 12-15 tons of rice straw as a by-product. These abundant by-products are typically disposed of and burned in the rice field, contributing severe air pollution issues.

Rice straw is rich in silica, organic, and inorganic nutrients (see Table 2) (Minu, Jiby, & Kishore, 2012). The organic nutrients in the rice straw mainly contain cellulose (32-47%), hemicellulose (19-27%), lignin (5-24%), and ash (13-20%), (Zaky et al., 2008) whereas the inorganic component is mostly alkali and earth alkali. As the rice straw contains high amount of silica (specifically, rice straw ash can contain 60-80 wt% of silica) (Nandiyanto et al., 2016b), the price is low and the existence of this waste is abundant, and rice straw can be used as an alternative source for silica.

How to Isolate Silica from Rice Straw?

Figure 1 shows illustration mechanism for the isolation of silica from rice straw waste, in which the process requires at least three steps. The steps include the burning process, alkali treatment, and conversion of silicic acid into silica.

The first step is a burning process. Since most of the rice straw components are organic-related compounds (Zaky et al., 2008), burning rice straw allows increases in surface area and decreases in volume. Most of the organic components turn into carbon dioxide gas and some are turned into ash. Thus, the burning process plays an important role for enriching the percentage of silica.

Next, the second step is the alkali treatment that is used for converting silica component (embedded in the ash) into silicic acid ($Si(OH)_4$). Silicic acid is water-dissolved chemical. The solubility of silicic acid increases in the pH of higher than 10 (Kalapathy, Proctor, & Shultz, 2000). The optimum pH to get excellent extraction process is 12-13 (Minu, Jiby, & Kishore, 2012). In general, the alkali treatment involves the following reaction:

$$Si\text{-}O\text{-}Si + H_2O \rightarrow 2Si\text{-}OH$$

$$Si\text{-}OH + M^+ + OH^- \rightarrow Si\text{-}O\text{-}M^+ + H_2O$$

where M is the alkali component such as sodium and potassium.

The third step is the conversion process from silicic acid into silica (SiO_2). Many techniques can be applied for achieving this conversion process such as acid and heat treatment. In this study, although many papers have reported the possibility of the conversion process using acid precipitation process (Lu & Hsieh, 2012), we limited our process using the heat treatment. The heat treatment is selected because this method is better than the acid treatment for maintaining the inorganic element (potassium and other inorganic components usually leach during the washing process in the acid treatment).

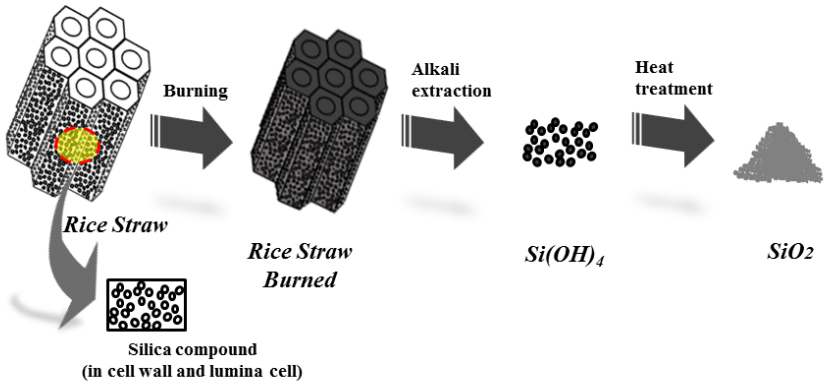


Figure 1. Illustration mechanism of the isolation of silica from the rice straw waste

Technical Information about Flame-assisted Spray-pyrolysis Method

Flame-assisted spray-pyrolysis method is a method for generating particles from liquid-phase solution via employment of the heat-shock treatment (Nandiyanto & Okuyama 2011). The production process involves generation of droplet from the initial solution. Next, the generated droplets are introduced to heat for evaporating solvent in the droplet (Chiang, Aroh, & Ehrman, 2012). Since heat-shock treatment is applied in the evaporation process, the final sizes of the produced particles are in the range of nano to submicrometer size (Nandiyanto & Okuyama, 2011).

Detailed information on the mechanism during the flame-assisted spray-pyrolysis method is shown in Figure 2. In general, the process (shown in Figure 2a) is started by generating droplets from liquid precursor. The generated droplets are then introduced into the reactor and put in direct contact with the flame. The flame itself is produced from the combustion of fuel with air. During the droplet travel, several steps occur (see Figure 2b), including solvent evaporation, droplet break-up, nucleation and growth process. The steps occur subsequently to form nanoparticles. In addition, the main important step for the successful production of nanoparticles is the droplet break-up step. This step can be achieved only due to the existence of heat-shock (that is achieved from the direct contact of droplet with the flame).

In addition, the flame process is conducted in the reactor system, as shown in Figure 2a. Thus, the heat in the process is typically maintained. The transfer of heat in the reactor is due to the heat radiation and convection that are gained from the combustion process of fuel and oxygen.

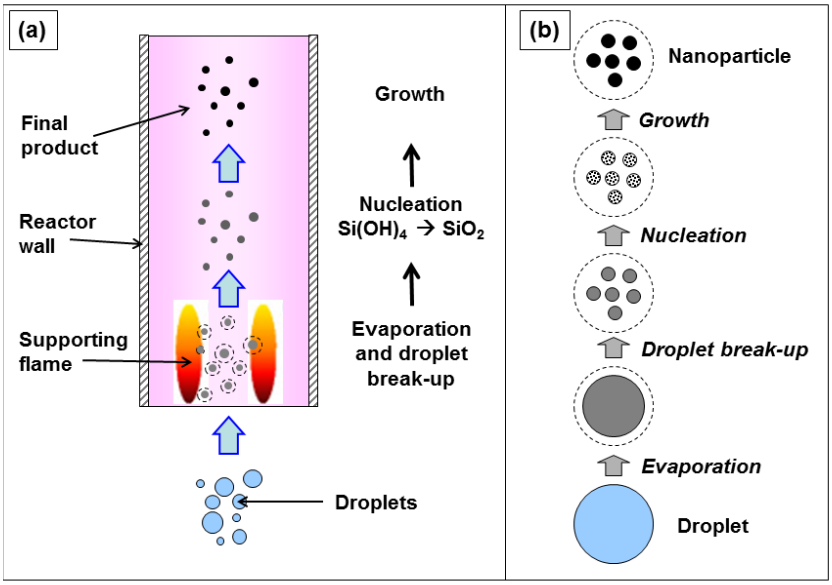


Figure 2. Illustration image of the particle formation mechanism. Figure (a) is the image for the particle travel phenomenon in the flame reactor, whereas Figure (b) is the particle formation mechanism

EXPERIMENTAL METHOD

Raw Materials

The following raw materials were used in this study: rice straw waste (rice field in Cimahi, Indonesia) and potassium hydroxide (KOH, 98%, PT. Bratachem, Indonesia). Rice straw was washed with ion-exchanged water, dried naturally for nine days, and then cut to get the size of about 1 cm.

Synthesis of Potassium Silicate Particles

Rice straw was heated in furnace at 973 K to produce rice straw ash in room condition. To ensure the conversion of organic material into ash, the heating process was maintained for about 7 hours in the furnace. Then, the rice straw ash was ground and put into the potassium hydroxide solution. The mixture was then stirred at 900 rpm and heated at 338 K for 2 hours. Next, the extraction solution was put into the flame-assisted spray-pyrolysis apparatus (equipped with commercial liquid petroleum gas (LPG; Pertamina) and flow of air to introduce droplets into flame apparatus).

In the flame-assisted spray-pyrolysis apparatus (see Figure 3), the system consisted of an ultrasonic nebuliser (NE112; Omron Corp., Japan, used for generating droplet with diameter of about 4.5 um; the ultrasonic nebuliser was equipped with cooling system and cyclone), a flame burner (used for drying as well as polymerising silica component), a flame reactor (stainless steel; diameter of 20 cm and total height of 50 cm), and a filter. The apparatus was also introduced by a flow of LPG (as a fuel source of flame burner) and flows of air. There are two types of air used in this study. One is carrier gas for introducing droplet from the

ultrasonic nebuliser to the flame reactor, and the other is total air that acts as an oxidiser combustion process in the burner. In this study, the flow of LPG in the flame-assisted spray pyrolysis was varied (in the range of 0.30 and 1.20 L/min), whereas the flows of air for the carrier and the oxidiser were fixed at 0.60 and 1.50 L/min, respectively. Based on our previous report (Nandiyanto, Fadhlulloh, Rahman, & Mudzakir, 2016a), fuel with gas flow rate below 0.30 L/min allows unstable flame, whereas fuel with gas flow rate exceeding 1.20 L/min gives impurities in the final product.

Characterisation

Chemical and functional groups, morphology, and particle size of potassium silicate particles were analysed using Transmission Electron Microscopy (TEM, JEOL JEM-1400, JEOL Ltd., Japan) and Fourier Transform Infrared (FTIR, FTIR-8400, Shimadzu Corp., Japan). To analyse the composition of gas, we used gas chromatography mass spectrometry (GC-MS, GCMS-QP2010 Ultra, Shimadzu Corp., Japan). To analyse the chemical component in the sample, we used several analysis methods including gravimetric (for analysing silica) and atomic absorption spectroscopy (AAS; Varian Spectra 240 FS, Varian Inc., US; for analysing metal oxide component).

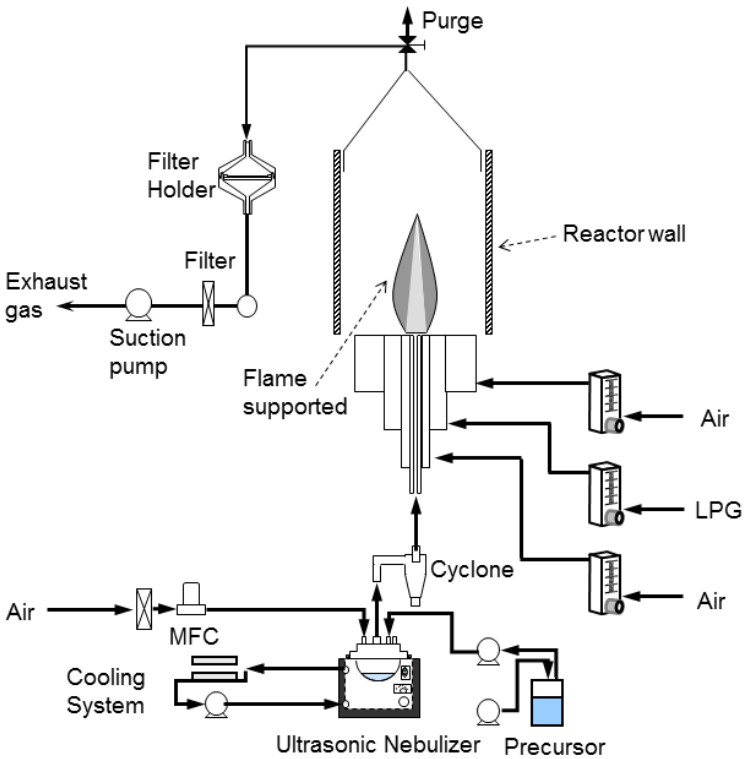


Figure 3. Illustration of the flame-assisted spray-pyrolysis apparatus. MFC is mass flow controller, and LPG is commercially available liquid petroleum gas

RESULTS AND DISCUSSION

Physicochemical Properties of Raw Materials

Table 2 shows the contents of silica and other component contained in the rice straw ash. The results identified that most of the components in the rice straw ash are silica, in which the amount of silica was 74.60%. Other components contained in the rice straw ash were ash, K_2O, CaO, MgO, Na_2O, Fe_2O_3, and Al_2O_3. As silica is the main component in the rice straw ash, we can conclude that rice straw is a potential source of silica material.

As a comparison for ensuring the effectiveness of our method for enriching silica, we added information on the chemical composition of rice straw adopted from the references (van Soest, 2006; Minu, Jiby, & Kishore, 2012; Zaky et al., 2008). We found that the concentration of silica increased from 13 to 74%. The increase in silica was followed by a great decrease in the composition of organic component. This revealed that most of the organic components decomposed into carbon dioxide, with some of them into ash. In addition, the table below presents information for the initial rice straw from the reference, while detailed information on the composition of silica in the initial raw straw will be given in our future work.

Table 2
The chemical contents of rice straw before and after the burning process

Composition	Before burning process	After burning process	
	Quantity (%)	Quantity (%)	Method
SiO_2	~13[a]	74.60	Gravimetric
Ash	13 – 20[c]	~ 5.00	
Oxide component		~ 12.00	AAS
K_2O	1.40 – 2.00[b]	9.06	AAS
CaO		1.87	AAS
MgO		1.48	AAS
Na_2O		0.45	AAS
Fe_2O_3		0.35	AAS
Al_2O_3		0.32	AAS
Organic component	~75		
Cellulose	32 – 47[c]		
Hemicellulose	19 – 27[c]		
Lignin	5 – 24[c]		
N	0.50 – 0.80[b]		
P_2O_5	0.16 – 0.27[b]		
S	0.05 – 0.10[b]		

Note: [a]Data were obtained from the reference (van Soest, 2006); [b]Data were obtained from the reference (Minu, Jiby, & Kishore, 2012). [c]Data were obtained from the reference (Zaky et al., 2008)

Figure 4 shows the GC-MS analysis results of the commercially available LPG used in this study. Based on the GC-MS analysis results, the LPG contained several combustible organic components including propane (58.90%), isobutane (18.35%), butane (22.26%), and 2-methyl

butane (0.48%). Since the component in the commercially available LPG is combustible, the LPG is prospective as a fuel source for flame-related process.

Based on the GC-MS results, complete combustion process is achieved if it only fits the following reaction:

$$LPG_{(g)} + 5.86\ O_{2(g)} \rightarrow 3.60\ CO_{2(g)} + 4.60\ H_2O$$

where LPG is the mixed gas containing propane (58.90%w), isobutane (18.35%w), butane (22.26%w) and 2-methyl butane (0.48%w). In simplification from the above chemical reaction, the mole ratio of oxygen and LPG must be 5.86. Thus, to completely combust 1 mole of LPG, the minimum number of oxygen is 5.86 mole.

Taking into account that the percentage of oxygen in the air is 21% and the densities of propane, iso butane, butane, 2-methyl butane, and oxygen are 0.55; 0.573; 0.573; 0.626; and 1.429 kg/m^3 (data obtained from www.chemicalbook.com; accessed on 24 December 2016), respectively, the minimum volume air required to reach a complete combustion of 1 L/min of LPG is about 10 L/min. In general, when using the flow rate of air of less than that value, the complete combustion cannot be obtained and the formation of carbon in the final product cannot be avoided as well.

Using the above correlation, the process was conducted using a gas carrier flow rate (air) of 0.45 L/min and a total air gas flow rate of 1.50 L/min, whereas the flow rate of LPG varied between 0.30 and 1.20 L/min, and the minimum total flow rate of air must be in the range between 3 and 12 L/min. Thus, we can conclude that the combustion process is supposed to give some carbon components in the product. Although the process must utilise higher flow rate of air, the use of higher rate of air can disturb the flame process. The flame is unstable and the possibility of fire to be ceased is high.

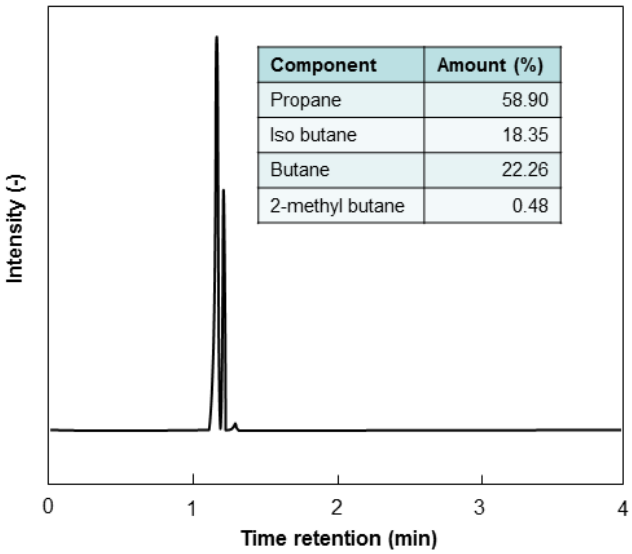


Figure 4. The GC-MS analysis result of commercially available LPG. The panelled table is the detailed information for the component in the LPG. The result was adopted from the following reference (Nandiyanto, Fadhlulloh, Rahman, & Mudzakir, 2016a)

Process Condition of Flame-assisted Spray-pyrolysis Apparatus: Effect of Fuel Flow Rate on the Reactor Temperature Profile

Figures 5a show the photograph image of flame for the production of particles. Figures 5a(i) and 5a(ii) are the images for the process conducted using 0.45 and 1.20 L/min of fuel flow rate, respectively. The images clearly reveal that the combustion process resulted in yellow flame. Some blue flames were formed only when conducting the experiments with the low rate of fuel. The higher rate for the fuel allowed the formation of greater fire with more yellow flame.

The existence of yellow flame is related to the production of incomplete combustion process. Indeed, it also indicates that carbon product will be created during the combustion process. To clarify the discussion about the carbon product, the following reaction happens:

$$LPG(g) + m\ O_{2(g)} \rightarrow n\ CO_{2(g)} + p\ CO_{(g)} + q\ C(s) + r\ H_2O$$

where LPG is the mixed gas containing propane (58.90%w), isobutane (18.35%w), butane (22.26%w), and 2-methyl butane (0.48%w). *m, n, p, q*, and *r* are, respectively, the reaction coefficient, depending on the availability of oxygen. In short, the more oxygen is supplied, the less values of p and q will be obtained. In general, in the constant air flow rate, the higher fuel flow rate allows less fuel molecules to contact and to be oxidised completely. Indeed, less oxidised fuel creates the decomposition of fuel without oxygen. As a consequence, carbon-related component is formed (the more value of q is formed) (Ribeiro & Pinho, 2004).

Figure 5b is the temperature distribution inside the reactor at a LPG flow rate of 0.90 L/min. We selected 0.9 L/min of LPG flow rate because the flame produced is typically stable. The blue flame was also found in this flow rate, while higher flow rate of LPG fuel would create more yellow flame in the process. To make discussion easier, the process temperature distribution was signed by the red colour, which was from transparent red to dark red. The figure clearly revealed that the temperature depended on the position from the flame reactor. The maximum temperature was reached in the specific position inside the flame (shown by the dark red area), whereas the farer from the flame resulted in the lower temperature (shown by the transparent red area). We also found that the temperature of wall in the reactor was relatively the same, which was in the range of 300 and 400°C.

In the centre of the reactor (shown as a dashed-and-dot line), the temperature was varied. This centre can be known as the main path of droplet travel because most of the droplets go through this line. In the position of 0 m distance, the temperature was between 100 and 300°C, in which, this was due to two reasons: the initial step for contacting fuel and oxygen (known as dark zone) and the interaction of flame with the droplet [Droplet evaporates some water (as a solvent)]. Indeed, this water cools the temperature because the evaporation of solvent absorbs some heat from its surrounding. The maximum temperature was reached at the distance between 0.125 and 0.15 m from the burner. The maximum temperature was about 700°C. Passing the position for the maximum temperature, the temperature decreased along the centre of the reactor to the outlet. Indeed, the decreases in the temperature were because the position was outside of the main flame (passing the **outermost zone**). In addition, the reactor wall temperature was almost the same in the range of 300 and 400°C due to the heat radiation (from the combustion process) and the force convection (in the wall itself).

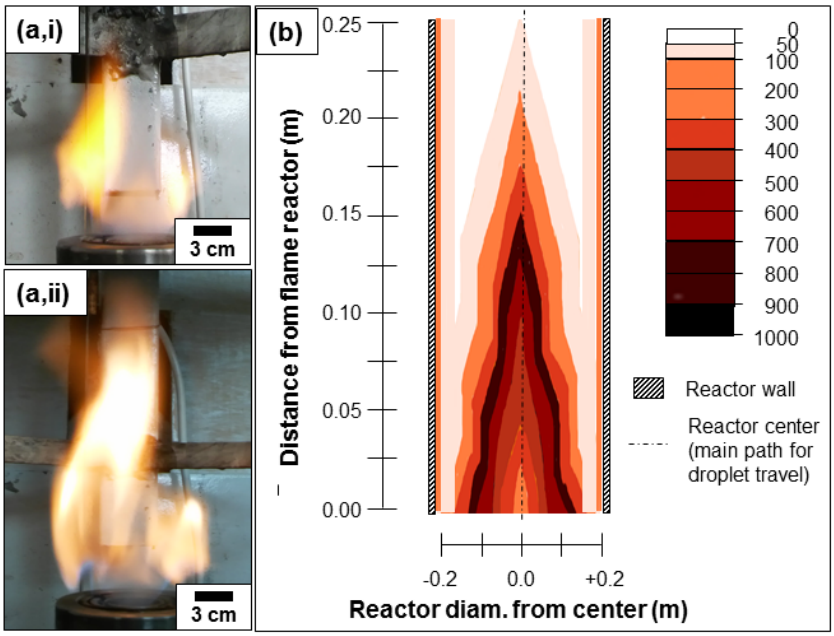


Figure 5. Photograph image of flame-supported pyrolysis method. Figures (a,i) and (a,ii) are the photograph images of flame produced using 0.45 and 1.20 L/min of LPG, respectively. Figure (b) is the reactor temperature profile at 0.90 L/min of LPG. All processes were conducted using a carrier gas flow rate (air) of 0.45 L/min, and a total air gas flow rate of 1.50 L/min

Figure 6 shows the temperature profile at the centre of reactor. The temperature was analysed using a conventional high-temperature-purpose thermocouple. To confirm the position of flame in the reactor, we added a photograph image of the flame (shown in the insert photograph image), and the position of the outer most zone was shown as a dashed line. As shown in the figure, the present flame is effective for heat-shock treatment that is required for droplet break-up phenomenon. The temperature increased extremely at the beginning of the process (from 0 to 10 cm distance from the burner). The temperature rose to more than 700°C. Then, the temperature increased gradually until it reached the maximum. After passing the fire interface line (see the dashed line), the temperature gradually decreased.

The maximum temperatures for 0.45; 0.60; 0.90; and 1.20 L/min of fuel flow rate were about 700; 785; 815; and 905°C, respectively. The increase in the fuel gas correlated with the generation of higher temperature. This is related to the number of breakage of fuel compound bond. The more broken fuel bond results in the more energy to be created.

During the flame, fire can be divided into several zones: dark zone, blue zone, luminous zone, and outermost zone. Dark zone (0-3 cm distance from the burner) is the zone when the reaction is started. The next zone is the blue zone, where the fuel has contacts with oxygen. Fuel compound's bond is broken and combined with oxygen. Some fuel compounds are radicalised. The third zone is the luminous zone (yellow zone), where the reaction continued from the blue zone. The radicalised fuel compounds combined with oxygen molecules further, breaking more bond. Indeed, higher temperature is created compared with the blue zone. The

final zone is the outermost zone. This zone is the hottest areas, with the temperature can take the maximum. However, passing this zone, the temperature drops significantly. Based on the experiments, the outermost zones for the burning process using 0.45; 0.60; 0.90; and 1.20 L/min of LPG fuel flow rate was about 10, 14, 16, and 17 cm distance from the burner, respectively.

Although the increase in the fuel gas is proportional to generate higher temperature, in which this temperature is crucial for evaporation and reaction process, the rate of fuel must be considered. The rate of fuel must be optimised and compared with the number of air flow rate. The ratio of fuel and air must be in the equal number in stoichiometry. Thus, to get a complete combustion process where all the fuel sources are converted into energy with carbon dioxide and water, the amount of air must be in excess. Otherwise, when using less amount of air, the product will have carbon impurities.

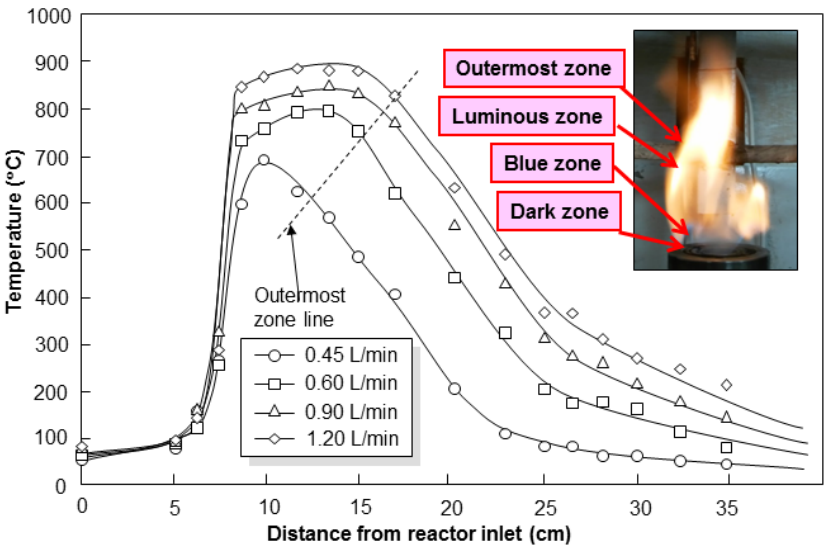


Figure 6. Reactor temperature profile at the centre of the reactor at various LPG fuel flow rates in the range of 0.45 and 1.20 L/min. The process was conducted using a carrier gas flow rate (air) of 0.45 L/min, and a total air gas flow rate of 1.50 L/min. Panelled figure was the photograph image of fire with information about the position of fire interface

Production of Potassium Silicate Nanoparticles with Controllable Size: Effects of LPG Flow Rate on the Final Particle Size

Figures 7a-c show the TEM analysis results of the particles prepared by using the flame-assisted spray-pyrolysis method. The image shows that the potassium silicate particles have a spherical morphology in the sizes of less than 100 nm (see Figure 7d).

As shown in Figure 7a, the particles in the sizes of about 25 nm were produced when the process conducted at an LPG fuel flow rate of 0.60 L/min. Then, the particles in the sizes of 75 nm were produced when the process was performed using the fuel flow rate of 0.90 L/min. These results give information that the fuel affects the final particle size.

A high magnification of TEM image in Figure 7c shows the spherical particle. The particle seems to have a composite structure due to the appearance of smaller white dots in the TEM image. The composed component as small dots (see arrow in the image) is from the potassium compound. However, to confirm our hypothesis regarding the existence of potassium compound as the small dot, we will do further analysis in our future work.

To confirm the effects of fuel flow rate on the final particle size, we varied the LPG fuel gas flow rate in the production process by using the flame-assisted spray-pyrolysis method. The result showed that the higher fuel gas flow rate allowed the production of larger particles (see Figure 7b). Less fuel flow rate did not result in the production of particles with several nanometres (see Figure 7a). Further, too slow results lead to unstable flame. Thus, the formation of particles was not successful.

In addition, based on the results above, the higher LPG flow rate can create larger particles. The main possible reason is the existence of carbon in the particle. The more fuel gas flow rate applied has a correlation with the existence of more carbon component as impurities in the final product.

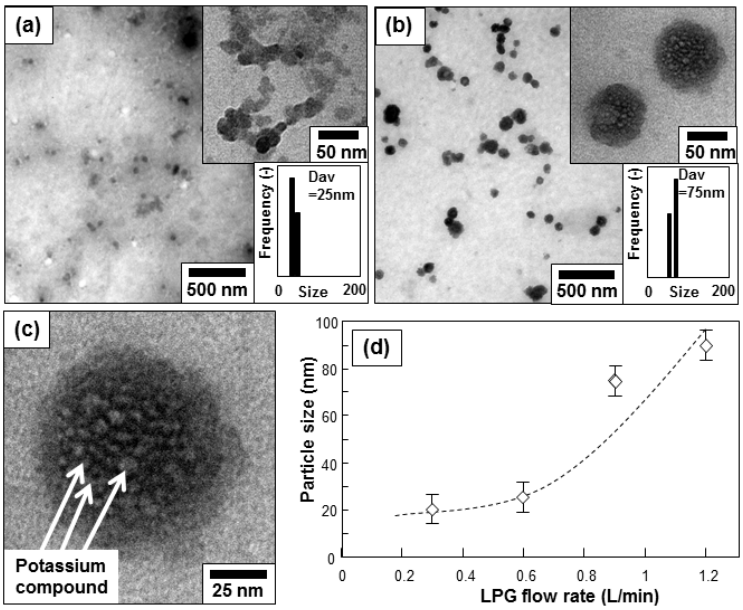


Figure 7. The TEM images of particles produced at various LPG flow rates. Figures (a) and (b) are the samples prepared at the LPG fuel flow rate of 0.60 and 0.90 L/min, respectively. Figure (c) is the high magnification of the TEM image of sample prepared at the fuel flow rate of 0.9 L/min. Figure (d) is the effect of LPG flow rate on the particle size. All the processes were conducted using a carrier gas flow rate (air) of 0.45 L/min, and a total air gas flow rate of 1.50 L/min

Figure 8 shows the FTIR analysis results of the flame spray-pyrolysed samples. In particular, the results showed almost similar absorption peaks and patterns for all the samples. The peak intensities were almost the same for all the samples, except for the peaks in the range of 2000-2100 and 2500-3000 cm^{-1} (see the dashed line area). Absorption peaks of silica-related

compound were identified, including the peaks at 806.25, 883.40, 1058.92, and 3194 cm^{-1} that ascribed to the existence of Si-O-Si symmetric stretching, Si-OH, Si-O-Si asymmetric stretching, and SiO-H, respectively. Further, FTIR also detected the peaks at 1128.36 and 1662.64 cm^{-1}, in which these belonged to Si-O-X (X = K, Si, atau H) stretching and K_2SiO_3, respectively. Detection of these potassium-related peaks strengthened our hypothesis that the present process is effective for maintaining potassium element in the product.

In addition to the above peaks, unique absorption peaks in the range of 2000-2100 and 2500-3000 cm^{-1} have different intensities for all the samples (as shown in the dashed line area in the figure). The intensities of the peaks increased with the increasing LPG fuel flow rate. The most different peak at in the range of 2000-2100 was obtained for the sample with 1.20 L/min of LPG flow rate. The appearance of these peaks corresponds to the existence of carbon-related component, which matches with our hypothesis (explained in Figure 4) and matches with our previous work (Nandiyanto, Fadhlulloh, Rahman, & Mudzakir, 2016a).

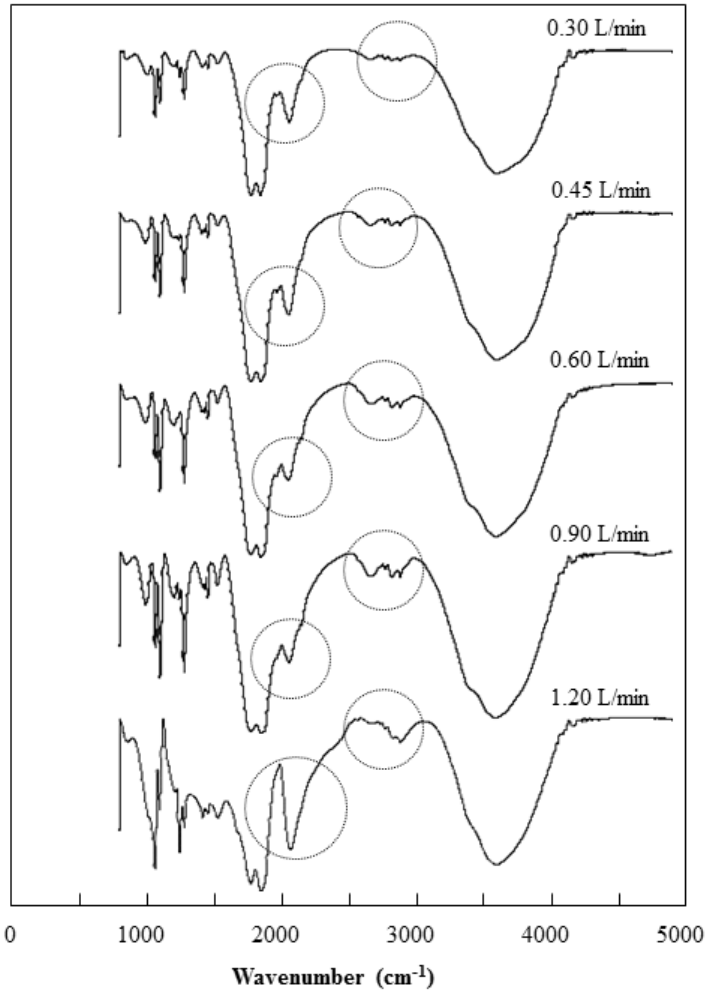


Figure 8. The FTIR analysis results of the samples produced at various LPG fuel flow rates in the range of 0.30 and 1.20 L/min. All the processes were conducted using a carrier gas flow rate (air) of 0.45 L/min, and a total air gas flow rate of 1.50 L/min

To ensure the peaks existed in the FTIR analysis, we presented a list of related peaks in comparison to the references (see Table 3) (Simonsen, Sønderby, Li, & Søgaard, 2009; Scarano, Bertarione, Spoto, Zecchina, & Arean, 2001; Cotton & Kraihanzel, 1962; Skoog & West, 1980; Nandiyanto et al., 2016b). As shown in the table, the peaks detected in the range of 2500 and 3000 cm^{-1} in Figure 8 are carbon-related material. The existence of carbon-related material is proportional with the flow rate of fuel. Indeed, in the constant flow rate of air, the higher flow rate of fuel correlated to the higher possibility to form carbon in the incomplete combustion process (Nandiyanto, Fadhlulloh, Rahman, & Mudzakir, 2016a).

Based on the FTIR data, the optimum condition to get potassium silicate particles is when applying process with less than 0.90 L/min of LPG fuel flow rate. Applying more than 0.90 L/min of LPG fuel rate would result in carbon component impurities in the particle.

Table 3
The FTIR spectra for silica and carbon-related material based on the references

Function Group	FTIR peak (cm^{-1})	Reference
Si-O-Si symmetric bending	461.34	Simonsen, Sønderby, Li, & Søgaard (2009)
Si-H	623	Simonsen, Sønderby, Li, & Søgaard (2009)
Si-O-Si symmetric stretching	795	Simonsen, Sønderby, Li, & Søgaard (2009)
Si-OH	895-880	Simonsen, Sønderby, Li, & Søgaard (2009)
Si-O-Si asymmetric stretching	1107	Simonsen, Sønderby, Li, & Søgaard (2009)
SiO-H	3400-3200	Simonsen, Sønderby, Li, & Søgaard (2009)
Si-O-X (X= K, Si, atau H) stretching	1100	Simonsen, Sønderby, Li, & Søgaard (2009)
K_2SiO_3	1625	Simonsen, Sønderby, Li, & Søgaard (2009)
Si-O stretching	1000-1200	Simonsen, Sønderby, Li, & Søgaard (2009)
Si-O bending	890-975	Simonsen, Sønderby, Li, & Søgaard (2009)
C-H (Alkanes)	2850-2970	Scarano, Bertarione, Spoto, Zecchina, & Arean (2001)
C-H (Alkenes)	3010-3095	Scarano, Bertarione, Spoto, Zecchina, & Arean (2001)
C-H (Alkynes)	3000	Scarano, Bertarione, Spoto, Zecchina, & Arean (2001)
C-H (aromatic rings)	3010-3100	Scarano, Bertarione, Spoto, Zecchina, & Arean (2001)
C-O	1050-1300	Cotton & Kraihanzel (1962)
C=O	1690-1760	Cotton & Kraihanzel (1962)
C-C	2100-2260	Cotton & Kraihanzel (1962)
C=C(alkenes)	1610-1680	Skoog & West (1980)
C-C(aromatic rings)	1500-1600	Skoog & West (1980)
C≡C(Alkynes)	2100-2260	Skoog & West (1980)

CONCLUSION

We have successfully demonstrated the technique to produce potassium silicate nanoparticles with controllable particle size (from 20 to 80 nm). This is different from other methods that used expensive raw chemicals, in this study, rice straw waste was used as a source of silica that is relatively low in price and largely available. To assist the production of nanoparticles, we used flame-assisted spray-pyrolysis method that utilised LPG as the fuel source for burner. We found that the prepared particles are spherical in sizes of nanometer range. To control the particle size, we found that managing fuel flow rate is the best option. In addition to the successful production of potassium silicate nanoparticles, since the product is generated from the rice straw waste that has remained underutilised until now, a further development of the present study would have a great positive impact for the reduction of rice straw waste emission and conversion of waste into useful and valuable products.

ACKNOWLEDGEMENT

Authors would like to acknowledge RISTEK DIKTI for supporting the research fund. Authors would also like to thank Novie Permatasari and Transmissia Noviska Sucahya for the assistance and support given in conducting this research.

REFERENCES

Baghramyan, V., Sargsyan, A., & Harutyunyan, V. (2016). Synthesis sodium, potassium silicates for thermoregulating coatings space vehicles by microwave method. *Journal of Armenia, 1*(68), 2016.

Beneke, K., & Lagaly, G. (1989). A hydrated potassium layer silicate and its crystalline silicic acid. *American Mineralogist, 74*(1-2), 224-229.

BPS. (2016). *Produksi padi tahun 2015 naik 6,37 persen*. Retrieved December, 2016, from https://www.bps.go.id/brs/view/id/1271

Chiang, C. Y., Aroh, K., & Ehrman, S. H. (2012). Copper oxide nanoparticle made by flame spray pyrolysis for photoelectrochemical water splitting–Part I. CuO nanoparticle preparation. *International Journal of Hydrogen Energy, 37*(6), 4871-4879.

Cotton, F. A., & Kraihanzel, C. S. (1962). Vibrational spectra and bonding in metal carbonyls. I. Infrared spectra of phosphine-substituted group VI carbonyls in the CO stretching region. *Journal of the American Chemical Society, 84*(23), 4432-4438.

Fadhlulloh, M. A., Rahman, T., Nandiyanto, A. B. D., & Mudzakir, A. (2014). Review tentang sintesis SiO2 Nanopartikel. *Jurnal Integrasi Proses, 5*(1), 30-45.

Kalapathy, U., Proctor, A., & Shultz, J. (2000). A simple method for production of pure silica from rice hull ash. *Bioresource Technology, 73*(3), 257-262.

Karthikeyan, J., Berndt, C. C., Tikkanen, J., Wang, J. Y., King, A. H., & Herman, H. (1997). Nanomaterial powders and deposits prepared by flame spray processing of liquid precursors. *Nanostructured Materials, 8*(1), 61-74.

Kikuchi, R. (1999). Application of coal ash to environmental improvement: Transformation into zeolite, potassium fertilizer, and FGD absorbent. *Resources, Conservation and Recycling, 27*(4), 333-346.

Lu, P., & Hsieh, Y. L. (2012). Preparation and characterization of cellulose nanocrystals from rice straw. *Carbohydrate Polymers, 87*(1), 564-573.

Minu, K., Jiby, K. K., & Kishore, V. V. N. (2012). Isolation and purification of lignin and silica from the black liquor generated during the production of bioethanol from rice straw. *Biomass and Bioenergy, 39*, 210-217.

Mori, H. (2003). Extraction of silicon dioxide from waste colored glasses by alkali fusion using potassium hydroxide. *Journal of Materials Science, 38*(16), 3461-3468.

Nandiyanto, A. B. D., Fadhlulloh, M. A., Rahman, T., & Mudzakir, A. (2016a). Synthesis of carbon nanoparticles from commercially available liquified petroleum gas. IOP Conference Series: *Materials Science and Engineering, 128*(1), 012042.

Nandiyanto, A. B. D., Rahman, T., Fadhlulloh, M. A., Abdullah, A. G., Hamidah, I., & Mulyanti, B. (2016b). Synthesis of silica particles from rice straw waste using a simple extraction method. I*OP Conference Series: Materials Science and Engineering, 128*(1), 012040.

Nandiyanto, A. B. D., & Okuyama, K. (2011). Progress in developing spray-drying methods for the production of controlled morphology particles: From the nanometer to submicrometer size ranges. *Advanced Powder Technology, 22*(1), 1-19.

Novotny, R., Hoff, A., & Schuertz, J. (1993). *U.S. Patent No. 5,238,668*. Washington, DC: U.S. Patent and Trademark Office.

Permatasari, N., Sucahya, T. N., & Nandiyanto, A. B. D. (2016). Review: Agricultural wastes as a source of silica material. *Indonesian Journal of Science and Technology, 1*(1) 82-106.

Rastsvetaeva, R. K., Aksenov, S. M., & Taroev, V. K. (2010). Crystal structures of endotaxic phases in europium potassium silicate having a pellyite unit cell. *Crystallography Reports, 55*(6), 1041-1049.

Ribeiro, L., & Pinho, C. (2004). Generic behaviour of propane combustion in fluidized beds. *Chemical Engineering Research and Design, 82*(12), 1597-1603.

Santmyers, D. (1957). *U.S. Patent No. 2,784,060*. Washington, DC: U.S. Patent and Trademark Office.

Scarano, D., Bertarione, S., Spoto, G., Zecchina, A., & Arean, C. O. (2001). FTIR spectroscopy of hydrogen, carbon monoxide, and methane adsorbed and co-adsorbed on zinc oxide. *Thin Solid Films, 400*(1), 50-55.

Skoog, D. A., & West, D. M. (1980). *Principles of instrumental analysis* (Vol. 158). Philadelphia: Saunders College.

Simonsen, M. E., Sønderby, C., Li, Z., & Søgaard, E. G. (2009). XPS and FT-IR investigation of silicate polymers. *Journal of Materials Science, 44*(8), 2079-2088.

Tokunaga, Y. (1991). Potassium silicate: a slow-release potassium fertilizer. *Nutrient Cycling in Agroecosystems, 30*(1), 55-59.

van Soest, P. J. (2006). Rice straw, the role of silica and treatments to improve quality. *Animal Feed Science and Technology, 130*(3), 137-171.

Zaky, R. R., Hessien, M. M., El-Midany, A. A., Khedr, M. H., Abdel-Aal, E. A., & El-Barawy, K. A. (2008). Preparation of silica nanoparticles from semi-burned rice straw ash. *Powder Technology, 185*(1), 31-35.

Pertanika J. Sci. & Technol. 26 (1): 409 - 426 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Morphodynamics of Coastal Lagoons: An Analysis of Multitemporal Landsat Images of Segara Anakan Lagoon Area

Nandi

Department of Geography Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi, Bandung 40154 West-Java, Indonesia

ABSTRACT

Lagoons are shallow coastal bodies of water separated from the ocean by a series of barrier islands which lie parallel to the shoreline. Sediment supplied to enclosed lagoons will influence the distribution and morphodynamics of coastal lagoons. Lagoons are re-shaped by erosion and deposition by the accumulation of material washed or blown over the enclosing barriers, and by the accretion of in-flowing river sediment. The purposes of this study is to describe the methodology and attain results for identification of multitemporal Landsat image data of the Segara Anakan Lagoon for the years 1978, 1994, 2003, and 2013. The method used in this study is a visual interpretation of Landsat by using GIS and remote sensing methods to produce maps of morphodynamics of a lagoon and multitemporal of a coastline change. The overall results for the identification of morphodynamics of coastal lagoons are the decrease of areas of water bodies, increase of areas of a land accretion, and coastline changes. The results are used to examine ecological management of coastal lagoons.

Keywords: Coastal Lagoon, Landsat Images, Morphodynamic, Multitemporal, Segara Anakan

INTRODUCTION

Environmental changes over the coastal areas are a result of regular processes that are continuous or periodically fluctuated such as oceanographic parameters and the dynamics of local climate. Coastal erosion, which is caused by the dominant factors of natural influence such as waves, currents, tides and sediment supply, leads to the changes of the coastline (Bird & Lewis, 2015).

The characteristics of these coastal zones are very dynamic (Woodroffe, 2002). There is an interaction between the land and sea through the role of the waves and winds. Erosion and sedimentation frequently occur in the coastal area, when rocks are eroded by the waves and winds. They will then deposit

ARTICLE INFO

Article history:
Received: 25 April 2017
Accepted: 28 November 2017

E-mail address:
nandi@upi.edu (Nandi)

ISSN: 0128-7680

at such zones, which depends on the energy of the waves and winds. In such a place, it will change the form of a landscape (Davidson-Arnott, 2010).

By the time, these dynamic processes will generate a new landform or change an existing form of that place. In this term, coastal geomorphology focuses on explaining landform in the coastal zone by examining the shapes, sediment and depository history at the modern shoreline. It includes the study of shallow marine environment that is influenced by terrestrial factors and the land. The influence of the sea is also present and the long-term evolution of the coast is appropriately understandable (Woodroffe, 2002).

One of the coastal landforms that is influenced by natural processes and human activities is coastal lagoon. Even though its position is inland of water bodies, the lagoon is still affected by waves and winds from the sea. The lagoon environment is a closed and semi-closed environment formed by the interaction between marine and terrestrial processes, which has a complex resource that comes from the land and sea. Sources of water in the lagoon are the river and the sea, where tidal currents and waves still have the influence (Nichols & Boon, 1994). Additionally, coastal lagoon will be the depository place for sedimentation from the upland area (Magni et al., 2008). All human activities at the upstream, particularly agriculture and cultivation, bring soil, waste and other materials to the downstream area through the river.

Therefore, lagoon environment is very important to study. It is unique, not only because it has a variety of aspects (geomorphology, oceanography and climatology) but it also contains large mineral and biological resources (Zonta, Guerzoni, Pérez-Ruzafa, & de Jonge, 2007). Segara Anakan (SA), which is located in Indonesia, is an example of a coastal area which has a unique biophysical characteristic. The region has a great natural ability to ensure the sustainability of the interrelationships between terrestrial, estuarine and marine ecosystems in harmony and balance as a habitat for flora and fauna. The region is an area of migration of various types of protected animals and it is a place of breeding for diverse species of the shrimp and fish, which has a high economic value. Furthermore, the region is also a source of livelihood for the local community at large. Hence, it is reasonable for the government to consider the SA area as a natural resource, which needs to be conserved and as a capital base of regional development (Nandi, 2014).

Natural processes and human intervention on the environment of SA can cause ecological dynamics to occur. Therefore, monitoring the activities of the phenomenon needs to be done. One of the effective ways is through the application of remote sensing. The application of remote sensing technologies for monitoring environmental conditions may provide optimal results. Several studies have been conducted to show the effectiveness of the use of remote sensing for monitoring coastal lagoons (Baldock, Weir, & Hughes, 2008; Ardli & Wolff, 2009; Amos, Umgiesser, Tosi, & Townend, 2010; Duck & da Silva, 2012). Moreover, the use of remote sensing data can provide spatial analysis repetitively and continuously cover a relatively large area with a reasonable low-cost. In addition, it is also fast compared to terrestrial surveys. The data obtained from remote sensing can provide objective information, and the method is reliable and economical in the effort for inventory, monitoring and evaluation of resources (Klemas, 2011). Thus, in order to optimise the results of research, it can use multitemporal remote sensing data from Landsat MSS, TM and ETM (Alves, Venerando & Helenice, 2003).

The lagoon ecosystem of SA has experienced a severe environmental degradation due to the high rate of sedimentation and human activities in the forms of land uses that do not take into account sustainable development. The impacts of the human activities can clearly be seen from the number of mangrove forests in SA that have been converted into agricultural lands, settlements and aquaculture, which lead to increasingly depleted marine life (Jennerjahn & Yuwono, 2009). Moreover, abrasion and erosion are greater, and sedimentation is getting worse. As a result, it is now more difficult for the local communities to catch fish so that fish production is decreasing/declining. To make things worse, a reduced accumulation of water in the water reservoir above the estuary can cause flooding in downstream areas, which subsequently leads to multiplier effects such as damage of clean water, pollution of water resources, land degradation, as well as destructions of agricultural areas and community settlements.

Based on the above discussion, silting, environment degradation and possibility development of the lagoon's existence must be anticipated and addressed. Moreover, a suitable policy is needed to ensure that human activities can match the temporal and spatial dynamics of the coastal resources. More importantly, it is necessary to study in-depth concerning these complex issues. The study will focus on the analysis of multitemporal images by using Landsat data to explore the dynamic of SAL.

The investigation of morphodynamics of coastal lagoons made use of the GIS and remote sensing methods, which was intended by the interpretations of the multitemporal Landsat images produced by USGS-NASA. Basically, there are three data of Landsat images (MSS, TM, and ETM+) used for the analysis of ecological changes of the lagoon area. There are also several steps involved in producing the maps of morphodynamics and coastline changes (buffering and overlaying of maps).

The morphodynamics of coastal lagoons was obtained by assessing various Landsat images. The results show that there have been several changes in the morphology of the lagoon during the period of 1978-2013. This can be observed by the declining lagoon areas, which are caused by the sedimentation from surrounding rivers of that area of lagoon. The total area of the lagoon in 1978 was 4,665 hectares. In 1994, it decreased dramatically to 1758 hectares and the increasing of accreted land area was 219 hectares. From 1994-2003, the total area of the lagoon declined to approximately 767 hectares with the increased accreted land area of 258 hectares. In 2013, the total area of the lagoon was only 347 hectares, with a total area of 837 hectares of accreted land. Thus, within 35 years, the change of lagoon areas involved 4,318 hectares or a decrease of approximately 123.37 hectares/year, and the increased accreted land area of 22.5 hectares/year. The morphodynamics changes depend on factors such as the type of sediment, morphology and geology of the beaches (Stephenson & Brander, 2003).

MATERIALS AND METHODS

Multitemporal satellite images from USGS (Landsat MSS, TM, and ETM+) were used to obtain the necessary data for the years 1978, 1994, 2003 and 2013. Geological Map scale was 1:100.000, Topographical Map (*Rupa bumi*) scale 1:25.000, and ground truth. The results of satellite imagery interpretation include landform, land use and data. Geological map interpretation will generate data on main constituent material landform, while Topographical

map generating slopes data through the interpretation of contour lines. Meanwhile, Ground truth was conducted with the location of points on the field based on a predetermined identification of landforms to make observations, measurements and interpretation of the matching results. Field check was performed to validate the results from the interpretation of imagery data and thematic maps that matched with the real condition. Besides, observations and data measurements that cannot be intercepted from the interpretation of imagery and thematic maps were conducted during the fields check.

Visual interpretation is intended to produce maps of morphodynamics of lagoons area and multitemporal land cover/use changes. Landforms of the lagoon maps are manually interpreted from the multitemporal Landsat imagery exposed with SRTM hill shade in the process. Recognition and identification of landforms from remotely sensed imagery was based on relief or shape, density and location. Then, to delineation the area of lagoon, it was derived from the images by digitising directly on screen or using digital elevation model (DEM). Manual digitisation on screen can help to determine the region of interest by making a polygon area, which distinguishes the surrounding region. Relief texture is very important to consider in interpretation.

Basic identification of landforms refers to the density that can be demonstrated by the appearance of which varies by utilising multispectral imagery. In this case, density becomes very important indication of relief through the identification of a shadow object. Location and landscape ecology in particular situations is an important part in the identification of landforms. It can be recognised from the structure or pattern of an appearance. Appearance characters of hydrology, vegetation, land use and the others are in a group, which can identify landforms easily. The location of those characters can be shown by the characteristics of the image through relief or density (Sulistyo, 2011).

The prior step before interpretation of landforms is drafting a colour composite image that will be used to perform manual interpretation of landforms. Once a composite colour image is compiled, the image enhancement is then performed, in this case using contrast-stretching techniques. It aims to improve the quality of the image so that it is easy to do interpretation manually.

RESULTS AND DISCUSSION

The Morphodynamics of a Coastal Lagoon

The morphodynamics of a coastal area involves the characteristics and time-series actions of coastal landforms and the trends controlling the actions within specific temporary and spatial data. A survey conducted on the morphodynamics of the lagoon's morphology has helped to understand the characteristic changes as a response to periodic and episodic activities resulting in changes in water courses, sedimentation transportation from watersheds, bathymetry and coastline alignment.

According to the multi-temporal Landsat data, it can be derived that there were serious morphological and coastline changes in the lagoons between 1978 and 2013 due to accretion and erosion. The accumulation occurred due to extreme sediment from the run-off in the

Citanduy and Segara Anakan watersheds and small tributaries surrounding the Segara Anakan Lagoon (SAL) area.

Changes to the lagoon's morphology (1978-2013)

Changes in the morphological of lagoons from 1978 to 2013 are presented in Figures 1-4, and the development of lagoon area is presented in Table 1.

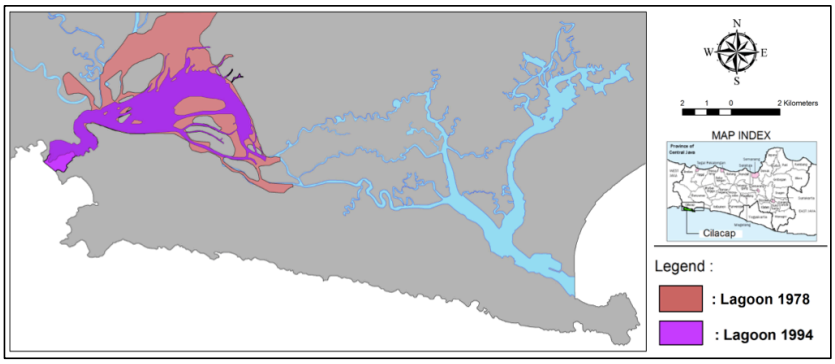


Figure 1. Changes to the lagoon's morphology (1978-1994)
(*Source:* MSS and TM images map, System coordinate WGS 1984 zone 49s; author's own results)

Figure 1 shows how the area of the lagoon changed from 1978 to 1994. This was a first phase of changing area at the lagoon. In 1978, the area of SAL was 4665 hectares, which was reduced dramatically to merely 1758 hectares in 1994. In other words, over 16 years, there was a loss of 2907 hectares. At that time, the first settlement was also built by farmer immigrants from the main land such as Java and Sumatera islands (Ardli & Wolff, 2009). Moreover, during that time, the area of accretion also increased significantly due to sedimentation transport from the upland watersheds. The area of land accretion in 1978 was 49 hectares, which increased by 170 hectares in 1994 to 219 hectares. This means that the increase in land accretions per year over that period was approximately 10.63 ha.

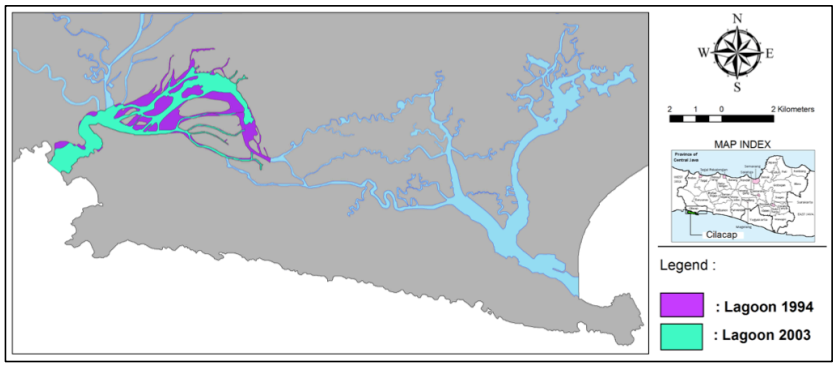


Figure 2. Changes to the lagoon's morphology (1994-2003)
(*Source:* TM and ETM images map, system coordinate WGS 1984 zone 49s; author's own results)

Figure 2 shows changes in the lagoon morphology from 1994 to 2003. The area of SAL did not only change in 1994, but also in 2003, with additional land accretion of approximately 39 hectares. The total area of SAL in 2003 was 991 hectares, which means that over a period of nine years, another 767 hectares were lost from the lagoon. The main cause is the high rate of sedimentation of the Citanduy River, which carries more than 9 million tons m^3 per year of sedimentation matter. Over this same period, an additional resettlement complex has also more than 2000 ha established. This further accelerated the process of environmental degradation in SA. Not only settled but also they stated the management of rice farming, which was running quite successful with the technical assistance provided by community organisation.

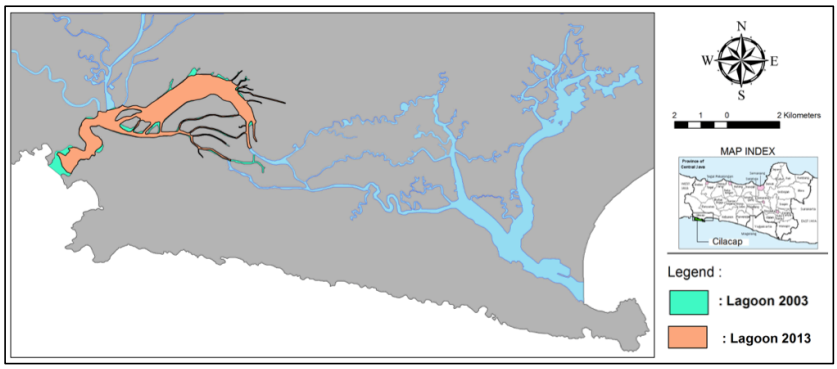


Figure 3. Changes to the lagoon's morphology (2003-2013)
(*Source:* ETM and Landsat 8 images map, System coordinate WGS 1984 zone 49s; author's own results)

Likewise, Figure 3 shows the changes to the lagoon's morphology from 2003 to 2013. In 2013, the SA area was reduced from 644 hectares to 347 only hectares, with additional 579 hectares of accreted land, reflecting that over ten years, the lagoon became narrow with an increase in the accreted land of 57.9 ha/year. It was clear that the extensive lagoon would diminish with a tendency to disappear in the coming year, covered by the sediment. However, the main factor has also increased at that time, i.e. illegal logging of mangrove in this region (Jennerjahn & Yuwono, 2009).

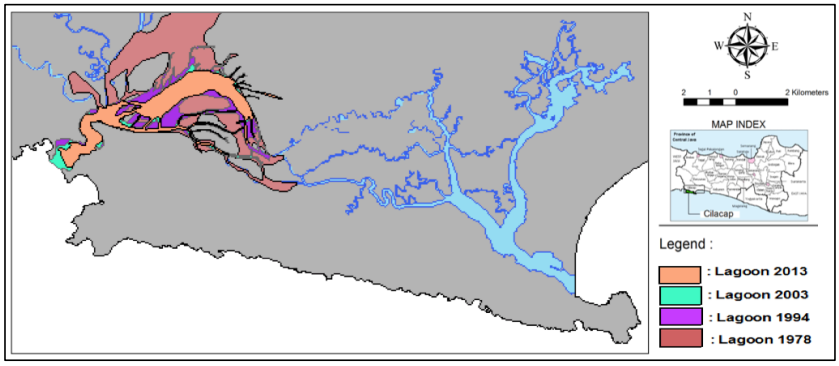


Figure 4. Changes to the lagoon's morphology (1978-2013)
(*Source:* MSS, TM, ETM and Landsat 8 images map, System coordinate WGS 1984 zone 49s; author's own results)

Figure 4 shows the overall changes of the lagoon due to sedimentation and accretion. From 1973 (the first phase of changes), the lagoon area was still in natural condition, i.e. there was not disturbing factor that caused area of the lagoon to shrink. In the second phase (1994), the area of lagoon tended to decrease as an effect of natural phenomenon such as the eruption of Galunggung Mountain volcanic in 1982.

Table 1
The development of the SAL area over the period 1978 – 2013 (Source: Author's own results, 2015)

Years	Lagoon Area (Ha)	Decrease (Ha)	Accreted Land (Ha)	Increase (Ha)
1978	4665	-	49	-
1994	1758	2907	219	170
2003	991	767	258	39
2013	347	644	837	579

Figure 5 shows the shrinkage of the water area in SAL. It can be seen that those changes from the emergence of new land or an increase in the extent of land that had previously been in the lagoon, as well as a reduction in the water area of the SAL.

The shrinkage of the water area of SAL can be seen in Figure 6, which shows that the emergence of land accretions occurred in the middle, western and northern parts of the lagoon. The shrinkage of the water area of SAL is strongly associated with the occurrence of erosion and sedimentation.

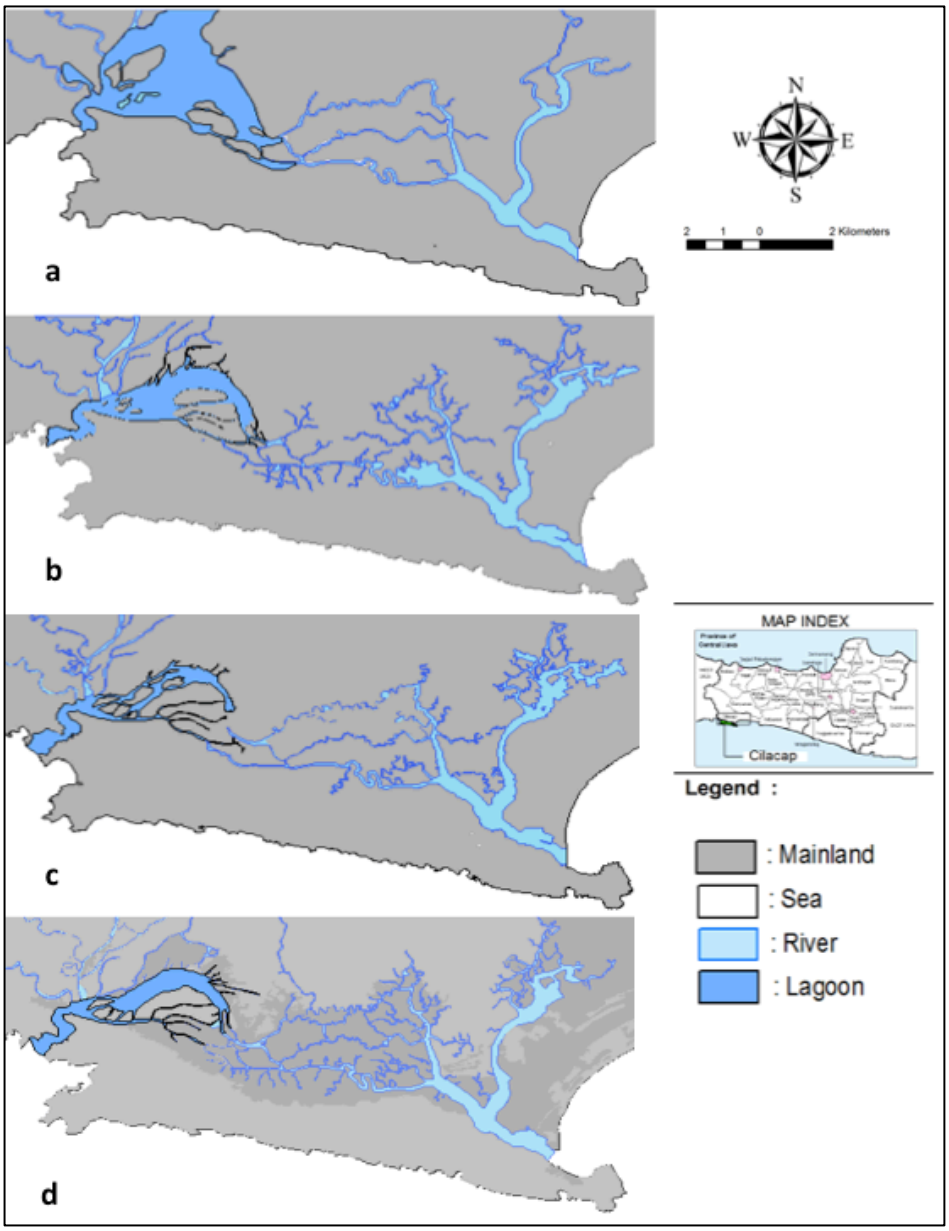


Figure 5. Changes to Segara Anakan lagoon's morphodynamics since 1978 (a-d)
(*Source:* MSS, TM, ETM and Landsat 8 images map, System coordinate WGS 1984 zone 49s; author's own results)

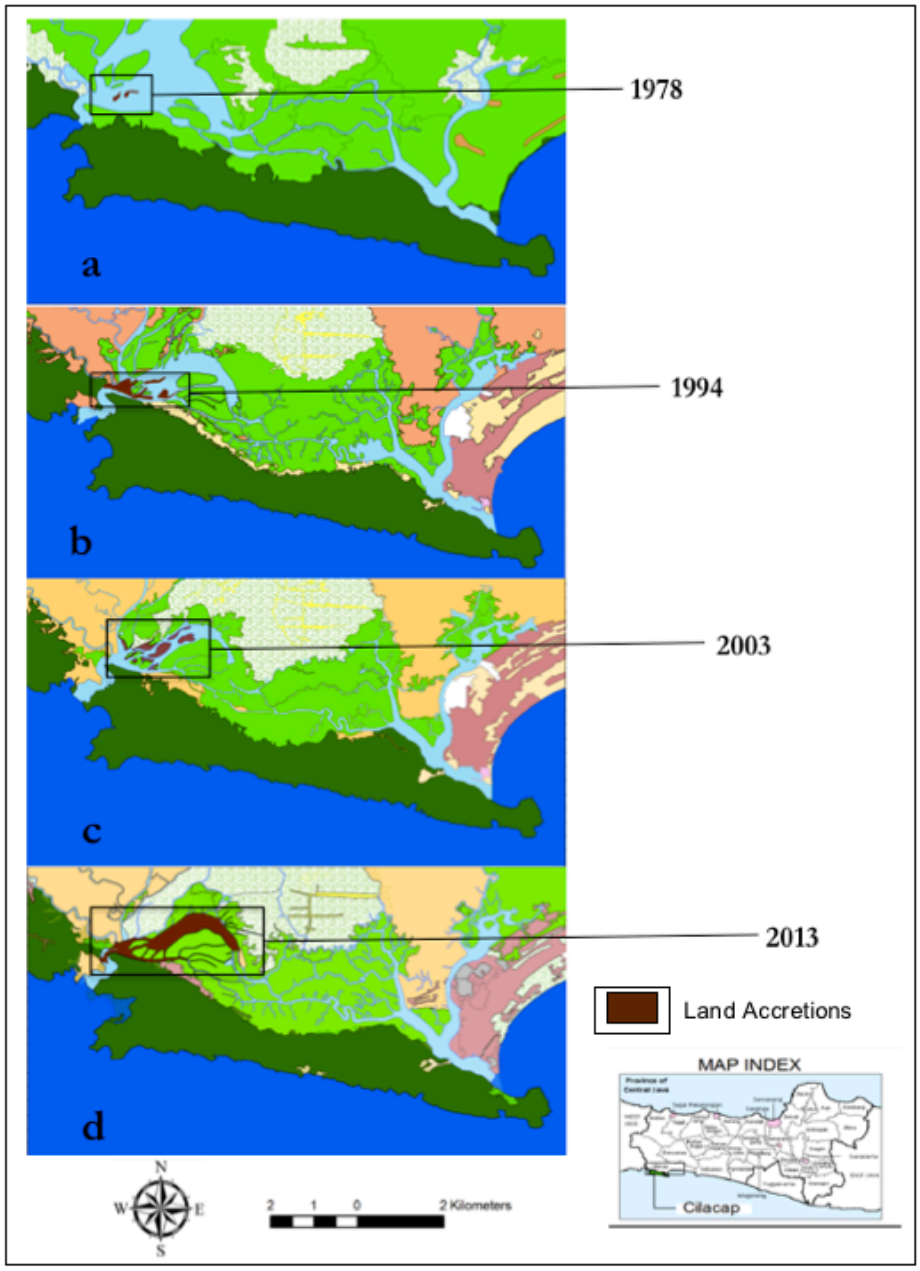


Figure 6. The development of land accretions in the Segara Anakan Lagoon since 1978 (a-d) (*Source:* MSS, TM, ETM and Landsat 8 images map, system coordinates WGS 1984 zone 49s; author's own results)

Figure 7. Sedimentation matter brought from Citanduy River

Figure 7 shows that the materials deposited into the SAL consisted of clay, silt and sand. Sediment from Citanduy River contributed to more or less 9,000,000 tons m^3/year. Citanduy River located in Western Java, where it flows through two provinces of West Java and East Jawa through Ciamis Regency and Cilacap Regency.

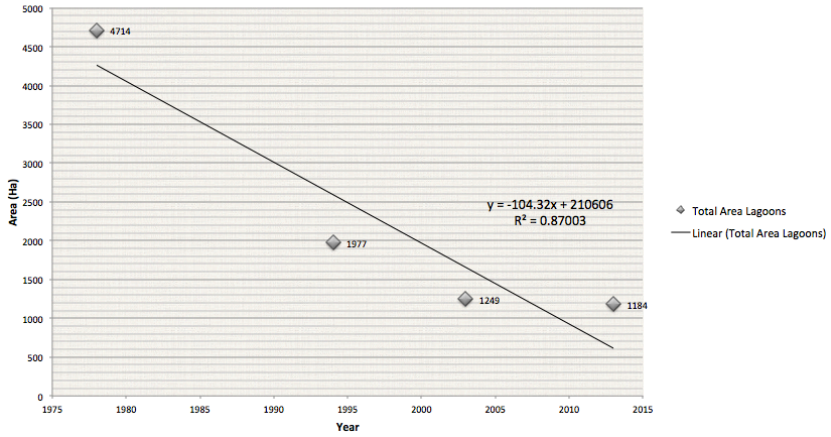


Figure 8. The formula used for the linear regression of the total area of Segara Anakan Lagoon, 1978-2013 (*Source:* Author's own results, 2015)

Figure 8 is the formula used for the linear regression of the total area of Segara Anakan Lagoon in 1978-2013. It shows how much of a change there is between the beginning of the trendline and the end. Regression lines also give us useful information about the data they were collected from. They show how one variable changes on average with another. For instance, the trends of the area drecreased during the development of lagoon, but on the other hands, the area of land accretion was significantly increased.

In addition to the reduction in the area of water found in the SAL, there has also been an additional area of land accretion as shown in Figure 6, which illustrates the extent of the area from land accretion in 1978 (Part a). Meanwhile, Part b shows the changes to the area of the SAL and the increase in the area for land accretion in 1994. Likewise, parts c and d show the changes in the land accretion area in 2003 and 2013. Hence, over a 35-year period, there has been an increase in the area of land accretion, covering 788 hectares, with the growth of land accretion at approximately 22.5 ha/year as a result of sedimentation.

Coastline Changes (1978-2013)

Sedimentation from the soil resulted in legal land tenure. In contrast, damages to coastal areas, due to abrasion in areas that are less stable against water erosion, caused a critical state of the land, damaging road infrastructure.

Besides the effects of sedimentation processes that were taking place in the SA water body, the abrasion process also influenced morphological changes to the SAL. The abrasion process taking place had impacts on the shape and coastline in the western part of the SAL (The comparison of coastline changes from 1978 to 2013 can be seen in Figure 9).

Coastline changes were analysed from Landsat images from 1978, 1994, 2003, and 2013. Coastline extraction from the Landsat images was processed based on the composites from band 543. To record coastline changes, multitemporal coastlines were overlapped. These changes were then used to determine where accretion and abrasion had taken place.

The area under investigation was limited to the shoreline of the SAL. It was divided into three profile areas: A, B, and C. Each profile area was representative for a particular coastline area, i.e. A for the West Channel (*Palawangan Barat*), B for *Nusakambangan Island*, and C for the East Channel (*Palawangan Timur*).

Based on the geomorphology of those area, the coastal area of Cilacap is a coastal region with an average height of less than 2 m. With alluvial soil types and textures of sandy, it is possible that the coastline changes from southern to the northern coastal area of Cilacap with an average rate of shoreline change of 60 m per year in the past 10 years.

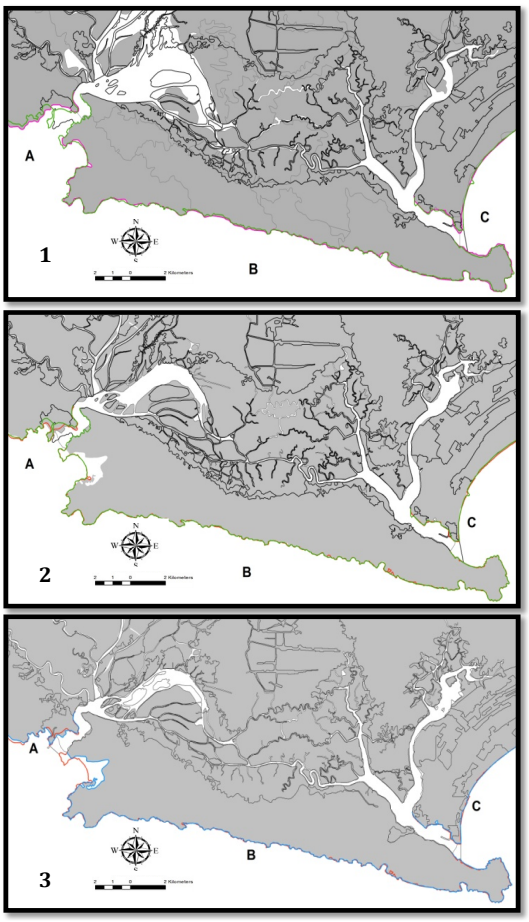


Figure 9. Coastline Changes from 1978-2013: (A) West channel, (B) Nusakambangan, and (C) East Channel (*Source:* MSS, TM, ETM and Landsat 8 images map, system coordinates WGS 1984 zone 49s; author's own results)

Table 2
Area (Ha) of sedimentation changes in the SAL (Source: Author's own results, 2015)

Zone	Change in area (hectares) 1978 – 1994	Change in area (hectares) 1994 - 2003	Change in area (hectares) 2003 - 2103
A (West Channel)	14.19	2.95	0.83
B (Nusakambangan)	0.95	0.47	0.79
C (East Channel)	1.64	1.41	0.37

In the meantime, sedimentation processes have also started to occur in the West Channel. They have changed the mouth of the lagoon. Table 2 shows that the area of this change was approximately 14.9 hectares over the period of 1978 - 1994, 2.95 hectares (1994-2003), and

0.83 hectares (2003-2013), respectively. Hence, the total area of sedimentation from 1978 to 2013 was 18.68 hectares (see Table 3), with the distance from the coastline to the mainland (zone A) changing from 1.4 km over the period 1978-1994, 32 m in (1994-2003) and 88 m in 2003-2013. Hence, the total change in the distance from the sea to the mainland was 1.5 km.

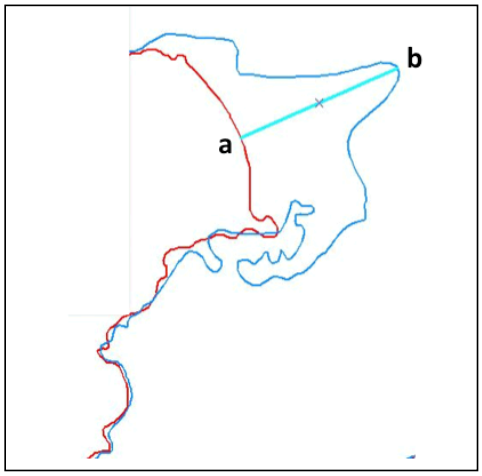


Figure 10. Coastlines changes 1978-2013 in the West Channel (A) sample area
(*Source:* Author's own results, 2014; afer MSS, TM, ETM and Landsat 8 images map, System coordinates WGS 1984 zone 49s)

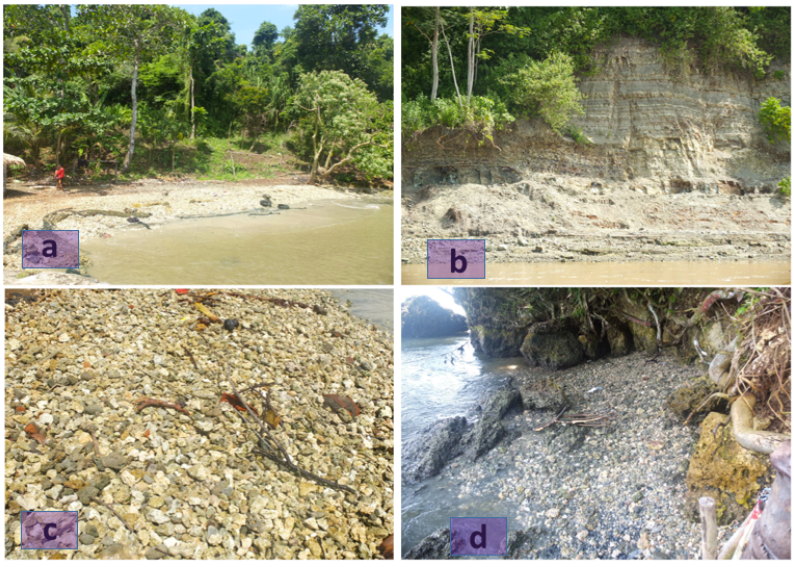


Figure 11. Geomorphological processes in Segara Anakan Lagoon
(*Photo:* Author's, 2013)

Figure 11 shows the abrasion, erosion, and deposition processes in the area of investigation. Based on the observation and analysis of type of rocks, it could be described that: (a) Alluvial plains, found on the northern coast; (b) Erosion in the river valleys in the southern part of Nusa Kambangan Island, which consists of mud, sand, gravel and clay results deposition of broken breccia stone; (c-d). Deposition of limestone and marl in the Northern and Southern Coast of Segara Anakan.

Morphodynamics and Implications

SA is a lagoon located in Cilacap in the Central Java Province. SAL has a very important function as an estuary for rivers such as Citanduy Cibereum, Palindukan, Cikonde, and other small tributaries. SAL was once a lagoon with a unique ecosystem consisting of the water body (the lagoon), which is brackish, mangrove forests, and the lowlands that are tidally influenced. For this reason, the lagoon ecosystem of SA has naturally served as a spawning ground for shrimp and fish, as well as a habitat for migrant and non-migrant water birds, various types of reptiles and mammals, and different species of flora. However, the lagoon ecosystem of SA has recently experienced severe environmental degradation caused by the high rates of river sedimentation and human activities in the form of land uses that do not take into account sustainable development.

The classification of multitemporal remote sensing images has been presented. Such an approach allows the classification of a remote-sensing image acquired for a specific geographical area at a given time, in the cases where training data are not available. The classification is performed using the statistical parameters estimated for an image acquired in the same area before the one under analysis. In this case, the results of the classification for morphodynamics and land use by the Landsat interpretation were used to analyse the spatial problems of the SAL area.

The Multitemporal analysis of the morphology and land use changes using remote sensing data is a method of analysis that is both effective and efficient (Klemas, 2011). With this analysis-based method, the multitemporal remote sensing of multitemporal images with a medium spatial resolution was used to derive information on landforms, land cover or land use. The details derived from the information can be tailored to the characteristics of the ability of the remote-sensing image data and standardised information needs and objectives, as well as the level of analysis to be performed and determined by the method of analysis used. In this study, the overall accuracy of the results was found to be 85 to 96% for the thematic accuracy of the results and interpretation of the lagoon's landform classification at a spatial resolution of 30 m.

SAL is a product of tectonic activities taking place within the depression zone. The formation of the SA waters took place because it is on the lower part below sea level. SA was formed from dikes and rocks composed essentially of sandstone from Tapak formation, from the upper Miocene - Pliocene. The conditions of the landscape of this area take on the form of the Southern-Mountains, including Nusa Kambangan, Jampang Formation, Pamali Formation, and Pamutuan Formation. These three formations are much older than the Oligo-Miocene period, as can be seen from the bedrock that is far below Tapak Formation.

The morphological condition of the lagoon changes in line with time and processes occurring within the area. From the data processing and results obtained, the water areas of SAL are shrinking from year to year. The shrinkage indicator of the water area of SAL can be seen from the emergence of new land or an increase in the extent of land that had previously been in the lagoon, as well as a reduced water area of SAL. Every year Citanduy, Cimeneng and Cikonde Rivers carried 5 million m^3 and 770,000 m^3 of sediment, of which 740,000 m^3 and 260,000 m^3 were deposited in SA, resulting in increased siltation of new land area in the waters of the lagoon. Thus, these conditions did not only affect natural communities and habitats, but also the culture of the people of Kampung Laut. For example, fishermen used to build their houses on stilts above the sea, but most of them now build brick houses on the mainland.

Based on the morphological comparisons of SAL from the Landsat image interpretations in 1978, 1994, 2003, and 2013 (see Figure 5 and Table 1), it can be clearly seen that the morphological changes of the lagoon have occurred in the western part of the lagoon. The area of the lagoon narrowed due to intensive sedimentation from the river Citanduy that will eventually create a lagoon in the future.

Extensive lagoon narrowing can also cause a reduction in the area of mangrove forests. Because of the equally important elements of lagoon in maintaining biodiversity and marine life, its presence must therefore be conserved and preserved. One solution to overcome this problem is to divert the course of the river Citanduy, which minimises the number of streams carrying sediment directly into the Indian Ocean. However, this method is also controversial as it results in other areas being affected by the rate of sedimentation. In addition, sediment dredging has been carried out in a very shallow area. Dredging is done so that the flow of tide will enter the lagoon because SAL is also influenced by the tides of the Indian Ocean through the western and eastern channels.

Sedimentation causes a variety of problems such as the declines in fishermen's incomes, conversion of mangroves into agricultural land, emergence of conflicts of interest between the local community and the Forest Agency regarding the use of land accretion, and the threat to SA from offshore fishing. Accreted land also causes disputes between those who want to keep it as a forest and the local government who wants to convert it for cultivation. At the rate of land accretion at approximately 22.5 ha/year due to sedimentation, this creates a high potential for conflict.

Sedimentation from the soil resulted in legal land tenure. In contrast, damages to the coastal areas due to abrasion in areas that are less stable against water erosion caused a critical state of the land, damaging road infrastructure. The process of coastal erosion (abrasion) in the western part of SA lasted quite a while, so the shoreline has basically retreated quite far away from the old shoreline mainland, which at this time is approximately 2 km from the edge of the sea. The coastline generally changes from time to time according to the changes in natural activities such as that of the waves, wind, tides and currents and river delta sedimentation area.

Inland and coastal sediments are essentially dynamic, moving according to the dimensions of space and time. Breaking waves, tidal streams, rivers, coastal vegetation and human activities are the factors that cause changes to the dynamics of the coast to establish a new equilibrium beach. Not every coastal area can respond to the whole process of changes, depending on

several factors such as the type of sediment, as well as morphology and the geology of the beaches (Stephenson & Brander, 2003).

Symptoms of shoreline change need urgent attention, given the major impacts on social life and the environment in order to realistically determine the likelihood of land use changes of the coastal areas of SA. Shoreline changes on the whole have changed from time to time in line with the changes in natural activities such as that of waves, wind, tides and currents and river delta sedimentation area (Davies, 2011).

Shoreline changes also occur due to the interferences to coastal ecosystems such as the construction of dikes and canals, as well as the buildings that surround the beach. Coastal mangrove forests, as a buffer function, have been greatly revamped to serve as the regional farms, residential and reclaimed areas that result in changes in the coastline. The development of the coastline, based on the pattern of sedimentation on the West Coast of SA, is likely to cause the formation of a bay.

CONCLUSION

The multitemporal satellite image can be used to evaluate the morphodynamic of the SAL area. The morphodynamics of coastal lagoons was obtained by assessing various Landsat images. During the period of 1978-2013, there were several changes in the morphology of the lagoon. These could be observed by the declining lagoon areas, which is caused by the sedimentation from Citanduy, Cimeneng, and Cibeureum Rivers, as well as other tributaries surrounding the SAL area. The coastline changes during 1978 – 2013 were due to the processes of erosion and sedimentation alternated within a relatively close distance and were parts of the dynamics of the beach. The total change of distance from sea to the mainland was 1.5 km. Coastline changes occurred in the western part of the lagoon are due to the currents and waves that enter the western channel of SAL, which is being greatly restrained by an accumulation of the sedimentation in the western part of the lagoon. Thus, the Indian Ocean tidal waves veer towards the edge of the mouth of the lagoon, while a deposited area has also been formed surrounding the mouth of the lagoon with the total area of 18.68 hectares of sedimentation from 1978 to 2013.

Resource degradation is expressed by the emergence of the lagoon's ecosystem damage. Seasonal factors such as the emergence of oceanographically characteristics also become the factors that accelerate sedimentation. In addition, the soluble waste that empties into the lagoon system also influences the lagoons ecosystem damage. Consequently, synergic coordination measures between upstream and downstream regions are necessary in the future.

REFERENCES

Alves, A. L., Venerando, E. A., & Helenice, V. (2003). Proceedings of the Brazilian symposium on sandy beaches: Morphodynamics, ecology, uses, hazards and management. *Journal of Coastal Research, 35*, 279-283.

Amos, C. L., Umgiesser, G., Tosi, L., & Townend, I. H. (2010). The coastal morphodynamics of Venice Lagoon, Italy: An introduction. *Continental Shelf Research, 30*, 837–846.

Ardli, E. R., & Wolff, M. (2009). Land use and land cover change affecting habitat distribution in the Segara Anakan lagoon, Java, Indonesia. *Regional Environmental Change, 9*(4), 235-243.

Baldock, T. E., Weir, F., & Hughes, M. G. (2008). Morphodynamic evolution of a coastal lagoon entrance during swash overwash. *Geomorphology, 95*(3), 398-411.

Bird, E., & Lewis, N. (2015). *Beach renourishment*. New York: Springer.

Davidson-Arnott, R. (2010). *Introduction to coastal processes and geomorphology*. Cambridge: Cambridge University Press.

Davies, M. (2011). *Geomorphic Shoreline Classification of Prince Edward Island*. Ottawa: Coldwater Consulting Ltd.

Duck, R. W., & da Silva, J. F. (2012). Coastal lagoons and their evolution: a hydromorphological perspective. *Estuarine, Coastal and Shelf Science, 110*, 2-14.

Jennerjahn, T. C., & Yuwono, E. (2009). Segara Anakan, Java, Indonesia, a mangrove-fringed coastal lagoon affected by human activities. *Regional Environmental Change, 9*, 231-233.

Klemas, V. (2011): Remote sensing techniques for studying coastal ecosystems: an overview. *Journal of Coastal Research, 27*(1), 2-17.

Magni, P., De Falco, G., Como, S., Casu, D., Floris, A., Petrov, A. N., & Perilli, A. (2008). Distribution and ecological relevance of fine sediments in organic-enriched lagoons: the case study of the Cabras lagoon (Sardinia, Italy). *Marine Pollution Bulletin, 56*(3), 549-564.

Nandi. (2014). Coastal Conservation Policies and Integrated Coastal Zone Management (ICZM) in Indonesia. *International Journal of Conservation Science, 5*(3), 387-396.

Nichols, M. M., & Boon, J. D. (1994). Sediment transport processes in coastal lagoons. *Elsevier Oceanography Series, 60*, 157-219.

Stephenson, W. J., & Brander, R. W. (2003). Coastal geomorphology into the twenty-first century. *Progress in Physical Geography, 27*(4), 607-623.

Sulistyo, B. (2011) *Remote sensing digital.* Yogyakarta: Locus.

Woodroffe, C. D. (2002). *Coasts: Form, process and evolution.* Cambridge: Cambridge University Press.

Zonta, R., Guerzoni, S., Pérez-Ruzafa, A., & de Jonge, V. N. (2007). Measuring and managing changes in estuaries and lagoons: morphological and eco-toxicological aspects. *Marine Pollution Bulletin, 55*, 403-406.

Pertanika J. Sci. & Technol. 26 (1): 427 - 440 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Multi-Hop Wireless Sensor Network Performance and Energy Simulation

Lilik Hasanah[1]*, Heru Yuwono[2], Ahmad Aminudin[1], Endi Suhendi[1], Yuyu Rachmat Tayubi[1] and Khairurrijal[3]

[1]Departmen Pendidikan Fisika, Universitas Pendidikan Indonesia, Jl. Setiabudi, Bandung 40154, Indonesia
[2]ALC Group, Jl. Iskandarsyah II, Blok M Kebayoran Baru, DKI Jakarta, Indonesia
[3]Physics of Electronic Materials Research Division, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung 40132, Indonesia

ABSTRACT

Simulation of a multi-hop Wireless Sensor Network (WSN) with different topologies and analysis of its performance in terms of number of messages exchanged and energy usage was done in this study. Sensor nodes in the simulation were modelled after an Arduino hardware system equipped with compatible radio transceiver for communication. The sensor nodes were configured in two network topologies, grid and random topology, for performance comparisons. Network sizes varied between 9 nodes and 256 nodes. Simulation was stopped when the communication link between the sensor nodes and their sink node broke down. It was obtained that grid topology has better performance, especially in small network size. Moreover, when the number of nodes in the network is higher, the performance of random topology network exceeds the grid's performance. Nonetheless, the lifetime span of the sensor network does not depend on the networks size or topology, rather on the available energy in each of the sensor nodes. We also have successfully improved the energy consumption model to account for more parameters of radio transceiver used in a WSN node. The energy needed to turn on and off the radio transceiver plays a significant part in the energy consumption of the sensor node.

Keywords: Energy consumption, grid network topology, multi-hop routing, random network topology, wireless sensor network

ARTICLE INFO

Article history:
Received: 25 April 2017
Accepted: 28 November 2017

E-mail addresses:
lilikhasanah@upi.edu, lilik-hasanah@yahoo.com (Lilik Hasanah),
heru.yuwono@gmail.com (Heru Yuwono),
aaminudin@upi.edu (Ahmad Aminudin),
endis@upi.edu (Endi Suhendi),
rachmat@upi.edu (Yuyu Rachmat Tayubi),
krijal@fi.itb.ac.id (Khairurrijal)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

General usage of a Wireless Sensor Network (WSN) is for environmental monitoring of a specific area of interest. Researchers need to collect data from sensors that are installed

in the area for a certain time duration (Culler, Estrin, & Srivastava, 2004). Development of Wireless Sensor Network (WSN) has been improved by the presence of alternative smart hardwares like Arduino and Raspberry Pi. It replaces traditional hardware such as Berkeley or Mica motes, especially for general purposes sensor nodes. Arduino offers low-cost, open hardware to researchers in studying WSN and it can be scaled up to hundreds or thousands of nodes at reasonable costs. Before this can be done, however, it is important to create a simulation environment or a simulator for studying sensors network with different sizes, topologies and routing strategies.

Previously, some researchers have done research on a wide range of simulation tools such as WSN to enable researchers to choose the most competent tool for the simulation of WSN and test the proposed research (Nayyar & Singh, 2015). At present, available WSN simulators only supports specific set of hardwares, but simulators with the support of Arduino hardware are not widely available. OMNET++ (http://www.omnetpp.org) is a family of libraries that can be used for general network simulation. Castalia, (http://castalia.npc.nicta.com.au), made by National ICT of Australia, is a network simulator derived from OMNET++. Meanwhile, TOSSIM (http://docs.tinyos.net/index.php/TOSSIM) is a discrete simulator for the sensor network that uses TinyOS for its operating system and Raspberry Pi's based hardware. NS-2 (http://www.isi.edu/nsnam/ns/) is a discrete simulator that focuses on network research. However, those simulators do not specifically support WSN. The simulator that supports Arduino-based processors is ATEMU, but the development of this simulator has been stopped by its developers.

Similar research on energy consumption of WSN nodes explored various strategies to achieve better energy usage. Some reports used coloured Petri-Net (CPN) to create power consumption model of WSN and perform energy usage simulation (Dâmaso, Freitas, Rosa, Silva, & Maciel, 2013). Li et al. (2014) proposes an event-driven QPN (Queueing Petri Net)-based modeling technique to simulate the energy behaviors of nodes. Some reports achieved longer network lifetime results by efficiently using available energy, which was obtained by controlling and maintaining the WSN topology (Zabi, Yousuf, & Manikonda, 2014). Meanwhile, some groups described the energy consumption of WSN on various layers of the WSN model (AboZahhad, Farrag, & Ali, 2015). However, while the above strategies utilise multi-hop routing and can be easily implemented in simulations, they are not easy to be applied in a real WSN.

In this research, we proposed a simpler approach to achieve energy efficiency, while maintaining the degree of applicability low by using decision-tree-based multi-hop routing strategy and Dijkstra algorithm to determine the shortest path. In order to determine the energy consumption of WSN, the calculation used is as presented in the previous work of AboZahhad et al. (2015).

In our previous research, the characteristics of SiGe-based nanoelectric devices were simulated as supporting the WSN's system (Hasanah, Abdullah, & Winata, 2008; Hasanah, Noor, Jung, & Khairrurijal, 2013). In this research, the simulation of a multi-hop Wireless Sensor Network with different topologies and analysis of its performance were done in terms of the number of messages exchanged and energy usage. To perform the simulation, a new

WSN simulator that supports hardware was developed using the Matlab programming tool. The simulator supports specific parameters derived from Arduino's electrical characteristics, network topologies, and multi-hop routing strategy. The sensor nodes in the simulation were modelled after the Arduino hardware system which was equipped with a compatible radio transceiver for communication. Meanwhile, the multi-hop routing strategy was employed and messages exchanged between nodes were counted. In addition, energy used by the sensor node to send and receive message was calculated. Using these data, the performance and energy usage of the WSN network can therefore be assessed.

METHODS

Multi-hop Routing

In general WSN application, each sensor node is equipped with limited energy source like small battery. Using continuous energy supply is not practical for WSN application, except for the sensor node that acts as a collector for data sent by all other nodes in the network, or which is usually called sink node. Due to this energy limitation, the sensor nodes have to be smart enough to manage its energy usage during normal operation with proper routing strategy.

Routing is a process of finding the optimum way delivering data from a start point to the destination. There are several protocols that can be used for routing such as LEACH, HEED, PEGASIS, TEEN, and APTEEN (Milan & Moravek, 2011). Routing with LEACH and HEED is usually considered as single-hop routing. One disadvantage with single-hop routing is that when the distance between the nodes increases, the energy needed to send data to other nodes increases siginificantly (Biradar, Sawant, Mudholkar & Patil, 2011; Farooq, Dogar, & Shah, 2012).

In this simulation, multi-hop routing was realised using tree-based routing (Milan & Moravek, 2011). This routing is a simplified way of general M-LEACH routing protocol, where a group head node or base station node was chosen prior to simulation. In this routing scheme, a sensor node sends message to the base station node through a series of neighbour nodes. A node can relay several messages from its neighbour nodes, which consumes more energy.

The simulation used Dijkstra algorithm in determining the closest distance between nodes in a graph structure. This algoritm was implemented in Matlab by grTheory toolbox. Then, it was used to develop our simulation software.

Energy Consumption Model

After the routes from every node to base station nodes are clearly defined, the number of messages processed by each node can then be calculated (Milan & Moravek, 2011). In Figure 1, node n_1 sends data to the next nodes until node n_4. Node n_2 receives the data, performs checking on route information, and then sends the data to node n_3. Sensor nodes use omnidirectional antenna for radio communication, so the same data sent by n_2 are also received (overhear) by n_1. By node n_1, the received message is ignored because the data are not intended for n_1.

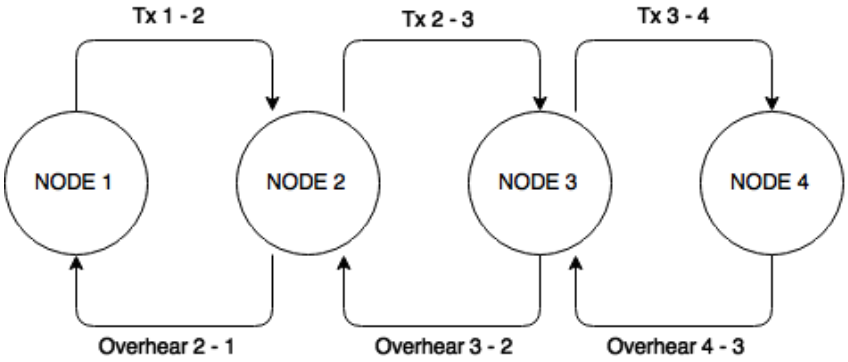


Figure 1. Data Transmission Diagram

Thus, node n_1 uses energy to send the initial message ETX and energy E_{RX} to receive overhear message from n_2. The energy used by n_1 is:

$$E_{n1} = E_{TX} + E_{RX} \quad (1)$$

As relay nodes, n_2 and n_3 use energy E_{RX} to receive messages from the previous node, E_{TX}, to send the relayed message, and E_{RX} to receive overheard message. Thus, energy used by relay nodes is:

$$E_{nr} = E_{TX} + 2E_{RX} \quad (2)$$

The last node n_4 as base station node is only responsible for receiving messages, so the energy used is equal to E_{RX}.

$$E_{SB} = E_{RX} \quad (3)$$

Energy consumption for a sensor node is defined as (AboZahhad et al., 2015):

$$E_{tran} = P_{ont}T_{ont} + P_{onr}T_{onr} + P_{tr}T_{tr} + P_{sm}T_{sm} + P_{dc}T_{on} \quad (4)$$

where the total time duration is divided between the communication active mode T_{on}, transient mode T_{tr} and sleep mode T_{sm}. During sleep mode, the leaking current of switching transistors dominates the power consumption P_{sm}. This term is often neglected, i.e. P_{sm} can be set to zero. The transient-mode time arises mainly from the frequency synthesiser settling time, and the settling time for other devices such as a mixer and power amplifier can be neglected and power consumption at this mode is P_{tr}. During the active mode, power is consumed at digital circuits (P_{dc}), analog circuits at the transmitter (P_{ont}) and receiver side (P_{onr}). P_{dc} is usually neglected

because power consumption of a digital signal processing is relatively small compared to that of the analogue circuits, especially in a wireless radio application. Equation (4) then can be simplified as follows:

$$E_{tran} = P_{ont}T_{ont} + P_{onr}T_{onr} + P_{tr}T_{tr} \quad (5)$$

Note that in Equation (5), every sensor node is treated equal so that its energy consumption is calculated by adding the energy needed to transmit and receive radio, as well as the energy needed by radio transceiver during settling time (time needed to change from standby state to working state).

In our research, we expanded equation (5) by treating sensor node depending on its function during the operation of WSN. For the sensor node that only transmits messages as shown by Node 1 in Figure 1, its energy consumption is defined by equation (1). E_{TX} and E_{RX} can be defined as:

$$E_{TX} = P_{ont}T_{ont} + P_{tr}T_{tr} + P_{tr,down}T_{tr,down} \quad (6)$$

$$E_{RX} = P_{onr}T_{onr} + P_{tr}T_{tr} + P_{tr,down}T_{tr,down} \quad (7)$$

Where T_{tr},down is the time needed by the transceiver sistem to change from transmit/receive state to standby state. By using equation (6) and (7) into equation (1), we can obtain:

$$E_{tran,source} = P_{ont}T_{ont} + P_{onr}T_{onr} + 2(P_{tr}T_{tr} + P_{tr,down}T_{tr,down}) \quad (8)$$

Using the same approach, equation (2) for relay node can be expanded into:

$$E_{tran,relay} = P_{ont}T_{ont} + 2P_{onr}T_{onr} + 3(P_{tr}T_{tr} + P_{tr,down}T_{tr,down}) \quad (9)$$

And finally, equation (3) will change into:

$$E_{tran,BS} = P_{onr}T_{onr} + P_{tr}T_{tr} + P_{tr,down}T_{tr,down} \quad (10)$$

By using equations (7), (8), (9), we developed a WSN simulator using MATLAB to investigate a sensor node's energy consumption. By differentiating sensor node's functions as a source, relay, and base station node, we can learn more about how energy consumption varies between nodes in a WSN.

Hardware used to build sensor node is Arduino Uno R3 with nRF24L01 radio transceiver module. In this simulation, energy consumption by Arduino is not included in the calculation because its value is the same for all the sensor nodes. The simulation focuses on the energy used for data communication between the nodes. Important parameters from nRF24L01 radio transceiver model were obtained from its factory datasheet, as shown in Table 1. The important parameters are electrical current needed for transmitting data at 11,3 mA, receiving data at

11,8 mA, and for standby at 22 uA. Meanwhile, the time needed to change between standby state to ready-to-transmit state and ready-to-receive data is 130 uS (micro seconds), and the same value for the opposite state changes. Other important parameter is data transfer rate of the radio transceiver, which is at 1 Mbps (Milan & Moravek, 2011).

Table 1
nRF24L01 radio transceiver parameters according to factory datasheets

Parameter	Value
T_{tr}	130 μS
$T_{tr,down}$	130 μS
$I_{tr,receive}$	8.3 mA
$I_{tr,down,receive}$	8.3 mA
$I_{tr,transmit}$	8.0 mA
$I_{tr,down,transmit}$	8.0 mA
Io_{nr}	11.8 mA
I_{ont}	11.3 mA
$T_{onr} = T_{ont}$	0.4 mS
I_{rxtx}	8.0 mA
T_{rxtx}	130 μS

Length of data message sent by every sensor nodes was assumed to be similar to the length calculated by some reports (Milan & Moravek, 2011). In IEEE 802.15.4 protocol, data payload length is 30 bytes. Sensor reading data are stored in this data payload. Additional information added into the message is Data Request message at 12 bytes length, while ACK message used as communication control at 5 bytes length. Thus, the total length of data messages is 47 bytes.

Simulation Scope

Simulation software was created using Matlab. For every network topology used, network size or the number of sensor nodes in the network varies according to n^2, where n=3, 4,.., 16 nodes. The software will create the network and assign all possible communication links between each node. The software assigns which node will act as base station node.

Area where the network resides was assumed to be 500 m wide and 500 m length. Nodes are placed algorithmically in the area according to the choosen topology. Nodes are placed in various points in the Cartesian coordinate which has its zero or centre point in the bottom left corner of the area.

Each sensor node was assumed to have 1 Joule of energy at the beginning of the simulation, except for base station node. Base station node has much more energy to prevent it from running out of energy. The next assumption is that Arduino's energy consumption is not included in the simulation.

RESULTS AND DISCUSSION

Network Performance Analysis

Simulation was run until all nodes that positioned at 1 hop before the base station node had run out of its energy. All messages sent, relayed and received by every node were recorded for further analysis. Figure 2 shows the simulation result for network with 9 sensor nodes.

For random topology in Figure 2, if node 5 is assigned as base station node, then nodes 2, 3, 4, and 9 are positioned at 1 hop before node 5. If all four of those nodes died, the simulation would stop. Similar rules applied to the grid topology in Figure 2.

As shown in Figure 2, the simulation stopped because all the sensor nodes in the network became dead at the same time. For comparison, on another simulation run with 9 sensor nodes and random topology with different nodes position, the simulation stopped when all the nodes positioned at 1 hop before the base station nodes died, as shown in Figure 3.

In Figures 3, nodes 2 and 3 died earlier than other nodes, which caused the simulation to be stopped at 113th run. Here, it could be seen that sensor nodes placement in the networks is very important to maximise the lifetime of the WSN network.

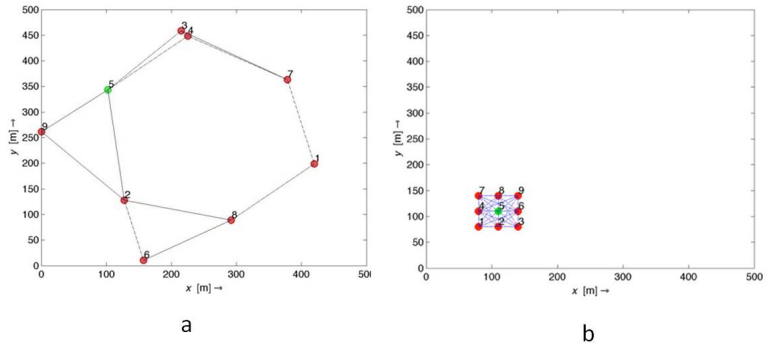


Figure 2. Random Topology (a) and Grid Topology (b) Every node is labeled with number. Both graph axes represent distance in metres. Red nodes represent dead node. Green nodes represent base station nodes

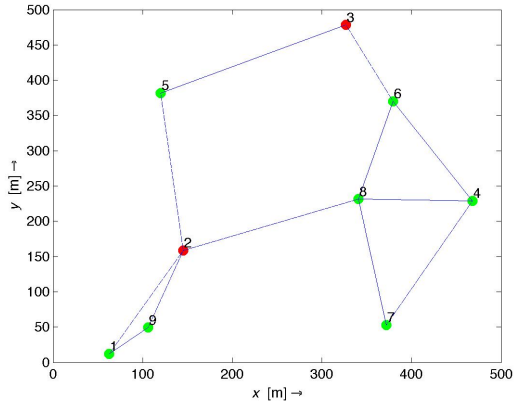


Figure 3. Simulation stopped because nodes 2 and 3 had run out of energy, closing the route to base station node 5

For network with random topology and 9 nodes, the simulation stopped at 222nd run. Grid topology had the same number of run as well. The difference between random and grid topology, however, is clearly shown in Figure 4.

It can be seen from Figure 4(a) that in random topology, nodes died gradually from the 100th run until 222nd run, whereby 6 nodes died at the same time. Different cases happened to the grid topology, where all 9 nodes died at the same time at 222nd run, as shown in Figure 4(b). This can be explained as follows. In the grid topology with 9 nodes, 8 nodes were placed at 1 hop before the base station node, as shown in Figure 2. Energy used by the 8 nodes is calculated using Equation (2). However, for the random topology shown in Figure 2, less number of nodes is placed at 1 hop before the base station node. These nodes act as relay nodes, in which the energy usage is calculated using Equation (3). The energy used for relaying messages is higher than energy used for sending messages, so relay nodes in random topology will run out of energy sooner than other nodes. The same pattern was also found to happen to network with higher number of nodes, with few differences.

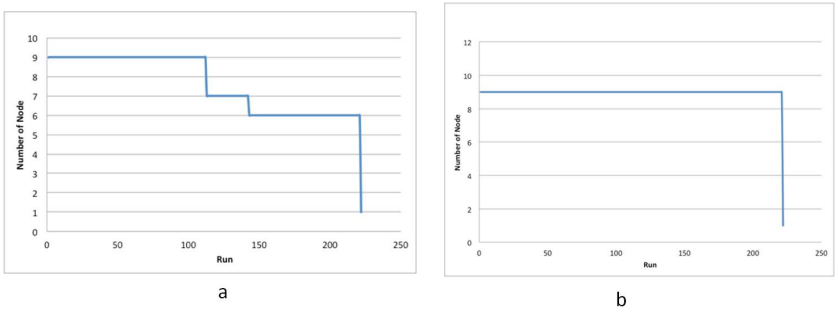


Figure 4. (a) Number of alive nodes at each simulation run for random topology with 9 nodes, and (b) number of alive nodes at each simulation run for the grid topology with 9 nodes

In Figure 5, for the random topology with 144 nodes, it can be seen that from around 140th run until 222nd run when the simulation stopped at the 82 run, there were around 80 nodes still up and running. These nodes were still sending messages, but they could not reach the base station node because there were no routes available.

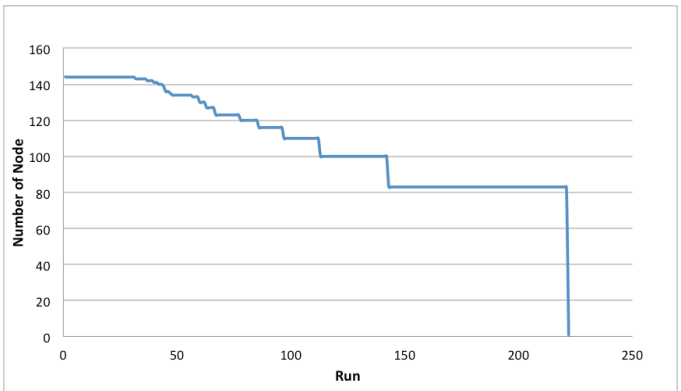


Figure 5. Number of alive nodes at each simulation run for the random topology with 144 nodes

In Figure 6, for the grid topology with 144 nodes, it can be seen that from around 140th run until 222nd run when the simulation stopped, there were only 60 nodes still up and running, which is lower compared to 80 nodes for the random topology. The WSN network with random topology seems to have advantage especially for a large number of sensor nodes compared to the grid topology network.

Interestingly, simulation stopped at the same run, i.e. 222nd run, for both the random and grid topology and was not affected by the number of nodes in the network. Using this fact, it can be concluded that the lifetime of the WSN network does not depend on the size of the network and topology, but rather determined by the available energy in each sensor nodes.

From Figure 4 through Figure 6 above, we can analyse which topology performs better than the other. As stated before, the criterion to stop the simulation is if there is no more possible route to go from the sensor node to the base-station node, this is because all the nodes positioned 1 hop before base-station node are run out of energy. Thus, we can calculate the number of messages sent by sensor nodes and received by base-station node. The topology with a higher number of received messages by base-station node for all simulation runs is the best topology. Table 2 shows the number of messages received by the base-station node.

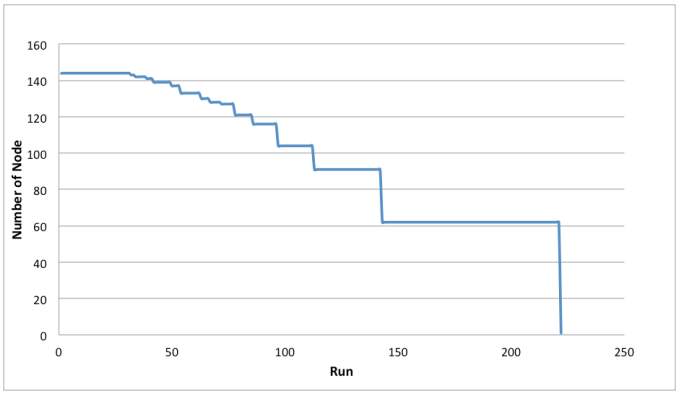


Figure 6. Number of alive nodes at each simulation run for grid topology with 144 nodes

From Table 2, it is clear that with higher number of nodes in a WSN, there are more messages received by the base-station node. In the grid topology, the number of messages received by base-station node is higher than in the Random topology, except for the number of sensor nodes above 144. This is because in a network with more than 144 nodes and with the same area size, more nodes using the same routes cause the intermediate nodes to run out energy quickly, especially in the grid topology. However, if we calculate the average of received messages compared to the number of nodes available in the WSN (Table 2), then what we can obtain for the same topology is that the number of messages per node is almost identical regardless of the number of nodes in the WSN. In the random topology, the average number of messages/node is 167 messages/node, whereas the average number is 184 messages/node in the grid topology. We summarised these in Figure 7.

Table 2
Number of messages received by the base-station node for Random and Grid topology

Number of Nodes	Topology					
	Random			Grid		
	Number of Message	Run	Message/ Node	Number of Message	Run	Message/ Node
9	1479	222	164.33	1776	222	197.33
16	2630	222	164.38	3172	222	198.25
25	4435	222	177.40	5012	222	200.48
36	5896	222	163.78	6874	222	190.94
49	8463	222	172.71	9904	222	202.12
64	10962	222	171.28	11354	222	177.41
81	13501	222	166.68	14721	222	181.74
100	16337	222	163.37	16766	222	167.66
121	19826	222	163.85	20658	222	170.73
144	24044	222	166.97	22120	222	153.61
169	26285	222	155.53	26866	222	158.97
196	30928	222	157.80	29267	222	149.32
225	35729	222	158.80	34044	222	151.31
256	40494	222	158.18	35772	222	139.73

From Figure 7, it can be seen that the smaller number of sensor nodes, the higher the number of messages/node for grid topology compared to the random topology. As shown in Figure 7, however, the graph for grid topology has a significant and sudden decreasing pattern than that of the random topology, and for the number of nodes above 200, the graph for the grid topology is lower than the random topology. We can also see in Figure 7 that the graph for random topology is more consistent compared to the grid topology for all range of the number of nodes.

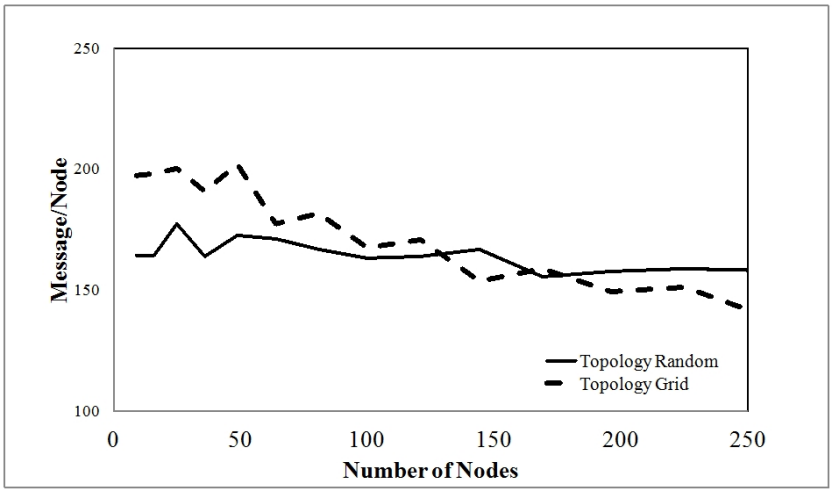


Figure 7. Number of messages per node in Random (blue line) and Grid (red line) topology

Energy Consumption Analysis

For the network with random topology and 9 nodes, simulation stopped at 222nd round. The energy consumed by each nodes during the whole run, calculated using Equation (5), (9), (10), and (11) above, was recorded and plotted. For a source node with energy consumption defined in Equation (9), the energy consumption Etran,source compared to Etran calculated using Equation (5) is shown in Figure 8.

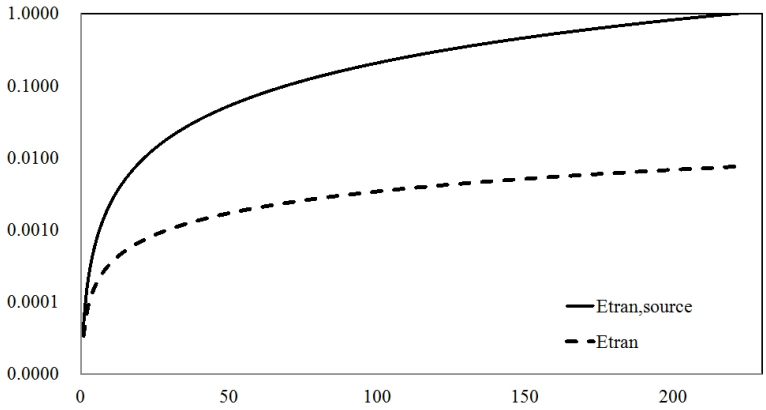


Figure 8. Energy consumption of a source node. Energy use shown is in the logarithmic scale

It can be seen from Figure 8 that the energy consumption for a source node in our simulation differs greatly from the energy consumption calculated in Milan and Moravek (2011). The difference comes from the additional term of $P_{tr,down}T_{tr,down}$ introduced in Equation (9) as compared to the terms used Equation (2) above. The added term represents energy needed by the transceiver chip to go from the working state (transmit or receive state) to standby state, or in other words, to sleep. $P_{tr}T_{tr}$ represents the energy needed to go from the standby state to working state or to wake up state.

For relay node shown in Figure 9, the difference between $E_{tran,relay}$ and E_{tran} become more significant. $E_{tran,relay}$ shown in Equation (10) uses more energy to receive messages and to wake up and sleep. Relay nodes process more messages compared to source node. It needs to wake up to receive messages from its neighbouring nodes, check its destination, transmit the messages to the next node, and then goes back to sleep. That is why relay node spends much of its energy to wake up and sleep. It can be seen in Figure 5 that the relay nodes run out of energy at 97th round of simulation compared to the simulation of source node (Figure 4) that stops at 222nd round. On average, the relay node used 58.69 times more energy than the source node.

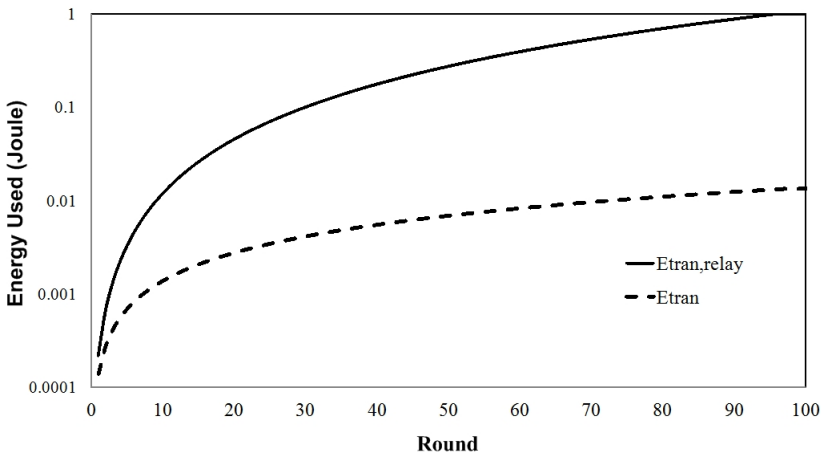


Figure 9. Energy consumption of a relay node. The energy use shown is in the logarithmic scale

Figure 10 shows the energy consumption of a base station node, the node that receives all messages transmitted by source nodes in the network. Again, the difference between $E_{tran,BS}$ and E_{tran} is significant. To see which type of sensor node (source, base-station, relay node) consumes more energy, the energy profile of each type is plotted in Figure 11.

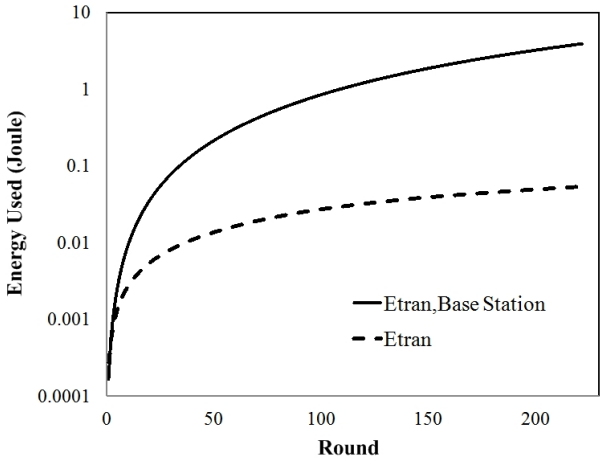


Figure 10. Energy consumption of a base station node. Energy use shown is in the logarithmic scale

As shown in Figure 11, energy consumption of the relay node is the highest compared to the other nodes, followed by the base station node and source node. From the second and third term of Equation (10) above, we can see that the relay nodes spent most of its energy to receive messages and wake up and sleep, respectively. Depending on how many messages it needs to relay to the other nodes, the relay node consumes more energy compared to other types of node.

Based on Figure 11, energy consumption of the base station node is significantly higher than the source node. Base station node has a duty to receive all messages from all the other nodes, so it spend most of its energy in receiving messages. According to Table 1 above, I_{onr} value or current in the receive mode is bigger than I_{ont} or current in the transmit mode, indicating that it needs more energy to receive messages than to transmit them. Furthermore, current is needed to wake up the radio transceiver to receive state Itr,receive is higher than the current needed to wake up to transmit state $I_{tr,transmit}$. This means it needs more energy to wake up into receive state than that to $t_{ransmit}$ state. The same thing happens to $I_{tr,down,receive}$ and $I_{tr,down,transmit}$, which need more energy to sleep from the receive state than that from transmit state. Compounding all the effects into the total $E_{trans,BS}$ we can see that energy consumption of base station nodes is higher than source node.

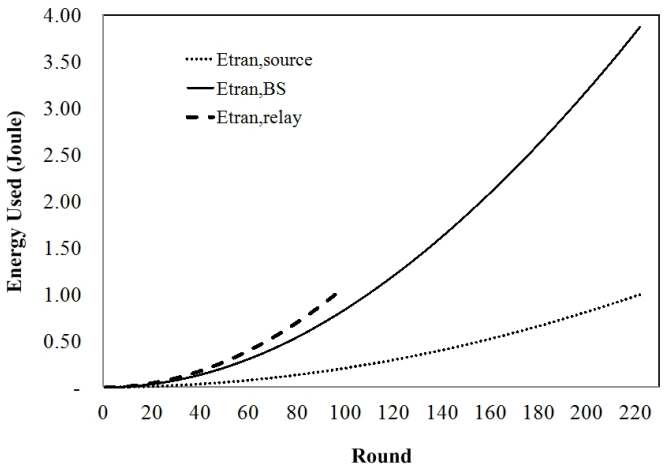


Figure 11. Energy consumption of various node types (source, base station, relay). Energy Use shown is in te logarithmic scale

CONCLUSION

Simulation of the WSN network using multi-hop routing protocol was done in this study. For the random topology network, sensor nodes placement is very important to maximise the lifetime of the network. Relay nodes consume more energy than ordinary nodes, so sensor placement that efficiently reduces the number of relay nodes in the network is needed.

It was found that the grid topology has better performance, especially in small network size. However, when the number of nodes in the network increases, the performance of Random topology network exceeds the grid's performance.

The lifetime of the WSN network does not depend on the size of the network and topology, but rather determined by the available energy in each sensor node. We also have successfully improved the energy consumption model proposed in the previous research to account for more parameters of radio transceiver used in a WSN node. The energy needed to wake up and sleep the radio transceiver plays a significant part in the energy consumption of the sensor node.

ACKNOWLEDGEMENT

The authors would like to thank to the Ministry of RISTEKDIKTI, Indonesia, who supported this research through its Hibah Bersaing research grant for 2015 – 2016 period.

REFERENCES

AboZahhad, M., Farrag, M., & Ali, A. (2015). A comparative study of energy consumption sources for wireless sensor networks. *International Journal of Grid and Distributed Computing, 8*(3), 65-76.

Biradar, R. V., Sawant, S. R., Mudholkar, R. R., & Patil, V. C. (2011). Inter-intra cluster multihop-LEACH routing in self-organizing wireless sensor networks. *Networks, 12*(13), 14.

Culler, D., Estrin, D., & Srivastava, M. (2004). Overview of sensor networks. Special issue in sensor networks. *IEEE Computer, 37*(8), 41-49.

Dâmaso, A., Freitas, D., Rosa, N., Silva, B., & Maciel, P. (2013). Evaluating the power consumption of wireless sensor network applications using models. *Sensors, 13*(3), 3473-3500.

Farooq, M. O., Dogar, A. B., & Shah, G. A. (2010, July). MR-LEACH: multi-hop routing with low energy adaptive clustering hierarchy. In *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on* (pp. 262-268). IEEE.

Hasanah, L., Abdullah, M., & Winata, T. (2008). Model of a tunneling current in an anisotropic Si/Si1−xGex/Si heterostructure with a nanometer-thick barrier including the effect of parallel–perpendicular kinetic energy coupling. *Semiconductor Science and Technology, 23*(12), 125024.

Hasanah, L., Noor, F. A., Jung, C. U., & Khairrurijal, K. (2013). Verification of theoretical model for collector current in SiGe-based heterojunction bipolar transistors. *Electronic Letters, 49*(21), 1347-1348

Li, J., Zhou, H. Y., Zuo, D. C., Hou, K. M., Xie, H. P., & Zhou, P. (2014). Energy consumption evaluation for wireless sensor network nodes based on queuing Petri net. *International Journal of Distributed Sensor Networks, 10*(4), 1-11.

Milan, S., & Moravek, P. (2011). Modeling of energy consumption of Zigbee devices in Matlab tool. *ElektroRevue, 2*(3), 41-46.

Nayyar, A., & Singh, R. (2015). A Comprehensive review of simulation tools for wireless sensor networks (WSNs). *Journal of Wireless Networking and Communications, 5*(1), 19-47.

Zabi, A., Yousuf, T., & Manikonda, A. (2014). A new method for controlling and maintaining topology in wireless sensor networks. *International Journal of Computer Networks and Communications, 6*(4), 91.

Pertanika J. Sci. & Technol. 26 (1): 441 - 462 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Parallel Exponential Smoothing Using the Bootstrap Method in R for Forecasting Asteroid's Orbital Elements

Lala Septem Riza[1]*, Judhistira Aria Utama[2], Syandi Mufti Putra[1], Ferry Mukharradi Simatupang[3] and Eddy Prasetyo Nugroho[1]

[1]Department of Computer Science Education, Universitas Pendidikan Indonesia, Jl. Setiabudhi, Bandung, Indonesia
[2]Department of Physics Education, Universitas Pendidikan Indonesia, Jl. Setiabudhi, Bandung, Indonesia
[3]Astronomy Research Division, Institut Teknologi Bandung, Jl. Ganesha, Bandung, Indonesia

ABSTRACT

Nowadays, large datasets become main intentions of researchers in many areas. However, a challenge that still remains mainly unresolved is the lack of strategies used for analysing large time-series datasets in parallel. Therefore, this research aims to design a model of exponential smoothing working on parallel computing by using the bootstrap method. Three parts will be considered in the model: data pre-processing using the bootstrap methods, parallel exponential smoothing, and aggregation of results to be the final predicted values. To implement the processes, some packages available in the R environment such as "foreach", "forecast" and "doParallel" are utilised. R environment provides many packages for scientific computing, data analysis, time-series analysis and high performance computing. For testing and validating the proposed model and implementation, a case study in astronomy, i.e. the prediction of asteroid's orbital elements, was done. Moreover, a comparison and analysis with the results produced by algorithm of Regularized Mix Variable Symplectic 4 Yarkovsky Effect (RMVS4-YE) is also presented in this paper to provide a high level of confidence on the proposed model.

Keywords: Exponential smoothing, orbital element, parallel computing, R programming language, time series analysis

ARTICLE INFO

Article history:
Received: 25 April 2017
Accepted: 28 November 2017

E-mail addresses:
lala.s.riza@upi.edu, lala_s_riza@yahoo.com (Lala Septem Riza),
j.aria.utama@upi.edu (Judhistira Aria Utama),
syandi.mufti@student.upi.edu (Syandi Mufti Putra),
fmsimatupang@yahoo.com (Ferry Mukharradi Simatupang),
eddypn@upi.edu (Eddy Prasetyo Nugroho)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Currently, the flood of data can no longer be stopped and thus, it inundates every corner of our daily activities. In other words, data have been released in a high amount and at a high speed with complex formats and from various sources. Based on a report from International Data Corporation in 2011, the overall digital

data volume in the world was 1.8 zettabytes, which was expected to grow by nearly nine times within the next five years (Gantz & Reinsel, 2011). A survey by Troester (2015) revealed that Facebook handles more than 250 million photo uploads and the interactions of 800 million active users with more than 900 million objects (pages, groups, etc.) on a daily basis. Wal-Mart processes more than a million customer transactions each hour and imports those into databases that are estimated to contain more than 2.5 petabytes of data. This phenomenon offers two opposing sides. First, it brings emerging problems since available tools and algorithms have difficulties in handling such large data efficiently. At the same time, however, this condition offers great advantages if we are able to extract knowledge from data, such as in making a better decision, predicting a future action, and describing a current situation, etc.

The term 'Big Data' has been widely used for expressing the above phenomenon. It was introduced by computer scientists several years ago. In 2012, Gartner, Beyer & Laney (2012), stated that "Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." According to this definition, there are 3Vs (which are volume, velocity, and variety) that should be taken into account. Basically, there are two issues related to the volume of Big Data. First, current computers and algorithms cannot handle massive datasets efficiently. Second, current storage management systems are also facing the same issue. In relation to the second issue, which is the aim of this research, there are many techniques that can be used, including data sampling and memory managements (i.e., to create, store, access and manipulate massive matrices) by allocating shared memory and using memory-mapped files (Kane, Emerson, & Weston, 2013), parallel and distributed computing (Zomaya, 1996), and Big Data platform (Dean & Ghemawat, 2008; Murthy, Vavilapalli, Eadline, Niemiec, & Markham, 2013).

In this research, we attempted to design a model and implement it on parallel computing to deal with forecasting large-time series datasets in the R programming language. In order to apply for prediction, we modified exponential smoothing so that it could be run in parallel. In short, we constructed three phases in parallel exponential smoothing, as follows: (i) data pre-processing by utilising the bootstrap method, (ii) conducting exponential smoothing in a parallel way by using three software libraries in R: "forecast", "foreach", and "doParallel", and then (iii) aggregating results to obtain the final predicted values.

Moreover, some experiments are presented to validate the model. These experiments used datasets in astronomy, which are asteroid's orbital elements. The data involved more than 400,000 rows and 6 important parameters in Kepler Components (i.e., a, e, i, ω, Ω and M representing semi major axis, eccentricity, inclination, argument of perihelion, longitude of ascending node and mean anomaly, respectively). The position and velocity of celestial bodies in their orbit at a certain time is expressed by 4 elements orbits. The size and shape of the orbit are represented by the element of a and e, respectively. Meanwhile, the element of $\varpi(=\omega+\Omega)$, which is also known as the longitude of perihelion, specifies the orientation, and the other element, M, specifies the position or phase of celestial bodies in their orbit.

TIME SERIES ANALYSIS

Introduction to Time Series Analysis

Time series datasets can be generated in econometrics, finance, weather station, earthquake, astronomy, medical clinic, energy, and other domains. For example, in finance, we obtained time series datasets produced by the S&P500 daily stock index, as illustrated in Figure 1 below.

Figure 1. S&P500 Daily Stock Index (January 4, 2006, to December 16, 2016)

To study these datasets thoroughly, it is a common way to consider the logarithm return, which is defined as follows (Palma, 2016).

$$r_t = \log \frac{P_t}{P_{t-1}} = \log P_t - \log P_{t-1} \tag{1}$$

where P_t represents the price or the index value at time, t. The log returns are displayed in Figure 2, in which the great volatility could be seen as happening sometime in January 2009.

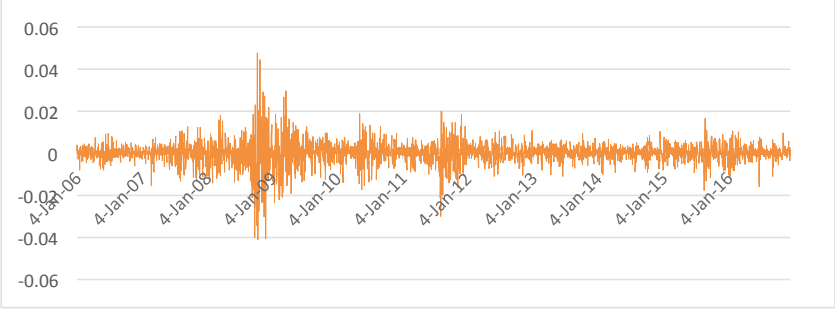


Figure 2. S&P500 Daily Log Returns (4 January 2006 to 16 December 2016)

According to the example above, we can define time series datasets as a time-ordered sequence of observation values of a defined variable at a certain time interval, Δt (Palit & Popovic, 2006). The data can be represented as a set of discrete values (i.e., $x_1, x_2, ..., x_n$). The main properties of time series are stationarity, linearity, trend, and seasonality. The first term means that the mean value and variance of data should be constant over time and the covariance value between x_t and x_{t-d} is dependent on the distance between those two data points and constant over time. Linearity refers to the sequence of observation values can be represented by a linear function. The trend component of time series datasets is a change on the local and global increases or decreases of data values in long-term periods. The last property of time series, which is seasonality, refers to changing on pattern periodically (Palit & Popovic, 2006).

Moreover, time series analysis focused on studying patterns or structures of the datasets so that we could make a better decision and solve the issues on the hand. Basically, the analysis involves the following activities: (i) definition and description of time series, (ii) model construction, (iii) forecasting or prediction of future data, and (iv) clustering to determine the characteristics of data (Riza, 2016b). In this research, we focused on forecasting or predicting of future data by considering historical data. Forecasting can be defined as a given set of observed values $x_1, x_2, x_3, ..., x_n$ of a time series, the future value $x_{n+1}, x_{n+2}, \ldots$, should be estimated (Palit & Popovic, 2006). Moreover, Palit and Popovic (2006) divided strategies on forecasting into the following groups: using trend analysis, regression approaches, the Box-Jenkins methods and smoothing methods. Firstly, the trend analysis utilises linear or nonlinear regression (e.g., quadratic and exponential functions). In regression analysis, we can perform a mathematical tool mapping input variables and its output. Recently, in addition to using approaches on the mathematics field, machine-learning techniques are conducted. For example, a research by Zhang (2012) provides a survey on neural network applications in time series forecasting such as air pollutant concentration, stock index option price, etc. A family of forecasting methods proposed by Box and Jenkins (Box, Jenkins, Reinsel, & Ljung, 2015) consists of Autoregressive Model (AR), Moving-average Model (MA), etc. The last group being focused in this paper, which is smoothing method, is a set of techniques based on reduction of irregularities or random fluctuations in time series data so that we could obtain a clean time series pattern out of contaminated observation data (Palit & Popovic, 2006). Furthermore, useful explanations on time series analysis can be found from in the works of Kirchgässner, Wolters, and Hassler (2012) and Derryberry (2014).

Smoothing Methods for Time Series Forecasting

As mentioned previously, smoothing can be defined as a technique by conducting reduction of irregularities or random fluctuations in time series datasets to obtain a clean time series pattern out of contaminated observation data (Palit & Popovic, 2006). There are two types of smoothing methods, namely moving-average smoothing and exponential smoothing.

The following is a formula of moving-average smoothing for prediction of future values:

$$x_m(t+1) = \frac{x(t) + x(t-1) + \cdots + x(t-n)}{n}. \quad (2)$$

It can be seen that the equation basically averages the past values for reducing the random variations present in *n* observation data. Another variant of the equation is by using weigh on each the past values, as illustrated in the following equation:

$$x_m(t+1) = w_1 x(t) + w_2 x(t-1) + \cdots + w_n x(t-n). \quad (3)$$

Thus, the method can be regarded as easy to understand and simple to use. However, in the case of the weighted moving average, it is difficult to choose the optimal values for each weight of the past values.

The second method, which is exponential smoothing, is a set of techniques where the weights are decreased exponentially as the observations get older (Hyndman, Koehler, Ord, & Snyder, 2008). According to the taxonomy of proposed by Pegels (1969), classification of exponential smoothing methods was obtained. It was improved by Gardner Jr. (1985), modified by Hyndman, Koehler, Snyder, and Grose (2002), and later extended by Taylor (2003), giving the fifteen methods in Table 1 below.

Table 1
Classification of exponential smoothing methods

Trend Component	Seasonal Components		
	N (None)	A (Adaptive)	M (Multiplicative)
N (None)	N, N	N, A	N, M
A (Adaptive)	A, N	A, A	A, M
A_d (Adaptive Damped)	A_d, N	A_d, A	A_d, M
M (Multiplicative)	M, N	M, A	M, M
M_d (Multiplicative damped)	M_d, N	M_d, A	M_d, M

There is another name for some cells such as follows:

- cell (N,N) is the simple exponential smoothing (or SES) method, which is defined as (Brown, 1959):

$$\hat{x}_{t+1} = \hat{x}_t + \alpha(x_t - \hat{x}_t), \quad (4)$$

  where α is a constant between 0 and 1.

- cell (A,N) is named Holt's linear method, which is defined as (Holt, 1957):

  ◦ Level:

$$l_t = \alpha x_t + (1-\alpha)(l_{t-1} + b_{t-1}), \quad (5)$$

- Growth:

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (6)$$

- Forecast:

$$\hat{x}_{t+h|t} = l_t + b_t h, \quad (7)$$

where α and β^* are a contant between 0 and 1.

For a complete list of the exponential smoothing included in Table 1, please refer to Hyndman et al. (2008).

A Brief Survey on the Implementations of Parallel Computing in Time Series Analysis

In this section, we present a brief review on some strategies and implementations of parallelisation of methods in time series analysis, especially for forecasting. It is presented to provide some related works instead of providing a comprehensive survey.

Table 2 shows a summary of the survey conducted in Scopus related to the following keywords: "time series and parallel computing" and "exponential smoothing and large data." It can be seen that researchers have been attempting to find suitable strategies to deal with large datasets or Big Data. For example, Big Data platform, MapReduce, is used in Mirko and Kantelhardt (2013), Sheng, Zhao, Leung and Wang (2013), etc., whereas Message Passing Interface (MPI) is utilised in the research conducted by Górriz, Algeciras, Puntonet, Salmerón, and Martin-Clemente (2004).

INTRODUCTION TO R AND ITS ECOSYSTEM

In this section, we briefly provide an introduction to R and its ecosystem. In addition, some examples of implementations R packages used for parallel computing and time series analysis will also be given in this section. Therefore, some reasons why we are using R in this research are presented.

Introduction to R Programming Language

R is an open-source programming language and software environment used for scientific computing, data analysis, visualisation, time series analysis, high performance computing, etc. (Ihaka & Gentleman, 1996). Furthermore, regarding a survey conducted by KDnuggets (Piatetsky, 2016), R takes the first place for the programming language used for an analytics/ data mining/data science in 2016 and 2015, as shown in Figure 3.

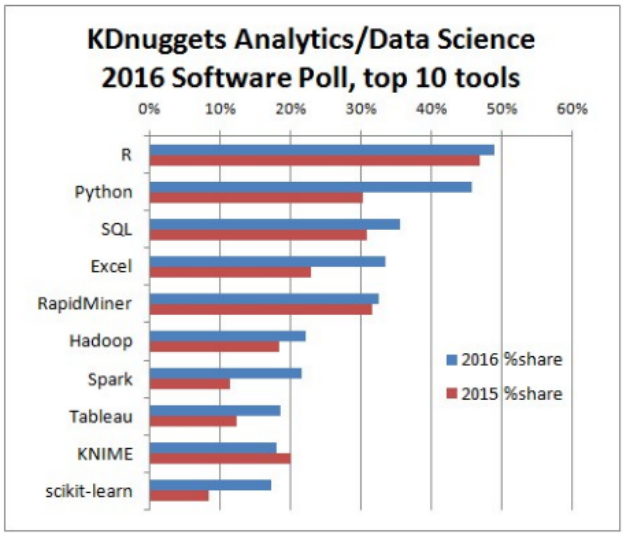


Figure 3. KDnuggets analytics/data science 2016 software poll: top 10 most popular tools (Piatetsky, 2016)

Table 2
A short review on implementations time series analysis in parallel computing

No	Refs	Objectives	Implemented Methods	Strategies for Parallelization
1	Mirko & Kantelhardt, (2013)	To implement statistical time series analysis, such as correlation, autocorrelation, in Big Data/large time series datasets	Correlation and autocorrelation	MapReduce and Hadoop platform
2	Sheng et al. (2013)	To design and implement Echo state network (ESN) for prediction time series by utilizing Extended Kalman Filter (EKF) in parallel computing by using MapReduce	Extended Kalman Filter	MapReduce
3	Górriz et al. (2004).	To implement a Parallel Neural Network (Cross-over Prediction Model) for time series forecasting	Neural network	PVM ("Parallel Virtual Machine") and MPI ("Message Passing Interface"
4	Zhao, Bryan, King, Song, & Yu (2012)	To implement an array-based algorithm to calculate summary statistics for long time-series daily grid climate data sets	Spatial analysis	The Parallel Python (PP) package
5	Liu & He (2012)	To find the optimal segmentation scheme of time series with a low execution time.	a modified Ant Colony Optimization algorithm (WACOS)	OPenMP library in C++

Table 2 *(continue)*

6	Xie, Wulamu, Wang, (2014)	To perform clustering analysis for time series data	Canopy and K-means based on SVD	MapReduce and Hadoop platform
7	Qian et al. (2014)	To present our parallel design of time series analysis and implementation of ARIMA modelling in MADlib's framework.	ARIMA	MADlib on Greenplum database and PostgreSQL database

From the programming language perspective, the R language has some characteristics that render it to offer some advantages (Wickham, 2014) such as providing complete data structures to ease data processing like list, *matrix, vector*, and *data.frame*.

Mostly, packages developed in the R framework are included in the following repositories: CRAN and the Bioconductor project. CRAN (Wickham, 2014) can be found at http://cran.r-project.org/, which is maintained through the efforts of volunteers (the "CRAN team") and the resources of the R Foundation and the employers of those volunteers (WU Wien, TU Dortmund, U Oxford, AT\&T Research). Meanwhile, Bioconductor (http://www.bioconductor.org/) is an open source, open development software project to provide R tools for the analysis and comprehension of high-throughput genomic data. Now, there are over 8000 packages available in CRAN, and these are classified into more than 30 task views. For instance, the task view of Time Series Analysis contains more than 30 R packages such as, the "forecast" package which involves methods and tools for displaying and analysing univariate time series forecasts (Hyndman & Khandakar, 2008).

Parallel Computing in R: "foreach" and "doParallel"

Firstly, we need to define what the definition of parallel computing is. It refers to a kind of computation in which many calculations or the execution of processes is conducted simultaneously (Gottlieb & Almasi, 1989). In other words, the processes need to be divided into some smaller modules so that these parts are executed at the same time. There are at least four types of parallel computing: bit-level parallelism, instruction-level parallelism, task parallelism, and data parallelism.

In the R ecosystem, there are more than 30 packages available at CRAN (https://CRAN.R-project.org/view=HighPerformanceComputing), which are used for high performance and parallel computing with R. In this CRAN Task View, we can find several groups of parallel computing such as Explicit Parallelism (e.g., the "pbdMPI" and "foreach" package), Grid Computing (e.g., the "multiR" package), etc.

The "foreach" package was developed by Calaway et al. (2015). It provides a new looping mechanism for executing R code simultaneously. Hence, the main reason for using the "foreach" package is that parallel computing on multiple processors/cores and multiple nodes of a cluster are available. Moreover, the "doParallel" package has a role as a backend engine for the "foreach" package. It provides a mechanism needed to execute the *foreach* command in parallel. It means that the "foreach" package must be used together with "doParallel" to execute code in parallel.

The following R code is a simple example to show how the "doParallel" and "foreach" perform a task in parallel:

```
R> library(foreach)
R> library(doParallel)
R> registerDoParallel(cores=2)
R> foreach(i=1:3) %dopar% sqrt(i)
```

The code on the first and second lines means we load the following packages: "foreach" and "doParallel". The second line shows that we use multicore-like functionality by defining 2 cores, whereas the last one is that we perform the square root command (i.e., *sqrt()*) of the three objects in parallel with 2 cores. Therefore, it yields the following results, as follows: 1, 1.414214, and 1.732051.

Time Series Analysis in R: "forecast"

The next package used in this research is "forecast." It provides automatic forecasting using exponential smoothing, ARIMA models, and other common forecasting methods (Hyndman & Khandakar, 2008). Related to this research, the following is the signature of one function included in the "forecast" package:

```
holt(y, h=10, damped=FALSE, level=c(80,95), fan=FALSE, initial=c("optimal","simple"),
exponential=FALSE, alpha=NULL, beta=NULL, lambda=NULL, biasadj=FALSE, x=y,
...)
```

The *holt()* function is used for predicting time series datasets based on the exponential smoothing methods. It can be seen that this function has several arguments such as: y is a numeric vector or time series, *h* is number of periods for forecasting/lead time, etc.

Design and Implementation of Parallel Exponential Smoothing

Figure 4 shows the model of parallel exponential smoothing utilising the package "forecast", "foreach", and "doParallel" for forecasting time series data.

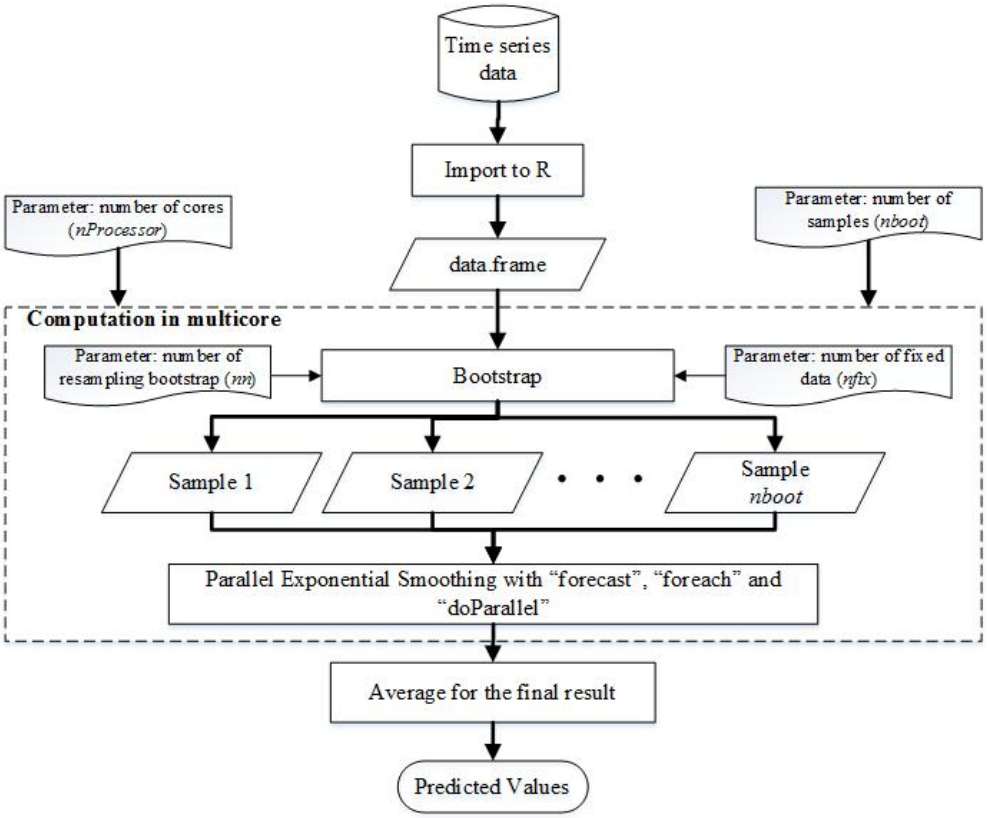


Figure 4. The model of parallel exponential smoothing for forecasting time series data in multicore

It can be seen that basically there are three main steps in this model, as follows:

1. Importing the original data (e.g., *.csv, .xls*, etc.) into an R object, which is *data.frame*. The implementation of this step is as given in the following code:

   R> alldata <- read.table("D:/bootstrap/follow1.out",header=TRUE)

   The *read.table()* function is used for reading the data saved in "D:/bootstrap/follow1.out" and saving into the variable *alldata* as *data.frame*.

2. This step is actually a main part of the model, which is, the computation of bootstrap and exponential smoothing in parallel by utilising the packages: "forecast", "foreach", and "doParallel". In this step, we can divide them into three phases, as follows:
   a. Loading and setting the computation in multicore. Firstly, we load the libraries by executing the following code:

      library(foreach)

      library(doParallel)

After importing the packages, the next code is to define the number of processors/ cores, as follows:

```
R> cl <- makeCluster(nProcessor=2)
R> registerDoParallel(cl)
```

It can be seen from the above code that the number of cores is 2.

b. Developing the procedure of bootstrap on time series data. Bootstrap is a process for selecting some data randomly with replacement so that this sample can represent the whole dataset. It is aimed to make parallelisation by splitting the data. Before explaining the implementation, we depict the processes on the bootstrap as in Figure 5. First, we need to define the number of fixed values (*nfix*) to be the beginning and ending parts of the sample. Then, bootstrap is done to assign the values between the fixed values. The processes will be repeated according to the defined number of sample (*nboot*).

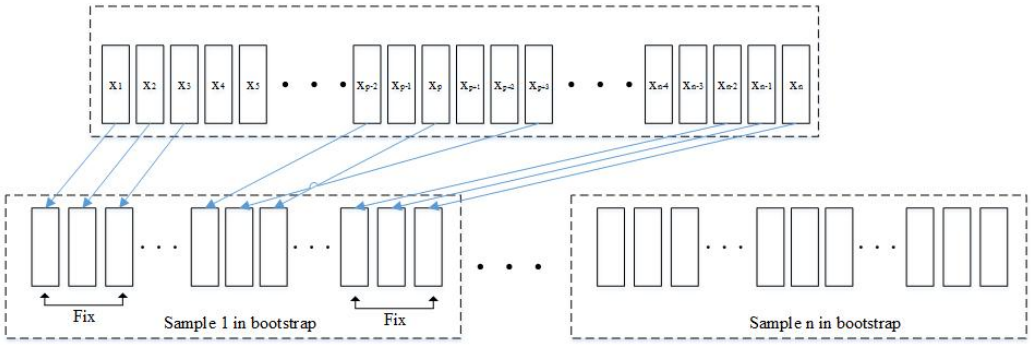


Figure 5. The procedure on the bootstrap method

The implementation of bootstrap can be seen in the following code:

```
boots <- function(alldata,nfix,nn){
    n<-nrow(alldata)
    a<-c(1:nfix)
    z<-c((n-nfix+1):n)
    i<-c(a,sample(c((nfix+1):(n-nfix-1)),nn,replace=T),z)
    x<-c(i[!duplicated(i)])
    ii<-c(1:length(x))
    temp<-list()
    tempData<-NA
    tempData[ii]<-NA
    temp$a<-tempData
```

```
temp$e<-tempData
temp$i<-tempData
temp$O<-tempData
temp$w_omega<-tempData
temp$a[x]<-alldata$a[x]
temp$e[x]<-alldata$e[x]
temp$i[x]<-alldata$i[x]
temp$O[x]<-alldata$O[x]
temp$w_omega[x]<-alldata$w_omega[x]
temp$a<-temp$a[!is.na(temp$a)]
temp$e<-temp$e[!is.na(temp$e)]
temp$i<-temp$i[!is.na(temp$i)]
temp$O<-temp$O[!is.na(temp$O)]
temp$w_omega<-temp$w_omega[!is.na(temp$w_omega)]
return (temp)
}
```

It can be seen that the computation of bootstrap is actually on the fifth line. Then, the process is repeated along with all columns of the data. In this case, we consider five columns, *a, e, i, O*, and *w_omega*, which represent semi major axis, eccentricity, inclination, argument of perihelion, and longitude of ascending node.

c. Prediction using exponential smoothing in parallel. This step is aimed to execute the "forecast" package for forecasting by using exponential smoothing in the "foreach" and "doParallel" package. It can be done by the following code:

```
predValue <- foreach(icount(nboot), .combine=rbind, .export=ls(envir=globalenv()))
%dopar%{
databaru<-boots(alldata,nfix,nn)
 numPoint<-nrow(databaru)
a<-holt(databaru$a,h=nForecast)
e<-holt(databaru$e,h=nForecast)
i<-holt(databaru$i,h=nForecast)
O<-holt(databaru$O,h=nForecast)
w_omega<-holt(databaru$w_omega,h=nForecast)
 return(c(a,e,i,O,w_omega,numPoint,.combine=rbind))
}
```

Thus, after executing the *boots()* function, we call the *holt()* function included in the "forecast" package on each column (i.e., *a, e, i, O,* and *w_omega*) on the *foreach()* function.

3. Average the results for obtaining predicted values. As we have some results that are according to the number of bootstrap samples, it is necessary to calculate the average value. This can be done by executing the *mean()* function built in R.

EXPERIMENTATION ON THE CALCULATION OF ASTEROID'S ORBITAL ELEMENT

In this section, we discuss three topics related to the experiment in this research. Firstly, we briefly explain the problem statement, and then illustrate how we collect the data. Finally, the experimental design is presented.

Problem Statement

Orbit calculations of celestial bodies in astronomy for a simple case is in the form of the two isolated objects (two-body problem), i.e., a single object with a smaller mass is orbiting other object that has a greater mass and under the influence of their mutual gravitational attraction only. Indeed, the two objects are orbiting their common centre of mass, where the orbital velocity and distance of each object from the common centre of mass are determined by each object's mass and their centre-to-centre distance. In the case of more than 2 objects are considered (generally known as the N-body problem), the same equations of motion can be expanded to the number of objects being simulated. In the implementation, the N-body problem is solved using restricted tree-body problem approach, where the third object is considered to have a negligible mass relative to the first and second objects. This approach is quite accurate such as when the systems considered are the Sun-planet-natural/artificial satellite and sun-planet-asteroid/comet as in this work.

The equation of motion for the N-body problem is (see, for example, Murray & Dermott, 1999) as follows:

$$\frac{d^2\boldsymbol{x}_i}{dt^2} = -\sum_{i,j=1;j\neq i}^{N} G\frac{m_j(\boldsymbol{x}_i-\boldsymbol{x}_j)}{|\boldsymbol{x}_i-\boldsymbol{x}_j|^3} \tag{8}$$

In equation (8), G is the universal gravitational constant and on the right hand side is the total gravitational attraction of the whole bodies considered. Equation (8) cannot be solved analytically but numerically for a certain time step in order to obtain a new position vector until the desired time. Some integration techniques available can be employed to obtain the numerical solution of equation (8), such as Wisdom-Holman Mapping (Wisdom & Holman, 1991), Regularised Mix Variable symplectic (Levison & Duncan, 1994), A fourth order T+U symplectic (TU4) method (Candy & Rozmus, 1991), Bulirsch-Stoer method (Press, Teukolsky, Vetterling, & Flannery, 1992), and so on.

Between the orbits of Mars and Jupiter (2.0 to 3.3 Astronomical Unit; 1 AU is defined as the average distance between the Earth-Sun), there is a population of asteroids, space rocks of various sizes that are remnants of the Solar system formation which failed to become a planet due to gravitational perturbations of Jupiter. This area is known as the Main Belt. It is strongly believed that this asteroid population is a major source of near-Earth asteroids (NEAs) (Bottke et al., 2002; Greenstreet, Ngo, & Gladman, 2012). The population of asteroids with orbit such

that [q (asteroids' perihelion distance – the closest distance of asteroids from the Sun) < 1.3 AU and Q (asteroids' aphelion distance – the furthest distance of asteroids from the Sun) > 0.98 AU] so the orbit bring them to near-Earth space. The mechanism that contributes to deliver the asteroids in the Main Belt toward the near-Earth space could be a collision among the asteroids which with the right post-collision velocity and trajectory immediately puts asteroids in the resonance zone (Farinella, Gonczi, Froeschlé, & Froeschlé, 1993), which will further change the orbit to become more elliptical or via slowly drift under the influence of the non-isotropic thermal force (Yarkovsky effect) (Bottke et al., 2002).

While in the near-Earth space, asteroids could experience close encounter with particular planets (Mercury, Venus, Earth and Mars). As a result of this close encounter, the asteroids may be fragmented due to strong tidal force, or if they can survive, their orbit can change drastically in a short time. Drastic changes in orbit occurred can change the fate of the asteroids in the future, i.e. whether they will still be orbiting the Sun or even collide with massive objects in the Solar system. Therefore, observational survey to find the presence of new asteroids and continually observations post-close encounter of asteroids with massive objects is essential to assess the probability of collision with the planets and the Sun in the future. The occurrence of an asteroid's close encounters with massive objects in the Solar System in the future can be predicted by computing the orbit for a certain span of time. Given that the orbits of asteroids in near-Earth space are very chaotic due to their close encounters with planets, orbit computation for asteroids' final fate prediction purpose is commonly done by generating some virtual asteroids (VA) which has a slightly different orbital elements from the real one. Evolution of the real and virtual asteroids' orbital elements obtained during the considered span of time will determine the orbital elements distribution, and where the most heavily populated cells are likely located. These are shown in Figure 6 for asteroid 2012 DA14 (see, for example, Wlodarczyk, 2012; Utama, Dermawan, Hidayat, & Fauzi, 2015) during our 106 years of orbital computation forward experiment after its close encounter with the Earth on 15th February 2013.

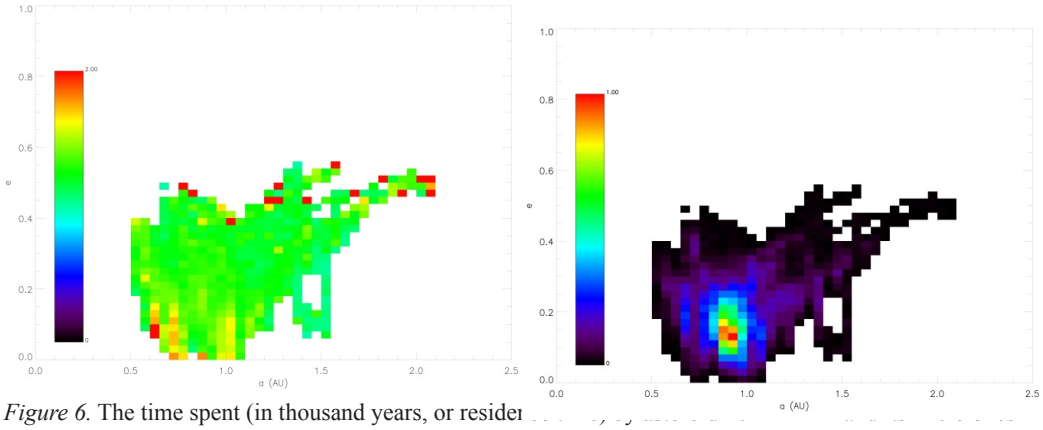


Figure 6. The time spent (in thousand years, or reside[illegible] cell (a, e) during 10^6 years orbital computation represents the expected orbital distribution. The colour scale gives the average time (thousands of years per particle) spent in the different cells; black being the shortest, while red is the longest. White corresponds to unvisited regions (left panel). Number of asteroids fall into each cell (a, e). Red depicts the most frequent cells visited by asteroids, while black depicts cell most rarely visited. White corresponds to unvisited region (right panel)

From the orbital computation, there are 9 clones of 120 that will end their life when colliding with Venus (4 clones) and Earth (5 clones).

Data Gathering

Asteroid's orbital elements data used in this work were obtained from NASA database (http://ssd.jpl.nasa.gov/sbdb.cgi) with the initial value of orbital elements (see Table 3), while description of each orbital element shown in Figure 7. The orbital evolution executed using SWIFT integrator package (Levison & Duncan, 1994) had been modified, namely SWIFT RMVS4YE, which has included non-isotropic thermal force in it (Dermawan, Hidayat, & Utama, 2013). The RMVS4 scheme implemented in this integrator helps us to do computation of orbital elements accurately, especially when a close encounter between asteroid with massive objects occurs. By including the thermal force (Yarkovsky effect), the accurate orbit predictions for long duration ($> 10^5$ years) computation can be reached.

Table 3
Orbital element of Asteroid used as a test case for parallel computing in this work

Orbital Element	Value
a	1.2456 AU
e	0.33552
I	13°.3358
Ω	337°.2327
ω	276°.8451
M	191°.0565

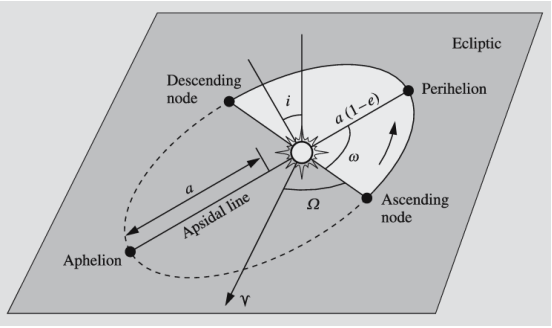


Figure 7. Six integration constants are needed to describe celestial objects' orbit. The constants are the longitude of ascending node Ω, the argument of perihelion ω, the inclination i, the semi major axis a, the eccentricity e and the time of perihelion passage τ [= (M x P/2π) + τ_0, where P is orbital period]
(*Source:* Karttunen, Kröger, Oja, Poutanen, & Donner, 2007)

Experimental Design

In order to do the experiments efficiently, we designed the following two scenarios of prediction: single computing and parallel computing. For the first scenario, we predicted 10 next point of

time, whereas in the second, several simulations were conducted with different as illustrated in Table 4.

Table 4
Parameters on the parallel-computing scenario

Parameters	Description	Values
nProcessor	Number of processors/cores	1 and 4
nboot	Number of samples on bootstrap	4, 12, and 20
Nn	Number of data point on each sample	1000, 10000, and 25000
nForecast	Lead time	1 and 10
nfix	Number of fixed values for the beginning and ending parts of time series data after bootstrap	2 and 20

Moreover, we compared the results with the algorithm RMVS4-YE by calculating Mean Absolute Percentage Error (MAPE) defined, as follows:

$$\text{MAPE} = \frac{\sum \frac{|x_i - f_i|}{x_i}}{n} \times 100\% \quad (9)$$

where x and f are the true and predicted values, while i and n are lead time and number of observations/samples.

RESULTS AND DISCUSSION

As explained previously, in this experiment, there are two scenarios: prediction by using exponential smoothing on single processor and prediction by using parallel exponential smoothing on multicore. Therefore, results of the MAPE calculation in comparison with those obtained from the Algorithm RMVS4-YE are presented in Table 5. In this case, we predicted for 10 periods for five components of asteroid's orbital element: *a, e, i, O, w_omega*,

Table 5
MAPE calculation between exponential smoothing with single processor and the algorithm RMVS4-YE

Period	*A*	*e*	*I*	*O*	*w_omega*	Average (%)
1	0.000809240	0.044868466	0.010442664	0.039505848	0.011069138	0.021339071
2	0.022884474	0.058763748	0.003809124	0.037560265	0.009701406	0.026543803
3	0.037606712	0.104317766	0.063138274	0.070033749	0.008846505	0.056788601
4	0.045043056	0.068458498	0.065173303	0.062258119	0.007350644	0.049656724
5	0.037680306	0.064277164	0.060649023	0.054457305	0.007521384	0.044917036
6	0.037680306	0.055576884	0.058092212	0.063763096	0.013759502	0.045774400
7	0.022958057	0.083037792	0.062095493	0.078176277	0.019996522	0.053252828
8	0.022958057	0.078855233	0.070624528	0.074059252	0.016621183	0.052623651
9	0.023031641	0.076933502	0.073315612	0.062859932	0.008204252	0.048868988
10	0.030392229	0.043376160	0.064199768	0.063112270	0.013159997	0.042848085

representing semi major axis, eccentricity, inclination, argument of perihelion, and longitude of ascending node. The last column is average of MAPE of all parameters. Moreover, the average computation time of all simulation, which is 84.48 seconds, was also calculated.

Then, for the second scenario (i.e., using parallel exponential smoothing) 72 simulations were performed based on the combinations of parameter given in Table 4. For example, Table 6 shows MAPE calculation on the first simulation that assigns the following parameters: *nProcessor* = 1, *nboot* = 4, *nfix* = 2, *nn* = 1000, and *nforecast* = 1. It can be seen that the average of MAPE in this case is around 0.312%.

Table 6
MAPE calculation of the first simulation in the second scenario

a	*E*	*i*	*O*	*w_omega*	Average (%)
0.008680939	0.251733092	0.597202154	0.694654647	0.005684152	0.311590997

Because of the limited space in this paper, we recapitulated the complete results, which are average of MAPE from 72 simulations on the second scenario, as illustrated in Table 7 below.

Table 7
The complete recapitulation of MAPE calculation of 72 simulations on scenario 2

No	*nProcesor*	*nboot*	*nfix*	*nn*	*nforecast*	Average *numPoint*	Computation Cost (s)	MAPE Average (%)
1	1	4	2	1000	1	1003.750	1.7453090	0.31159100
2	1	4	2	1000	10	1003.000	1.8960360	1.94565918
3	1	4	20	1000	1	1039.750	1.7257320	0.15078242
4	1	4	20	1000	10	1038.500	1.6875560	1.05304777
5	1	4	2	10000	1	9908.750	4.5730320	0.04971548
6	1	4	2	10000	10	9902.000	4.8302820	0.28135093
7	1	4	20	10000	1	9939.250	4.4239750	0.02676445
8	1	4	20	10000	10	9940.750	4.4637740	0.14651301
9	1	4	2	25000	1	24394.750	10.5053300	0.02695131
10	1	12	2	25000	10	24403.750	9.9289680	0.12089122
11	1	12	20	25000	1	24441.500	10.1047500	0.02123174
12	1	12	20	25000	10	24426.250	9.1230160	0.09271730
13	1	12	2	1000	1	1002.667	3.1559090	0.41470090
14	1	12	2	1000	10	1002.917	2.8936640	2.07270682
15	1	12	20	1000	1	1039.000	2.9690490	0.15143420
16	1	12	20	1000	10	1039.333	2.9958390	0.82286152
17	1	12	2	10000	1	9907.083	12.6789100	0.04737498
18	1	12	2	10000	10	9903.500	12.9946800	0.27593724
19	1	12	20	10000	1	9941.500	11.6467000	0.02591236
20	1	12	20	10000	10	9943.333	12.6069800	0.15076421
21	1	12	2	25000	1	24378.080	29.9228700	0.02606734

Table 7 *(continue)*

22	1	12	2	25000	10	24394.000	28.3534500	0.12098353
23	1	12	20	25000	1	24421.000	28.0681100	0.02086259
24	1	12	20	25000	10	24439.080	31.4522200	0.09423594
25	1	20	2	1000	1	1002.900	4.4074640	0.39646641
26	1	20	2	1000	10	1003.000	4.8229670	2.14794240
27	1	20	20	1000	1	1039.050	4.4783750	0.16141778
28	1	20	20	1000	10	1039.050	4.3665020	0.85253999
29	1	20	2	10000	1	9902.550	20.7596900	0.04828453
30	1	20	2	10000	10	9903.350	18.8794600	0.26617271
31	1	20	20	10000	1	9943.250	18.5950500	0.02625648
32	1	20	20	10000	10	9937.750	17.6448800	0.14852876
33	1	20	2	25000	1	24388.100	44.4893600	0.02582092
34	1	20	2	25000	10	24388.450	45.3040700	0.12037558
35	1	20	20	25000	1	24423.150	43.3132700	0.02086241
36	1	20	20	25000	10	24429.350	46.4674400	0.09522702
37	4	4	2	1000	1	1003.000	3.2107390	0.45777004
38	4	4	2	1000	10	1002.750	3.0782100	2.22245284
39	4	4	20	1000	1	1039.000	3.0190890	0.15731712
40	4	4	20	1000	10	1038.500	3.1023910	0.82419055
41	4	4	2	10000	1	9899.000	4.8375500	0.04813014
42	4	4	2	10000	10	9892.750	4.7776220	0.26066381
43	4	4	20	10000	1	9934.250	4.6702120	0.02545224
44	4	4	20	10000	10	9942.250	4.5244820	0.14652436
45	4	4	2	25000	1	24399.750	7.5436390	0.02593090
46	4	12	2	25000	10	24403.000	6.8495680	0.11601128
47	4	12	20	25000	1	24420.500	7.1595970	0.02139856
48	4	12	20	25000	10	24450.000	7.4199600	0.09610011
49	4	12	2	1000	1	1002.583	3.6817990	0.40538402
50	4	12	2	1000	10	1003.333	3.6715830	2.21894915
51	4	12	20	1000	1	1039.333	3.7104400	0.16964136
52	4	12	20	1000	10	1038.833	3.6743280	0.82289333
53	4	12	2	10000	1	9903.750	8.2461690	0.04858334
54	4	12	2	10000	10	9907.333	8.5302590	0.27667333
55	4	12	20	10000	1	9940.250	7.5553620	0.02569687
56	4	12	20	10000	10	9937.583	8.1135430	0.27667333
57	4	12	2	25000	1	24386.920	15.5334500	0.02618081
58	4	12	2	25000	10	24390.670	15.8121500	0.12152408
59	4	12	20	25000	1	24429.580	16.7012900	0.02051017
60	4	12	20	25000	10	24424.250	15.8905900	0.09819543
61	4	20	2	1000	1	1002.900	4.4920760	0.37927788
62	4	20	2	1000	10	1001.950	4.4137490	2.17249141
63	4	20	20	1000	1	1039.350	4.3392870	0.15577406

Table 7 *(continue)*

64	4	20	20	1000	10	1039.050	4.4213620	0.88915073
65	4	20	2	10000	1	9904.150	12.3716900	0.04753417
66	4	20	2	10000	10	9903.900	12.0911500	0.29628936
67	4	20	20	10000	1	9937.800	11.0131900	0.02718215
68	4	20	20	10000	10	9942.750	11.2667100	0.15365591
69	4	20	2	25000	1	24388.600	24.6162400	0.02656839
70	4	20	2	25000	10	24391.200	23.8743000	0.12202390
71	4	20	20	25000	1	24423.650	23.8025900	0.02093147
72	4	20	20	25000	10	24432.800	25.6407900	0.09519576

As shown in Table 7, it can be stated that the best results and computation cost is the 72nd simulation, where the MAPE average is about 0.095% at the computation cost around 25.6 seconds. Of course, it is also faster than the first scenario (i.e., exponential smoothing with single processor), where its computation cost is about 84.48 seconds. Moreover, the MAPE average of both simulation is relatively close, which is 0.095% is for the 72nd simulation in the second scenario and 0.044% for the first scenario. This means that the result of parallel exponential smoothing is reasonable. We also focused on analysing the second scenario. For example, the result shows that the higher number of bootstrap causes longer computation time, while higher number of resampling bootstrap makes the lower MAPE, even though certain number of samples cannot make MAPE significantly better.

CONCLUSION AND RECOMMENDATIONS FOR FUTURE WORK

In this research, a model of time series analysis for conducting exponential smoothing with the bootstrap methods in parallel computing by utilising the R packages: "forecast," "foreach," and "doParallel" has been presented. The next contribution is that the implementation of the proposed model has been technically explained in detail so that the code is reproducible. Furthermore, we conducted some experiments for forecasting asteroid's orbital elements in order to evaluate the model and its implementation, along with the analysis of the results.

As for future work, we plan to convert the problem of time series analysis on asteroid's orbital elements into regression one. Then, we will solve it by utilising machine-learning methods, such as fuzzy rule based systems (Riza, Bergmeir, Herrera, & Benitez, 2015), rough set theory and fuzzy rough set theory (Riza et al., 2014; Nazir, Shahzad, & Riza, 2016b), and gradient descent (Riza, Nasrulloh, Junaeti, Zain, & Nandiyanto, 2016a).

REFERENCES

Beyer, M. A., & Laney, D. (2012). *The importance of 'Big Data': A definition.* Retrieved from https://www.gartner.com/doc/2057415/importance-big-data-definition.

Bottke, W. F., Morbidelli, A., Jedicke, R., Petit, J. M., Levison, H. F., Michel, P., & Metcalfe, T. S. (2002). Debiased orbital and absolute magnitude distribution of the near-Earth objects. *Icarus, 156*(2), 399-433.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. New Jersey: John Wiley & Sons.

Brown, R. G. (1959). *Statistical forecasting for inventory control*. New York: McGraw-Hill.

Calaway, R., Analytics, R., & Weston, S. (2015). *Foreach: Provides foreach looping construct for R*. Retrieved from https://cran.r-project.org/web/packages/foreach/index.html.

Candy, J., & Rozmus, W. (1991). A symplectic integration algorithm for separable hamiltonian functions. *Journal of Computational Physics, 92*(1), 230-256.

Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM, 51*(1), 107-113.

Dermawan, B., Hidayat, T., & Utama, J. A. (2013). *Pengembangan integrator swift_rmvs4 dengan melibatkan efek termal*. Prosiding Seminar Himpunan Astronomi Indonesia.

Derryberry, D. R. (2014). *Basic data analysis for time series with R*. New Jersey: John Wiley & Sons.

Farinella, P., Gonczi, R., Froeschlé, C., & Froeschlé, C. (1993). The injection of asteroid fragments into resonances. *Icarus, 101*(2), 174-187.

Gantz, J., & Reinsel, D. (2011). *Extracting value from chaos*. Retrieved from http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf.

Gardner, E. S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting, 4*(1), 1-28.

Górriz, J. M., Algeciras, E. P. S., Puntonet, C. G., Salmerón, M., & Martin-Clemente, R. (2004). Parallelization of time series forecasting model. In *Proceeding of 12th Euromicro Conference* (pp. 103-110). Vienna, Austria: Ustrian Computer Society.

Gottlieb, A., & Almasi, G. S. (1989). *Highly parallel computing*. Redwood City, Calif.: Benjamin/Cummings.

Greenstreet, S., Ngo, H., & Gladman, B. (2012). The orbital distribution of near-Earth objects inside Earth's orbit. *Icarus, 217*(1), 355–366.

Holt, C. C. (1957). *Forecasting trends and seasonals by exponentially weighted averages*. O.N.R. Memorandum 52/1957, Carnegie Institute of Technology.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 26*(3), 1-22.

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Berlin: Springer Science & Business Media.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting, 18*(3), 439-454.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics, 5*(3), 299-314.

Kane, M. J., Emerson, J., & Weston, S. (2013). Scalable strategies for computing with massive data. *Journal of Statistical Software, 55*(14), 1-19.

Karttunen, H., Kröger, P., Oja, H., Poutanen, M., & Donner, K. J. (2007). *Fundamental Astronomy* (5th Ed). Berlin: Springer.

Kirchgässner, G., Wolters, J., & Hassler, U. (2012). *Introduction to modern time series analysis.* Berlin: Springer Science & Business Media.

Levison, H. F., & Duncan, M. J. (1994). The long-term dynamical behavior of short-period comets. *Icarus, 108*(1), 18–36.

Liu, H., & He, Z. (2012). Parallel ant colony optimization algorithms for time series segmentation on a multi-core processor. In *Proceeding of 2012 4th International Conference of Intelligent Human-Machine Systems and Cybernetics (IHMSC)* (pp. 340-343). Nanchang, China: IEEE.

Mirko, K., & Kantelhardt, J. W. (2013). Hadoop. TS: large-scale time-series processing. *International Journal of Computer Applications, 74*(17), 1-8.

Murray, C. D., & Dermott, S. F. (1999). *Solar system dynamics.* Cambridge: University Press.

Murthy, A. C., Vavilapalli, V. K., Eadline, D., Niemiec, J., & Markham, J. (2013). *Apache Hadoop YARN: Moving beyond MapReduce and batch processing with Apache Hadoop 2.* New Jersey: Pearson Education.

Nazir, S., Shahzad, S., & Riza, L. S. (2016). Birthmark-based software classification using rough sets. *Arabian Journal for Science and Engineering, 42*(2), 859-871.

Palit, A. K., & Popovic, D. (2006). *Computational intelligence in time series forecasting*. London: Springer-Verlag London Limited.

Palma, W. (2016). *Time series analysis.* New Jersey: Wiley.

Pegels, C. C. (1969). Exponential forecasting: Some new variations. *Management Science, 15*(5), 311-315.

Piatetsky, G. (2016). *R, python duel as top analytics, data science software – kdnuggets 2016 software poll results.* Retrieved from http://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN77*. Cambridge: Cambridge University Press.

Qian, H., Yang, S., Iyer, R., Feng, X., Wellons, M., & Welton, C. (2014). Parallel time series modeling-a case study of in-database big data analytics. In *Proceeding of Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 417-428). Berlin: Jerman Springer International Publishing.

Riza, L. S., Bergmeir, C., Herrera, F., & Benitez, J. M. (2015). frbs: fuzzy rule-based systems for classification and regression in R. *Journal of Statistical Software, 65*(1), 1-30.

Riza, L. S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Slezak, D., & Benıtez, J. M. (2014). Implementing algorithms of rough set theory and fuzzy rough set theory in the R package RoughSets. *Information Sciences, 287*, 68-89.

Riza, L. S., Nasrulloh, I. F., Junaeti, E., Zain, R., & Nandiyanto, A. B. D. (2016a). gradDescentR: An R package implementing gradient descent and its variants for regression tasks. In *Proceeding of 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)* (pp. 125-129). Yogyakarta, Indonesia: IEEE.

Riza, L. S., Wihardi, Y., Nurdin, E. A., Ardi, N. D., Asmoro, C. P., Wijaya, A. F. C., & Nandiyanto, A. B. D. (2016b). Analysis on atmospheric pressure, temperature, and wind speed profiles during total solar eclipse 9 March 2016 using time series clustering. *Journal of Physics: Conference Series, 771*(1), 012009.

Sheng, C., Zhao, J., Leung, H., & Wang, W. (2013). Extended kalman filter based echo state network for time series prediction using mapreduce framework. In *Proceeding of Ninth International Conference of Mobile Ad-hoc and Sensor Networks (MSN)* (pp. 175-180). Dalian, China: IEEE.

Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting, 19*(4), 715–725.

Troester, M. (2015). *Big data meets big data analytics.* Retrieved from http://www.sas.com/resources/whitepaper/wp_46345.pdf.

Utama, J. A., Dermawan, B., Hidayat, T., & Fauzi, U. (2015). Dinamika orbit 2012 DA14 pascapapasan dekat dengan bumi. *SPEKTRA: Jurnal Fisika dan Aplikasinya, 16*(1), 1-5.

Wickham, H. (2014). *Advanced R.* Florida: CRC Press.

Wisdom, J., & Holman, M. (1991). Symplectic maps for the n-body problem. *The Astronomical Journal, 102*(4), 1528-1539.

Wlodarczyk, I. (2012). The potentially dangerous asteroid 2012 DA14. *Monthly Notices of the Royal Astronomical Society, 427*(2), 1175-1181.

Xie, Y., Wulamu, A., Wang, Y., & Liu, Z. (2014). Implementation of time series data clustering based on SVD for stock data analysis on hadoop platform. In *Proceeding of 2014 9th IEEE Conference on Industrial Electronics and Applications* (pp. 2007-2010). Hangzhou, China: IEEE.

Zhang, G. P. (2012). Neural networks for time-series forecasting. In *Proceeding of Handbook of Natural Computing* (pp. 461-477). Berlin, Germany: Springer Berlin Heidelberg.

Zhao, G., Bryan, B. A., King, D., Song, X., & Yu, Q. (2012). Parallelization and optimization of spatial analysis for large scale environmental model data assembly. *Computers and Electronics in Agriculture, 89*, 94-99.

Zomaya, A. Y. (1996). *Parallel and distributed computing handbook* (Vol. 204). New York: McGraw-Hill.

Pertanika J. Sci. & Technol. 26 (1): 463 - 478 (2018)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Omnidirectional MIMO Antenna with Collinear Array for LTE Applications

Enjang Akhmad Juanda, Tommi Hariyadi*, Abdul Aziz Reguna and Arjuni Budi Pantjawati

Departemen Pendidikan Teknik Elektro, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi, Bandung 40154, Indonesia

ABSTRACT

To increase reception in Long Term Evolution (LTE) network inside a building, a repeater is needed. The antenna used in the repeater inside the building is usually a high gain antenna with omnidirectional radiation pattern. Meanwhile, to increase data rate in LTE, one of methods used is by using Multiple Input Multiple Output (MIMO) antenna. In this paper, the omnidirectional MIMO antenna at 1.8 GHz for LTE applications has been designed and realised. The single element of this MIMO antenna is a collinear microstrip antenna array. The design and simulation were done using 3D electromagnetic simulator software, while antenna realisation was done using FR4 microstrip with a thickness of 1.6 mm and permittivity of 4.4. The measurement results showed that this antenna has 359 MHz bandwidth in frequency range at 1.6-1.9 GHz, with a return loss less than -10 dB. The antenna gain is around 7.4 to 8.7 dBi with omnidirectional radiation pattern and mutual coupling is around -22 dB to -27 dB.

Keywords: Collinear, LTE, Microstrip Antenna, MIMO, Omnidirectional

INTRODUCTION

To increase reception in Long Term Evolution (LTE) network inside a building, a repeater is needed (Yeo, Lee, Hwang, & Kim, 2013). The function of a repeater is to receive radio signal from outside the building, strengthen the received signal, and emit it inside the desirable building (Xue, 2008). The antenna used in the repeater inside the building is usually the antenna with omnidirectional radiation pattern and high gain (Shankari & Gowtham, 2014). To increase data rate in LTE, one of methods used is by using the Multiple Input Multiple Output (MIMO) antenna (Abdullah & Yonis, 2012).

ARTICLE INFO

Article history:
Received: 25 April 2017
Accepted: 28 November 2017

E-mail addresses:
juanda@upi.edu (Enjang Akhmad Juanda),
tommi.hariyadi@upi.edu, tommi@ieee.org (Tommi Hariyadi),
abd.azizreguna@gmail.com (Abdul Aziz Reguna),
arjunib@yahoo.com (Arjuni Budi Pantjawati)
*Corresponding Author

ISSN: 0128-7680

Omnidirectional antenna chosen here is monopole antenna, while to get the antenna with high gain, we make the array of the monopole antenna in collinear (Balsley & Ecklund, 1972; Wang, Yang, Li, Li, & Chen, 2011). Collinear antenna array is chosen to get antenna with high gain with omnidirectional radiation pattern. There are many types of monopole antenna and today, monopole antenna made from microstrip is being developed in vast amount. Microstrip antenna has advantages such as light, simple, and easy to be manufactured because it can be made with etching process, which is the same as the process of making PCB. Some research has been done related to collinear microstrip antenna array by several researchers (Jiao, Xueguan, Xinmi, & Huiping, 2013; Litva, Zhuang, & Liang, 1993; Peng-Fei, Li-Ming, Yong, & Xin, 2013; Polivka & Holub, 2006).

In LTE network, MIMO antenna is commonly used as a way to increase data rate. MIMO antenna suitable for indoor repeater is MIMO antenna that has omnidirectional radiation pattern (Ching-song, Hsu, Chun, Engineering, Rd, & Shiang, 2013; Fang, Sun, & Chuang, 2014; Malik, Kartikeyan, & Nagpal, 2016; Moradikordalivand, Rahman, & Khalily, 2014). Therefore, we proposed a MIMO omnidirectional antenna that has high gain, where in every antenna is microstrip monopole collinear antenna.

Mutual coupling is a critical problem in the design of MIMO antennas because it deteriorates the performance of MIMO systems, which not only affects the antenna efficiency but also influences the correlation (Li, Du, Takahashi, Saito, & Ito, 2012). Some research has been done to reduce mutual coupling in a MIMO antenna but the mutual coupling is less than -15 dB (Nguyen, Le, Le, Tran, & Yamada, 2016) and -16 dB (Ghafor, Hameed, Abdullah, Sabbagh, Al, & Bashir, 2015).

This research made use of the experimental method which consists of design and simulation using 3D Electromagnetic Simulator, fabrication of antenna, and antenna measurement. The proposed antenna is implemented using FR4 substrate, with a thickness of 1.6 mm and permittivity of 4.4. The expected specifications of the antenna are that it works at a frequency of 1.8 GHz (LTE band 3) with a frequency range from 1.7 to 1.9 GHz for return loss less than -10 dB, gain more than 5 dBi with mutual coupling less than -20 dB (Li et al., 2012) and has omnidirectional radiation pattern.

ANTENNA DESIGN

The initial step in antenna design is to determine the characteristics of the antenna that we expected. In this design, we proposed an antenna that meets the characteristics listed in Table 1 below.

Table 1
Specifications of the proposed antenna

Frequency Range	1.7 – 1.9 GHz
Gain	> 5 dBi
Bandwidth	≥ 200 MHz
Mutual Coupling	< -20 dB
Radiation Pattern	Omnidirectional
Return Loss	≤ -10dB

The first step in designing microstrip MIMO collinear antenna array is designing a single microstrip collinear antenna as shown in Figure 1. The designing process of the microstrip collinear antenna aims to increase antenna gain with omnidirectional radiation pattern. Parameters of the antenna dimensions used in the first step of collinear antenna design are given in Figure 2 and Table 2.

The second step in designing the antenna is optimisation towards the size of collinear antenna dimension so that the results of simulation will match the characteristics of the antenna to be obtained.

Changes of antenna dimension after optimisation include parameters a, c and d. After optimising those parameters, the size of antenna is obtained, followed by obtaining the characteristics of the antenna. These parameters are also shown in Table 2.

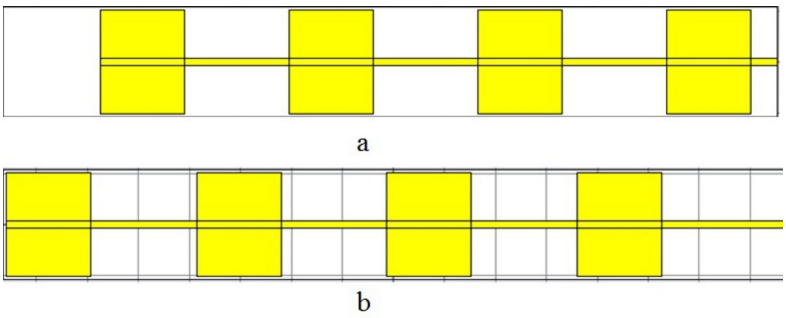


Figure 1. (a) Layout of collinear antenna, top view (b) Layout of collinear antenna, bottom view

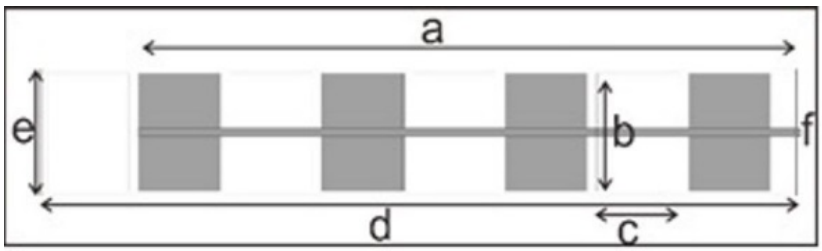


Figure 2. Initial parameters of collinear antenna array

Table 2
Parameters of the collinear antenna array before and after optimisation

Antenna Parameters	Dimension (mm)	
	Before Optimisation	After Optimisation
Length of Feeder (a)	266.87	265.5
Length of Patch (b)	46	46
Width of Patch (c)	30	33
Length of Substrat (d)	294	305
Width of Substrat (e)	49	49
Width of Feeder (f)	3.065	3.065

MIMO ANTENNA DESIGN

The third step of antenna design is designing MIMO antenna from the collinear antenna that has been made before. The design of the antenna is shown in Figure 3. To get the characteristics of MIMO antenna that we expected, the modification of antenna dimension should be done. Therefore, optimisation process, the same process with the optimisation of single collinear antenna, should be done. In this optimisation, there are a few parameters that will be adapted towards antenna dimension so that the characteristics of the antenna produced match the characteristics of the antenna that we expected. Dimension and distance parameters of the MIMO antenna, before and after the optimisation, are shown in Table 3. Based on data in Table 3, all the parameters have changed except for patch width and feeder width.

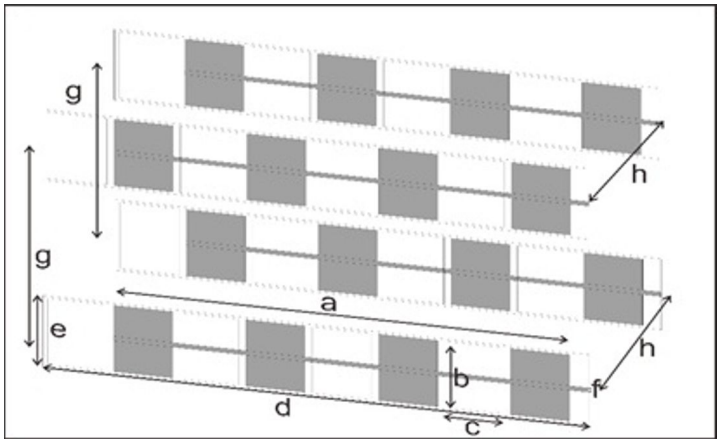


Figure 3. The Antenna MIMO design

Table 3
Dimension and distance parameters of the MIMO antenna before and after the optimisation

Antenna Parameters	Dimension (mm)	
	Before Optimisation	After Optimisation
Length of Feeder (a)	265.5	267.6
Length of Patch (b)	46	41
Width of Patch (c)	33	33
Length of Substrat (d)	305	306
Width of Substrat (e)	49	43.25
Width of Feeder (f)	3.065	3.065
Antenna Distance (g)	170	160
Antenna Distance (h)	170	160

RESULTS AND DISCUSSION

Early Simulation

In this step, early simulation of microstrip collinear antenna is done with dimension from the literature review. Feeding technique used is the feeding with microstrip transmission line. The return loss of the early simulation does not match with the expected frequency range. Figure 4 shows the former design of microstrip collinear antenna with the parameter from Table 2, while Figure 5 shows the return loss value of the antenna and Figure 6 shows gain and radiation pattern of the antenna at 1.8 GHz.

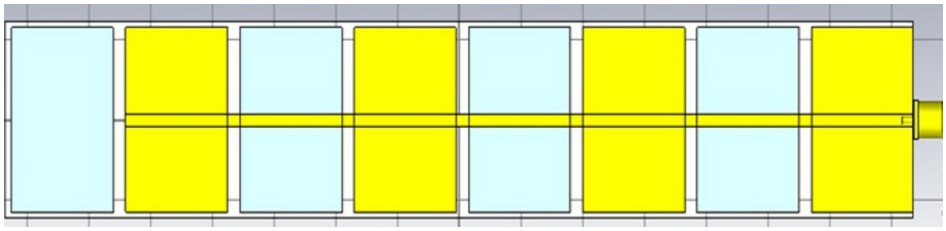


Figure 4. Former design of the microstrip collinear antenna

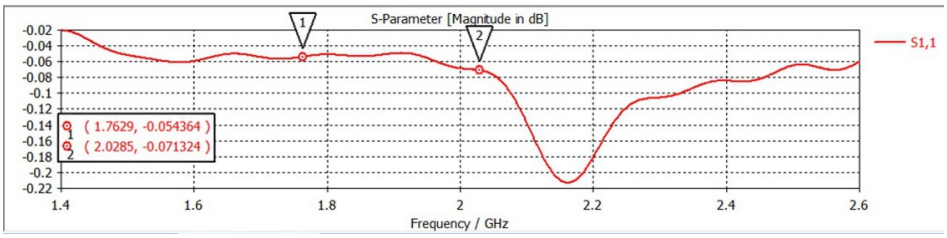


Figure 5. Return loss of the antenna in the early design

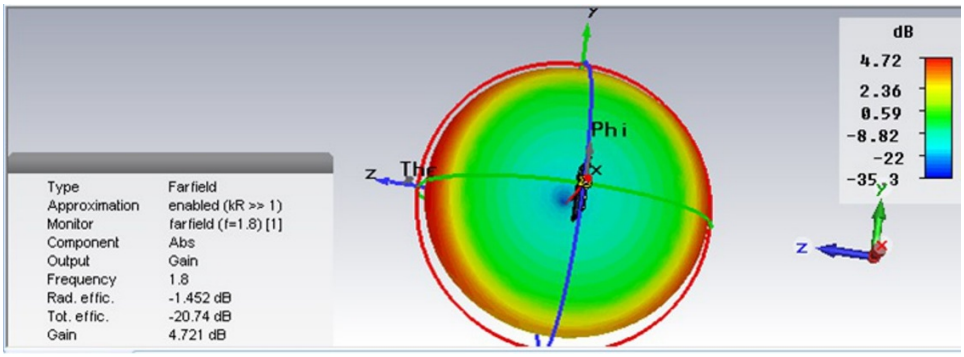


Figure 6. Gain and radiation pattern of the antenna at 1.8 GHz

Result of design simulation of early microstrip collinear antenna shows that the characteristics of the antenna in Figure 5 do not match with the characteristics of the expected antenna. Therefore, the optimisation towards antenna dimension should be done. Dimensional change is done by changing parameter c or patch width from 30-36 mm using 5 samples. The result of return loss simulation is presented in Figure 7.

Based on the data in Figure 7, we can see that when the patch width is 33 mm, the resonant frequency of the antenna is around 1.8 GHz, while the dimension is close to the characteristics of the expected antenna when the patch width is 33 mm. If we extend the patch width of antenna, the resonant frequency shifts to the lower frequency. The resonant frequency of the antenna shifts to the upper frequency if we shrink the patch width of the antenna. Antenna dimension is proportional with the wavelength or inversely proportional with the frequency (Balanis, 2005; Kraus, Marhefka, & Khan, 2010).

The result of return loss simulation from antenna with desirable dimension obtained from optimisation (Table 2) is depicted in Figure 8. The resonant frequency is around 1.8 GHz with the frequency range of 1.6 GHz to 1.9 GHz and return loss less than -10 dB. Gain and radiation pattern of the single microstrip collinear antenna are shown in Figure 9. From Figure 9, we can see that the gain of single microstrip collinear antenna is 5.459 dBi with omnidirectional radiation pattern.

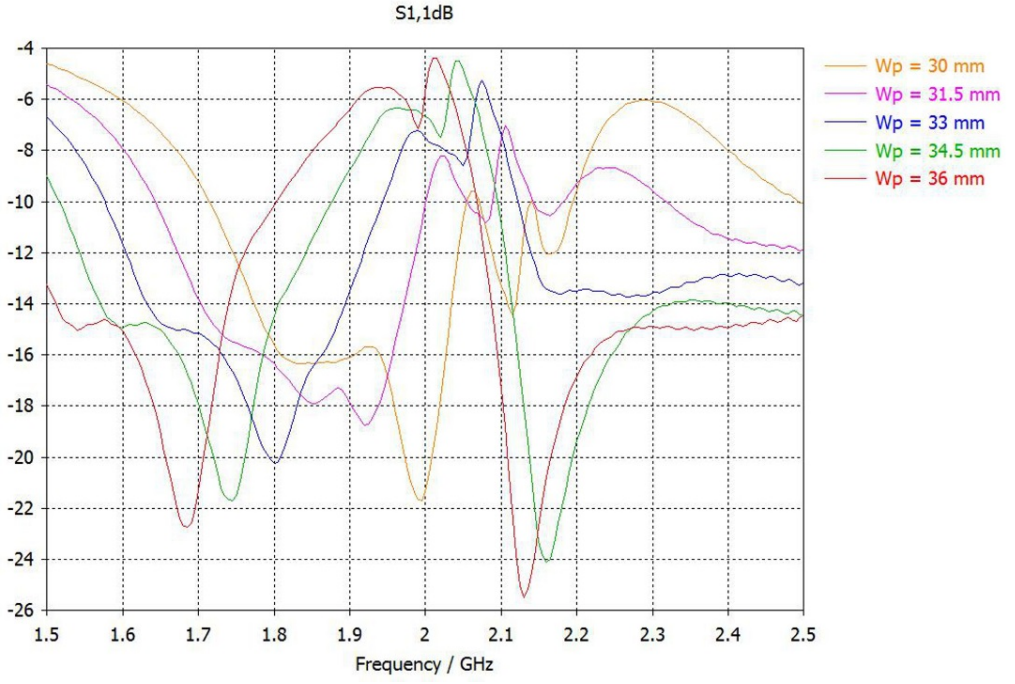


Figure 7. The simulation result of return loss for five different patch width values

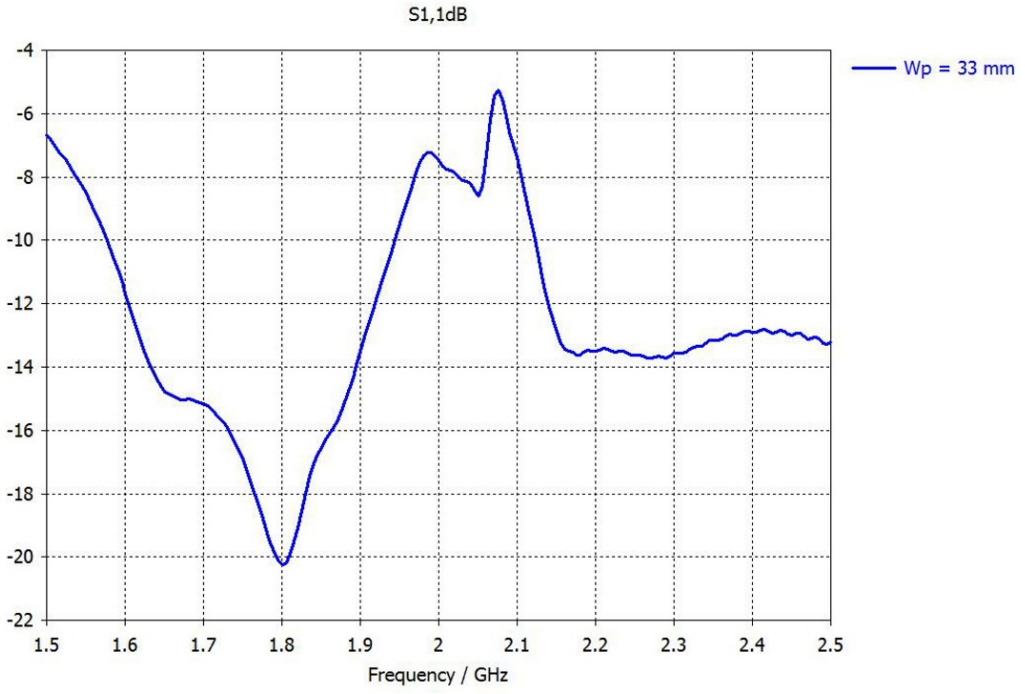


Figure 8. The simulation result of antenna return loss with optimal dimension obtained from optimisation

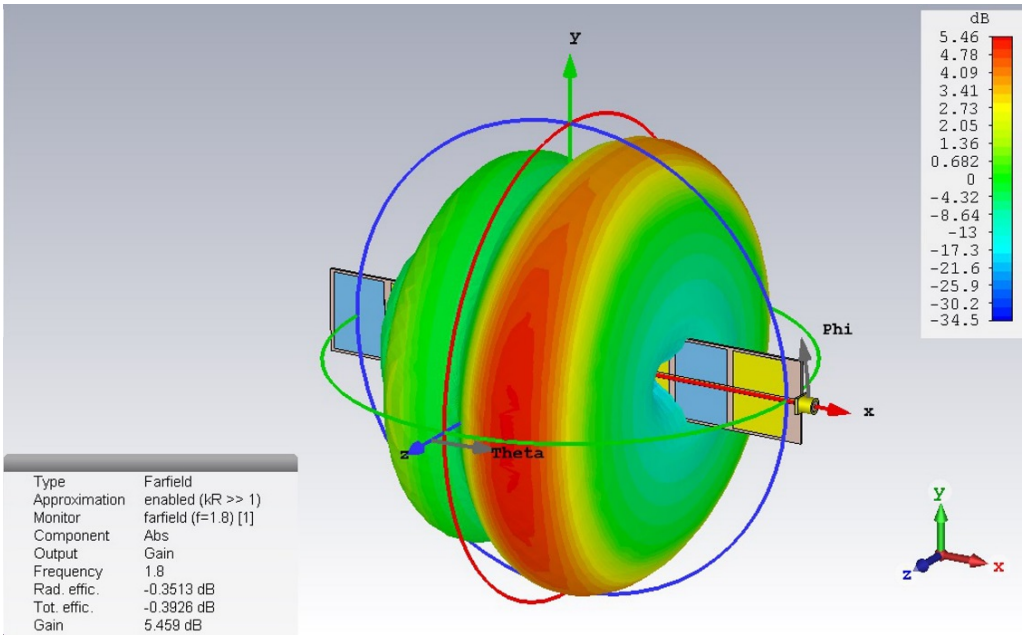


Figure 9. Gain and radiation pattern of the single microstrip collinear antenna

Results of the MIMO Antenna Simulation

The proposed MIMO antenna consists of four microstrip collinear antenna with specific distance. To get the final specification expected, optimisation toward the distance between antennas is done. The distance expected is as close as possible so that the total dimension is not too big but with mutual coupling value less than -20 dB in the expected frequency range.

Distance optimisation between MIMO antennas is done by changing parameter g and h (Ja) with a distance of 140-170 mm to get the expected simulation result. The result of return loss simulation with antenna distance 140-170 mm is given in Figure 10.

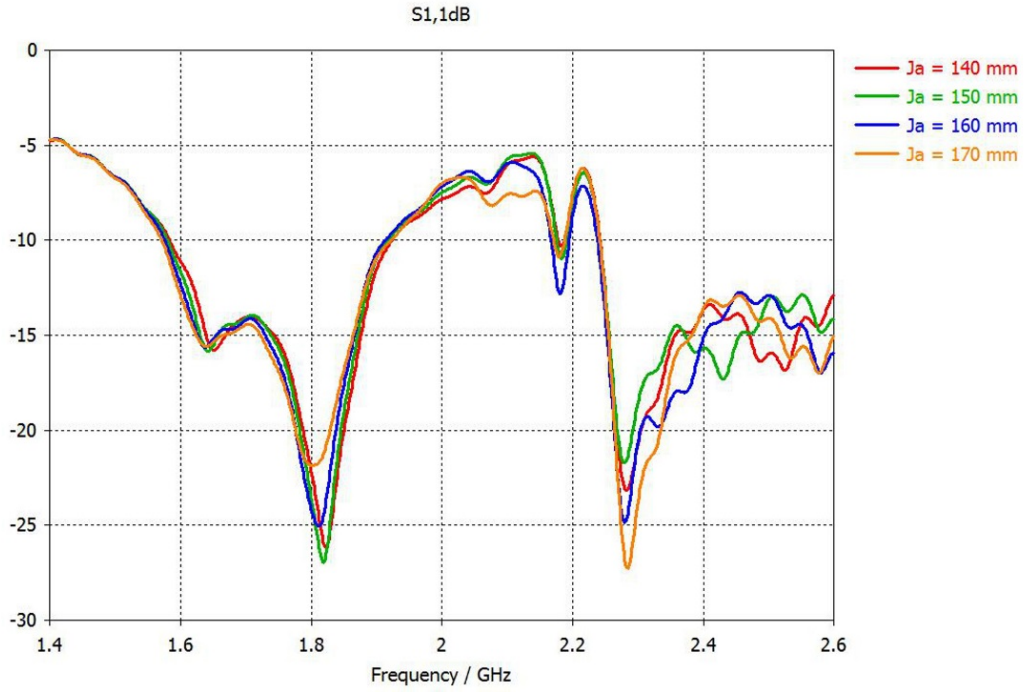


Figure 10. The simulation result of return loss with the antenna distance of 140-170 mm

The simulation result of return loss in Figure 10 shows that the variation in the antenna distance from 140 mm to 170 mm does not have any significant difference. From Figure 11, however, we can see that the ideal antenna distance or the closest distance is 160 mm. If we shrink the antenna distance, the total dimension of MIMO antenna will become smaller but the mutual coupling will increase. Meanwhile, if we extend the antenna distance, the mutual coupling will decrease but the total size of MIMO antenna will be larger. Thus, the ideal antenna distance in this case is 160 mm. This value is close to 1 wavelength (1λ) where the λ at frequency 1.8 GHz is 166.67 mm. Linnartz (2008) said that mutual coupling would add extra degradation to the outage capacity of the MIMO system. Therefore, to guarantee minimal loss in capacity, the antennas should be spaced about 1 wavelength apart. The final result of the return loss simulation of MIMO antenna is presented in Figure 12.

The final result of the return loss simulation of MIMO antenna in Figure 12 shows that the specifications already match with the specification set at the beginning, i.e. 1.8 GHz with bandwidth 359 MHz (1.576-1.935 GHz). The mutual coupling value from the final simulation is shown in Figure 13.

The average mutual coupling value in Figure 13 is under -20 dB in frequency range 1.6-1.9 GHz. So, mutual coupling value is match with the value in the beginning. While antenna gain 1 to 4 from final simulation is shown in Figure 14.

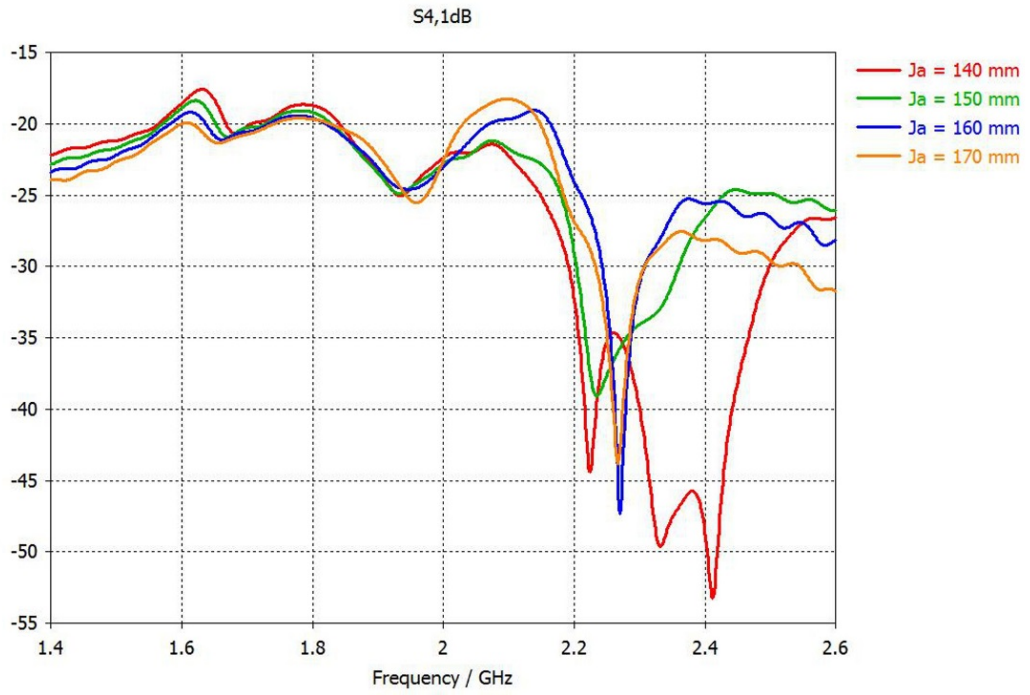


Figure 11. Mutual coupling from antenna distance optimisation from 140 mm to 170 mm

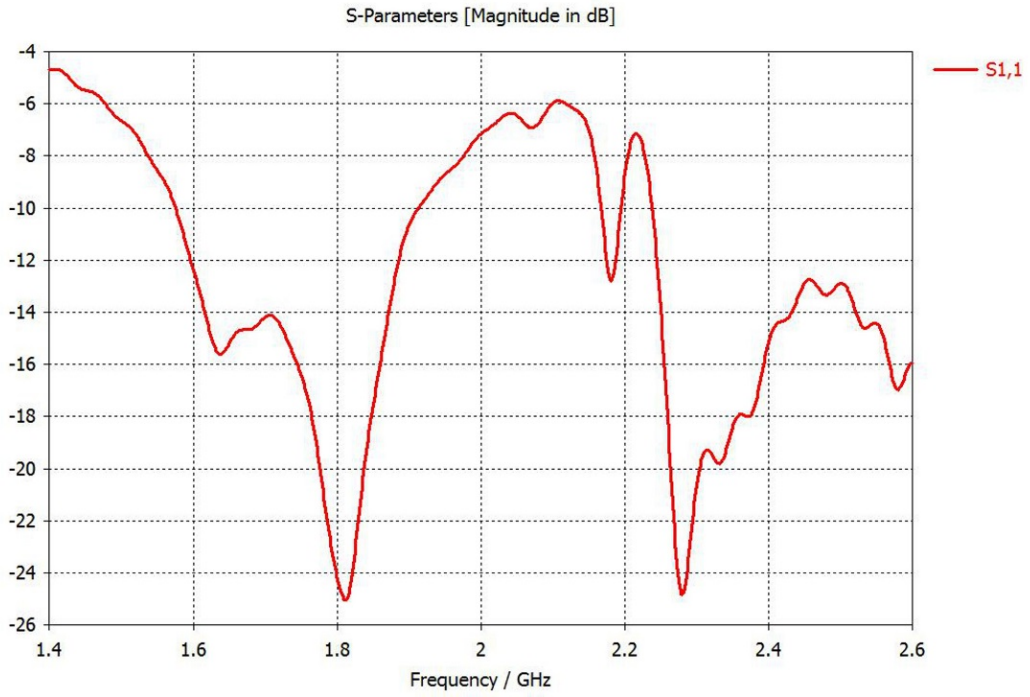


Figure 12. The final result of the return loss simulation of MIMO antenna

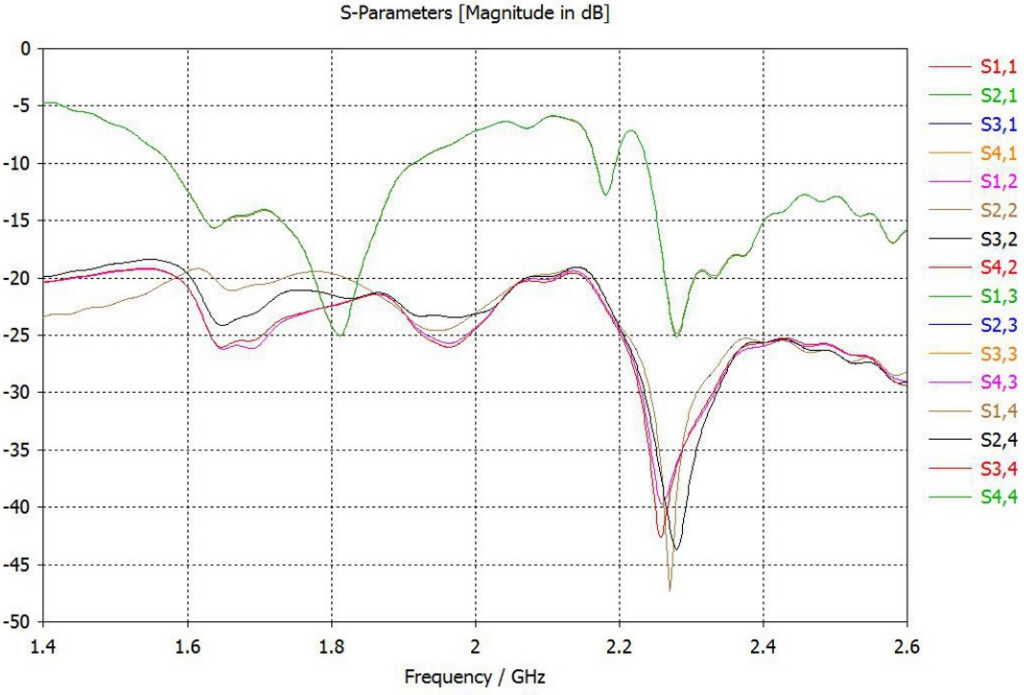


Figure 13. Mutual coupling of the MIMO antenna from the final simulation

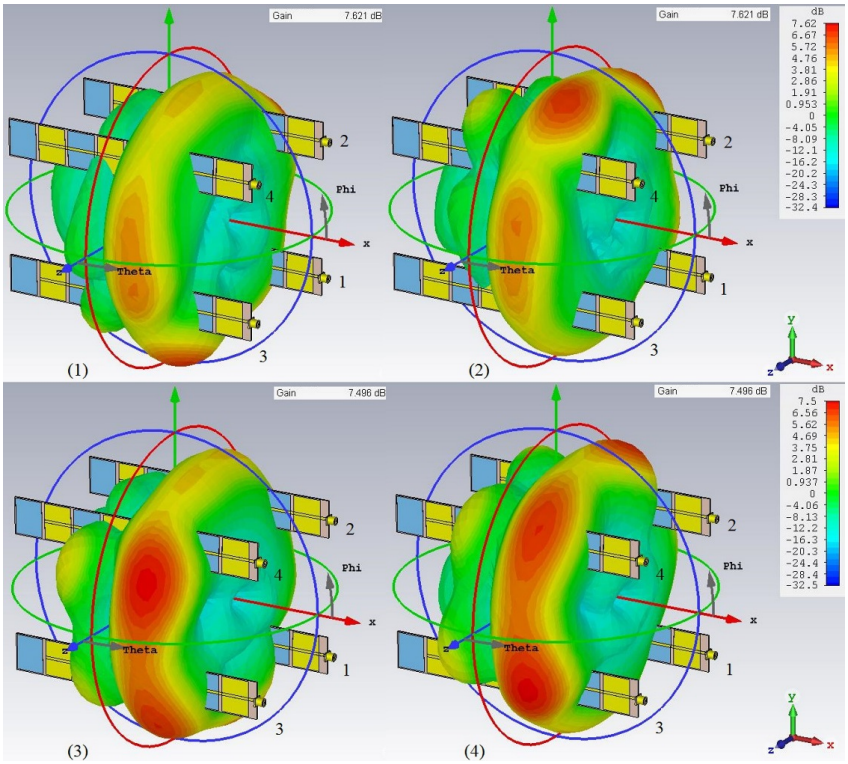


Figure 14. Simulation result of the MIMO antenna gain: (a) gain of antennas 1 and 2 (7.621 dBi), and (b) gain of antennas 3 and 4 (7.496 dBi)

The results for the MIMO antenna radiation pattern for azimuth and elevation plane are given in Figure 15. The radiation pattern of microstrip MIMO antenna from the simulation shows that the antenna has omnidirectional radiation pattern. Hence, the radiation pattern matches the expected radiation pattern.

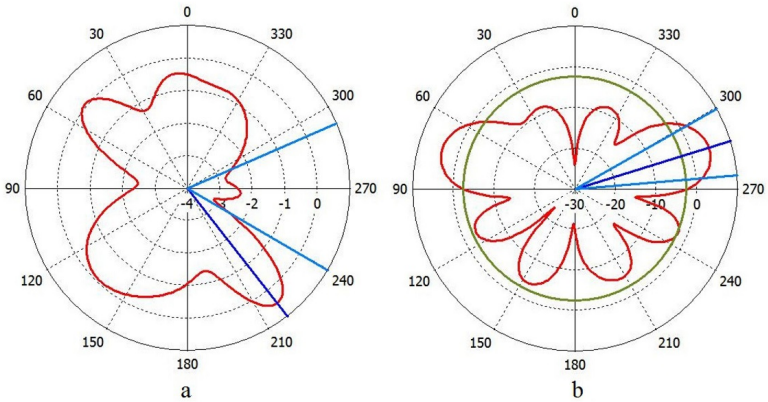


Figure 15. Radiation pattern of the antenna: (a) Azimuth plane, and (b) Elevation plane

Fabrication of Antenna

Based on the final simulations of MIMO antenna, the antenna is manufactured. Figure 16 shows the result of antenna fabrication. Antenna substrate used is FR4 with dielectric constant 4.4 and 1.6 mm of thickness, while the antenna mounting is using acrylic.

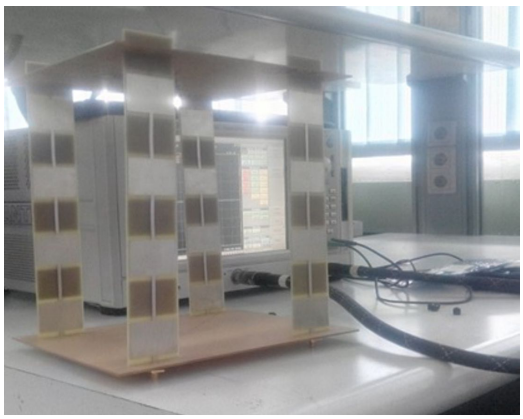


Figure 16. Results of the MIMO antenna manufacturing

The Result of Antenna Measurement

Antenna return loss measurement. Results for measuring return loss of MIMO antenna are presented in Figure 17. From this figure, we can see that the expected return loss value matches with the results of the return loss measurement, which is less than -10 dB in the frequency range of 1.6-1.9 GHz.

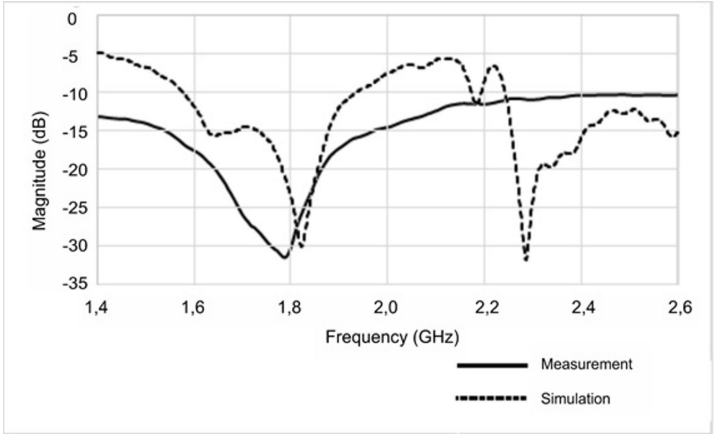


Figure 17. Result of measuring return loss

Mutual coupling measurement. Mutual coupling of MIMO antenna is shown in Table 4. It shows that mutual coupling at 1.8 GHz is around -22 dB to -27 dB. Hence, the mutual coupling of this MIMO antenna matches the specification that is less than -20 dB.

Table 4
Mutual coupling of MIMO antenna at 1.8 GHz

	Parameters					
	S12/S21	S13/S31	S14/S41	S23/S32	S24/S42	S34/S43
Mutua Cloupling (dB)	-26.4	-22.7	-23.6	-23.6	-22.7	-27.5

Results of Radiation Pattern Measurement

The result of measuring radiation pattern is given in Figure 18. In Figure 18, we can see that the result of radiation pattern measurement has a shape that is almost similar with the simulation result of radiation pattern. From Figure 18, we also can see that the radiation pattern azimuth base is omnidirectional. Hence, the radiation pattern of the MIMO antenna matches the specification, i.e. omni directional.

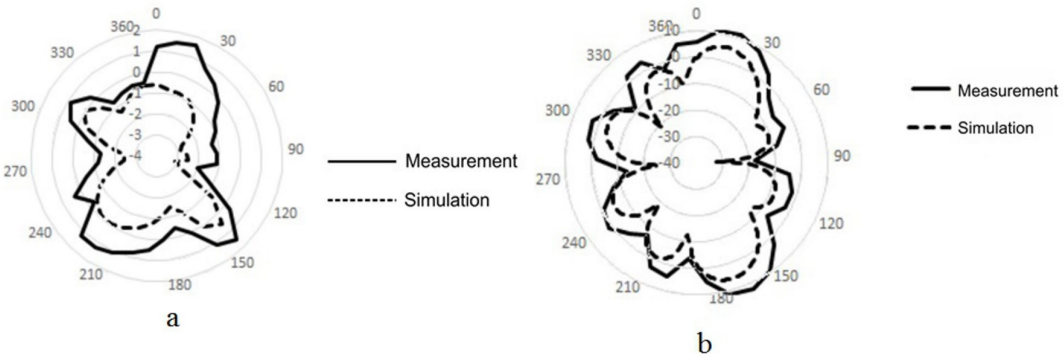


Figure 18. The result of measuring radiation pattern: (a) Azimuth plane (b) Elevation plane

The Results of Gain Measurement

The measuring of gain is done using two similar antenna methods. We can see the differences in Table 5. From Table 5, we can see that MIMO antenna gain is between 7.44-8.77 dBi. The gain is obtained from measurement using two similar antenna methods. The result matches with the expected value that is more than 5 dBi. Although the MIMO antenna uses monopole antenna, the gain obtained is higher compared to when common monopole antenna is used. This happened because we used collinear antenna that consists of monopole antenna arranged in a line order for every element of the MIMO antenna, so that the gain obtained is higher than that of the common monopole.

Table 5
The result of antenna gain measurement

Frequency (GHz)	Gain (dBi)			
	Antenna 1	Antenna 2	Antenna 3	Antenna 4
1.7	8.1	7.9	8.2	8.38
1.75	8.29	8.11	7.98	7.89
1.8	8.29	7.71	7.44	7.97
1.85	8.67	8.5	7.96	8.21
1.9	7.91	8.13	8.7	8.77

CONCLUSION

In this research, we have done the designing, arrangement and measurement processes of the Microstrip MIMO antenna. The arrangement of the microstrip MIMO antenna can work in 1,6-1,9 GHz frequency with return loss less than -10 dB. Antenna gain obtained from the measurement is about 7435-8765 dBi, which matches the expected specifications. The mutual coupling value of micro-strip MIMO antenna array is less than -20 dB, which also matches the specification. Based on the results of the measurement of radiation pattern, the arrangement of the microstrip MIMO antenna is already omnidirectional. Therefore, the proposed antenna can be used for the LTE application.

REFERENCES

Abdullah, M. F. L., & Yonis, A. Z. (2012). Performance of LTE release 8 and release 10 in wireless communications. In *Proceedings of the International Conference on Cyber Security, Cyber Warfare and Digital Forensic* (pp. 236-241). Kuala Lumpur, Malaysia: IEEE.

Balanis, C. A. (2005). *Antenna theory, analysis and design* (3rd Ed.). Hoboken, NJ: John Wiley and Sons, Inc.

Balsley, B. B., & Ecklund, W. L. (1972). A portable coaxial collinear antenna. *IEEE Transactions on Antennas and Propagation, 20*(4), 513-516.

Ching-song, C., Hsu, W., Chun, L. M., Engineering, E., Rd, W., & Shiang, G. (2013). A compact multi band MIMO antenna. In *Proceedings of the International Conference on Electrical and Electronics Engineering* (pp. 435-438). Johor, Malaysia: IEEE.

Fang, H. S., Sun, J. S., & Chuang, C. S. (2014). A compact dual-band MIMO planar antenna. In *Proceedings of the International Conference on Systems, Man, and Cybernetics* (pp. 1827-1830). California, USA: IEEE.

Ghafor, A., Hameed, A. A., Abdullah, A. K. S., Sabbagh, H. M., Al & Bashir, H. K. (2015). Mutual coupling reduction of a (2 × 1) MIMO antenna system using parasitic element structure for WLAN applications. *Journal of Emerging Trends in Computing and Information Sciences, 6*(11), 605-613.

Jiao, W., Xueguan, L., Xinmi, Y., & Huiping, G. (2013). A planar coaxial collinear antenna with rectangular coaxial strip. In *Proceedings of the International Symposium on Antennas and Propagation* (pp. 716-719). Orlando, FL, USA: IEEE.

Kraus, J. D., Marhefka, R. J., & Khan, A. S. (2010). *Antenna and wave propagation* (4th Ed.). New Delhi: Mcgraw Hill Education.

Li, Z., Du, Z., Takahashi, M., Saito, K., & Ito, K. (2012). Reducing mutual coupling of MIMO antennas with parasitic elements for mobile terminals. *IEEE Transactions on Antennas and Propagation, 60*(2), 473-481.

Linnartz, J. P. (2008). Effects of antenna mutual coupling on the performance of MIMO systems. In *Proceedings of the 29th Symposium on Information Theory* (pp. 1-8). Auckland, New Zealand: IEEE.

Litva, J., Zhuang, Y. Z. Y., & Liang, A. (1993). Modelling study of coaxial collinear antenna array. In *Proceedings of Canadian Conference on Electrical and Computer Engineering* (pp. 920-924). Vancouver, BC, Canada: IEEE.

Malik, J., Kartikeyan, M. V., & Nagpal, D. (2016). MIMO antenna with omnidirectional pattern diversity. *Electronics Letters, 52*(2), 102-104.

Moradikordalivand, A., Rahman, T. A., & Khalily, M. (2014). Common elements wideband MIMO antenna system for WiFi/LTE access-point applications. *IEEE Antennas and Wireless Propagation Letters, 13*, 1601-1604.

Nguyen, Q. D., Le, T. T., Le, D. T., Tran, X. N., & Yamada, Y. (2016). A compact MIMO ultra-wide band antenna with low mutual coupling. *ACES Journal, 31*(3), 252-260.

Peng-Fei, Z., Li-Ming, S., Yong, L., & Xin, L. (2013). A novel omnidirectional microstrip coaxial collinear antenna array. In *Proceedings of the International Radar Conference 2013* (pp. 0285-0285). Adelaide, Australia: IEEE.

Polivka, M., & Holub, A. (2006). Planar version of collinear microstrip patch antenna. In *Proceedings of the International Conference on Microwaves, Radar and Wireless Communications* (pp. 2-5). Krakow: IEEE.

Shankari, S. S., & Gowtham, A. G. (2014). Signal strength enhancer. *International Journal of Industrial Electronics and Electrical Engineering, 2*(2), 16-18.

Wang, Q., Yang, X. L., Li, Y. T., Li, Z. B., & Chen, S. T. (2011). Design of a high gain and low sidelobe coaxial collinear antenna array. In *Proceedings of the International Conference on Computational Problem-Solving* (pp. 26–28). Chengdu, China: IEEE.

Xue, Y. (2008). Amplify-and-forward repeater enhanced random access in single-cell wireless communications. In *Proceedings of the International Symposium on Personal, Indoor and Mobile Radio Communications* (pp.7–11). Cannes, France: IEEE.

Yeo, W. Y., Lee, S. M., Hwang, G. H., & Kim, J. H. (2013). Coexistence of 3G repeaters with LTE base stations. *The Scientific World Journal, 2013*. http://doi.org/10.1155/2013/185372.

REFEREES FOR THE PERTANIKA
JOURNAL OF SCIENCE AND TECHNOLOGY

VOL. 26 (1) JAN. 2018

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 Jalil Nordin
(UPM, Malaysia)

Adi Irfan Che Ani
(UKM, Malaysia)

Agus Arsad
(UTM, Malaysia)

Ahmad Ramli Mohd Yahya
(USM, Malaysia)

Ahmed Jalal Khan Chowdhury
(IIUM, Malaysia)

Aminuddin Ab. Ghani
(USM, Malaysia)

Anil K. Philip
(UNIZWA, Oman)

Arifah Bahar
(UTM, Malaysia)

Asep Suhendi
(Telkom University, Indonesia)

Binu, K. G.
(SJEC, India)

Chua Han Bing
(Curtin University, Malaysia)

Daud Mohamad
(UiTM, Malaysia)

Edriyana Abd. Aziz
(UMP, Malaysia)

Ezza Syuhada Sazali
(UTM, Malaysia)

Hazrul Abdul Hamid
(USM, Malaysia)

Ibrahim Venkat
(USM, Malaysia)

Isa Bala Muhammad
(FUTMINNA, Nigeria)

Kanchan Sandhu
(PAU, India)

Kek Sie Long
(UTHM, Malaysia)

Leong Siow Hoo
(UiTM, Malaysia)

Liawas Barukang
(UMS, Malaysia)

Liong Choong Yeun
(UKM, Malaysia)

Low Heng Chin
(USM, Malaysia)

Marina Yusoff
(UiTM, Malaysia)

Mohamadariff Othman
(UM, Malaysia)

Mohamed Abd Rahman
(IIUM, Malaysia)

Mohammed Alias Yusof
(UPNM, Malaysia)

Mohd Adzir Mahdi
(UPM, Malaysia)

Mohd Azlan Mohd Ishak
(UiTM, Malaysia)

Mohd Zalisham Jali
(USIM, Malaysia)

Nor Asiah Ramli
(UiTM, Malaysia)

Nor Fazlida Mohd. Sani
(UPM, Malaysia)

Nor Hayati Othman
(USM, Malaysia)

Noraini Abdullah
(UMS, Malaysia)

Norhayati Mohd Tahir
(UMT, Malaysia)

Nur Izura Udzir
(UPM, Malaysia)

Nurain Mohd Noor
(HPJ, Malaysia)

Nurin Haniah Asmuni
(UiTM, Malaysia)

Nurul Sima Mohamad Shariff
(USIM, Malaysia)

Papita Das
(Jadavpur University, India)

Rizka Zulhijah
(Universitas Brawijaya, Indonesia)

Roslinazairimah Zakaria
(UMP, Malaysia)

Sathyashankara Sharma
(Manipal University, India)

Sayang Mohd Deni
(UiTM, Malaysia)

Simon Fong
(UMAC, China)

Srikumar Chakravarthi
(PU, Malaysia)

Subhash Bhatia
(USM, Malaysia)

Suehazlyn Zainudin
(UKM, Malaysia)

Syamsiah Mashohor
(UPM, Malaysia)

Syamsul Rizal Abd Shukor
(USM, Malaysia)

Tay Chia Chay
(UiTM, Malaysia)

Thamer Ahmed Mohamed
(UPM, Malaysia)

Thaned Rojsiraphisal
(CMU, Thailand)

Yumn Suhaylah Yusoff
(USIM, Malaysia)

Zamalia Mahmud
(UiTM, Malaysia)

CMU	– Chiang Mai University
FUTMINNA	– Federal University of Technology Minna
HPJ	– Hospital Putrajaya
IIUM	– International Islamic University Malaysia
PAU	– Punjab Agricultural University
PU	– Perdana University
SJEC	– St. Joseph Engineering College
UiTM	– Universiti Teknologi MARA
UKM	– Universiti Kebangsaan Malaysia
UM	– Universiti Malaya
UMAC	– University of Macau
UMP	– Universiti Malaysia Pahang
UMS	– Universiti Malaysia Sabah
UMT	– Universiti Malaysia Terengganu
UNIZWA	– University of Nizwa
UPM	– Universiti Putra Malaysia
UPNM	– Universiti Pertahanan Nasional Malaysia
USIM	– Universiti Sains Islam Malaysia
USM	– Universiti Sains Malaysia
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

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.M[3].

Authors' addresses. Multiple authors with different addresses must indicate their respective addresses separately by superscript numbers:

George Swan[1] and Nayan Kanwal[2]
[1]Department of Biology, Faculty of Science, Duke University, Durham, North Carolina, USA.,
[2]Office 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*

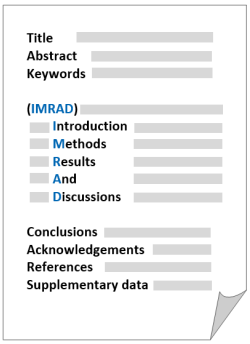
This page should **repeat** the **full title** of your paper with only the **Abstract** (the abstract should be less than 250 words for a Regular Paper and up to 100 words for a Short Communication), and **Keywords**.

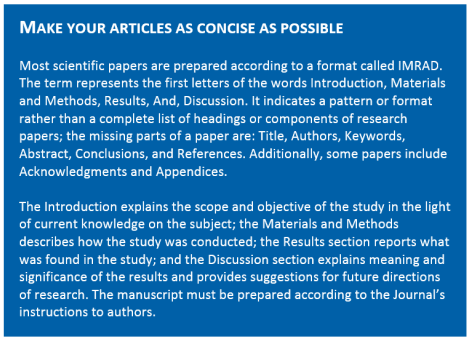
Keywords: Not more than eight keywords in alphabetical order must be provided to describe the contents of the manuscript.

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 avavailble) in this order.





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/).

PERTANIKA
JOURNALS

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).

PERTANIKA JOURNALS

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

Pertanika Journal of Science & Technology
Vol. 26 (1) Jan. 2018

Contents

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

PENERBIT UPM PRESS
http://penerbit.upm.edu.my
E-mail : penerbit@putra.upm.edu.my
Tel : +603 8946 8855/8854
Fax : +603 8941 6172

ISSN 0128-7680
01
9 770128 768083