Pertanika Journal of

SCIENCE & TECHNOLOGY

JST

VOL. 25 (1) JAN. 2017

PERTANIKA JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

Journal of Science & Technology

About the Journal

Overview

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

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

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

The Journal is available world-wide.

Aims and scope

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

History

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

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

Goal of *Pertanika*

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

Quality

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

Abstracting and indexing of *Pertanika*

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

Future vision

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

Citing journal articles

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

Publication policy

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

Code of Ethics

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

International Standard Serial Number (ISSN)

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

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

Lag time

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

Authorship

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

Manuscript preparation

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

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words **I**ntroduction, **M**aterials and 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 | 1619 | 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 35 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. 25 (1) Jan. 2017

Contents

Foreword

Welcome to the **First Issue 2017** 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 **31 articles,** of which **ten** are regular research articles. This issue also features **seven** selected papers from the World Research & Innovation Convention in Engineering and Technology 2014 (WRICET 2014), **nine** papers were chosen from the International Conference on Computational Methods in Engineering & Health Sciences (ICCMEH 2015) and **five** papers from the International Conference of Translational Molecular Imaging (ICTMIPs 2016). The authors of these articles vary in country of origin, coming from **Malaysia, New Zealand, Canada, Japan, India, North Africa, United Kingdom, Australia and Iraq**.

The regular articles cover a wide range of topics. The first article is on the analysis on the effect of shop floor parameters on the effectiveness of preventive maintenance through discrete event simulation (*Sheng, L. W., Basri, E. I. and Kamaruddin, S.*). The following articles look at: the analysis of energy consumption in pineapple cultivation in Malaysia (*Nazri, A. M. and Pebrian, D. E.*); a combined analytical method for optimal location and sizing of distributed generation considering voltage stability and power loss in a power distribution system (*Maryam Mirzaei, Jasronita Jasni, Hashim Hizam, Noor Izzri Abdol Wahab and Ehsan Moazami*); a novel adaptive neuro fuzzy inference system based classification model for heart disease prediction (*Abdu Masanawa Sagir and Saratha Sathasivam*); an ID-based secure and flexible buyer-seller watermarking protocol for copyright protection (*Ashwani Kumar, S. P. Ghrera and Vipin Tyagi*); modified Kohonen network algorithm for selection of the initial centres of Gustafson-kessel algorithm in credit scoring (*Sameer, F. and Abu Bakar, M. R.*); wireless sensor network and internet of things (IoT) solution in Agriculture (*Zulkifli, C. Z. and Noor, N. N.*); a comparative study of the behaviour of treated and untreated tyre crumb mortar with Oil palm fruit fibre addition (*Farah N. A. Abd. Aziz, Sani M. B., Noor Azline M. N. and Jaafar M. S.*); Diabetes knowledge, attitudes, self-management, and quality of life among people with type 2 diabetes mellitus – a comparison between Australia- and Malaysia-based samples (*Yee Cheng Kueh, Tony Morris and Aziz-Al-Safi Ismail*); and hybrid genetic algorithm in the Hopfield network for Logic Satisfiability problem (*Mohd Shareduwan Mohd Kasihmuddin, Mohd Asyraf Mansor and Saratha Sathasivam*).

I conclude this issue with 21 articles arising from the **WRICET 2014**, **ICCMEH 2015** and **ICTMIPs 2016** international conferences: maximising potential of methane production from biogas for power generation (*Tinia Idaty, Mohd Ghazi and Ismail, Muhammad*

Nasir); anaerobic batch digestion of cattle manure under various oscillatory flow mixing (*Ismail, Muhammad Nasir, Tinia Idaty, Mohd Ghazi, Rozita, Omar and Wan Azlina, Wan Abd Karim Ghani*); car annual vehicle Kilometer travelled estimated from car manufacturer data – an improved method (*A, Shabadin, N, Megat Johari and H, Mohamed Jamil*); intelligent surveillance system for street surveillance (*Y. M. Mustafah, N. A. Zainuddin, M. A. Rashidan, N. N. A. Aziz and M. I. Saripan*); modelling the effect of sediment coarseness on local scour at wide bridge piers (*Nordila, Ahmad Thamer, Mohammad, Melville, Bruce W., Faisal, Ali and Badronnisa, Yusuf*); electromyography analysis during lifting tasks: a pilot study (*Siti Anom, Ahmad, Nor Hafeezah, Kamarudin, Rosnah, Mohd Yusoff, Mohd Khair, Hassan and Siti Zawiah, Md Dawal*); anaerobic digestion of domestic wastewater in different salinity levels: the adaptation process (*Syazwani, Idrus, Nik Norsyahariati, Nik Daud and Amimul Ahsan*); experimental and numerical study of the heat transfer characteristics of aluminium metal foam (with/ without channels) subjected to steady water flow (*Bayomy, A. M. and Saghir, M. Z.*); molecular dynamics study of the diffusion behaviour of Li in $Li_{10}GeP_2S_{12}$ (*Kawano, S. and Iikubo, S.*); temperature measurement and optimisation in machining magnesium alloy using RSM and ANOVA (*Viswanathan R, Ramesh S, Elango N and Kamesh Kumar D*); clique regular graphs (*Bhat, R. S., Bhat, Surekha, R., Bhat, Smitha, G. and Udupa, Sayinath.*); multi–band, multi–mode microstrip filters for RF, WLAN, WIMAX, and wireless communication by using stepped impedance resonator (SIR) (*Mohammed K. A. and Ratna K. Z. S.*); wireless structural health monitoring (shm) system for damage detection using ultrasonic guided waveform response (*A. H. S. Shirazi, M. K. H. Muda, M. J. Thirumurthy and F. Mustapha*); yawing force of electric trimmers of a hybrid buoyant aerial vehicle (*Haque, A. U., Asrar, W., Omar, A. A., Sulaeman, E. and Ali, J. S. M.*); impact characterisation of glass fibre-reinforced polymer (GFRP) type c-600 and e-800 using a single stage gas gun (SSGG) (*Syafiqah Nur Azrie, S., Mohamed Thariq, H. S. and Francisco, C.*); drop-weight impact test on laminated composite plate of flax (linum usitatissimum) using rice husk ash from paddy *(Oryza Sativa)* as a natural binder (*M. S. Yaakob, N. A. Z. Abidin, M. N. Abdullah, M. K. H. Muda and F. Mustapha*); different germanium dopant concentration and the thermoluminescence characteristics of flat ge-doped optical fibres (*Muhammad Safwan Ahmad Fadzil, Ung Ngie Min, David Andrew Bradley and Noramaliza Mohd Noor*); characterization of adnexal masses using multidetector contrast-enhanced CT scan – recognising common pitfalls that masquerade as ovarian cancer (*Suppiah S, Kamal SH, Mohd Zabid AZ and Abu Hassan H*); botox injections to the face: a mimic of richter's transformation (*Shafik-Eid, R, Nandurkar, D and Ramdave, S*); one-time prolonged ultrasound exposure during early pregnancy affects bone strength in young aged *Oryctolagus cuniculus* (*Che Isa, I. N., Md Dom, S., Abdul Razak, H. R. and Hashim, U. F.*); and popliteal artery to tibial plateau distance at the knee level: a radiological study to assess injury risks in osteoarthritic knees using dual source dual energy CT scan (*Ezamin, A. R., Hasyma, A. H., Suppiah, S., Suriani, M. S., Arifaizad, A., Paisal, Hussin., Nasir, M. Nizlan., Sidique, S. F. and Hariati, J.*).

I anticipate that you will find the evidence presented in this issue to be intriguing, thought-provoking, and, hopefully, useful in setting up 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. 25 (1): 1 - 16 (2017)

SCIENCE & TECHNOLOGY

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

Analysis on the Effect of Shop Floor Parameters on the Effectiveness of Preventive Maintenance through Discrete Event Simulation

Sheng, L. W.[1], Basri, E. I.[1] and Kamaruddin, S.[2]*

[1]*School of Mechanical Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia*

[2]*Mechanical Engineering Department, Universiti Teknologi Petronas, Seri Iskandar, 36210 Perak, Malaysia*

ABSTRACT

Preventive maintenance (PM) is a predetermined task that constitutes any maintenance actions performed before the quantity or quality of product equipment deteriorates. PM is primarily aimed at protecting assets, improving system reliability, and decreasing system downtime. Recently, the implementation of PM is becoming more challenging due to the increase in complexity of manufacturing systems. A case study on the effect of different parameters from a maintenance perspective on industrial shop floor performance is presented. The parameters include number of technicians, number of operations in each machine before maintenance actions, and volume of parts ordered by customers. This present research demonstrates the application of WITNESS simulation software to develop shop floor simulation model and the use of analysis of variance (ANOVA) to measure the significance of such parameters on performance.

Keywords: Analysis of Variance (ANOVA), models, parameters, performance measure, Preventive Maintenance (PM), WITNESS simulation software

INTRODUCTION

Preventive maintenance (PM) plays an important role in keeping equipment in excellent condition. The prevention of sudden malfunction and downtime during production, especially in the manufacturing industry, will improve productivity and reduce loss. PM involves various activities such as inspection, servicing, and repairing or replacing component after a specified service life. Apart from preventing equipment or machine from malfunctioning, PM improves reliability and reduces maintenance costs (Oyedepo & Fagbenle, 2011). Moreover, enhanced PM planning can lead to a globally competitive production system by reducing malfunctions and hence, downtime. This scenario can be

Article history:
Received: 10 June 2014
Accepted: 08 October 2015

E-mail addresses:
lwsheng615@gmail.com (Sheng, L. W.),
ernebasri@gmail.com (Basri, E. I.),
shahrul.k@utp.edu.my (Kamaruddin, S.)
*Corresponding Author

ISSN: 0128-7680

achieved by knowing the optimal period and frequency in carrying out PM (Eti et al., 2006). Discrete Event Simulation (DES) is a computational technique to analyse, design and operates complex systems for better understanding without affecting the real system (Alabdulkarim et al., 2013). It is an indispensable problem solving tool for addressing real-time problems (Banks, 1998). With the development of highly-advanced model, simulations are widely used in engineering, production, medical, and economic systems. It enables designers to obtain a thorough perspective of the shop floor behaviour system, especially in production (Ghasemi et al., 2012).

In the early years of this research, Eloranta and Raisanen (1987) proposed simulation models to characterise the shop floor buffer size requirement. Wallace and Mills (1988) investigated the effects of manufacturing strategy through a simulation and proved the importance of simulation in the field of manufacturing. Furthermore, O'Kane (2000) stated that simulation can be applied to justify and design production technology programmes. Viharos and Monostori (2002) conducted optimisation of production systems. By linking simulation model and the concept of key performance indicator, Bataineh et al. (2010) were able to measure the performance of public departments. Its applicability was also extended to network modelling as explained by Sukhroop et al. (2012) in wired and wireless network performance. The wide range of application shows the versatility of the DES.

In PM perspective, various methods have been proposed to determine the optimal frequency of performing the PM planning. As stated by Ab-Samat et al. (2012), developing the PM planning includes preparing for PM activities, prioritising spare parts, setting up systems and components, and preparing time estimation for maintenance operations in order to aid the decision making for actions to be taken. In addition, the documentation of the system that provides valuable information such as maintenance recommendation on the system from the manufacturer, production operations, maintenance personnel, breakdown record, and maintenance costs are important to be considered in order to properly maintain the system. Knapp and Mahajan (1998) optimised a manpower model which aims to reduce the cost of the maintenance resources by optimising the allocation of the cost of manpower based on the workload demand. The optimisation model was implemented by using a simulation analysis (SIMAN). The model provides statistical information such as the utilisation of the workers and queue length of failed systems in each area and for each craft-type. The results showed the improvement of the overall performance due to the allocation of workers, worker utilisation and queue length, which can help in making decisions for PM planning in regard of the workers' assignments. Caldeira and Guedes (2007) designed an algorithm that calculates the optimisation of PM planning through Weibull hazard function. On the other hand, Su and Tsai (2010) proposed a flexible PM planning analytical model for two parallel machines. In a more recent study, Abogrean and Latif (2012) presented a new approach using Witness simulation in combating issues related to maintenance in a cement industry. The study developed a simulation model and demonstrated on a single-unit system by focusing on the spare parts and maintenance personnel, with reference to the time-effective aspect. Thus, a better PM planning can be implemented in the actual operation based on the result of a simulation by improving the stock control system and ensuring efficient communication and teamwork throughout the facility. However, there are significant knowledge gap in the optimisation of PM planning (Almomani et al., 2012).

This present research outlines an approach that uses simulation to measure shop floor performances toward the effectiveness of PM. WITNESS simulation software is applied in conducting a case study on various variables on the shop floor with different parameters from the PM perspective. This study aims to analyse the significance of throughput rate, machine utilisation and technician utilisation towards the effectiveness of PM.

DEVELOPMENT AND ANALYSIS OF SIMULATION MODEL

The shop floor maintenance performance model approach utilised in this research is aimed at analysing the impact of conducting PM on the shop floor. WITNESS simulation software is used for the model development, verification and validation. WITNESS is a DES software that includes the visual interactive simulation. This software offers a visual display and a statistical report of the replication results and thus, ascertains errors in the flow logic during model building. The shop floor parameters can be amended at any interval throughout the simulation runs, which are prominent for an easy model-building platform. WITNESS also allows models to be built and tested in small incremental stages. Therefore, by adopting the conceptual model development (Banks, 1998), a comprehensive shop floor maintenance performance model methodology is developed. The adopted methodology is shown in Figure 1. The methodology consists of five phases. In phase 1, shop floor analysis is executed to obtain a good understanding of the shop floor environment. In phase 2, shop floor data are collected to develop the simulation model. Computations of the data, incorporated in the development and verification of the developed simulation model, are included in this phase. Performance measures are then identified and selected. Next, the experiment is designed to study the effects of performance measures. In phase 5, ANOVA is used to analyse the significant effects of the different shop floor models on performance measures. The framework is discussed in detail in the subsequent section.

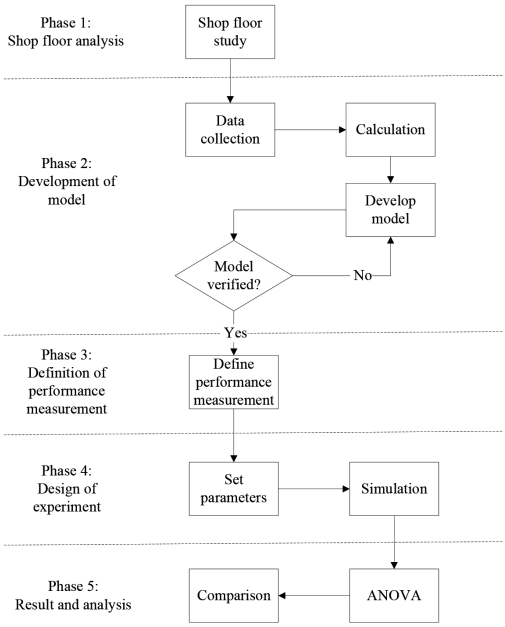


Figure 1. Development and analysis of the simulation model

Phase 1: Shop floor analysis

Shop floor analysis is a preliminary step prior to the development of the simulation model. In this phase, the product, process and layout of the department are studied. A good understanding of various variables and elements such as the flow of materials, products, human interaction and processes ensures the development of realistic simulation model to represent the real world scenario.

Phase 2: Development of the simulation model

Data are collected to develop the simulation model. The data include machine cycle time, output per shift, machine setup time and breakdown time. The collected data are then analysed to determine the parts demand ratio, output per shift, working time per shift, and average breakdown time that are crucial in the model architecture. The simulation model is developed using WITNESS. Next, the computed data are structured into the numerical elements to mimic the real-world environment. The simulation model is verified before proceeding to the stage in which the parameters are set. This step is vital in ensuring that the developed simulation model can be used in the design of experiment phase.

Phase 3: Identifying performance measures

Several authors defined performance measurement in different ways. Magretta and Stone (2002) defined performance measurement as the critical elements in translating missions or strategies of an organisation into reality. On the other hand, Tagen (2004) stated that performance measurement is a complex issue that incorporates four disciplines, namely, economics, management, accounting, and information technology. Performance measurement is also essential for an organisation to gauge its progress towards its predetermined objectives (Purbey et al., 2007). Hence, a significant performance measure must be identified and used.

Phase 4: Design of experiment

In this phase, different parameters are determined and several different models are constructed to investigate the effects of shop floor parameters on the identified performance measures. The experiments are designed by incorporating elements from the data and the determined performance measures. Three systematic experimental setups are identified to obtain a clear depiction of the problem to be analysed. Given the problem to be analysed, a good, apparent solution can be obtained, especially those related to PM. The experimental setup will be discussed in detail in the following section.

Phase 5: Result and analysis

In this phase, statistical reports obtained from the WITNESS simulation are tabulated. Based on the simulation results from each experiment, the model that performs better is selected. Next, the chosen models are analysed by using ANOVA to identify the effects of varying parameters to the shop floor performance.

CASE STUDY

The case study was conducted in a printed circuit board (PCB) manufacturing company, which is referred as Company A. Company provides a full range support and services, from circuit design, prototype fabrication and mass production up to full board assembly. This case study focuses on the punching department of Company A. The punching department generally handles single-sided printed circuit board and double-sided printed circuit board. The printed circuit board undergoes punching process to create holes for further processes or to fulfil customers' requirements. As mentioned previously, this paper focuses on the imperative effects of different shop floor parameters such as throughput rate, machine utilisation and technician utilisation on the effectiveness of PM.

Phase 1: Shop floor analysis

Shop floor observation and check are the first steps in the shop floor analysis. Observation and check is one of the qualitative means for accruing data in order to identify unlooked-for, but latent issues and variables related to the actual shop floor environment. The observation and scrutiny are performed to gain information and understand the whole shop floor processes and procedure. In this phase, the observation and scrutiny is important in the attainment of the real production operations, culture and working procedure of both the management and production point of views for this research. Therefore, before initiating the maintenance performance analysis phase, this research is directed by looking at the analysis of the shop floor a case company. This analysis is conducted on a shop floor that has multifaceted manufacturing processes and a large number of systems to perform the production operations. Observation of the actual production operations of a related manufacturing environment is the main step in this phase. The observation normally refers to the Toyota terms such as Gemba, where observers are required to go and find out the actual production floor in order to gain a better understanding of the problem and perspective. Gemba can be carried out in various ways, such as interviewing related personnel, reviewing the documentation and understanding the whole production operations in detail. Thus, the information acquired can assist in defining the problems that arise in the related study. From the observation in the simple case study, the breakdown of the system is a major problem that occurs in the manufacturing plant.

In this research, the punching department is essential for the case study prior to the development of the simulation model. Therefore, a detailed analysis was carried out on the aforementioned department. The punching department currently uses a total of 17 machines, which consist of three SWA machines, four CF machines and 10 LC machines. As indicated, the processes consign to the each machine are related to punching a through hole or a chamfer hole on the PCB. Each machine is operated by one operator. There are two dedicated technicians that are responsible for breakdown and PM. The layout of the punching department is divided into two sections, namely, the incoming line and the looping line, as shown in Figure 2. The incoming and looping lines involve different punching processes, including panel trim, subpanel, outline, and piercing. Another difference is that the looping line is used for parts that need to be processed in the punching

department for the second time. Some parts usually need to undergo other punching processes in between the incoming line and looping line. The following are examples of the processes involved in the incoming line and the looping line based on machines:

- **Incoming line:** Other processes → Incoming line punching process (SWA3→CF1→ LC12) → Other processes
- **Looping line:** Incoming line punching process (SWA2 → LC14 → LC13) → Other processes → Looping line punching process (LC8 → LC1) → Other processes

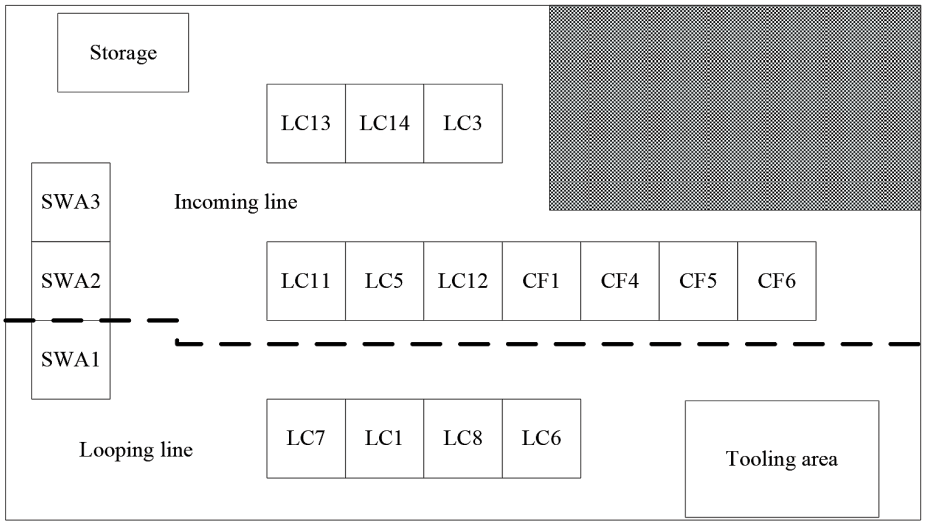


Figure 2. Shop floor layout of the punching department

Phase 2: Development of the simulation model

The initial development of the model development before it is translated into the WITNESS simulation software begins with the hand simulation. This is to ensure that the structure of the model has been well designed. A model was initially built on paper and used to analyse with WITNESS later. In creating a model through the WITNESS simulation software, the primary step was to establish the elements and variables to be included. Then, the elements and variables were detailed with their individual characteristics such as the cycle time and set-up time for each machine. The elements and variables were linked with rules and action. For example, as depicted in Figure 2, part inputted to machine LC7 was either pushed to or pulled from another machine. Then, labour (operator and technician) element was incorporated into the machine with details, such as assigning the labour to operate or set-up the workstation. As mentioned previously, there are two types of parts; single-sided part and double-sided part. The lot sizes for the single-sided part and double-sided part are 300 pieces and 80 pieces, respectively. Both parts have the same processing time of 14.4 seconds per piece in the SWA machines and 40.8 seconds per piece in the CF or LC machines. In this stage, a base model

of the shop floor is constructed by using WITNESS simulation software. The path of each part processed through different machines is set by using the PUSH and PULL rules. IF rules are applied when conditional flow is needed. PM activities are set as the setup of machines. Labour rules are programmed to ensure that specified operators are responsible for traveling and processing the parts on specified machines. Labour rules are also integrated to ensure that technicians are involved in PM activities and repair during breakdown. The verification of this model is evaluated by comparing the output per shift that resulted from the simulation and manual calculation. The following assumptions are made during the development of the simulation model:

1. The ratio of part arrivals from previous processes is constant.
2. The processing flow is continuous.
3. Changeover of tool at each machine takes place with every change in part number.
4. The production runs 24 hours a day, seven days a week, in two shifts.
5. No scrap and rejected parts are produced.
6. Operators and technicians have similar working experiences and efficiencies.

A warm-up period is likely to start empty in the simulation. However, in reality, the punching department does not start each shift without any work-in-progress. Hence, the warm-up period is set as 1440 minutes, which is equivalent to two shifts of production runs.

Phase 3: Identifying performance measures

In this study, the simulation is used to measure shop floor performances by varying shop floor parameters from the maintenance perspective. Three performance measures are chosen; throughput rate of the department, machine utilisation and technician utilisation.

Throughput rate. In this experiment, throughput rate is the number of output units per unit time going through the punching process. The throughput rate is as in equation [1]:

$$\text{Throughput rate} = \sum_{i=1}^{T} \frac{Pi}{T} \quad [1]$$

Where,
Pi : the number of parts completed in a unit time i
T : the total completion time

Machine utilisation. Machine utilisation is the occupied percentage of a machine throughout the entire run of the simulation model. In this case study, the machine utilisation per shift can be obtained from the WITNESS simulation report, and average machine utilisation can be calculated by using the summation of machine utilisation divided by the number of machines. The machine utilisation is expressed as equation [2]:

$$\text{Average machine utilisation} = \frac{\sum machine\ utilization}{number\ of\ machines} \quad [2]$$

Technician utilisation. Technician utilisation is the occupied percentage of a technician throughout the entire run of the simulation model. Similar to machine utilisation, the technician utilisation per shift can be obtained from the WITNESS simulation report, and the average technician utilisation can be calculated by using the summation of technician utilisation divided by the number of technicians. The technician utilisation is expressed as equation [3]:

$$\text{Average technician utilisation} = \frac{\sum technician\ utilization}{number\ of\ technicians} \quad [3]$$

Phase 4: Design of the experiment

Three experiments were conducted. The first experiment was conducted to study the effects of the number of technicians who were responsible for maintenance activities on the shop floor performance. The second experiment was conducted to study the effects of the number of machine operations before maintenance is performed on the shop floor performance. The last experiment was conducted to study the effects of different volumes of parts on the shop floor performance. A total of seven different models were used in this study. The first model is the base model, applied for comparison purposes, while the other six models were used for experimental purposes. All the models used the same layout, number of machines and number of operators. Each model consists of different parameters, namely, the number of technicians, maintenance schedule, and demand of volume of parts as tabulated in Table 1.

Table 1
Simulation models with specified parameters

Model	Number of technicians	Number of operations before maintenance	Volume of parts
Base model	2	100	Low
Model 1	1	100	Low
Model 2	3	100	Low
Model 3	2	80	Low
Model 4	2	120	Low
Model 5	2	100	Medium
Model 6	2	100	High

The number of maintenance technicians in the punching production area varied from one to three. The maintenance schedule is another variable considered in this experiment. The experiments are designed in which maintenance activities are carried out by one to three technicians at different time intervals considering the number of operations on each machine. The parameters varied between 80, 100 and 120 operations before maintenance the activities were carried out. Moreover, the experiments were conducted based on the volume of parts in the punching department. The part volume varied from low (1050 lots), medium (4200 lots) and high (7350 lots), respectively.

Experiment 1: *Test on significance of number of technicians on shop floor performance*
In experiment 1, base model 1 and model 2 were selected to test the significance of the number of technicians on the shop floor performance. In model 1, the number of technicians is one and in model 2, the number of technicians is three. It is assumed that 100 operations are run prior to PM activities and low volume production.

Experiment 2: *Test on significance of operation number before PM activities are conducted on shop floor performance*
In experiment 2, base model, model 3 and model 4 were selected to test the significance of the operation number on the shop floor performance before PM activities were conducted. In model 3, PM was conducted after 80 operations, whereas in model 4, PM was conducted after 120 operations. In this experiment, two technicians were used with low volume production.

Experiment 3: *Test on significance of the demand of volume parts on the shop floor performance*
In experiment 3, base model, model 5 and model 6 were selected to test the significance of the volume of parts on the shop floor performance. The volume was set to 1050, 4200 and 7350 parts for model 1, model 5 and model 6, correspondingly. In this simulation, it is assumed that 100 operations are required with two maintenance technicians.

Phase 5: Results and analysis

Once the simulation had been carried out, the WITNESS software was used to compute the statistical data from each model. Specific performance data are tabulated for further analysis. The results of the simulation runs, ANOVA, and model comparisons are shown in the following sections.

Simulation Results for Experiment 1. Table 2 outlines the results of three performance measures based on the models involved in experiment 1.

Table 2
Results of experiment 1

Model	Throughput rate (lot/minute)	Machine utilisation (%)	Technician utilisation (%)
Base model	0.023	44.35	11.59
Model 1	0.023	44.22	23.29
Model 2	0.023	44.45	3.92

The throughput rate in the base model, model 1, and model 2 are similar at 0.023 lots per minute. Slight difference was observed in terms of machine utilisation, in which base model obtained 44.35%, model 1 obtained 44.22%, and model 2 obtained 44.45%. The technician utilisation in each model varies significantly. Model 1 achieves the highest technician utilisation among all the models in this experiment (23.29%), followed by the base model (11.59%) and finally, model 2 (3.92%). ANOVA was used to test the impacts of the number of technicians on the shop floor on performance measures to decide whether to reject the null hypothesis, h_0,

or to accept the alternative hypothesis, h_1, or vice versa. In this section, although the machine utilisation in model 1 is not the highest, model 1 was selected because it achieved the highest technician utilisation. Model 1 was further analysed by using ANOVA.

Simulation Results for Experiment 2. Table 3 shows the results of three performance measures based on the models involved in experiment 2.

Table 3
Results of experiment 2

Model	Throughput rate (lot/minute)	Machine utilisation (%)	Technician utilisation (%)
Base model	0.023	44.35	11.59
Model 3	0.023	53.53	15.99
Model 4	0.023	44.42	10.99

In contrast to experiment 1, the throughput rate in base model, model 3 and model 4 are similar at 0.023 lots per minute. In order to determine the influence of operation number before PM activities were conducted on shop floor performance, ANOVA was used. It is to decide whether to reject the null hypothesis, h_0, or to accept the alternative hypothesis, h_1, or vice versa. Model 3 has the highest machine utilisation at 53.53%, compared with machine utilisation of the base model at 44.35%, and model 4 at 44.42%. Moreover, model 3 achieves the highest technician utilisation, which is 15.99%, and is 4.4% and 5% higher than the technician utilisation of the base model and model 4, respectively. Obviously, model 3 performed the best among all the models in experiment 2. Hence, model 3 was selected for further analysis by using ANOVA.

Simulation Results for Experiment 3. Table 4 shows the results of three performance measures based on the models involved in experiment 3.

Table 4
Results of experiment 3

Model	Throughput rate (lot/minute)	Machine utilisation (%)	Technician utilisation (%)
Base model	0.023	44.35	11.59
Model 5	0.027	69.67	19.63
Model 6	0.024	68.46	18.41

Differences in throughput rate, machine utilisation and technician utilisation among the models were observed in experiment 3. Model 5 has the highest throughput rate at 0.027 lots per minute, followed by the throughput rate in the base model and model 6 at 0.023 lots per minute and 0.024 lots per minute, respectively. Furthermore, the machine utilisation in model

5 is optimum among the three models in experiment 3 at 69.67%. In model 6, the machine utilisation was slightly lower than model 5 at 68.46%, whereas the base model obtained the lowest utilisation at 44.35%.

In terms of technician utilisation, model 5 achieved the highest utilisation with 19.63%, which is 1.22% higher than that of model 6 and 8.04% higher than that of the base model. ANOVA is adopted to define the impacts of demand of volume parts on shop floor performance. It is to establish whether to reject the null hypothesis, h_0, or to accept the alternative hypothesis, h_1. Overall, model 5 achieved the best performance compared with the other models in this experiment. Hence, model 5 was chosen for the ANOVA analysis.

ANALYSIS OF VARIANCE (ANOVA)

ANOVA is used to test the effects of the different models with specified parameters on the performance measures. In this phase, base model, model 1, model 3, and model 5 were chosen for ANOVA based on their significant performances in the previous experiments. One-way ANOVA at 95% confidence level is used to test the null hypothesis (h_0) in which the different models have no significant effects on performance measures and the alternative hypothesis (h_1) in which the different models significantly affect performance measures. The null hypothesis (h_0) is rejected and the alternative hypothesis (h_1) is accepted when the *F* value computed from ANOVA is higher than the *F* critical value, and vice versa.

ANOVA for throughput rate

In this section, ANOVA is used to test the following hypotheses:

h_0: The different models have no significant effects on the throughput rate.

h_1: The different models have significant effects on the throughput rate.

After ANOVA has been carried out, the result is tabulated in Table 5.

Table 5
ANOVA results for throughput rate

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	0.000727	3	0.000242	1.186224	0.315666	2.642851
Within groups	0.048238	236	0.000204			
Total	0.048966	239				

The obtained *F* value is 1.186224. This value is lower than the *F* critical value, 2.642851. In this scenario, the hypothesis h_1 is rejected, while hypothesis h_0 is accepted. The result indicates that different shop floor models have no significant effects on the throughput rate.

ANOVA for Machine Utilisation

In this section, ANOVA is used to test the following hypotheses:

h_0: The different models have no significant effects on machine utilisation.

h_1: The different models have significant effects on machine utilisation.

After ANOVA has been completed, the results are shown in Table 6.

Table 6
Results of ANOVA on machine utilisation

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	25807.85	3	8602.618	4.496066	0.004334	2.642851
Within groups	451554.2	236	1913.365			
Total	477362.1	239				

The obtained *F* value, 4.496066, is greater than the *F* critical value, 2.642851. Thus, the hypothesis h_0 is rejected, but hypothesis h_1 is accepted. The result indicates that different shop floor models significantly affect machine utilisation.

ANOVA for Technician utilisation

In this section, ANOVA is used to test the following hypotheses:

h_0: The different models have no significant effects on technician utilisation.

h_1: The different models have significant effects on technician utilisation.

After the ANOVA test has been carried out, the results are tabulated in Table 7 below.

Table 7
Results of ANOVA on technician utilisation

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	4509.809	3	1503.27	2.71367	0.045587	2.642851
Within groups	130735	236	553.9618			
Total	135244.8	239				

In this analysis, the obtained *F* value, 2.71367, is greater than the *F* critical value of 2.642851. The hypothesis h_0 is rejected, but hypothesis h_1 is accepted. The results indicate that the different shop floor models significantly affect technician utilisation.

MODEL COMPARISON

In this section, the performance measures of the chosen models are plotted in the column chart for comparison purpose. Based on the ANOVA test performed previously, the different models were found to significantly affect machine utilisation and technician utilisation without any significant effect on the throughput rate. Thus, the comparison is only based on machine utilisation and technician utilisation.

Comparison of machine utilisation

Figure 3 shows a comparison of machine utilisation between the base model, model 1, model 3, and model 5. The comparison was done to determine the effects of different parameters, including the number of technicians, number of operations at each machine before PM was conducted, and volume of parts on the machine utilisation of the shop floor. Each model of 1, 3 and 5 represents certain parameters with reference to machine utilisation. In Figure 3, it can be seen that model 3 and 5 have better performances compared to the other models.

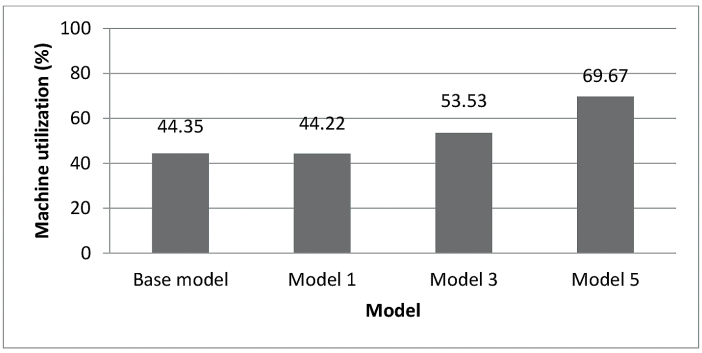


Figure 3. Graph of models against machine utilisation

Model 5 shows the highest machine utilisation, followed by model 3, base model, and model 1, with machine utilisation readings of 69.67%, 53.53%, 44.35%, and 44.22%, respectively. A higher machine utilisation percentage can eliminate more wastage during production. Hence, model 5 is selected among the models.

For a system that is highly dependent on manpower for maintenance, machine reliability is an important factor to ensure the resilience of maintenance systems should opposing scenarios occur for PM. It is vital that PM is planned precisely for optimal result. Machine reliability can be quantified by computing the probability of a technician to successfully complete a maintenance task without any mistakes or errors.

Comparison of technician utilisation

Figure 4 shows a comparison of technician utilisation among base model, model 1, model 3, and model 5. The comparison was done to determine the effects of different parameters, including the number of technicians, number of operations at each machine before PM was conducted, and volume of parts, on the technician utilisation of the shop floor.

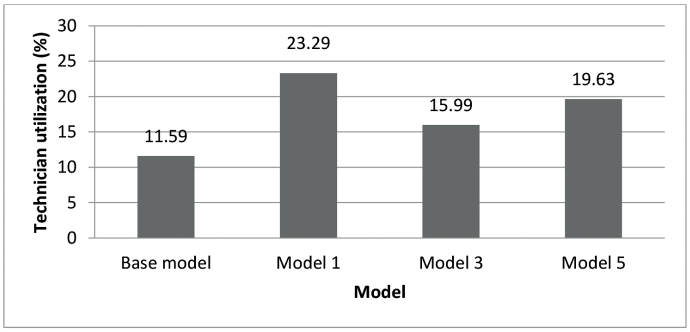


Figure 4. Graph of models against technician utilisation

Based on the chart, model 1 has the highest technician utilisation at 23.29%. Although the maintenance takes place after 80 operations for each machine in model 3 compared with 100 operations for model 1, the technician utilisation of model 1 is still 7.33% higher than that in model 3. The result could be due to the difference in the number of technicians; model 1 has only one, and model 3 has two. In model 1, the technician performs all the maintenance work, whereas in model 3, two technicians share the task.

The utilisation of technicians is highly dependent on their skills and job scope. It can be translated to working-hour productivity. A highly skilled technician may reduce the machine downtime and improve machine availability. The number of tasks the personnel performed in a week can be compared to others in determining the efficiency of a technician.

The results of this case study can be used by Company A to make informed decisions about improving the performances of the punching department. This study creates the shop floor design alternatives for Company A, which could then enable the company to deal with PM issues in enhancing machine or technician utilisations. Model 5 can be adopted to improve machine utilisation performance, in which machines undergo PM after every 100 lots of operations with two technicians at 4200 lots volume. Model 1 can be used to improve technician utilisation performance. In this case, a single technician is employed and PM takes place after every 100 lots of operations at 1050 lots production volume.

CONCLUSION

This study demonstrates the employment of DES as a tool to measure shop floor performance, especially when the shop floor parameters need to be changed due to maintenance issues. The simulation can be used by companies to obtain cost saving designs in predicting future shop floor performance rather than implementing a new design of maintenance parameters and measuring performances through trial and error. Thus, a systematic approach can be employed to address unforeseen issues. Based on the case study in the punching department of Company A, shop floor parameters have no significant effects towards its throughput rate. However, machine and technician utilisations are highly dependent on shop floor parameters. In the punching shop floor, model 5 can be used to obtain a higher machine utilisation percentage, while model 1 can be considered to obtain a higher technician utilisation percentage. The

methodology developed in this research is simple and easy for implementation, and suited for applications in actual industries. The five phases in the plan can be adopted in various shop floors. Emphasis should be given on understanding the current situation of the company and translating the situation into a replicated scenario, with a list of employees responsible for each task in the system model. Improvement is then implemented by including as much as possible information for each task to achieve overall shop floor effectiveness. Thus, the maintenance planning protocol is continuously revised to detect imperfections. This study can be further improved by including different shop floor layouts and inspection operations. These variables can enhance the simulation model, which can lead to the advancement of performance measures in a shop floor.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the Government of Malaysia through the CREST grant scheme for funding this research.

REFERENCES

Abogrean, E., & Latif, M. (2012). Effective maintenance enabled by the use of witness simulation in a Libyan cement factory. *International Journal of Advances in Management and Economics, 1*(5), 138-147.

Alabdulkarim, A. A., Ball, P. D., & Tiwari, A. (2013). Applications of simulation in maintenance research. *World Journal of Modelling and Simulation*, *9*(1), 14 – 37.

Almomani, M., Abdelhadi, A., Seifoddini, H., & Xiaohang, Y. (2012) Preventive maintenance planning using group technology: A case study at Arab Potash Company, Jordan. *Journal of Quality in Maintenance Engineering*, *18*(4), 472 – 480.

Banks, J. (Ed.). (1998). *Handbook of simulations principle, methodology, advances, application and practice*. John Wiley and Sons. Inc.

Bataineh, O., Al-Aomar, R., & Shakra, A. A. (2010). Simulation-based optimization for performance enhancement of public department. *Jordan Journal of Mechanical and Industrial Engineering, 4*(3)*,* 346-351.

Caldeira, D. J., & Guedes, S. C. (2007). Optimisation of the preventive maintenance plan of a series components system with Weibull hazard function. *Electronic Journal of International Group on Reliability, 2*(3-4), 33-39.

Eloranta, E., & Raisanen, J. (1987). Evaluation and design of plant layout by simulation. In *Proceedings of the Third International Conference on Simulation in Manufacturing* (pp. 11-22).

Eti, M. C., Ogaji, S. O. T., & Probert, S. D. (2006). Development and implementation of preventive-maintenance practices in Nigerian industries. *Applied Energy, 83*(10), 1163-1179.

Ghasemi, F., Momeni, M., Sharifi, A., & Piran, M. (2012). A Simulation Model of Production Scheduling. *World Applied Sciences Journal, 18*(6), 861-867.

Knapp, G. M., & Mahajan, M. (1998) Optimization of maintenance organization and manpower in process industries. *Journal of Quality in Maintenance Engineering*, *4*(3), 168 – 183.

Magretta, J., & Stone, N. (2002). *What management is: How it works and why it's is everyone's business*. New York, NY: Free Press.

O'Kane, J. F., Spencekley, J. R., & Taylor, R. (2000). Simulation as an essential tool for advance manufacturing technology problems. *Journal of Materials Processing Technology*, *107*(1), 412-424.

Oyedepo, S. O., & Fagbenle, R. O. (2011). A study of implementation of preventive maintenance programme in Nigeria power industry–Egbin thermal power plant, case study. *Energy and Power Engineering*, *3*(03), 207-220.

Purbey, S., Mukherjee, K., & Bhar, C. (2007). Performance measurement system for healthcare processes. *International Journal of Productivity and Performance Management, 56*(3)*,* 241-251.

Samat, H. A., Jeikumar, L. N., Basri, E. I., Harun, N. A., & Kamaruddin, S. (2012). Effective preventive maintenance scheduling: a case study. In *Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management* (pp. 1249-1257). Istanbul, Turkey.

Su, L. H., & Tsai, H. L. (2010). Flexible preventive maintenance planning for two parallel machines problem to minimize makespan. *Journal of Quality in Maintenance Engineering, 16*(3)*,* 288-302.

Sukhroop, P., Singla, K., & Singla, A. (2012). Simulate the performance parameters of wired and wireless networks by soft computing technique. *International Journal of Engineering Research and Application (IJERA)*, *1*(2), 25-35.

Tagen, S. (2004). Performance measurement: from philosophy to practice. *International Journal of Productivity and Performance Management*, *53*(8), 726-737.

Viharos, Z. J., & Monostori, L. (2002). Optimisation of production system using simulation and artificial intelligence techniques. In *Third Conference on Mechanical Engineering* (pp. 579-583).

Wallace, R., & Mills, R. I. (1988). Simulation as an essential element of manufacturing strategy. In *Proceedings of Fourth International Conference on Simulation in Manufacturing* (pp. 227-237).

Pertanika J. Sci. & Technol. 25 (1): 17 - 28 (2017)

SCIENCE & TECHNOLOGY

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

Analysis of Energy Consumption in Pineapple Cultivation in Malaysia: A Case Study

Nazri, A. M.[1] and Pebrian, D. E.[2]*

[1]*Faculty of Plantation and Agrotechnology, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia*

[2]*Faculty of Plantation and Agrotechnology, Universit Teknologi MARA Melaka, 77300 UiTM, Merlimau, Melaka, Malaysia*

ABSTRACT

Pineapple (*Ananas Comosus)*, which has been identified as high-value non-seasonal tropic fruit, is one of the top five fruits in Malaysia that shows a promising demand in the local and export markets. The current study was conducted to analyse the efficiency energy inputs used in every activity involved in the process of pineapple production. The study also aimed to contribute in initiating a database for energy used in producing pineapples in Malaysia, which would hopefully become a guideline for applications in related policies in pineapple plantations. Data were collected based on direct field observations at a pineapple plantation under the management of Koperasi Serbaguna Anak-Anak Selangor Bhd. (KOSAS Bhd.) at Kg. Kundang in Selangor, Malaysia. Oral interviews were also carried out to gather any relevant information. The highest energy input used in pineapple production is fertiliser, contributing to 45.65% from the total energy equivalents used in the production. Fuel is the second highest energy source with a segment of 20.21% of the total energy, followed by planting material, agrochemicals, human labour and machine at 17.33%, 12.76%, 3.34 % and 0.69%, respectively. The ratio of energy output/inputs in pineapple production in the study area was 3.56. Conclusively, it means the energy input is effectively used in the pineapple production in the study area since the energy output/inputs ratio is greater than 1. Involvement of machinery in the pineapple cultivation practices in Malaysia is recommended since it reduces human drudgery and optimises farm's capability to do work more efficiently and also to offset the labour shortage problem.

Keywords: Pineapple plantation, energy consumption, energy efficiency, sustainable agriculture, mechanisation

Article history:
Received: 24 February 2016
Accepted: 17 June 2016

E-mail addresses:
darius@melaka.uitm.ed.my (Nazri, A. M.),
nazriemanap@gmail.com (Pebrian, D. E.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

In Malaysia, pineapple *(Ananas Comosus)*, which has been identified as a high-value non-seasonal tropic fruit, is one of the top five

fruits that shows a promising demand in the local and export markets. The focus in pineapple cultivation is to increase its production on mineral and peat soils, mechanise pineapple production and add value to the fruit before it reaches the market. There is a growing interest among growers by planting a fresh fruit variety as it seems to have better demand in both domestic and foreign markets. The rising star for the pineapples is the MD2 variety, a hybrid variety that originates from Hawaii.

Statistically, about 90% of the pineapples in Malaysia are cultivated on peat soil, while the rest are planted on mineral soil (Chan, 2000). Based on the statistics by the Malaysian Pineapple Industry Board (2012) shown in Table 1, the pineapple production fluctuated between 2008 and 2010. At that time, the pineapple production from pineapple small holders increased, while the production from the plantations decreased by years.

Table 1
Production of pineapples (in metric tons)

Year	Small holder	Portion of total production (%)	Plantation	Portion of total production (%)	Total
2007	12,109	17.40	57,498	82.60	69,607
2008	98,895	63.35	57,216	36.65	156,111
2009	59,164	51.47	55,794	48.53	114,958
2010	79,158	62.13	48,257	37.87	127,415

Source: MPIB (2012)

Nowadays, the agricultural sector relies heavily on the input from the fossil fuels to produce outputs. It is a well-known fact that a high consumption of this source of input results in negative environmental effects through the release of CO_2 and other greenhouse gases that leads to global warming. Ironically, even though we are aware of these bad effects, the continuous use of fossil based fuels is inevitable in modern farming practices to increase yields and reduce the risks of production loss. This is due to the intensification of agricultural practices to ensure food security for the growing human population. Mechanisation, agrochemicals, fertilisers and herbicides are among farm inputs produced by fossil fuel sources. Thus, searching for agricultural methods that require lesser energy input with higher energy productivity has become one of the important issues today (Refsgaard, 1998). As human activities are closely related to energy use and the cost it incurs, energy analysis can be a good tool for providing an overview of the energy used for an activity. Effective energy use in agriculture is one of the conditions for sustainable agricultural production as it leads to financial saving, better fossil fuel preservation and lesser air pollution (Pimentel & Patzek, 2005).

Many relevant studies have been done so far; however, these works generally focused on world's main crops production such as wheat, rice, soybean, cotton and sugarcane to improve the energy output–input analyses and investigate their relationships (Ricaud, 1980; Mandal et al., 2002; Safa & Tabatabaeefarz, 2002; Bockari et al., 2005; Nuray, 2009). Yet, in Malaysia, there is not much work done to investigate energy input in crop cultivation. Meanwhile, several related studies reported by Bockari et al. (2005) and Pebrian et al. (2014) focused on certain main crops such as rice and oil palm cultivations. A similar study on energy consumption in

pineapple cultivation in Hawaii was reported by Duke (1983). However, the available data from the previous research could not be utilised in analysing energy consumption for pineapple production in Malaysia because of the differences in the cropping system, soil characteristics and nature of works. Thus, energy input/output analyses are urgently needed to determine the energy efficiency and energy pattern, as well as their distribution in pineapple plantations in Malaysia.

The study aimed to analyse energy input/output used in the pineapple production in Malaysia. The distribution of energy input used in the production of pineapple, energy efficiency and energy pattern were also studied. The study could contribute to initiation of a database for energy consumption in pineapple production that can serve as a guideline for applications within related strategic farm management policies.

MATERIAL AND METHODS

The data of this study were obtained through field observations carried out from April to July 2012 at a pineapple plantation under the management of Koperasi Serbaguna Anak-Anak Selangor Bhd. (KOSAS Bhd.) located in Kg. Kundang in Selangor, Malaysia. With a planting area of 300 acres, this plantation is among the largest pineapple plantations in Selangor and it can be considered as a representative of a typical pineapple plantation in Malaysia. The type of soil in the study area is peat soil with a good natural water supply system (ditch). Thus, with this condition, irrigation activity was excluded in the data collection. The data were collected for all the field operations involved in pineapple production process in the field such as land preparation, planting, fertilising, weeding, spraying and harvesting.

Description of field operations

The description of field operations involved in pineapple production process at the Koperasi Serbaguna Anak-Anak Selangor Bhd. (KOSAS Bhd.) plantation is as follows:

Land preparation. Since the type of soil in this area is peat soil with a good natural water supply system, thus land preparation is started by constructing drainage to lower the water level for the requirement of pineapple cultivation. A ditcher was used to make ditches to remove excessive water for the plantation areas. The ditches were also cleared from weeds and any hindrance material once every three months. In replanting areas, a 37 kW tractor-mounted sprayer, a pair of *2000 litres herbicide* tanks was used to kill the crop residues left in the areas before planting begins. After a week, slashing and burning were employed to clear and clean the entire areas from any debris and crop residues so that the areas are ready to be planted with pineapples.

Planting. This plantation employs a high density planting system with a total of 22000 pineapples planted per acre with planting distances of 30 cm x 60 cm. Planting was started with lining task. Tied wood is used to make lines in planting rows. A special tool made from wood, which is locally known as "Tugal", is utilised to dig planting holes. The workers used

the "Tugal" wood dig planting holes at a depth of 10-15 cm while walking in the planting areas. Prior to planting, all the planting materials were delivered to a location near the planting areas using a 37 kW tractor-mounted trailer.

Fertilising. There two types of fertiliser, namely, foliar fertiliser and granular fertiliser, applied in this plantation. The foliar fertiliser is chemical fertilisers such as ferrous sulphate, zinc sulphate, hydrate lime, copper sulphate and urea that are mixed with water. They were applied to the planting areas when the pineapples are 6 and 18 weeks old. Meanwhile, the granular is a compound fertiliser or NPK fertiliser utilised to the planting areas when the pineapples are 12, 14 and 34 weeks old. The granular fertiliser was distributed manually by the workers using their hands, while the foliar fertiliser was spread to the planting areas through the use of a 37 kW tractor-mounted high pressure sprayer. The sprayer tank can accommodate 2000 litres of liquid solution at a time.

Spraying hormone. Spraying hormone was performed when pineapples are 40 weeks old for the purpose of flower induction. Similar to the fertilising operation, this plantation also uses a 37 kW tractor-mounted high pressure sprayer with the tank capacity of 2000 litres to spray hormone to targeted crops.

Weeding. In this plantation, weeding was carried out every month until the pineapples are at age 40 weeks. The weeding was done manually using simple hand tools to remove weeds off the ground. This operation is considered as a laborious task because the planting distances are too narrow.

Crop Protection. A "cap" was used to protect the pineapple's crown from sunburn. The cap was made from high density paper and fitted on crop when it has formed fruit and crown. Like weeding, this operation was also a painstaking job and undertaken manually.

Harvesting. The harvesting operation is accomplished manually. The worker brings along a basket to harvest ripe fruits by using a sharp machete while walking in the harvesting areas. The basket was made from bamboo material and utilised to store the harvested fruits during the harvesting operation. When the basket is fully loaded, the worker moves while carrying along the basket to the temporary collecting point at the main road side to unload the harvested fruits. A 37 kW tractor-mounted trailer waiting at the temporary collection point was employed to transport the harvested fruits to the packing house.

These operations were observed for 12 (twelve) experimental plots based on an acreage basis. Oral interviews were also conducted with the plantation manager and workers to gather all the relevant information.

Data Analysis

Energy analysis was performed based on the above field operations, as well as on the direct and indirect energy sources involved in the pineapple production process. Firstly, measurements of inputs such agrochemicals (kg), human labour (h), machinery (h), fertilisers (kg), fuel (L) and seeds (kg) used in the production of pineapple were specified in order to calculate energy

equivalence in the study. For the operation using a self-propelled machine, fuel consumption was measured by filling the machine's fuel tank twice, before and after performing each operation based on Alcock (1996). All of the experimental plots were managed with the same field practices including land clearing and planting method, amount of fertiliser and agrochemicals used, crop protection and harvesting system in order to reduce the significance of differential on crop yield.

Fuel consumption for tractor was estimated based on the formula from ASABE Standards (2009), as follows:

The average diesel consumption (l/h) = rated PTO power (kW) X 0.305 literkWh^{-1}. (1)

Energy was calculated according to the obtained input, multiplied with the coefficient of energy equivalent from the previous research literature, as indicated in Table 2. The unit of the results was in megaJoule (MJ) per hectare term. The energy contributed by machine can be calculated according to Moerschner and Gerowitt (2000), as follows:

Energy of machinery = [weight of machine (kg) X coefficient of energy for machine (MJkg^{-1}) X working hour (hha^{-1}) X number of application] / wear-outlife of machinery (2)

Using the summation of energy equivalences of all inputs in MJ terms, the total input equivalent can be calculated. Based on the energy equivalent of the inputs and output (Table 2), energy ratio (energy use efficiency) and energy productivity were calculated according to Singh *et al*. (1998) and Mandal *et al.* (2002) , as follows:

Energy use efficiency = Energy output (MJha^{-1}) /Energy input (MJha^{-1}) (3)

Energy productivity = Pineapples output (kgha^{-1}) /Energy input (MJha^{-1}) (4)

Table 2
Energy equivalents for different inputs and outputs in agricultural production

	Unit	Energy equivalent (MJ unit^{-1})	Reference
Input			
1) Labour	h	1.96	Safa and Tabatabaeerfar (2002)
2) Machinery			
Tractor 50 hp	kg	109.00	Pimentel (1992)
Petrol engine, 5hp	kg	109.00	Pimentel (1992)
High pressure sprayer	kg	109.00	Pimentel (1992)
3) Fertiliser			
Nitrogen (N)	kg	61.53	Pimentel and Patzek (2005)
Phosphorus (P_2O_5)	kg	12.56	Pimentel and Patzek (2005)
Potassium (K_2O)	kg	6.70	Pimentel and Patzek (2005)
4) Agrochemicals			
Hydrated lime	kg	1.17	Pimentel and Patzek (2005)
Micronutrient*	kg	20.90	Anon (2004)
Ethrel	kg	255.00	Anon (2004)
Paraquat	kg	459.00	Anon (2004)
Glyphosate	kg	453.00	Anon (2004)
5) Fuel			
Diesel	l	56.31	Safa and Tabatabaeerfar (2002)
Petrol	l	46.30	Safa and Tabatabaeerfar (2002)
6) Suckers	kg	1.90	Ricaud (1980)
Output			
7) Fruits	kg	1.90	Singh and Mittal (1992)

Finally, the energy input was divided into direct, indirect, renewable and non-renewable forms. The indirect energy includes the chemical and farm fertilisers, seeds and machinery. The direct energy includes human labour and fuel. The non-renewable energy sources include fuel, fertiliser, pesticides and machinery, while the renewable energy sources include human labour and seeds.

RESULTS AND DISCUSSION

Operational energy consumption based on field operations

The operational energy consumption in the pineapple production system was distributed according to the following field operations: land preparation, planting, fertilising, hormone spraying, weeding, crop protection and harvesting. As shown in Table 3, the average operational energy consumption for fertilising of 16264.33 $MJha^{-1}$ was the highest, which accounted for about 54.61% of the computed total operational energy consumption of 29780.91 $MJha^{-1}$, followed by planting of 6512.03 $MJha^{-1}$ (21.87%), land preparation of 4056.18 $MJha^{-1}$ (13.62%), hormone spraying of 1371.74 $MJha^{-1}$ (4.61%) and harvesting of 1170.73 $MJha^{-1}$ (3.39%) of the total energy. Nonetheless, weeding and crop protection did not make any significant contribution to the operational energy consumption. The high operational energy for fertilising operation observed in the study was due to the fact that the fertilisers (N, P, K) incurred a high energy equivalent in the process, especially nitrogen (N). In this study area, most of the field operations were done manually. The involvement of machinery in the field operations was only for spraying activities. In Malaysia, the use of machinery for pineapples is still very low due to some circumstances like lacking the financial capability, uneconomical size of planting areas for mechanisation and problematic characteristics of peat soil. These are among the reasons for the continuous reliance on human labour in the plantation. As shown in Table 4, the planting operation was claimed to be sustaining the highest labour input with energy equivalent to 343.43$MJha^{-1}$, followed by weeding, crop protection and harvesting. As mentioned earlier, all these operations were done manually.

Table 3
Distribution of operational energy consumption by field operations

Field Operation	Energy used (MJ ha-1)	Portion (%)
Land preparation	4056.18	13.62
Planting	6512.03	21.87
Fertilising	16264.33	54.61
Hormone spraying	1371.74	4.61
Weeding	236.55	0.76
Crop protection	169.34	0.57
Harvesting	1170.73	3.39
Total	29780.91	100

Total energy consumption based on energy sources

The allocation of inputs and their energy equivalences used in the production of pineapples are shown in Table 4. The highest share of input equivalent energy is indicated by the fertiliser usage which accounted about 45.65% of the total equivalent energy in the pineapple production. In specific, the fertiliser used comprised of 79.4 kg of nitrogen, 19.20 kg of phosphorus and 56.40 kg of potassium. The second highest equivalent energy used in the production of pineapples was the fuel consumption, which was about 20.22% of the total equivalent energy. These were followed by other inputs such as planting material or sucker (17.33%), chemical application (12.76%), human labour (3.34%) and machinery (0.69%). Fertilisers and fuel were instigated as the highest energy simply because the energy equivalents for these two inputs were also very high. For fertiliser, nitrogen energy equivalent was 61.53 MJ kg^{-1}, whereas for fuel, diesel's energy equivalent was 56.31 MJ kg^{-1} and petrol's energy equivalent was 46.30 MJ kg^{-1}. Even though chemicals like flowering hormone also have a high energy equivalent (255 MJ kg^{-1}), the small quantity of hormone used in the cultivation has caused it to not giving any significant increment to the input. As mentioned earlier on, human labour is the backbone in the process of pineapple production since almost all of the field operations are done manually. However, as the calculation goes on, the low energy equivalent for labour (1.90 MJ h^{-1}) is simply because of its small contribution to the total equivalent energy.

Table 4
Allocation of physical inputs used in pineapple production

Input	Unit	Amount of input used per hectare	Energy equivalent ($MJunit^{-1}$)	Energy equivalent ($MJha^{-1}$)	Portion (%)
Labor (h)					3.34
Land preparation	h	13.34	1.96	26.13	
Planting	h	175.22	1.96	343.43	
Weeding	h	120.68	1.96	236.55	
Fertilizing	h	25.61	1.96	50.19	
Spraying hormone	h	2.91	1.96	5.71	
Crop protection	h	86.40	1.96	169.34	
Harvesting	h	83.88	1.96	164.40	
Machinery(kg)					*0.69*
Tractor, 50 hp	kg	4520.1	109	205.28[a]	
Petrol engine, 5 hp	kg	37.05	109	0.49[a]	
High pressure sprayer	kg	24.7	109	0.32[a]	
Fertiliser (kg)					*45.65*
Nitrogen (N)	kg	196.12	61.53	12067.14	
Phosphorus (P_2O_5)	kg	47.42	12.56	595.64	
Potassium (K_2O)	kg	139.31	6.70	933.36	

Table 2 (*continue*)

Input	Unit	Amount of input used per hectare	Energy equivalent ($MJunit^{-1}$)	Energy equivalent ($MJha^{-1}$)	Portion (%)
Agrochemicals (kg)					*12.76*
Hydrated lime	kg	50.00	1.17	144.50	
Micronutrient[b]	kg	7.50	20.90	387.17	
Ethrel	kg	0.48	255.00	302.33	
Paraquat	kg	0.84	459.00	952.33	
Glyphosate	kg	1.80	453.00	2014.04	
Fuel (L)					*20.22*
Diesel	l	41.64	56.31	5791.53	
Petrol	l	2.00	46.30	228.72	
Suckers(kg)	*kg*	*1100.00*	*1.90*	*5162.30*	*17.33*
Total energy input ($MJ\ ha^{-1}$)				29780.91	100
Yield of pineapple ($kg\ ha^{-1}$)		55749.78	1.90	105924.59	
Ratio of energy output/input				3.56	
Energy productivity ($kg\ MJ^{-1}$)				1.87	

[a] Based on calculation using equation [2] and the average usage of machinery is 1.2 hours.
[b] Ferrous sulfate, copper sulfate, zinc sulfate

In specific, the energy equivalent used for labour, machinery, fertiliser, agrochemical, fuel (including the plant material) was about 29780.91 MJ ha^{-1} for 14 months of pineapple cultivation or equivalent to 2127.21 MJ ha^{-1} per month. The monthly input of 2127.21 MJ ha^{-1}energy needed for the pineapple production at KOSAS Bhd. plantation was 47.29% lower compared to the total of 4035.65 $MJ.ha^{-1}$ energy per month required by the pineapple plantations in Hawaii (Duke, 1983). Duke (1983) reported that the pineapple production in Hawaii requires monthly inputs of 60.71 $MJ.ha^{-1}$ for manual labour, 205.15 $MJ.ha^{-1}$ for machines, 2271.76 $MJ.ha^{-1}$ for fuel, 1418.91 $MJ.ha^{-1}$ for fertilisers and 79.13 $MJ.ha^{-1}$ for pesticides. The different soil characteristics and nature of works have caused pineapple production in KOSAS Bhd plantation requires lesser energy inputs than that of the plantations in Hawaii. As mentioned earlier, in KOSAS Bhd. plantation, pineapple is cultivated on peat soil area, while in Hawaii, pineapple is mainly grown silt loams, silty clay loams and silty clay (Hepton, 2003). Problematic characteristics of peat soil, with an extreme low bearing capacity, have limited the use of heavy machinery in KOSAS Bhd. plantation. Thus, fuel usage is less as a result of limited machinery use. Besides, KOSAS Bhd. plantation also adopted no-tillage method and minimum fertiliser input for growing pineapples due to higher fertility of prepared peat soil in this plantation. These conditions have enable KOSAS Bhd. plantation to save more energy inputs in pineapple production.

Energy use efficiency

Based on the yield per hectare, the total energy output in pineapple cultivation in KOSAS Bhd. plantation was 55749.78 kg ha^{-1}. The energy equivalent to produce the yield is 29780.91 $MJha^{-1}$. Therefore, the energy productivity indicates that the product amount obtained from the unit area in return for the used energy amount is 1.87 MJ kg^{-1} (Table 4). In other words, for 1.87 kg of produce, 1 MJ of energy is used in the farm. The energy use efficiency is determined as the ratio of output energy to input energy. This in line with the statement by the International Energy Agency (2016), who states that something is more energy efficient if it delivers more services for the same energy input, or the same services for less energy input. Thus, it is assumed that if the energy ratio is greater than 1, the production system is therefore gaining energy; otherwise, it is losing energy. The energy use efficiency in the pineapple production was 3.56 (Table 4). It is shown that the pineapple cultivation by KOSAS Bhd. earned at least 3.56 times of the energy inputs given into the production process. This can be considered as a very efficient use of inputs. Thus, the calculated energy use efficiency of the pineapple production in the study area is more efficient as compared to the respective output/input ratio of 1.10 for apple production in Turkey (Nuray, 2009), 1.24, 1.31 and 3.37 for apricot production (Gezer et al., 2003; Esengun et al., 2007). Moreover, it is also higher than the ratio of 1.25, 1.06 and 1.17 for the production of orange, lemon and mandarin, respectively (Ozkan et al., 2004), as well as of rice production range, i.e. from 1.03 to 1.76, in US (Duke, 1983). However, the calculated energy use efficiency of pineapple production was lower compared to the energy use efficiency in the rice cultivation in Malaysia, with a ratio of 8.86 (Bockari et al., 2005). Nonetheless, the irrigation input was not included in their study.

Energy of direct, indirect, renewable and non-renewable

Finally, the energy input in the production of pineapple was divided into direct and indirect, and also renewable and non-renewable forms as presented in Figures 1 and 2. Direct energy is the energy that is invested physically in the farm like labour, fuel and electricity, while indirect energy is the energy that comes from the manufacturer of the inputs like fertilisers, agrochemicals and machinery. The direct energy was 23.56% lower than the indirect energy resource, which was 76.44% of the total energy input. The same thing also goes to the portion of renewable and non-renewable energy, where it was fairly different from each other. The portion of renewable energy inputs within the total energy was 20.68%, which is considered as very low as compared to the overall portion of energy from renewable resources in Iceland (72%), Norway (64%), New Zealand (32%), Chile (29%), Canada (24%) and Switzerland (24%) (Eurostat, 2016).

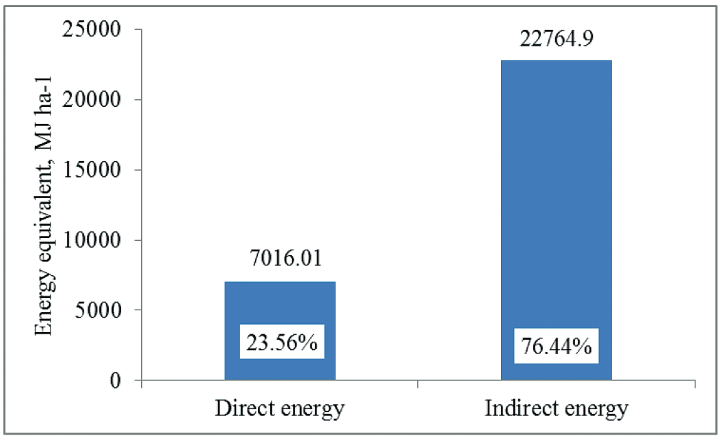


Figure 1. Distribution of direct and indirect energy inputs in the production of pineapples

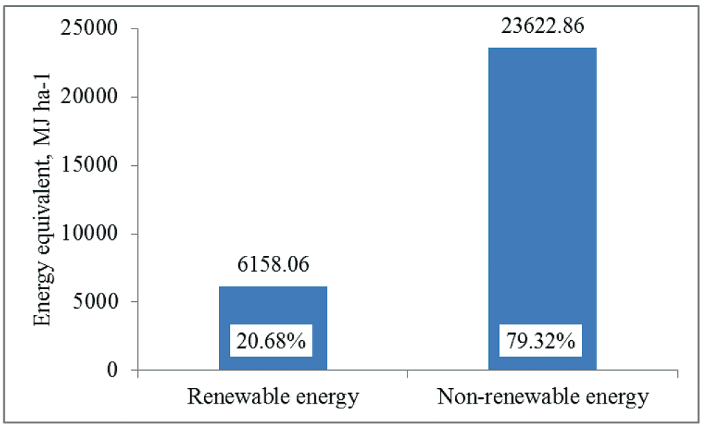


Figure 2. Distribution of renewable and non-renewable energy inputs in the production of pineapples

Generally, the findings of this study are very useful to enrich the database of energy consumption in the production of pineapples, particularly in the production of pineapples in peat soil areas, in which about 90% of the total pineapples planted areas in Malaysia are on peat soil (Chan, 2000). Similar findings published by Duke (1983) from past research conducted in Hawaiian pineapple plantations are limited to pineapple cultivation on the soil in Hawaii only, which is characterised mostly by silt loams, silty clay loams and silty clay (Hepton, 2003).

CONCLUSION

The distribution of energy consumption in pineapple production in Malaysia has successfully been audited and analysed through a case study carried out at a pineapple plantation under KOSAS Bhd. located at Kg. Kundang in Selangor, Malaysia. The calculated energy used for the pineapple production in KOSAS Bhd. was 29780.91 MJ ha^{-1}. A very large portion of this energy (54.61%) was provided for fertilising operation, where compound NPK fertiliser, urea and other micronutrient fertilisers supplied about 45.65% of the total production energy. The

ratio of energy output/input was 3.56, whereby the input is highly and effectively used at this pineapple plantation. It is essential to use the resources effectively for a sustainable agricultural production. Basically, a sustainable agriculture production requires sustainable supply of energy resources. Even though labour input has a small portion in energy consumption at the pineapple plantation, the use of appropriate machinery in cultivation practices in Malaysia is recommended as an alternative to reduce human drudgery and to optimise the capability of plantation to do work efficiently and also to offset the labour shortage problem.

REFERENCES

Anon. (2004). *Units, Equivalents and energy Constants. Cooperative Extension Service.* Institute of Food and Agricultural Sciences, University of Florida.

Alcock, R. (1986). *Tractor-Implement Systems*. Westport, Connecticut: AVI Publishing Co.

ASABE. (2009, June 6). *ASABE Standards: Agricultural Machinery Management Data*. ASAE D497. St. Joseph, Mich.: ASABE. Retrieved December 29, 2012, from http://asae.frymulti.com/standards.asp

Chan, Y. K. (2000). Status of pineapple industry and research and develpment in Malaysia. In *III International Pineapple Symposium 529* (pp. 77-84). Pattaya, Thailand.

Duke, J. A. (1983). *Handbook of Energy Crops*. Retrieved November 22, 2012, from http://www.hort.purdue.edu/newcrop/ duke_energy.

Esengun, K., Gunduz, O., & Erdal, G. (2007). Input-output energy analysis in dry apricot production of Turkey. *Energy Conversion and Management, 48*(2), 592-598.

Eurostat. (2016). *Energy from renewable sources*. Retrieved May 10, 2016, from http://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_from_renewable_sources

Gevao, S. M. B., Ishak, W. I. W., Azmi, Y., & Chan, C. W. (2005). Analysis of energy consumption in lowland rice based cropping system of Malaysia. Songklanakarin. *Journal of Science and Technology, 27*(4), 819-826.

Gezer, I., Acaroglu, M., & Haciseferogullari, H. (2003). Use of energy and labour in apricot agriculture in Turkey. *Biomass Bioenergy, 24*(3), 215–219.

Hepton, A. (2003). Culture system. In D. P. Bartholomew, R. E. Paul, & K. G. Rohrbach (Eds), *The Pineapple: Botany, Production and Uses*. Wallingford, UK: CABI Publishing.

International Energy Agency (2016). *Energy efficiency*. Retrieved May 10, 2016, from http://www.iea.org/topics/energyefficiency/.

Kızılaslan, N. (2009). Energy use and input-output energy analysis for apple production in Turkey. *Journal of Food, Agriculture & Environment, 7*(2), 419-423.

MIPB. (2012). *Pineapple info*. Malaysian Pineapple Industry Board. Retrieved August 20, 2012, from www.mpib.gov.my

Mandal, K. G., Saha, K. P., Gosh, P. L., Hati, K. M., & Bandyopadhyay, K. K. (2002). Bioenergy and economic analyses of soybean-based crop production systems in central India. *Biomass Bioenergy, 23*(5), 337–345.

Moerschner, J., & Gerowitt, B. (2000). Direct and indirect energy use in arable farming - An example on winter wheat in Northern Germany. In B. P. Weidema & M. J. G Meeusen (Eds.), *Agricultural Data for Life Cycle Assessments* (p. 195). The Hague, Agricultural Economics Research Institute (LEI).

Ozkan, B., Akcaoz, H., & Karadeniz, F. (2004a). Energy requirement and economic analysis of citrus production in Turkey. *Energy Conversion and Management, 45*(11), 1821-1830.

Pimentel, D., & Patzek, T. W. (2005). Ethanol production using corn, switchgrass and wood; biodiesel production using soybean and sunflower. *Natural Resources Research, 14*(1), 65-76.

Refsgaard, K., Halberg, N., & Kristensen, E. S. (1998). Energy utilization in crop and dairy production in organic and conventional livestock production systems. *Agricultural Systems, 57*(4), 599–630.

Ricaud, R. (1980). Energy input and output for sugarcane in Louisiana. In D. Pimentel (Ed.), *Handbook of Energy Utilization in Agriculture* (pp. 135-136). Boca Raton, FL.: CRC Press.

Safa, M., & Tabatabaeefar, A. (2002). Energy consumption in wheat production in irrigated and Dry Land Farming. *World Journal of Agricultural Sciences, 4*(1), 86-90.

Singh, M. K., Pal, S. K., Thakur, R., & Verma, U. N. (1997). Energy input–output relationship of cropping systems. *Indian Journal of Agricultural Sciences, 67*(6), 262-264.

Pertanika J. Sci. & Technol. 25 (1): 29 - 42 (2017)

SCIENCE & TECHNOLOGY

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

A Combined Analytical Method for Optimal Location and Sizing of Distributed Generation Considering Voltage Stability and Power Loss in a Power Distribution System

Maryam Mirzaei[1]*, Jasronita Jasni[2], Hashim Hizam[3], Noor Izzri Abdol Wahab[3] and Ehsan Moazami[1]

[1]Department of Electrical Power Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[2]Centre of Excellence on Lightning Protection, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[3]Centre for Advanced Power and Energy Research, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

In this paper, a multi-objective analytical method to evaluate the impacts of optimal location and sizing of distributed generation is presented. This method is based on an analysis of the exact loss formula and continuous power flow in a radial distribution system. Based on two methods of analysis, power loss and weakest voltage buses and lines are calculated and then the optimal size of distributed generation is determined. After that, by considering the minimum power losses and the maximisation of voltage stability, the proposed index determines and ranks positions to decide the optimal distributed generation location in the system. This method allows us to find the best places and size to connect a number of distributed generation units by optimising the objective functions. The simulation results were obtained using a 33-bus radial distribution system to determine the location and size of the distributed generation units. The results show the effectiveness of voltage profile improvement, loading factor improvement and power loss reduction. Further, the problems of a single objective function and the placement of the distributed generation unit using analytical methods are solved by the proposed approach.

Keywords: Distributed generation, continuous power flow, voltage stability, exact loss formula, optimum size, optimum location

Article history:
Received: 30 September 2016
Accepted: 22 April 2016

E-mail addresses:
m.mirzaei.eh@gmail.com (Maryam Mirzaei),
jas@upm.edu.my (Jasronita Jasni),
hhizam@upm.edu.my (Hashim Hizam),
izzri@upm.edu.my (Noor Izzri Abdol Wahab),
e.moazzami@gmail.com (Ehsan Moazami)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Due to the recent widespread use of distributed generation, the power industry has experienced significant changes in the distribution power system. In general,

distributed generation (DG) can be described as electric power generation that is integrated within distribution networks by utilising a number of smaller generating units, especially at places which are close to the point of consumption (Ugranli & Karatepe, 2012). DG can be an alternative for industrial, residential and commercial applications. DG units are smaller scale based on locally available resources and therefore, they are mostly connected at the distribution level. The penetration of DG is different. When it is high, the generated power of DG units not only changes the power flow in the distribution system, but also in the transmission system. Hernandez, Velasco, and Trujillo (2011) presented some advantages of DG as follows: "(a) reduces power losses; (b) improves the voltage profile and reliability of the system; (c) Reduces greenhouse gas emission; (d) More flexible energy solution due to the small size; and (e) Mitigates environmental concerns."

The connection of DG to the network may influence the stability of the power system, i.e., angle, frequency and voltage stability (Donnelly, Dagle, Trudnowski, & Rogers, 1996; Reza, Slootweg, Schavemaker, Kling, & Van der Sluis, 2003). If DG is connected to the most suitable location in the power system, DG integrated power systems have many advantages in comparison to the classical power systems (Ugranli & Karatepe, 2012). The problem of finding the optimal placing and sizing of DG units is a high priority issue, which if installed in non-optimal places, may lead to increasing power loss, reducing reliability level and growing cost in the system. Therefore, it is important to allocate and determine the optimal size and placing of DG to maximise the system efficiency. Various approaches have been proposed by several researchers to find the optimal and fast methods for placing and sizing of the DG units on the basis of improving the voltage profile, reducing power system losses, maximising system loadability and maximising bus and line stability. Generally, these methods can be classified into three broad categories, which are conventional optimisation methods, meta-heuristic based optimisation methods and a hybrid method. There are different conventional approaches to find the optimal size and locations of DG units in power distribution systems including the analytical approaches (Borges & Falcao, 2003; Acharya, Mahat, & Mithulananthan, 2006; Tuba Gozel & Hocaoglu, 2009; Hung, Mithulananthan, & Bansal, 2010; Hosseini & Kazemzadeh, 2011; Aman, Jasmon, Mokhlis, & Bakar, 2012; Hung, Mithulananthan, & Bansal, 2013), Continuation Power Flow based iterative (Hedayati, Nabaviniaki, & Akbarimajd, 2008; Ettehadi, Vaez-Zadeh, & Ghasemi, 2012; Ettehadi, Ghasemi, & Vaez-Zadeh, 2013), the grid search method (T Gozel, Eminoglu, & Hocaoglu, 2008), reactive power optimisation (Khalesi, Rezaei, & Haghifam, 2011), Optimal Power Flow (Dent, Ochoa, Harrison, & Bialek, 2010), Mixed Integer Nonlinear Programming (Ochoa & Harrison, 2011), Primal-Dual Interior Point Method (Rueda-Medina, Franco, Rider, Padilha-Feltrin, & Romero, 2013), and Deterministic optimisation techniques (AlHajri, AlRashidi, & El-Hawary, 2010). On the other hand, the meta-heuristic methods have been widely used by researchers in recent years as these approaches are intuitive, easy to understand, simple to implement and address the integer variable very well as compared to the conventional method and analytical programming. However, the results produced are not guaranteed to be optimal and the process speed can be slower.

Some examples of the meta-heuristic optimisation class are genetic algorithms (GA) (Celli, Ghiani, Mocci, & Pilo, 2005; Abou El-Ela, Allam, & Shatla, 2010). The Evolutionary Programming (Rahman, Rahim, & Musirin, 2004), tabu search (Katsigiannis & Georgilakis,

2008), simulated annealing (Injeti & Prema Kumar, 2013), particle swarm optimisation (Lalitha, Reddy, & Usha, 2010; El-Zonkoly, 2011; Kayal & Chanda, 2013), ant colony optimisation (Falaghi & Haghifam, 2007), Cuckoo Search (Moravej & Akhlaghi, 2013), bacterial foraging optimisation (S. Devi & Geethanjali, 2014), which have all been applied in most optimisation problems, as well as DG optimal placement and sizing problems. Tabu Search is an efficient combinatorial method which can be used to achieve an optimal or a sub optimal solution within a reasonably short duration. Meanwhile, researchers have not paid much attention on other methods of Evolutionary Programming, Ant Colony Algorithm, Simulated Annealing and Cuckoo Search.

GA is competent at obtaining a solution near the global minima which is computationally intensive. The hybrid method is also a combination of the above approaches, which include the Genetic-Fuzzy (Akorede, Hizam, Aris, & Ab Kadir, 2011), Genetic-Particle Swarm (Moradi & Abedini, 2012), Genetic-Optimal Power Flow (Naderi, Seifi, & Sepasian, 2012), Particle Swarm Optimisation-Optimal Power Flow (Gomez-Gonzalez, Lopez, & Jurado, 2012), and Analytical-Fuzzy (Devi & Subramanyam, 2007) techniques. The performance of the hybrid algorithm is fast and efficient because of the improved versions. The analytical techniques are easily implemented with high efficiency in terms of computational time and best precision factor while OPF, CPF and meta-heuristic approaches are the next advanced level. The problems related to the difficulties of DG placement and sizing issue by use of the Hybrid method remained unsolved problem despite many improvements. This paper presents a combined analytical method to evaluate the impacts of optimal location and sizing of DG. This method is based on the combination of the exact loss formula and continuous power flow (CPF) in a radial distribution system. This method allows finding the best places and size to connect a number of DG units by optimising the objective functions. The simulation results are obtained on a 33-bus radial distribution system in the Matlab environment to determine the location and size of DG units.

DISTRIBUTED GENERATION

DG can have a significant impact on the power-flow, voltage profile, stability, continuity, and quality of power supply (Van Thong, Driesen, & Belmans, 2005). Other aspects of DG are the DG rating, purpose, power delivery area, technology, environmental impact, mode of operation and penetration DG (Ackermann, Andersson, & Soder, 2001). Due to some complex issues of DG, compared to the traditional network, the consideration of the benefits of DG is important. These benefits include power loss reduction, voltage profile improvement, reliability and security improvement, decreasing congestion in feeders, power quality improvement and stability enhancement (Ackermann et al., 2001; Chiradeja & Ramakumar, 2004; Rao, Ravindra, Satish, & Narasimham, 2013).

DG impact on power losses

The design of a distribution system is basically based on the power flows from the source substation (sending end) to the load (the consumer side) and over this direction of flow, the size

of the conductors is gradually decreased. Moreover, if a DG is installed with a high capacity to export power beyond the local substation, the losses will be very high. Therefore, determining the size of DG is important when considering the distribution system size in terms of load (MW) (Wang & Nehrir, 2004). According to the above explanation, the relation of the higher DG capacity with the consequent higher losses can be defined. Thus, a higher capacity of the DGs will increase the power losses. Otherwise, the system restrictions must be considered. Figure 1 shows this relationship for a 10 bus radial distribution system (Hosseini & Kazemzadeh, 2011). Based on the figure, the effect of the DG size on system losses is minimal at first, and then the power loss is very sensitive to the size of DG at the lateral buses. Obviously, increasing the DG penetration leads to the increase in the power losses. Therefore, it can be concluded that the correct estimation of the size of DG is an important rule to decrease the power loss in the system. Moreover, to determine the optimal size of the DG from the loss curve, the minimum value of power loss is considered for each bus.

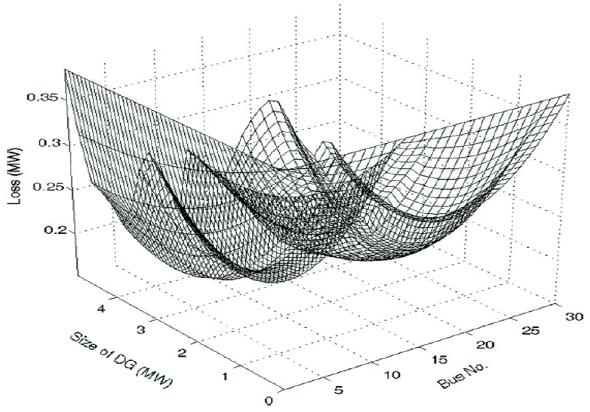


Figure 1. The effects of size and location of DG on system loss.

DG impact on voltage stability

The installation of DG has a positive impact on improving the stability of the voltage in the power system. In order to analyse the static voltage stability, a PV curve is used, which is achieved using the CPF approach (Canizares & Alvarado, 1993). Based on the PV curve in Figure 2, to evaluate the performance of the system from the current loading level or operating point (λ) to the voltage collapse point or the critical point (CP), the voltage stability margin (VSM) is defined. Clearly, the reactive power loss is decreased by increasing the active power injected by a DG unit. Therefore, by raising the voltage to the operating point from (V1) to (V2), then CP1 is increased to CP2. Therefore, the maximum loadability (λmax1) is a growth to (λmax2) for which the relationship is:

$$P_i = (\lambda+1)\, P_{oi}\,, \qquad Q_i = (\lambda+1)\, Q_{oi} \tag{1}$$

where Poi and Qoi are the base case load active and reactive power demands, respectively.

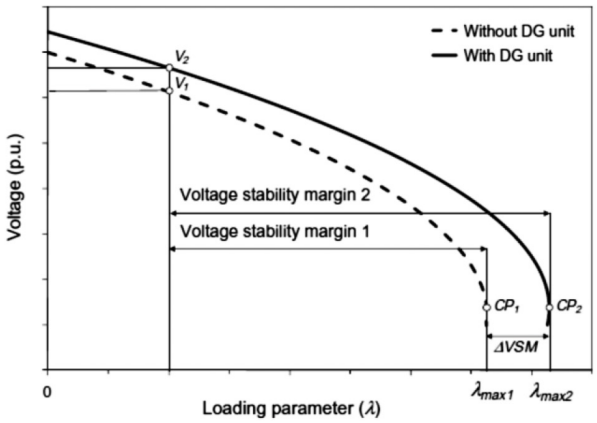


Figure 2. DG impacts on maximum loadability and voltage stability margin.

PROPOSED METHODOLOGY

The total real power loss (P_L) with N buses is calculated by "the exact loss formula" (Elgerd, 1983) in a distribution system.

$$P_L = \sum_{i=1}^{N}\sum_{j=1}^{N}[\alpha_{ij}(P_iP_j + Q_iQ_j) + \beta_{ij}(Q_iP_j - P_iQ_j)] \quad (2)$$

where $\alpha_{ij} = \frac{r_{ij}}{V_iV_j}\cos(\delta_i - \delta_j)$ and $\beta_{ij} = \frac{r_{ij}}{V_iV_j}\sin(\delta_i - \delta_j)$, N is the bus number and P_i, Q_i, P_j and Q_j are the active and reactive power injections at buses i and j, respectively. $R_{ij} + j\,X_{ij} = Z_{ij}$ is the *ij*th element of impedance matrix [Zbus] = $[Ybus]^{-1}$.

The sensitivity factor of real power loss with respect to real power injection from DG is given as follows:

$$\alpha_i = \frac{\partial P_L}{\partial P_i} = \sum_{i=1}^{N}(\alpha_{ij}P_j - \beta_{ij}Q_j) \quad (3)$$

If the partial derivative of Equation (3) becomes zero, the total active power loss will then be minimised.

The active power injection of the bus i, where the DG unit is installed, is given by:

$$P_i = P_{DGi} - P_{Di} \quad (4)$$

where, P_{Di} is the load demand at node i and P_{DGi} is the injection power from DG placed to the node i.

On the other hand, in CPF, by applying the loading parameter of (λ), loads and generations are increased from their base case values, as follows:

$$P_D = P_D^{oi} + \lambda.P_D^{CPF} \quad (5)$$

where P_D^{CPF} is power increment directions of loads in the CPF and P_D^{oi} is active powers of loads at the base case (Milano, 2005). Substituting Equation (4) and Equation (5) into Equation (2), the total active power losses with a DG unit can be obtained as follows:

$$P_{DGi} = (P_{Di}^{oi} + \lambda_i . P_{Di}^{CPF}) + \frac{1}{\alpha_{ii}} \left[\beta_{ii} Q_i - \sum_{j=1, j\neq i}^{N} (\alpha_{ij} P_j - \beta_{ij} Q_j)\right] \quad (6)$$

The optimum size of the DG at bus i is calculated by P_{DGi}.

PROBLEM FORMULATION

The objective functions (OF) are achieved by finding the optimal size and location of the DG in a power system. To achieve this goal, minimise point of power losses from the Equation (2) and the most sensitive bus from running the CPF is considered to sit the DG as the first candidate. The optimal size of DG is measured from Equation (6); then, to select a place for DG, the Find minimum of single-variable function is used as:

$$\min_x f(x) \quad that\ x_1 < x < x_2 \quad (7)$$

After sitting, the DG at the initial candidate place and the loop starts from the first step again. The OF to locate the DG in optimal place is firstly calculated by minimising the power losses index based on the ratio of the total active power losses, with and without the DG unit, as:

$$\text{OPL} = \frac{\sum_{i=1}^{N} PL_{base\ case} - \sum_{i=1}^{N} PL_{with\ DG}}{\sum_{i=1}^{N} PL_{base\ case}} \quad (8)$$

After that, the maximum of the loading factor improvement, with and without the DG unit, is defined as a second index by:

$$\text{OVS} = \frac{\sum_{i=1}^{N} \lambda\ max_{withDG} - \sum_{i=1}^{N} \lambda\ max_{base\ case}}{\sum_{i=1}^{N} \lambda\ max_{base\ case}} \quad (9)$$

Consequently, the following OF is defined as a combination of both indices to optimise by:

$$\text{Minimise OF } (P_{DG}) = a \times \min\{\text{OPL}\} + b \times \frac{1}{\max\{\text{OVS}\}} \quad (10)$$

Where, a and b are weighting factors that are considered to reduce loss and improve the loading factor. They are determined based on which OF has a higher value. Here, the value for each one is considered as 0.5. The algorithm for locating and sizing the DG, based on the proposed method, is illustrated in Figure 3.

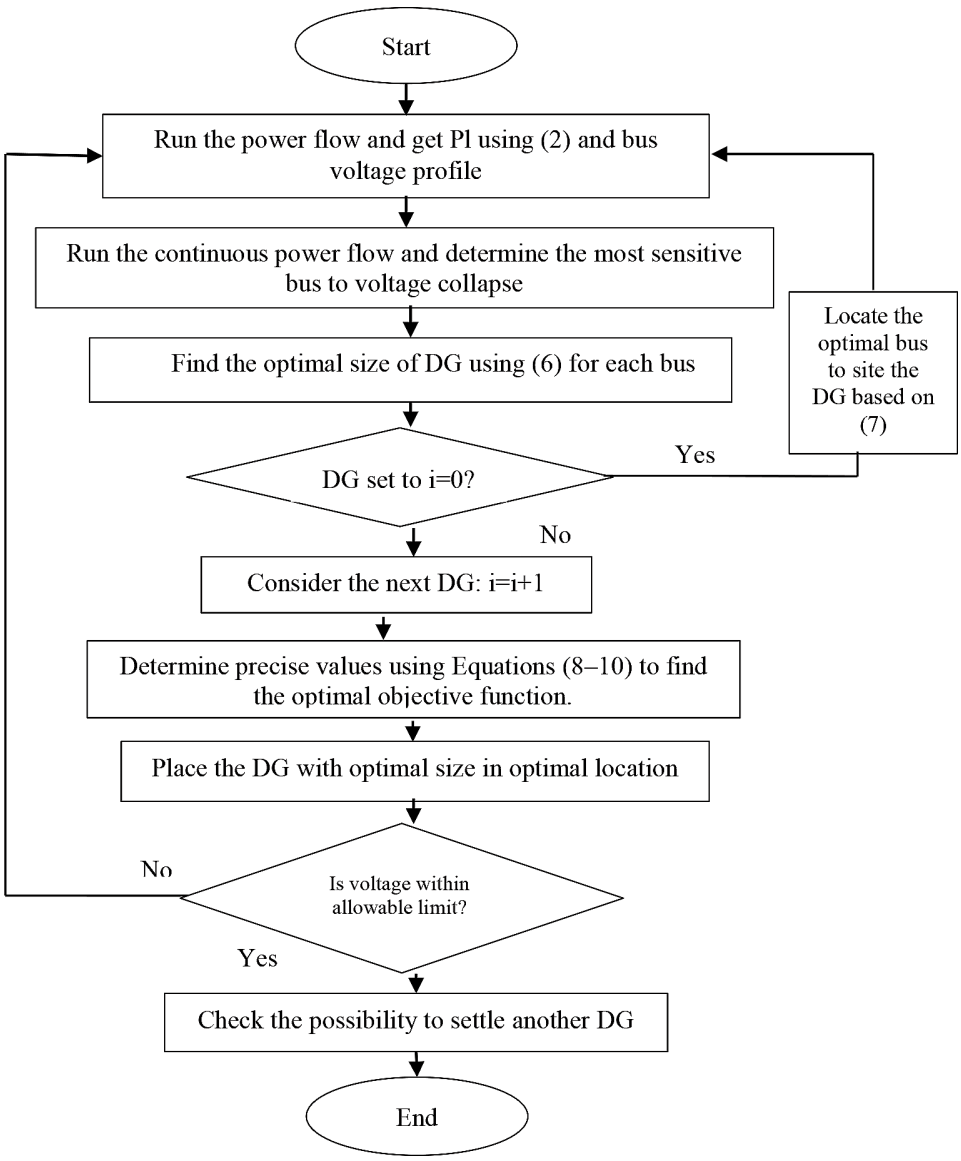


Figure 3. Flowchart of the proposed method.

NUMERICAL RESULTS AND DISCUSSION

The proposed algorithm was tested on a 33-bus radial distribution network, which has 33 buses and 32 branches. The radial system has a total load of 3.72 MW and 2.3 MVAR. The voltage level should be in the allowable voltage level ($0.95 < Vbus < 1.05$). The system is illustrated in Figure 4. After simulating the test system in the Matlab environment, and writing the necessary computer programme, the optimum DG locations and sizes have been calculated. The total power loss and optimal size of DG, based on the proposed

formula for each bus, are illustrated in Figure 5 and Figure 6, respectively. Meanwhile, the voltage profile at the base case and the sensitive bus to the voltage collapse is shown in Figure 7.

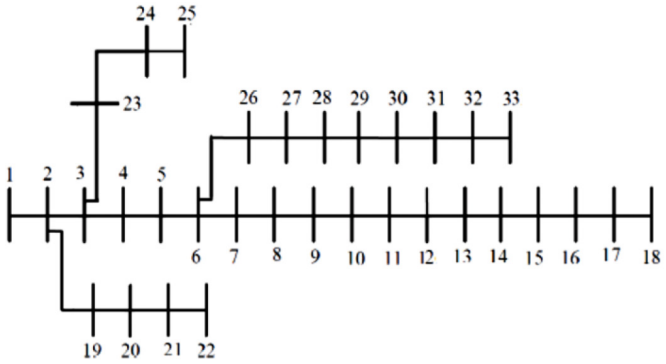


Figure 4. Single line diagram of a 33-bus radial distribution network

As illustrated in Figure 7, the most sensitive bus to voltage collapse is bus number 18, which also has the lowest voltage profile. After calculating the size of DG and sitting in the initial place, the optimal sizing and location of DG are evaluated, as shown in following figure and tables. Figure 8 shows the voltage profile, with and without DG. The voltage level is achieved by installing the DG completely lands in the allowable voltage level, thus presenting the effectiveness of the proposed method on the performance of the system when comparing cases with and without the DG unit condition (base case). The optimum DG sizes and places are shown in Table 1. This table indicates a comparative analysis for the cases with and without DG unit.

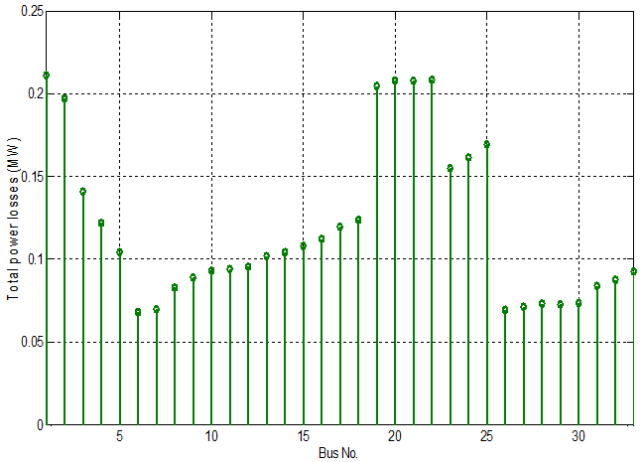


Figure 5. Total power loss of a 33 bus distribution system

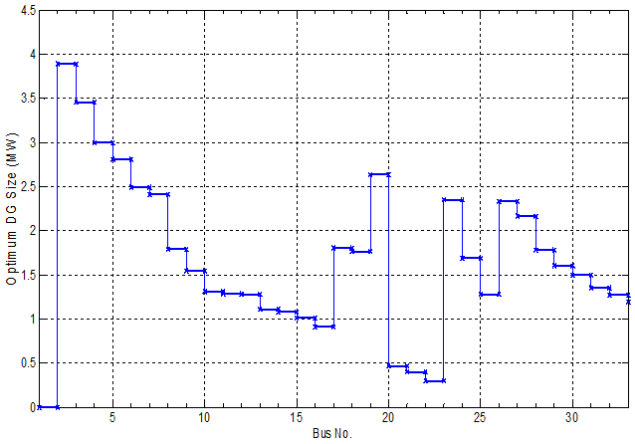


Figure 6. Optimal size of DG of a 33 bus distribution system

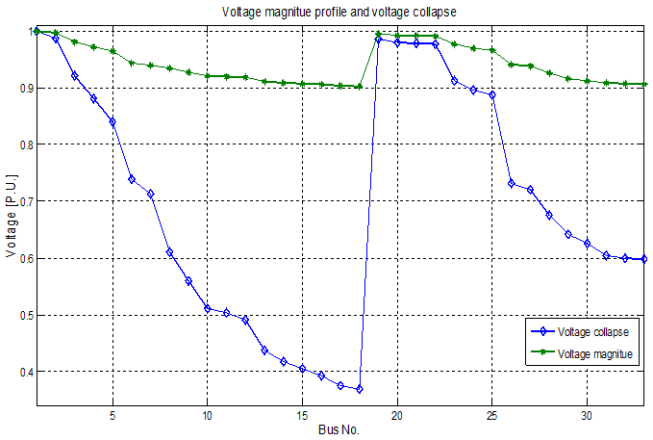


Figure 7. Voltage magnitude profile and voltage collapse

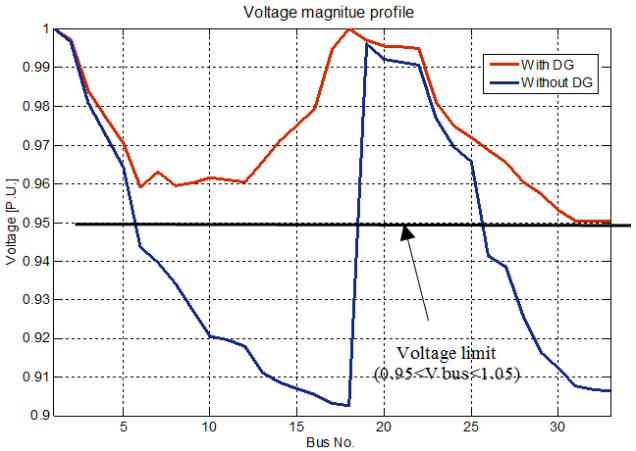


Figure 8. Voltage profile curve with and without DG unit

Based on Figure 3, after sizing DG using (6), the load flow function should be executed two times (one time with the system without DG unit and another with the system with DG unit) to achieve the final solution. Therefore, firstly, DG will be placed in bus 18, which is the most sensitive bus to voltage collapse. The result of the initial step showed voltage profile growth from 0.9089 to 0.9474 and power loss is decreased about 34.6% (see Table 1). Moreover, the loading factor increased after installing the DG unit. In the next step, the second load flow is run to obtain the exact value of results. Obviously, it can be seen that by placing DG, the minimum voltage profile and loading factor parameter were again increased to 4.58% and 11.39% respectively. Moreover, the power loss was dropped to 0.1097. Therefore, it is possible to conclude that by locating the DG in the optimal place with optimal size. The stability of the system was improved and this has a significant effect on improving the loading parameter. To evaluate the performance of the proposed method, it is compared with other three methods (Gozel & Hocaoglu, 2009; Moradi & Abedini, 2012; Ettehadi & Ghasemi, 2013) as shown in Table 2. As shown in Table 2, results of the proposed method show that by siting the one DG unit in the system, the minimum voltage magnitude is increased to the acceptable voltage limit compared to the techniques of Gozel and Ettehadi. In Moradi's method, however, the minimum voltage magnitude reached 9808. Nonetheless, he used three DG units.

Moreover, loading factor difference with and without DG in proposed method is higher compared with the approaches of Gozel and Ettehadi, while Moradi did not consider it at all. Also, the power loss value of the proposed method is lower than the two other methods with one DG unit. Besides, in Moradi's method, by installing three DG units in the system, the power loss just reduced %5.8 more in comparison to the proposed technique. Therefore, the above comparison shows the efficiency of the proposed approach.

Table 1
Effects of optimising DG location and sizing on the loading factors

	DG bus	V max (p.u.)	V min (p.u.) (×10^-4)	System loading (×10^-3)	Power loss (MW) (×10^-4)
Without DG	---	1	9089	3073	2127
With DG at initial step	18	1	9474	3397	1198
With DG based on OF	18	1	9505	3423	1097

Table 2
Comparison of the effects of DG Placement on the system performance

	V max (p.u.)	V min (p.u.) (×10^-4)	Number of DG	Difference of loading (×10^-3)	Power loss (MW) (×10^-4)
Proposed method	1	9505	1	350	1097
Gozel & Hocaoglu (2009)	1	9410	1	300	1112
Ettehadi & Ghasemi (2013)	1	9370	1	240	1420
Moradi & Abedini (2012)	1	9808	3	---	1034

CONCLUSION

This paper proposed a combined analytical method based on the exact loss formula and continuous power flow to find the optimal size and location of the DG unit so as to reduce power loss and increase the stability of voltage. The proposed method presented a new formula to calculate the optimal size of DG unit. Then, the multi-objective index is presented based on the combination of the power losses and loading factor improvement indices by assigning a weight factor for each index. Moreover, the single-variable function is used to find the minimum value of the objective function. The proposed technique is tested on a typical distribution test system and its results are discussed and compared with three previous research. The results show that by finding the optimal size and place of DG unit, the total power losses is reduced %48.43, while voltage profile is enhanced %4.38 and voltage stability is increased %10.23. Consequently, it can be concluded that by installing a DG unit with accurate size and location, the system condition can be improved significantly.

REFERENCES

Abou El-Ela, A., Allam, S., & Shatla, M. (2010). Maximal optimal benefits of distributed generation using genetic algorithms. *Electric Power Systems Research, 80*(7), 869-877.

Acharya, N., Mahat, P., & Mithulananthan, N. (2006). An analytical approach for DG allocation in primary distribution network. *International Journal of Electrical Power & Energy Systems, 28*(10), 669-678.

Ackermann, T., Andersson, G., & Soder, L. (2001). Distributed generation: a definition. *Electric Power Systems Research, 57*(3), 195-204.

Akorede, M., Hizam, H., Aris, I., & Kadir, M. A. (2011). Effective method for optimal allocation of distributed generation units in meshed electric power systems. *Generation, Transmission & Distribution, IET, 5*(2), 276-287.

AlHajri, M. F., AlRashidi, M. R., & El-Hawary, M. E. (2010). *Improved Sequential Quadratic Programming approach for optimal Distribution Generation sizing in distribution networks.* Paper presented at the Electrical and Computer Engineering (CCECE), 23rd Canadian Conference.

Aman, M., Jasmon, G., Mokhlis, H., & Bakar, A. (2012). Optimal placement and sizing of a DG based on a new power stability index and line losses. *International Journal of Electrical Power & Energy Systems, 43*(1), 1296-1304.

Borges, C. L., & Falcao, D. M. (2003). *Impact of distributed generation allocation and sizing on reliability, losses and voltage profile.* Paper presented at the Power Tech Conference Proceedings, 2003 IEEE Bologna.

Canizares, C. A., & Alvarado, F. L. (1993). Point of collapse and continuation methods for large AC/DC systems. *Power Systems, IEEE Transactions on, 8*(1), 1-8.

Celli, G., Ghiani, E., Mocci, S., & Pilo, F. (2005). A multiobjective evolutionary algorithm for the sizing and siting of distributed generation. *Power Systems, IEEE Transactions on, 20*(2), 750-757.

Chiradeja, P., & Ramakumar, R. (2004). An approach to quantify the technical benefits of distributed generation. *Energy Conversion, IEEE Transactions on, 19*(4), 764-773.

Dent, C. J., Ochoa, L. F., Harrison, G. P., & Bialek, J. W. (2010). Efficient secure AC OPF for network generation capacity assessment. *Power Systems, IEEE Transactions on, 25*(1), 575-583.

Devi, A. L., & Subramanyam, B. (2007). Optimal DG unit placement for loss reduction in radial distribution system-a case study. *Journal of Engineering and Applied Sciences, 2*(6), 57-61.

Devi, S., & Geethanjali, M. (2014). Application of Modified Bacterial Foraging Optimization algorithm for optimal placement and sizing of Distributed Generation. *Expert Systems with Applications, 41*(6), 2772–2781.

Donnelly, M., Dagle, J., Trudnowski, D., & Rogers, G. (1996). Impacts of the distributed utility on transmission system stability. *Power Systems, IEEE Transactions on, 11*(2), 741-746.

El-Zonkoly, A. (2011). Optimal placement of multi-distributed generation units including different load models using particle swarm optimization. *Swarm and Evolutionary Computation, 1*(1), 50-59.

Elgerd, O. I. (1983). *Electric energy systems theory: an introduction*: Tata McGraw-Hill Education.

Ettehadi, M., Ghasemi, H., & Vaez-Zadeh, S. (2013). Voltage Stability-Based DG Placement in Distribution Networks. *IEEE* Transactions on Power Delivery*, 28*(1), 171-178.

Ettehadi, M., Vaez-Zadeh, S., & Ghasemi, H. (2012). *DG placement in distribution networks considering voltage stability.* Paper presented at the Environment and Electrical Engineering (EEEIC), 11th International Conference.

Falaghi, H., & Haghifam, M. R. (2007). *ACO based algorithm for distributed generation sources allocation and sizing in distribution systems.* Paper presented at the Power Tech, 2007 IEEE Lausanne.

Gonzalez, M. G., Lopez, A., & Jurado, F. (2012). Optimization of distributed generation systems using a new discrete PSO and OPF. *Electric Power Systems Research, 84*(1), 174-180.

Gozel, T., Eminoglu, U., & Hocaoglu, M. (2008). A tool for voltage stability and optimization (VS&OP) in radial distribution systems using Matlab graphical user interface (GUI). *Simulation Modelling Practice and Theory, 16*(5), 505-518.

Gozel, T., & Hocaoglu, M. H. (2009). An analytical method for the sizing and siting of distributed generators in radial systems. *Electric Power Systems Research, 79*(6), 912-918.

Hedayati, H., Nabaviniaki, S., & Akbarimajd, A. (2008). A method for placement of DG units in distribution networks. *Power Delivery, IEEE Transactions on, 23*(3), 1620-1628.

Hernandez, J., Velasco, D., & Trujillo, C. (2011). Analysis of the effect of the implementation of photovoltaic systems like option of distributed generation in Colombia. *Renewable and Sustainable Energy Reviews, 15*(5), 2290-2298.

Hosseini, R. K., & Kazemzadeh, R. (2011). *Optimal DG allocation by extending an analytical method to minimize losses in radial distribution systems.* Paper presented at the Electrical Engineering (ICEE), 19th Iranian Conference.

Hung, D. Q., Mithulananthan, N., & Bansal, R. (2010). Analytical expressions for DG allocation in primary distribution networks. *Energy Conversion, IEEE Transactions on, 25*(3), 814-820.

Hung, D. Q., Mithulananthan, N., & Bansal, R. (2013). Analytical strategies for renewable distributed generation integration considering energy loss minimization. *Applied Energy, 105*, 75-85.

Injeti, S. K., & Kumar, N. P. (2013). A novel approach to identify optimal access point and capacity of multiple DGs in a small, medium and large scale radial distribution systems. *International Journal of Electrical Power & Energy Systems, 45*(1), 142-151.

Katsigiannis, Y., & Georgilakis, P. (2008). Optimal sizing of small isolated hybrid power systems using tabu search. *Journal of Optoelectronics and Advanced Materials, 10*(5), 1241- 1245.

Kayal, P., & Chanda, C. (2013). Placement of wind and solar based DGs in distribution system for power loss minimization and voltage stability improvement. *International Journal of Electrical Power & Energy Systems, 53*, 795-809.

Khalesi, N., Rezaei, N., & Haghifam, M. R. (2011). DG allocation with application of dynamic programming for loss reduction and reliability improvement. *International Journal of Electrical Power & Energy Systems, 33*(2), 288-295.

Lalitha, M. P., Reddy, V. V., & Usha, V. (2010). Optimal DG placement for minimum real power loss in radial distribution systems using PSO. *Journal of Theoretical and Applied Information Technology, 13*(2), 107-116.

Medina, A. C. R., Franco, J. F., Rider, M. J., Feltrin, A. P., & Romero, R. (2013). A mixed-integer linear programming approach for optimal type, size and allocation of distributed generation in radial distribution systems. *Electric Power Systems Research, 97*(0), 133-143.

Milano, F. (2005). An Open Source Power System Analysis Toolbox. *Power Systems, IEEE Transactions on, 20*(3), 1199-1206.

Moradi, M., & Abedini, M. (2012). A combination of genetic algorithm and particle swarm optimization for optimal DG location and sizing in distribution systems. *International Journal of Electrical Power & Energy Systems, 34*(1), 66-74.

Moravej, Z., & Akhlaghi, A. (2013). A novel approach based on cuckoo search for DG allocation in distribution network. *International Journal of Electrical Power & Energy Systems, 44*(1), 672-679.

Naderi, E., Seifi, H., & Sepasian, M. S. (2012). A dynamic approach for distribution system planning considering distributed generation. *Power Delivery, IEEE Transactions on, 27*(3), 1313-1322.

Ochoa, L. F., & Harrison, G. P. (2011). Minimizing energy losses: Optimal accommodation and smart operation of renewable distributed generation. *Power Systems, IEEE Transactions on, 26*(1), 198-205.

Rahman, T., Rahim, S., & Musirin, I. (2004). *Optimal allocation and sizing of embedded generators.* Paper presented at the Power and Energy Conference, PECon. Proceedings. National.

Rao, R., Ravindra, K., Satish, K., & Narasimham, S. (2013). Power loss minimization in distribution system using network reconfiguration in the presence of distributed generation. *IEEE transactions on power systems, 28*(1), 317-325.

Reza, M., Slootweg, J., Schavemaker, P., Kling, W., & Van der Sluis, L. (2003). *Investigating impacts of distributed generation on transmission system stability.* Paper presented at the Power Tech Conference Proceedings, 2003 IEEE Bologna.

Thong, V. V., Driesen, J., & Belmans, R. (2005). Power quality and voltage stability of distribution system with distributed energy resources. *International Journal of Distributed Energy Resources, 1*(3), 227-240.

Ugranli, F., & Karatepe, E. (2012). Convergence of rule-of-thumb sizing and allocating rules of distributed generation in meshed power networks. *Renewable and Sustainable Energy Reviews, 16*(1), 582-590.

Wang, C., & Nehrir, M. H. (2004). Analytical approaches for optimal placement of distributed generation sources in power systems. *Power Systems, IEEE Transactions on, 19*(4), 2068-2076.

Pertanika J. Sci. & Technol. 25 (1): 43 - 56 (2017)

SCIENCE & TECHNOLOGY

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

A Novel Adaptive Neuro Fuzzy Inference System Based Classification Model for Heart Disease Prediction

Abdu Masanawa Sagir* and Saratha Sathasivam

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

ABSTRACT

Adaptive Neuro Fuzzy Inference System (ANFIS) is among the most efficient classification and prediction modelling techniques used to develop accurate relationship between input and output parameters in different processes. This paper reports the design and evaluation of the classification performances of two discrete Adaptive Neuro Fuzzy Inference System models, ANFIS Matlab's built-in model (ANFIS_LSGD) and a newly ANFIS model with Levenberg-Marquardt algorithm (ANFIS_LSLM). Major steps were performed, which included classification using grid partitioning method, the ANFIS trained with least square estimates and backpropagation gradient descent method, as well as the ANFIS trained with Levenberg-Marquardt algorithm using finite difference technique for computation of a Jacobian matrix. The proposed ANFIS_LSLM model predicts the degree of patient's heart disease with better, reliable and more accurate results. This is due to its new feature of index membership function that determines the unique membership functions in an ANFIS structure, which indexes them into a row-wise vector. In addition, an attempt was also done to specify the effectiveness of the model's performance measuring accuracy, sensitivity and specificity. A comparison of the two models in terms of training and testing with the Statlog-Cleveland Heart Disease dataset have also been done.

Keywords: Adaptive neuro fuzzy inference system, Classification, Grid Partitioning Method, Levenberg-Marquardt algorithm, Prediction

INTRODUCTION

Heart attack disease remains the main cause of death rate worldwide. The World Health Organisation estimated 17.5 million people died from cardiovascular diseases in 2012, representing 31% of all deaths around the globe. An estimate of 16 million deaths under the age of 70 were due to non-communicable diseases, 82% of which are in low- and middle-income countries. About 7.4 million were due to coronary heart disease, and 6.7 million were due to stroke (WHO, 2015).

Article history:
Received: 02 December 2015
Accepted: 30 August 2016

E-mail addresses:
ams13_mah013@student.usm.my (Abdu Masanawa Sagir),
saratha@usm.my (Saratha Sathasivam)
*Corresponding Author

ISSN: 0128-7680

In order to investigate the misfortune of heart attack, certain factors that are associated with lifestyle need to be addressed. Therefore, people with heart disease due to the presence of chest pain, resting blood pressure, cholesterol, fasting blood sugar resting electrocardiographic and maximum heart rate need early detection and prediction for better counselling and appropriate medicine. Some factors make physicians' work even more difficult to be analysed by evaluating the existing test results of patients. As such, some complicated measures are not easy to perform when considering large number of factors. Anooj (2012) and Hedeshi and Abadeh (2014) stated that the decision about the presence or absence of a patient with certain diseases depends on the physician's intuition, experience and skill in comparing with the previous ones than on knowledge-rich data hidden in the database. This measure is a challenging task with regards to the large number of factors that has to be considered. In order to achieve our goals in this complex stage, the physician may need accurate and efficient hybrid fuzzy expert systems that can classify and predict the likelihood of a patient getting a heart disease problem and being able to help in diagnosing disease.

Classification is a process used to find a model that describes and differentiates data classes or concepts for the purpose of using the model to predict the class of objects whose class label is unknown.

Over a decade, the literatures about the use of intelligent methods in the medical sector had a vast number of related works (Muthukaruppan & Er, 2012; Sikchi et al., 2012; Kumar, 2013; Sikchi et al., 2013). The medical practitioners make use of computerised technologies to assist in diagnosis and give suggestions as medical diagnosis is full of uncertainty. According to Opeyemi and Justice (2012), the best and most efficient techniques for dealing with uncertainty is by incorporating fuzzy logic and neural network.

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

There are diverse types of studies based on ANFIS methodologies (Palaniappan & Awang, 2008; Patil & Kumaraswamy, 2009; Abdullah et al., 2011; Zhu et al., 2012; Kar & Ghosh, 2014; Mayilvaganan & Rajeswari, 2014; Yang et al., 2014).

This research work involves developing a framework that incorporates hybrid learning algorithms least square estimates with gradient descent and Levenberg-Marquardt algorithm on the training Statlog-Cleveland Heart Disease Dataset.

The remaining part of this paper is organised as follows; in section 2, the designs of newly adaptive neuro-fuzzy models are presented. This led us to section 3, in which simulation results are described, while the discussion and conclusion part of the work is given in section 4.

METHODS AND MATERIALS

According to Nguyen et al. (2003), Takagi Sugeno Kang Fuzzy model's rules are given in the form of:

$$R_i: \text{if } x_i \text{ is } A_i \text{ then } f_i(x), i = 1,2, \dots, n \tag{1}$$

where

$$f_1, f_2, \dots, f_n \text{ are functions } X = X_1 * X_2 * \dots X_k \rightarrow R \text{ and } A_i = \bigwedge_{j=1}^{k} A_{ij} \tag{2}$$

These rules are combined to get a function:

$$R(x) = \frac{A_1(x)f_1(x) + A_2(x)f_2(x) + \cdots + A_n(x)f_n(x)}{A_1(x) + A_2(x) + \cdots + A_n(x)} \tag{3}$$

This TSK fuzzy model produces a real-valued function.

ANFIS was first introduced by Jang (1993). NFIS is a framework of adaptive techniques to assist learning and adaptation. To illustrate the ANFIS structure, two fuzzy IF-THEN rules according to a first order Sugeno model are to be considered for simplicity based on the following algorithms:

Layer 1: Calculate the Membership Functions values for inputs.

$$O_i^1 = \mu_{Ai}(x) = e^{-\frac{1}{2}\left(\frac{x-c_i}{\sigma_i}\right)}, i = 1,2 \tag{4}$$

$$O_i^1 = \mu_{Bi-2}(y) = e^{-\frac{1}{2}\left(\frac{y-m_i}{\beta_i}\right)}, i = 3,4 \tag{5}$$

Layer 2: Calculate the rule firing strengths.

$$O_i^2 = w_i = \mu_{Ai}(x) * \mu_{Bi}(y) = e^{-\frac{1}{2}\left(\frac{x-c_i}{\sigma_i}\right)} * e^{-\frac{1}{2}\left(\frac{y-m_i}{\beta_i}\right)}, i = 1,2 \tag{6}$$

Layer 3: Determine the normalised firing strengths

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} = \frac{\mu_{Ai}(x) * \mu_{Bi}(y)}{\mu_{Ai}(x) + \mu_{Bi}(y)}, i = 1,2 \tag{7}$$

Layer 4: Calculate the rules outputs for rule consequent layer

$$O_i^4 = \bar{w}_i f_i = w_i(p_i x + q_i y + r_i), i = 1,2 \tag{8}$$

Layer 5: Calculate the overall output

$$O_i^5 = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}, i = 1,2 \tag{9}$$

Image of the ANFIS Structure

Figure 1 shows the structure of Adaptive Neuro Fuzzy Inference System (ANFIS), as described in equations (4) – (9). The structure of the proposed model contains five layers, input and output layers, and three hidden layers that represent membership functions and fuzzy rules.

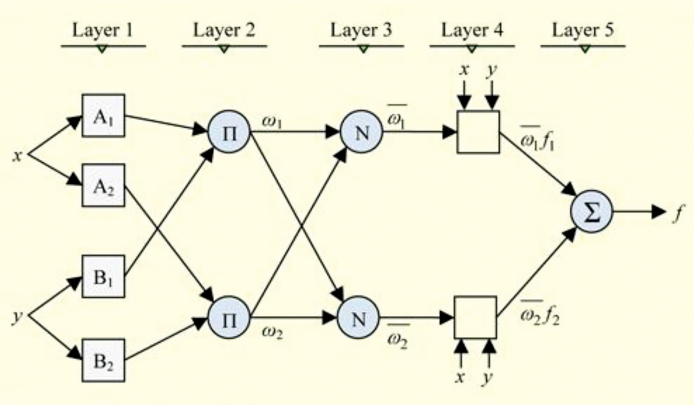


Figure 1. The ANFIS structure

Rules Index Vector

Index Membership function is the index vector that keeps track of the unique MFs. This function determines the unique MFs in the ANFIS structure and indexes them row-wise. In case two rules use the same MF, then their indices will be the same (rows are the rules, columns are the inputs). For Inputs, rule list is a Nr * Ni matrix that identifies the membership functions for the ith rule & jth input, and for outputs that include index MF, which is the index table of the unique MF used in the rules, while Nf is the number of unique MFs. Where Nr = Number of Rules and Ni = Number of Inputs. Therefore, the final index vector collects the indices found in MF "row-wise" according to the rules.

We consider a rule list of:

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$, and [3, 4, 3, 4, 2, 3, 3] the number of membership function for each input, respectively.

HYBRID LEARNING ALGORITHM OF ANFIS

The main idea of the learning algorithm is to adjust all of the modifiable parameters such as $[a_i, b_i, c_i]$ and $[p_i, q_i, r_i]$ for the purpose of matching the ANFIS output with the training data. There are two passes for hybrid algorithm, forward pass and backward pass. In the forward pass of hybrid algorithm, when the values of premise algorithm are fixed, the overall output can be expressed as a linear combination of the consequent parameters by Ziasabounchi and Askerzade (2014). For unchanging the parameters of the membership function, the output of the ANFIS model can be written as:

$$F = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}, i = 1,2$$

$$= (\bar{w}_1 x)p_1 + (\bar{w}_1 y)q_1 + (\bar{w}_1)r_1 + (\bar{w}_2 x)p_2 + (\bar{w}_2 y)q_2 + (\bar{w}_2)r_2 \quad (10)$$

From (10), the parameters p_i, q_i & r_i are to be updated by a least square estimate using Moore-Penrose pseudo-inverse, which incorporates SVD decomposition for robustness that minimises the errors $||AX - B||^2$ by approximating X with X^*.

A = Output produced by O_i^3, B = Target output, X = Unknown consequent values related to the set of consequent parameters p_i, q_i & r_i.

$$X^* = (A^T A)^{-1} A^T B \quad (11)$$

The ANFIS Models Design

Adaptive Neuro Fuzzy Inference System is one of the hybrid neuro fuzzy inference expert systems that has the potential to capture the benefits of both artificial neural network learning rules to conclude and adjust the fuzzy inference systems, particularly in Takagi Sugeno Kang type fuzzy inference system. Grid Partition method was used to create initial membership functions. At the very beginning of the training, this method divides the data space into rectangular sub-spaces using axis-paralleled partition based on a predefined number of membership functions and their types in each dimension (Wei et al., 2007).

Grid Partition method generates rules by enumerating all possible combinations of the membership functions of all inputs. Most of the researchers used less input variables (less than 5) in grid partitioning method. In our case, we used Gaussian membership functions for each of the input variables. The number of these membership functions is shown in Table 1. Seven inputs with these membership functions result in 2,592 fuzzy if-then rules. In these proposed ANFIS models (ANFIS_LSGD and ANFIS_LSLM), we set the initial learning rate, mu = 1e-1, h = 1e-8 and number of epoch = 1800.

The Proposed ANFIS_LSGD Model Design

In designing this model, a hybrid learning technique based on the ANFIS Matlab's built-in model using Least squares estimate and backpropagation gradient descent training algorithm is used.

Forward Pass

Least squares estimate (LSE) is used at the very beginning to get the initial values of the conclusion parameters, and then at backward pass for the gradient descent take over to update all parameters. When the values of the premise parameters are fixed, the overall output can be expressed as in (10), which is linear in the consequent parameters.

Let

$$B = AX \tag{12}$$

If X is invertible matrix then

$$X = A^{-1}B \tag{13}$$

$$X^{*} = (A^{T}A)^{-1}A^{T}B \tag{14}$$

Otherwise a pseuso-inverse is used to solve for X

Backward Pass

The error signal propagate backward and premise parameters are updated by gradient descent,

$$\alpha_{ij}(t+1) = \alpha_{ij}(t) - \eta.\frac{\partial E}{\partial \alpha} \tag{15}$$

where η is the learning rate for α_{ij}, which can be further expressed as:

$$\eta = \frac{k}{\sqrt{\sum_{\alpha}\left(\frac{\partial E}{\partial \alpha}\right)^{2}}} \tag{16}$$

where k is the step size, the length of each gradient transition in the parameter space.

The chain rule is used to calculate the partial derivatives to update the membership function parameters:

$$\frac{\partial E}{\partial \alpha_{ij}} = \frac{\partial E}{\partial f}.\frac{\partial f}{\partial f_i}.\frac{\partial f_i}{\partial w_i}.\frac{\partial w_i}{\partial \mu_{ij}}.\frac{\partial \mu_{ij}}{\partial \mu_{ij}} \tag{17}$$

The Proposed ANFIS_LSLM Model Design

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

Forward Pass

Least squares estimate (LSE) was used at the very beginning to get the initial values of the conclusion parameters, and then at backward pass for the Levenberg-Marquardt algorithm take over to update all the parameters. When the values of the premise parameters are fixed, the overall output can be expressed as in (10), followed by (11) – (14).

Backward Pass

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

$$F(w) = \frac{1}{2}E^T E \tag{18}$$

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

$$\mathbf{W}_k(t+1) = \mathbf{W}_k - \left(J_k^T J_k + \mu I\right)^{-1} J_k^T E(w) \tag{19}$$

$$\Delta \mathbf{W}_k = (J_k + \mu I)^{-1} J_k^T E(w) \tag{20}$$

Get the parameters of unique MF_s of current FIS as presented in rules index vector; and Obtain Cumulative Current Error vector and RMSE

$$E(w) = [e_{11} \dots e_{k1}\, e_{12} \dots e_{k2} \dots e_{kp}]^T \tag{21}$$

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

$$rmse1 = \sqrt{\frac{E_p}{p}} \tag{22}$$

Built up Jacobian matrix column–wise, which contains 1st order partial derivatives of network error using the central difference method

$$f'(x_0) = \frac{f_1 - f_{-1}}{2h} + E_{trunc}(f, h), \tag{23}$$

where $f_1 = f(x + h)$

Therefore,

$$J_{i,j} = \frac{\partial f_i}{\partial x_j} \tag{24}$$

Transform Jacobian into sparse matrix to speed things up

$$J_{i,j} = \text{Sparse}\left(\frac{\partial f_i}{\partial x_j}\right) \tag{25}$$

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

$$H \approx J^T J \tag{26}$$

Therefore,

$$H_{i,j} = \frac{\partial^2 f_i}{\partial x_i \, \partial x_j} \tag{27}$$

Compute the error gradient

$$g = J^T E \tag{28}$$

Update the Hessian matrix

$$H^* = [H + \psi I] \tag{29}$$

where I is the sparse identity matrix and $\psi = 0.1$ is the learning parameter; and the network parameter needs to be updated using (20).
Recalculate the RMSE using the updated parameters,

$$rmse2 = \sqrt{\frac{E_p}{p}} \tag{30}$$

Adjust the learning parameters; if total error is decreased as a result of the update, then go to the next epoch; otherwise, if total error is increased, then increase learning rate. Finally, if rmse2 is less than rmse1, accept; otherwise reject.

Description of Input attributes

The dataset is available from the University of Strathclyde, Glasgow, Scotland, U.K., via Ross (1992). Detailed information about the input variables is shown in Table 1.

Table 1
Information about input variables

Variable Name	Min	Max	No. of MF	Description of Input Variable	Type
AGE	29	77	3	Age (very young, young & old)	Real
CP	1	4	4	Chest pain type (1-typical angina, 2- atypical angina, 3- non-anginal pain, 4-asymtomatic)	Nominal
TRESTBPS	94	200	3	Resting blood pressure (Low, Normal & High)	Real
CHOL	126	564	3	Cholesterol (low, medium, high & very high)	Real
FBS	0	1	2	Resting blood sugar (0=false, 1=true) it is true when fbs>120	Binary
RESTECG	0	2	3	Resting electrocardiographic (0-normal, 1-having ST-T & 2-showing definite left VH)	Nominal
THALACH	71	202	2	Maximum heart rate (low, normal & high)	Real

Accuracy, Sensitivity and Specificity

To be confidently used in medical decision-making, the test methods must meet tough standards of statistical measurements: sensitivity, specificity and accuracy, which are the terms most commonly associated with the Binary classification test and they statistically measure the performance of the test by Saed (2015).

RESULTS AND DISCUSSION

In this research work, an attempt was made to develop and examine the ability of two ANFIS models for predicting heart disease.

Performance Evaluation

After training the system, it has to be tested with a set of testing dataset so as to verify the capacity of the models. This determines how well the ANFIS models are worked. Accuracy was calculated based on the correct classified instances divided by the total number of instances (Patil et al., 2010).

System Validation

Validation is the process of determining the degree to which a model is an accurate real world representation from the perspective of the intended uses of the model (Thacker et al., 2004). Therefore, to measure the stability of performance, the data are divided into training and testing using the validation method. Hold-out validation method is used for testing of results.

Hold-out Validation Method

The data were split into two groups, namely, training set and test set. The total number of samples used was 270, out of which the first 180 were for training and the remaining 90 for testing.

Results of Accuracy, Sensitivity and Specificity

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

Classification results of the correct and incorrect number of predicted values, as well as the performance of our results of accuracy are shown in Table 4 and Table 5, respectively.

Table 2
Obtained Statlog-Cleveland Heart Disease classification results

	ANFIS_LSGD	ANFIS_LSLM
Train-correct predicted value	135	135
Train-incorrect predicted value	45	45
Test-correct predicted value	68	69
Test-incorrect predicted value	22	21

Table 3
Accuracy Results of the proposed models

	ANFIS_LSGD		ANFIS_LSLM	
	Train (%)	Test (%)	Train (%)	Test (%)
Sensitivity	68.29	71.05	68.29	71.05
Specificity	80.61	78.85	80.61	80.77
F-measure	71.34	71.05	71.34	72.00
Precision	74.67	71.05	74.67	72.97
Accuracy	75.00	75.56	75.00	76.67

Root Mean Square Error (RMSE)

Root Mean Square Error is one of the most acceptable indicators that describes the differences between the actual data and the predicted values. The values of the premise and consequent parameters can be obtained after network training by directly minimising the RMSE performance criterion (Ho et al., 2009).

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y - y')^2} \tag{31}$$

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

Table 4
The values of RMSE

	ANFIS_LSGD	ANFIS_LSLM
RMSE	0.40344	0.40327

Comparison of the Results

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

Table 5
A comparison of the results with other algorithms based on error rates (train & test) (Michie et al., 1994).

Algorithm	Error (Train)	Error (Test)	Reference
Proposed (ANFIS_LSLM)	**0.25**	**0.233**	**This study**
Proposed (ANFIS_LSGD)	**0.25**	**0.244**	**This study**
k-NN,k=30,eucl,std	-	0.344	KG
NaiveBay	0.351	0.374	Statlog
Discrim	0.315	0.393	Statlog
Logdisc	0.271	0.396	Statlog
ALLOC80	0.394	0.407	Statlog
Quadisc	0.274	0.422	Statlog
CASTLE	0.374	0.441	Statlog
Cal5	0.330	0.444	Statlog
CART	0.463	0.452	Statlog
Cascade	0.207	0.467	Statlog
k-NN	0.000	0.478	Statlog
SMART	0.264	0.478	Statlog
DIPOL92	0.429	0.507	Statlog
ITrule	*	0.515	Statlog
Baytree	0.111	0.526	Statlog
Default	0.560	0.560	Statlog
Backprop	0.381	0.574	Statlog
LVQ	0.140	0.600	Statlog
IndCART	0.261	0.630	Statlog
Kohonen	0.429	0.693	Statlog
AC2	0.000	0.744	Statlog
CN2	0.206	0.767	Statlog
RBF	0.303	0.781	Statlog
C4.5	0.439	0.781	Statlog
NewID	0.000	0.844	Statlog
k-NN, k = 1, eucl, std	-	0.725	KG

Table 6
A comparison of the results with other classifiers for the Statlog-Cleveland Heart Disease (Department of Informatics, n.d.)

Method	Accuracy %	Reference
ANFIS_LSLM	**76.7**	**This study**
ANFIS_LSGD	**75.6**	**This study**
IR	71.4	WEKA, RA
T2	68.1	WEKA, RA
FOIL	64.0	WEKA, RA
RBF	60.0	ToolDiag, RA
InductH	58.5	WEKA, RA

Discussion of the Results

The number of correct and incorrect predicted values of the two classifiers is presented in Table 2. Both of the two classifiers predicted 135 and 45 as correct and incorrect values for the training dataset. For the test dataset, the two classifiers, ANFIS_LSGD and ANFIS_LSLM, classified the predicted values 68 & 22, and 69 & 21 as the correct and incorrect values, respectively.

The present research work further evaluates the performance of the proposed models using the Statlog-Cleveland benchmark dataset for Heart Disease prediction. The results showed that the performance measures were more important for interpreting the result of a classifier. The test performance of the classifiers was determined by the computation of sensitivity, specificity and total classification accuracy. As shown in Tables 3 and 4, the ANFIS_LSLM classifier yields better results when compared with the ANFIS_LSGD classifier.

Generally, the results of ANFIS_LSLM outperform the traditional ANFIS_LSGD even though the computational time for generating the ANFIS_LSGD was slightly faster when compared with ANFIS_LSLM. This is due to the complexity of computation of Jacobian matrix for each iteration, but it still yields better accuracy, which can be considered as efficient.

For the purpose of comparing with other algorithms, the results are shown in Table 5. Our proposed algorithms are found to be better than the accuracies of other algorithms in the literature for the Statlog-Cleveland heart disease dataset. From the previous research work and Table 6, it is clearly confirmed that none of the research studies has success rates higher than 71.4% for the mentioned algorithms on the Statlog-Cleveland Heart Disease Dataset. Based on the comparison of the results, it can be seen that the proposed models produced reasonable results in classifying the possible heart disease patients.

CONCLUSION

The objective of this study was to design two different ANFIS based classification models for heart disease prediction. It was observed that the classifiers learnt how to classify the dataset. Their performances were evaluated based on training, testing and accuracy of classification. We further conclude that this research work has so many features. We used grid partition technique, seven input variables with Gaussian membership functions resulting in 2,592 rules and achieved an accuracy of about 70% - 80% level. The root mean square error with the LS+LM and LM+GD algorithms after 1,800 iterations was found to be 0.40327 and 0.40344, respectively. The programme automatically compares the Matlab's built-in training to the Levenberg-Marquardt training and displays the results in a table format. The total classification accuracies obtained were 76.67% and 75.56% for the ANFIS_LSLM and ANFIS_LSGD classifiers, respectively.

The Levenberg-Marquardt algorithm has the problem of computational complexity of Jacobian matrix J at each iteration step by taking first order partial derivative and the inversion of J^TJ square matrix, the dimension of which is NxN. In the present work, the Jacobi is computed via central difference, which is made for a faster convergence speed by using sparse structure.

The results belong to the first attempt of study and confirmed that our proposed models were better than other models in the literature as they have the potential for classifying and predicting heart diseases. We thought if we adapted another means of computation of Jacobian matrix, the results would be improved for ANFIS_LSLM.

FUTURE WORK

The proposed ANFIS models could be enhanced in the future by:

1. Applying another means of derivation in computation of Jacobian matrix in order to increase convergence speed of the results.
2. Adapting another training algorithm such as Scaled Conjugate Gradient algorithm to produce better results.

ACKNOWLEDGEMENTS

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

REFERENCES

Abdallah, M., Warith, M., Narbaitz, R., Petriu, E., & Kennedy, K. (2011). Combining fuzzy logic and neural networks in modelling landfill gas production. *International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering, 5*(6), 273-279.

Anooj, P. K. (2012). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University-Computer and Information Sciences*, *24*(1), 27-40.

Hedeshi, N. G., & Abadeh, M. S. (2014). Coronary artery disease detection using a fuzzy-boosting PSO approach. *Computational Intelligence and Neuroscience*, *2014*(2014), 1-12.

Ho, W. H., Tsai, J. T., Lin, B. T., & Chou, J. H. (2009). Adaptive network-based fuzzy inference system for prediction of surface roughness in end milling process using hybrid Taguchi-genetic learning algorithm. *Expert Systems with Applications*, *36*(2), 3216-3222.

Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, *23*(3), 665-685.

Kahramanli, H., & Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. *Expert Systems with Applications*, *35*(1), 82-89.

Kar, S., Das, S., & Ghosh, P. K. (2014). Applications of neuro fuzzy systems: A brief review and future outline. *Applied Soft Computing*, *15*, 243-259.

Kathy, L. T. (2013). *Classification*. In Oracle Data Mining Concepts (Part II). Retrieved September 10, 2015, from http://docs.oracle.com/cd/E11882_01/datamine.112/e16808/classify.htm#DMCON035.

Kumar, A. S. (2013). Diagnosis of heart disease using Advanced Fuzzy resolution Mechanism. *International Journal of Science and Applied Information Technology (IJSAIT)*, *2*(2), 22-30.

Mayilvaganan, M., & Rajeswari, K. (2014). Risk Factor Analysis to Patient Based on Fuzzy Logic Control System. *Blood Pressure*, *2*(5), 185-190.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). *Statlog Datasets: Comparison of Results*. Retrieved September 09, 2015, from http://www.is.umk.pl/projects/datasets-stat.html.

Muthukaruppan, S., & Er, M. J. (2012). A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease. *Expert Systems with Applications*, *39*(14), 11657-11665.

NCU. (n.d.). *Logical rules extracted from data.* Department of Informatics, Nicolaus Copernicus University. Retrieved September 09, 2015, from http://www.fizyka.umk.pl/kmk/projects/rules.html.

Nguyen, H. T., Prasad, N. R., Walker, C. L., & Walker, E. A. (2003). *A First Course in Fuzzy and Neural Control* (p. 112). United States of America, USA: Chapman and Hall/CRC Press.

Opeyemi, O., & Justice, E. O. (2012). Development of Neuro-fuzzy System for Early Prediction of Heart Attack. *International Journal of Information Technology and Computer Science (IJITCS)*, *4*(9), 22-28.

Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. In *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on* (pp. 108-115). IEEE.

Patil, S. B., & Kumaraswamy, Y. S. (2009). Intelligent and effective heart attack prediction system using data mining and artificial neural network. *European Journal of Scientific Research*, *31*(4), 642-656.

Patil, B. M., Joshi, R. C., & Toshniwal, D. (2010). Hybrid prediction model for Type-2 diabetic patients. *Expert Systems with Applications*, *37*(12), 8102-8108.

Ross, D. K. (1992). *Statlog Databases*. Retrieved September 07, 2015, from https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/.

Saed, S. (2015). *Model Evaluation*. Retrieved September 10, 2015, from http://www.saedsayad.com/model_evaluation.htm.

Sikchi, S. S., Sikchi, S., & Ali, M. S. (2012). Design of fuzzy expert system for diagnosis of cardiac diseases. *International Journal of Medical Science and Public Health*, *2*(1), 56-61.

Sikchi, S. S., Sikchi, S., & Ali, M. S. (2013). Fuzzy Expert Systems (FES) for Medical Diagnosis. *International Journal of Computer Applications (0975–8887)*, *63*(11), 7-17.

Thacker, B. H., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., & Rodriguez, E. A. (2004). *Concepts of model verification and validation* No. LA-14167). Los Alamos National Lab., Los Alamos, NM (US).

Wei, M., Bai, B., Sung, A. H., Liu, Q., Wang, J., & Cather, M. E. (2007). Predicting injection profiles using ANFIS. *Information Sciences*, *177*(20), 4445-4461.

WHO. (2015). *Cardiovascular diseases*. Retrieved August 28, 2015 from http://www.who.int/mediacentre/factsheets/fs317/en/.

Yang, J., Kim, J., Park, Y., & Lee, Y. (2014). A study on a prediction model for coronary heart disease risks through optimizing ANFIS and RBFN. *Experimental and Clinical Cardiology, 20*(9), 5749-5760.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338-353.

Zadeh, L. A. (2004). *Fuzzy Logic Systems: Origin, Concepts, and Trends*. Retrieved November 29, 2015, from http://wi-consortium.org/wicweb/pdf/Zadeh.pdf.

Zhu, B., He, C. Z., Liatsis, P., & Li, X. Y. (2012). A GMDH-based fuzzy modeling approach for constructing TS model. *Fuzzy Sets and Systems*, *189*(1), 19-29.

Ziasabounchi, N., & Askerzade, I. (2014). ANFIS Based Classification Model for Heart Disease Prediction. *International Journal of Electrical & Computer Sciences IJECS-IJENS*, *14*(02), 7- 12.

Pertanika J. Sci. & Technol. 25 (1): 57 - 76 (2017)

SCIENCE & TECHNOLOGY

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

An ID-based Secure and Flexible Buyer-seller Watermarking Protocol for Copyright Protection

Ashwani Kumar[1]*, S. P. Ghrera[1] and Vipin Tyagi[2]

[1]*Department of CSE, Jaypee University of Information Technology, Waknaghat, P.O. Waknaghat, Teh Kandaghat, Distt. Solan, 173234, India*

[2]*Department of CSE, Jaypee University of Engineering and Technology, A-B Road, Raghogarh, 473226, India*

ABSTRACT

Digital watermarking protocols are the one, which have combined fingerprinting technique with watermarking, for embedding digital signal or watermark into an original multimedia object. Buyer-seller watermarking protocol is fundamentally applied to continue the digital rights of both buyers and seller. We proposed an identity-based buyer-seller watermarking protocol that encounters various weaknesses of Zhang et al.'s watermarking protocol. We ensured that by pointing out these weaknesses, inaccuracy can be minimised for further implementing the buyer-seller watermarking protocol. The suggested protocol uses ID-based public key cryptography and digital watermarking scheme to place the ownership of digital content. Hence, copyright protection is attained. We claim that our suggested protocol is efficient and has adequate security as compared to traditionally proposed protocols, and therefore suitable for any practical buyer-seller watermarking scheme.

Keywords: Digital content, Identity-based technique, Public key cryptography, Digital watermarking, Copyright protection

INTRODUCTION

The speedy development of internet and e-commerce needs a copyright protection mechanism for multimedia data. Digital watermarking becomes an important technique for protecting the digital rights. The principal object of digital watermarking technique is: (Mintzer & Braudaway, 1999) to retain digital copyright or watermark, embedded into the cover object. The desirable secure digital watermarking scheme is one, which integrates public key cryptosystem and digital watermarking technique for protecting the buyers and seller in a digital content transaction. Digital watermarking

Article history:
Received: 14 January 2016
Accepted: 26 July 2016

E-mail addresses:
ashwani.kumarcse@gmail.com (Ashwani Kumar),
sp.ghrera@juit.ac.in (S. P. Ghrera),
dr.vipin.tyagi@gmail.com (Vipin Tyagi)
*Corresponding Author

ISSN: 0128-7680

(Memon & Wong, 2001) techniques use encrypted domain for embedding and extracting the watermarks. The rapid growth of the internet encourages some bad usage too; these include operations such as transformation, duplication and redistribution of digital content. With the avail of some software tools, we can easily identify these bad users and redistribution of digital content can also be placed. In general, secure digital watermarking (Zeng et al., 2011) scheme should satisfy the following requirements:

Robustness
The capability of the watermark to resist various image processing attacks such as rotation, scaling, cropping, etc.

Imperceptibility
The optical aberration of the watermarked image should not bear on the characters of the original picture.

Effectiveness
The algorithms for embedding and extracting the watermark from the digital content should be effective.

Memon et al. (2001) proposed the very first buyer-seller watermarking protocol in 2001, and Ju et al. (2002) modified this protocol with various advances. In history, various protocols have been proposed (Choi et al., 2003; Goi & Phan, n.d; Hu & Zhang, 2009; Hu & Li, 2009). Digital watermarking (Zeng et al., 2011) algorithm is divided into two parts; first non-blind watermarking scheme and second, blind watermarking scheme. The non-blind watermarking scheme needs original cover object as well as watermark and watermark key for extracting the watermark, while blind watermarking scheme does not require cover object, watermark key and watermark for detection or extraction of the watermark. The buyer-seller watermarking protocol (Kumar et al., 2011a; Kumar et al., 2011b) is a three-party protocol among a service provider, a customer, and a trusted watermark certificate authority. This protocol combines fingerprinting and encryption techniques for protecting the participants into any transaction. A very common buyer-seller watermarking protocol consists of four sub-protocols the registration protocol, the watermarking protocol, the identification and arbitration protocol, and the dispute resolution protocol. The buyer-seller watermarking protocol (Kumar et al., 2011a; Kumar et al., 2011b) is expected to solve the problems, which are given below.

Certification authority problem
In this, a digital certificate is given for the participants involving in a transaction.

The conspiracy problem
Malicious parties may collude with each other and mount attacks to cast an innocent buyer or to confound the tracing by removing the watermark from the digital content.

The customer's rights problem
Customer's right problem states that when the service provider embeds a watermark information into digital content, and have the advantage to frame the customer.

The piracy-tracing problem
In this, the service provider is supposed to trace illegal copies of digital content. Therefore, the redistribution of digital content can be controlled by the seller.

The unbinding problem
Unbinding means unable to bind watermark for the digital rights. In this problem, service providers can fabricate piracy of the customer by manipulating customer's watermark.

The anonymity problem
In the anonymity problem, during the transaction, the customer identity should be hidden until the customer is declared as the culprit.

According to history, Hwang et al. (2005) introduced a time stamping protocol in 2005. In their protocol, a TTP (trusted third party) was introduced for checking the verification and signing phase. Ju et al. (2002) proposed an anonymous buyer-seller watermarking protocol with anonymity control in 2002. In the paper, the authors have identified the anonymity problem. They discussed that a buyer could purchase digital content anonymously, but the anonymity can be controlled. Zhang et al. (2006) proposed a secure buyer–seller watermarking protocol in 2006. In his paper, no assistance is needed, so that it avoids the conspiracy problem, piracy tracing problem and customer's right problem. There are only two participants, a seller and a buyer. The protocol can simultaneously resolve many problems. However, there is a drawback in Zhang et al.'s protocol, i.e. the buyer's assistance is needed to solve the piracy dispute. Therefore, dispute resolution and unbinding problems exist in the protocol by Zhang et al. (2006).

We proposed an identity-based buyer-seller watermarking protocol and encountered various existing weaknesses of Zhang et al.'s (2006) protocol such as dispute resolution and unbinding problem. Here, we proposed a new identity-based buyer-seller watermarking protocol to prove the ownership of digital content. Our proposed protocol enables the seller to produce the watermarked content with their private key. The watermark certificate authority (WCA) is responsible for issuing the digital signature that corresponds to ID of the seller, timestamp (Hwang et al., 2005) used for watermark content, watermark and cover object. WCA is maintaining its own table and keeping the requested IDs of both buyer and seller; suppose if dispute occurs, the buyer can communicate or confirm to the WCA to checkout that whether he/she is the original buyer or not. If any dispute occurs at a later stage, with the help of arbiter, it can also be resolved to check the correctness of information used by the seller. Timestamps are compared by the arbiter to identify the appropriate seller of digital content and with the help of timestamps, the unbinding problem is also solved. Some key details of our proposed watermarking protocol are identified below:

1. In our proposed protocol, we adopt wavelet and principal component analysis based techniques (kumar et al., 2015) with identity-based public key cryptography.
2. This watermarking protocol must be autonomous of all watermarking schemes.
3. Our protocol makes use of a tamper resistance device, which is embedded into seller's computer and reduces the overhead on WCA as TTP.

The rest of the paper is structured as follows. In Section Two, we review the scheme of Zhang et al. (2006) and identify previously unpublished problems. Section Three describes the proposed ID-based buyer-seller, watermarking protocol. Section Four discusses the security analysis. Section Five shows the experimental results. Finally, Section Six concludes our paper.

REVIEWING THE SCHEME OF ZHANG *ET AL.*

Zhang et al. (2006) proposed a secure buyer-seller watermarking protocol in 2006. The authors proposed a secure buyer-seller watermarking protocol without the assistance of a TTP in which there are only two participants, the seller and buyer. Zhang et al.'s paper is based on the Lei et al. (2004) and in this, no third party is brought in; therefore, the proposed protocol is more childlike and more dependable than the existing watermarking protocol. Zhang et al.'s protocol resolves the conspiracy problem, piracy tracing problem and customer's right problem. However, there is a drawback in Zhang et al.'s protocol, i.e. the buyer's assistance is needed to solve the piracy dispute problem.

The protocol of Zhang et al. composes of three sub-protocols: the registration protocol, the watermarking protocol and the identification and arbitration protocol. Here, we show the notations of Zhang et al.'s protocol.

$E_{pk*}(X')$ = encrypted watermark image

$E_{pk*}(X'')$ = second round encrypted watermark image

$E_{pk*}(W)$ = encrypted watermark

$Cert_{CA}(pk_B)$ = digital certificate of CA

pk_B, sk_B = random key pair

ARB = Arbiter

SEC_B = secret key of buyer

SEC_S = secret key of seller

$E_{pk^*}(SEC_B)$ = encrypted secret key

$Sign_{sk^*}(E_{pk^*}(SEC_B))$ = sign encrypted secret key

$Cert_{pk_B}(pk^*)$ = anonymous certificate

In the protocol of Zhang et al., a seller randomly generates a secret $\mathrm{SEC_S}$ key. In the encrypted domain, the seller obtains the encrypted watermark $\mathrm{E_{pk*}(W)}$, as follows.

$$E_{pk*}(W) = E_{pk*}(SEC_S) \otimes E_{pk*}(SEC_B)$$
$$E_{pk*}(SEC_S \oplus SEC_B) \quad (2.1)$$

Seller S then inserts the second round watermark through the following formula:

$$E_{pk*}(X'') = E_{pk*}(X') \otimes E_{pk*}(W)$$
$$E_{pk*}(X' \oplus W) \quad (2.2)$$

Zhang et al. claimed that their proposed secure buyer-seller watermarking protocol provides a solution for conspiracy problem, piracy-tracing problem and customer's right problem. We have identified that the protocol is unable to solve the dispute resolution problem and unbinding problem. Figure 1 shows a simplified trading model, based on the protocol by Lei et al.

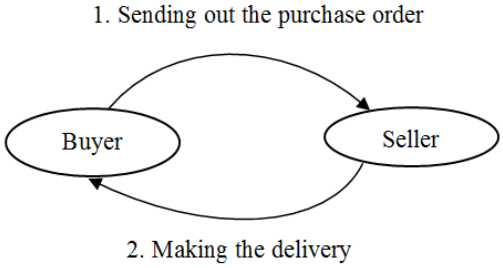


Figure 1. A simplified trading model

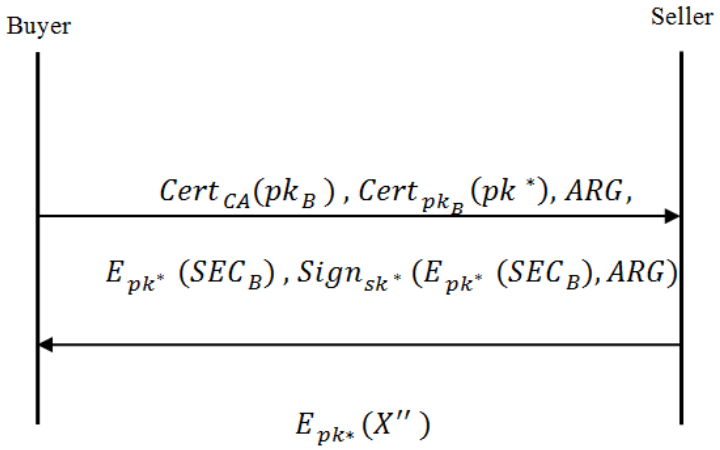


Figure 2. The encryption phase of J. Zhang secure buyer-seller watermarking protocol (Zhang et al., 2006).

In Zhang *et al.*'s protocol, the unbinding problem arises because once the seller finds out a pirated copy, there is a potential for a seller to transplant the watermark embedded in the pirated copy into another copy of digital content. Dispute resolution problem also exists because the data encrypted by the judge may not be equal to the data furnished by the seller. To level out these topics, we have proposed identity-based cryptographic scheme (Zeng et al., 2011) into the watermarking algorithm. ID-based techniques were introduced by Shamir in 1984.

PROPOSED ID-BASED BUYER-SELLER WATERMARKING PROTOCOL

This paper is an extension of our previous work (Kumar et al., 2011a; Kumar et al., 2011b). In our suggested protocol, the same trust model used by Memon et al. (2001) and Lei et al. (2004) is employed. The proposed protocol is based on public key infrastructure, arbiter, ID-based public key cryptography (Cox et al., 1997; Paillier, 1999) and digital watermarking scheme. The watermarking scheme involves secret key and digital signature certificate issued by WCA. WCA maintains its own table and keeps the requested IDs of both buyer and seller because it contains a database. Our protocol is flexible because it makes use of tamper resistance device, which is utilised to reduce the overhead on WCA and also solves the problems listed in section 1. Now, in the digital signature verification phase, someone else can use the WCA public keys to validate that the watermarked content embedded at a certain time into the digital content. Our proposed digital watermarking protocol consists of three sub-protocols: the watermark embedding and signing protocol; watermark detecting and verifying protocol; and registration protocol, as presented in Figure 4. We first determine the roles and notations for various participants in our proposed protocol, as presented in Figure 3.

Seller

The owner of the digital content or from where the buyer wants to purchase the digital content.

Spurious buyer

Th person who wants to learn the rightful side of the digital capacity that does not belong to him.

WCA

Public, private, and shared secret key is issued by this authority. The valid watermark and digital signature are also generated by *WCA*.

ARB

ARB stands for an arbiter; if any dispute occurs between the buyer and seller, that dispute is resolved by arbitration. ARB also verifies the correctness of the digital certificates.

Tamper-resistant device

This device is detached from seller's computer and used to produce necessary watermarks and digital signature.

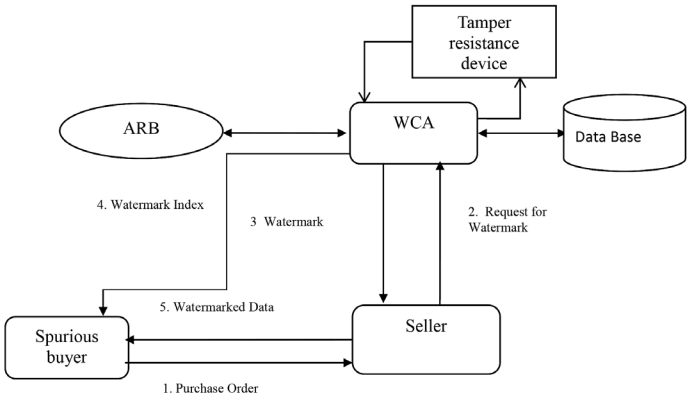


Figure 3. ID-based buyer-seller watermarking protocol.

We have shown assumptions of our suggested protocol.

1. Buyer-seller and *WCA* contain a matched clock. This clock is held securely.
2. The cover object is an image in which the watermark is applied.
3. WCA is assumed to be trustworthy.
4. The buyer-seller communicates through a secure channel.
5. The valid watermark is generated by the seller and WCA.

The goals of the proposed watermarking scheme are described below:

1. Our protocol solves the unbinding problem, the dispute resolution problem and it also identifies the spurious seller who claims the ownership of digital content.
2. The buyer interacts with the seller but one time.
3. The buyer does not possess any knowledge of cryptosystem and the embedded watermark.
4. Our protocol avoids the double watermark insertion and *WCA* is responsible for the generation of the watermark.

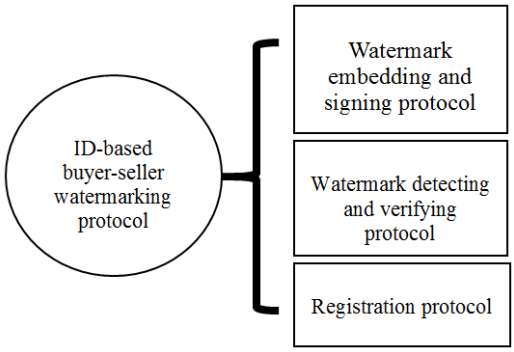


Figure 4. Three sub-protocols of ID-based buyer-seller watermarking protocol.

We have shown the roles and notations of our proposed ID-based secure and flexible buyer-seller watermarking protocol.

X	= original image
W	= watermark
W'	= forge watermark
Z	= forged digital content
X_W	= watermarked image or data
$E_{H(w_k)}$	= encrypted watermark key
$D_{H(w_k)}$	= decrypted watermark key
C	= cipher text

t = timestamp

$SE_{PC_{wca}}$ = seller performs encryption using WCA public key

ARG = arbiter

ID_S = seller credential

$SD_{K_{wca}}$ = seller performs decryption using WCA public key

$Ver_{PC_{wca}}(Ds)$ = verification of digital signature

$Ds(Sig_{K_{wca}})$ = digital signature Ds generated by WCA using its private-key.

Watermark embedding and signing protocol

The watermark embedding and signing protocol are described in Figure 5. The protocol is being executed multiple times for authentication of a buyer between the seller and WCA. If the seller wants to establish the lawful ownership of their digital content, i.e. image X, then the seller can carry out the embedding and then sign a protocol with WCA, as given in Figure 5. The deal between the seller and WCA is given below.

1. The seller selects a random and robust watermark *W*.

2. The seller embeds the watermark *W* into digital content *X* to obtain watermarked data X_w.

$$X_W = E_{H(w_k)}(X, W) \tag{3.1}$$

where *E* is the watermark embedding algorithm and $H(w_k)$ is the watermark key.

3. The seller then converts plaintext *P* into ciphertext *C* using the public-key cryptosystem, PC_{wca}.

$$C = SE_{PC_{wca}}(ID_S, t, X, W, X_W) \tag{3.2}$$

where X_W is provided by the seller and t is the creation of time of the watermarked content, i.e X_W.

4. The seller sends the ciphertext C to the WCA.

5. After receiving C, WCA performs some decrypt operation.

$$(ID_S, t, X, W, X_W) = SD_{K_{wca}}(C) \tag{3.3}$$

where K_{wca} is the private key of WCA and *D* is the decryption.

6. WCA checks whether ID_S is legitimate or not. If not, WCA aborts the sub-protocol.

7. If the time t is accepted, WCA then checks to confirm that the watermarked content X_W has been constructed by embedding the watermark W in X.

8. If watermark content X_W is valid, then WCA generates the digital signature Ds using WCA private key K_{wca}.

$$Ds = Sig_{K_{wca}}(ID_S, t, W, H(w_k)) \tag{3.4}$$

9. WCA sends this digital signature to the seller.

10. After receiving the digital signature Ds, seller verifies it using the public key of WCA.

$$Ver_{PC_{wca}}(Ds) = (ID_S, t, W, H(w_k)) \tag{3.5}$$

11. If the digital signature is valid, then the seller keeps Ds, t and W in their local database. After successfully completing the watermark embedding and signing protocol, the Seller can publicise the digital watermarked content X_W.

Watermark detecting and verifying protocol

The watermark detecting and verifying protocol is described in Figure 6. This protocol takes place between the buyer and the arbiter (ARB). If the arbiter receives a forge digital content, let say, Z and seller consist K_S, t, Ds, W, X. Then, seller can claim the rightful ownership of Z by executing the watermark detecting and verifying protocol.

1. The seller sends $ID_S, t, W, H(w_k), Ds$ and X to arbiter.

2. After getting all information from the seller, arbiter uses the watermark detection algorithm:

$$D_{H(w_k)}(Z, X, W) \tag{3.6}$$

Where D belongs to watermark detecting scheme. If we receive the result of the above equation, equal to 1, then Z consist watermark W and if the effect of above equation equal to 0 then Z does not contain watermark W and arbiter performs next step.

3. After step 2, arbiter verifies the validity of the digital signature Ds using the following equation:

$$Ver_{PC_{wca}}(Ds) = (ID_S, t, W, H(w_k)) \tag{3.7}$$

where PC_{wca} is the public-key cryptosystem of WCA. If Eq. (3.7) is true, the arbiter returns their own key $ID_S, t,$ otherwise, arbiter returns 0.

Registration Protocol

The registration protocol takes place between the customer and the WCA. If a buyer wants to hide his identity to a transaction of digital content, then the buyer randomly selects a pair of key Rpk_B, Rsk_B and sends Rpk_B to a trustworthy WCA [11]. After receiving Rpk_B, WCA generates an anonymous digital certificate $Cert_{WCA}(Rpk_B)$ and sends it to the buyer. If the buyer does not require anonymity, the entire registration process can be skipped and normal digital certificate can be practiced by the buyer.

Seller		WCA
1. Seller select watermark W 2. $X_W = E_{H(w_k)}(X, W)$ 3. $C = SE_{PC_{wca}}(ID_S, t, X, W, X_W)$	4. C →	5. $(ID_S, t, X, W, X_W) = SD_{K_{wca}}(C)$ 6. WCA checks ID_S, t 7. Checks X, X_W
10. $Ver_{PC_{wca}}(s) = (ID_S, t, W, H(w_k))$	9. Ds ←	8. $Ds = Sig_{K_{wca}}(ID_S, t, W, H(w_k))$
11. Stores Ds, t, W into the database		

Figure 5. Watermark embedding and signing protocol

Seller		Arbiter (ARB)
1. Seller sends $(ID_S, t, W, H(w_k), Ds, X)$		2. Perform decryption $D_{H(w_k)}(Z, X, W)$ 3. Arbiter verifies $Ver_{PC_{wca}}(s) = (ID_S, t, W, H(w_k))$

Figure 6. Watermark detecting and verifying protocol

Table 1
A comparison of our proposed scheme with the existing protocols.

	[11]	**[6]**	**[4]**	**[12]**	**[Our Scheme]**
The customer's rights problem	Solved	solved	solved	solved	solved
The piracy-tracing problem	Solved	solved	not tested	solved	solved
The unbinding problem	Solved	not solved	not solved	solved	Solved
The anonymity problem	partially solved	solved	not solved	partially solved	Solved
The dispute resolution problem	Solved	not solved	not solved	solved	Solved
Tamper-resistant WCA device With Database	No	No	No	No	Yes

SECURITY ANALYSIS OF THE PROPOSED SCHEME

In this section, the security of the proposed ID-based secure and flexible buyer-seller watermarking protocol is analysed. The security of the proposed protocol depends on the security and robustness of the underlying watermarking embedding and detecting scheme. We have examined the security of our protocol and compared it with the scheme (Memon & Wong, 2001; Choi et al., 2003; Lei et al., 2004; Zhang et al., 2006). The proposed watermarking protocol is secure and flexible for the reason that buyer has no idea about the original digital content **X**, and hence, is unable to remove the watermark. Since seller gets no access to the watermarked copy of the digital content **X**, hence, the seller cannot distribute illegal replicas

of digital content X. Our proposed protocol has successfully solved the problems identified in the protocol of Zhang et al., which are listed in section 1. The security of the proposed protocol is examined and compared with the previously published work (Memon & Wong, 2001; Choi et al., 2003; Lei et al., 2004; Zhang et al., 2006) in tabular form.

Table 2
A comparison of computation cost with existing protocol

Encryption Operation	3	2k+1	2	3	1
Decryption Operation	1	4	1	1	1
⊕ operation	2	2	2	2	2
Signing Operation	1	k	2	2	1
k is the number of watermarks					

Table 1 and Table 2 show the comparison of various results. In Table 1, the seller uses a tamper resistant device, which produces necessary watermarks and digital signature. Table 1 shows that our protocol can withstand all of the known problems which are listed above and identify the true owner of the digital content. Table 2 shows the various encryption and decryption operation used in our proposed protocol. We used three sub-protocols watermark embedding and signing the protocol, watermark detection and verifation protocol and registration protocol. In our scheme, the number of communication rounds takes one encryption, one decryption, two watermark embedding, and one signing operation respectively, which minimises the passing time and also reduces overhead on WCA, and hence is better when compared to others such as that of Memon & Wong, 2001; Choi et al., 2003; Lei et al., 2004; Zhang et al., 2006. Furthermore, if buyer sends a request to purchase a product with anonymity, the then seller publishes the encrypted product to the buyer and the seller is not able to trace the identity of the buyer. Hence, during the entire transaction, the privacy of the buyer is protected against the seller. In the case of WCA, it only has the credentials of the buyer, but WCA is not aware of the product or digital content which the buyer has bought, and hence, the buyer is also protected against WCA.

Dispute resolution problem

In the dispute resolution problem, if the seller takes evidence to the judge, i.e. arbiter that the buyer is responsible for copyright violation. The seller does not know exactly where the watermark is embedded in the digital content, X. The seller is unable to frame the buyer. When the arbiter asks the buyer for the watermark, W, the buyer can send some random watermark $W_{\square'}$ instead of original, W. The seller has presented the Judge with a signed and encrypted copy of the watermark W and this watermark W will not match the watermark W′ presented by the buyer. Then, the buyer would be considered as the spurious buyer. For that, WCA finds the value of watermark W in place of watermark W′ with the help of Equation 3.3. WCA takes the final decision based on this equation.

$$(ID_S, t, X, W, X_W) = SD_{K_{wca}}(C)$$

Unbinding problem

The unbinding problem is solved because in this, first, the seller does not know the buyer's watermark W_B, because the watermark is embedded by a trusted third party, i.e. WCA under encryption algorithm. The buyer's signature binds to the ARG that uniquely identifies a particular digital content, X. These aspects make it impossible for the seller to transplant the watermark into another copy of the forged digital content.

When both buyer and seller argue to prove the ownership of a similar media *Z*, then the arbiter executes the watermark detecting and verifying protocol to specify the lawful possessor of the digital content *Z*. For determining the robustness of the underlying watermarking algorithm, we check the result of equation 3.6. If the digital signature, i.e. Ds is generated by the WCA then the equation no. (3.7) should be reliable.

$$D_{H(w_k)}(Z, X, W)$$

$$Ver_{PC_{wca}}(Ds) = (ID_S, t, W, H(w_k))$$

In the case of watermark embedding and signing protocol ciphertext C and digital signature Ds are transmitted between WCA and seller. As ciphertext C is encrypted using the public key of PC_{wca} of WCA, an unauthorised person cannot obtain the digital content X and X_W from the ciphertext C and digital signature Ds because the original digital content X and watermarked data X_W are kept secret in the watermark embedding and signing protocol. Hence, the buyer can obtain X_W only if after seller publicises the watermarked data, and then arbiter can determine whether that the seller is the rightful owner or not.

From the above analysis, our proposed protocol can solve the common problems, which are presented in Section 1 and design goals are also achieved, which are given in Section 3. Our protocol has come at some modification based on the previously published protocol. We did not embed the second watermark into the original digital content, the buyer needs to interact with the seller, arbiter and the WCA in the transaction process and the seller and WCA are used for issuing the valid watermark. Hence, the seller is unable to bind a watermark for framing the innocent buyer i.e. the unbinding problem is resolved and if any disputed occurs between buyer and seller, and then WCA and arbiter can solve that issue using time stamp based technique to establish the use of timestamp at what time the digital content or signal was created, signed or verified, i.e. dispute resolution problem is resolved.

EXPERIMENTAL RESULTS

All previously proposed buyer-seller watermarking protocol uses Cox (Cox et al., 1997) method to gain robust watermarking. In our proposed protocol, however, we adopted wavelet and principal component analysis based techniques (Kumar et al., 2015) with identity-based public key cryptography for achieving high robustness. Hence, we claimed the novelty of our proposed scheme as our protocol is more robust and has very high imperceptibility. We have presented various parameters for analysing the performance of the proposed protocol.

Peak Signal-To-Noise Ratio (PSNR)

Peak Signal-To-Noise Ratio is generally applied to analyse the quality of a picture.

$$PSNR = 10 * log \frac{255^2}{MSE} \tag{5.1}$$

Mean Square Error (MSE)

The MSE represents the cumulative squared error between the compressed image and the original image. In order to calculate the PSNR, first, the mean-squared error (MSE) is calculated using the following equation.

$$MSE = \sum_{i=1}^{x} \sum_{j=1}^{y} \frac{(|A_{i,j} - B_{i,j}|)^2}{x * y} \tag{5.2}$$

Where x is the width of the image and y is height, and $x * y$ is the number of pixels.

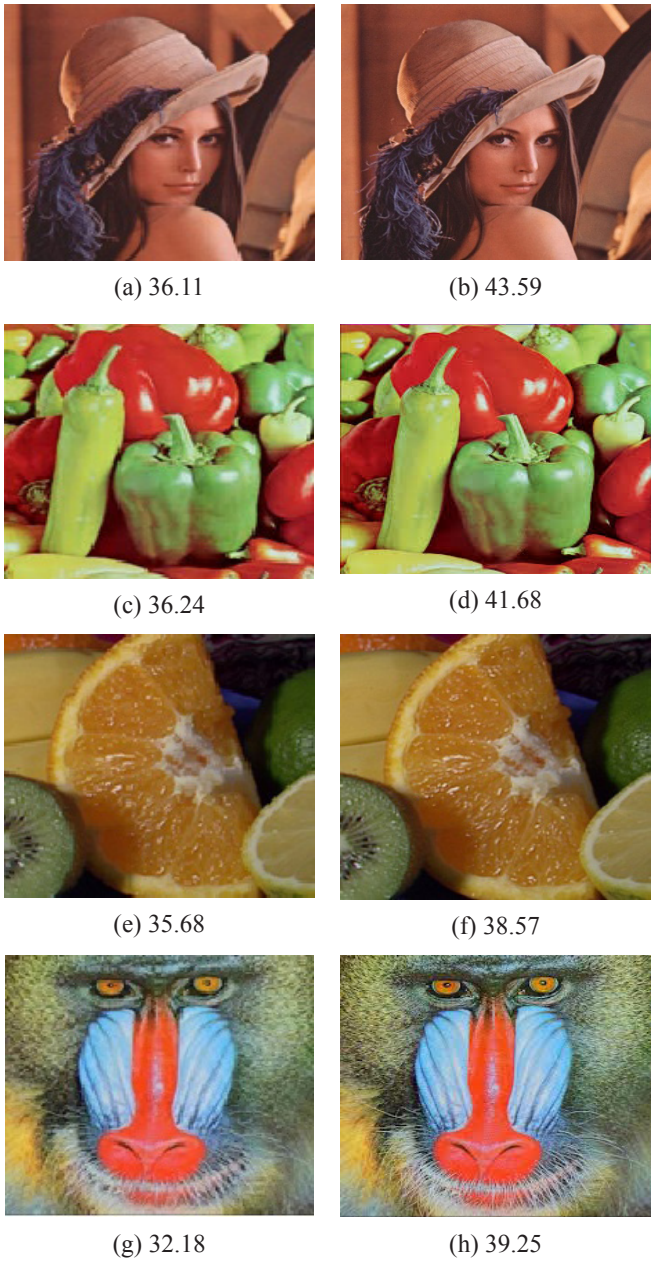
Normalized Correlation Coefficient (NCC)

It is used for calculating the robustness of the algorithm.

$$NC = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j} B_{i,j}}{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} \tag{5.3}$$

Where $A_{i,j}$ and $B_{i,j}$ denote the pixel values in row i and line j of the original watermark and the exacted watermark respectively.

The correctness of our proposed approach depends on the robustness of watermarking embedding and extracting scheme. In our scheme, we set α =0.01 i.e. watermark embedding coefficient factor. For instance, we have chosen Lena and Baboon images for producing our results. Figure 8 shows the original test images and watermarked test images. The various watermark logos, i.e. JNU logo and copyright logo, are shown in Figure 9. These watermark logos are embedded into the original images to prove the owner of the digital content. Some attacks are applied to the watermarked images for checking the robustness of the proposed scheme. The primary objective of our protocol is to solve the entire problem, which is solved by Zhang *et al*., as and as well dispute resolution problem and unbinding problem. The embedding method uses wavelets and principal component analysis technique (Kumar et al., 2015) with identity-based public cryptography for getting the watermark, while the existing protocol uses Cox's embedding method (Cox et al., 1997), which is based on DCT transform. In the previously proposed protocol, each element is processed independently, and thus, the computation cost and overhead increase linearly. In our proposed scheme, however, we adopted wavelet and principal component analysis based scheme (Kumar et al., 2015), which has only one asymmetric operation by the buyer or seller and two asymmetric operations by TTP as WCA. Consequently, communication overhead is almost constant.



(a) 36.11 (b) 43.59

(c) 36.24 (d) 41.68

(e) 35.68 (f) 38.57

(g) 32.18 (h) 39.25

Figure 7. The watermarked images on the left side are created by the buyer and right sides are created by the seller.

In the watermarking process, WCA generates the valid watermark images seller using wavelet and principal component analysis transform to create correlation coefficient, and buyer executes these correlation coefficients for generating the watermarked image. Figure 7 shows the watermarked images created by both the buyer and seller. It is clear that the PSNR value of Figure 7 on the left side is lower than the right side. The left side of Figure 7

shows the watermarked images generated by the buyer and the right side of the figure show watermarked images generated by the seller. Table 3 shows that the PSNR values corresponding to seller and buyer. Here, in order to calculate the robustness of the watermark embedding scheme, we have applied several types of attacks to the watermarked images.

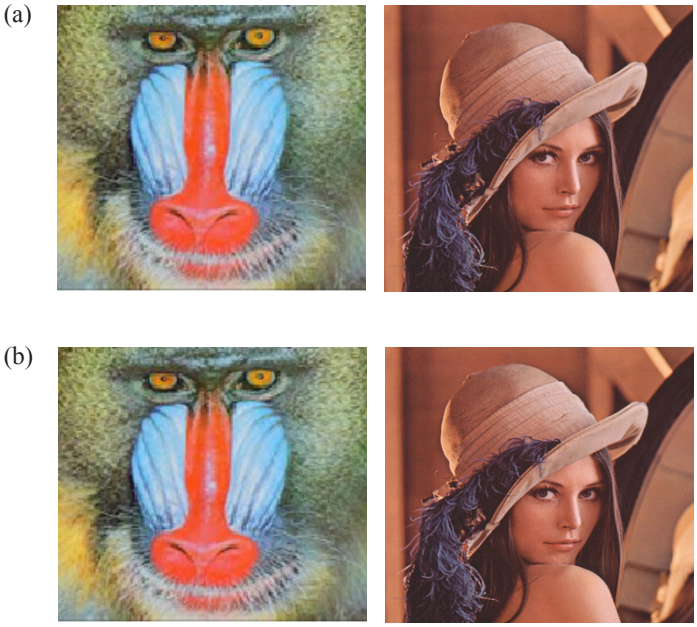
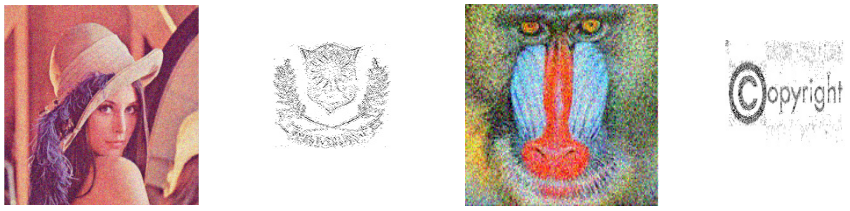


Figure 8. (a) Original Test Image; (b) Watermark Test Images

Table 3
Peak signal to noise ratio (PSNR) dB created by both buyer and seller for each original colour image.

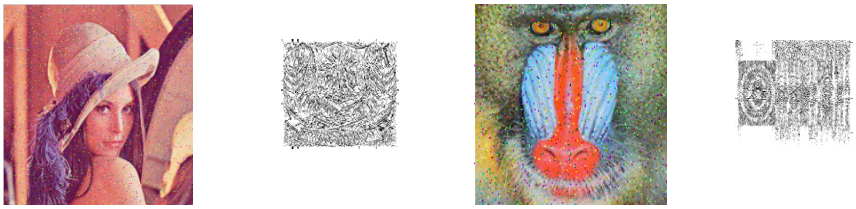
PSNR(dB)	Lena	Pepper	Fruit	Baboon
Buyer	36.11	36.24	35.68	32.18
Seller	43.59	41.68	38.57	39.25

Figure 9. Watermarks logos (a) JNU (b) Copyright (c) gray scale JNU (d) gray scale Copyright



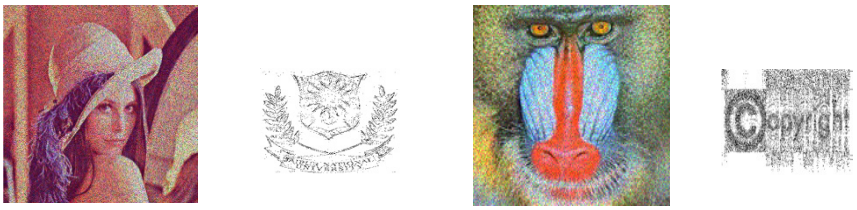
(a) (b) (c) (d)

Figure 10. (a) Watermarked Lena Image after Gaussian Noise at 0.02; (b) Extracted Watermark; (c) Watermarked Baboon Image after Gaussian Noise at 0.02; (d) Extracted Watermark.



(a) (b) (c) (d)

Figure 11. (a) Watermarked Lena Image after Salt & Pepper Noise at 0.02; (b) Extracted Watermark; (c) Watermarked Baboon Image after Salt & Pepper Noise at 0.02; (d) Extracted Watermark.

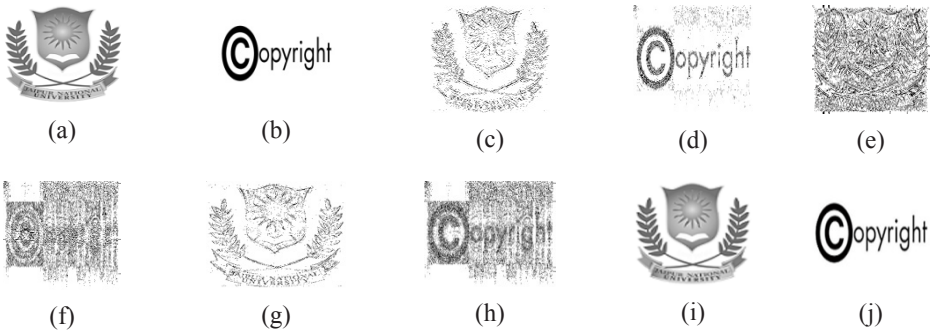


(a) (b) (c) (d)

Figure 12. (a) Watermarked Lena Image after Speckle Noise at 0.03; (b) Extracted Watermark; (c) Watermarked Baboon Image after Speckle Noise at 0.03; (d) Extracted Watermark.

(a) (b) (c) (d)

Figure 13. a) Watermarked Lena Image after Median Filter at [5 5] b) Extracted Watermark c) Watermarked Baboon Image after Median Filter at [5 5] d) Extracted Watermark.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 14. Extracted Watermark Images

In Figure 10, we have applied Gaussian noise with a density of 0.02 to the watermarked Lena and Baboon images, with the JNU watermark logo and copyright logo are embedded respectively. The quality of the extracted watermark logos is good with the presence of attacks. To check the quality of these watermarks, we have calculated correlation coefficient by using equation 5.3. Meanwhile, the corresponding correlation coefficients are shown in Table 4.

Figure 11 shows the performance of our scheme against the salt and pepper noise. We have applied this noise with a density of 0.02 on the Lena and baboon watermarked images and then extracted the corresponding watermarks.

In Figure 12, we have applied speckle noise with 0.03 density on the Lena and Baboon watermarked images, and then the corresponding watermark are extracted. In this, we got good results and the watermarks are still extracted.

Filtering is the most common attacks on digital images. Thus, we have applied the median filter to both the watermarked images with the filter size M=5 in Figure 13. The results show that watermarks are easily recognised. If we increase the filter, size normalised correlation will also decrease. Figure 14 (a,b) shows the original watermarks and (c, d, e, f, g, h, i, j) with the extracted watermarks from the Lena and Baboon images. In order to measure the quality of watermarked images, we have PSNR by equation 5.1. Figure 8(b) shows the watermarked images and their corresponding PSNR values are shown in Table 4.

Table 4

The PSNR values of all the test images.

Images	Lena	Pepper	Fruit	Baboon
PSNR	43.59	41.68	38.57	39.25

Table 5 shows the various results of our proposed scheme. We have successfully extracted the watermark from Gaussian noise, salt and pepper noise, Speckle noise and Median filter attacks. It is noticeable that in the case of Median filter and Gaussian noise, the performance of our scheme is quite impressive. Hence, our scheme is very robust against Salt and pepper noise and Median filter attack, and it also shows a better performance. Table 5 shows that the correlation coefficient values for the extracted watermark and PSNR values for the attacked watermark images. The imperceptability and robustness our watermark embedding scheme is very high.

Table 5
PSNR values and normalised corelation coefficient of all watermarked images and extracted logos after the attacks.

Images	**Lena**		**Baboon**	
Attacks	**PSNR**	**NC**	**PSNR**	**NC**
Gaussian Noise with Noise 0.02	40.62	0.8463	37.39	0.8131
Salt & pepper Noise with density 0.02	39.63	0.6338	35.10	0.5321
Speckle Noise with density 0.03	39.72	0.7842	37.61	0.8741
Median Filter with filter size [5, 5]	41.01	0.9762	36.49	0.7153

CONCLUSION

In this paper, we have presented an identity-based buyer-seller watermarking protocol which can solve the various problems of the previously published protocol and free from all known attacks. In addition, we made use of a tamper resistance device, which is embedded into seller's computer and reduces the overhead on WCA. WCA maintains its own table and keeps the requested IDs of both buyer and seller. Hence, WCA is required to participate in each transaction of the digital contents between buyer and seller. We also adopted wavelet and principal component analysis based techniques (Kumar et al., 2015) to increase the robustness and imperceptibility of our embedding scheme. The watermark certificate authority is responsible for issuing the digital signature corresponding to ID of the seller. If the problem of multiple ownership occurs, then it is the duty of an arbiter to decide on it. The arbiter checks the correctness of data used by the seller, and then the arbiter compares the timestamps for determining the original possessor of the digital content. These changes enable our proposed protocol to become really secure, feasible and efficient.

REFERENCES

Choi, J. G. Sakurai, K., & Park, J. H. (2003). Does it need trusted third party? Design of buyer-seller watermarking protocol without trusted third party. In *International Conference on Applied Cryptography and Network Security* (pp. 265-279). Springer Berlin Heidelberg.

Cox, I. J., Kilian, J., Leighton, T., & Shamoon, T. (1997). Secure spread spectrum watermarking for multimedia. *IEEE Trans. Image Process, 6*(12), 1673–1687.

Goi, B. M., & Phan, R. C. W. (2004). Cryptanalysis of two anonymous buyer seller watermarking protocols and an improvement for true anonymity. In *International Conference on Applied Cryptography and Network Security* (pp. 369-382). Springer Berlin Heidelberg.

Hu, D., & Li, Q. (2009). A secure and practical buyer-seller watermarking protocol. In *2009 International Conference on Multimedia Information Networking and Security* (Vol. 2, pp. 105-108). IEEE.

Hu, Y., & Zhang, J. (2009). A Secure and Efficient Buyer-Seller Watermarking Protocol, *Journal of Multimedia, 4*(3), 161-168.

Hwang, M. S., Hwang, K. F., & Chang, C. C. (2005). A time-stamping protocol for digital watermarking. *Applied Mathematics and Computation, 169*(2), 1276–1284.

Ju, H. S., Kim, H. J., Lee, D. H., & Lim, J. I. (2002). An anonymous buyer-seller watermarking protocol with anonymity control. In *International Conference on Information Security and Cryptology* (pp. 421-432). Springer Berlin Heidelberg.

Kumar, A., Tyagi, V., Ansari, M. D., & Kumar, K. (2011). A Practical Buyer-Seller Watermarking Protocol based on Discrete Wavelet Transform. *International Journal of Computer Applications, 21*(8), 46-51.

Kumar, A., Ansari, M. D., Ali, J., & Kumar, K. (2011). A New Buyer-Seller Watermarking Protocol with Discrete Cosine Transform. In *International Conference on Advances in Communication, Network, and Computing* (pp. 468-471). Springer Berlin Heidelberg.

Kumar, A., Ghrera, S. P., & Tyagi, V. (2015). Modified Buyer Seller Watermarking Protocol based on Discrete Wavelet Transform and Principal Component Analysis. *Indian Journal of Science and Technology, 8*(35), 1-9.

Lei, C. L., Yu, P. L., Tsai, P. L., & Chan, M. H. (2004). An Efficient Anonymous Buyer-Seller Watermarking Protocol. *IEEE Transactions on Image Processing, 13*(12), 1618– 1626.

Memon, N. D., & Wong, P. W. (2001). Secret and public key image watermarking schemes for image authentication and ownership verification. *IEEE Transactions on Image Processing, 10*(10), 1593–1601.

Memon, N. D., & Wong, P. W. (2001). A buyer-seller watermarking protocol. *IEEE Transactions on Image Processing, 10*(4)*,* 643–649.

Mintzer, F., & Braudaway, G. W. (1999). If one watermark is good, are more better? In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)* (pp. 2067–2069).

Paillier, P. (1999). Public key cryptosystems based on composite degree residuosity classes. In *International Conference on the Theory and Applications of Cryptographic Techniques* (pp. 223-238). Springer Berlin Heidelberg.

Rial, A., Deng, M., Bianchi., T., Piva, A., & Preneel, B. (2010). A Provably Secure Anonymous Buyer–Seller Watermarking Protocol. *IEEE Transactions on Information Forensics and Security, 5*(4), 920-931.

Shamir, A. (1985). Identity-based cryptosystems and signature schemes. In *Workshop on the Theory and Application of Cryptographic Techniques* (pp. 47-53). Springer Berlin Heidelberg.

Zeng, P., Cao, Z., & Choo, K. R. (2011). An ID-based digital watermarking protocol for copyright protection. *Computers and Electrical Engineering, 37*(4)*,* 526–531.

Zhang, J., Kou, W., & Fan, K. (2006). Secure Buyer-Seller Watermarking Protocol. *IEEE Proceedings of Information Security, 153*(3), 15–18.

Pertanika J. Sci. & Technol. 25 (1): 77 - 90 (2017)

SCIENCE & TECHNOLOGY

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

Modified Kohonen Network Algorithm for Selection of the Initial Centres of Gustafson-Kessel Algorithm in Credit Scoring

Sameer, F.[1, 2]* and Abu Bakar, M. R.[1]

[1]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

[2]*Faculty of Science, Universiti of Baghdad, 98798 Baghdad, Iraq*

ABSTRACT

Credit risk assessment has become an important topic in financial risk administration. Fuzzy clustering analysis has been applied in credit scoring. Gustafson-Kessel (GK) algorithm has been utilised to cluster creditworthy customers as against non-creditworthy ones. A good clustering analysis implemented by good Initial Centres of clusters should be selected. To overcome this problem of Gustafson-Kessel (GK) algorithm, we proposed a modified version of Kohonen Network (KN) algorithm to select the initial centres. Utilising similar degree between points to get similarity density, and then by means of maximum density points selecting; the modified Kohonen Network method generate clustering initial centres to get more reasonable clustering results. The comparative was conducted using three credit scoring datasets: Australian, German and Taiwan. Internal and external indexes of validity clustering are computed and the proposed method was found to have the best performance in these three data sets.

Keywords: Credit Scoring, Decision-making, Clustering Techniques, Fuzzy Clustering Algorithms, Gustafson-Kessel Algorithm, Kohonen Network

INTRODUCTION

Banks' databases contain information about their customers and the financial history of their payments. The databases can be utilised to assess the credit risk by investigating whether it can be a good basis on which to predict borrowers' ability to repay their loans on time.

The credit-scoring technique is commonly used to evaluate the creditworthiness of credit clients. The credit risk evaluation system plays an important role in decision making to enable faster decisions for credit, lessen the possible risk and reduce the cost

Article history:
Received: 30 December 2015
Accepted: 04 July 2016

E-mail addresses:
fadfhaa55@gmail.com (Sameer, F.),
drrizam@gmail.com (Abu Bakar, M. R.)
*Corresponding Author

ISSN: 0128-7680

of credit analysis. The 2008 financial crisis revealed the importance of credit risk evaluation decisions, not only for financial institutions and banks, but also for both the global and local economy (Wu, 2008).

The credit models built with a credit risk evaluation technique ought to fulfil two essential criteria: precision, which means that they are capable of predicting the behaviour of customers, and transparency, which means that the model is able to describe the input-output relationship in an understandable way. Credit scoring classifies credit applicants as 'bad' or 'good' customers by considering features like age, monthly income, and marital status (Yang, 2007). Statistical methods have been used most oftenly for assessing credit risk for customers. Logistic regression and linear discriminate analysis are most commonly used (Thomas, 2000). Accuracy of credit scoring for several neural networks was investigated (West, 2000). Results were benchmarked against traditional statistical methods like linear discriminant analysis, logistic regression, k-nearest neighbour, and decision trees. Clustering techniques provide distinct new options for measurable routines for building credit scoring models. Recently, more pragmatic approaches have been adopted, and several classification techniques have appeared to perform well for credit scoring. Since the introduction of Fuzzy logic by Zadeh in 1965, it has successfully been implemented in many fields like credit scoring. Clustering is a technique that is used in data analysis. This method is used to find groups in a data set such that there are the most similarities in each group and the most dissimilarity between different groups. Gustafson–Kessel (GK) algorithm is one of fuzzy clustering techniques used in credit scoring. Gustafson–Kessel (GK) algorithm also has its drawbacks in relation to choosing initial centres (Gustafson & Kessel, 1979). In order to overcome the sensitivity of the initial point choice and increase the accuracy of credit decisions. Another method was used to obtain better initial centres.

In this paper, we modified the kohonen algorithm for selecting the centres of clusters of Gustafson-Kessel algorithm. The paper is organised as follows: In section 2, the Clustering analysis techniques and Self-Organising Map are described. Section 3 describes the modified kohonen algorithm, while Sections 4 provides a brief overview of the measures of cluster validity and Section 5 presents the experimental analysis and results. Discussion and conclusion are given in Sections 6 and 7, respectively.

MATERIAL AND METHODS

Clustering Analysis Techniques

Clustering is used to assign a set of objects to groups (called clusters) and the objects in the same cluster are more similar than to the objects in other clusters. Based on the similarities between the objects, cluster analysis is the classification and the organisation of the objects into groups (Gan, Ma, & Wu, 2007). Clustering partition methods can be fuzzy or hard. Hard clustering partition methods are based on the classical set theory, which requires an object to belong to only one cluster. Fuzzy clustering partition methods allow objects to belong to many clusters simultaneously, with different degrees of membership (Touzi, 2010).

The Gustafson – Kessel Algorithm

The GK algorithm is a powerful fuzzy clustering technique, with a large number of applications, such as image processing and classification systems. The importance of this algorithm lies in its ability to estimate the cluster covariance matrix to adapt the distance metric to the shape of the cluster (Gustafson & Kessel, 1979).

The GK algorithm needs a set of N samples in the n dimensional space and the number of clusters as the input parameters. A fuzzy partition of the data set X can be represented by an $(N * k)$ matrix $\boldsymbol{U} = [u_{ij}]$, where u_{ij} denotes the degree of membership, with which the i^{th} object belongs to the j^{th} cluster, where $(1 \leq i \leq N)$ and $(1 \leq j \leq k)$. $\boldsymbol{U}$ is the fuzzy partition matrix, and it must satisfy the following constraints:

$$\begin{aligned} &\bullet \quad 0 \leq u_{ij} \leq 1 \qquad \text{for } i \in \{1,\ldots,N\},\ j \in \{1,\ldots,k\}. \\ &0 < \sum_{i=1}^{N} u_{ij} < N \quad \text{for } j \in \{1,\ldots,k\} \\ &\bullet \quad \sum_{j=1}^{k} u_{ij} = 1 \qquad \text{for } i \in \{1,2,\ldots,N\}. \end{aligned} \tag{1}$$

The final constraint expresses that the sum of the memberships of an object over the all set of clusters must be equal to 1. The number of clusters is at least two. The objective function of the *GK* algorithm is defined as follows:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{k} {u_{ij}}^m d_{ij}^2 \tag{2}$$

where

$$d_{ij}^2 = (x_{i-}c_j)\, A_j\, (x_i - c_j)^T \tag{3}$$

$C=(c_1,c_2,\ldots..,c_k)$ represents the cluster centre in R^n, m represents the fuzziness exponential, where $1 \leq m < \infty$, $d_{ij} = d(c_j, x_i)$, is the distance between the centre c_j and the data point x_i, and u_{ij} is the degree of membership of point x_i in the j^{th} cluster. A_j is the Mahalanobis distance matrix for the j^{th} cluster; if A_j is the identity matrix, then the square Euclidean distance measure is obtained. A_j is defined as follows:

$$A_j = V_j[det(F_j)]^{\frac{1}{n}} F_j^{-1} \tag{4}$$

Here n is the number of attributes or features, V_j is the volume of the j^{th} cluster and F_j is the cluster covariance matrix from the following formula:

$$F_j = \frac{\sum_{j=1}^{k} (u_{ij})^m (x_i - c_j)(x_i - c_j)^T}{\sum_{j=1}^{k} (u_{ij})^m} \tag{5}$$

In the GK algorithm, the cluster shape changes depending on the data, and can be in an ellipsoidal form or a hyper ellipsoidal form. For this reason, the GK algorithm employs the covariance matrix.

The steps of the GK algorithm are as follows:

1. Given a dataset $X=\{x_1, x_2, ..., x_N\}$.

2. Select the k which represent the number of clusters, ($2 \leq k \leq N$), and select the termination condition $\epsilon > 0$.

3. Choose the initial centre c_j from the dataset.

4. Compute the cluster covariance matrix using the formula in equation (5).

5. Compute the Mahalanobis distance using equation (3).

6. Compute the partition matrix (u_{ij}), as follows:

$$u_{ij} = \frac{1}{\sum_{r=1}^{k} (\frac{d_{ij}}{drj})^{\frac{2}{m-1}}} \tag{6}$$

where, i=1, 2, …., N, and j=1, 2, …., k

7. Update the C-means matrix (c_j), as follows:

$$c_j = \frac{\sum_{i=1}^{N} (u_{ij})^m x_i}{\sum_{i=1}^{N} (u_{ij})^m} \quad for\ 1 \leq j \leq k \tag{7}$$

8. Repeat the above steps until the centre matrix for two sequential iterations is stable, in the following sense:

$$\|C^{I+1} - C^I\| < \epsilon \tag{8}$$

where (I) represents the number of iterations.

Self-Organising Map

Networks with supervised training techniques are networks with a target output for every input pattern. The networks learn how to produce the outputs that are required; in unsupervised training, however, the networks learn to form classifications of the training data without external supervision (Yin, 2008). When input patterns and features are shared, the network is able to identify those features across the input patterns. An unsupervised system is based on competitive learning, in which the neurons that are the output compete among themselves to be activated. Only one neuron is activated at any one time, and this activated neuron is called the winning neuron (Kohonen et al., 2009). This competition can be implemented if the neurons have

parallel inhibition connections (negative criticism ways) among them; therefore, the neurons learn to organize themselves. This neural network is called a self-organising map (SOM) and projects high-dimensional data onto a low-dimensional grid (Kohonen, 1982).

Kohonen Network

A Kohonen network is a classification method that forms the basis of self-organising maps (SOM) (Kohonen, 1982). The Kohonen network method, which was proposed by Kohonen, has a single computational layer arranged in rows and columns, and is a feed-forward structure. Every neuron is connected to all the nodes in the input layer or source (Kohonen, 1998). The Kohonen algorithm is an unsupervised classified network and it deals with inputs that are unable to be led without overlapping in the classes. It is robust (that is, it is able to resist noise); therefore, the Kohonen algorithm has interesting properties.

MODIFIED KOHONEN ALGORITHM

The GK algorithm chooses points as initial clustering centres randomly, and different points may lead to different solutions. In order to overcome the sensitivity of the initial point choice, we modified the Kohonen Network method to obtain better initial centres. The steps of the algorithm are as follows:

1. Compute the similar neighbourhood of each point x of data set X; it is denoted by *simneighbor (x, r)*, take x as the centre and r is the threshold value of degree similarity. The objects with large degree of (r) are in similar neighbourhood of point x.

$$Simneighbor(x,r)=\left\{\begin{matrix} x_i | r \leq sim_{x_i \epsilon X}(x_i,x) \leq 1,0 \leq r \leq 1, \\ X=\{x_1,\dots,x_n\} \end{matrix}\right\} \quad (9)$$

where $sim_{x_i \epsilon X}(x_i,x)$ is the similarity degree between x_i and x can be denoted by the formula

$$sim_{x_i \epsilon X}(x_i,x) = \lambda d_m(x_i,x) + (1-\lambda) cos(x_i,x) \quad (10)$$

and $d_{m_{x_i,x \epsilon X}}(x_i,x) = \frac{d_{x_i,x \epsilon X}(x_i,x) - min_{x_i,x \epsilon X}\{d(x_i,x)\}}{max_{x_i,x \epsilon X}\{d(x_i,x)\} - min_{x_i,x \epsilon X}\{d(x_i,x)\}}$ is normalised Euclidean distance (Zhang, 2013). The symbol $cos\ (x_i,x)$ is the cosine of the intersection angle of two points and can be computed by the formula:

$\frac{\sum_{i=1}^{p}(x_i x_j)}{\sqrt{\sum_{i=1}^{p} x_i^2}\sqrt{\sum_{i=1}^{p} x_j^2}}$ and its value ranging from -1 to 1.

2. Compute the similarity density of each point and it is denoted by Density (x_i) where x_i belong to *X*. The formula of density is as

$$Density_{x_i \in X}(x_i) = \frac{\sum_{i=1}^{|p_{neighbor(x_i)}|} sim(x_i, p_{neighbor(x_i)})}{|p_{neighbor(x_i)}|} \quad ,$$
$$X = \{x_i, \ldots, x_n\} \tag{11}$$

The symbol $p_{neighbor(x_i)}$ denoted the points that satisfy the threshold r in similar neighbourhood of x_i.
The $|p_{neighbor(x_i)}|$ is the number of points in the neighbourhood of x_i.

3. Choose the points that have high means of max density points as the weights of kohonen algorithm. And input the number of nodes (k, which equals the number of weights) and let I=1 represent the time or number of iterations.

4. Compute the distance to nodes by determining the Euclidean distance d_j between the input data point and each weight:

$$d_j = \sum_{i=1}^{N-1}(x_i - x_j), \text{ for } j = 1,.., K \text{ ; } i = 1,\ldots,N \tag{12}$$

where x_j points that have high density in dataset *X*.

5. Select the winning node j^* that produces the minimum d_j and update the weights at iteration I for node j^* and its neighbours:

$$x_j(\mathrm{I}+1) = x_j(\mathrm{I}) + \eta(\mathrm{I})\left(x_i - x_j\right) \tag{13}$$

where $\eta(\mathrm{I})$ is the learning rate parameter, with the initial learning rate parameter being set (usually to a figure between 0.2 and 0.5). The learning rate is initialised at 0.5, and will decrease at each iteration by the following expression:

$$\eta(\mathrm{I}+1) = 0.5\eta(I) \tag{14}$$

The nodes in the neighbourhood of j^* become more similar to the input vector x_i, after these updates.
After optimising the modified Kohonen Network method of selection the initial cluster centres, the Gustafson–Kessel Algorithm begins with these centres clustering analysis, as follows:

6. Input the centres retrieved from the modified Kohonen Network method as the initial centres of the clusters.

7. Compute the cluster covariance matrices using equation (5).

8. Compute the distances using equation (3).

9. Compute the partition matrix using equation (6).

10. Update the C-means matrix (c_j) from the following expression:

$$c_j = \frac{\sum_{i=1}^{N}(u_{ij})^m x_i}{\sum_{i=1}^{N}(u_{ij})^m}$$

11. Repeat the above steps (6 to 10) until the centres matrix for two sequential iterations (I,I+1) are stable in the following sense:

$$\|C^{I+1} - C^I\| < \epsilon$$

At this step, the partition matrix gated by fuzzy clustering is used to classify the creditworthiness of credit clients.

MEASURES OF CLUSTER VALIDITY

The clustering algorithm always seeks to find the best fit for a fixed number of clusters and the shapes of the parameterised cluster. Cluster validity refers to whether a given fuzzy partition fits the data at all. The number of clusters is application-specific and is usually identified by a user. Cluster validity criteria are applied to determine the optimal number of clusters and a good clustering algorithm (Wang & Zhang, 2007). Although there are many cluster validity measures that can be used for this purpose, none is perfect. There are mainly two types of validity measures:

- External measures: using the class label for cluster analysis.
- Internal measures: use the vectors for analysis.

Internal Measures

Several internal indices are used simultaneously and the most important ones are described below:

Partition Coefficient (PC). This measures the amount of "overlapping" between clusters. Bezdek (Bezdek, 1981; Pal & Bezdek, 1995) defines it as follows:

$$PC(k) = \frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{N} \quad (12)$$

Here, u_{ij} is the membership degree of the i^{th} data point to the j^{th} cluster. The best algorithm for partitioning the data is the one that produces the highest value of PC.

Classifications Entropy (CE) (Pal & Bezdek, 1995). This measures only the fuzziness of the clusters partition, so it has similarity to the Partitions Coefficient.

$$CE(k) = -\frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{N} u_{ij}\log(u_{ij}) \quad (13)$$

The best clustering algorithm is the one with the lowest value for this index.

Partitions Index (SC) (Bensaid, 1996). This is the ratio between the sum of the separation and the compactness of the clusters. It is the sum of the cluster validity measures for each individual divided by the fuzzy cardinality for each cluster.

$$SC(k) = \sum_{j=1}^{k} \frac{\sum_{i=1}^{N} (u_{ij})^m ||x_i - c_j||^2}{\sum_{i=1}^{N} u_{ij} \sum_{d=1}^{k} ||c_d - c_j||^2} \quad (14)$$

When comparing different partitions with an equal number of clusters, SC is useful. A better partition can be obtained by a lower value of SC.

Separation Index (S) (Bensaid, 1996). In contrast to the partition index (SC), the separation index uses a minimum-distance separation for partition validity, and a lower value of S indicates a better partition.

$$S(k) = \sum_{j=1}^{k} \frac{\sum_{h=1}^{N} (u_{ij})^2 ||x_h - c_j||^2}{N min_{i \neq j, i=1,..,k} ||c_i - c_j||^2} \quad (15)$$

Xie and Beni's Index (XB) (Xie & Beni, 1991). This aims to measure the proportion between the total variation within clusters and the separation of clusters, and it is defined as follows:

$$XB(k) = \sum_{j=1}^{k} \frac{\sum_{i=1}^{N} (u_{ij})^m ||x_i - c_j||^2}{N min_{ij} ||x_i - c_j||^2} \quad (16)$$

This index focuses on separation and compactness properties. The clusters are more separated if Index (*XB*) has a smaller value.

Dunn's Index (DI) (Xie & Beni, 1991). This index aims to recognise dense and well-separated clusters. It is defined as the proportion between the minimal intra-cluster distance and the maximal inter-cluster distance. For each cluster partition, this index can be identified as follows:

$$DI(k) = min_{j \in k} \left\{ min_{j \in k i \neq j} \left\{ \frac{min_{x \in k_i, y \in k_j} d(x,y)}{max_{x,y \in k} d(x,y)} \right\} \right\} \quad (17)$$

A high Dunn's index indictes a desirable algorithm for producing clusters.

Davies-Bouldin index (DB) (Davies & Bouldin, 1979). This index can be identified as follows:

$$DB = \frac{1}{n} \sum_{i=1}^{n} max_{i \neq j} \left(\frac{d_i + d_j}{d(c_i, c_j)} \right) \quad (18)$$

Here n is the number of clusters, c_j and c_i are the centres of cluster, while d_j *and* d_i are the average distances of all elements in clusters j and i respectively, and $d(c_i,c_j)$ is the distance between the centres c_i and c_j. The best algorithm is the clustering algorithm that produces a collection of clusters with the smallest DB index.

External Measures

These methods give an indication of the quality of the resulting partitioning and thus they can only be considered as a tool at the disposal of the experts in order to evaluate the clustering results. Fuzzy Rand Index is a well-known measure of similarity between two partitions of a data set (Hullermeier, 2012). Given a fuzzy partition P = {P1, P2 ,…, P_k} of X, each element $x \in X$ can be characterised by its membership vector.

$$P(x) = (P_1(x), P_2(x) , ..., P_k(x)) \in [0; 1]^k \quad (19)$$

where $P_i(x)$ is the degree of membership of x in the *i-th* cluster P_i. A similarity measure on the associated membership vectors (19) can be formed as:

$$E_P(x, x') = 1 - ||P(x) - P(x')|| \quad (20)$$

Where, ||.|| is a proper metric on $[0; 1]^k$ if two fuzzy partitions P and Q are given. To generalise the concept of concordance, a pair (x; x') is defined and the degree of concordance is:

$$conc(x, x') = 1 - ||E_P(x, x') - E_Q(x, x')|| \in [0\ 1] \quad (21)$$

the degree of discordance is:

$$disc(x, x') = ||E_P(x, x') - E_Q(x, x')|| \quad (22)$$

the distance measure on fuzzy partitions is then defined by the normalised sum of degrees of discordance:

$$d(P,Q) = \frac{\sum_{(x,x')\epsilon X} ||EP(x,x') - EQ(x,x')||}{N(N-1)/2} \quad (23)$$

Likewise,

$$RE(P,Q) = 1 - d(P,Q) \quad (24)$$

Corresponding to the normalised degree of concordance, and therefore, it is a direct generalisation of the original Rand index. Rand index is a similarity measure which assumes values between 0 and 1. If near 1 means that the *i-th* cluster in P and the *i-th* cluster in Q are identical, thus P=Q.

EXPERIMENTAL ANALYSIS AND RESULTS

Data Sample

We run experiments on three real-life data sets: Australian credit data, German credit data and Taiwan credit data (UCI machine learning databases). The Australian credit data are composed of 690 entries, of which 307 match creditworthy clients and 383 match bad clients. Each entry is described by 14 features or attributes, 6 of which are continuous, while the remainders are

categorical. The German credit data consist of 1000 entries, 70% of which correspond to clients who are a good credit risk and 30% of which relate to clients to whom credit should be refused or who are bad credit risk. Each client is described by 20 features, including personal information such as age, sex and marital status, existing accounts, credit history records, loan amount and purpose and employment status. Seven features are continuous, and the rest are categorical. Taiwan data are composed of 30000 entries, with 23 features and of which 23364 match creditworthy clients and 6636 match bad clients. Table 1 shows the characteristics of the datasets.

Table 1
Characteristics of the data sets used in the experiment

Dataset	Number of features	Number of data	Number of groups
Australian	14	690	2
Germany	20	1000	2
Taiwan	23	30000	2

Implementation

The work was implemented using MATLAB version R2010a by creating a programme to perform the GK algorithm, GK+MK) algorithm and modified kohonen algorithm. Tables 2, 3 and 4 show the results for the objective function, number of iterations, as well as internal indexes and external indexes.

Table 2
The validity measures of Australian credit data

Algorithm / validity measures	*Gustafson-Kessel(GK)*	Gustafson-Kessel with Kohonen Network method(GK+K)	Gustafson-Kessel with modified Kohonen Network method (GK+MK)
PC	0.8277	0.8487	0.8959
CE	0.3897	0.3269	0.3247
SC	2.9343	1.8472	1.7238
S	1.7605	1.5121	1.2952
XB	1.5017	0.9084	0.3133
DI	1.4938e-005	1.4938e-005	1.2282
DB	1.4972	1.4972	0.8140
***J* (objective Function)**	94.2459	90.3452	80.3230
No. iteration	31	25	20

Table 3
The validity measures of German credit data

Algorithm / validity measures	Gustafson-Kessel (GK)	Gustafson-Kessel with Kohonen Network method (GK+K)	Gustafson-Kessel with modified Kohonen method (GK+MK)
PC	0.5698	0.8007	0.8854
CE	0.6084	0.6374	0.5394
SC	0.5694	0.4421	0.3869
S	0.2952	0.2480	0.2039
XB	0.1721	0.1678	0.1216
DI	0.2741	0.2489	0.1234
DB	0.7487	0.7387	0.6387
***J*(objective Function)**	417.3321	316.3481	250.3240
No. iteration	19	15	12

Table 4
The validity measures of Taiwan credit data

Algorithm / validity measures	Gustafson-Kessel (GK)	Gustafson-Kessel with Kohonen Network method (GK+K)	Gustafson-Kessel modified with Kohonen Network method (GK+MK)
PC	0.5000	0.5035	0.7432
CE	0.6931	0.6894	0.5437
SC	6.1690	5.8680	1.8760
S	8.1324	1.1483	0.5630
XB	1.5423	1.5212	0.5481
DI	0.0043	0.0045	0.0033
DB	0.8709	0.8709	0.7540
***J*(objective Function)**	155	153	140
No. iteration	10	6	5

Table 5
The fuzzy rand validity measure of three credit data

Algorithm / data	Gustafson-Kessel (GK)	Gustafson-Kessel with Kohonen Network method (GK+K)	Gustafson-Kessel modified with Kohonen Network method (GK+MK)
Australian	0.8600	0.8803	0.9410
Germany	0.7888	0.8406	0.9032
Taiwan	0.8476	0.8530	0.8942

DISCUSSION

As shown in the summarised results, the modified Kohonen method with Gustafson-Kessel algorithm (GK+MK) has the best performance for the three data sets, since it has a smaller distance function (objective function) and a lower number of iterations for the three data sets.

Table (2) shows the internal index values for the Australian data for GK, GK+K and GK+MK. The value of the first index (the partitions index, PC) for the GK+MK algorithm is near to 1, which is greater than its value for the GK and GKK. The value of the second index (the Classification Entropy, CE) for the GK+MK algorithm is less than the value for the GK and GKK. It can be seen that the values of the other indexes (SC, S, XB) for the GK+MK algorithm are lower than the values for the GK algorithm, but DI and DB for two algorithms are equal, which mean the algorithms well-separated clusters. The objective function GK+MK algorithm is less than the value for the GK and GK+K Algorithms. The number of iterations is 31 for the GK Algorithm, which is greater than the iterations of *GKK* algorithm 25.

In Table 3, the index values of the German data are showed. The partitions index value (PC) for the new GK+MK algorithm is greater than the value of GK and GKK algorithm, which means the GK+MK algorithm is the best. The Classification Entropy, CE in GK+MK algorithm is less than the value in GK and GKK algorithm, which means the fuzziness of the clusters partition for GK+MK algorithm is less than its coordinate for the GK and GKK algorithm. The other indexes (SC, S, and XB) for the GK+MK algorithm are lower than the values for the GK and GKK algorithm.

DI and DB for the two algorithms are unequal, which means the GK+MK algorithms are well-separated clusters than the GK and GKK algorithm. In addition, the objective function and number of iterations for GK+MK algorithm are less than the value of the two algorithms.

Table 4 shows the internal index values for the Taiwan data for the algorithms. The values of the two indexes (the partitions index, PC and Dunn's index DI) for the new GK+MK algorithm are high, while the values of the other indexes (SC, S, XB , DB) are low. The results indicate that the new GK+MK algorithm is the best.

Table 5 shows the fuzzy rand validity measure of three credit data. The results show that the values of GK+MK are greater than values of the two other method fuzzy partitions are robust.

CONCLUSION

In this paper, a new modified Kohonen method to centres selection of fuzzy clustering was proposed. Developing and improving the GK algorithm to identify the centres of clusters, the three algorithms were applied to three datasets. A comparative study among the algorithms was carried out.

The cluster internal validity indexes confirmed that the performance of the proposed algorithm (GK+MK) is better than that of the GK and GKK algorithms. A fuzzy validity index is applied in this paper for evaluating the fitness of clustering to data sets.

ACKNOWLEDGEMENT

We are thankful to God who has given us the power, good sense and confidence to complete this research analysis successfully, and I am grateful to my research supervisor and to the Department of Mathematics for support, valuable suggestions and guidance. I also would like to express gratitude to the editor and to others who have made suggestions and given help with comments to improve the paper.

REFERENCES

Bensaid, A. (1996). Validity–guided (Re) Clustering with Application to Image Segmentation. *IEEE Transactions on fuzzy Systems, 4*(2)*,* 112-123.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function*. New York: Plenum Press.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224-227.

Gan, G., Ma, C., & Wu, J. (2007). *Data Clustering: Theory, Algorithm and Application* (Vol. 20). Society for Industrial and Applied Mathematics (SIAM).

Gustafson, D. E., & Kessel, W. (1979). Fuzzy clustering with a Fuzzy Covariance Matrix. In *Proceeding of 55th IEEE Conference on Decision and Control* (pp. 761–766). Las Vegas, USA.

Hullermeier, E., Rifqi, M., Henzgen, S., & Senge. (2012). Comparing Fuzzy Partitions: A Generalization of the Rand Index and Related Measures. *IEEE Transactions on Fuzzy Systems, 20*(3), 546-555.

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics, 43*(1), 59–69.

Kohonen, T. (1998). *Self Organization and associative memory* (2nd Ed.). Springer Series in Information Sciences. Berlin: Springer.

Kohonen, T., Nieminen, I., & Honkela, T. (2009). On the quantization error in SOM vs. VQ: a critical and systematic study. In *International Workshop on Self-Organizing Maps* (pp. 133-144). Springer Berlin Heidelberg.

Lahsasna, A., Ainon, R. N., & Wah, T. Y. (2010). Credit scoring models using soft computing methods: A survey. *The International Arab Journal of Information Technology, 7*(2), 115-123.

Pal, N. R. & Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model, IEEE Trans. *Fuzzy System, 3*(3), 370–379.

Thomas, L. C. (2000). A survey of credit and behavioral scoring: Forecasting financial risks of lending to customers. *International Journal of Forecasting, 16*(2), 149-172.

Touzi, A. G. (2010). An Alternative Extension of the FCM Algorithm for Clustering Fuzzy Databases. In *Advances in Databases Knowledge and Data Applications (DBKDA), 2010 Second International Conference on* (pp. 135-142). IEEE.

Wang, W. N., & Zhang, Y. J. (2007). On fuzzy cluster validity indices. *Fuzzy sets and systems, 158*(19), 2095–2117.

West, D. (2000). Neural network credit scoring models. *Computers and Operations Research, 27*(11–12), 1131–1152.

Wu, X. (2008, June 29). *Credit Scoring Model Validation*. (Master Disertation). Faculty of Science, Korteweg-de Vries Institute for Mathematics, Universiti Van Amsterdam.

Xie, X. L. & Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13*(4), 841-846**.**

Yang, Y. (2007). Adaptive credit scoring with kernel learning methods. *European Journal of Operational Research, 183*(3), 1521–1536.

Yin, H. (2008). The self-organizing maps: background, theories, extensions and applications. In *Computational intelligence: A compendium* (pp. 715-762). Springer Berlin Heidelberg.

Zhang, Y., & Cheng, E. (2013). *An Optimized Method for Selection of the Initial Centres of K-Means Clustering* (pp. 149-156). Springer-Verlag Berlin Heidelberg.

Pertanika J. Sci. & Technol. 25 (1): 91 - 100 (2017)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Wireless Sensor Network and Internet of Things (IoT) Solution in Agriculture

Zulkifli, C. Z.* and Noor, N. N.

Department of Computing, Faculty of Art, Computing and Creative Industry, Sultan Idris Education University, 35900 Tanjung Malim, Perak, Malaysia

ABSTRACT

This paper presents the technology of Active Radio Frequency Identification (RFID) and Wireless Mesh Sensor Network (WMSN) that will be used in agriculture. In this paper, ZigBee technology platform is applied in 2.45 GHz and active RFID to sustain the WSN by developing a fully automated IoT solution in agriculture for irrigation system. The system includes a plurality of sensor nodes installed in a crop field sending an ID, which are embedded sensor and WSN that work on ZigBee 2.4 GHz platform. The ID was sent to act as a signal of soil in dry condition of a specific area to a reader at base station. The pump stations will use information from base station to sprinkling water in the specific area of the dry state automatically. The automatic control system is very practical in agriculture but most of it is based on schedule and timer regardless of soil condition and temperature. Therefore, wireless automated irrigation system for efficient water use and production is proposed.

Keywords: WSN, RFID, ZigBee, Agriculture, Sensor

INTRODUCTION

Agriculture is an industry that uses a lot of water throughout the world. This resource should be used in an efficient way without affecting the production (Jiber et al., 2011). The obstacles in measuring and monitoring water usage and inefficient irrigation systems due to human control are the main contributors to this situation. The farmers are aware that water shortage or over watering may damage the yield. They need to understand when and the amount of water is needed for specific crops (Jiber et al., 2011).

Most farmers have little knowledge of their farm and they are unaware of the methods to improve their productivity of agricultural practices. All these conflicts make it necessary to think of resolve support systems for agriculture (Jiber et al., 2011). In

Article history:
Received: 12 January 2016
Accepted: 17 June 2016

E-mail addresses:
chezalina@fskik.upsi.edu.my (Zulkifli, C. Z.),
nssa3900@gmail.com (Noor, N. N.)
*Corresponding Author

ISSN: 0128-7680

order to overcome this problem, Wireless Mesh Sensor Network (WMSN) and active Radio Frequency Identification (RFID) for agriculture monitoring control are applied. In this study, we proposed an automated irrigation system with full real-time remote monitoring and control system in the farm. The system replaces human-to-human (H2H) and human-to-machine (H2M) to machine-to-machine (M2M) architecture, which is embedded with active RFID. It has moisture sensor and monitoring devices that are required for the farming data such as soil moisture and condition.

WMSN combines the reliability of hardwiring with the versatility of wireless networking in spite of having to compromise the speed. In addition, WMSN consists of cost efficient, battery-powered sensor modules and embedded networking intelligence (Zulkifli Et al., 2011). The communication that hinders optimal production output ZigBee is a growing technology that will create advantages in the agriculture industry (Zulkifli Et al., 2011).

The development WMSN application in agriculture gives it potential to increase efficiency, productivity and profitability while decreasing unintended effects on crops and the environment in agriculture production (Ruiz-Garcia et al., 2009). The real time information from the irrigation area will contribute a solid base for farmers to change consideration at any time rather of taking decisions based some assumed average condition (Ruiz-Garcia et al., 2009).

MATERIAL AND METHOD

Radio Frequency Identification (RFID)

RFID is one of an operative automatic identification technology for different things. The ultimate function of RFID is the capability to trace the position of the tagged things. RFID Technologies composes of tags, reader and computer, which acts a host and comes in all shapes, sizes and read ranges. It is also thin, flexible and can penetrate between paper and plastic. The tag has an identification number and a memory that stores data such as manufacturer, product type and environmental data such as temperature, humidity of an object (Mustafa et al., 2013). In the RFID applications, the tags are attached into objects that are to be tracked (Mustafa et al., 2013).

RFID is the most utilised in the real-time locating system in agriculture applications. It becomes a choice for farmers due to its low cost. RFID tags come in two forms, active and passive. In this, in spite of using the same RFID technology, they are dissimilar in many forms. In this system, active RFIDs are used to send ID that works on Zigbee platform to readers at the base station. It can be seen that active tags are controlled by a battery formed into the tag, which allows data to be transmitted over long distances compared to passive RFID. The read and write distances are much longer than for passive tags. The active RFID has a small battery built–in to the tag, which works as an internal power source. The batteries can sometimes be replaceable or the unit will be replaced after certain period of time, which is normally between 1 year and 7 years.

Active tags can operate at higher frequencies such as 455MHz, 2.45 GHz, and 5.8 GHz. The active RFID broadcasts by itself. The passive RFIDs are of low cost and low range,

while relics on the reader supply the energy to power the tag. The read range is limited and it is difficult to read through metal or liquid. By comparing these two active and passive tags, in this research we provided active RFID to use in the real-time irrigation monitoring system.

The RFID systems operate in low frequency, high frequency, and ultra-high frequency. Frequency relates to the radio wave sizes is used to transmit between RFID systems components. The Low Frequency (LF) band covers frequencies from 30 KHz to 300 KHz. Generally, the LF RFID systems operate at 125 KHz or 134 KHz. This frequency band provides a short read range less than 0.5m, and has slower read speed than the higher frequencies, yet is not very sensitive to radio wave interference. While the High Frequency (HF) band ranges are from 3 to 30 MHz, most HF RFID systems operate at 13.56 MHz with read range up to 1.5 m (Lee et al., 2009).

RFID is not only used for human to machine or machine to human, but the requirement for to machine communication has expended. Thus, the attribute is suitable to be used for monitoring agriculture environment. In the irrigation system, monitored embedded system comes into a new platform for farmers to spend their energy, money and time, which will take place only when there is a requirement of water. In this proposed system, an active RFID based on 2.4 GHz Zigbee Platform is used to send ID to the reader to recognise the node that sends data for irrigation and fertilisation processes without human intervention.

ZigBee Technology

ZigBee, which was originated in 1998, is based on the IEEE 802.15.4 standard and pioneered by ZigBee Alliance, which was formed by several companies interested in defining low cost, low power, and wireless network standard (Lee et al., 2009). ZigBee can support large number of nodes providing a low cost global network. The IEEE defines only the PHY and MAC layers in its standard, while ZigBee defines the network and application layers, application profile and security mechanism. Due to this design, the consumption of power is minimal and the battery life span is longer.

Zigbee supports three topologies, which are star, mesh and cluster-tree, as shown in Figure 1. In star topology, each end node is connected to the coordinator and communication is carried out by the Zigbee Coordinator (ZC). In mesh topology, each device communicates with any other device within its radio range or through multi-hop. In cluster tree topology, there is a single routing path between any devices (Kalaivani et al., 2011). In the Zigbee application, it is mostly used for mesh topology. In spite of that, for the proposed system monitoring mesh topology was chosen. The various sense data from moisture sensors go to WMSN, that integrates with RFID tag and sprinkler will turn into a node. On the farm, there are plenty of nodes and each node will communicate through this ZigBee technology platform. Based on that, the reader will read the sensor data and stores the data to the server, which is used by a farmer for monitoring. In the proposed system, field monitoring uses 2.4 GHz operating frequency nodes for the purpose of study.

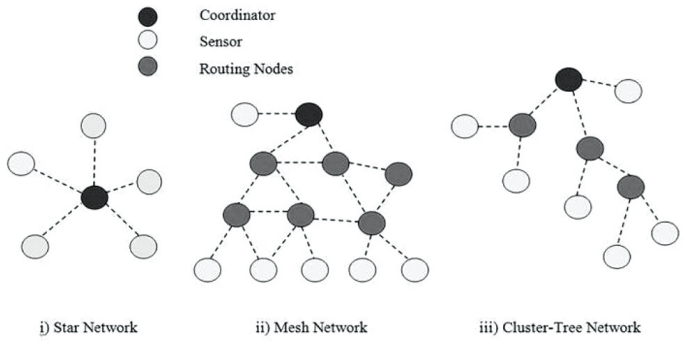


Figure 1. ZigBee Topologies [9]

Wireless Sensor Network (WSN)

Wireless network refers to the technology to communicate and access the internet without cable connection between computers and other electronic devices. Sensor Network has contributed to several applications, and awareness has expended to implement the technology into the agriculture environment. WSN is one of the most important technologies in the 21st century (Mendez & Mukhopadhyay, 2013). WSN is an assembly of a number of low-power, low-cost, multipurpose sensor nodes communicating wireless upon a short distance (Sazak et al., 2013).

The difference between a WSN and a RFID system is that RFID devices have no cooperative capabilities, while WSN allows different network topologies and multihop communication (Ruiz-Garcia et al., 2009). WSN can cut down the effort and time needed for monitoring environment (Mendez & Mukhopadhyay, 2013). As a result, money, water and labour costs can be reduced. The technology allows for remote measurements such as temperature, humidity, soil moisture and water level (Mendez & Mukhopadhyay, 2013). There seems to be increased development towards wireless outcomes in comparison to wired-based systems. Figure 2 shows the concept of wireless monitoring that is to be applied in the agriculture environment.

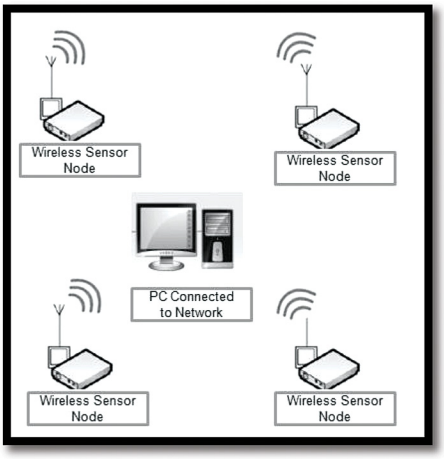


Figure 2. The Concept of Wireless Monitoring

This systems provides a full network coverage in large facilities such as a big farm, typically massive lengths of cabling that leads to remarkable return on investment. WSN provides an intelligent platform to gather and collect data from the sensor nodes that can detect and interact with the physical environment (Zulkifli et al., 2011). Using a wireless mesh as a backbone network simplifies installation and provides an affordable medium (Zulkifli et al., 2011). WSN can be used to identify soil moisture to admit the irrigation system and identify where and when to irrigate. It helps in maximising crop yield and elevates profit (Mendez & Mukhopadhyay, 2013). Irrigation system requires temperatures, water level, humidity sensors and moisture sensor; in this proposed system, however, moisture sensor and temperature sensor are used to detect soil moisture.

Moisture Sensor

Measuring the soil moisture is important in agriculture to help farmers to conduct their irrigation systems effectively. For this reason, farmers are able to use less water to irrigate crop as it is able to grow yields and the quality of the crop. Moisture sensor can read the amount of moisture nearby in the soil surrounding it properly. In this proposed system, sensor node and sprinkler will be attached together. Table 1 shows the functionality of the sensor. When a sensor detects low water level in the soil, sprinklers will supply more water. If the sensor detects excess water in the soil, sprinklers will supply less water. More water is needed when the sensor is dry and this causes the soil to conduct electricity easily (less resistance), while dry soil conducts electricity poorly (more resistance). Embedding the technology with moisture sensor can save and reduce water consumption. Using the moisture sensor, water does not need to function or irrigate when the sensor has the right amount of water.

Table 1
How the sensor works

Item	Condition	Min	Typical	Max	Amount of water
Output	Sensor in dry soil	0%	~	30%	High water
	Sensor in humid soil	30%	~	70%	Medium water
	Sensor in water	70%	~	85%	Low water
	Sensor in water	85%	~	95%	No water

IOT – THE PROPOSED REMOTE MONITORING SYSTEM

The remote monitoring systems are promoting IoT solution working on WSN embedded with RFID technology. The system communicates with hardware and software automatically to send data in the farm. The solution is proven and can therefore be implemented from planting to harvest as a tool for appropriate irrigation tactic to improve crop yields. Besides, the WSN nodes can effectively collect data as well. Remote monitoring for irrigation and fertilising using WSN and RFID can ensure a good quality crop yield. In spite of the stressful environmental conditions, it increases the application efficiency of irrigation systems by 50%. The collaboration has been made with local farmer company that runs the herbaceous plants on a farm located in Ipoh, Perak. This collaboration facilitates research and development of this project, while helping the company to increase productivity and reduce operating costs.

In this system, automatic irrigation systems are developed in the farm to collect the data from moisture sensors placed in the field. The farm will be monitored through the wireless sensor network that is integrated with the active RFID at the field. WSN will sense and monitor the environment like soil moisture and temperature. The coverage area for the experiment is 10 acres, in which 20 nodes are required in this Roselle farm. The systems proposed are very intelligent where the node always sleeps in standby mode. If the sensor senses soil in dry, the node will be activated to work in the mesh network between the other nodes to send ID to the reader. The end device of active RFID shown in Figure 3 is embedded with the sensor that represents wireless network sensor ID that works on Zigbee 2.4 GHz platform. The ID sent to the reader at the base station is used to recognise and allocate which nodes are sending data to the irrigation process automatically.

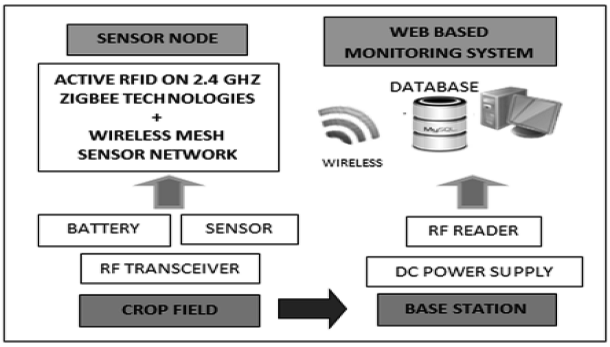


Figure 3. The reader at the base station receives ID from the sensor node

The full concept of the system shown in Figure 4 are active RFID on 2.4 GHz Zigbee platform and moisture sensor are embedded together to become one sensor mesh node. The moisture sensor collects data from the soil, which will be processed before sending via wireless to the controller for further action. The sprinkler will supply water based on the condition of soil. The data that are processed will be sent through the computer for monitoring by the farmers. The farmers can monitor their farm anywhere using internet connectivity by phone or computer. All the systems in the farm are connected to each other via wireless. The messy cabling like conventional method is not used anymore because it will be disturbing an irrigation process.

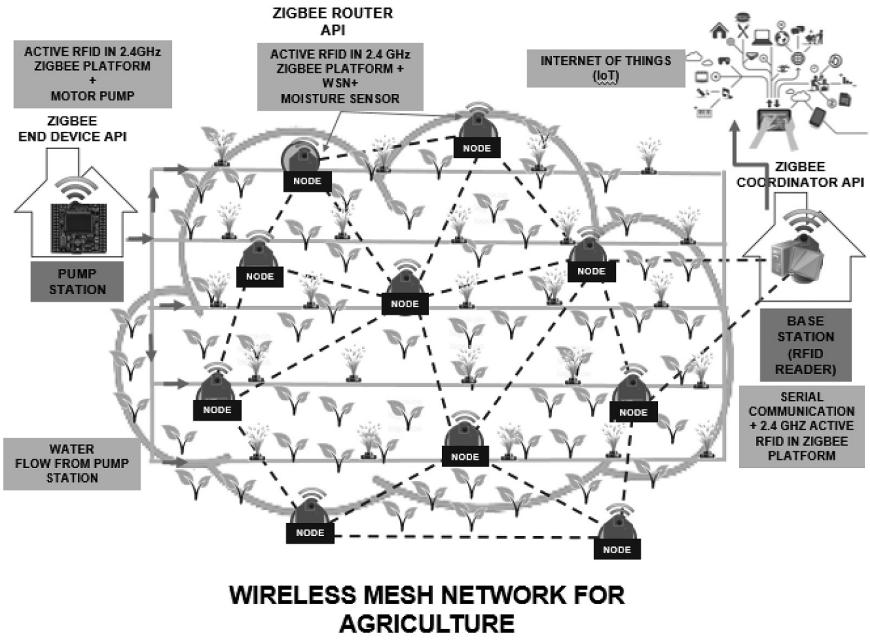


Figure 4. The concept of the system

RESULTS AND DISCUSSION

Figure 5 illustrates the data that will be shown in monitoring base station. A comparison was made within five months at two different areas; one using automatic irrigation and the other using manual irrigation. From the data collected, it can be concluded that using the proposed system on the farm has its benefits. Water usage can be reduced approximately up to 50% when the embedded technology is used compared to the conventional method. In this system, the sprinkler will supply water when the moisture sensors give a signal with the right amount. Sensor in range 0-30% makes the sprinkler supply a large volume of water because the soil is in a dry state. Therefore, it needs a 100% amount of water. Meanwhile, when the sensor is in the range of 30-70%, the sprinkler will reduce water intake by 50% and supply an average volume of water to the soil. This range saves a whole sum of water. The sprinkler will stop the water supply when the moisture sensor sends data of about 85-95%. In this condition, the soil is wet so there is no need for water to be supplied. Thus, farmers can reduce water consumption. The conventional method uses the same amount of water when it needs to irrigate every day. Over irrigation can cause the death of plants and production of farm to be affected badly. In particular, this can affect the revenue of farmers as well since water is wasted and over irrigation may cause damages to the plants. Besides that, the irrigation processes need a number of workers for the conventional method, as it is time consuming.

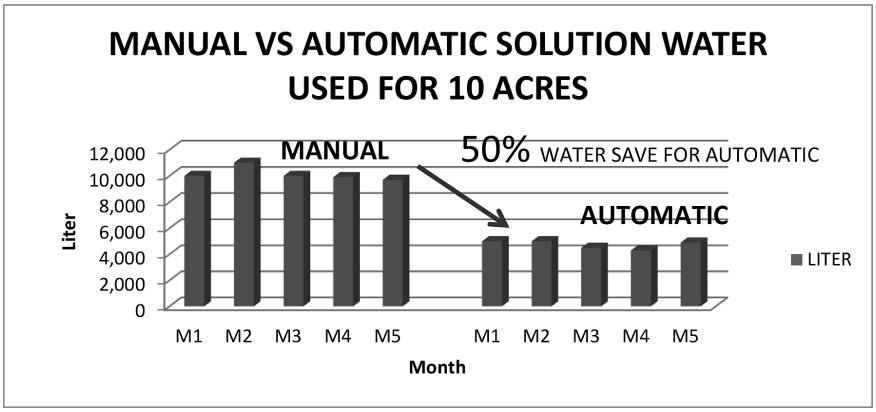


Figure 5. Data collected from the wireless sensor node

WSN has a capability to represent inherent soil variability, which is present in the fields with more accuracy than the currently available system. Thus, the benefit for the farmer is a real-time support system that allows maximising their productivity while saving water. WSN also removes the obstacle to wire sensor stations over the field and decreases the maintenance cost (Zulkifli et al., 2011). The advancement of WSN applications can affect agriculture by increasing profitability, productivity and efficiencies, while minimising unconscious impacts to the environment in many agricultural production systems (Ruiz-Garcia et al., 2009). The real-time information from the fields will afford a solid base for farmers to plan strategies at any time. Instead of taking decisions based on some hypothetical average condition, a precision farming approach recognises differences and adjusts management actions accordingly (Ruiz-Garcia et al., 2009).

CONCLUSION

By introducing RFID Technology and WMSN in the farming industry, growing crops and plants can be greatly optimised. WMSN reduces the wiring and piping costs, and facilitates installation and maintenance in large areas. The use of technology in agriculture is important, particularly to increase production apart from decreasing labour cost and water requirements. Thus, the WMSN technology obviously performs the most technology to improve the current irrigation systems. Soil moisture sensors are constantly improving and becoming affordable and appropriate for massive deployment in the WMSN applications.

ACKNOWLEDGEMENTS

The authors would like to thank to Herbal Development Office (HDO), Ministry of Agriculture and Agro-Based Industry (NKEA Research Grant Scheme-NRGS) for sponsoring this work.

REFERENCES

IJECSCSE. (2015). *International Journal of Electronics, Communication and Soft Computing Science and Engineering*. Retrieved June 11, 2015, from http://www.ijecscse.org/

IMPINJ. (n.d). *The Different Types of RFID Systems*. Retrieved October 11, 2015, from http://www.impinj.com/resources/about-rfid/the-different-types-of-rfid-systems/

Jiber, Y., Harroud, H., & Karmouch, A. (2011). Precision agriculture monitoring framework based on WSN. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International* (pp. 2015-2020). IEEE.

Jose, D., & Gutierrez, A. (2005). IEE Std. 820.15.4 Enabling Pervasive Wireless Sensor Networks. In *Innovation Centre* (pp. 2-54).

Kalaivani, T., Allirani, A., & Priya, P. (2011). A survey on Zigbee based wireless sensor networks in agriculture. In *Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on* (pp. 85-89). IEEE.

Lee, J. S., Chuang, C. C., & Shen, C. C. (2009). Applications of short-range wireless technologies to industrial automation: a ZigBee approach. In *Telecommunications, 2009. AICT'09. Fifth Advanced International Conference* (pp. 15-20). IEEE.

Mendez, G. R., & Mukhopadhyay, S. C. (2013). A Wi-Fi Based Smart Wireless Sensor Network for an Agricultural Environment. In *Wireless Sensor Networks and Ecological Monitoring* (pp. 247-268). Berlin Heidelberg: Springer.

Mendez, G. R., & Mukhopadhyay, S. C. (2013). A Wi-Fi Based Smart Wireless Sensor Network for an Agricultural Environment. In *Wireless Sensor Networks and Ecological Monitoring* (pp. 247-268). Springer Berlin Heidelberg.

Mustafa, M. Y., Eilertsen, S. M., Hansen, I., Pettersen, E., & Kronen, A. (2013). Matching mother and calf reindeer using wireless sensor networks. In *Computer Science and Information Technology (CSIT), 2013 5th International Conference on* (pp. 99-105). IEEE.

Ruiz-Garcia, L., Lunadei, L., Barreiro, P., & Robla, I. (2009). A review of wireless sensor technologies and applications in agriculture and food industry: state of the art and current trends. *Sensors*, *9*(6), 4728-4750.

Ruiz-Garcia, L., Lunadei, L., Barreiro, P., & Robla, I. (2009). A review of wireless sensor technologies and applications in agriculture and food industry: state of the art and current trends. *Sensors*, *9*(6), 4728-4750.

Sazak, N., Erturk, I., & Ukaya, E. K. Ö. (2013). An event-driven WSN MAC protocol design based on active node and dynamic time slot allocation. *Turkish Journal of Electrical Engineering and Computer Science*, *21*(3), 812-824.

Zulkifli, C. Z., Ismail, W., & Rahman, M. G. (2011). Implementation of Embedded Active RFID with Wireless Mesh Sensor Network for Industrial Automation. *Electronics World*, *117*(1908), 28-36.

Pertanika J. Sci. & Technol. 25 (1): 101 - 120 (2017)

SCIENCE & TECHNOLOGY

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

A Comparative Study of the Behaviour of Treated and Untreated Tyre Crumb Mortar with Oil Palm Fruit Fibre Addition

Farah N. A. Abd. Aziz*, Sani M. B., Noor Azline M. N. and Jaafar M. S.

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

ABSTRACT

An incorporation of waste tyre particles in concrete has been established to produce a green concrete. However, despite its advantages, strength reduction is an obvious handicap. To improve the strength, pre-treatments of the waste tyre particles and addition of Oil Palm Fruit Fibre (OPFF) were chosen and reported in this study. The addition of OPFF was to influence the internal structure in order to improve shrinkage and other strength properties. Performance of the composites in compressive, split tensile and flexural strengths, as well as shrinkage and microstructure were observed. Results showed better behaviour of the treated tyre crumb mortar rather compared to the untreated tyre, with the replacement of up to 40% by volume of the treated tyre crumb particles and 0.5- 1.0% OPFF addition by mass of cement content.

Keywords: Lightweight aggregates, treated, untreated, tyre crumb, oil palm fruit fibre, mechanical properties, unrestrained shrinkage

INTRODUCTION

Waste tyre material is regarded as one of the world's most persisting problems and needs to be utilised or recycled properly (Pedro et al., 2013). Several research studies have been on-going for the past decade to find a sustainable method of disposing of the waste tyres that are accumulating globally (Huang et al., 2004). The non-bio-degradable nature of the material has posed a serious challenge when establishing disposal sites and has thus become an environmental problem. Waste tyres comprise a significant portion of solid waste management activities which is a global environmental challenge. In many parts of the world, tyres are being burned to extract some important materials such as steel. In some cases, they are being burned as fuel, which could be harmful to human life and the environment at large (Brown et al.,

Article history:
Received: 06 January 2016
Accepted: 26 April 2016

E-mail addresses:
farah@upm.edu.my (Farah N. A. Abd. Aziz),
informsani@yahoo.com(Sani M. B.),
nazline@upm.edu.my (Noor Azline M. N.),
msj@upm.edu.my (Jaafar M. S.)
*Corresponding Author

ISSN: 0128-7680

2001). Recycling offers a promising avenue and the incorporation of this material in concrete could be an important solution due to the size of the construction industry. The construction industry is one of the largest avenues for recycling waste tyre materials, and hence, should be given appropriate attention.

Laboratory investigations have indicated that the substitution of waste tyre particles in concretes and mortars considerably increases their impact resistance, the toughness, plastic deformation, energy absorption, and fracture criteria of concrete. These offer many advantages when used in sound/crash barriers, retaining structures and pavements (Eldin & Senouci, 1993; Topcu, 1995; Lee et al., 1998; Khatib Bayomy, 1999; Taha, 2008). However, previous investigations have shown a considerable decrease in concrete strength properties when waste tyre particles are incorporated (Topcu 1995; Chung & Hong, 1995; Segre & Joekes, 2000; Kosievith et al., 2001; Li et al., 2004; Aeillo & Leuzzi, 2010; Chou et al., 2010). Chou et al. (2010) reported that the reduction of compressive and tensile strengths of the waste tyre incorporated concrete is an important problem to be solved. Some efforts have been made to improve some of these properties by adding other materials in concrete and mortars such as glass, ashes, bricks, polymers, silica fume, etc. (Nehdi & Khan 2001; Cairns & Kenny, 2004; Sánchez et al., 2008; Bektas et al., 2009; Turki et al., 2009; Correia et al., 2010; Silva et al., 2010; Wang & Huang 2010; Neno et al., 2011; Ahmad et al., 2012; Braga et al., 2012).

Onuaguluchi and Panesar (2014) investigated the mechanical and durability properties of concrete composites containing limestone powder pre-coated crumb tyre at 0%, 5%, 10% and 15% of fine aggregate by volumetric replacement. Also, silica fume at 15% cement replacement by volume was added to the matrix. It was observed that the strength improvement due to the coating of crumb tyre alone was negligible; instead silica replacement improves the mechanical properties of the mixtures significantly. Eldin and Senouci (1993), among the pioneers in this research area, reported that the compressive and tensile strengths of the composites decrease with the increase in size of the waste tyre aggregate. Moreover, the replacement of aggregate with coarse tyre aggregate was found to record lower workability. This is supported by Huang et al. (2004), who reported that a smaller diameter of waste tyre aggregate increases the mechanical strength of the composites and that stiffer coarse tyre aggregate produced from truck tyres could also add to the strength. Similarly, Benazzouk et al. (2007) reported that particles of waste tyre of less than 1 mm could significantly increase the strain capacity and toughness, reduce water absorption and decrease the sorptivity-value of the tyre infused concrete.

Also, the strength losses could be reduced by proper surface treatment of the waste tyre particles to improve the bond strength between the waste tyre particles and the paste (Li et al., 1998). The surface treatment should roughen the tyre particles to provide a better bond with the matrix which may yield higher strength properties (Segre & Joekes, 2000). Chou et al. (2010) reported that the surface treatment of waste tyre surfaces achieved with waste organic sulphur compounds is an easy and economical approach to improve strength properties. Apart from this, partial oxidation of the crumb tyre surfaces can also be used to improve the mechanical properties of waste tyre mortar (Chou et al., 2010b).

Raghavan et al. (1998) are amongst the few researchers to have carried out an investigation into the shrinkage of waste tyre concretes. The effect of incorporating two different shapes of rubber particles (granules 2 mm in diameter, and shredded of two sizes of 5.5 mm ×1.2 mm

and 10.8 mm×1.8 mm) into cement mortar have been studied. It was discovered that plastic shrinkage cracking was reduced in all types of tyre incorporated samples in comparison to the control specimen, and the control specimen developed a crack earlier than the waste tyre incorporated sample at 5% by mass. The average crack widths were 0.9 mm and 0.4-0.6 mm for the control and waste tyre incorporated samples, respectively. It was also observed that the initial time of cracking started after 30 minutes for the control, while that of the 15% waste tyre was delayed until after an hour. Moreover, the higher the rubber shreds content, the smaller the width and length of crack.

Jingfu et al. (2009) investigated the shrinkage and strength behaviour of roller compacted concrete containing a waste tyre additive by studying the crack resistance behaviour of the roller compacted waste tyre concrete specimens. The drying shrinkage of roller compacted rubberised concrete was observed to be greater than that of the control sample, and the sample containing different amounts of tyre rubber exhibited similar shrinkage behaviour. The rate of drying shrinkage in the first month was found to be higher than in the later ages of the samples. The waste tyre concrete containing 50 kg/m^3 and the control without rubber were observed to have a similar shrinkage pattern, whereas the shrinkage increased with waste tyre contents of 100 and 120 kg/m^3. It was concluded that the drying shrinkage of waste tyre concrete does not reduce by increasing the tyre content, which is contrary to the observation made by Raghavan et al. (1998).

Most efforts made to improve the strength properties of waste tyre incorporated concretes and mortars involved the use of chemicals and additives. The use of the chemical could add to the overall cost of the resulting products due to its high cost. Natural fibres, mainly of oil palm fruit fibre (OPFF), have shown beneficial potentialities in concretes at lower dosage Aziz et al. (2014). However, it has not been utilised in waste tyre incorporated concretes and mortars despite its abundance at little or no cost.

Ismail and Hashim (2008) investigated the properties of concrete by using oil palm fruit fibre (OPFF) at 0.25%, and 0.5% of fibre by mass of cement, and the length of the fibre was 10, 30 and 50 mm. It was observed that the optimum fibre length is 30 to 50 mm, which increased the compressive strength by 39% when compared with the control and above 1% fibre lowered the strength. Ahmad et al. (2010) studied the influence of oil palm trunk fibre (OPTF) in concrete by addition of OPTF at 0%, 1%, 2% and 3%, water cement ratio of 0.5 and cement sand and aggregate ratio of 1:1.47:3.0. It was observed that the slump decreased with the increase in fibre content. The compressive and split tensile strength increased by 30% and 25% respectively at 1% fibre content, which decreased below the control when the fibre content increased beyond 1%. It was also observed that the resistance to attack by NaOH and NaCl was higher at 1% OPTF, and at 3%, it showed higher resistance to HCl attack.

An addition of waste tyre particles in concrete and mortars have been reported to yield low mechanical properties. Therefore, pre-treatment and additives could be used to enhance the mechanical performance of the specimens. Hence, this research was carried out to compare the performance of the pre-treated and untreated tyre crumb incorporated mortars both with the addition of oil palm fruit fibre (OPFF) in the specimens on the compression, split tensile and flexural strength, as well as the shrinkage and its microstructure. The mortar composite studied is a potential material for use in applications such as masonry and precast lightweight concrete walls.

MATERIAL PROPERTIES

The mortar composite was made from fine aggregate, tyre crumb, cement, OPFF and water. Composite Portland cement Type II - conforming to ASTM C150 (2001) - was used in all mixes. The fine aggregate was obtained from a quarry in Selangor, Malaysia, with properties that are in accordance with ASTM C 33 and C 127. The specific gravity, water absorption, density and fineness modulus of the fine aggregates are 2.63, 3%, 1702 kg/m^3 and 0.9, respectively.

The crumb tyre aggregate, with a maximum aggregate size of 2.36 mm, was supplied by Arayaja Enterprise Sdn. Bhd., Malaysia. The tyre crumb was graded in the same particle distribution as the fine aggregate. The untreated tyre crumb was supplied and graded from the factory, while the treated tyre crumb went through the cement powder pre-treatment, as adopted by Li et al. (1998). Pre-treatment was carried out by soaking the tyre crumb in water for 48 hours, and then surface dried (SSD) before mixing it with cement powder Type II Portland cement. The compacted density and fineness modulus values of the untreated tyre crumb aggregate are 589 kg/m^3 and 0.9, respectively, while the treated tyre crumb values are 668 kg/m^3 and 0.9, respectively. Figures 1(a) and (b) show the untreated and treated tyre crumb aggregates used in this work, respectively.

Natural fibre from an oil palm fruit bunch was supplied by the palm oil extraction factory Seri Ulu Langat Palm Oil Mill Sdn. Bhd. at Dengkil, Malaysia. The oil palm fruit fibre (OPFF) was washed to clean it from oil, dirt and fungi. Then, the OPFF were dried at room temperature and cut into 30-50 mm lengths.

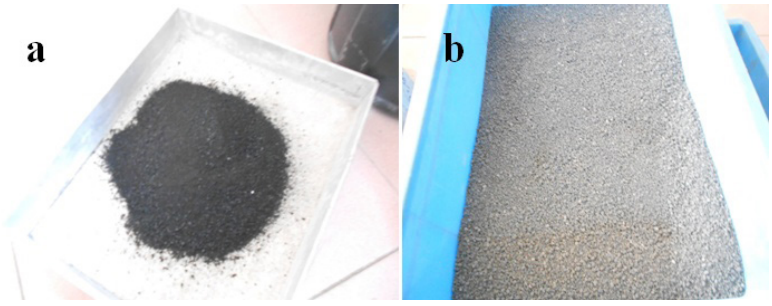


Figure 1. Tyre crumb aggregates (a) untreated; (b) treated

MIX PROPORTIONS

In line with ASTM C 109 (2005), the mortar mixes were prepared using a water-cement ratio of 0.485 and cement-to-sand ratio of 1: 2.75. The total water contents in all mixes include 3% moisture from the aggregates. The replacement of fine aggregates with tyre crumb was made at 10%, 20%, 30% and 40% by volume, and the addition of OPFF was made at 0.5%, 1% and 1.5% by weight of cement. The mixing was carried out in a bowl mixer with a maximum capacity of 0.005 m^3. The process was started by mixing the cement, aggregate and tyre crumb aggregates thoroughly for about 2 mins. Then, 1/3 of the water was added and allowed to mix for 2 more minutes before the remaining water was added for the final 2-min

mixing operation. The first part of the water was added to enable the fibre to loosen and mix with the other constituents. A total of 32 mix designs were prepared with a targeted compressive strength of 17 MPa.

EXPERIMENTAL PROGRAMME

The tests were carried out to compare the performance of the treated and untreated tyre crumb materials under similar conditions. In all cases, the averages of three specimens were used. The samples made of mortars containing crumb tyre of 0-40% by volume and oil palm fruit fibre of 0-1.5% by weight of cement were tested in this research work. The workability test for the various mixes was measured in accordance with ASTM C1437 (2007). The compression and density/water absorption tests were performed using 50 mm mortar cubes in accordance with ASTM C109 (2005) and ASTM C642 (1997), respectively. In total, 288 cubes were tested at 3, 7, and 28-days. A three-point flexural test was carried out in accordance with ASTM C348 (2008). In total, 64 specimens, each with a prism size of 40 x 40 x 160 mm, were tested at a loading rate of 5 mm/m. The split tensile test was carried out on 128 cylindrical samples of 100 x 200 mm size at 7 and 28 days respectively as in ASTM C496/496M (2004).

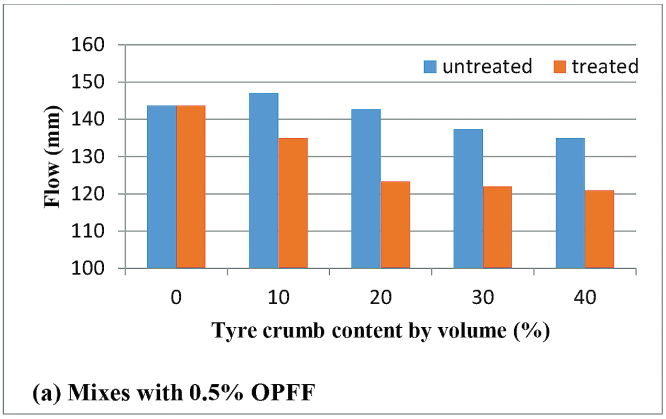
The unrestrained shrinkage test was conducted on a specimen size of 70 x 70 x 280 mm in accordance with the ASTM C 341/341M (2006) specification. A total of 64 specimens were produced; 32 were observed in the laboratory at room temperature, while the remaining specimens were kept in a climate chamber at 23^0C and 50% humidity for 90 days. The change in length was measured using an extensometer every day for the first 21 days and subsequently three times a week for 39 days and finally at the 90th day.

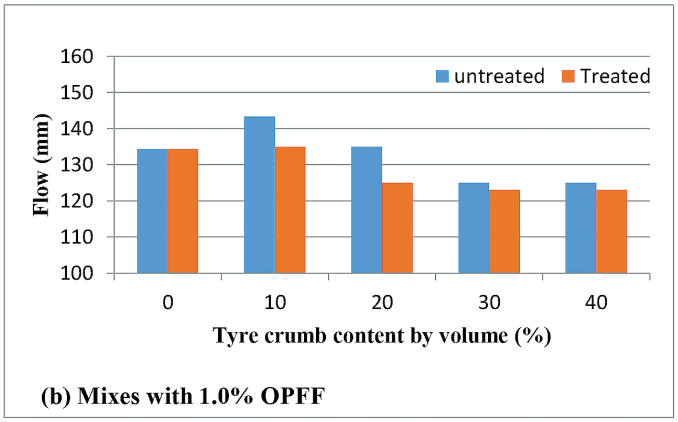
Scanning Electron Microscopy (SEM) analysis was carried out to examine the microstructure of the mortar samples using a Hitachi 5.00kv machine. A mortar sample size of 40 x 40 x 10 mm was used. A total of 5 samples were taken from the 10% and 40% tyre crumb contents for the treated and untreated tyre crumb mortars at 0.5% oil palm fruit fibres and the control specimen. The upper and lower limits of 10 and 40% of the tyre crumb contents were considered for comparison and the 0.5% OPFF was chosen due to its overall performance in the composite.

RESULTS AND DISCUSSION

Workability

The workability of the treated and untreated mortar mixtures are shown in terms of flow values as in Figure 2(a), (b) and (c) for 0.5%, 1% and 1.5% OPFF, respectively. In all cases, the workability flow achieved is more than the required workability for the mortar of 110 mm. In general, increases in tyre crumb content will reduce the workability of the mixtures, regardless of whether they are treated or untreated. However, the losses for the treated tyre crumb mixtures are higher than for untreated mixtures because the coating is able to absorb more water. On the contrary, the untreated tyre crumb particles repel water, hence increasing workability. The addition of OPFF into the mixes further decreased the workability for both the untreated and treated tyre crumb mortars because of its higher absorption capacity.





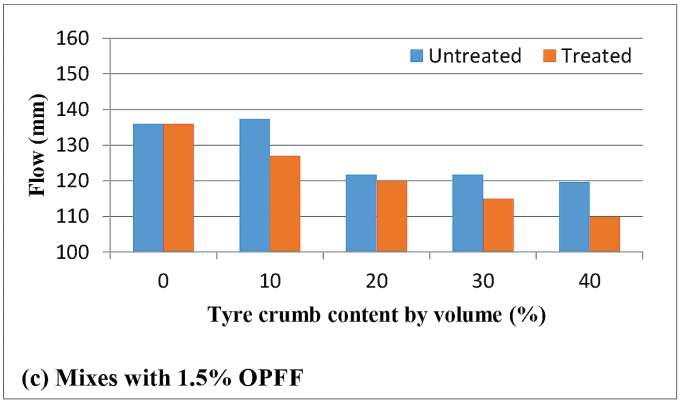


Figure 2. Workability of mortar composites (a) 0.5% OPFF; (b) 1% OPFF; (c) 1.5% OPFF

Density

The replacement of aggregates with tyre crumb resulted in a reduction in the unit weight of the hardened mortar specimens because of the low densities of the tyre crumbs. Figure 3 (a) and (b) show the density versus percentage of tyre crumb for untreated and treated samples,

respectively. The density of the treated tyre crumb mortar showed a slightly higher value compared with the untreated tyre crumb mortars. The higher density of the treated samples came from the cement powder addition on the outer surfaces of the tyre crumbs that increased the unit weight of the tyre particles from 589 to 668 kg/m^3 from the untreated to treated tyre particles respectively. Generally, the untreated tyre crumb particles entrap air in their surface, which makes the mortars exhibit a lower density compared to the treated tyre crumbs which have fewer voids and hence entrap less air. Meanwhile, the presence of a low dosage of OPFF in the matrix did not affect the density of the specimens significantly, as it does not occupy space in the specimens and also it is light in weight. The density ranging between 1675-1950 kg/m^3 for all tyre crumb replacement mortar mixes, which is in a range of lightweight concrete of 1675-1950 kg/m^3, as reported by Neville (1995). However, it is higher than lightweight mortar made of polymer of 1346 kg/m^3 reported by Uygunoglu et al. (2013) and similar range of lightweight mortar using pumice aggregate of 1800 kg/ m^3 according to Hossain et al. (2011).

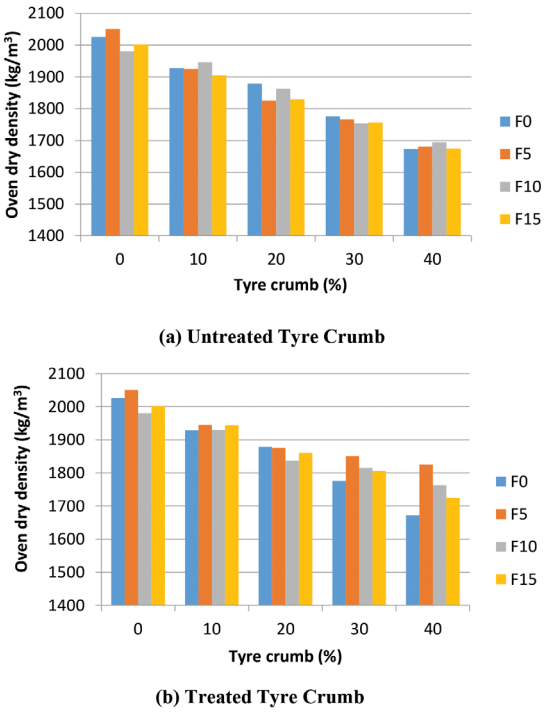


Figure 3. Density of mortars with OPFF and (a) Untreated tyre crumb; (b) treated tyre crumb

Water Absorption

The water absorption capacities of OPFF were compared to the control specimen as shown in Table 1. The addition of 0-1.5% OPFF showed increases of 4.2% to 8.5% on the percentage of water absorption of the mixes as compared to the control sample. Natural fibre is well known for its high water absorption capacity; however, because only a small amount is used,

its effect can be considered negligible. Obvious differences for the absorption capacity of the treated and untreated crumb tyre mortar are shown for the specimens with 0.5% OPFF added, as shown in Figure 4. In general, the treated tyre crumb mortars showed higher absorption than the untreated sample. However, the differences are not significant. The increased water absorption capacity of the treated tyre crumb mortars is a result of the coating treatment itself, which allows more water absorption.

Table 1
OPFF effect on water absorption

Sample Ref.	% of OPFF	Water Absorption (%)
F0	0.0	9.4
F5	0.5	9.8
F10	1.0	10.1
F15	1.5	10.2

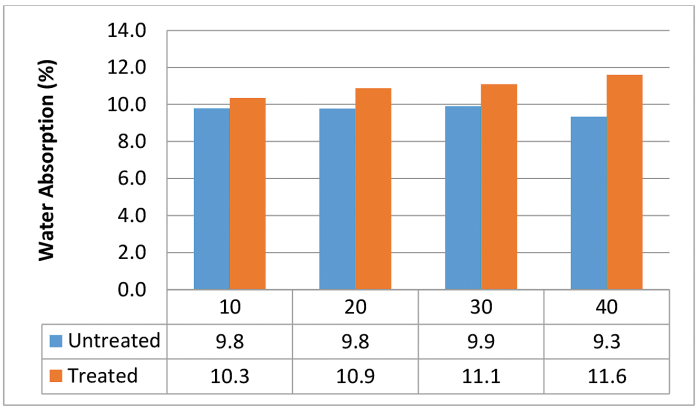


Figure 4. Water absorption of tyre crumb addition on mortar mixes with 0.5% OPFF

Compressive Strength

A compression test was first carried out on the samples with OPFF alone, as shown in Figure 5. In general, the addition of OPFF did reduce the strength of the mortar at 3, 7 and 28 days, except with the addition of 0.5%, which contributed to strength addition of about 4% at 28 days, revealing the best performance. Although this is not very significant, its advantage could be great when shrinkage is studied. The effects of the treated and untreated tyre crumbs were then studied on mortars with 0.5% OPFF, as in Figure 6. The replacement of aggregate with treated tyre crumb particles showed an improvement in the 7 days compressive strength by 0.11, 12.3, 12.4, and 14.4% for the replacement levels of 10, 20, 30, and 40%, respectively. Similarly, the same replacement at 28 days resulted in an improvement in the 28-days compressive strength by 5.2, 24.4, 22.5, and 20.1% for the replacement levels of 10, 20, 30, and 40%, respectively.

An improvement of the compressive strengths of treated tyre crumb resulted from a better adhesion of the roughly textured surface of the treated tyre particles, which provides better interlocking behaviour in the mortar matrix (Rangaraju & Gadka, 2012; Segre & Joekes, 2000). The SEM images of the treated and untreated crumb tyre mortar mixes obtained (Figure 7) indicated a similar texture. Furthermore, a clearer and larger air gap was observed at the interfacial zones (ITZ) of the untreated tyre crumb aggregate and the hardened cement paste, as shown in Figure 7(a), indicating a poorer bond than the treated one (see Figure 7b).

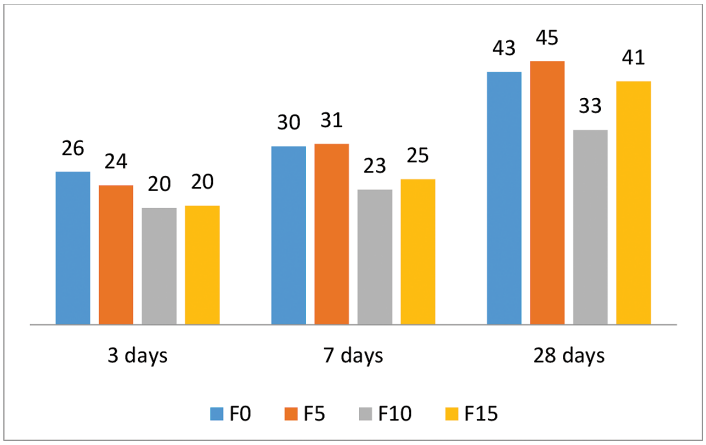


Figure 5. Compressive strength of mortar mix with 0%, 0.5%, 1.0% and 1.5% OPFF additions

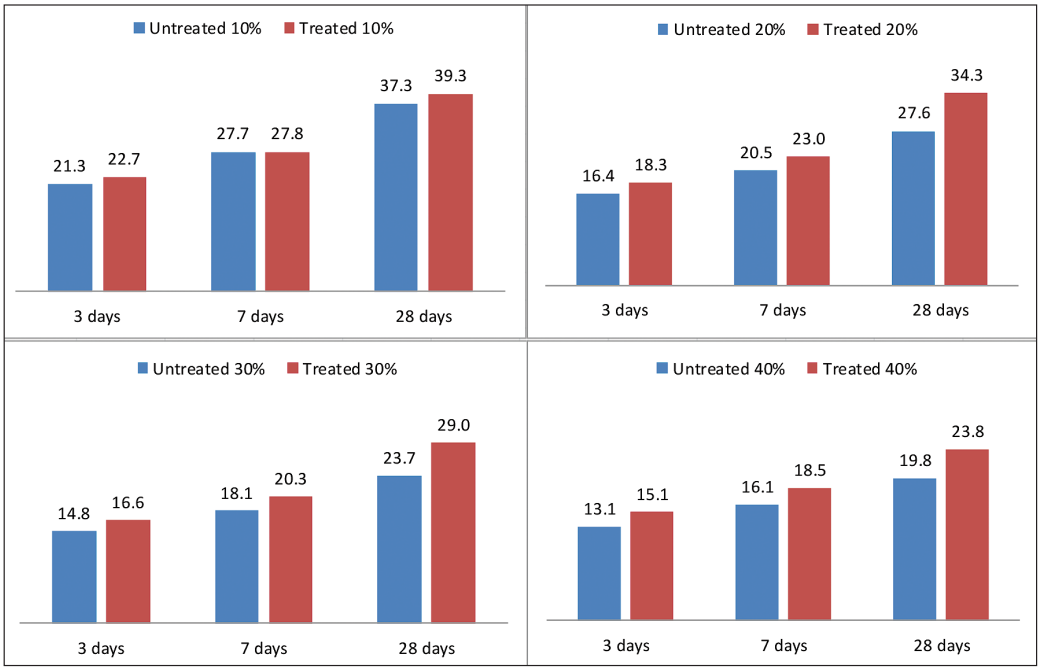


Figure 6. Compressive strength of mortars with untreated (CR) and treated (MR) tyre crumb additions

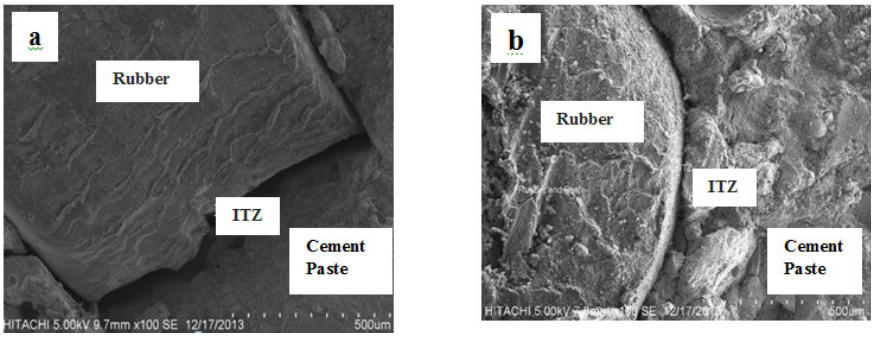


Figure 7. The SEM images of tyre crumb mortars (a) untreated; (b) treated tyre.

Split Tensile Strength

The split tensile strength of OPFF additions in the mortar mixes is shown in Figure 8. This shows that an addition of OPFF from 0.5% to 1.5% by volume contributed significantly to the strength of the composites. The maximum strength improvement was achieved by the addition of 0.5% OPFF, showing the best bond between the particles in the composite. As more OPFF was added, the tensile strength reduced due to the loosening bonds between the particles and the presence of more air voids.

Figure 9 shows the treated and untreated crumb replacement specimens with a 0.5% OPFF addition of mortar mixes after a 28-day curing period. The degradation of strength of the untreated crumb tyres is more obvious, particularly at 10% and 20% replacements. The strength reductions of untreated crumb tyre mortars are expected and these range from 28% to 50%. The same behaviour was reported by Edlin and Serouci (1993), Ganjian et al. (2009), and Najmi and Hall (2012). On the other hand, the treated crumb tyres showed a lower reduction by 1%, 13%, 38% and 47%, for the addition of 10%, 20%, 30% and 40% crumb tyres, respectively. Based on the compression, flexural and split tensile strength improvement in all the mixes containing treated tyre crumb particles, it can be concluded that the influence of the better surface texture of the treated tyre crumb, coupled with the interaction between the rough surfaces of the OPFF, helps in increasing a resistance to early failure, whereas the untreated tyre crumb mixes could easily slip on the surface due to the smoothness.

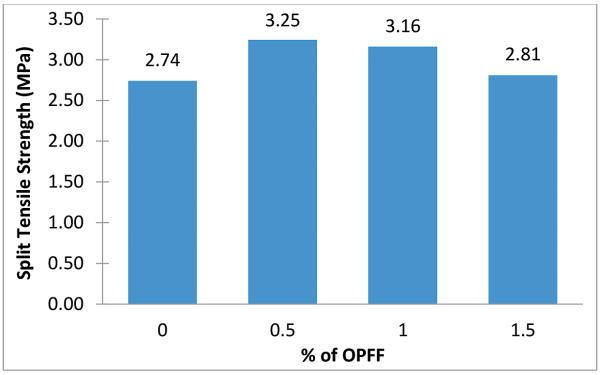


Figure 8. Effects of OPFF on the split tensile strength of the mortar mixes

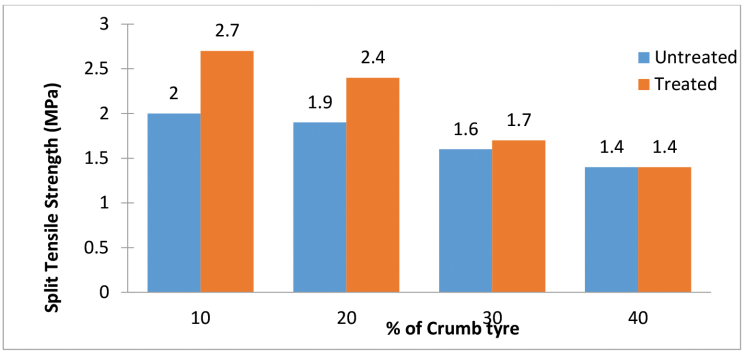


Figure 9. Effects of the treated and untreated tyre crumb on the split tensile strength of the mortar mixes

Flexural Strength

Figure 10 shows the effects of OPFF at 0.5, 1.0 and 1.5% by mass of cement content on the mortar's flexural strength. The influence of OPFF had less effect on the strength but extended the ductility of the material with the best percentage of OPFF addition being 0.5%, and this produced better strength and ductility as compared to the plain mortar, as shown in Figure 8.

The flexural test was conducted on the specimens with four different percentages of OPFF and four different percentages of tyre crumb (both treated and untreated), but only the best percentages of OPFF addition (i.e. 0.5%) for both the untreated and treated tyre crumb mortars were included. This is because the behaviour patterns of the samples, in terms of deflection, stiffness and ductility are the same and only the strength showed slight differences. Apart from this, all the composite mix specimens achieved a larger deflection than the normal weight mortar specimens, presenting more ductile behaviour due to the tyre crumb and oil palm fruit fibre additions. It was observed that the flexural strength and ductility of the treated tyre crumb yielded a better result than the untreated specimens with regards to flexure at all levels of tyre crumb additions. Table 2 shows the peak flexure load of treated and untreated tyre crumb and the percentage of improvement of the treated specimens as compared to the untreated one. The higher improvements observed in the treated tyre crumb incorporated mortars against the untreated was due to the textured surface of the treated tyre crumb particles, which resulted in a better interfacial bond between the tyre crumb particles and the matrix. This means that the OPFF is more effective in the distribution of internal stress in the treated tyre crumb mortars. The OPFF tends to bridge any tendency to initiate cracking due to higher tension forces, which results in higher flexural strength and ductility until the resistance is overcome.

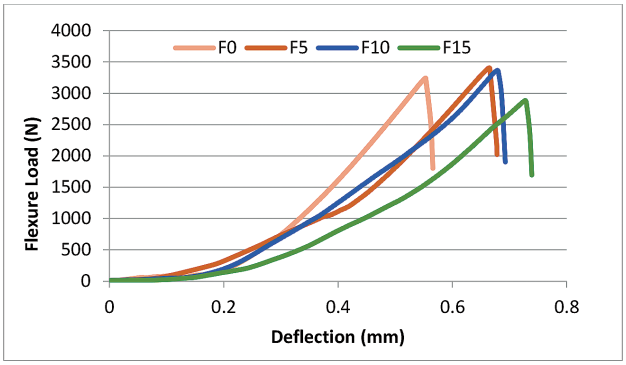


Figure 10. Effects of OPFF in the flexural strength of the mortar mixes

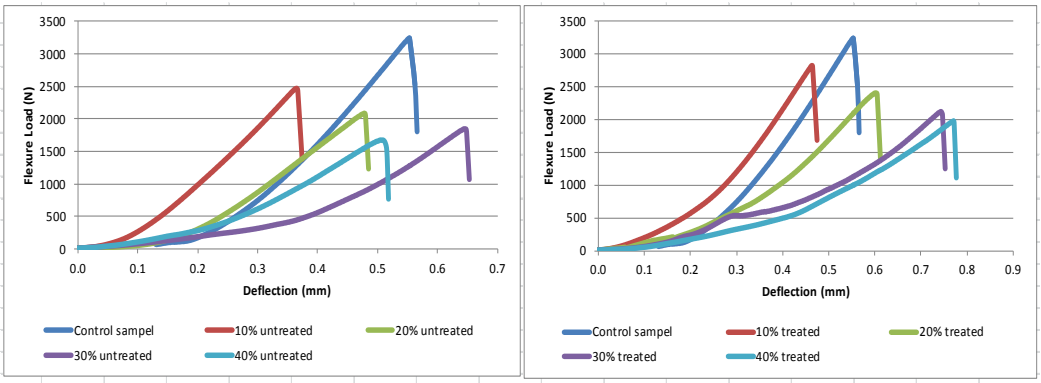


Figure 11. Flexural load vs. deflection of the untreated and treated tyre crumb and with the addition of 0.5% OPFF mortar mixes

Table 2

Flexural load improvement between the treated and untreated tyre crumb samples with 0.5% OPFF addition

Tyre crumb Content (%)	Untreated Tyre crumb (N)	Treated Tyre crumb (N)	Improvement (%)
10	2467	2824	14.5
20	2084	2405	15.4
30	1844	2118	14.9
40	1672	1979	18.4

Unrestrained Shrinkage

The shrinkage strains of the samples were monitored in either the controlled condition (i.e., climatic chamber) or in the uncontrolled laboratory condition. The effect of OPFF was first examined, whereby no tyre crumb was added. The addition of OPFF was compared to the control specimen, i.e. plain mortar in both the controlled and uncontrolled conditions. The results in Figure 12 show that the shrinkage strain increases with the increase in OPFF. This finding contradicts that of Ahmad et al. (2010), who reported that the OPFF fibre improved the shrinkage properties of the mortar.

The addition of OPFF in mortar composites containing untreated tyre crumb was observed to increase the shrinkage strain between 380-800 micro-strains, as shown in Figures 13 to 16, in both the climate chamber and at laboratory temperature. The maximum shrinkage strains of 680 and 710 micro-strains at 90-days were observed for climate chamber and laboratory climatic conditions, respectively. This result is in agreement with Toledo et al. (2005) and Ahmad et al. (2010), who reported on the shrinkage increment due to the addition of natural fibre. Natural fibre shows an increase in concrete shrinkage due to its higher porosity.

The shrinkage strains of the treated tyre crumb mortars containing OPFF are also shown in Figures 13 to 16. The shrinkage strain increases with an increase in the OPFF content without any particular pattern. A shrinkage strain of about 350-1000 micro strains was recorded for all the samples measured, regardless of the exposure conditions. In conclusion, there is no obvious difference in the shrinkage behaviour of the treated and untreated tyre crumb mortars. Also, in both specimens, no regular pattern was observed in terms of the maximum and minimum shrinkage strains.

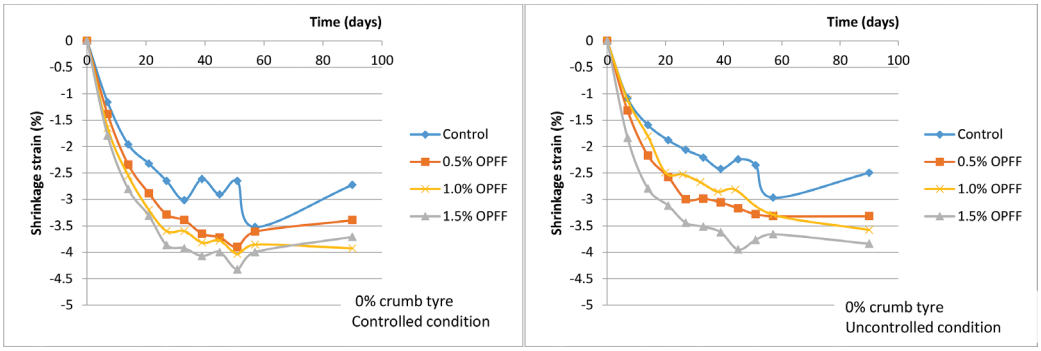


Figure 12. Effects of OPFF on the shrinkage properties of plain mortar in controlled and uncontrolled conditions

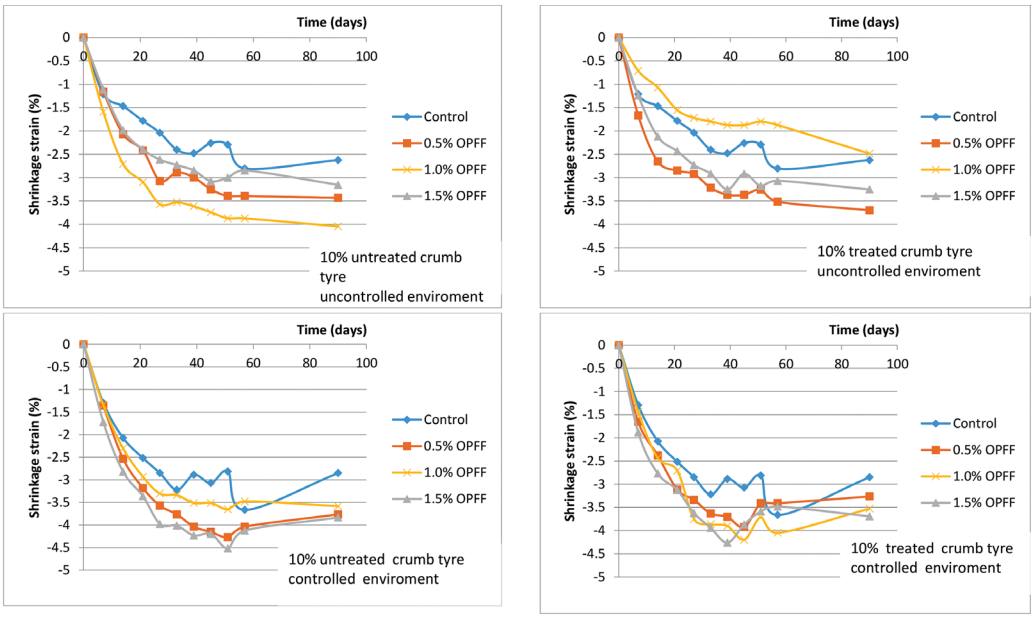


Figure 13. Shrinkage behaviour of 10% treated and untreated tyre crumb samples under controlled and uncontrolled conditions

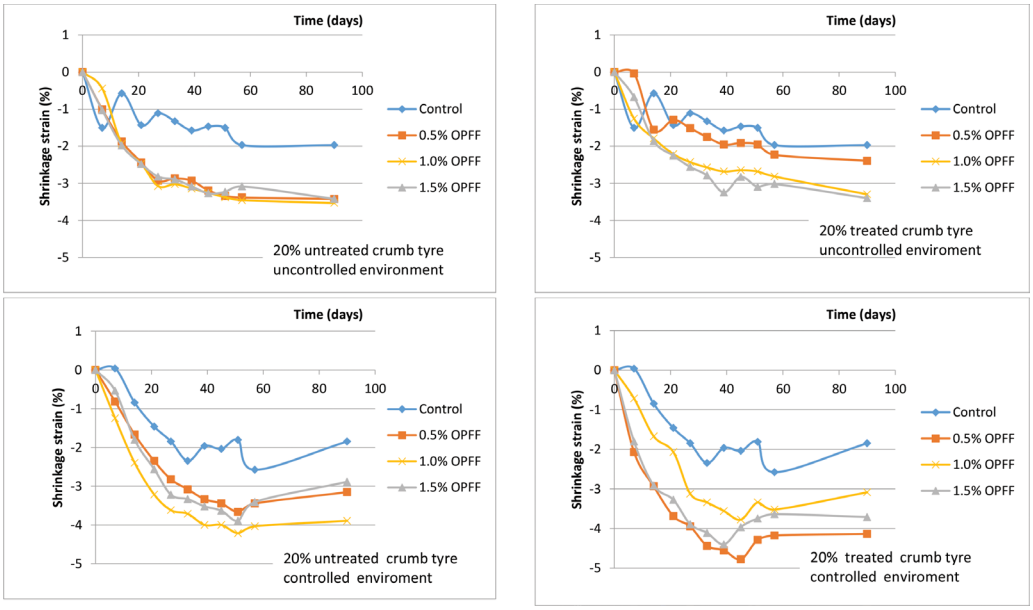


Figure 14. Shrinkage behaviour of 20% treated and untreated tyre crumb samples under controlled and uncontrolled conditions

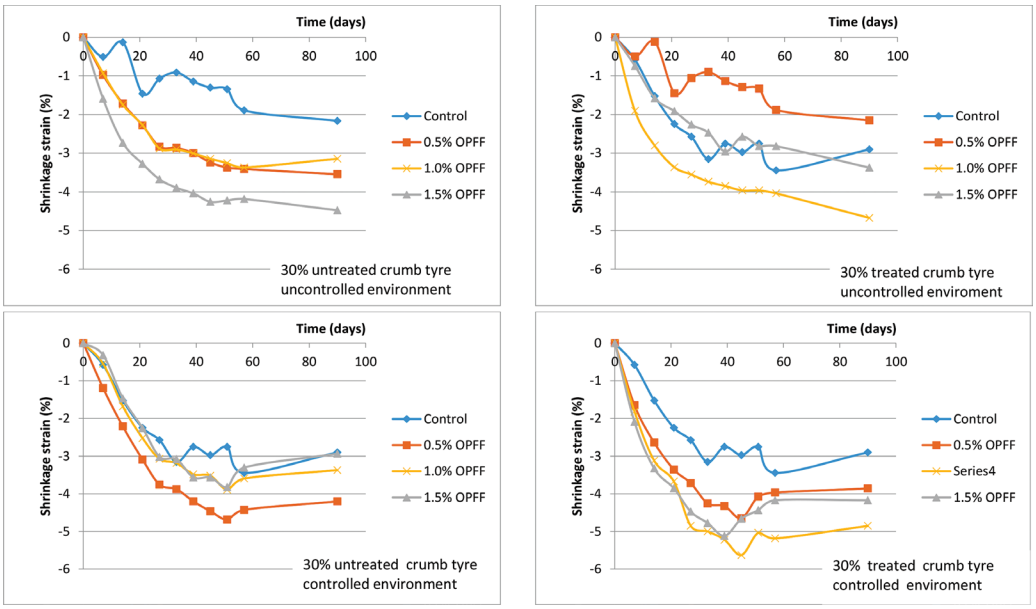


Figure 15. Shrinkage behaviour of 30% treated and untreated tyre crumb samples under controlled and uncontrolled conditions

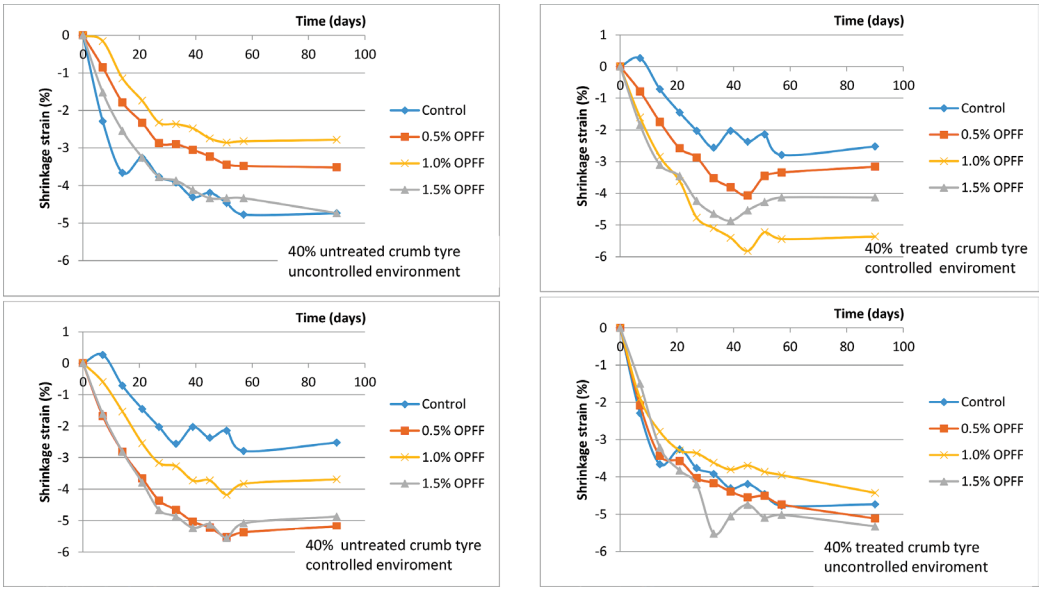


Figure 16. Shrinkage behaviour of 40% treated and untreated tyre crumb samples under controlled and uncontrolled conditions

CONCLUSION

The utilisation of tyre crumb particles as a replacement for aggregate in mortar composites is necessitated by the need to recycle waste tyres that are abundant and non-bio-degradable. However, the use of this material causes losses in strength properties. The losses were reduced by surface treatment and coating the tyre crumb particles with cement powder, coupled with the use of OPFF to redistribute the internal stress in the hardened samples to prevent sudden failure and improve performance. The results showed a promising avenue for producing a new composite mortar which demonstrates improved compressive, split tensile and flexural strengths compared with untreated tyre crumb incorporated mortars. Incorporating tyre crumb particles into the mortar matrix, as a partial replacement of the mineral aggregate, showed a decrease in the fresh concrete flowability and unit weight.

At hardened state, on the other hand, the results showed a promising avenue for producing a new composite mortar with pre-treated tyre crumb and OPFF addition that demonstrated better compressive, split tensile, and flexural strengths compared with the untreated tyre crumb incorporated mortars. Hence, it can be concluded that the replacement of up to 40% by volume of aggregate with pre-treated tyre crumb particles containing 0.5 to 1.0% OPFF addition have proven the mechanical properties of the new mortar composite. However, the addition of OPFF in both the untreated and treated tyre crumb specimens gave no particular pattern in terms of the shrinkage of mortar composite. Meanwhile, the microstructural investigation revealed that the decrease in the strength properties of the untreated tyre crumb mortars is due to soft, elastic and smooth surfaces of the tyre crumb particles, which reduce the bond between the tyre particles and the cement matrix.

REFERENCES

Ahmad, S., Elahi, A., Barbhuiya, S. A., & Farid, Y. (2012). Use of polymer modified mortar in controlling cracks in reinforced concrete beams. *Construction and Building Materials*, *27*(1), 91–96.

Ahmad, Z., Ibrahim, A., & Tahir, P. M. (2010). Drying Shrinkage Characteristics of Concrete Reinforced With Oil Palm Trunk Fibre. *International Journal of Engineering Science and Technology*, *2*(5), 1441-1450.

Ahmad, Z., Saman, H., & Tahir, T. (2010). Oil palm trunk fiber as a bio-waste resource for concrete reinforcement. *International Journal of Mechanical and Materials Engineering (IJMME), 5*(2), 199-207.

ASTM C33, A. (2004). Standard Specification for Concrete Aggregates. In *American Society for Testing and Material* (pp. 1-11). West Conshohocken, PA.

ASTM C109, A. (2002). Standard test method for compressive strength of hydraulic cement mortars (Using 2-in. or [50-mm] cube specimens). *American Society for Testing and Material*. West Conshohocken, PA.

ASTM C127, A. (1993). Standard Test Method for Specific Gravity and Absorption of Coarse Aggregate. *American Society for Testing and Material.* West Conshohocken, PA.

ASTM C150, A. (2001). Standard Specification for Portland Cement. *Annual Book of ASTM Standards*.

ASTM C341. (2006). Standard practice for length change of cast, drilled, or sawed specimens of hydraulic-cemet mortar and concrete. *American Society for Testing and Material*. West Conshohocken, PA.

ASTM C348. (2008). Standard specification for flexural strength of hydraulic-cement mortars. *American Society for Testing and Material*. West Conshohocken, PA.

ASTM C496. (2004). Standard specification for split tensile strengthof cylinderical concrete specimens. *American Society for Testing and Material*. West Conshohocken, PA.

ASTM C642, A. (2001). Standard Test Method for Density, Absorption, and Voids in Hardened Concrete. *American Society for Testing and Material.* West Conshohocken, PA.

ASTM C1437. (2007). Standard test method for flow of hydraulic cement mortar. *American Society for Testing and Material.* West Conshohocken, PA.

Aziz, F. N. A. A., Bida, S. M., Nasir, N. A. M., & Jaafar, M. S. (2014). Mechanical properties of lightweight mortar modified with oil palm fruit fibre and tire crumb. *Construction and Building Materials*, *73*, 544-550.

Benazzouk, A., Douzane, O., Langlet, T., Mezreb, K., Roucoult, J., & Quéneudec, M. (2007). Physico-mechanical properties and water absorption of cement composite containing shredded rubber wastes. *Cement and Concrete Composites, 29*(10), 732-740.

Bektas, F., Wang, K., & Ceylan, H. (2009). Effects of crushed clay brick aggregate on mortar durability. *Construction and Building Materials*, *23*(5), 1909–1914.

Braga, M., de Brito, J., & Veiga, R. (2012). Incorporation of fine concrete aggregates in mortars. *Construction and Building Materials*, *36*, 960–968.

Brown, K. M., Cummings, R., Mrozek, J. R., & Terrebonne, P. (2001). Scrap tire disposal: three principles for policy choice. *Natural Resources Journal*, *41*(1), 9-22.

Cairns, R., & Kenny, M. (2004). *The use of recycled rubber tyres in concrete*. Paper presented at the Sustainable Waste Management and Recycling: Used/Post-Consumer Tyres. Concrete and Masonry Research Group and Held at Kingston University-London.

Chou, L. H., Lin, C. N., Lu, C. K., Lee, C. H., Lee, M. T. (2010). Improving rubber concrete by waste organic sulfur compounds. *Waste Management & Research, 28*(1), 29-35.

Chou, L. H., Yang, C. K., Lee, M. T., & Shu, C. C. (2010b). Effects of partial oxidation of crumb rubber on properties of rubberized mortar. *Composites Part B: Engineering*, *41*(8), 613-616.

Chung, K. H., & Hong, Y. K. (1999). Introductory behavior of rubber concrete. *Journal of Applied Polymer Science*, *72*(1), 35-40.

Correia, S. L., Partala, T., Loch, F. C., & Segadães, A. M. (2010). Factorial design used to model the compressive strength of mortars containing recycled rubber. *Composite Structures*, *92*(9), 2047–2051.

Eldin, N. N., & Senouci, A. B. (1993). Rubber-tire particles as concrete aggregate. *Journal of Materials in Civil Engineering, 5*(4), 478-496.

Filho, R. D. T., Ghavami, K., Sanjuán, M. A., & England, G. L. (2005). Free, restrained and drying shrinkage of cement mortar composites reinforced with vegetable fibres. *Cement and Concrete Composites*, *27*(5), 537-546.

Ganjian, E., Khorami, M., & Maghsoudi, A. A. (2009). Scrap-tyre-rubber replacement for aggregate and filler in concrete. *Construction and Building Materials*, *23*(5), 1828-1836.

Hossain, K., Ahmed, S., & Lachemi, M. (2011). Lightweight concrete incorporating pumice based blended cement and aggregate: mechanical and durability characteristics. *Construction and Building Materials, 25*(3), 1186–95.

Huang, B., Li, G., Pang, S. S., & Eggers, J. (2004). Investigation into waste tire rubber-filled concrete. *Journal of Materials in Civil Engineering*, *16*(3), 187-194.

Ismail, M. A., & Hashim, H. (2008). Palm oil fiber concrete. In *Proceedings of the 3th ACF International Conference Sustainable Concrete Technology and Structures in Local Climate and Environmental Conditions.* Ho Chi Minh City, Vietnam.

Jingfu, K., Chuncui, H., & Zhenli, Z. (2009). Strength and shrinkage behaviors of roller-compacted concrete with rubber additives. *Materials and structures*, *42*(8), 1117-1124.

Khatib, Z. K., & Bayomy, F. M. (1999). Rubberized Portland cement concrete. *Journal of Materials in Civil Engineering*, *11*(3), 206-213.

Kosievith, V., Pinto, C., & Hamassaki, L. (2001). *Rubber powder and Portland cement composites.* Brazil: Polythenic School, University of Sao Paulo.

Lee, H. S., Lee, H., Moon, J. S., & Jung, H. W. (1998). Development of tire added latex concrete. *ACI Materials Journal, 95*(4), 356–364.

Li, G., Stubblefield, M. A., Garrick, G., Eggers, J., Abadie, C., & Huang, B. (2004). Development of waste tire modified concrete. *Cement and Concrete Research, 34*(12), 2283-2289.

Li, Z., Li, F., & Li, J. (1998). Properties of concrete incorporating rubber tyre particles. *Magazine of Concrete Research*, *50*(4), 297-304.

Najim, K. B., & Hall, M. R. (2012). Mechanical and dynamic properties of self-compacting crumb rubber modified concrete. *Construction and Building Materials*, *27*(1), 521-530.

Nehdi, M., & Khan, A. (2001). Cementitious composites containing recycled tire rubber: An overview of engineering properties and potential applications. *Cement Concrete and Aggregates, 23*(1), 3–10.

Neville, A. M., & Neville, A. M. (1995). *Properties of concrete* (Vol. 4). Longman, London.

Neno, C., Brito, J. D., & Veiga, R. (2014). Using fine recycled concrete aggregate for mortar production. *Materials Research*, *17*(1), 168-177.

Onuaguluchi, O., & Panesar, D. K. (2014). Hardened properties of concrete mixtures containing pre-coated crumb rubber and silica fume. *Journal of Cleaner Production*, *82*, 125-131.

Pedro, D., de Brito, J., & Veiga, R. (2012). Mortars made with fine granulate from shredded tires. *Journal of Materials in Civil Engineering*, *25*(4), 519-529.

Raghavan, D., Huynh, H., & Ferraris, C. (1998). Workability, mechanical properties, and chemical stability of a recycled tyre rubber-filled cementitious composite. *Journal of Materials Science*, *33*(7), 1745-1752.

Rangaraju, P. R., & Gadkar, S. (2012). *Durability evaluation of crumb rubber addition rate on portland cement concrete*. Clemson University Report. USA.

Sánchez, E., Massana, J., Garcimartín, M. A., & Moragues, A. (2008). Mechanical strength and microstructure evolution of fly ash cement mortar submerged in pig slurry. *Cement and Concrete Research*, *38*(5), 717–724.

Segre, N., & Joekes, I. (2000). Use of tire rubber particles as addition to cement paste. *Cement and Concrete Research*, *30*(9), 1421-1425.

Silva, J., de Brito, J., & Veiga, R. (2010). Recycled red-clay ceramic construction and demolition waste for mortars production. *Journal of Materials in Civil Engineering*, *22*(3), 236–244.

Taha, M. M. R., El-Dieb, A. S., El-Wahab, M. A. A., & Hameed, M. E. A. (2008). Mechanical, fracture, and microstructural investigations of rubber concrete. *Journal of Materials in Civil Engineering*, *20*(10), 640-649.

Topcu, I. B. (1995). The properties of rubberized concretes. *Cement and Concrete Research*, *25*(2), 304-310.

Turki, M., Bretagne, E., Rouis, M. J., & Quéneudec, M. (2009). Microstructure, physical and mechanical properties of mortar–rubber aggregates mixtures. *Construction and Building Materials*, *23*(7), 2715–2722.

Uygunoğlu, T., Brostow, W., Gencel, O., & Topçu, İ.B. (2013). Bond strength of polymer lightweight aggregate concrete. *Polymer Composites*, *34*(12), 2125-2132.

Wang, H., & Huang, W. (2010). Durability of self-consolidating concrete using waste LCD glass. *Construction and Building Materials*, *24*(6), 1008–1013.

Pertanika J. Sci. & Technol. 25 (1): 121 - 138 (2017)

SCIENCE & TECHNOLOGY

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

Diabetes Knowledge, Attitudes, Self-management, and Quality of Life among People with Type 2 Diabetes Mellitus – A Comparison between Australia- and Malaysia-Based Samples

Yee Cheng Kueh[1, 2]*, Tony Morris[1] and Aziz-Al-Safi Ismail[3]

[1]*Institute of Sport, Exercise and Active Living, College of Sport and Exercise Science, Victoria University, 14428 Melbourne, Australia*

[2]*Unit of Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, 16150 USM, Kubang Kerian, Kelantan, Malaysia*

[3]*Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, 16150 USM, Kubang Kerian, Kelantan, Malaysia*

ABSTRACT

The present study aimed to examine the levels of diabetes knowledge, attitudes, self-management and quality of life (QoL) between two countries with different cultural and language backgrounds. Data collection was conducted in two hospitals in Melbourne, Australia, and a hospital in Kelantan, Malaysia. Participants with type 2 diabetes mellitus (T2DM) were asked to complete four questionnaires, measuring diabetes knowledge, attitudes, self-management, and QoL. The differences between the samples were examined using chi-square and independent samples *t*-tests. The variables of gender and type of treatment (using insulin or not using insulin for treatment) based on groups were analysed using one-way ANOVA. All analyses were conducted using SPSS 22.0. The results highlighted some similarities and differences between the Australia-based sample and the Malaysia-based sample. In general, the Australia-based participants scored significantly higher in diabetes knowledge and reported more regular self-management of T2DM in exercise, blood glucose testing and foot care. The Australia-based sample also scored higher on attitudes compared to the Malaysia-based participants. On the other hand, Malaysia-based participants reported a lower level of impacts of T2DM on QoL. There was no significant difference between self-management of T2DM in terms of diet and satisfaction as an aspect of QoL related to living with diabetes between the two samples. The present study highlighted the levels of diabetes knowledge, attitudes, self-management, and QoL among Australia-based and Malaysia-based people with T2DM.

Keywords: Type 2 diabetes mellitus, knowledge, attitudes, self-management, quality of life, insulin, gender

Article history:
Received: 20 January 2016
Accepted: 13 April 2016

E-mail addresses:
yckueh@usm.my (Yee Cheng Kueh),
Tony.Morris@vu.edu.au (Tony Morris),
aziz.alsafi@usm.my (Aziz-Al-Safi Ismail)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

The prevalence of type 2 diabetes mellitus (T2DM) is causing a major public health burden across the world. Diabetes is a health, social and economic burden for individuals with the progressive, chronic condition and for their families and the community. It is also associated with various disease complications, so it impacts on the quality of life (QoL) and life expectancy for people diagnosed with diabetes. Diabetes mellitus has been one of the top ten leading causes of death in Malaysia since 2002 (World Health Organization, 2006). The latest statistics released by Diabetes Australia Victoria (DAV) in 2011 stated that diabetes is the sixth leading cause of death in Australia (Diabetes Australia Victoria, 2011). Malaysia was reported to have approximately 1.2 million people with diagnosed diabetes (Zanariah et al., 2009), which is higher than Australia with 898,800 people reported to have diabetes (Australian Institute of Health and Welfare, 2011). However, DAV has reported that an estimated total of 1.7 million Australians have diabetes, but 50% are undiagnosed (Diabetes Australia Victoria, 2011). As diabetes has become the fastest growing chronic disease and one of the leading causes of death in Australia and Malaysia, it is important for researchers to identify ways to help people manage their diabetes more effectively to reduce the progress of complications for individuals and minimise the burden for health services.

Although medical and health services must play a major role in diagnosis, treatment and monitoring of T2DM day-to-day management have to lie with the individuals with diabetes, with the support of their families (Haas et al., 2013). Thus, self-management is critical to maintaining health and slowing down the progress of the disease. Self-management includes lifestyle factors including diet and exercise, as well as taking prescribed medication, and monitoring of blood sugar and foot care. For most individuals with little or no medical background, the nature and progress of diabetes and what can or must be done to manage it is not self-evident.

Being diagnosed with a progressive, chronic condition like T2DM can be a shock and people can react with negative feelings of varying intensities. Individuals' attitudes to having diabetes are likely to impact on their response to the demands of the condition for self-management (Nam, Chesla, Stotts, Kroon, & Janson, 2011). Research has shown that people who were diagnosed with diabetes have lower QoL compared to the general population (Solli, Stavem, & Kristiansen, 2010). The impacts of having diabetes and how people with T2DM address the condition by finding out how to manage it, perceiving ways through to a healthier lifestyle and managing the condition from day to day will affect their QoL. However, the perception of QoL and level of knowledge, attitudes and self-management among people with T2DM may vary across different cultures and countries. Therefore, these variables are the important aspects in diabetes research globally.

One factor that has been widely been cited as distinguishing between cultures in the western countries and many Asian countries is how strongly individualist or collectivist they are (Hofstede, 2001). Most western countries have individualist cultures in which people think first about themselves. Many Asian countries have collectivist cultures in which personal interests are subservient to the needs of the community. This creates cultural differences that can indirectly influence people's self-management of health conditions like diabetes, as well

as how they interact with health professionals (Tripp-Reimer, Choi, Kelley, & Enslein, 2001). Individualistic cultures, such as Australia, which are associated with independent ideas of self, prioritise personal goals and uphold individual attitudes as key determinants of behaviour, including health-related behaviour (Matsumoto, Yoo, & Fontaine, 2008). Therefore, privacy and assertiveness are highly valued in individualist cultures. Conversely, collectivistic cultures, such as Malaysia, are associated with interdependent selves and group-focused goals. They value the group's or community's norms as primary determinants of behaviour, and they value group harmony. These distinctive characteristics that are influenced by the culture of the country people live in may affect how they act on their illnesses and value their life as an individual with T2DM.

Gender differences have been reported in diabetes knowledge, attitudes, self-management and QoL. For instance, Kamel, Badawy, El-Zeiny, and Merdan (1999) showed that women generally had a poorer level of diabetes knowledge compared to males. In a study conducted in Pakistan, Rafique, Azam, and White (2006) reported that males had more positive attitudes in comparison with females. Misra and Lager (2009) also reported that women had more difficulty in managing their diet and self-management than men. This could be because women are mostly responsible for preparing meals for their family and because of their diabetic condition, they have to prepare their own meal separately (Samuel-Hodge et al., 2000). Thus, women are most likely to experience difficulties related to their self-management of diet. Women were also found to be more anxious about their diabetes condition, thus, they generally experienced lower QoL than males (Misra & Lager, 2009).

In addition, diabetes treatment, such as insulin therapy, can have a significant impact on the life of people with T2DM. Insulin therapy usually occurs against a background of many years of habitual diabetes self-management that may have to be changed to improve the health condition of people with T2DM. This can be a reminder of the progressive nature of the condition that requires more complex and invasive treatment. Furthermore, the level of impact of insulin therapy on people with T2DM can vary in different cultural contexts. Some researchers have reported that insulin therapy was not associated with QoL (UK Prospective Diabetes Study Group, 1999). Other researchers have shown the effects of insulin on well-being or QoL, but they have found different relationships, ranging from beneficial (Chow, Tsang, Sorensen, & Cockram, 1995; Pibernik-Okanovic, Szabo, & Metelko,1998) to deleterious (Hanninen, Takala, & Keinanen-KiuKaanniemi, 1998; Goddijn et al., 1999). One reason why the results of studies undertaken in different cultures are not consistent is probably that the research methods varied between the studies. Therefore, it is important to undertake cross-cultural studies in which the same methods are employed in different cultures, allowing direct comparison. Furthermore, we could not find studies in the literature that examined comparisons between countries and gender differences in diabetes knowledge, attitudes, self-management, and QoL among people with T2DM. Such comparative assessments are important for us to understand disparities in those aspects between different cultural groups so that appropriate strategies can be utilised to improve the health condition of vulnerable people in the diabetes population.

In the present study, we aimed to examine the differences in the levels of diabetes knowledge, attitudes, self-management and QoL between individuals based in Australia and

Malaysia, who have been diagnosed with T2DM. Further, we examined differences in the diabetes knowledge, attitudes, self-management, and QoL between adult males and females, as well as between insulin users and non-users in Australia and Malaysia.

METHOD

Participants

Participants for this study were males and females, aged over 18 years. We included only individuals who were diagnosed with T2DM at least a year and above by medical practitioners and were registered with the specific hospitals that collaborated in this research. The participants were identified through the patient database of the hospitals or they were referred to the researchers by staff nurses.

Measures

The Demographic and Health Measure Form included several demographic and health treatment questions. These questions assessed the participants' personal attributes (e.g., age, gender) and their health condition (e.g., how long they have been diagnosed with T2DM, how it has been treated) at the time when they completed the measures.

The Diabetes Knowledge (DKN) scale is a reliable diabetes knowledge assessment in people with diabetes. Furthermore, it is easily administered, being short with 15 multiple choice questions (Beeney, Dunn, & Welch, 1994). All the questions require a single correct answer, except for items 13 to 15, for which several answers are correct and all must be checked to obtain a score of 1. The total score for each form is the sum of correct answers in a score range of 0 to 15, with higher scores indicating higher levels of diabetes knowledge. Then, the scores were converted into percentages. The DKN has internal consistency reliability of Cronbach's alpha 0.82 (Beeney et al., 1994).

The Diabetes Integration Scale-19 (ATT19) consists of 19 self-administered attitudinal statements on the perception of diabetes among people with diabetes. Responses are made on a 5-point Likert scale with possible answers from 1 (*I disagree completely*) to 5 (*I agree completely*). In the present study, the total score was computed by summing the scores from all 19 items. Therefore, the potential score ranged from 19 to 95. The ATT19 has an internal consistency reliability of Cronbach's alpha 0.84 (Welch, Beeney, Dunn, & Smith, 1996).

Summary of Diabetes Self-Care Activities (SDSCA) is a brief self-report measure of diabetes self-management developed by Toobert, Hampson, and Glasgow (2000). It measures of the frequency of practicing different diabetes regimen activities over the past seven days. The SDSCA consists of 11 core items that were used in the present study. These core items measure self-care behaviour in terms of diet (general and specific), exercise, blood glucose testing, foot care and smoking status. In psychometric validation, the average inter-item correlations were high and generally exceeded 0.5 (Toobert & Glasgow, 1994). In the present study, the two items that refer to specific diet were omitted from the data analysis. This is because the specific diet items developed for use in America were not suitable for the Malaysia-based population, where diet is quite different, so they do not provide a meaningful comparison between the samples.

The Diabetes Quality of Life questionnaire (DQoL) measures the relative burden of diabetes treatment, with the goal of maintaining blood glucose levels as close as possible to those of people without diabetes (Diabetes Control and Complications Trial Research Group, 1988). The DQoL questionnaire used in this study consists of 15 items measuring satisfaction, 20 items measuring impacts, and one item measuring self-rated general health. Items are scored on a 5-point Likert scale rated from 1 (very satisfied, no impact) to 5 (very dissatisfied, very impacted). The total scores for each subscale were converted to percentages, where the highest percentage of satisfaction is considered to reflect highest QoL, and the highest percentage of impact is considered to represent lowest QoL. The DQoL has good reliability with test-retest correlations in the range of 0.78 to 0.92 (Diabetes Control and Complications Trial Research Group, 1988).

Procedure

A cross-sectional design study was conducted. For the Australia-based sample, two hospitals, the Alfred Hospital and Western Hospital, agreed to take part in this study. For the Malaysia-based sample, the participants were recruited via the Diabetes Health Clinic in Hospital Universiti Sains Malaysia (HUSM). A non-probability, convenience sampling method was applied in recruiting the participants. For administration to the Malaysia-based sample, the original English-language questionnaires were translated into the local familiar language, which is Malay (Kueh, 2014; Kueh, Morris, & Ismail, 2014). Content and construct validity to ensure the questionnaires were appropriate for use in Malaysia had been conducted in a previous study (Kueh, 2014). People with T2DM were invited to participate in this study by completing the questionnaire pack during their routine clinical appointment with their physician. The participants were provided with a research information sheet and an informed consent form. Written consent was obtained from the participants who had agreed to take part in this study by completing the questionnaires provided to them. The participants were informed they were allowed to withdraw from the study at any stage or to restrict their data for use in the analysis and report. The research was approved by the Human Research Ethics Committees of Victoria University (ethics code: HRETH 08/139), Alfred Hospital and Western Hospital in Melbourne, Australia, and Universiti Sains Malaysia in Malaysia.

Data Analysis

The independent samples t-test was used to identify the difference between the Australia-based and Malaysia-based samples on numerical variables. Chi-square statistic was used to identify any differences between the Australia-based and Malaysia-based samples on categorical variables. The differences between the gender and insulin treatment of both groups were tested using the One-way ANOVA. The differences between each pair of means were then determined by using Post-hoc Tukey tests, which are frequently used in multiple group comparisons. All the statistical analyses were conducted using SPSS 22.0 and significant level was set as 0.05.

RESULTS

Characteristics and Differences of the Participants

The total participants included 284 Australia-based participants and 276 Malaysia-based participants. The mean age for the participants in the Australia-based and Malaysia-based samples were 56.0 (*SD* = 11.01) and 57.1 (*SD* = 8.47), respectively. Table 1 shows the respondents' characteristics in the Australia-based and Malaysia-based samples and the differences between the two samples. The results revealed that there were significant differences between the two samples in terms of duration of diabetes since diagnosis, levels of diabetes knowledge, frequency of conducting exercise, blood glucose testing, foot care and level of impacts on QoL.

Table 1
Independent t-test of Differences in Characteristics of Respondents with T2DM.

Characteristics (numerical variables)	Australia-based sample Mean (SD)	Malaysia-based sample Mean (SD)	*t*-statistic	*p*-value
Age	56.0 (11.01)	57.1 (8.47)	1.30	0.196
Duration of diabetes	12.0 (9.01)	10.4 (7.53)	2.24	0.025
Diabetes knowledge	59.1 (19.69)	52.3 (17.27)	5.80	<0.001
Attitude	63.4 (11.68)	60.5 (9.44)	3.23	0.001
Self-management:				
Diet	4.9 (1.91)	5.2 (2.13)	1.50	0.135
Exercise	3.5 (2.26)	2.4 (2.34)	5.78	<0.001
Blood glucose testing	4.9 (2.48)	1.2 (1.80)	20.38	<0.001
Foot care	3.4 (2.58)	2.9 (2.65)	2.17	0.030
QoL:				
Impact	26.0 (16.30)	22.7 (14.30)	3.98	<0.001
Satisfaction	68.3 (17.99)	70.68 (14.43)	1.73	0.085

Table 2 presents the significant differences between participants from the two cultures in terms of their diabetes treatment, education background, working status, smoking, and self-rated general health. Results showed that there was a significant difference between the cultures in all these areas. In general, the Australia-based participants had experienced higher levels of education compared to the Malaysia-based participants. The results also showed that the majority of the participants were non-smokers for both the samples. More Australia-based participants rated themselves as poor in general health compared to the Malaysia-based participants.

Table 2
Chi-square Test of Differences in the Characteristics of Respondents with T2DM.

Characteristics (categorical variables)	Australia-based sample n (%)	Malaysia-based sample n (%)	χ^2 value	*p*-value
Type of treatment:				
Diet	25 (8.8%)	6 (2.2%)	57.63	<0.001
Diet and Tablet	107(37.7%)	163 (59.1%)		
Diet and Insulin	82 (28.9%)	23 (8.3%)		
Diet, tablet, and insulin	70 (24.6%)	84 (30.4%)		
Education level:				
Less than high school	74 (26.1%)	70 (25.4%)	13.05	0.005
High school graduate	108 (38.0%)	140 (50.6%)		
College graduate	45 (15.8%)	33 (12.0%)		
University graduate	57 (20.1%)	33 (12.0%)		
Working status:				
Yes	183 (64.4%)	113 (40.9%)	29.10	<0.001
No	101 (35.6%)	163 (59.1%)		
Smoking:				
No	230 (81.0%)	248 (89.9%)	8.69	0.003
Yes	54 (19.0%)	28 (10.1%)		
General health:				
Excellent	12 (4.2%)	21 (7.6%)	46.10	<0.001
Good	117 (41.2%)	112 (40.6%)		
Fair	104 (36.6%)	139 (50.4%)		
Poor	51 (18.0%)	4 (1.4%)		

The results from both the tables indicate that when compared with the Malaysia-based participants, the Australia-based participants on average showed greater diabetes knowledge, more positive attitudes towards T2DM, practised self-management more regularly (i.e., exercise, blood glucose testing, and foot care), included a higher percentage of university graduates and a larger proportion of the Australia-based sample were still working. However, a larger percentage of the Malaysia-based participants rated themselves as either excellent or good in general health, and only a small percentage viewed their general health as poor compared to the Australia-based participants.

Results Based on Gender

The ANOVA results in Table 3 for the differences based on gender indicated that there was a significant difference between the males and females in terms of their diabetes knowledge, attitudes, self-management in exercise, blood glucose testing and impact of diabetes.

Table 3
One-Way ANOVA Results for the Differences Based on Gender

Study variables	Australia-based sample		Malaysia-based sample		*F*–statistic	*p*-value
	Female, n=99	Male, n=185	Female, n=147	Male, n=129		
Diabetes Knowledge	66.2 (17.62)	58.8 (20.29)	52.4 (18.86)	52.3 (15.34)	14.79	<0.001
Attitudes	64.2 (11.25)	63.0 (11.92)	60.4 (8.95)	60.7 (10.01)	3.73	0.011
Self-management:						
Diet	5.0 (1.76)	4.9 (2.00)	5.4 (2.06)	5.0 (2.20)	1.52	0.208
Exercise	3.5 (2.10)	3.6 (2.36)	2.1 (2.27)	2.8 (2.35)	13.63	<0.001
Blood glucose testing	5.0 (2.48)	5.0 (2.50)	1.1 (1.85)	1.3 (1.73)	137.33	<0.001
Foot care	3.7 (2.50)	3.2 (2.61)	2.8 (2.57)	3.0 (2.75)	2.26	0.081
QoL:						
Impact	26.6 (16.30)	28.6 (16.30)	22.1 (14.26)	23.4 (14.36)	5.82	0.001
Satisfaction	67.8 (17.86)	68.6 (18.11)	72.0 (13.95)	69.2 (14.87)	1.67	0.173

The Post-hoc Tukey test revealed that the Australia-based females had a significantly higher level of diabetes knowledge than the Australia-based males ($p = 0.007$) and both genders for the Malaysia-based sample (female, $p < 0.001$; male, $p < 0.001$). The Australia-based females had significantly higher positive attitudes than the Malaysia-based females ($p = 0.032$). In self-management of T2DM, the Australia-based participants of both genders conducted self-care activities such as exercise and blood glucose testing more frequently than both genders of Malaysia-based participants. The Australia-based males had a significantly higher mean of diabetes impacts compared to the Malaysia-based males ($p = 0.019$) and females ($p = 0.001$).

Table 4
Post-Hoc Tukey Test for Gender

Study variables	*Difference between groups, p*-value					
	AF vs AM	AF vs MF	AF vs MM	AM vs MF	AM vs MM	MF vs MM
Diabetes knowledge	0.007	<0.001	<0.001	0.009	0.011	1.000
Attitudes	0.811	0.032	0.066	0.117	0.225	0.996
Self-management:						
Exercise	0.995	<0.001	0.114	<0.001	0.022	0.039
Blood glucose testing	0.996	<0.001	<0.001	<0.001	<0.001	0.743
QoL:						
Impact	0.721	0.112	0.418	0.001	0.019	0.888

AF = Female of Australia-based sample, AM = Male of Australia-based sample, MF = Female of Malaysia-based sample, MM = Male of Malaysia-based sample

Results Regarding Treatment with Insulin

Table 5 shows the ANOVA results for the differences based on the treatment with insulin. The findings indicated that there was a significant difference between insulin users and non-insulin users in diabetes knowledge, attitudes, self-management in diet, exercise, blood glucose testing, foot care and impacts of QoL on T2DM.

Table 5
One-Way ANOVA Results for Differences Based on Insulin Treatment

Study variables	Australia-based sample		Malaysia-based sample		*F* –statistic	*p*-value
	Without insulin, n=132	With insulin, n=152	Without insulin, n=169	With insulin, n=107		
Diabetes Knowledge	55.7 (19.10)	66.4 (18.87)	47.6 (16.83)	59.7 (15.29)	31.15	<0.001
Attitudes	62.8 (11.81)	63.9 (11.58)	61.2 (9.36)	59.3 (9.51)	4.41	0.004
Self-management:						
Diet	4.67 (2.09)	5.2 (1.71)	5.3 (2.10)	5.2 (2.18)	2.59	0.052
Exercise	3.6 (2.24)	3.5 (2.30)	2.4 (2.27)	2.5 (2.44)	11.30	<0.001
Blood glucose testing	3.87 (2.59)	5.8 (1.98)	0.7 (1.28)	2.0 (2.16)	192.71	<0.001
Foot care	3.06 (2.61)	3.67 (2.52)	2.6 (2.61)	3.4 (2.64)	4.96	0.002
QoL:						
Impact	25.4 (15.67)	30.1 (16.57)	20.5 (13.21)	26.2 (15.29)	10.70	<0.001
Satisfaction	67.4 (19.02)	69.1 (17.07)	71.4 (14.40)	69.5 (14.46)	1.54	0.202

The results of Post-hoc Tukey test revealed that no matter which culture they are from, people with T2DM using insulin in their treatment had a significantly higher level of diabetes knowledge compared to their counterparts who were not using insulin in their treatment. The Australia-based participants who were not using insulin treatment practised more frequent exercises ($p < 0.001$) compared to the Malaysia-based participants who were not using insulin in their treatment. Regardless of which country they are from, people using insulin treatment practised significantly more frequent blood glucose testing compared to their counterparts who were not using insulin in their treatment. In terms of QoL, people who used insulin in their treatment experienced higher mean impacts of diabetes compared to those who were not using insulin within the Australia-based sample ($p = 0.048$) and Malaysia-based sample ($p = 0.014$).

Table 6
Post-hoc Tukey Test for the Insulin and Non-Insulin Users

Study variables	*Differences between groups, p*-value					
	ANI vs AI	ANI vs MNI	ANI vs MI	AI vs MNI	AI vs MI	MNI vs MI
Diabetes knowledge	<0.001	0.001	0.284	<0.001	0.017	<0.001
Attitudes	0.789	0.591	0.068	0.100	0.004	0.495
Self-management:						
Exercise	0.927	<0.001	0.001	<0.001	0.006	0.963
Blood glucose testing	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Foot care	0.189	0.430	0.753	0.001	0.825	0.066
QoL:						
Impact	0.048	0.031	0.977	<0.001	0.181	0.014

ANI = Australia-based insulin users, AI = Australia-based non-insulin users, MNI = Malaysia-based non-insulin users, MI = Malaysia-based insulin users

DISCUSSION

To date, research has been conducted in population-based samples. There has been a range of research on diabetes knowledge, attitudes, self-management and QoL among people with T2DM, but each study is typically limited to the health-care setting of a single country. In addition, research on these variables, especially diabetes self-management and QoL, has been limited to western countries, and thus, patterns of these variables in non-western cultures remain largely unknown. In comprehensive reviews of diabetes research such as those related to diabetes self-management (Clar, Barnard, Cummins, Royle, & Waugh, 2010) and diabetes QoL (Cochran & Conn, 2008; Schram, Baan, & Pouwer, 2009), the authors found no reports of comparison between two cultures or countries. Thus, this is the first QoL study identified in the literature that examined differences between western and non-western cultures in terms of knowledge, attitudes, self-management and QoL among people with T2DM. It is important to compare critical health-care variables across cultures so as to identify limitations in some countries that could be reduced by adopting practices that are employed in other countries.

In the present comparison, we found that the Australia-based participants with T2DM had a significantly higher level of understanding about diabetes than their Malaysia-based counterparts. This could probably be due to the fact that the Australia-based participants scored higher on the DKN measure of knowledge about diabetes because more Australia-based (36.4%) participants were either college graduates or university graduates than the Malaysia-based participants (24.0%). This is consistent with the research by Maxwell, Hunt, and Bush (1992), who reported a positive correlation between education level and DKN scores. There is either a lack of research or no consistency among the research findings on the relationship between genders and diabetes knowledge. In some developing countries, females were found to have lower levels of knowledge about diabetes than males (Kamel et al., 1999; Rafique et al., 2006). Some researchers, however, found no significant differences between males and

females in terms of their knowledge about diabetes (Tham, Ong, Tan, & How, 2004; Garrett et al., 2005). In the present study, the Australia-based females had a significantly higher level of knowledge about diabetes than males, while the Malaysia-based females and males had almost the same mean level of knowledge about diabetes.

In measuring the attitudes to T2DM, the *t*-test results indicated that the mean scores of attitudes varied significantly across cultures. The Australia-based participants with T2DM scored significantly higher in their mean attitudes than their Malaysia-based counterparts. This finding also indicated that the Australia-based sample showed a more positive attitude toward their illness condition than their Malaysia-based counterparts. This study also revealed that the Australia-based females reported higher levels in attitudes than their male counterparts. In their study, Nielsen et al. (2006) explained that female participants showed more adaptive attitudes towards lifestyle modification as part of diabetic management compared to males. In the present study, however, there was no significant difference between Malaysia-based females and males in their attitude scores.

In self-management, as reflected in their responses to the SDSCA, the results indicated that the Australia-based participants practised exercises more regularly within the week compared to the Malaysia-based participants. Lack of culturally-appropriate exercise facilities available to the community in Kelantan, especially for women, may contribute to low motivation for exercise among people with T2DM. For example, in a study of Arab women, where the majority are Muslim, researchers found that the lack of culturally-sensitive exercise facilities and socio-cultural norms restricted the participants' motivation to engage in exercises (Ali, Baynouna, & Bernsen, 2010). Moreover, some barriers to exercises have been found among Malaysians of Malay ethnic background with T2DM. For example, in a qualitative study conducted among the Malay participants with T2DM in Malaysia, Ali and Jusoff reported that one of the participants explained that she could not do more exercise because she prepared food for her family (Ali & Jusoff, 2010). Other researchers reported that some people with T2DM even perceived exercises that increased heart rate and breathlessness to be inducing illness states, so they regarded such exercises as something they should try to avoid (Lawton, Ahmad, Hanna, Douglas, & Hallowell, 2006).

In their responses to the SDSCA, the Australia-based participants also reported that they checked their blood glucose level and foot care more frequently than the Malaysia-based participants during the week prior to completing the self-report measures. Foot care seems to be the least practised self-management practice by Australia-based participants compared to other self-management regimens. Tapp et al. (2004) observed that foot screening appears to be poor in Australia, with less than one-half of the T2DM population having a regular examination for foot complications. The results indicated that self-management of blood glucose testing and foot care among T2DM people differed significantly depending on whether they were on insulin treatment. The Post-hoc test showed that people with T2DM, who were using insulin in their treatment, had a significantly higher mean number of days on which they practised self-management of blood glucose testing and foot care than those who were not using insulin treatment. This result is consistent across both cultures. However, the Malaysia-based participants had a significantly lower mean number of days of practising blood glucose testing

than observed in the Australia-based sample. This is primarily due to the fact that most of the Malaysia-based participants did not have their own blood glucose test equipment when the data were collected; thus, they were only able to do blood glucose testing during their monthly clinic visit for diabetes. Evidence indicates that self-monitoring of blood glucose is a cornerstone of the treatment of diabetes (Harris, 2001). Thus, the instrument for measuring blood glucose that can be used for self-monitoring should be used as widely as possible by people with T2DM.

In the present study, males in both countries reported to experiencing a higher impact of diabetes than females within their cultural groups. The Malaysia-based females experienced the lowest impact of diabetes among genders in both cultures. A possible reason for this situation may be that the majority of the Malaysia-based participants hold particular religious beliefs that influence their view of diabetes illness in their life (e.g., Mir & Sheikh, 2010; Padela, Killawi, Forman, DeMonner, & Heisler, 2012; Peterson, Nayda, & Hill, 2012). Previous study indicated that there is an association between spirituality and depression in adults with T2DM, which revealed that females with lower educational levels and lower income had greater spirituality, and greater spirituality was associated with less depression (Lynch, Hernandez-Tejada, Strom, Egede, 2012). This indicates that having stronger religious belief plays an important role in self-reported well-being or depression, which may influence QoL, especially in terms of impact, and reflect a negative response to living with T2DM. The Malaysia-based participants were drawn from the state of Kelantan, where Malays with strong Muslim belief are the dominant group, forming 95 percent of the total population (Kelantan Properties, 2006). Thus, it is not surprising that the Malaysia-based participants held spiritual belief related to their religious background. This may explain why females reported lower mean impacts of QoL. However, spirituality or religious belief was not explicitly examined in the present research and evidence regarding faith-based personal motivation (fatalism, religiosity, spirituality) as a mechanism for greater QoL among people with diabetes is still very limited (Lynch et al., 2012). Future research using qualitative methods would be an interesting approach to explore this issue.

Undertaking insulin therapy can be a significant event in the life of people with T2DM. In the Malaysia-based and Australia-based samples in the present research, the participants being treated with insulin reported to have higher mean scores of impact within both cultures. This indicates that people treated with insulin experienced greater impacts of diabetes on QoL compared to those who were not using insulin in their treatment. This result is consistent with other studies (Bradley, Todd, Gorton, Symonds, Martin, & Plowright, 1999; Davis, Clifford, & Davis, 2001). Davis et al. (2001) reported that after adjusting for variables, including diabetes duration, fasting plasma glucose, efficiency in English and ethnicity, insulin treatment was significantly associated with impacts. Davis et al. (2001) also explained that insulin therapy could be a significant adjustment in the life of people with T2DM and this method is relatively complex and invasive. Furthermore, people tend to interpret the use of insulin as a failure on the part of individuals with T2DM. For instance, people with T2DM perceived shifting to insulin treatment as a threat, which emphasised that they had failed in managing their condition through lifestyle changes and oral medication (Gomersall, Madill, & Summers, 2011). This may contribute to the lower QoL reported among those who were on insulin treatment, such

as reported in a study by Shiu et al. in the Hong Kong Chinese population (Shiu, Thompson, & Wong, 2008). Shiu et al. also found that people using insulin treatment had poorer QoL than those who were not prescribed insulin to manage their diabetes.

In the present research, the sample from Australia comprised people who live in a major city (Melbourne), whereas the sample from Malaysia consisted of people who live in a more provincial state in Malaysia (Kelantan). Thus, conclusions about these two samples need to be carefully drawn. Meanwhile, the statistical differences between the Australia-based sample and the Malaysia-based sample might be explained by the differences in the health and medical provisions in the different locations where the samples were examined. We acknowledge that there were some limitations involved in choosing the samples from two different countries because the health system in each country is totally different. However, HUSM is a University's hospital where the health system is considered as advanced within the Malaysian context, given that the University has a diabetes clinic that specialises in providing treatment to people with diabetes in Kelantan, and it should be comparable to the health system in Melbourne, Australia, in terms of its diabetes care.

There are various benefits and strengths of the research comparing Australia-based and Malaysia-based people with diabetes reported in this study. This is the first research that to have identified and compared the levels of diabetes knowledge, attitudes, self-management and QoL among two cultural groups. This increases the understanding and reveals comparisons on the current situation of diabetes knowledge, attitudes to T2DM, self-management of T2DM and QoL of T2DM among people with T2DM who live in a developed country, Australia, and a developing country, Malaysia. For example, as mentioned earlier, this research has provided insights that people with T2DM living in a more advanced country with better facilities of medical care do not necessary experience less impacts of their illness compared to those with T2DM residing in a less advanced country with less opportunity of receiving the most sophisticated diabetes care for their illness. However, that less impact in the Malaysia-based participants reflects lower expectations of care and of health perhaps based on the lack of knowledge. The results suggest that there could be great benefits in exploring the basis of differences between the variables in this study. Among other, this could lead to changes in health-care practices in countries like Malaysia that should lead to improved management of T2DM.

CONCLUSION

The present research has uncovered interesting differences between two samples derived from differing cultures. These results suggest that differences can appear in two cultures or people from different countries in terms of how much they know about diabetes, how positive their attitude is in regard to their illness, how regularly they self-manage some aspects of their treatment regimens and their perceptions of the impacts of diabetes on their QoL. Although Malaysia-based participants scored significantly lower in the areas of diabetes knowledge, attitudes, exercise, blood glucose testing, and foot care, they experienced significantly lower impacts of diabetes than their Australia-based counterparts. This might be due to the lifestyle differences, social-economic status and religious beliefs and practices held within Malaysian culture, which remain an area for future research in the cross-cultural study of diabetes.

REFERENCES

Ali, H. I., Baynouna, L. M., & Bernsen, R. M. (2010). Barriers and facilitators of weight management: perspectives of Arab women at risk for type 2 diabetes. *Health and Social Care in the Community, 18*(2), 219-228.

Ali, S. M., & Jusoff, K. (2009). Barriers to optimal control of type 2 diabetes in Malaysian Malay patients. *Global Journal of Health Science, 1*(2), 106-118.

AIHW. (2011). *How common is diabetes*? Australian Institute of Health and Welfare. Retrieved from http://www.aihw.gov.au/how-common-is-diabetes/

American Diabetes Association. (1999). Quality of life in type 2 diabetic patients is affected by complications but not by intensive policies to improve blood glucose or blood pressure control (UKPDS 37). UK Prospective Diabetes Study Group. *Diabetes Care, 22*(7), 1125-1136.

Beeney, L. J., Dunn, S. M., & Welch, G. (1994). Measurement of diabetes knowledge - The development of the DKN scales. In C. Bradley (Ed.), *Handbook of psychology and diabetes* (pp. 159-190). Australia: Harwood Academic Publishers.

Bradley, C., Todd, C., Gorton, T., Symonds, E., Martin, A., & Plowright, R. (1999). The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. *Quality of Life Research, 8*(1-2), 79-91.

Chow, C. C., Tsang, L. W., Sorensen, J. P., & Cockram, C. S. (1995). Comparison of insulin with or without continuation of oral hypoglycemic agents in the treatment of secondary failure in NIDDM patients. *Diabetes Care, 18*(3), 307-314.

Clar, C., Barnard, K., Cummins, E., Royle, P., & Waugh, N. (2010). Self-monitoring of blood glucose in type 2 diabetes: Systematic review. *Health Technology Assessment, 14*(12), 1-140. doi: 10.3310/hta14120

Cochran, J., & Conn, V. S. (2008). Meta-analysis of quality of life outcomes following diabetes self-management training. *The Diabetes Educator, 34*(5), 815-823.

Davis, T. M., Clifford, R. M., & Davis, W. A. (2001). Effect of insulin therapy on quality of life in type 2 diabetes mellitus: The fremantle diabetes study. *Diabetes Research and Clinical Practice, 52*(1), 63-71.

DAV. (2011). *Diabetes epidemic 10 years on 2001-2011*. Diabetes Australia Victoria. Retrieved from http://www.diabetesvic.org.au/media-centre/diabetes-epidemic

DCCT Research Group. (1988). Reliability and validity of a diabetes quality-of-life measure for the diabetes control and complications trial (DCCT). *Diabetes care, 11*(9), 725-732.

Garrett, N., Hageman, C. M., Sibley, S. D., Davern, M., Berger, M., Brunzell, C.,...Richards, S. W. (2005). The effectiveness of an interactive small group diabetes intervention in improving knowledge, feeling of control, and behavior. *Health Promotion Practice, 6*(3), 320-328.

Goddijn, P. P., Bilo, H. J., Feskens, E. J., Groeniert, K. H., van der Zee, K. I., & Meyboom de Jong, B. (1999). Longitudinal study on glycaemic control and quality of life in patients with type 2 diabetes mellitus referred for intensified control. *Diabetes Medicine, 16*(1), 23-30.

Gomersall, T., Madill, A., & Summers, L. K. M. (2011). A metasynthesis of the self-management of type 2 diabetes. *Qualitative Health Research, 21*(6), 853-871.

Haas, L., Maryniuk, M., Beck, J., Cox, C. E., Duker, P., Edwards, L., ... & Youssef, G. (2013). National standards for diabetes self-management education and support. *Diabetes care*, *36*(Supplement 1), S100-S108.

Hanninen, J., Takala, J., & Keinanen-KiuKaanniemi, S. (1998). Quality of life in NIDDM patients assessed with the SF-20 questionnaire. *Diabetes Research and Clinical Practice, 42*(1), 17-27.

Harris, M. I. (2001). Frequency of blood glucose monitoring in relation to glycemic control in patients with type 2 diabetes. *Diabetes Care, 24*(6), 979-982.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). CA: Sage, Thousand Oaks.

Kelantan Properties. (2006). *Kelantan - case study.* Kelantan Properties. Retrieved from http://kelantan-properties.com/kelantan.htm

Kamel, N. M., Badawy, Y. A., El-Zeiny, N. A., & Merdan, I. A. (1999). Sociodemographic determinants of management behaviour of diabetic patients part II. Diabetics' knowledge of the disease and their management behavior. *Eastern Mediterranean Health Journal, 5*(5), 974-983.

Kueh, Y. C. (2014). *Modelling knowledge, attitudes, self-management, and quality of life in type 2 diabetes mellitus.* (Doctoral dissertation). Victoria University, Melbourne, Australia.

Kueh, Y. C., Morris, T., & Ismail, A. A. S. (2014). Assessment of attitudes, self-management, and quality of life of people with type 2 diabetes mellitus in the malay population. *Asian Pacific Journal of Health Science, 1*(4), 550-558.

Lawton, J., Ahmad, N., Hanna, L., Douglas, M., & Hallowell, N. (2006). I can't do any serious exercise: barriers to physical activity amongst people of Pakistani and Indian origin with type 2 diabetes. *Health Education Research, 21*(1), 43-54.

Lynch, C. P., Hernandez-Tejada, M. A., Strom, J. L., & Egede, L. E. (2012). Association between spirituality and depression in adults with type 2 diabetes. *The Diabetes Educator, 38*(3), 427-435.

Matsumoto, D., Yoo, S., & Fontaine, J. (2008). Mapping expressive differences around the world: the relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology, 39*(1), 55-74.

Maxwell, A. E., Hunt, I. F., & Bush, M. A. (1992). Effects of a social support group, as an adjunct to diabetes training, on metabolic control and psychosocial outcomes. *The Diabetes Educator, 18*(4), 303-309.

Mir, G., & Sheikh, A. (2010). Fasting and prayer don't concern the doctors... they don't even know what it is: communication, decision-making and perceived social relations of Pakistani Muslim patients with long-term illnesses. *Ethnicity and Health, 15*(4), 327-342.

Misra, R., & Lager, J. (2008). Predictors of quality of life among adults with type 2 diabetes mellitus. *Journal of Diabetes and Its Complications, 22*(3), 217-223.

Nam, S., Chesla, C., Stotts, N. A., Kroon, L., & Janson, S. L. (2011). Barriers to diabetes management: patient and provider factors. *Diabetes Research and Clinical Practice*, *93*(1), 1-9.

Nielsen, A. B. S., Olivarius, N. D. F., Gannik, D., Hindsberger, C., & Hollnagel, H. (2006). Structural personal diabetes care in primary health care affects only women's HbA1c. *Diabetes Care, 29*(5), 963-969.

Padela, A. I., Killawi, A., Forman, J., DeMonner, S., & Heisler, M. (2012). American Muslim perceptions of healing: Key agents in healing, and their roles. *Qualitative Health Research, 22*(6), 846-858.

Peterson, S., Nayda, R., & Hill, P. (2012). Muslim persons' experiences of diabetes during Ramadan: information for health professionals. *Advances in Contemporary Community and Family Health Care, 41*(1), 41-47.

Pibernik-Okanovic, M., Szabo, S., & Metelko, Z. (1998). Quality of life following a change in therapy for diabetes mellitus. *Pharmacoeconomics, 14*(2), 201-207.

Rafique, G., Azam, S. I., & White, F. (2006). Diabetes knowledge, beliefs and practices among people with diabetes attending a university hospital in Karachi, Pakistan. *Eastern Mediterranean Health Journal, 12*(5), 590-598.

Samuel-Hodge, C. D., Headen, S. W., Skelly, A. H., Ingram, A. F., Keyerserling, T. C., & Jackson, E. J. (2000). Influences on day-to-day management of type 2 diabetes among African American women. *Diabetes Care, 23*(7), 928-933.

Schram, M. T., Baan, C. A., & Pouwer, F. (2009). Depression and quality of life in patients with diabetes: A systematic review from the European Depression in Diabetes (EDID) Research Consortium. *Current Diabetes Reviews, 5*(2), 112-119.

Shiu, A. T. Y., Thompson, D. R., & Wong, R. Y. M. (2008). Quality of life and its predictors among Hong Kong Chinese patients with diabetes. *Journal of Clinical Nursing, 17*(5A), 125-132.

Solli, O., Stavem, K., & Kristiansen, I. S. (2010). Health-related quality of life in diabetes: The associations of complications with EQ-5D scores. *Health and quality of life outcomes*, *8*(1), 1-18. doi: 10.1186/1477-7525-8-18.

Tapp, R. J., Zimmet, P. Z., Harper, C. A., Courten, M. P., Balkau, B., & McCarty, D. J. (2004). On behalf of the Australian Diabetes Obesity and Lifestyle Study Group: Diabetes care in an Australian population. *Diabetes Care, 27*, 688-693.

Tham, K. Y., Ong, J. J. Y., Tan, D. K. L., & How, K. Y. (2004). How much do diabetic patients know about diabetes mellitus and its complications? *Annals Academy Medicine Singapore, 33*(4), 503-509.

Toobert, D. J., & Glasgow, R. E. (1994). Assessing diabetes self-management: The summary of diabetes self-care activities questionnaire. In C. Bradley (Ed.), *Handbook of psychology and diabetes* (pp. 351-378). Australia: Harwood Academic Publishers.

Toobert, D. J., Hampson, S. E., & Glasgow, R. E. (2000). The summary of diabetes self-care activities measure. *Diabetes Care, 23*(7), 943-950.

Tripp-Reimer, T., Choi, E., Kelley, L. S., & Enslein, J. C. (2001). Cultural barriers to care: Inverting the problem. *Diabetes Spectrometry, 14*(1), 13-22.

Welch, G., Beeney, L. J., Dunn, S. M., & Smith, R. B. W. (1996). The development of the diabetes integration scale: a psychometric study of the ATT39. *Multivariate Experimental Clinical Research, 11*(2), 75-88.

Wild, S., Roglic, G., Green, A., & King, H. (2004). Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. *Diabetes Care 27*(5), 1047-1053. doi: 10.2337/diacare.27.5.1047

WHO. (2006). *Mortality country fact sheet 2006*. World Health Organization. Retrieved from http://www.who.int/whosis/mort/profiles/mort_wpro_mys_malaysia.pd

Zanariah, H., Chandran, L. R., Mohamad, W. B. W., Nazaimoon, W. M. W., Letchuman, G. R., Jamaiyah, H., ... Rodi, I. M. (2009). *Prevalence of diabetes mellitus in Malaysia in 2006 - results of the 3rd national health and morbidity*. Retrieved from http://www.crc.gov.my/documents/abstract/prevalenceOfDiabetes_ppt.pdf

Pertanika J. Sci. & Technol. 25 (1): 139 - 152 (2017)

SCIENCE & TECHNOLOGY

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

Hybrid Genetic Algorithm in the Hopfield Network for Logic Satisfiability Problem

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

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

ABSTRACT

In this study, a hybrid approach that employs Hopfield neural network and a genetic algorithm in doing k-SAT problems was proposed. The Hopfield neural network was used to minimise logical inconsistency in interpreting logic clauses or programme. Hybrid optimisation made use of the global convergence advantage of the genetic algorithm to deal with learning complexity in the Hopfield network. The simulation incorporated with genetic algorithm and exhaustive search method with different k-Satisfiability (k-SAT) problems, namely, the Horn-Satisfiability (HORN-SAT), 2-Satisfiability (2-SAT) and 3-Satisfiability (3-SAT) will be developed by using Microsoft Visual C++ 2010 Express Software. The performance of both searching techniques was evaluated based on global minima ratio, hamming distance and computation time. Simulated results suggested that the genetic algorithm outperformed exhaustive search in doing k-SAT logic programming in the Hopfield network.

Keywords: Genetic Algorithm, Exhaustive Search, Hopfield network, Satisfiability, Logic Programming, HORN-SAT, 3-SAT, 2-SAT

INTRODUCTION

Artificial intelligence is one of the most eminent and staple fields in Mathematics, Physics and Computer Science (Bekir & Vehbi, 2010). Strictly speaking, the integration of neural network, logic programming, satisfiability problem and neuro-searching approach as a single hybrid network is a brand new paradigm in artificial intelligence. There are numerous types of neural networks, from modest to intricate, such as the Hopfield neural network introduced by Hopfield and Tank (1985). The Hopfield neural network is a simple recurrent network which can serve as an efficient associative memory and store definite

Article history:
Received: 26 January 2016
Accepted: 30 August 2016

E-mail addresses:
iwanmaidin@gmail.com (Mohd Shareduwan Mohd Kasihmuddin),
asyrafalvez@live.com (Mohd Asyraf Mansor),
saratha@usm.my (Saratha Sathasivam)
*Corresponding Author

ISSN: 0128-7680

memories in a manner rather similar to the brain (Rojas, 1999; Sathasivam et al., 2013). Moreover, it is a branch of neural network that has been applying a vast Mathematical problems such as Travelling Salesperson problem (TSP) and hard satisfiability problem (Haykin, 1992). Next, logic programming can be treated as a problem in combinatorial optimisation perspective (Kowalski, 1979). Hence, it can be implemented and assimilated in a neural network to hunt desired solutions (Hamadneh et al., 2013). Recently, the conventional and common model is Wan Abdullah's logic programming (Wan Abdullah, 1993). In this paper, we will combine the advantages of the Hopfield network, logic programming and neuro-searching methods to do the satisfiability problem.

Traditionally, neuro-searching methods, such as exhaustive search and metaheuristic, can be implemented as a mechanism of doing the k-SAT problems. The most widely used technique is exhaustive search as it is a simple algorithm (Hooker, 2005). Theoretically, this technique considers the whole search space in order to check the clause satisfaction for the k-SAT problems, namely, Horn-SAT, 2-SAT and 3-SAT. However, the exhaustive search can be applied only if the problem size or the number of clause is limited (Mark & Lee, 1992). Another limitation is that the exhaustive search typically consumes more time to complete the whole searching process (Tobias & Walter, 2004; Kaushik, 2012). In this paper, we proposed a metaheuristic approach, the genetic algorithm (GA), to obtain the satisfied interpretation within acceptable timescales. On the other hand, genetic algorithm is an evolutionary algorithm that involves iterative procedures combining the exploration and exploitation process within any search space (Holland, 1975; John, 2005; Aiman & Asrar, 2015). Hence, sense of balance in both processes is vital to improve the convergence of the algorithm (Siddique, & Adeli, 2013).

This paper is organised as follows. Section 2 introduces the important concept of the Hopfield neural network, logic programming, satisfiability problem, exhaustive search (ES) and genetic algorithm (GA). In section 3, the fundamental theory of k-satisfiability (k-SAT) problems, namely, Horn-Satisfiability (HORN-SAT), 2-Satisfiability (2-SAT) and 3-Satisfiability are discussed. Section 4 covers a brief discussion of neuro-logic, which basically revolves around the Hopfield neural network and logic programming. Meanwhile, section 5 presents the neuro-searching methods involved in this research including Exhaustive search (ES) and Genetic algorithm (GA). In section 6, the theory implementation of the networks is discussed. Finally, sections 7 and 8 enclose the experimental results and offer conclusion of this research.

SATISFIABILITY (SAT) PROBLEM

Satisfiability or SAT is a rudimentary problem in computer science. The problem is to determine whether a truth assignment to variables appearing in a Boolean formula ϕ is satisfied (Kowalski, 1979). Therefore, a Boolean formula is satisfiable if an assignment of true and false values renders the entire expression true. For a problem of size n, there will be 2^n such assignments and l literals to set for each assignment (Sathasivam & Sagir, 2014), in which such an approach requires $O(l.2^n)$ operations (Gu, 1999). Hence, SAT is an NP-complete problem in general. For instance, the satisfiability problem concerns Boolean variables or expressions in conjunctive normal form (CNF). CNF comprises of conjunction of clauses, where clauses are disjunctions of literal (Sathasivam *et al.,* 2013). Meanwhile, literal is a variable or its negation. For example:

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

Here x_1, x_2, x_3, x_4 are Boolean variables to be assigned, $\neg$ which refer to negations (logical NOT), $\vee$ means negations (logical OR), $\wedge$ means negations (logical AND). One may note that the formula above is satisfiable when $x_1 = true, x_2 = false, x_3 = false, x_4 = true$, where it takes on the value of true. However, if a formula is not satisfiable, it is called unsatisfiable, which means that it takes on the value false on any value of its variable. The basic form of k-SAT is simplified, as follows:

$$\text{k-SAT} : P \rightarrow \{\text{YES}, \text{NO}\} \quad (2)$$

2-Satisfiability (2-SAT)

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

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

The Boolean values are $\{1, -1\}$. Researchers have emphasised the True and False in the neural networks by 1 and -1. Due to this, the goal of the 2-SAT problem is to determine whether there exists an assignment of truth values to variables that makes the following formula satisfiable.

$$P = \bigwedge_{i=1}^{n} C_i \quad (3)$$

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

$$C_i = \bigvee_{i=1}^{n} (x_i, y_i) \quad (4)$$

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

3-Satisfiability (3-SAT)

In this paper, we emphasize a paradigmatic NP-complete problem namely 3-Satsifiability (3-SAT). Generally speaking, 3-SAT can be defined as a formula in conjunctive normal form, where each clause is limited to at most or strictly three literals (Vilhelm et al., 2005), thus

the problem is an example of a non-deterministic problem (Tobias & Walter, 2004). In our analysis, the following 3-SAT logic programme consisting of 3 clauses and 3 literals will be used. For instance:

$$P = \left(A \vee B \vee \overline{C}\right) \wedge \left(\overline{A} \vee \overline{B} \vee C\right) \wedge \left(\overline{A} \vee B \vee D\right) \tag{5}$$

We represent the above 3 CNF formula with P. The formula can be in any combination as the number of atoms can be varied, except for the literals that are strictly equal to 3, which is vital for the combinatorial optimisation problem. Thus, the higher number of literals in each clause will increase the possibilities or chances for a clause to be satisfied. The general formula of 3-SAT for conjunctive normal form (CNF):

$$P = \bigwedge_{i=1}^{n} Z_i \tag{6}$$

Hence, the value of k denotes the number of satisfiability. In our case, k-SAT is 3-SAT.

$$Z_i = \bigwedge_{j=1}^{k} (x_{ij}, y_{ij}, z_{ij}), k > 3 \tag{7}$$

Horn-Satisfiability (HORN-SAT)

Horn-Satisfiability can be defined as a clause with at most one positive literal. It comprises of any individual atom in the head and literals in their body. Therefore, the general form of horn clauses is illustrated as follows:

$$P = C_i \leftarrow \bigwedge_{j=1}^{m} D_j \tag{8}$$

For the HORN-SAT clause, we can construct the logic in a form of $P = C_1, C_2, \mathrm{C}_3 ...., \mathrm{C}_n \leftarrow D_1, D_2, \mathrm{D}_3 ...., \mathrm{D}_m$. In this study, P should be satisfied. The truth assignment is vital as in 3-SAT previously. Thus, the HORN-SAT formula comprised at most one positive literal is as follows:

$$P = \left(A \vee \neg B \vee \neg C\right) \wedge \left(D \vee \neg B\right) \wedge C \tag{9}$$

Works on HORN-SAT have been done by Wan Abdullah (1992) and Sathasivam (2010). One of the important features of HORN-SAT is that the formula is always satisfiable.

NEURO-LOGIC IN THE HOPFIELD NEURAL NETWORK

The Hopfield Model

The Hopfield model is a standard model for content addressable memory (CAM). The units in Hopfield nets are binary threshold unit (Haykin, 1992), which can only take binary values such as 1 and -1. The possible definitions for unit I's activation, a_i are:

$$a_i = \begin{cases} 1 \quad if \sum_j W_{ij} S_j > \xi_i \\ -1 \; Otherwise \end{cases} \tag{10}$$

Where, W_{ij} is the connection strength from units j to i. S_j is the state of unit j and ξ_i is the threshold of unit i. The connection in the Hopfield net typically has no connection with itself, $W_{ij} = 0$ and connections are symmetric or bidirectional (Sathasivam et al., 2013). The system consists of N formal neurons, each is described by an Ising variable. Neurons are bipolar. $S_i \in \{1, -1\}$ is obeying the dynamics $S_i \to \text{sgn}(h_i)$, where the local field h_i is. The connection model can be generalised to include higher order connection. This modifies the field to:

$$h_i = \sum_j W_{ij}^{(2)} S_j + J_i^{(1)} \tag{11}$$

The weight in the Hopfield network is always symmetrical. The updating rule maintains:

$$S_i(t+1) = \text{sgn}\left[h_i(t)\right] \tag{12}$$

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

$$E = ... - \frac{1}{2}\sum_i \sum_j W_{ij}^{(2)} S_i S_j - \sum_i W_i^{(1)} S_j \tag{13}$$

Logic Programming in the Hopfield network

In essence, logic programming can be seen as a problem in combinatorial optimisation and it can be carried out on a Hopfield network (Sathasivam et al., 2013). Furthermore, this can be done by using the neurons to store the truth values of the literal and writing a cost function, which is minimised when all clauses are satisfied (Wan Abdullah, 1993). In other words, the main task is to find the 'models' corresponding to the given logic programme. The fundamental of the Hopfield network in doing logic programming was brought up due to its unique content on addressable memory properties.

Implementation of k-SATGA in the Hopfield neural network (HNN-kSATGA)

i. Translate all the k-SAT clauses into Boolean algebra. Identify a neuron to each ground neuron.

iii. Initialise all connection strengths to zero.

iv. Derive a cost function that is associated with the negation of all k-SAT clauses. For example, $X = \frac{1}{2}(1 + S_X)$ and $\overline{X} = \frac{1}{2}(1 - S_X)$. $S_X = 1$ (True) and $S_X = -1$ (False). Multiplication represents CNF and addition represents DNF.

v. Compare the cost function with energy, E, by obtaining the values of the connection strengths.

vi. Check clause satisfaction by using k-SATGA. Satisfied clauses will be stored.

vii. Randomise the states of the neurons. The network undergoes a series of network relaxation. Calculate the corresponding local field $h_i(t)$ of the state. If the final state is stable for 5 runs, we consider it as the final state.

viii. Find the corresponding final energy E of the final state by using lypunov equation. Verify whether the final energy obtained is a global minimum energy or local minima. Calculate the corresponding hamming distance and computation time.

THE NEURO-SEARCH TECHNIQUES

Exhaustive search algorithm

Strictly speaking, exhaustive search is the simplest algorithm but can be computationally super expensive. In this algorithm, it exhaustively searches for the entire possible clause even though the search space is getting tremendous. The main advantage of this algorithm is the guarantee to obtain a solution (satisfied clause) by taking into a consideration the entire search space (Rojas, 1999). However, the exhaustive search consumes more computation time or CPU time to hunt for the satisfied interpretation completely (Kaushik, 2012; Asrar & Aiman, 2015). In exhaustive search, we will check the clause satisfaction directly for the k-SAT problem until a satisfying one is found (Tobias & Walter, 2004). In our case, we are required to seek for the satisfied clause during the training phase for HORN-SAT, 2-SAT and 3-SAT problem.

In this paper, we emphasise the complexity of the network as we venture with a higher number of neurons. Hence, the computation time will be slower if the complexity of the network increases. Generally for k-SAT problem, there are potentially 2^n satisfying assignments. Thus, the run-time complexity is equivalent to $O(2^n)$. The correct interpretation will be stored into Hopfield's brain as content addressable memory (CAM). In this paper, we will implement this algorithm together with the Hopfield neural network, logic programming and satisfiability problem.

Genetic algorithm

Genetic algorithm is a staple computational paradigm inspired from the Darwin's model, namely, survival for the fittest model. Darwin stated that the survival of an organism can be maintained through the processes of reproduction, crossover and mutation (Holland, 1975). Hence, it can be implemented in the Mathematical model and become one of good heuristic methods. For instance, every generation is represented by an array of bit strings analogous to the chromosomes of DNA. The core impetus of our approach is to hunt for the fittest assignment or satisfied clause given any k-SAT clauses. Besides, the fittest assignment gives the maximum number of satisfied clauses, which depends on the number of satisfied clauses and can be calculated as follows:

$$f_{k-SAT}(x) = c_1(x) + c_2(x) + c_3(x) ..... + c_{total\,NC}(x) \tag{14}$$

Theoretically, HNN-kSATGA is able to scan and search for the satisfied interpretation with the highest fitness value systematically. On the other note, the procedures do not require complex Mathematics to execute and are easy to implement to solve any constrained optimisation

problem. In GA, the initial set of population will be generated by using bit string as a chromosome. Moreover, the fitness value will be computed according to the truth value of each chromosome. During the crossover process, two chromosomes were randomly selected and broken from a randomly selected crossover locus (John, 2005). In addition, a child chromosome is produced by linking the first part and second part of the parent chromosome. Correspondingly, the second child is formed by joining the second part of the first chromosome and the first part of the second chromosome (Luke, 2013). Thus, we can compute the fitness value for the newly formed chromosomes. The crossover operator mimics the biological combination between two single-chromosomes (haploid) in organisms (Rojas, 1999).

During mutation, a newly child chromosome is selected to form a set of bit strings. If the fitness value is still lower than the maximum fitness, the random bit will be improved during this stage. Besides that, the number of bits being complemented or flipped depends on the mutation rate. In this paper, we are using the crossover rate equal to 0.9. According to Aiman and Asrar (2015), the perfect mutation rate is usually less than 1. Then, the fitness will be computed for the new chromosome (bit strings). If the fitness value matches the maximum fitness, we print the output as our desired satisfied interpretation. Figure 1 shows the algorithm for this paradigm.

THEORY IMPLEMENTATION

For the implementation, firstly, random k-SAT clauses were generated. From there, the initial states were initiated for the neurons in the clauses. The network will evolve until the final state is reached. Once the programme has reached the final state, the neuron state is updated via equation (11). As soon as the network is relaxed to an equilibrium state, the final state obtained for the relaxed neuron is tested to determine whether it is a stable state. Furthermore, stable state will be considered, provided the state remains unchanged for five runs. According to Pinkas and Dechter (1995), letting an ANN to evolve will eventually lead to stable state where the energy function obtained does not change further. In this case, the corresponding final energy for the stable state will be calculated. If the difference between the final energy and the global minimum energy is within the given tolerance value, then the solution is considered as a global solution.

In addition, the simulations are performed on Microsoft Visual C++ 2010 Express Software. Each satisfiability problem will be repeated with 100 different combination neurons and each neuron combination will undergo 100 trials to reduce the statistical error, in which the selected tolerance value is 0.001. According to Sathasivam et al. (2013), 0.001 was selected because it gave a better performance than other values. Other than that, connection strength can be obtained by using either Sathasivam's method or Hebbian Rule. Both methods will result in similar strength (Sathasivam & Sagir, 2014) because both methods consider the same knowledge base (clauses).

RESULTS AND DISCUSSION

Global Minima Ratio

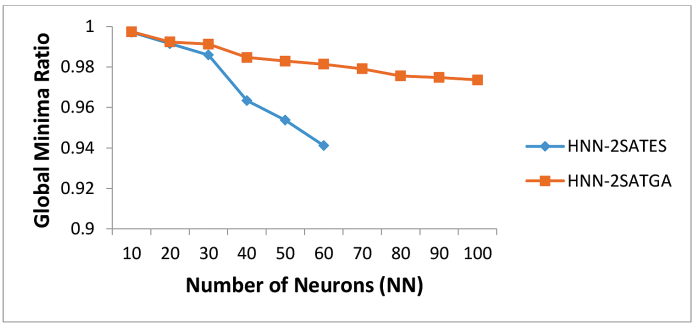


Figure 2. Global Minima Ratio for HNN-2SATES and HNN-2SATGA

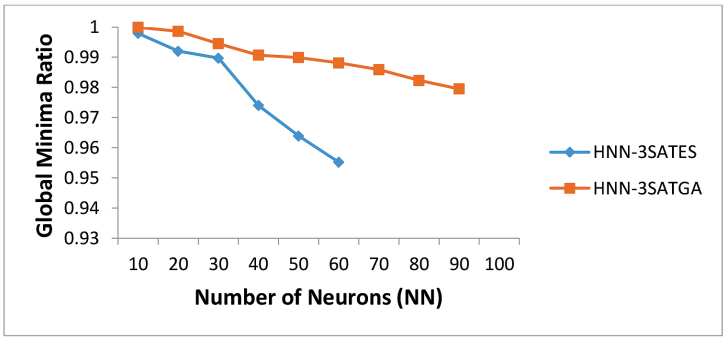


Figure 3. Global Minima Ratio for HNN-3SATES and HNN-3SATGA

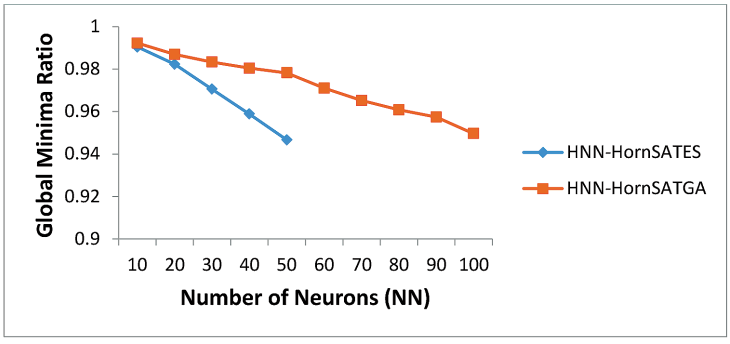


Figure 4. Global Minima Ratio for HNN-HornSATES and HNN-HornSATGe

The results of the HNN-kSATES and HNN-kSATGA are summarised in Figure 2 through Figure 4. Based on these figures, HNN-kSATGA outperforms HNN-kSATES in terms of global minima ratio. Also, HNN-kSATGA is able to recall more correct states compared to

HNN-kSATES. When the number of neurons increased, HNN-kSATGA is able to sustain more neurons. The limit for HNN-kSATES is 60 neurons. After 60 neurons, the network is stuck in trial and error state in a long period of time. The whole bitstring in exhaustive search method will be collapsed when one of the neuron-clauses is not satisfied. In addition, it might take a longer time for the network to search the correct neuron states and proceed with the relaxation state, as the network spends more time in training state. On the other hand, HNN-kSATGA can sustain more neurons because genetic algorithm reduces the complexity of the network to find the correct states. Apart from that, unsatisfied neuron clauses will be improved through crossover among the best offspring and undergo mutation until we found the correct neuron state. Besides, giving more relaxation time for the network will help the network to retrieve the state more effectively. Thus, less relaxation time will create spurious minima which will cause the retrieved solution to achieve local minima.

Hamming Distance

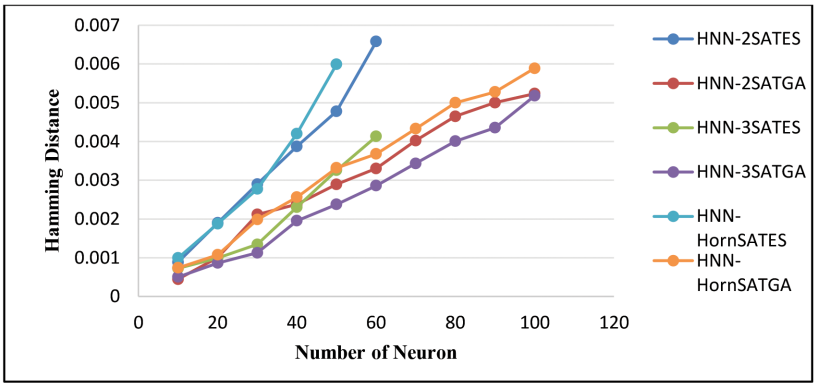


Figure 5. Hamming distance for different HNN- kSATGA and HNN-kSATES.

Figure 5 depicts the obvious variation in Hamming distance obtained by HNN-kSATGA and HNN-kSATES. In this study, Hamming distance is the closeness of bits between the training state and global state (retrieved state) of the neurons upon relaxation process. The plot shows that HNN-kSATGA consistently performed better than HNN-kSATES. Hence, the ability to recall accurate interpretations (training phase) and correct states (testing phase) was improved drastically by implementing Genetic algorithm (GA) in the Hopfield neural network for any k-SAT problem. This is due to the power of GA in ascertaining the satisfied clause, especially during the crossover stage, where the clause was being improved by certain rate to achieve the highest fitness value. Additionally, HNN-kSATGA would be able to recall the correct states that contributed to the lower hamming distance. Conversely, the exhaustive search algorithm emphasised the trial and error process during clause satisfaction process. When the complexity increased, HNN-2SATES and HNN-HORNSAT were able to sustain up to 50 neurons and HNN-3SATES with the limitation until 60 neurons. The main reason is due to the nature of exhaustive search that increases the computation burden in searching

for the correct neuron states. On the contrary, HNN-HornSATGA, HNN-2SATGA and HNN-3SATGA were able to sustain up to 100 neurons. Hence, the ability to sustain huge number of neurons is due to the special ability of GA that reduces the computation burden in hunting the correct states.

Computation Time

Table 1
Computation time for different HNN-kSATGA and HNN-kSATES.

Number of Neurons (NN)	Computation Time					
	HNN-2SATES	HNN-2SATGA	HNN-3SATES	HNN-3SATGA	HNN-HornSATES	HNN-HornSATGA
10	4	3	2	1	8	4
20	72	8	30	9	145	57
30	208	18	88	22	4566	149
40	759	33	202	46	33890	280
50	8036	53	4601	99	260953	372
60	75472	78	19455	134		455
70		111		268		652
80		143		371		899
90		181		404		1004
100		231		450		2345

The computation time or CPU time can be defined as the time taken for a logic programme to generate the global solutions including the training process. According to Table 1, the computation time for the HNN-Ksatga is faster than HNN-kSATES. For instance, the complexity of the network increased as the network got massive. We can see that the computational time increased when the number of neuron was getting higher. This is due to the condition when the network was getting larger and complex, the network was likely to get stuck in local minima and devour more computation time. As a consequence of these arguments, extra time was needed to relax to global solution as the number of neurons increased. Moreover, the neurons needed to jump enormous energy barrier to reach the global solutions. On a separate note, the training process by using exhaustive search usually consumes more computational time due to the trial and error processes in hunting for the satisfied interpretation. However, when genetic algorithm was implemented, the computation time was faster due to the crossover and mutation processes that turned the unsatisfied clause into satisfied clause systematically. The computation time obtained for HORN-SAT, 2-SAT and 3-SAT indicated that HNN-kSATGA outperformed HNN-kSATES in terms of computation time.

CONCLUSION

Inspired by the fundamental of artificial intelligent and nature inspired optimization, an optimal k-SAT model was developed in the Hopfield network. In general, genetic algorithm incorporated with the Hopfield network can sustain k-SAT patterns. Computer simulations were carried out to verify the ability of the Hopfield network doing k-SAT logic. HNN-kSATGA gives us global minima ratio of approximately 1 and hamming distance, which is approximately 0 compared to HNN-kSATES. On the basis of the above illustration, it has been shown in computer simulation that genetic algorithm successfully reduced the complexity of the network, produced more ideal solutions and had a smaller error compared to the traditional exhaustive search method in the Hopfield network.

REFERENCES

Abdullah, W. A. T. W. (1993). The logic of neural networks. *Physics Letters A*, *176*(3), 202-206.

Aiman, U., & Asrar, N. (2015). Genetic algorithm based solution to SAT-3 problem. *Journal of Computer Sciences and Applications*, *3*(2), 33-39.

Bekir, K., & Vehbi, A. O. (2010). Performance analysis of various activation functions in generalized MLP architectures of neural network. *International Journal of Artificial Intelligence and Expert Systems*, *1*(4), 111-122.

Fernandez, W. (2001). Random 2-SAT: Result and Problems. *Theoretical Computer Science*, *265*(1), 131-146.

Gu, J. (1999). The Multi-SAT algorithm. *Discrete Applied Mathematics*, *96*, 111-126.

Hamadneh, N., Sathasivam, S., Tilahun, S. L., & Ong, H. C. (2013). Prey-Predator Algorithm as a New Optimization Technique Using in Radial Basis Function Neural Network. *Research Journal of Applied Sciences*, *8*(7), 383-5.

Haykin, S. (1992). *Neural Networks: A Comprehensive Foundation*. New York: Macmillan College Publishing.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Michigan: The University of Michigan Press.

Hooker, J. N. (2005). Unifying local and exhaustive search. In L. Villasenor & A. I. Martinez (Eds.), *Proceeding of ENC 2005 - Sixth Mexican International Conference on Computer Science* (pp. 237-243). IEEE Press.

Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernatics*, *52*(3), 141-152.

Iwama, K. (1989). CNF Satisfiability Test by Counting and Polynomial Average Time. *SIAM, Journal of Computer, 4*(1), 385-391.

John, M. (2005). Genetic algorithm for modelling and optimization. *Journal of Computational and Applied Mathematics*, *184*, 205-222.

Kaushik, M. (2012). Comparative analysis of exhaustive search algorithm with ARPS algorithm for motion estimation. *International Journal of Applied Information Systems*, *1*(6), 16-19.

Kowalski, R. A. (1979). *Logic for Problem Solving*. New York: Elsevier Science Publishing.

Luke, S. (2013). *Essentials of metaheuristics* (2nd edition). United States: Lulu.

Mark, W. G., & Lee C. G. (1992). Routing in random multistage interconnections networks: Comparing exhaustive search, greedy, and neural network approaches. *International Journal of Neural System*, *2*(3), 125-142.

Pinkas, G., & Dechter, R. (1995). Improving energy connectionist energy minimization. *Journal of Artificial Intelligence Research*, *3*, 223-237.

Razgon, I., & O' Sullivan, B. (2009). Almost 2-SAT is Fixed Parameter Tractable. *Journal of Computer Sciences and System Sciences*, *75*(8), 435-450.

Rojas, R. (1999). *Neural Networks: A Systematic Introduction*. Berlin: Springer.

Sathasivam, S. (2010). Upgrading Logic Programming in Hopfield Network. *Sains Malaysiana*, *39*(1), 115-118.

Sathasivam, S., Ng, P. F., & Hamadneh, N. (2013). Developing agent based modelling for reverse analysis method. *Journal of Applied Sciences, Engineering and Technology*, *6*(22), 4281-4288.

Sathasivam, S., & Sagir, A. M. (2014). An Overview of Hopfield Network and Boltzmann Machine. *International Journal of Computational and Electronics Aspects in Engineering*, *1*(1), 20-26.

Siddique, N., & Adeli, H. (2013). *Computational Intelligence Synergies of Fuzzy Logic, Neural Network and Evolutionary Computing.* United Kingdom: John Wiley and Sons.

Tobias, B., & Walter, K. (2004). An improved deterministic local search algorithm for 3-SAT. *Theoretical Computer Science*, *329*(1), 303-313.

Vilhelm, D., Peter, J., & Magnus, W. (2005). Counting models for 2SAT and 3SAT formulae. *Theoretical Computer Science*, *332*(1), 265-291.

APPENDIX

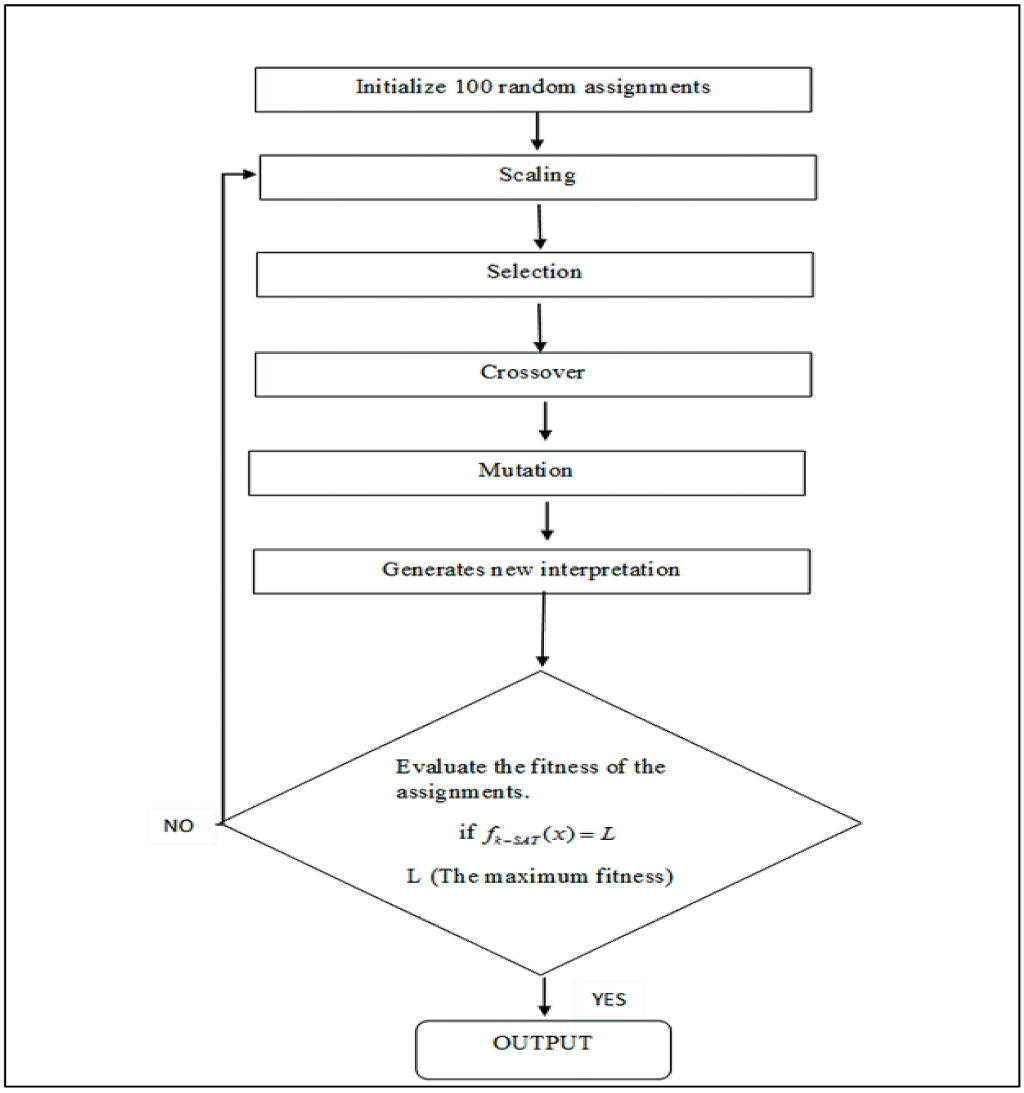


Figure 1. Flowchart for Genetic Algorithm

Pertanika J. Sci. & Technol. 25 (1): 153 - 160 (2017)

SCIENCE & TECHNOLOGY

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

Maximising Potential of Methane Production from Biogas for Power Generation

Tinia Idaty, Mohd Ghazi* and Ismail, Muhammad Nasir

Department of Chemical and Environmental Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Renewable energy projects in many developing countries need financial and legal back up from governments and other supportive bodies. There is a viable alternative to finite energy via usage of biomass waste a renewable energy source. The electrical energy production analysis on biomass waste presented in this paper is based on the experimental analysis carried out using the laboratory and pilot scale bioreactors. Electrical energy generated with oscillatory flow bioreactor (OFBR) was 10.12 kWh or up to 91% higher than the 10 L lab scale bioreactor (0.9 kWh), demonstrating that the novel OFBR has a great potential for renewable electricity. Also, the pilot scale plant achieves a value of 12.3 kWh, which the difference is not quite significant with that of OFBR. These results illustrate that the generation of the renewable electricity is feasible especially with the OFBR thereby achieving high methane potential during the treatment of manure and food waste. Nevertheless, energy recoveries should be enhanced to improve the entire operational performance.

Keywords: Biogas, electrical power generation, methane production, renewable energy

INTRODUCTION

The advancements of large feedlots for livestock have generated an economic opportunity for commercialised agribusiness in Malaysia. Beef and dairy cattle, and poultry are gradually fed in close proximity to maximize proficient production and reduce costs. This practice produces large amounts of animal manure that will exude odour, methane, hydrogen sulphide, nitrous oxide, carbon dioxide, and ammonia (Monteny, Bannink, & Chadwick, 2006). Manure run-off if unrestrained can also produce water pollution due to its phosphorus and nitrates content (Miner, 1999). In recent times, such unprecedented passion in the renewable energy, as a sustainable energy

Article history:
Received: 02 March 2016
Accepted: 14 December 2016

E-mail addresses:
tinia@upm.edu.my; tiniaghazi@gmail.com
(Tinia Idaty, Mohd Ghazi),
naseer.ismaeel@gmail.com (Ismail, Muhammad Nasir)
*Corresponding Author

ISSN: 0128-7680

source due to the increasing ecological concerns coupled with high energy costs have prompted an increased interest in utilizing biomass waste, such as animal manure, for biogas production (Omer, 2008; Nasir et al., 2013). This can be achieved either by burning manure specifically for fuel or to produce heat, and/or by turning it into "biogas" through anaerobic digestion (Panwar et al., 2011). Anaerobic digestion is a biological process whereby microorganism breakdown the degradable organic content in the absence of oxygen to produce biogas, and a nutrient-rich digestate used as fertiliser (Nasir, Mohd Ghazi, & Omar, 2012). Biogas is made up of basically methane and carbon dioxide, and may have traces of hydrogen sulphide and moisture. Methane has a global warming effect that is 21 times that of carbon dioxide; so using the methane for energy production considerably reduces the greenhouse gas emissions (Jorgenson, 2006). Also, since the manure used in the anaerobic methane digesters is not runoff over the surface of land, this benefit the local drainage basin as well. Manure-based digesters can improve rural economic growth and gives livestock farmers and feedlot operators with diversified revenue source, or at best lower the cost of disposal (Chen, Yang, Sweeney, & Feng, 2010).

In Europe, the number of operational anaerobic digestion plants has increased to over 14,500 as reported by the European Biogas Association's (EBA) Biogas Report 2014 (EBA, 2014), which focus mainly on the electricity generation, supplying heat and/or transport fuel. Similarly, the rapid growth for this technology is recently experienced, especially because of the essential factor of the greenhouse gases emission reduction adopted at the Kyoto Summit (Mata-Alvarez, Mace, & Llabres, 2000). On the other hand, in Asia, the most common application of biogas is for households. For example, in China, it is estimated that up to 15 million households use biogas energy in rural areas, and effective programmes have been set up in Indian and Nepal (Van Nes, 2006). Now, moves are going on to unveil biogas development programmes right across South East Asia, particularly in Malaysia and Vietnam. In Malaysia, there are huge prospect for utilizing biogas resources from palm oil mills, wastewater treatment facilities, landfills and agriculture industry (Sumathi, Chai, & Mohamed, 2008). Highlighting the significance of the country's sustainable biomass feedstock in addition to the government's attention on the green renewable technologies is the country's strategic plan for developing Malaysia as a biogas centre in Asia (Jaafar, Kheng, & Kamaruddin, 2003). In the 10th Malaysia plan, the new energy set to achieve is of 985 MW by 2015, which will contribute to 5.5% of Malaysia's total electricity generation mix. In order to achieve its target, the National Renewable Energy Policy (NREP) was launched, together with various new initiatives attached upon the Renewable Energy Policy 2010 and Action Plan will be embarked on (Hashim & Ho, 2011).

This paper describes how the biogas energy anaerobic digesters create system-wide operational efficiencies in the energy production and the operational costs. Also, the efficiency of utilizing biogas for power generation for an anaerobic digestion system-treating animal and food waste is presented. The findings presented here are based mainly on the laboratory experimental results published. It should be mentioned that this work would not address the economic assessment of the biogas production in general but focus on the possible electrical generation.

TECHNO-ECONOMIC ANALYSIS

Energy production efficiency

Many in the anaerobic digester industries use a cow per kW ratio to denote the power generation efficiency, but this measure has a number of limitations. Firstly, the methane production differs significantly depending on system downtime, so technologies that check or quickly recover from shock have an obvious advantage as per enhanced energy production (Chae, Jang, Yim, & Kim, 2008). An important factor in methane production is the time duration the substrates expend in the digesters, referred to as the hydraulic retention time. It should be noted that shorter retention time depicts an ineffective methane production, so complete revenue is not realized. Whereas, longer retention time depicts more was spent on excess capacity or not sufficient substrate is being added to maximize revenue (Ahring, 2003). Therefore, an optimized digester retention time is needed to maximize revenue with the most suitable capital costs. Secondly, co-digesting manure, by adding other substrates has been found to increase the methane production (Mata-Alvarez, Mace, & Llabres, 2000). Hence, it has been found out that different substrates have various methane productions potential (Table 1).

Table 1
Methane production potential of various organic wastes

Substrates	Methane yield (L/g VS_{added})
Chicken manure	0.295
Dairy manure	0.500
Swine manure	0.322
Corn stover	0.241
Wheat straw	0.245
Rice straw	0.281
Kitchen waste	0.541
Fruit and vegetable waste	0.342
Used animal oil	0.776
Used vegetable oil	0.811
Yard waste	0.183
Switch grass	0.246
Vinegar residue	0.253
Rice husk	0.049

Source: Li, Zhang, Liu, Chen, He, & Liu (2013)

It can be seen form Table 1 that the highest methane production for the various organic substrates studied was achieved by the vegetable oil and used animal oil with a yield of 0.811 and 0.776 L/g VS_{added}, respectively. This is followed by the kitchen waste with a methane yield of 0.541 L/g VS_{added} possibly due to its high content of lipids and protein. Dairy manure also showed a bit higher yield of 0.500 L/g VS_{added}, whereas rice husk demonstrated the lowest methane yield of 0.049 L/g VS_{added}. Similarly, lignocellulosic biomass showed low yields, which were all below 0.300 L/g VS_{added}.

Cost of systems

As it is known that the conventional cows per kW ratio is not perfect for describing the energy production in a digester, similarly, a cost per cow figure is insufficient due to the variables concerned. Hence, a more convenient way to approximate cost is to find out all potential substrates for a given operation, which will set up the capital costs as well as revenue from the energy production and tipping fees (Theodore, 1994). After that, the power generation rates can be estimated in addition to downtime estimates (Ghadfoori & Flynn, 2006). In this way, a financial model can be developed to provide a clearer picture than the conventional cost of capital in dollars per cow.

Efficiency of electrical energy production from biogas

Generally, biogas anaerobic digesters produce a system-wide operational efficiency in the renewable energy production from agricultural biomass. Energy production from biogas can be a very effective technique for electric energy generation from an alternative energy source (McKendry, 2002). Biogas is usually used as fuel in combined heat and power plants (CHPs) for generating electricity and heat, although it can also be processed and transferred to the natural gas grid (Plochl & Heiermann, 2006). However, this is suitable only if the upcoming heat from the biogas generator can be utilized in a cost-effective and environment-friendly manner (Demirbaş, 2001). The energy content of a biogas mixture is directly proportional to the methane content in it. Hence, in order to convert the biogas to pure methane, the biogas must go through scrubbing and cleaning processes to remove carbon dioxide, hydrogen sulfide, and other traces of gases (Zhao, Leonhardt, MacConnell, Frear, & Chen, 2010). The calorific energy content of each cubic meter of biogas is about 6 kWh (21-23.5 MJ/m^3), which is equal to 0.5-0.6 L of diesel fuel (Ahmad, 2010). However, due to conversion losses, approximately 1m^3 of biogas (65% methane) can be converted only to approximately 1.7 kWh of the useable electric energy (Oleszkiewicz & Barnard, 2006). Whereas, the rest is converted into heat which can further be used for heating applications. This 1.7 kWh is sufficient energy to power a 100 W light bulb for 15 hours or a 2000W hair dryer for an hour. Therefore, the electrical energy generation from biogas is a suitable technology suitable even for quite minor applications between range of 10 and 100kW (Ahmad, 2010). Similarly, however, the electrical energy generation costs of a biogas plant decrease with an increase in plant size, which automatically reduces the efficiency and the operational cost (Persson, Jönsson, & Wellinger, 2006).

The biogas electrical conversion efficiency is the amount of electrical energy from the total energy available. Conventionally, biogas has been used to fuel engine-driven generators, which attains up to 30-35% energy efficiency, while, much older systems achieve only 20-30% electrical energy efficiency (Galitsky, Worrell, & Ruth, 2003). Lately, the biogas engine generator systems have improved and will achieve 35 to 40% electrical energy efficiency (Jacobs III & Schneider, 2009). In recent years, micro turbines have been tested and used in the biogas applications. These systems are much like a jet engine and are presently being tested in a variety of biogas systems to assess use and durability (Demirbas, 2011).

Energy value of methane from the bioreactor

From the data collected on a batch and semi-continuous 10L lab scale bioreactor treating cattle manure, operating at hydraulic retention time (HRT) of 20 days with an average methane yield of 0.19 m^3/kg VS added and 55% methane content, the electrical energy production was calculated according to this formula as:

Electrical energy production
= [volume of Methane (m^3) × methane energy content (kWh) × efficiency of biogas engine (%)] ÷ 100 (Eqn. 1)

Assuming 30% electric generator efficiency, a 240 kW generator will be required to produce 0.9 kWh of electricity. In peninsular Malaysia, the average per household electricity consumption is 251 kWh per month (Alshafiqab, Nor Azizia, Shamsul & Amyra, 2015), and the average carbon emission factor of electricity use is about 0.684 kg CO_2e/kWh (Zaid, Myeda, Mahyuddin, Sulaiman, 2013). In terms of carbon dioxide emission (CO_2), this is interpreted to release: 0.684 kg CO_2e/kWh × 251 kWh per month = 171.68 kg of CO_2 emission per household per month. As already discussed, CO_2 emission contributes to the greenhouse gas effect responsible for global warming, also according to Advani, Bassi, Bowen, Fankhauser, Johnson, Leicester, Stoye et al. (2013) each kWh of electric energy consumption from the electrical grid produces 0.541 kg of CO_2 (Zaid, Myeda, Mahyuddin & Sulaiman 2013). Therefore, the 0.9 kWh produced from renewable source in this study will save about 0.541 kg of CO_2 emission, thereby reducing the output of carbon dioxide in the air, which makes it possible to minimize the trend of global warming. This will certainly pave the way to a healthy atmosphere and reduction of environmental foot print. Although the electrical energy value obtained was lower, it is expected that further optimization of the design and substrate will enhance the methane yield as such increase the energy production. Cuéllar & Webber (2008) reported that the biogas from livestock manure represents a saving of between 47.2 and 150.4 metric tons of CO_2 in the USA. They concluded that by co-digesting animal waste with other organic wastes will at least double, if not triple, the volume of biogas available, and because the biogas methane can be purified as a renewable fuel for mobile uses for cars as well as farm machinery. Hence, it can displace larger amounts of fossil fuels, thereby contributing even more to mitigating greenhouse gases and energy savings.

Similarly, from the experimental results obtained for a novel oscillatory flow bioreactor (OFBR) operated semi-continuously for cattle manure treatment. The bioreactor was initially operated at a loading rate of 2.5 g VS/L/day, after which it was increased step-wise from 2.5 to 3.5 g VS/L/day on day 33, and to 4.5 and 5.5 on day 85 and 124, respectively. The average HRT corresponding to these OLR were 20, 12, 11.3 and 9.3 days, respectively. By assuming 30% electric generator efficiency, a 240 kW generator will be required to generate 10.12 kWh of electricity from the result of the experiment. At a price of 21.8 sen/kWh (Abdullah, Abdullah, Hassan, & Hussin, 2012), the biogas generated from treating the cattle manure waste through the anaerobic digestion using the OFBR would save 220.6 sen on the electricity. In addition, CO_2 emission of 12.14 kg will be saved monthly using this novel reactor. On the other hand, the result of batch and semi-continuous anaerobic digestion of food waste at carbon: nitrogen (C/N) ratio of 17, assuming 30% electric generator efficiency, a 240 kW generator will be required to produce 10.7 kWh at lab scale. The pilot scale experiment operated under the same condition, achieved an electrical energy of 12.3 kWh. This will also save CO_2 emission of 12.8 and 14.7 kg for the lab scale and pilot scale, respectively.

Presently, there are quite a few biogas plants in Malaysia, most of which operated at palm oil mills for treating the palm oil mill effluent (POME). As of now, four oil palm biogas projects were permitted to be Feed-in Approval Holders (FIAHs) for grid connection under the feed-in-tariff (FiT) system [36]. Two of these biogas plants are Bell Eco Power Sdn. Bhd., and Achi Jaya Plantations Sdn. Bhd., both are located at Johor and were already connected to the grid with a total capacity of 3.25 MW (Chin, Poh, Tey, Chan, Chin, 2013). In addition, there are a number of firms that provide technical know-how for the application of biogas technology in palm oil mills such as International Asia Sdn. Bhd., Novaviro Technology Sdn. Bhd., Biogas Environmental Engineering Sdn. Bhd. by setting up mostly the sealed cover ponds over the present anaerobic digester systems to save cost (Chin, Poh, Tey, Chan, Chin, 2013). However, strong government commitment is significant by formulating the regulatory context of capturing methane gas from the anaerobic digestion of POME. This will generally encourage the shift from the conventional open pond system to closed digester biogas system for capturing methane gas. Also, government should provide incentives and possibly tax deduction to biogas producers particularly to the palm oil mills to support them with an increased capital expenditure of the biogas power plant. Therefore, with the improvement in the anaerobic digester system for anaerobic treatment of POME, it is certain that POME will turn into a viable biogas resource to enhance Malaysia's renewable energy sector in the nearest future.

CONCLUSION

In most developing countries power production from renewable sources has emerged to be particularly profitable in areas that are not in the national grid. There are signs of financial and legal back up for feeding in electricity from biogas power plants, and a considerable increase in feed-in tariff for biogas power in some countries such as Malaysia. As a major contributor to Malaysia's economy, palm oil industry can be expected to continue its prominence in the future making palm oil mill effluent (POME) a sustainable biogas energy source for Malaysia. Hence, anaerobic digestion is an excellent treatment practice for POME as it converts it to sustainable energy that will benefit the palm oil mill and the government with regard to the environment, climate change and profitable future. However, there are many problems that impede the advancement of renewable energy using POME as the main source. But the most important factor is the reluctance of the palm oil mills in Malaysia to undertake into advanced efficiency technologies such as closed anaerobic bioreactors due to cheaper operating costs and ease of operation. Thus, it is essential to mobilize market forces by setting up policies, regulatory framework, and appropriate incentives to encourage the palm oil industries. In addition, more financial support will be needed and researches should be performed to enhance the methane and biogas yield from the anaerobic digestion of POME. With the improvement in the anaerobic bioreactor technology of POME, like the application of the novel OFBR, it is certain that POME will act as a good biogas resource to boost Malaysia's renewable and sustainable energy sector.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the Universiti Putra Malaysia for funding under the Research University Grant scheme.

REFERENCES

Abdullah, A. S., Abdullah, M. P., Hassan, M. Y., & Hussin, F. (2012). Renewable energy cost-benefit analysis under Malaysian feed-in-tariff. In *Research and Development (SCOReD), 2012 IEEE Student Conference*, (pp. 160-165).

Ahmad, S. (2010). Energy and Bio-fertilizers for Rural Pakistan: Opportunities, Integrated Technology Applications, Vision and Future Strategy. *Managing Natural Resources for Sustaining Future Agriculture, 2*(17), 1-31.

Ahring, B. K. (2003). Perspectives for anaerobic digestion. In *Biomethanation* (pp. 1-30). Springer Berlin Heidelberg.

Alshafiqab, N. A., Shamsul, S., & Amyra, M. Y. (2015). Optimization Performance of Power Generation and Freshwater Production Utilizing Ocean Thermal Energy Application. In *3rd International OTEC Symposium,* Universiti Teknologi Malaysia Kuala Lumpur, Malaysia.

Chae, K. J., Jang, A. M., Yim, S. K., & Kim, I. S. (2008). The effects of digestion temperature and temperature shock on the biogas yields from the mesophilic anaerobic digestion of swine manure. *Bioresource Technology, 99*(1), 1-6.

Chen, Y., Yang, G., Sweeney, S., & Feng, Y. (2010). Household biogas use in rural China: a study of opportunities and constraints. *Renewable and Sustainable Energy Reviews, 14*(1), 545-549.

Cuellar, A. D., & Webber, M. E. (2008). Cow power: the energy and emissions benefits of converting manure to biogas. *Environmental Research Letters, 3*(3), 1-8.

Demirbaş, A. (2001). Biomass resource facilities and biomass conversion processing for fuels and chemicals. *Energy Conversion and Management, 42*(11), 1357-1378.

Demirbas, A. (2011). Competitive liquid biofuels from biomass. *Applied Energy, 88*(1), 17-28.

EBA. (2014). *European Biogas Association*. Retrieved November 12, 2015, from http://european-biogas.eu/

Elfeituri, F. E., & Taboun, S. M. (2002). An Evaluation of the NIOSH Lifting Equation: A Psychophysical and Biomechanical Investigation. *International Journal of Occupational Safety and Ergonomics, 8*(2), 243–258.

Galitsky, C., Worrel, E., & Ruth, M. (2003). Energy Efficiency Improvement and Cost Saving Opportunities for Corn Wet Milling Industry: An Energy star Guide for Energy and Plant Managers. *Ernest Orlando Lawrence Berkeley National Laboratory and US Environmental Protection Agency.* University of California Berkeley, CA.

Ghadfoori, E., & Flynn, P. C. (2006, March). Optimum sizing for anaerobic digestion. *BIOCAP Canada Foundation.* Department of Mechanical Engineering, University of Alberta.

Hashim, H., & Ho, W. S. (2011). Renewable energy policies and initiatives for a sustainable energy future in Malaysia. *Renewable and Sustainable Energy Reviews, 15*(9), 4780-4787.

Jaafar, M. Z., Kheng, W. H., & Kamaruddin, N. (2003). Greener energy solutions for a sustainable future: issues and challenges for Malaysia. *Energy policy*, *31*(11), 1061-1072.

Jacobs III, J. A., & Schneider, M. (2009). Cogeneration Application Considerations. *GE Energy*. Corporate Publication, Technical Evaluation and Marketing Analysis.

Jorgenson, A. K. (2006). Global warming and the neglected greenhouse gas: A cross-national study of the social causes of methane emissions intensity, 1995. *Social Forces*, *84*(3), 1779-1798.

Kuorinka, I., Lortie, M., & Gautreau, M. (1994). Manual Handling in Warehouses: The Illusion of Correct Working Posture. *Ergonomics*, *37*(4), 655 – 661.

Li, Y., Zhang, R., Liu, G., Chen, C., He, Y., & Liu, X. (2013). Comparison of methane production potential, biodegradability, and kinetics of different organic substrates. *Bioresource Technology, 149*, 565-569.

Marras, W. S., Lavender, S. A., Leurgans, S. E., Fathallah, F. A., Ferguson, S. A., Allread, W. G., & Rajulu, S. L. (1995). Biomechanical Risk Factors for Occupationally Related Low Back Disorders. *Ergonomics, 38*(2), 377-410.

Mata-Alvarez, J., Mace, S., & Llabres, P. (2000). Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives. *Bioresource Technology, 74*(1), 3-16.

McKendry, P. (2002). Energy production from biomass (part 2): conversion technologies. *Bioresource Technology, 83*(1), 47-54.

Miner, J. R. (1999). Alternatives to minimize the environmental impact of large swine production units. *Journal of Animal Science, 77*(2), 440-444.

Monteny, G. J., Bannink, A., & Chadwick, D. (2006). Greenhouse gas abatement strategies for animal husbandry. *Agriculture, Ecosystems and Environment, 112*(2), 163-170.

Nasir, I. M., Ghazi, T. I. M., & Omar, R. (2012). Production of biogas from solid organic wastes through anaerobic digestion: a review. *Applied Microbiology and Biotechnology, 95*(2), 321-329.

Nasir, I. M., Ghazi, T. I. M., Omar, R., & Idris, A. (2013). Batch and semi-continuous biogas production from cattle manure. *International Journal of Engineering and Technology*, *10*(1), 16-21.

Oleszkiewicz, J. A., & Barnard, J. L. (2006). Nutrient removal technology in North America and the European Union: a review. *Water Quality Research Journal of Canada, 41*(4), 449-462.

Omer, A. M. (2008). Green energies and the environment. *Renewable and Sustainable Energy Reviews, 12*(7), 1789-1821.

Panwar, N. L., Kaushik, S. C., & Kothari, S. (2011). Role of renewable energy sources in environmental protection: a review. *Renewable and Sustainable Energy Reviews, 15*(3), 1513-1524.

Persson, M., Jönsson, O., & Wellinger, A. (2006). Biogas upgrading to vehicle fuel standards and grid injection. *IEA Bioenergy task*, *37*, 1-34.

Plochl, M., & Heiermann, M. (2006). Biogas farming in central and northern Europe: a strategy for developing countries? Invited overview. *Agricultural Engineering International*, *8*(8), 1–15.

Sumathi, S., Chai, S. P., & Mohamed, A. R. (2008). Utilization of oil palm as a source of renewable energy in Malaysia. *Renewable and Sustainable Energy Reviews*, 12(9), 2404-2421.

Theodore, L. (1994). *Pollution prevention*. Switzerland: Gordon and Breach Science Publishers.

Van Nes, W. J. (2006). Asia hits the gas: Biogas from anaerobic digestion rolls out across Asia. *Renewable Energy World, 9*(1), 102-111.

Zaid, S. M., Myeda, N. E., Mahyuddin, N., & Sulaiman, R. (2013). The Need for Energy Efficiency Legislation in the Malaysian Building Sector. In T*he 3rd International Building Control Conference,* Kuala Lumpur, Malaysia.

Zhao, Q., Leonhardt, E., MacConnell, C., Frear, C., & Chen, S. (2010). Purification technologies for biogas generated by anaerobic digestion (pp. 1-24). *CSANR Research Report.*

Pertanika J. Sci. & Technol. 25 (1): 161 - 170 (2017)

SCIENCE & TECHNOLOGY

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

Anaerobic Batch Digestion of Cattle Manure under Various Oscillatory Flow Mixing

Ismail, Muhammad Nasir, Tinia Idaty, Mohd Ghazi*, Rozita, Omar and Wan Azlina, Wan Abd Karim Ghani

Department of Chemical and Environmental Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

The feasibility of an anaerobic digestion of cattle manure for biogas production is studied in this paper. A batch thermophilic oscillatory flow anaerobic bioreactor (OFBR) operated in thermophilic (55°C) condition was used. Within the experimental conditions set in this study, the effect of mixing intensity on volatile solids removal was found out to be significant. Results demonstrated that increasing the level of mixing decreased the digester performance. Low intensity mixing at oscillatory Reynolds number (Re_o) of 100 achieved an increase of 37% in biogas yields compared to high mixing intensity, Re_o of 500. It was observed that the mixing intensity effect interacts with the methane composition in the biogas. The benefit of decreasing mixing intensity emerges to significantly increase the methane composition in the biogas. These experiments established that high intensity mixing was not essential for good performance of oscillatory flow anaerobic bioreactor. In addition, the effect of mixing intensity might be reduced through the use of a slightly lower total solid concentration, hence, lowering the operational cost of the process. Although the study was lab scale a pilot-scale system where mixing retention times are longer would be useful.

Keywords: Anaerobic digestion, biogas, cattle manure, mixing intensity, oscillatory flow mixing

INTRODUCTION

In recent times eco-friendly reliable waste disposal methods have gained popularity in the animal husbandry industry to protect the environment. An alternative animal manure treatment approach is anaerobic biological treatment bioreactors. Anaerobic digestion is a biological operation that breakdown organic matter such as animal waste in the absence

Article history:
Received: 02 March 2016
Accepted: 14 December 2016

E-mail addresses:
naseer.ismaeel@gmail.com (Ismail, Muhammad Nasir),
tinia@upm.edu.my; tiniaghazi@gmail.com
(Tinia Idaty, Mohd Ghazi),
rozitaom@upm.edu.my (Rozita, Omar),
wanazlina@upm.edu.my (Wan Azlina, Wan Abd Karim Ghani)
*Corresponding Author

ISSN: 0128-7680

of free oxygen into biogas, and produces a digestate that can be used for soil conditioning (Nasir et al., 2012a). There are two major bioreactor designs common to farm-scale operations, which are the plug-flow, and mixed-flow bioreactors (Wu & Chen, 2008). The mixed-flow reactors are more advantageous than the plug-flow reactors, because of the homogeneous dispersion of the substrate induced by agitation. These, therefore, provides effective use of the total bioreactor volume, prevent stratification and temperature gradients, disperse metabolic end products and toxic materials in the influent, and sustain close contact between the bacteria, and their substrate (Wu & Chen, 2008). Adequate mixing provides a uniform environment for anaerobic bacteria, which is one of the major factors in achieving optimal digestion (Amani et al., 2010). Mixing is recognised as among the most common unit operations in process manufacturing industries and several categories of mixers and impellers have been formed for different operations (Nunhez et al., 2005). However, the major disadvantage of this type of bioreactors is the additional energy and mechanical complexity related with the pump or impeller.

The conventional types of mixing in digesters include slurry-recirculation, gas-recirculation, and mechanical agitation. However, mechanical agitation is leading the market for stirring fermenting substrates especially from agricultural origin (Wu & Chen, 2008). A study conducted by Karim, Hoffmann, Klasson & Al-Dahhan (2005) on the influence of mixing in the anaerobic degradation of animal waste showed that mechanical, hydraulic and pneumatic accounted for 29%, 22% and 15% higher biogas yields compared to the unmixed digesters. In addition, Stroot, McMahon, Mackie & Raskin (2001) reported that by treating animal manure, similar effects on biogas production rates and yields at steady-state conditions of four different mixing intensities (50, 350, 500 and 1500 per min) could be obtained in continuously stirred bioreactors. They observed a higher methane production by 1.3% and 12.5% with intermittent and minimal mixing strategies compared to continuous mixing of manure in the bioreactor. According to Kaparaju, Buendiaa, Ellegaardb, & Angelidaki (2007) mixing in manure fed anaerobic digesters is very important, and the mixing intensity was found to have a small effect on biogas yield. On the other hand, mixing showed to have an effect on the anaerobic digestion of manures.

Although, much research had been conducted in this field using different mixing, but it must be stated that oscillatory flow mixing (OFM) has never been reported in the anaerobic digestion of animal manure. OFM use a combination of flow oscillation and baffled tube geometry to ensure efficient mixing and effective heat transfer. Therefore, the objective of this study was to investigate the effect of different oscillatory flow mixing intensity on the anaerobic digestion of cattle manure for biogas production.

MATERIALS AND METHODS

Substrate

The cattle manure used in this study was collected from a dairy farm in Taman Pertanian Universiti (TPU) in the Universiti Putra Malaysia campus in Serdang, Selangor, Malaysia. The cattle manure was mixed (1:1) with water then filtered through a screen (0.5cm x 0.5cm) and stored at 4°C until use. The characteristics of the initial substrate are shown in Table 1.

Table 1
Chemical Composition of Cattle manure

Parameter	Unit	Cattle manure
pH	-	7.65
TS	%	11.7
VS	%TS	83.7
NH_3-N	mg/L	2733
COD	mg/L	640.3
DM	%dry base	16.7
CF	g/kg DM	34
CP	g/kg DM	11.8
NDF	g/kg DM	83.36
ADF	g/kg DM	71.77
Lignin	g/kg DM	30.61
Cellulose	g/kg DM	41.16
Hemicellulose	g/kg DM	11.59
Acetic acid	mg/L	399.4
Propionic	mg/L	277.5
Butyric	mg/L	120.4
Isovaleric	mg/L	83

Experimental Setup

Experiments were performed in a stainless steel jacketed oscillatory flow bioreactor (OFBR) mounted on an oscillator base unit, with a working volume of 4.5 litres (L). The internal diameter and length of the tube were 98 mm and 732 mm, respectively, and it is fitted with a baffle insert assembly using outside diameter of 80 mm baffles with 28 mm diameter orifices spaced at 1.5 tube diameter. A variable-frequency, variable-amplitude oscillator was used, in which a stainless steel pistons provided fluid oscillation at the base of the reactor tube (Figure 1). Operating temperature was maintained at 55±1 °C, controlled and monitored circulating heated water through the OFBR jacket. While, pH was set in the range between 6.8 and 7.1, by a pH controller that continuously monitors the pH and adds acid or base, as required.

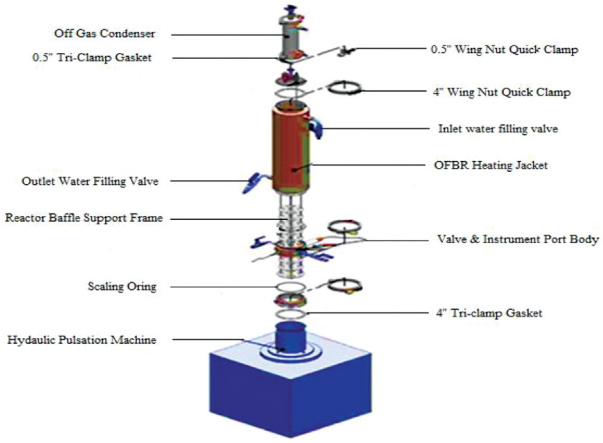


Figure 1. Schematic of the OFBR system

In the experimental period; three different mixing intensities were tested (Harvey, Mackley, & Stonestreet, 2001) as batch at different times. Three batches were processed at different oscillatory Reynolds number (Re_o) of 100, 300 and 500, which is equivalent to frequencies of 1Hz, 2Hz and 5Hz, respectively. Further, a control batch experiment without mixing was performed to provide a baseline for comparison. Within each batch, the effect of mixing on the bioreactor performance was evaluated. Composition and volume of the biogas produced from the bioreactor was measured daily until the biogas production seizes. Biogas was collected in tedlar bags (Supelco, Bellefonte, PA) and the volume produced was measured by water displacement.

Analytical Methods

Total solids (TS), volatile solids (VS), chemical oxygen demand (COD), and ammonia nitrogen (NH_3-N) were analysed according to the standard method (APHA, 1998). The fibre compositions such as neutral detergent fibre (NDF), acid detergent fibre (ADF), crude fibre (CF), crude protein (CP) and lignin were analysed using the reflux apparatus. Methane content was analysed using a gas chromatography (HP 6890N) equipped with a thermal conductivity detector (TCD) and a HP Molesieve column 30 m x 0.5 mm ID. The injector, detector, and oven temperatures were 60°C, 200°C and 70°C, respectively. Argon served as the carrier gas. The biogas yield was expressed as the volume of biogas produced based on the initial total VS in the feedstock. Samples for volatile fatty acid (VFA) determination was collected every other day and centrifuged at 11,000 rpm for 20 min. The supernatant was used in the analysis. A gas chromatography (Agilent 7890N), using a Supelco SP 2560 capillary column of 100m × 0.25 mm ID × 0.2-µm film thickness (Supelco, Bellefonte, PA, USA) was used for the analyses. It was equipped with a split/splitless injector, flame ionization detector (FID) and an auto sampler (Agilent Auto Analyzer 7683 B series, Agilent Technologies, Santa Clara, CA, USA). The temperature of the injector was 250 °C, and the detector temperature was 270 °C. The carrier gas was nitrogen at a flow rate of 1.2 mL/min. The peaks of samples were identified, and concentrations calculated based on the retention time and peak area of known standards (Sigma Chemical). The fatty acid concentrations are expressed as gram per 100 gram of the sum of identified peaks measured in each sample.

RESULTS AND DISCUSSION

Biogas and methane production

Biogas was generated from day 1 and varied significantly during the different batch studies. The daily biogas production increased steadily during the oscillatory Reynolds number (Re_o) of 100 experiment until day 10, after which the production dropped gradually until the digestion process was stopped at day 21 (Figure 2). Similarly, the biogas production during the Re_o of 300 and 500 experiments appeared to increase until day 6, and then decreased continuously until day 20 when the biogas production was observed to seize. However, the highest amount of average daily biogas was observed during the experiment using Re_o of 100, as compared

to Re_o of 300 (1.46 L/L/day) and Re_o of 500 (1.28 L/L/day). The methane content of biogas exceeded 50 % in all the experiments except for the control. The average biogas production of Re_o of 100 and Re_o of 300 operations amounted to 0.28 and 0.17 L/g VS added with an average methane content of 52 % and 44 %, respectively, which were higher than the values of Re_o of 500 and control operations (Figure 3).

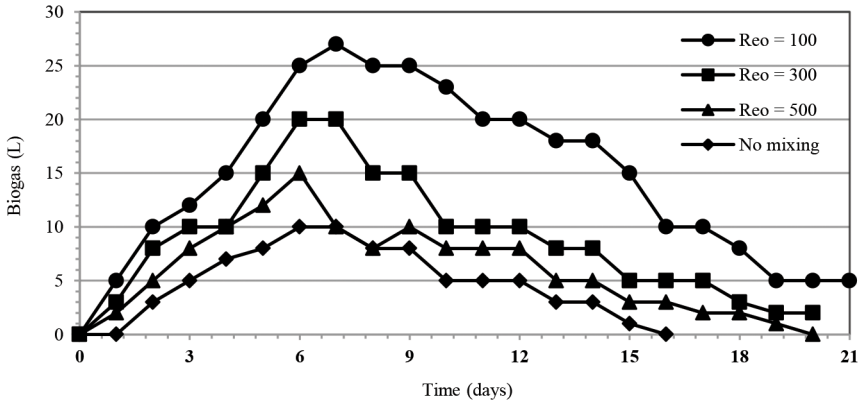


Figure 2. Biogas production in different mixing conditions

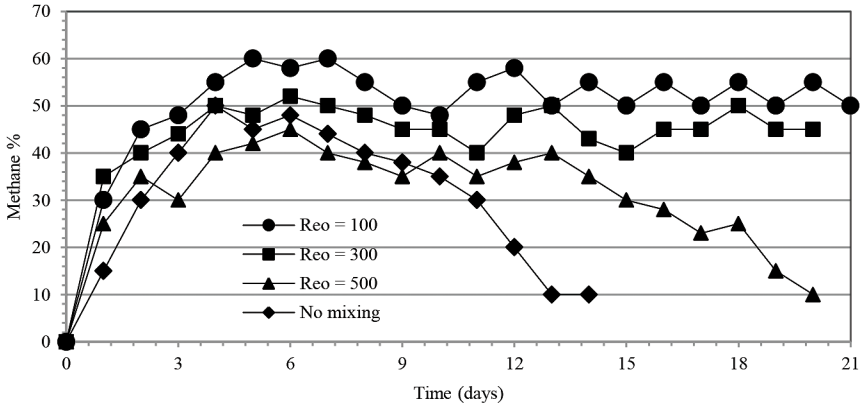


Figure 3. Methane composition in different mixing conditions

Additionally, over the 20 days of operation, the highest methane yield was obtained for Re_o of 100, followed by Re_o of 300 and 500 experiments with values of 0.144, 0.074 and 0.038 L CH_4/g VS added respectively, which was higher than the control of 0.034L CH_4/g VS added.

The significance of mixing in accomplishing effective substrate conversion has been reported in the literatures (Karim, Hoffmann, Klasson & Al-Dahhan, 2005; Kaparaju, Buendiaa, Ellegaardb & Angelidaki, 2007). Though, the readily obtainable facts available in the literature on the influence of mixing intensity on the performance of anaerobic bioreactors is contradictory. Adequate mixing was shown to enhance the uniform distribution of substrates, enzymes and microorganism all through the digester; on the other hand, inadequate mixing

leads to the non-uniform distribution of substrates, enzymes and microorganisms which was to result in stratification and formation of floating layer of solids (Amani, Nosrati & Sreekrishnan, 2010; Jha, Li, Nies & Zhang, 2013). In this study, the results obtained suggest that high intensity mixing may prevent excellent performance of the OFBR. Whereas, low intensity mixing exhibited higher methane yield and better performance, which is attributed to sufficient substrate distribution and producing direct contact with the microbes in the bioreactor. Results show that the biogas and methane yield decreased with increasing mixing intensity. This coincides with the results achieved by Kaparaju, Buendiaa, Ellegaardb & Angelidaki (2007) who found that the vigorous mixing would result in delaying and lowering the methane production.

Effects of performance parameters on OFBR stability

pH. As can be seen from figure 4, the pH was stable and remained in the neutral range (between 6.8 and 7.5) until the end of the experiment for Re_o of 100 and 300, but the pH for Re_o of 500 and the control decreased continuously at the beginning of the experiment from 6.8 to the range of 4.7–5.3 after 4 days (Figure 4). The observed pH was reported as unsuitable for anaerobic digestion as the suitable pH for methanogenic bacteria ranges from 6.5 to 7.5 (Nasir et al., 2012b). In addition, the ammonia nitrogen values in different experiments were all below 1000 mg/L indicating that it was in the safe range for anaerobic digestion as reported by Calli, Mertoglu, Inanc & Yenigun (2005).

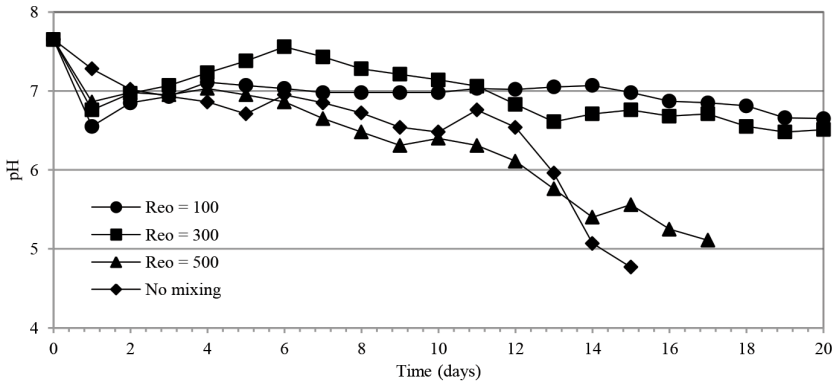


Figure 4. pH profile during different mixing conditions

Volatile solids (VS). Figure 5 shows the VS concentration profile with different Re_o and no mixing. The VS concentration reached stability after day 18 during the experiments for Re_o of 300 and 500, which indicates the microbial transformational activity. However, for Re_o of 100 there is still tendency for further degradation of VS. Further, it was observed that the VS degradation during the experiments except the control was not over after 20 days of operation. As evidence in Figure 2, the best values of biogas production were for all experiments except the control. This is because the VS degradation took place during days 0 to 5, after which the

degradation dropped drastically. Finally, the VS removal was the highest for Re_o of 100, with the value of 57 % after 20 days of operation, which was almost 3 times higher than that of the control, followed by Re_o of 300 and 500 with 42 % and 30 %, respectively.

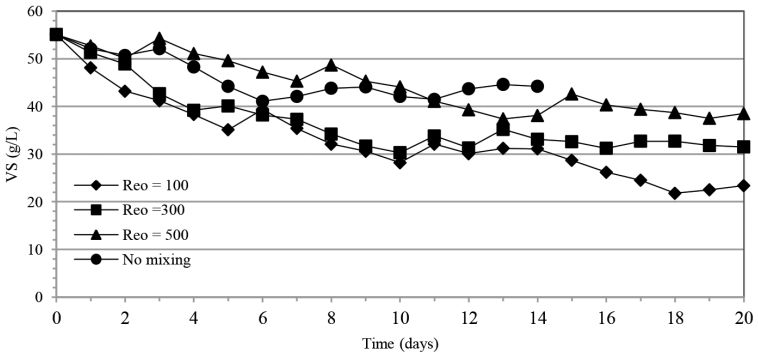


Figure 5. VS concentration during different mixing condition

Volatile fatty acid (VFA). The concentration of the total VFA (TVFAs) in a bioreactor is usually obtained by their rate of production and removal. During this experiment, the rate of acetic acid removal was higher to its production rate because almost 70 % of the methane generated is originated from acetate, while the remainder of the methane is originated essentially from the reduction of carbon dioxide with hydrogen (Chynoweth et al., 2001). During the experiments, it was observed that the concentration of propionic acid increased slowly for Re_o of 100, 300 and 500 with almost no accumulation of butyric and valeric acids. However, the butyric and valeric acids were observed to accumulate in the control experiment, this is evident as pH was observed to drop fast as such stop the digestion after day 14. Table 2 summarises the average concentration of VFAs data during the study.

Table 2
Reactor Performance Data on the volatile fatty acids concentration

	Concentration (mg/L)				
	Acetic acid	**Propionic acid**	**Butyric acid**	**Valeric acid**	**TVFA**
Re_o = 100	620.25	440.87	181	64.62	1306.7
Re_o = 300	645.2	467.5	194.1	73.1	1380
Re_o = 500	682.9	548.5	213.5	84.3	1529.2
No mixing	673.1	586.4	319	116.6	1695.1

CONCLUSION

The experiments with different mixing strategy at different oscillatory Reynolds number (Re_o) of 100 and 300 during the anaerobic digestion of cattle manure for biogas production had been performed successfully, except for Re_o of 500 and no mixing. Results demonstrated that Re_o

of 100 and 300 resulted in excellent performance after 20 days of the experiment. High biogas production rates of 1.46 L/L/day and 1.28 L/L/day, with methane yield of 0.144 L /g VS added and 0.074 L /g VS added were observed for these operational conditions, respectively. This shows that minimal mixing could enhance the digestion process by improving the concentration of VFAs in a safe range. On the other hand, Re_o of 100 and control experiments were found to be inhibitory during the experiment, possibly due to the disruption of syntrophic relationships between the microorganisms. This is evident as VFAs mainly of valeric and butyric acids were observed to accumulate, which degraded slowly and finally seized the digestion. In all the experiments, an optimum performance for the methane yield was achieved by adapting the mixing condition at Re_o of 100.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Taman Pertanian Universiti, UPM, for feedstock supply, and Universiti Putra Malaysia for financial support.

REFERENCES

Amani, T., Nosrati, M., & Sreekrishnan, T. R. (2010). Anaerobic digestion from the viewpoint of microbiological, chemical, and operational aspects-a review. *Environmental Reviews*, *18*(NA), 255-278.

APHA. (1998). *Standard methods for the examination of water and wastewater* (19th ed). New York, NY: American Public Health Association.

Calli, B., Mertoglu, B., Inanc, B., & Yenigun, O. (2005). Effects of high free ammonia concentrations on the performances of anaerobic bioreactors. *Process Biochemistry, 40*(3), 1285-1292.

Chynoweth, D. P., John M. O., & Robert, L. (2001). Renewable methane from anaerobic digestion of biomass. *Renewable Energy, 22*(1), 1-8.

Harvey, A. P., Mackley, M. R., & Stonestreet, P. (2001). Operation and optimization of an oscillatory flow continuous reactor. *Industrial and Engineering Chemistry Research, 40*(23), 5371-5377.

Jha, A. K., Li, J., Nies, L., & Zhang, L. (2013). Research advances in dry anaerobic digestion process of solid organic wastes. *African Journal of Biotechnology, 10*(64), 14242-14253.

Karim, K., Hoffmann, R., Klasson, T., & Al-Dahhan, M. H. (2005). Anaerobic digestion of animal waste: Waste strength versus impact of mixing. *Bioresource Technology, 96*(16), 1771–1781.

Kaparaju, P., Buendiaa, I., Ellegaardb, L., & Angelidaki, I. (2007). Effect of mixing on methane production during thermophilic anaerobic digestion of manure: Lab-scale and pilot-scale studies. *Bioresource Technology, 99*(11), 4919–4928.

Nasir, I. M., Ghazi, T. I. M., & Omar, R. (2012a) Production of biogas from solid organic wastes through anaerobic digestion: a review. *Applied Microbiology and Biotechnology*, *95*(2), 321-329.

Nasir, I. M., Ghazi, T. I. M., & Omar, R. (2012b). Anaerobic digestion technology in livestock manure treatment for biogas production: a review. *Engineering in Life Sciences, 12*(3), 258-269.

Nunhez, J. R., Cekinski, E., Joaquim, C. F., Jr., Fernandes, L. A. G., & Seckler, M. M. (2005). Design of a Static Mixer using CFD and Experiments (Vol. 1; pp. 1). In *Proceedings of Annual AIChE Meeting,* Cincinnati, OH, USA.

Stroot, P. G., McMahon, K. D., Mackie, R. I., & Raskin, L. (2001). Anaerobic codigestion of municipal solid waste and biosolids under various mixing conditions - I. Digester performance. *Water Research, 35*(7), 1804–1816.

Wu, B., & Chen, S. (2008). CFD simulation of non-Newtonian fluid flow in anaerobic digesters. *Biotechnology and Bioengineering*, *99*(3), 700-711.

Pertanika J. Sci. & Technol. 25 (1): 171 - 180 (2017)

SCIENCE & TECHNOLOGY

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

Car Annual Vehicle Kilometer Travelled Estimated from Car Manufacturer Data – An Improved Method

A, Shabadin*, N, Megat Johari and H, Mohamed Jamil

Road Safety Engineering and Environment Research Centre, Malaysian Institute of Road Safety Research, 43000 Kajang, Selangor, Malaysia

ABSTRACT

In 2013, 632 602 private cars were involved in road crashes, this amounting to 64% of all road crashes on Malaysian roads. Risk is often used to quantifying the level of road safety whilst exposure is an essential component of risk measurement. The calculation of VKT (vehicle kilometre travel) in this study using odometer reading data obtained from automobile manufacturer is used for cars. This average in 2013 was 24,129 kilometres, the highest recorded in Selangor with 28,575 kilometres and the lowest at 16,342 kilometres a year, in Johor. This method is believed to be reliable with a high yielding number of samples as well as a good representative set of samples for Malaysia.

Keywords: Average annual kilometre travelled, car, fatality index, motorcar, vehicle kilometre travelled

INTRODUCTION

Cars as the highest registered number of vehicles are undeniably the most preferred mode of transport for Malaysians. The number of registered vehicle increased by 100% within the past 10 years, where figures now reach more than 10 million in 2013. Aligned with the increasing number of cars on the road, the number of cars involved in crashes also increased. In 2013, 632 602 private cars were involved in road crashes. Figure 1 shows the number of registered private vehicle and number of private vehicle involved in crashes for the past ten years.

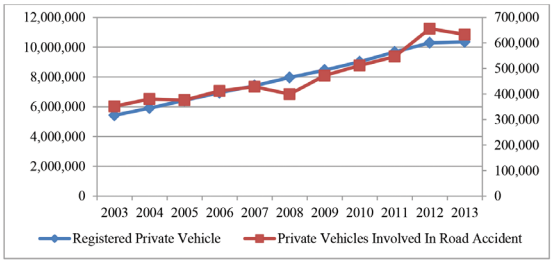


Figure 1. Profile of yearly registered private vehicle

Article history:
Received: 02 March 2016
Accepted: 14 December 2016

E-mail addresses:
akmalia@miros.gov.my (A, Shabadin),
nusayba@miros.gov.my (N, Megat Johari),
hawajamil@miros.gov.my (H, Mohamed Jamil)
*Corresponding Author

ISSN: 0128-7680

In applying or adopting necessary countermeasures, it is important to understand their exposure on the road. Road safety performance indicators play an important role in explaining the situation. Road safety performance indicators are defined as any measurements that are related to crashes or injury. These indicators are used to indicate road safety level and provide a clearer picture for road safety performance. Risk is often used as a way of quantifying the level of road safety (Hakkert & Braimaister, 2002) as well as to improve transport safety (Hakkert & Braimaister, 2002). Hauer defined risk as the probability of a crash occurring (Hauer, 1982), where the risk of a crash occurring can be estimated by dividing the number of crashes by the road user exposure to the opportunity for a crash to occur (Wundersitz & Hutchinson, 2008).

Exposure as a measure of the number of opportunities for crashes or injuries to occur is an essential component of risk measurement. Exposure is often used as the denominator when calculating crash rates to estimate crash or injury risk. The common measure of exposure is distance travel that enables disaggregation by travel and demographic group for significant comparisons. It is crucial to look into exposure measures to better identify countermeasures for improvement of transport safety. These countermeasures may be designed to reduce the risk of exposure, the risk of a crash, or the risk of an injury or death once a crash has occurred (Hakkert & Braimaister, 2002). An acceptable and useful distance travelled measurement is vehicle kilometre travel also known as VKT. It is the measurement of how far a vehicle normally travels in a year term. It is internationally a well-established indicator and is accepted by most developed countries. Total VKT provides a proxy measure of the overall pressure on the environment from all forms of road transport (NZ Ministry, 2009). However, the estimation of VKT is not as straightforward as the traffic flow. VKT has always been a difficult indicator, because it is not measured directly, rather it is always estimated (Collins, 2001).

The main VKT estimation methods can be classified into two categories namely traffic measurement methods and non–traffic measurement methods (Kumapley & Fricker, 1996). The traffic measurement VKT estimation methods are more preferable than the non-traffic measurement methods, because the former methods are based on actual data for vehicle movement (Environmental Protection Agency, 1992). Under these two categories, there are four basic methods. The two types of traffic measurement methods are odometer readings (vehicle-based method) and traffic counts (road-based method), while non–traffic measurement methods consist of household or driver survey method and fuel sales method (Azevedo & Cardoso, 2009).

White in 1976 estimated Vehicle Kilometer Travel by inspection receipt which includes previous and current odometer reading. Then inspection receipt was selected and the vehicle owners were surveyed via mail questionnaire for driving exposure information (White, 1976). Pekka on the other hand, used traffic count to estimate Vehicle Kilometer Travel and a study was done to introduce the method used to execute data collection at various level of a road network (Pekka & Pekka, 1999). The National Transport Commission Australia estimated Vehicle Kilometer Travel using volume count on arterial and municipal roads (ARRB, 2005). A study done in Australia estimated quarterly VKT by vehicle type by fuel type from the state fuel sales data for all eight states in Australia (Hossain & Gargett, 2011).

In Malaysia, the Vehicle Kilometer Travel indicator development started in 2004 and since then has improved on method of data collection in ensuring the reliability of the data. In 2004, the VKT data were collected using household survey in the state of Selangor (Mohd Fauzi, 2004). The survey was divided into two stages; the first stage involved a face-to-face interview with the respondents, and in the second stage respondents were followed up with a telephone call after a period of 3 months. Nurshaeza also adopted the same method carried out in the year 2006 (Nur Shaeza, Radin Umar, Kulanthayan & Dadang, 2006). Although the respondent rates were high (71%), the studies revealed several shortcomings such as high operating cost, time consuming and requires a significant number of manpower.

In 2007, Nurulhuda adopted the use of postcard survey to obtain the odometer readings (Nurulhuda, Akmalia, Ho & Jamilah, 2013). This method is recognized to be cost effective with the ability to reach a wider area of coverage, thus reaching out to all states in Malaysia. The same postcard method was again adopted for the calculation of the 2010 VKT value. However, the slow response rate (re-mailing of the postcards by respondents) is a concern, and is recorded as a weakness in the postcard survey method for VKT.

This paper discusses the use of a different method in the calculation of VKT where secondary data is acquired from car manufacturer in estimating the VKT value for motorcar. This study also highlights the fatality index values.

METHODOLOGY

The calculation of VKT for car is by using odometer reading data from car manufacturers. Based on new motor vehicle sales reported by Malaysian Automotive Association for 2009 until 2013, Perodua surpass with 30.6% car selling in Malaysia while Proton able to have 24.7% car sales for the past five years (Malaysian Automotive Association, 2013). The press release statement also stated that the imported Toyota car recorded up to 15.2% of car sales for the five years (Malaysian Automotive Association, 2013). Collectively, these three car brands recorded a total of more than 70% of all new car sales in Malaysia. 44 other car brands shared the remaining 30%.

The odometer reading of a car were recorded when the car is serviced at their respective service centre. The data were then sent to headquarters for storage. The odometer readings were collected from headquarters of Perodua, Proton and Toyota. These data were recorded by the respective car manufacturers from all their service centres. Throughout Malaysia, there are about 176 service centres for Perodua, 280 service centres for Proton and Toyota has 76 service centres. The variables requested from the service centres through respective car manufacturers were current odometer reading, current date, registration date, car model, and service centre branch. The data were requested for those who came for service from January to March 2013. This includes both cars that were sent for first time service, as well as older cars sent for service. Based on the data provided, on average, the age of cars that were sent to service centres were 2 years old. A total of 239,916 data was been provided by Proton while Perodua and Toyota provided 91,596 and 189,622 data respectively. The total number of sample collected is 521,134 data.

The Average Annual Kilometer Travelled (AAKT) value is calculated based on formula below:

$$\text{AAKT} = \frac{\text{Current Odometer Reading}}{\text{Current Service Data - Registration Date}} \times 365 days \quad (1)$$

The formula will give the average annual kilometre travelled based on their brands. The vehicle kilometre travelled (VKT) for a car will be calculated using expression below:

$$\text{VKT} = \frac{\text{Total Kilometer Travelled}}{\text{Total Samples}} \times \text{Registered Car Number} \quad (2)$$

The fatality indices were also calculated based on formula below:

$$\text{Death per 100,000 population} = \frac{\text{Fatality}}{\text{Population}} \times 100000 \quad (3)$$

$$\text{Death per 10,000 vehicles registered} = \frac{\text{Fatality}}{\text{Registered Vehicle}} \times 10000 \quad (4)$$

$$\text{Death per billion VKT} = \frac{\text{Fatality}}{\text{VKT}} \times 1000000000 \quad (5)$$

RESULTS AND DISCUSSION

Vehicle Kilometer Travelled

A total of 521,134 vehicles were involved in the calculation of the VKT index. Table 1 shows the average annual kilometre travelled (AAKT) by brands. On average, a Proton user drove 22,048 kilometres a year whilst a Perodua user drove a bit further with 27,994 kilometres a year. A Toyota user drove approximately 24,895 kilometres a year.

Table 1
Relationship between EMG Signal and Weight of Subject

Car Brands	Number of Sample	AAKT (km)
Proton	239,916	22,048
Perodua	91,596	27,994
Toyota	189,622	24,895
Total	521,134	24,129

The average annual kilometre car travelled for year 2013 is 24,129 kilometres. Average annual kilometre car travelled for the year 2007 is 19,135 (Nurulhuda, Akmalia, Sharifah Allyana & Hussain, 2014). The AAKT value for the year 2008 was also calculated based on odometer readings. In comparison, a difference of more than 4500 kilometres can be

observed. This increase is fairly high compared to National Travel Survey done by Department of Transport United Kingdom reported that their average distance travelled was reduced by 4% on 2012 (Lyndsey, 2013). An annual estimate of total VKT in Australia shows the total growth of 14.1 per cent during the last decade; 2001–2010 (Hossain & Gargett, 2011). In the 2008 postcard method, a response rate of 28% was obtained, giving a total of 1052 odometer readings. The difference on the method used to collect the odometer readings may considerably affect the ability to justly compare VKT readings between 2008 and 2013 without further data verification.

Table 2 shows the average annual kilometre travelled by engine size. The result shows that cars with bigger engine size travel more kilometres compared to smaller engine size. However, the average annual kilometre travelled for cars with engine sizes smaller or equal to 1.3 cubic centimetres (cc) is higher (25,133 km) than car that has engine size between 1.4 to 1.8 cc, where the AAKT is only 22,868. Engine size has always been associated with fuel consumptions where analysis of EPA data on 1977 automobiles shows that rate of fuel consumption increases with the increase of engine size, a consequences of the increase of automobile weight (Essenhigh et al., 1979). Based on the results, there is the possibility that cars with smaller engine size consume less fuel whilst travelling the same distance as cars that consumes more fuel. This translates into; for the same amount of fuel used, smaller engine cars travel more kilometres than bigger engine cars. Users with car engine size more than 1.8 cc, usually has higher income, which may give insight the higher kilometre travelled compared to those with less than 1.3 cc.

Table 2
Average Annual Kilometer Travelled by Engine Size

Engine Size	AAKT
≤ 1.3 cc	25,133
1.4-1.8 cc	22,868
>1.8 cc	31,605

Table 3 compiles the average annual kilometre travelled and the vehicle kilometre travelled by states. The highest average kilometre travelled is in Selangor with 28,575 kilometres while the lowest kilometre a car travelled is in Johor with only 16,342 kilometres a year. The number of registered cars in Selangor is higher, ranking at number three amongst all states in Malaysia. Considering that Selangor has the largest economic activity in Malaysia with good infrastructure development such as newer highways built, it is predictable that Selangor recorded the highest VKT reading. In addition, number of registered car in Selangor is also in the higher range, i.e. more than 1 million vehicle kilometre travelled, where Selangor ranked number three. On the other hand, Johor recorded the lowest number of kilometre travelled, which did not match the indicators Johor recorded, such as the high number of registered vehicles. The best assumption one can make is that people of Johor do not extensively travel out of state, which differs from the travel pattern of people in Selangor and Kuala Lumpur.

Table 3
Vehicle kilometre travelled by states

State	Kilometre Travelled (km)	AAKT (km)	Registered Car	VKT
Perlis	26,472,108.30	25,953.00	21,229	550,956,237
Kedah	621,423,233.30	22,692.90	300,868	6,827,567,437
Penang	620,520,721.40	20,979.80	1,024,197	21,487,448,221
Perak	970,021,871.30	24,933.10	699,651	17,444,468,348
Selangor	4,135,331,288.50	28,575.90	1,037,243	29,640,152,244
Wilayah Persekutuan	1,119,215,232.70	25,569.80	3,442,319	88,019,408,366
Negeri Sembilan	479,982,755.60	24,619.60	312,156	7,685,155,858
Melaka	473,455,909.80	23,551.50	310,169	7,304,945,204
Johor	1,046,923,415.90	16,342.30	1,339,446	21,889,628,366
Pahang	747,786,379.50	27,919.10	346,939	9,686,224,635
Kelantan	352,266,433.60	22,601.50	273,140	6,173,373,710
Terengganu	300,537,789.80	23,461.20	188,275	4,417,157,430
Sabah	935,051,240.30	22,821.20	556,699	12,704,539,219
Sarawak	745,488,271.20	20,737.40	683,244	14,168,704,126
Total	12,574,476,651.10	24,129.10	10,535,575	254,213,942,733

Vehicle kilometre travelled by states

Road Fatality Index

Fatality index shows trends in death rates and is used to measure road safety performance. The three types of fatality index indicators are extensively used in transport safety are fatalities per billion VKT, fatalities per 100,000 population and fatalities per 10,000 vehicles. Table 4 show the number of fatalities for the year 2013, the death per 100,000 populations and per 10,000 vehicles registered. When viewing fatality numbers alone, Johor recorded the highest number of road fatality while the least is Perlis. Nevertheless, death per 100,000 populations and 10,000 registered vehicles indicates different results. The result shows that Pahang had the highest index while Johor is placed in the bottom three. Index per 10,000 vehicles show the same results with Pahang ranked at first while Johor ranked at eighth. It is interesting to highlight that the use of a different denominator yields a different standpoint of road safety.

Table 4
Fatalities Index by State

State	Fatalities	Death per 100,000 population	Death per 10,000 vehicles registered	Death per billion VKT
Perlis	72	30.0	1.1	
Kedah	517	25.7	4.4	
Penang	381	23.1	1.8	
Perak	770	31.5	4.3	
Selangor	1019	17.6	4.8	
Wilayah Persekutuan	243	4.2	0.5	
Negeri Sembilan	396	36.7	4.5	
Melaka	258	30.4	2.5	12.2
Johor	1128	32.3	4.2	
Pahang	592	37.7	6.3	
Kelantan	378	22.6	5.0	
Terengganu	320	28.4	2.6	
Sabah	420	12.0	4.1	
Sarawak	421	16.1	2.0	
Total	6915	23.1	2.9	

Note: The death per billion VKT by state cannot be calculated due to unobtainability of VKT data by state for all vehicle types.

The fatality index per billion VKT for car was also calculated. This indicator is to describe the safety quality of road traffic (IRTAD, 2014). In general, for the year 2012, Malaysia ranked at number 21 out of 23 countries based on data recorded by IRTAD. Fatality index per billion VKT for all vehicles is 12.2 for the year 2013 whilst the government's target is to reduce the fatality index to 10 fatalities per billion VKT by year 2010 (Ministry of Transport, 2006). In the event that only cars are considered, the fatality index per billion VKT for car is 5.5. This was calculated to be that there are 6 deaths per every billion kilometre car travelled on Malaysian roads for the year 2013.

CONCLUSION

The average annual kilometre a car travelled in 2013 increased around 4994 kilometres from the average in 2007. The is due to ease of access and increased of distance between home and work place, encouraging people to travel more. An in-depth study on these factors needs to be done to better understand this trend.

ACKNOWLEDGEMENTS

This research was funded by a research grant from Malaysian Institute Road Safety Research (MIROS). We record our appreciation to Proton Holdings Berhad, the Perusahaan Otomobil Kedua Sendirian Berhad (PERODUA) and Toyota Malaysia Sdn. Bhd for providing all the odometer readings. A very special appreciation goes out to all drivers who contributed indirectly towards the completion of this study and to the staff of MIROS who helped with the preparation of facilities required for the study.

REFERENCES

ARRB Transport Research Ltd. (2005). *Estimation of Vehicle Kilometres Travelled On Arterial and Local Roads Information Paper.* ISBN: 1 877093 64 5.

Azevedo, C. L., & Cardoso, J. (2009). *Estimation of annual traffic volumes: A model for Portugal Papers of the Young Researchers Seminar 2009.* Torino Italy, June 3-5.

Collins, F., & LUTRAQ Team. (2001). *Estimation of VMT and VMT growth rate VMT reduction project*. Retrieved October 24, 2014, from www.fcgov.com/climateprotection/pdf/ctf-res-lutraq.pdf

Environmental Protection Agency. (1992, January). *VMT forecasting and tracking guidance*. EPA Section 187.

Essenhigh, R. H., Shull, H. E., Blackadar, T., & McKinstry, H. (1979). Effect of vehicle size and engine displacement on automobile fuel consumption. *Transportation Research Part A: General, 13*(3), 175-177.

Fauzi, M. W. M. (2004). *Developing New Road Safety Index for Malaysia*. (Master dissertation). Universiti Putra Malaysia, Malaysia.

Fukuda A., Satiennam T., Ito H., Imura, D., & Kedsadayurat, S. (2013). Study on Estimation of VKT and Fuel Consumption in Khon Kaen City, Thailand. In *Proceedings of the Eastern Asia Society for Transportation Studies*, *10*(0), 113-130.

Hakkert, A. S., & Braimaister, L. (2002). *The uses of exposure and risk in road safety studies, Leidschendam*. The Netherlands: SWOV Institute for Road Safety Research.

Hauer, E. (1982). Traffic conflicts and exposure. *Accident Analysis and Prevention*, *14*(5), 359-364.

Hossain, A., & Gargett, D., (2011). Road vehicle-kilometres travelled estimated from state/territory fuel sales. In *Australasian Transport Research Forum 2011 Proceedings,* Adelaide, Australia, September 28-30.

IRTAD. (2014). *Road Safety Annual Report 2014*. International Traffic Safety Data and Analysis Group.

Kumapley, R. K., & Fricker, J. D. (1996). Review of Methods for Estimating Vehicle Miles Traveled. *Transportation Research Record*, *1551*, 59-66.

Lyndsey, M. (2013). *National Travel Survey 2012-Statistical Release*. Department for Transport, Great Britain.

MAA. (2013). *Market Review 2009-2013*. Malaysian Automotive Association, Malaysia.

Ministry of Transport. (2006). *Road Safety Plan of Malaysia 2006 - 2010.* Kuala Lumpur: Road Safety Department.

Nurulhuda, J., Akmalia, S., Ho, J. S., & Jamilah, M. M. (2013). Exposure and Risk of Motorcyclist in Malaysia. *Journal of Society for Transportation and Traffic Studies, 4*(2), 35-40. ISSN: 1906-8360.

Nurulhuda, J., Akmalia, S., Allyana, S. M. R., & Hussain, H. (2014). Significance of Vehicle Kilometer Travelled (VKT) Data and Postcard Method for VKT Data Collection. In *Malaysian Road Conference,* Kuala Lumpur, Malaysia, November 10-12.

NZ Ministry for the Environment. (2009, March). *Vehicle kilometres travelled by road Environmental Report Card* (pp. 1-23). NZ Ministry for the Environment.

Pekka R., & Pekka, L. (1999). Estimating Vehicle Kilometers of Travel Using PPS Sampling Method. *Journal Transport Engineering, 125*(1), 8-14.

Umar, R.S. N. S. R, Kulanthayan, S., & Ma'soem, D. M. (2006). Fatality Indicator Based On Vehicle Kilometer Travelled (Preliminary Result). *Report to Ministry of Transport Malaysia.*

White, S. B. (1976). On The Use of Annual Vehicle Miles of Travel Estimates from Vehicle Owners? *Accident Analysis & Prevention*, *8*(4), 257-261.

Wundersitz, L. N., & Hutchinson, T. P. (2008). Identifying and improving exposure measure. *Case Report Series,* CASR053.

Pertanika J. Sci. & Technol. 25 (1): 181 - 190 (2017)

SCIENCE & TECHNOLOGY

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

Intelligent Surveillance System for Street Surveillance

Y. M. Mustafah[1]*, N. A. Zainuddin[1], M. A. Rashidan[1], N. N. A. Aziz[1] and M. I. Saripan[2]

[1]Mechatronics Engineering Department, Kulliyyah of Engineering, International Islamic University Malaysia, 50728 IIUM, Kuala Lumpur, Malaysia

[2]Department of Computer & Communication Systems, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

ABSTRACT

CCTV surveillance systems are widely used as a street monitoring tool in public and private areas. This paper presents a novel approach of an intelligent surveillance system that consists of adaptive background modelling, optimal trade-off features tracking and detected moving objects classification. The proposed system is designed to work in real-time. Experimental results show that the proposed background modelling algorithms are able to reconstruct the background correctly and handle illumination and adverse weather that modifies the background. For the tracking algorithm, the effectiveness between colour, edge and texture features for target and candidate blobs were analysed. Finally, it is also demonstrated that the proposed object classification algorithm performs well with different classes of moving objects such as, cars, motorcycles and pedestrians.

Keywords: Adaptive background modelling, distributed cameras tracking, object detection, object classification, intelligent surveillance

INTRODUCTION

Surveillance system is becoming more common in Malaysia. Many areas such as banks, shop areas, pedestrian streets and many more areas are monitored by CCTVs. Moreover, the surveillance systems are getting cheaper nowadays, making its easily deployable (Shearing & Johnston, 2013). What is lacking is, an active monitoring of the surveillance video footage to detect, track and classify the moving object accurately. Having human operator to monitor the video feed is very costly and inefficient. Human tend to become bored due to the dull nature of the monitoring

Article history:
Received: 02 March 2016
Accepted: 14 December 2016

E-mail addresses:
yasir@iium.edu.my (Y. M. Mustafah),
fiqahzainuddin@gmail.com (N. A. Zainuddin),
ariffrashidan@gmail.com (M. A. Rashidan),
nornadirahaziz89@gmail.com (N. N. A. Aziz),
iqbal.saripan@gmail.com (M. I. Saripan),
*Corresponding Author

ISSN: 0128-7680

activity. Thus, the development of intelligent surveillance technology becoming more critical as the number of CCTV is rising. Video processing of surveillance camera has many challenges such as variable illumination condition, inconsistent complex background, and differences of object appearances across camera views (Chen, Huang, & Tan, 2011). Therefore, an intelligent surveillance system that is able to solve background modelling problem, accurately track the objects and correctly classified the detected objects will produce a good prevention tool for security and good monitoring tool for surveillance purposes.

RELATED WORKS

Intelligent surveillance system for street scene may involve the application of monitoring the traffic, and handle congestion problem. There are many proposed techniques with different focus are proposed to solve the related in the street scene. One of the common approaches to be used is computer vision as taking the advantage of excess supply of CCTV on the streets. Normally for most of the street surveillance system, there are three major stages involved which are moving object detection, tracking and classification.

Particularly for moving object detections, the most common method used is background subtraction. Background subtraction is widely used method and there are a lot of proposed works related to background subtraction (Ridder, O. Munkelt, & Kirchner, 1995) (Stauffer & Grimson, 2000) (Elgammal, Duraiswami, Harwood, & Davis, 2002) (Mukerjee & Das, 2013) (Asaidia, Aarabb, & Belloukic, 2014) (Asif, Javed, & Irfan, 2014) (Hung, Pan, & Hsieh, 2014) (Lee & Lee, 2014) (Nimse, Varma, & Patil, 2014). Basically, in any background subtraction approach, the interested foreground is detected by image differencing and represented in binary form of foreground mask. Information of the foreground mask is essential to the higher level processing, which are tracking and classification. The main difference between each of the proposed methods in background subtraction is the background modelling stage. Stauffer et al. (2000), Mukerjee et al. (2013) and Nimse (2014) used Gaussian mixture model to model the background. Meanwhile, Elgammal et al. (2002) claimed that GMM is ideal for indoor scenes only. Thus, they introduced the use of kernel density estimation (KDE) for background reconstruction. Gao et al. (2009) and Lee et al. (2014) also reconstructed the background using KDE approach.

Meanwhile, for tracking stage, the appearance model gives a priori information about the interested object and can be updated for each new frame. The appearance can be modelled using shape, templates, histograms or parametric representations of distributions, extracted from the object detection stage. The second step in visual tracking is how to use the model to find the location of the object in the next frame. One of the simplest ways is by correlating the new frame with the appearance model and finding the maximum response. Recently, most of the surveillance systems rely on multiple features to model the object's appearance. A multi-feature fusion scheme has achieved high boosting performance or robustness in computer vision field (Deori & Thounaojam, 2014). The commonly used features for tracking are colour, texture, centroid, height, and width.

Similarly, in classification of moving objects, the features extracted from the object detection stage can be used to cluster the interested objects to different classes. There are several

remarkable methods that are proposed to classify the moving objects, which are, neuro-fuzzy classifier, support vector machine (SVM) classifier, K-nearest neighbourhood (kNN) and boosted classifier. Basically, the classification algorithm involves learning and testing phase. In learning phase, features extracted are learned by the classifier to recognize the distinctive attributes of the moving objects.

PROPOSED SYSTEM

Our proposed system consists of three real-time modules: (1) adaptive background modelling to effectively model the background of the scene for more accurate foreground object detection, (2) optimal trade-off features tracking to track moving objects across distributed overlapping and non-overlapping camera views and (3) detected objects classification to classify moving objects in the surveillance scene.

Object Detection

Object of interest in street surveillance are normally moving object which can be detected using background subtraction algorithm. We introduce a background subtraction algorithm which consist of Maximum Occurrence Patch based Background Modelling (MOP-BM). As pixel by pixel analysis is not preferable for real-time analysis, image frames are divided into several patches and the analysis of the patches will be done in order to reconstruct the background model.

MOP-BM utilizes the assumption that background pixels are the most frequently observed pixels in the entire video sequence. After the frames have been segmented into $m \times n$ patches, the first stage of MOP-BM is the maximum occurrence calculation of M patches. The process is repeated until $f_i = f_N$. Based on the maximum occurrence patches, the reconstruction of the background model is built. The patches with foreground objects will be omitted. Figure 1 shows the overall process of MOP-BM.

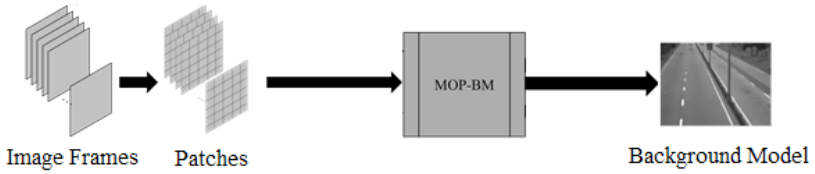


Figure 1. Details of MOP-BM algorithm

Let $f_1, f_2, f_3, \ldots, f_n$ represent the frames from the same video sequence. First, each of the frames must be segmented into $m \times n$ patches, where m and n can be the same number. Assume N patches are obtained, where we let each patch be marked as in the notation of $p_1, p_2, p_3, \ldots, p_N$. Experiments were conducted to find the optimum size of the patches. The result of the experiments shows that the size of the patches is not as sensitive as the sampling frame value. It does effect the result, but in a small percentage error. Then, at each frame, $f_1, f_2, f_3, \ldots, f_n$

the intensity value of pixel *(x,y)* in the area of specified patch, p_i is marked as i_1, i_2, i_3, …, in where $i = 1,2,3, \ldots, n$. Then, mean $(m_p f)$ at each frame for the specified patch is calculated as in the formula below:

$$m_{p,f} = \frac{\sum_{j=1}^{q} i_j(x,y)}{m \times n} \quad [1]$$

In order to calculate the MOP, the calculated mean is grouped into matrices called *mean_matrix*$_{p,f}$ according to the specified sampling time. Assume that the sampling time is denoted as *T*. Thus, the mode at that time interval for patch 1 is:

$$n_{mode} = \max(m_{1,1}, m_{1,2}, \ldots, m_{1,\tau-1}) \quad [2]$$

Multi-camera Tracking

Tracking the same object within different cameras' view is essential in many surveillance applications. In our tracking algorithm, colour, edges and local binary pattern (LBP) are chosen as the tracking features. For the colour feature, HSV colour space is chosen. The first step is to extract the HSV colour space from the intensity image. The second step is to compute the mean value for each HSV channel. The similarity of each target and candidate blob is computed based on Equation [3]. The object position is computed using Euclidean distance based on Equation [4].

$$dist(HSV) = \frac{\sqrt{(P(H_1) - P(H_2))^2 + (P(S_1) - P(S_2))^2 + (P(V_1) - P(V_2))^2}}{\sqrt{P(H_1)^2 + P(S_1)^2 + P(V_1)^2} + \sqrt{P(H_2)^2 + P(S_2)^2 + P(V_2)^2}} \quad [3]$$

$$d_{Euclidean} = (p_1 - q_1)^2 + (p_2 - q_2)^2 \quad [4]$$

where $p = (p_1, p_2)$ and $q = (q_1, q_2)$.

In order to make the system to be more robust, edge is used as an additional feature since it is insensitive to illumination. Object appearance may be differentiated in terms of its clothing such as checked shirt or plain shirt. It can be differentiated by using the edge feature of the clothing. The edge feature is obtained by using Canny Edge Detector. The frequency value of the edge is then used to calculate the similarity of tracked object. The similarity of each target blob and candidate blob are computed based on Equation [5].

$$dist(P(p_1) - P(p_2)) = \frac{\sqrt{(P(p_1) - P(p_2))^2}}{\sqrt{P(p_1)^2} + \sqrt{P(p_2)^2}} \quad [5]$$

The computational of LBP feature for each patch is represented by *mn* matrix, which is number of rows and columns respectively. The patch matrices are concatenated to obtain a set of decimal values. Then, the decimal values are divided into 5 groups of bins. The similarity between target and candidate blob are computed based on Equation [6].

$$dist(P(p_1) - P(p_2)) = \frac{\sqrt{(P(p_1) - P(p_2))^2}}{\sqrt{P(p_1)^2} + \sqrt{P(p_2)^2}} \quad [6]$$

The features are combined based on Equation [7], where *N* is the received candidates. Each candidate, has *K* different features, namely colour, texture and edge features. For each feature, similarity score is calculated and given a weight. The similarity scores are referred to Equation [3], [5], and [6].

$$O = \arg\max_{i \in N} \sum_{j \in K} (w_j s^i{}_j) \quad [7]$$

Classification Model

Our classification of moving objects is based on Adaptive Neuro-Fuzzy Inference System (ANFIS). ANFIS is a general artificial intelligence connectionist model of inference system, and its structure is shown in Figure 2. In this model, the reasoning system is based on Takagi-Sugeno-Kang (TSK) rules which constitutes of five layers, and the function of each layer is interdependent on each other.

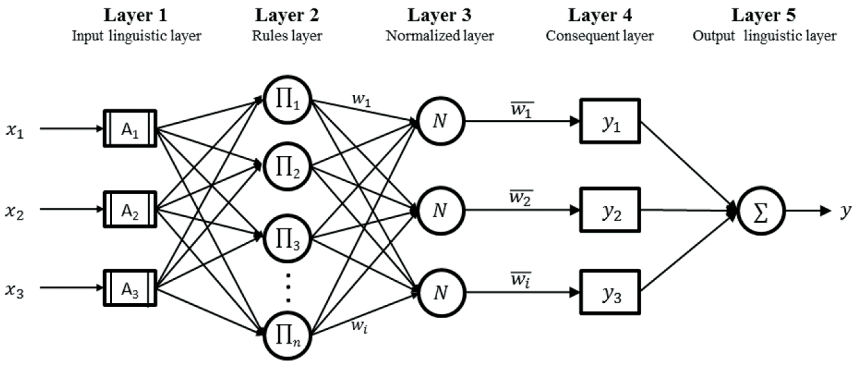


Figure 2. Structure of the ANFIS

In the training stage, three FIS were developed, in which each system is specifically for each output class. Each network has one output node, which will eventually be adjoined to recognized three different classes of moving objects including pedestrian, motorcycle, and car. Initially, the ANFIS training was done using our own datasets. All moving objects in the dataset are divided into three categories: pedestrian, motorcycle, and car. 200 positive samples, 1434 negative samples, and 200 test samples were randomly selected for each class in the ANFIS parameter training. The output of the network is in terms of probability value in the range of $[0, 1] \in \mathbb{Z}^+$. As the desired outputs were to be in finite value, a simple threshold is applied. Therefore, the outputs were (1, 0, 0), (0, 1, 0), and (0, 0, 1) for the pedestrian, motorcycle, and car respectively. During the testing stage, the feature vector of each region-of-interest that

detected by segmentation process will be extracted. This discriminant vector will be given to the network as an input. The output of linguistic layer will determine the belonging class of the moving object, which would be in the sequence of (y_1, y_2, y_3).

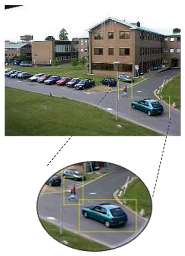
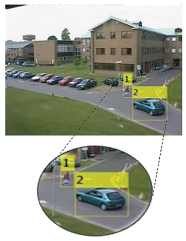
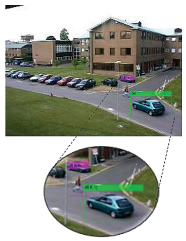
RESULTS AND DISCUSSION

PETS 2001 dataset is used to evaluate the performance of the proposed system. The dataset consists of 2688 frames with interested class of pedestrians and car. Table 1 shows the parameter setup for moving object detection, tracking and classification stage of the proposed system. Table 2 shows the result of proposed system for three random frames, 535, 578 and 596 respectively. From the table, we can deduce that, in the object detection stage, the moving objects are successfully detected and negative effect from the surrounding such as illumination variation are suppressed. The IDs for tracking the moving objects also managed to be assigned correctly. As for the moving classification, the car and pedestrian are annotated by the correct pre-defined blob colour for each classes.

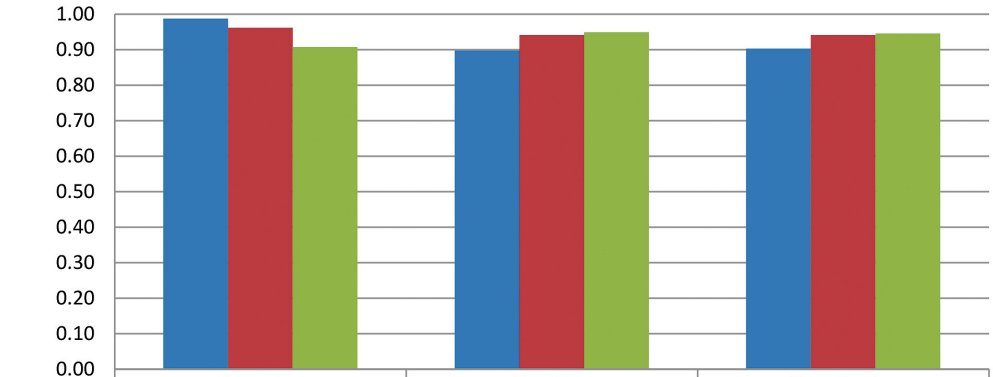
Table 1
Initial System Parameters

Stage	Parameter	Value
Pre processing	Frame spatial resolution	240×320 pixels
	Median filtering structuring element	5×5
Moving object detection	Sampling Frame	F_s= 80
	Patch Size	P_N=10×10
Moving object tracking	Threshold for Hue Colour	T_{HUE} = 0.27
	Threshold for Edge & Texture	$T_T T_E$ = 0.07
	Threshold for Top Weighted Ratio	T_{Cl_top} = 0.04
	Threshold for Bottom Weighted Ratio	T_{Cl_bottom} = 0.04

Table 2
Result of Proposed System for PETS 2001 Dataset

Image Frame	Detected Foreground	Object Detection	Object Tracking	Object Classification
#0535				
#0578				
#0596				

Graph in Figure 3 shows the overall performance for detection, tracking and classification for PETS 2001 dataset. From the quantitative result, we can see that the proposed system achieved more than 90% in accuracy, precision and recall value for detection stage. At same time, the proposed system also get high value of accuracy, precision and recall for tracking stage with 0.8974, 0.9412, and 0.9488 respectively. For the classification, the high value for accuracy, precision and recall indicates that the proposed system able to differentiate the moving objects accurately.



	Detection	Tracking	Classification
Accuracy	0.9875	0.8974	0.9026
Precision	0.9615	0.9412	0.9407
Recall	0.9067	0.9488	0.9453

Figure 3. Overall Performance of the Proposed System with PETS 2001 Dataset

CONCLUSION

This paper presents a solution for background modelling, object tracking and object classification for intelligent surveillance system. From the result of background modelling, it shows that the proposed system able to reconstruct the background successfully and there is no need to have the non-moving object in the video sequence to reconstruct the background. For tracking moving objects under distributed cameras' view, hybrid of features, namely position, edge, colour and texture are combined. From the experimental results, it shows that the tracking performance using multiple features benefit the system particularly on the accuracy value. For classification of the objects, the classifier was developed using Neuro-Fuzzy approach. From the experimental results, it shows satisfactory results in detecting different classes of objects which are the pedestrian and car under complex street scene condition. The proposed system would be very helpful for in increasing the efficiency of CCTV system.

REFERENCES

Asaidia, H., Aarabb, A., & Belloukic, M. (2014). Shadow elimination and vehicles classification approaches in traffic video surveillance context. *Journal of Visual Languages & Computing, 25*(4), 333–345.

Asif, S., Javed, A., & Irfan, M. (2014). Human Identification On the basis of Gaits Using Time Efficient Feature Extraction and Temporal Median Background Subtraction. *International Journal Image, Graphics and Signal Processing, 3*(2), 35-42.

Beale, M. H., Hagan, M. T., & Demuth, H. B. (2010). *Neural Network Toolbox 7. User's Guide.* MathWorks.

Breitenstein, M. D., Reichlin, F., & Leibi, B. (2011). Online Multiperson Tracking by detection from a Single, Uncalibrated Camera. *Pattern Analysis and Machine Intelligence, 33*(9), 1820-1833.

Chen, X., Huang, K., & Tan, T. (2011). Direction-based stochastic matching for pedestrian recognition in non-overlapping cameras. In *IEEE 18th International Conference Image Processing* (pp. 2065-2068). IEEE.

Deori, B., & Thounaojam, D. M. (2014). A Survey on Moving Object Tracking in Video. *International Journal on Information Theory (IJIT), 3*(3), 31-46.

Elgammal, A., Duraiswami, R., Harwood, D., & Davis, L. (2002). Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance. In *Proceedings of the IEEE*, *90* (pp. 1151-1163). IEEE.

Gao, T., Zhang, J., Gao, W., & Liu, Z. (2009). A Robust Technique for Background Subtraction in Traffic Video. In *15th International Conference on Neural Information Processing* (pp. 736-744). Springer Berlin Heidelberg, Auckland, New Zealand.

Hung, M. H., Pan, J. S., & Hsieh, H. C. (2014). A Fast Algorithm of Temporal Median Filter for Background Subtraction. *Journal of Information Hiding and Multimedia Signal Processing, 5*(1), 33-41.

Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks, 22*(9), 1341-1356.

Kim, K., Chalidabhongse, T. H., Harwood, D., & Davis, L. (2005). Real-Time Foreground-Background Segmentation using Codebook Model. *Real-Time Imaging, 11*(3), 172-185.

Lee, S., & Lee, C. (2014). Low-complexity background subtraction based on spatial similarity. *EURASIP Journal on Image and Video Processing, 2014*(30), 1-16.

Mukerjee, S., & Das, K. (2013). An Adaptive GMM Approach to Background Subtraction for Application in Real Time Surveillance. *International Journal of Research in Engineering and Technology, 2*(1), 25-29.

Nimse, M., Varma, S., & Patil, S. (2014). Shadow Removal Using Background Subtraction and Reconstruction. *International Journal of Emerging Technology and Advanced Engineering, 4*(4), 324-327.

Ridder, C., Munkelt, O., & Kirchner, H. (1995). Adaptive Background Estimation and Foreground Detection Using Kalman Filter. In *Proceedings of International Conference on Recent Advances in Mechatronics* (pp. 193-199). Istanbul, Turkey.

Shearing, C. D., & Johnston, L. (2013). Public and Private Interest Overlap. In *Governing Security: Explorations of Policing and Justice* (p. 121). Routledge.

Sonka, M., Hlavac, V., & Boyle, R. (2008). Image Processing, Analysis, and Machine Vision. In *Mathematical morphology* (pp. 559-597). Thomson.

Stauffer, C., & Grimson, W. E. (2000). Adaptive Background Mixture Models for Real-Time Tracking. In *IEEE Computer Society Conference* (Vol. 2). IEEE. Colorado.

Vilas, G. L., Spyrakos, E., & Palenzuela, T. (2011). Neural network estimation of chlorophyll from MERIS full resolution data for the coastal waters of Galician. *Remote Sensing of Environment, 115*(2), 524-535.

Zheng, Y., Ma, X., Zhao, X., & Wang, X. (2011). Mean shift target tracking algorithm based on color and edge features. *Journal of Optoelectronics, 8*(26), 1231–1235.

Pertanika J. Sci. & Technol. 25 (1): 191 - 200 (2017)

SCIENCE & TECHNOLOGY

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

Modelling the Effect of Sediment Coarseness on Local Scour at Wide Bridge Piers

Nordila, Ahmad[1]* Thamer, Mohammad[2], Melville, Bruce W.[3], Faisal, Ali[1] and Badronnisa, Yusuf[2]

[1]Department of Civil Engineering, Faculty of Engineering, National Defence University of Malaysia, Kem Sungai Besi, 57000 Kuala Lumpur, Malaysia

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

[3]Department of Civil and Environmental Engineering, University of Auckland, 20 Symonds St. Auckland1010, New Zealand

ABSTRACT

Experimental data from a physical model of scouring around a cylindrical wide pier embedded in two types of uniform sediment beds are presented. The effects of sediment sizes and various pier widths on scour development and equilibrium scour depth of wide bridge piers are described. Existing literature suggest that the empirical scour prediction equations based on laboratory data over-predict scour depths for large structures. The present study has attempted to fill this gap for a cylindrical wide pier. Further, equations for the estimation of non-dimensional maximum scour depth for a wide cylindrical pier embedded in uniform sediment were proposed as functions of the sediment coarseness.

Keywords: Physical model, scour, wide piers, uniform sediment, sediment coarseness

INTRODUCTION

Many studies have been carried out with the objective of developing maximum scour depth for bridge piers. However, studies on measuring the scour depth for wide bridge piers are limited. By using laboratory and field data for large piers, a correction factor, K_w, was introduced for wide piers in shallow flows (Johnson & Torrico, (1994; Arneson et al. (2012)). A wide pier is defined as a pier which is located in a shallow channel with low-velocity flow, and $y/b < 0.8$ for Froude

Article history:
Received: 02 March 2016
Accepted: 14 December 2016

E-mail addresses:
nordila@upnm.edu.my; nordilaahmad1@gmail.com (Nordila, Ahmad),
thamer@upm.edu.my (Thamer, Mohammad),
b.melville@auckland.ac.nz (Melville, Bruce W.),
fahali@gmail.com (Faisal, Ali),
nisa@upm.edu.my (Badronnisa, Yusuf),
*Corresponding Author

ISSN: 0128-7680

number < 0.8 (Johnson, 1999) while Sheppard et al. (2011) noted that piers have a width to sediment diameter ratio greater than 100. Sheppard et al. (2004) and Lee and Sturm (2009) showed that parameter b/d_{50} or sediment coarseness has a significant impact on equilibrium local scour depth at wide piers. The predictive equations in the literature are intended for large and small piers (Sheppard et al. (2011)). Therefore, this situation leads to the over-prediction of local scour at such piers, thus, leading to the use of unwarranted and costly foundations or countermeasures. Therefore, local pier scour experiments were performed in the laboratory to investigate the effect of sediment coarseness (b/d_{50}) using two uniform sediment sizes and five bridge pier models. The equilibrium local scour depth at wide piers can be expressed by an empirical relation and is discussed here. The aim of this study is to show the relationship between dimensionless pier scour depth and the ratio of pier width to sediment size over a large range of physical scales.

METHODOLOGY

Experiments reported in this paper were conducted in a flume 50.0 m long, 1.5 m wide, and 2.0 m deep located at the Hydraulics Laboratory of the National Hydraulic Research Institute of Malaysia (NAHRIM), which contains a 0.4-m-deep sediment recess. A test area is located in the brick-sided part of the flume in the form of a recess, 10 m long, filled with uniform sediment of up to 0.4 m depth. The flow water depth was maintained at 0.25 m and the water discharged was controlled by an inlet valve. Cohesionless uniform sediments were used as bed material with median particle sizes of d_{50} = 0.23 and 0.80 mm and geometric standard deviations of σ_g = 1.3 and σ_g = 1.26 respectively. As shown in Table 1, five different pier diameters, 0.06, 0.076, 0.102, 0.140, and 0.165 m, were chosen for this study. The study was conducted under clear-water conditions at a threshold flow intensity of $U/U_c \cong 0.95$. U_c was estimated using the Shields Function proposed by Melville and Coleman (2000).

RESULTS AND DISCUSSION

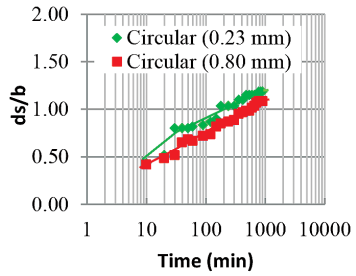
Ten experiments were carried out. All of them were run for a single cylindrical pier with various diameters. Table 1 records the main parameters of the 10 experiments.

Table 1
Summary of experimental results for the present study

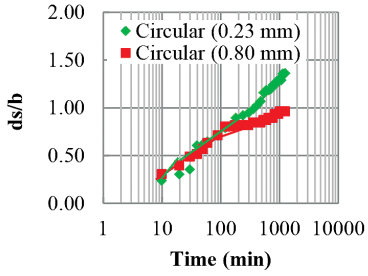
Pier	Velocity, U (m/s)	Critical velocity, U_c (m/s)	U/U_c	Time, t_e (h)	d_{50} (mm)	s_g	Diameter, b	d_s (m)
Pier 1	0.36	0.38	0.95	18	0.8	1.26	0.165	0.182
Pier 2	0.36	0.38	0.95	20	0.8	1.26	0.14	0.133
Pier 3	0.36	0.38	0.95	19	0.8	1.26	0.102	0.116
Pier 4	0.36	0.38	0.95	17	0.8	1.26	0.076	0.073
Pier 5	0.36	0.38	0.95	13	0.8	1.26	0.06	0.065
Pier 1	0.27	0.285	0.95	23	0.23	1.3	0.165	0.197
Pier 2	0.27	0.285	0.95	23	0.23	1.3	0.14	0.167
Pier 3	0.27	0.285	0.95	22	0.23	1.3	0.102	0.125
Pier 4	0.27	0.285	0.95	22	0.23	1.3	0.076	0.106
Pier 5	0.27	0.285	0.95	13	0.23	1.3	0.06	0.071

Scour Hole Development

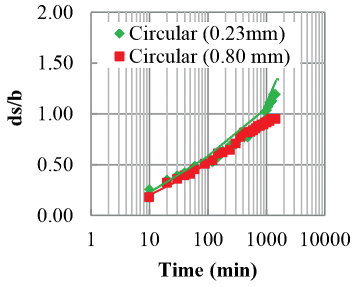
The temporal development of local scour with a value of flow intensity U/U_c = 0.95 in a uniform sediment bed around a cylindrical pier is shown in Figure 1 (a)–(e). The plotted graphs show that the scour development can be divided into three stages: (i) the initial stage, (ii) the main erosion stage, and (iii) the equilibrium stage. The results are analysed based on the abovementioned stages of local scour development.



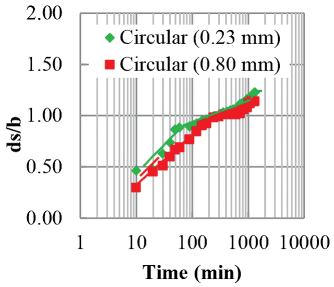
(a) b = 60 mm



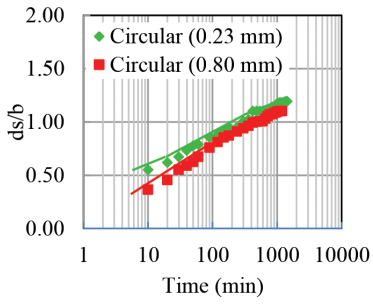
(b) b = 76 mm



(c) b = 102 mm



(d) b = 140 mm

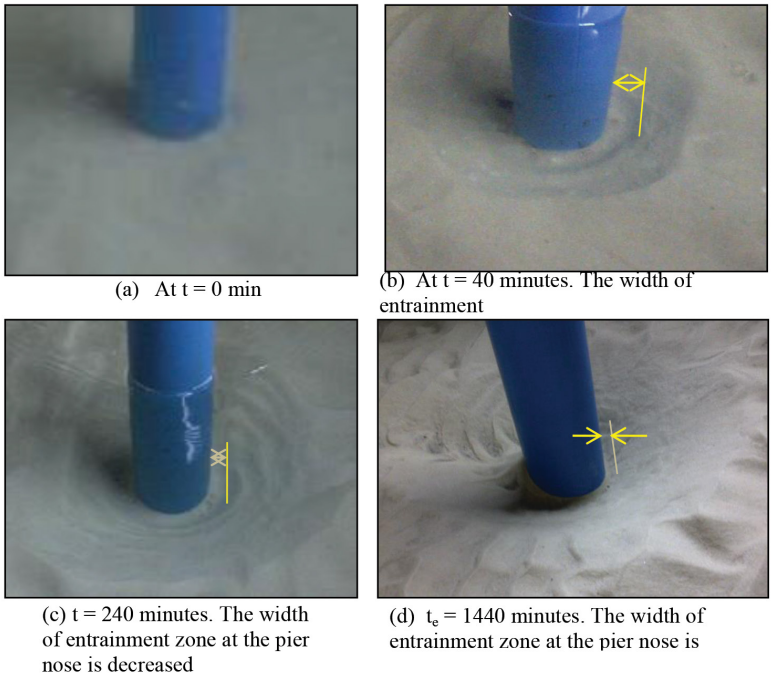


(e) b = 165 mm

Figure 1. Normalised local scour depth (d_s/b) versus time at the same value of b in sediment beds of d_{50}=0.23 and 0.80 mm

Initial Stage. In each experiment, the local scour began on the plane bed immediately around the pier and at locations approximately ± 45° from the direction of flow. At these locations, a small depression is formed in the sediment at either side of the pier. The scouring begins where the local shear stress exceeds the critical value for the entrainment of bed particles. From the observation, once the grooves have formed in the sediment at the sides of the pier, the bed shear stress pattern around the pier changes. The downflow near the base of the pier was observed to be deflected to the sides of the pier by the pressure gradients. Overall, at this stage the maximum depths of scour at the pier are $d_s/b \approx 0.3$ and 0.39 at $d_{50} = 0.80$ and 0.23 mm, respectively.

Main Erosion Stage. In order to observe the zone within the scour hole influenced by the bed sediment, the parameter b/d_{50} is used. This parameter compares the scale of the secondary flow within the scour hole to the size of the bed particles. Ettema (1980) analysed the influence of bed sediment on equilibrium scour depth within the range $b/d_{50} = 28–187$. He classified the sediment as fine relative to the pier width when $b/d_{50} > 130$ and intermediate size when $130 > b/d_{50}{}^3$ 8. Since two sediments were used in this study, it is possible to identify two regions of sediment coarseness, b/d_{50}, in which either the development of scour differed or the equilibrium depth of scour diminished with increasing values of b/d_{50}. The scour development of these regions was also observed using two different sizes of bed sediment. The regions can be classified as (a) $b/d_{50} > 230$, with d_{50}=0.23 mm; and (b) $b/d_{50} < 230$, with d_{50}=0.80 mm. The development of local scour for values of the sediment coarseness, b/d_{50}, corresponding to regions (i) and (ii) are shown in Figure 2 (a)–(d) and 3 (a)–(d). Figure 4 gives a more definitive illustration of the development of local scour for values of b/d_{50} in the two regions (i) and (ii).



(a) At t = 0 min

(b) At t = 40 minutes. The width of entrainment

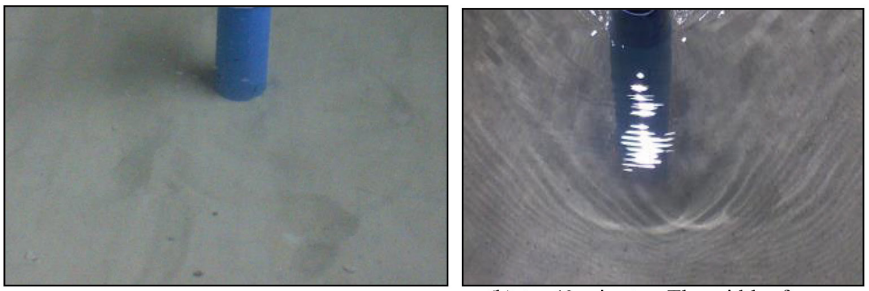
(c) t = 240 minutes. The width of entrainment zone at the pier nose is decreased

(d) t_e = 1440 minutes. The width of entrainment zone at the pier nose is

Figure 2. b = 140 mm, d_{50} = 0.23 mm, b/d_{50} = 609, U/U_c = 0.95

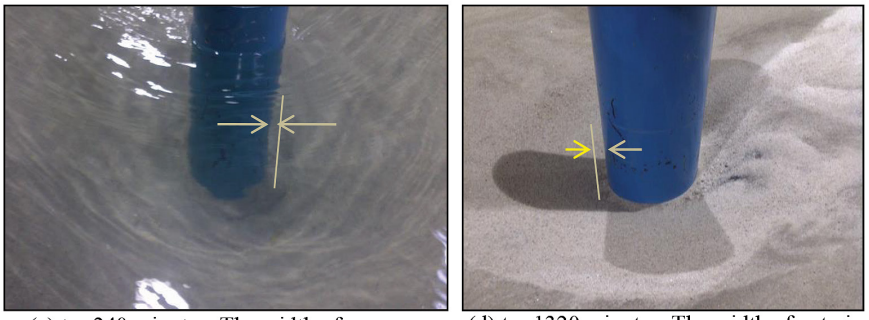
When b/d_{50}³ 230, the scour depth increases due to the strength of downflow in the vicinity of the pier nose. The downflow impinges on the base of the scour hole, eroding a nearly flat, narrow strip around the pier, and is deflected up the slopes of the scour hole to feed the horseshoe vortex. The measurement of the horseshoe vortex by Melville (1975) and observations of the motion of entrained sediment indicate that the entrainment zone corresponds to the contact area of the downflow impinging on the bed at the base of the scour hole with the underside of the horseshoe vortex on the slope of scour hole (Ettema et al., 2006). It is also observed that the erosion of sediment from the lower portion of the slopes in the scour hole disturbs the stability of the sediment so that it slides down into the entrainment zone each time the slope stabilises. Junliang and Xiong (2014) also added if the bed is erodible, such vortexes can enhance the transport of the sediment particles. The extent of the entrainment zone from the pier to about midway up the scour hole slope can be seen in Figure 2 (b). As the scour hole increases, the extent of the entrainment zone decreases until most of the sediment is entrained from the region in which the downflow near the pier impinges on the bed at the base of the scour hole. Ettema (1980) noted that from the depth of scour $d_s/b \approx 0.8–0.9$, the decrease in the extent of the entrainment zone indicates that the strength of the horseshoe vortex is decreasing. The photograph in Figure 2 (b)–(d) shows that the entrainment zone in this study decreases with time. An increasing strength of the horseshoe vortex from the flat bed ($d_s/b = 0–9$) followed by a decrease in the strength of the horse shoe vortex, was also observed in Melville (1975), where the author mentioned that the vortex expands in size but its rotational velocities decrease. Kirkil et al. (2008) also revealed the numerical solution where the interactions of the turbulences structures are particularly complex especially when the near wake region and horseshoe vortex system are merged. Figure 3 illustrates the scour process at b = 140 mm for a pier diameter placed in 0.80mm of bed sediment ($b/d_{50} = 175$). From the observations, at the beginning of the erosion stage the groove width increased to a value of approximately 0.11b, when the depth of scour was about $d_s/b \approx 0.5$, which is around t= 20–30 minutes for all pier sizes. As the scour depth increased to $d_s/b \approx 0.8$, the groove widened to approximately 0.24b and remained at this width until the equilibrium scour depth was attained for all pier sizes.

Equilibrium Stage. In uniform flow, the equilibrium scour depth is reached when the downflow impinging on the base of the scour hole is unable to erode particles from the scour hole. Figure 4 shows that the equilibrium scour depth decreased with increasing b/d_{50}. The curve indicates a high value of equilibrium scour depth achieved at $b/d_{50} = 330$ with b = 76 mm in fine sediment with $d_{50} = 0.23$ mm. The values of d_s/b for piers of 102, 140, and 165 mm show a little less than the peak value observed for the smaller piers at $d_{50} = 0.23$ mm.



(a) At t = 0 min

(b) t = 40 minutes. The width of entrainment is not clear



(c) t = 240 minutes. The width of entrainment zone at pier nose is indicated

(d) t = 1320 minutes. The width of entrainment zone at the pier nose is indicated

Figure 3. b = 140 mm, d_{50} = 0.80 mm, b/d_{50} = 175, U/U_c = 0.95

The smaller maximum equilibrium scour depths recorded for the larger pier sizes may be due to the greater localised scour of the bed surface around the rim of the scour hole at each of these piers than for the smaller piers. Ettema (1980) noted that the reductions of scour depth for larger pier sizes were influenced by the adjustment of bed level at the upstream rim of the scour hole. This corresponds to the occurrence of a small rotation in bed level immediately upstream of the scour hole for larger piers (140 and 165mm) with regard to the main erosion stage. However, this upstream erosion was not observed with smaller piers in the same sediments. The equilibrium scour depth at two smaller piers was attained before the localised scour of the bed surface around the scour hole could affect the development of the scour at the pier. This phenomenon did not happen with the large piers, for which a considerably longer period of time was required to reach an equilibrium depth. However, all pier sizes for cases where the sediment is coarse relative to b ($b/d_{50} < 230$) show lower values compared with cases where the sediment is fine relative to b ($b/d_{50} > 230$) except at b/d_{50} = 608. These trends may occur because the downflow effect in the local scour process is higher in finer sediment which creates a deep hole in the vicinity of the pier base. Nicolet (1971) reported a similar peak of d_s/b at a pier diameter < 0.1 m and slightly lower equilibrium scour depths for larger pier sizes.

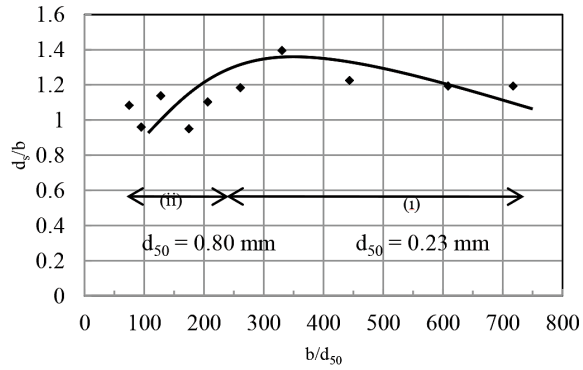


Figure 4. Equilibrium scour depth versus b/d_{50} for the present study

Effect of Sediment Coarseness

In order to demonstrate a wider range of applicability, laboratory data from the present study were merged with those obtained from other researchers. Figure 5 shows the equilibrium scour depth versus b/d_{50} for a large range of values of sediment coarseness. The continuous reduction in the dependence of d_s/b on b/d_{50} with increasing b/d_{50} differs from that noted by other researchers (Ettema, 1980; Jones & Sheppard, 2000; Sheppard & Miller, 2006; Lee & Sturm, 2009). The present study shows that the scour depth values were lower than those found by other researchers with a range of 0.95–1.39 and the maximum value of d_s/b occurred in ripple-forming sediment. Ettema (1980) showed that the range of maximum values of d_s/b in ripple-forming sediments with d_{50}< 0.7 mm were between 0.90 and 1.70. However, Rui et al. (2013) mentioned that the values was not so critical for sand bed d_{50} < 2mm and Melville (2008) showed that for a value of the shear velocity ratio U/U_c approaching unity or the threshold of motion condition for the bed sediment, a planar bed composed of fine sand can gradually develop ripples. This is because planar beds of ripple-forming sediments are unstable for bed shear stresses approaching the critical shear stress from particle entrainment. The formation of ripples on an initially planar bed surface increases the roughness of the bed and the mean bed shear stress, causing a low intensity of sediment transport into the scour hole. This could be the principal reason for the lower value of d_s/b indicated by this study compared with the findings of other researchers.

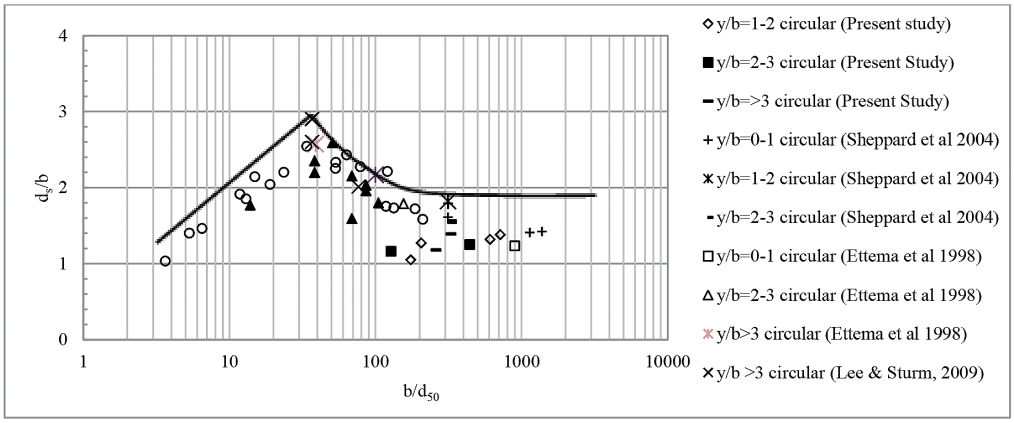


Figure 5. Effect of b/d_{50}on d_s/b

Data in Figure 5 are plotted as the upper envelope in the same figure and are represented by the following equation:

$$\frac{d_s}{b} = 0.05 \left(\frac{b}{d_{50}} \right) + 1.11 \qquad 4 \le b/d_{50} \le 37 \qquad [1]$$

$$\frac{d_s}{b} = \frac{2}{(0.027 \frac{b}{d_{50}} - 0.6)^{1.4} + 1.3} + 1.8 \qquad 37 \le b/d_{50} \le 1 \text{ x } 10^4 \qquad [2]$$

These plots also clearly show the reduced dependence of d_s/b on b/d_{50} with increased values of b/d_{50}, as mentioned by Sheppard et al., (2004); Lee and Sturm (2009); Rui et al., (2013) and Sheppard et al., (2014). All the values of the predicted d_s/b were compared with the values of the observed d_s/b obtained from the experimental work in this study and selected literature. Figure 6 compares d_s/b predicted with d_s/b observed using Equation [1] and [2]. It was demonstrated that values of d_s/b using Equation [1] and [2] are closer to the observed values with less under-prediction.

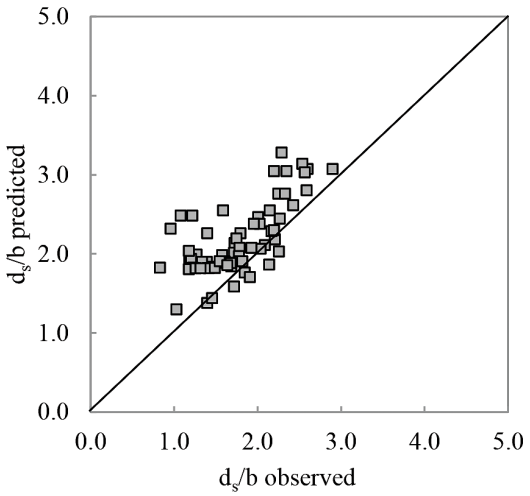


Figure 6. Comparison of observed values of ds/b around wide piers with those predicted using Equation [1] and Equation [2]

CONCLUSION

In conclusion, the temporal and scour hole development of local scour depth with new laboratory data is demonstrated for the flow intensity $U/U_c = 0.95$. The effects of local scour depth at wide piers on scour hole development and sediment coarseness were presented. It was found that the relative scour depth d_{se}/b decreased with increasing values of sediment coarseness

b/d_{50}. Two continuous upper envelope equations for pier scour depth as a function of b/d_{50} were developed according to the value of b/d_{50} (≤ 37 and > 37). It was proved that for large values of b/d50, the values of dse/b become reduced., and this was consistent with results of experiments using very large flumes as well as findings by previous researchers. Experimental results and analysis of the effects of pier width and bed sediment size on scour depth validate the theory of wide piers for the most part. However, the results of the present study are limited to clear-water conditions, steady flow, non-cohesive sediments, and subcritical flow conditions. An important practical limitation of laboratory experiments on pier configuration is flume width; the size of the flume is not wide enough to facilitate scour experiments for large values of b.

ACKNOWLEDGEMENTS

The financial support via the Research University Grant Scheme (RUGS), University Putra Malaysia (Grant No. 05-01-10-0904RU), for this project is acknowledged. The experiments were conducted at the Hydraulic Laboratory in the National Hydraulic Research Institute of Malaysia (NAHRIM).

REFERENCES

Arneson, L. A., Zevenbergen, L. W., Lagasse, P. F., & Clopper, P. E. (2012). Evaluating scour at bridges (4th Ed.). *Hydraulic Engineering Circular No. 18* (HEC-18). Federal Highway Administration, Washington, DC.

Ettema, R., (1980). *Scour around bridge piers.* Report No. 216, University of Auckland, Auckland, New Zealand.

Ettema, R., (1980). *Scour around bridge piers.* Report No. 216, University of Auckland, Auckland, New Zealand.

Ettema, R., Kirkil, G., & Muste, M. (2006). Similitude of large-scale turbulence in experiments on local scour at cylinders. *Journal of Hydraulic Engineering, 132*(1), 33-40.

Johnson, P. A., & Torrico, E. F. (1994). Scour around wide piers in shallow water. *Transportation Research Record,* (1471), 66-70.

Johnson, P. A. (1999). Scour at Wide Piers Relative to Flow Depth, Stream Stability and Scour at Highway Bridges. In E. V. Richardson & P. F. Lagasse (Eds.), *Compendium of ASCE Conference Papers* (pp. 280–287).

Jones, J., & Sheppard, D. (2000). *Scour at wide bridge pier* (pp. 1-10). Federal Highway Administration, Turner-Fairbank Highway Research Center, McLean, Virginia.

Junliang, T., & Xiong, Y. (2014). Flow and Scour Patterns around Bridge Piers with Different Configurations: Insights from CFD Simulations. In *Geo-Congress 2014 Technical Papers: Geo-characterization and Modeling for Sustainability* (pp. 2655-2664). ASCE.

Kirkil, G., Constantinescu, S. G., & Ettema, R., (2008). Coherent structures in the flow field around a circular cylinder with scour hole. *Journal of Hydraulic Engineering, 134*(5), 572–587.

Lança, R. M., Fael, C. S., Maia, R. J., Pêgo, J. P., & Cardoso, A. H. (2013). Clear-water scour at comparatively large cylindrical piers. *Journal of Hydraulic Engineering, 139*(11), 1117-1125.

Lee, S. O., & Sturm, T. W. (2009). Effect of sediment size scaling on physical modeling of bridge pier scour. *Journal of Hydraulic Engineering, 135*(10), 793-802.

Melville, B. W., & Coleman, S. E., (2000). *Bridge Scour*. United States of America, USA: Water Resources Publications.

Melville, B.W., (1975). *Local scour at bridge sites*. Report No.117. School of Engineering, University of Auckland, New Zealand.

Melville, B. W. (2008). The physics of Local Scour at Bridge Piers. In *Fourth International Conference on Scour and Erosion* (pp. 28-38). Tokyo.

Nicolet, G. (1971). *Deformation des lits alluvionaires affouillements autor des piles se ponts cylindriques*. Report No. HC 043 684. Laboratoire National d'Hydraulique, Chatou, France.

Sheppard, D. M., Huseyin, D., & Melville, B. W. (2011). *Scour at wide piers and long skewed piers*. Report (National Cooperative Highway Research Program); 682. Washington, D.C.: Transportation Research Board.

Sheppard, D. M., Melville, B., & Demir, H. (2013). Evaluation of existing equations for local scour at bridge piers. *Journal of Hydraulic Engineering, 140*(1), 14-23.

Sheppard, D. M., & Miller, J. W. (2006). Live-bed local pier scour experiments. *Journal of Hydraulic Engineering, 132*(7), 635-642.

Sheppard, D. M., Odeh, M., & Glasser, T. (2004). Large scale clear-water local pier scour experiments. *Journal of Hydraulic Engineering, 130*(10), 957-963.

Pertanika J. Sci. & Technol. 25 (1): 201 - 210 (2017)

SCIENCE & TECHNOLOGY

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

Electromyography Analysis during Lifting Tasks: A Pilot Study

Siti Anom, Ahmad[1]*, Nor Hafeezah, Kamarudin[1], Rosnah, Mohd Yusoff[2], Mohd Khair, Hassan[1] and Siti Zawiah, Md Dawal[3]

[1]Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[2]Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[3]Department of Engineering Design and Manufacture, Faculty of Engineering, Universiti Malaya, 50603 UM, Kuala Lumpur, Wilayah Persekutuan, Malaysia

ABSTRACT

Manual Material Handling (MMH) involves lifting, bending, and twisting motions of the torso. Poor lifting technique is often considered a major risk factor in low back injury associated with manual lifting tasks. Currently, there is little work on the effects of lifting on the Malaysian population. The MMH activities that was designed with the different lifting heights, frequency, weight of loads and the effect on of biceps and triceps muscle contraction of the subjects during the lifting tasks were studied. The parameters involved are weight of the loads lifted, height of the loads lifted and lifting frequency as the independent variables. Whereas the dependent variable is Electromyography (EMG) signal. The weight loads are varying from 10kg up to 24kg and the heights of the loads travels from the floor to 70cm and 130cm heights. The frequency of lifting is set to 1 lift and 6 lifts per minute. 14 healthy male and female subjects were recruited in this study. The questionnaires and consent form were used to identify the health condition of the subjects before performing the lifting tasks. The EMG activity was recorded and collected from biceps and triceps muscles using the Shimmer EMG system. This method is used in determining the maximum acceptable weight limit (MAWL) that can be lifted by the subjects in the lifting tasks. This research aims to design a lifting equation that suits for Malaysian people. Therefore, the effects of different manual lifting tasks on Malaysian physiological limits need to be identified.

Keywords: Electromyography, frequency, height, lifting tasks, load

Article history:
Received: 02 March 2016
Accepted: 14 December 2016

E-mail addresses:
sanom@upm.edu.my (Siti Anom, Ahmad),
nhafeezah@gmail.com (Nor Hafeezah, Kamarudin),
rosnahmv@upm.edu.my (Rosnah, Mohd Yusoff),
khair@upm.edu.my (Mohd Khair, Hassan),
sitizawiahmd@um.edu.my (Siti Zawiah, Md Dawal),
*Corresponding Author

INTRODUCTION

Manual material handling (MMH) involving acts such as lifting, lowering, bending, twisting, holding and others are the tasks

ISSN: 0128-7680

and activities that involved in our daily life either in one time or frequent as part of regular works. Manual handling tasks occurs in many industries such as constructions, manufacturing, healthcare, hotels, agriculture, restaurant and other industries. In 1981, the National Institute for Occupational Safety and Health (NIOSH) recognised the growing problem of work related back injuries. This problem is contributing to tremendous health care costs, human suffering and lost productivity in company (Elfeituri & Taboun, 2002), (Waters, Anderson, Garg & Fine, 1993), (Rud, 2011), (Medina & Vina, 2012). The repetitive manual lifting of objects and materials will affect potential injuries to the workers involving the musculoskeletal problems (Potvin, 2012). Improper lifting techniques will cause back pain of the workers and also can lead the major loss to the company. Back pain is one of the factors of increasing the medical and compensation cost of the company (Shy, 2008). The performance of the workers such as productivity and quality will also be decreased. The two most prevalent musculoskeletal problems are low back pain and upper extremity cumulative trauma disorders (Marras, Leurgans, Fathallah, Ferguson & Allread, 1995). In MMH, the ergonomic risk factors such as awkward working posture, excessive load, and extreme temperature can also contribute to occupational injuries to workers (Kuorinka, Lortie, & Gautreau, 1994). Risk and back injury are exposed to the workers when they are doing the manual material handling of loads such as transporting and supporting the loads in an unfavourable ergonomic condition (Marras, Leurgans, Fathallah, Ferguson & Allread, 1995), (Kuorinka, Lortie, & Gautreau, 1994).

METHODOLOGY

The material and method were setup for this experiment including posture, variables, procedures and also the equipment used for this experiment.

Subject Selection

A total of 14 healthy subjects between the ages of 23 to 27 years, with no history of back pain were studied. The average height and weight of the subjects were 1.66m and 64.5 kg respectively. A questionnaires and consent form prior to the data collection was given to the subjects. The procedure for conducting the experiment was approved by the Ethics Committee.

Experimental Setup

There are several equipment's used to measure the parameters such as Shimmer EXG Development kit, adjustable table, loads, weight scale and lifting box. Shimmer EMG kit is used as main equipment to display and record the EMG signal of the subjects for each lifting tasks. EMG signal records the electrical activity that associated with muscle contraction and muscle response collected from biceps brachii and triceps brachii.

Lifting Variable. Two types of variable were measured in this experiment, which are the independent and dependent variable.
For independent variable, they are:

a) Weight Loads (L)
   The loads are 10kg, 15kg, 20kg, 23kg and 24kg, placed at the centre of the lifting box.
b) Lifting Height (H)
   The load will be lift started from the floor and increase at two different heights; 70cm and 130cm.
c) Lifting Frequency (F)
   The number of lift; 1 lift per minute and 6 lifts per minute.

While for the dependent variable, EMG was used.
The fixed variables are squat lifting posture, distance of the subject to the table at 0.25m.

Lifting Posture. Squat lifting posture as in Figure 1 used during this lifting tasks. People mostly tend to use squat posture during lifting due to the less energy expenditure used.

Figure 1. Squat posture

Adjustable Table. The adjustable table as in Figure 3 can be adjusted to desired height for lifting tasks which are 70cm and 130cm.

Figure 2. Adjustable Table

Lifting Box and Loads. Lifting box used to carry the load. The box shown in Figure 4 has handles to reduce the grasp force and good coupling grasp. Various weights of loads (L) used in the experiment from 10 kg, 15kg, 20kg, 23kg, and 24kg.

Figure 3. Lifting box and loads

Lifting Procedures. The lifting procedures shown in Figure 5.

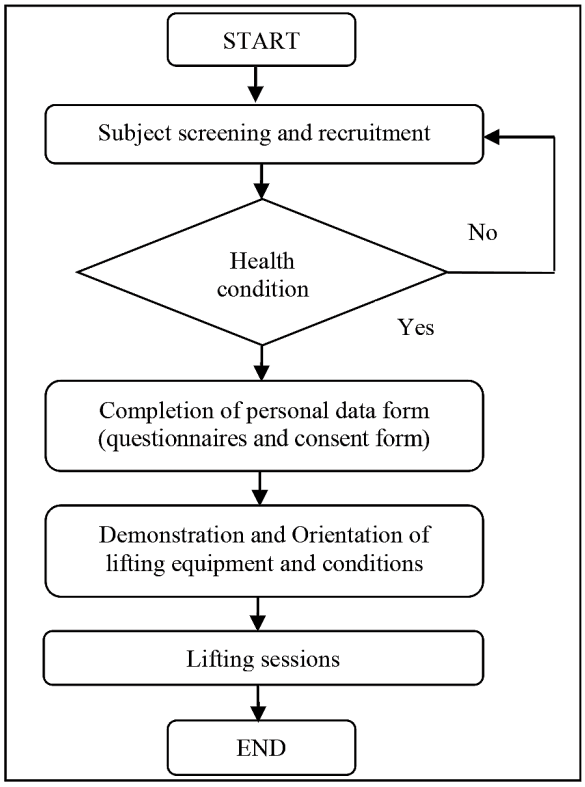


Figure 4. Flowchart of Lifting Procedures

Data Collection

The EMG signals were recorded during the lifting activities. EMG surface electrodes (Ag/AgCI, 10 mm diameter, 20mm inter-electrode distance) were placed near the centre of biceps brachii and triceps brachii muscle. The skin surface was prepared with alcohol prior to electrode attachment and the grounding electrode was placed on ulna bone since it has no effect of electrical biosignal as in Figure 5. The EMG signal shows the level of muscle contraction when the heights and loads of lifting task are varied

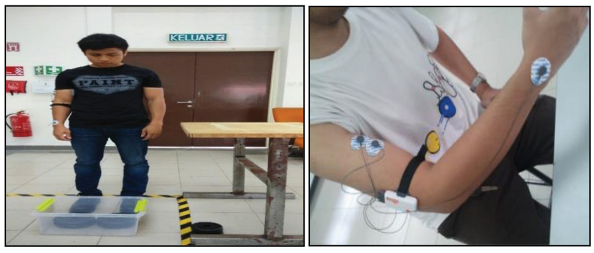


Figure 5. Shimmer EMG electrode placement and lifting setup

After placement of the electrodes, the subjects were asked to contract the muscle like bending the arm. The electrical activity can be seen on the monitor. The signal provides information about the muscle's ability to respond when the muscle nerves are stimulated. The more the muscle contract such as lift the heavier load, the electrical activity increases and a pattern can be seen. This electrical activity pattern can determine how the muscles respond during the lifting task while the loads of lifting were increased.

Lifting Session. Three lifting sessions involved in this experiment. Subjects will be screened by answering questionnaires and short briefing will be given on how to perform the tasks before each session started.

Subjects were given sufficient time to practice and familiarize with the task before actual data were recorded.

Table 1
Lifting Session

Lifting Parameters	**Lifting Session**		
	Session 1	**Session 2**	**Session 3**
Weight Loads	10 kg,15kg, 20kg, 23 kg and 24 kg		
Height	70cm	70cm	130cm
Frequency	1 lift / minute	6 lifts / minute	1 lift / minute 6 lifts / minute

The subjects required to do three major tasks of lifting which included the weight load varies from 10kg to 24kg as summarised in Table 1. The task performed by the subjects consisted of lifting the load from the floor to the desired height level of 70cm for the first and second tasks and increase to 130cm for the third tasks. The frequencies of lifting task were assigning to 1 lift per minute and 6 lifts per minute. During the lifting task, the subjects had the option to withdraw if they cannot proceed to the next stage of lifting tasks with heavier loads.

RESULTS AND DISCUSSION

The results of this study are to identify on how the muscle contraction during the lifting activities at each combination of height and loads together with different lifting frequency.

EMG Signal of Muscle Contraction

Figure 6 shows the relationship between mean of EMG signal and weight of subjects in Session 1. The EMG signal was inversely proportional to the weight of subjects. The force required to lift the load decreases as the weight of subject increases. Hence, the muscle contraction of bigger subjects was lower than the smaller subjects.

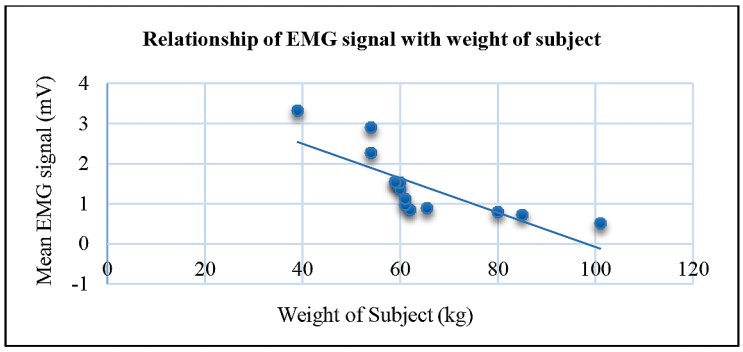


Figure 6. Relationship between EMG Signal and Weight of Subject

Figure 7 refers to Subject 1 with the height of 1.63m and weight of 80 kg in Session 1 lifting task at 70cm lifting height. The subject can lift the entire weight load from 10kg to 24kg. The amplitude of EMG signal increased constantly with the weight loads increased. In Figure 8 which represents the Subject 2 with height of 1.45 m and weight of 39 kg, it shows that the subject only can lift the loads of 10kg and 15kg. The EMG signals shows the subject needs a lot of energy from the biceps muscle to lift the heavy loads. The increasing of the EMG signal when the weight loads is increased associated with the energy that they use up to lift the loads.

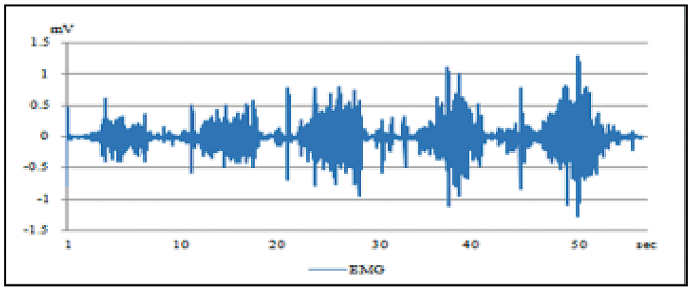


Figure 7. EMG signal of Subject 1 in Session 1

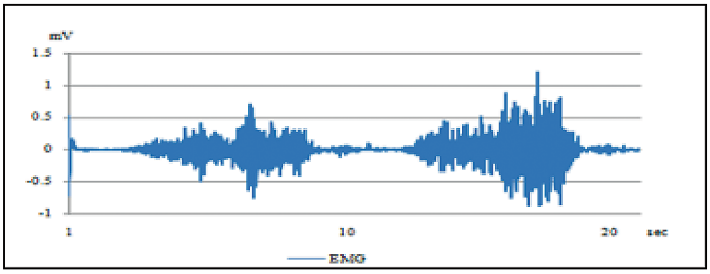


Figure 8. EMG signal Subject 2 in Session 1

Figure 9 and 10 represents the Session 2 for repetitive tasks when the loads were lift 6 lifts per minute of 10kg weight load. This result is based on the same subjects in Session 1. The different capability of different weight of subjects can be seen through the peak amplitude of the EMG signal produced. Figure 8 shows the peak amplitude of EMG signal for Subject 1 is less than 0.6mV. However, Subject 2 represented in Figure 9, the peak amplitude of EMG signals is nearly 1.0mV. The repetitive task of lifting frequencies also will influence the performance of muscle. This graph shows when the muscle contraction is increased, the biceps muscle consumes more energy and the higher the energy consumption, it will induce more fatigue to muscle of the subjects.

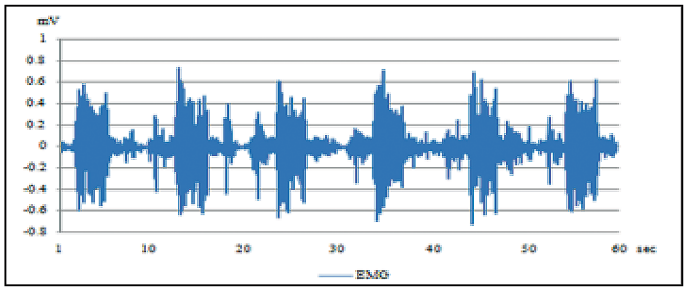


Figure 9. EMG signal Subject 1 in Session 2

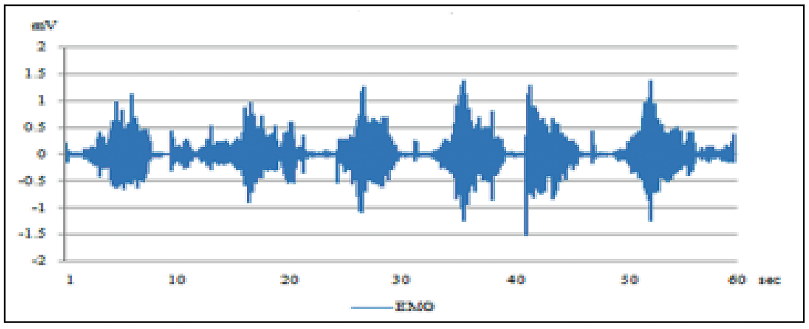


Figure 10. EMG signal Subject 2 in Session 2

Maximum Acceptable Weight Limit (MAWL)

Figure 11 shows the relationship between the weights of subject is directly proportional with the lifting weight loads. When the weight of subject is increased, the weight of loads increased. The bigger size of subjects able to lift the loads from 10kg to maximum of 24kg for 1 lift per minute task since they did not use much energy to lift the loads according to their size of body and strength. This load is within their ability without affecting any back pain problem.

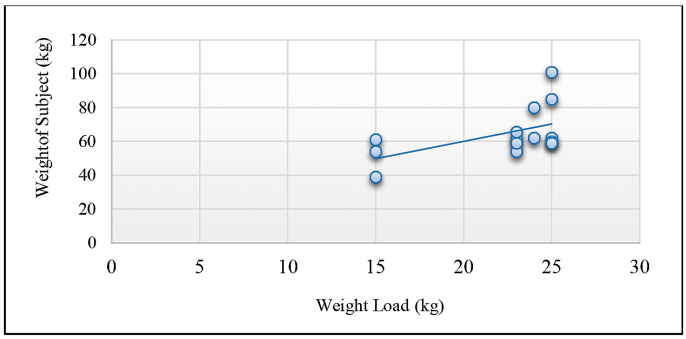


Figure 11. Weight of Subject and Maximum load in Session 1

Based on the Figure 12, it shows that the lifting frequency also effect the capability of the subjects to reach the maximum loads of lift since the number of subjects who can lift the maximum weight loads decrease in session of 6 lifts per minute. This is due to the force and energy are decreasing when the lifting frequency increases and the muscle become fatigued because of repetitive lifting task.

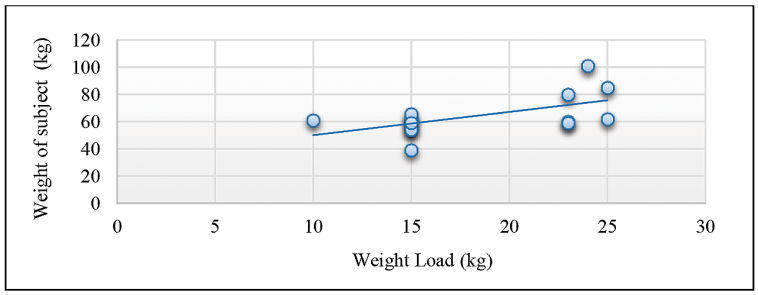


Figure 12. Weight of Subject and Maximum load in Session 2

CONCLUSION

Based on this preliminary experiment, the weight of the loads, height of lifting and lifting frequency could affect how the human body performs the lifting activities in terms of capability to lift the loads concerning of the weight of the subjects and energy used during the lifting tasks. The relationship between EMG signal was inversely proportional to the weight of subjects. The force required to lift the load decreases as the weight of subject increases. Since the smaller subjects need larger force to lift the load, they also required more energy to lift the load compared to the bigger subjects. Hence, the muscle contraction of EMG signal of smaller subjects will be higher than the bigger subjects. The increasing of lifting weight load and height have increased the contraction of EMG signal in the human muscle. The lifting frequencies also affect the muscle strength during the lifting tasks. Further analysis of this research is to define the recommended weight limit that suitable and safe for Malaysian people especially who work in industries to decrease the musculoskeletal disorder and injuries related to manual lifting task and also to increase the safety awareness in the workplace.

ACKNOWLEDGEMENTS

Thanks for the assistance of Siti Nur Rabani Abd Rahman, Amira Aminuddin, and technicians during the collection of data and evaluation of subjects. This research was funded by eScience Fund Ministry of Science, Technology and Innovation (MOSTI) Malaysia

REFERENCES

Elfeituri, F. E., & Taboun, S. M. (2002). An Evaluation of the NIOSH Lifting Equation: A Psychophysical and Biomechanical Investigation. *International Journal of Occupational Safety and Ergonomics, 8*(2), 243–258.

Kuorinka, I., Lortie, M., & Gautreau, M. (1994). Manual Handling in Warehouses: The Illusion of Correct Working Posture. *Ergonomics*, *37*(4), 655 – 661.

Marras, W. S., Lavender, S. A., Leurgans, S. E., Fathallah, F. A., Ferguson, S. A., Allread, W. G., & Rajulu, S. L. (1995). Biomechanical Risk Factors for Occupationally Related Low Back Disorders. *Ergonomics, 38*(2), 377-410.

Medina, Y. T., & Vina, S. (2012). Evaluation and Redesign of Manual Material Handling in a Vaccine Production Centre's Warehouse. *A Journal of Prevention, Assessment and Rehabilitation*, *41*(Supplement 1), 2487-2491.

Potvin, J. M, (2012). Predicting Maximum Acceptable Efforts for Repetitive Tasks: An Equation Based on Duty Cycle. *Journal on Human Factor*, *54*(2), 175-188.

Rud, S. (2011). *An Ergonomic Analysis of the Current Lifting Techniques in Height Restricted Cargo Bins at Company XYZ*. (Doctoral dissertation). University of Wisconsin-Stout.

Shy, L. H. (2008). Ergonomic Intervention to Reduce the Risk of Musculoskeletal Disorders (MSDs) for Manual Materials Handling Tasks. Project Report. UTeM Malaysia.

Waters, T. R., Anderson, V. P., Garg, A., & Fine, L. J. (1993). Revised NIOSH equation for the Design and Evaluation of Manual Lifting Tasks. *Ergonomics, 36*(7), 749-776.

Pertanika J. Sci. & Technol. 25 (1): 211 - 220 (2017)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Anaerobic Digestion of Domestic Wastewater in different Salinity Levels: The Adaptation Process

Syazwani, Idrus*, Nik Norsyahariati, Nik Daud and Amimul Ahsan

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

ABSTRACT

The effect of osmotic stress was carried out to determine the resistance to salt toxicity using 4 Continuous Stirred Tank Reactor (CSTR). A CSTR digestion study revealed that digesters seeded with an inoculum from a conventional mesophilic digester treating municipal wastewater and fed on domestic wastewater (DW) plus salts were able to acclimate successfully to a final salt concentration of 10 g l^{-1}. The digesters showed some disturbances during the acclimatisation period as indicated by reductions in specific methane production (SMP), specific biogas production (SBP), pH and increases in Intermediate Alkalinity /Partial Alkalinity (1A/PA) ratio and Volatile Fatty Acid (VFA) concentration. This study revealed the order of disturbance was Sodium Chloride (NaCl) > Potassium Chloride (KCl) > KCl + NaCl. The average values for SMP after stabilisation were below those in the controlled digester, at 0.335 (controlled), 0.323 (NaCl), 0.316 (KCl + NaCl) and 0.308 l CH4 g-1 COD added (KCl).

Keywords: Adaptation, chemical oxygen demand, inhibition, osmoregulation theory, specific methane production

INTRODUCTION

Salinity in wastewater is known to lead to low Chemical Oxygen Demand (COD) removals in the anaerobic digestion (AD) due to the loss of activity of organisms (Kargi & Dincer, 1996). Concentrations of salt more than 1% were reported to produce inhibitory or toxic effects on microbes not adapted to high salinity; high salt concentrations (>1%) have been shown to cause plasmolysis and/or loss of activity of cells. Some studies have reported a detrimental effect on COD removal efficiency immediately after addition of salt, due to leading to changes in surface charge leading to alterations in the community Reid

Article history:
Received: 02 March 2016
Accepted: 14 December 2016

E-mail addresses:
syazwani@upm.edu.my; syazwanii@yahoo.com (Syazwani, Idrus),
niknor@upm.edu.my (Nik Norsyahariati, Nik Daud),
amimul@upm.edu.my (Amimul Ahsan),
*Corresponding Author

ISSN: 0128-7680

et al. (2006). This finding is consistent with Abbott et al. (2015), Kumar et al. (2016) and Azman et al. (2016) who reported that shocking with salt caused a decrease in digestion efficiency.

Several studies have looked at the application of anaerobic treatment by ordinary (i.e. non-halophilic) methanogenic consortia for the removal of organic pollutants in highly saline wastewaters. Abbott et al. (2015) studied the effects of metal salt addition on odour and process stability during the anaerobic digestion of municipal waste sludge. Other research which employed non halophilic methanogens was conducted by Rovirosa et al. (2004) using down-flow anaerobic fixed bed reactors for the treatment of piggery effluent. The study revealed that 90% of COD removal at an Organic Loading Rate (OLR) of 0.5 kg COD m^{-3} day^{-1} at a salt concentration of 15 g l^{-1}. The used of non-halophilic groups in an up-flow anaerobic filter has been demonstrated by Guerrero et al. (1997) during the treatment of seafood processing effluent at 15 g l^{-1} of salts and revealed that 83% of COD removal at COD influent of 34 g l^{-1} and OLR of 2.8 kg COD m^{-3} day^{-1}. Furthermore, Astals et al. (2013) also reported thermophilic co-digestion of pig manure and crude glycerol. Kumar et al. (2015) reported correlations between chemical oxygen demand (COD) and pH and (Calcium Chloride) $CaCl_2$ where the reduction of $CaCl_2$ concentration caused pH to increase thus lead to reduction of COD. Anaerobic digestion has not worked well in all such cases, however, and there are reports of saline wastewaters being inhibitory in AD systems when employing non-halophilic methanogens. Gebauer (2004) investigated the recycling of fish farm wastewater with an influent COD of 70.1 g l^{-1} and salt concentration at 35 g l^{-1} using a Completely Stirred Reactor (CSTR). This study found only 50% of COD removal rate at OLR of 2.5 kg COD m^{-3} day^{-1}. Feijoo et al. (1995) demonstrated the effect of antagonism and the adaptation process in treating seafood processing wastewater in an AD system. This experiment was conducted in a continuous digester by introducing three different anaerobic digestates including; (1) sludge from an AF system after 2 years treating mussel wastewater, (2) suspended biomass from a central digester treating a mixture of seafood processing wastewater and (3) sludge from UASB which previously treated potato processing wastewater. This study found that 50% of inhibition at concentration of 3 - 16 g Na l^{-1} but reported adaptation of methanogens to the saline environment, with increased Na tolerance and a shorter lag phase before the onset of methane production.

Microorganisms requiring salt for growth are designated halophilic and halotolerant, and have developed more acidic enzymes that require the ion potassium (K) for optimal activity. Therefore, they can maintain an osmotic balance of their cytoplasm with the external medium by accumulating high concentrations of various organic osmotic solutes (Lai & Gunsalus, 1992). This suggests the potential of using salt tolerant microorganisms in biological saline wastewater treatment. Furthermore, halophilic and halotolerant microorganisms require a high salt concentration, particularly of K, for most biochemical reactions and lower concentrations may lead to cell disintegration (Wood, 2006).

Studies on the treatment of saline wastewater indicate that the salt concentration is not the main indicator in determining the level of inhibition. Apart from pre adapted sludge and population of the bacteria, the presence of other cations and OLR also need to be considered. These statements are supported by Feijoo et al. (1995) who noted that performance in treating

saline wastewater in AD system depends on nutrient in the feedstock, previous adaptation of the sludge, antagonistic or synergistic effect (due to the presence of other cations) and lower substrates to biomass ratio used.

Numerous studies on saline wastewater have highlighted the role of Na in contributing to inhibition in AD systems. The effect of other salts, particularly K, has been given relatively little attention. Fernandez and Foster (1994) investigated the threshold of K inhibition using glucose feed substrate (batch study) and observed that low concentrations of K (less than 400 mg l^{-1}) facilitated performance in both the thermophilic and mesophilic conditions, while at higher concentrations (greater than 2500 mg l^{-1}) toxicity of K was significant. Mouneimne et al. (2003) investigated the toxicity of K in an acetate batch assay test and found the toxicity threshold was 0.43 mol l-1. Most of the K toxicity thresholds which have been reported were conducted in batch studies, which implied that further investigation is required in continuous system.

None of the previous research has focused on the potential for anaerobic digestion of wastewater at various types of salts and level of salinity. Recent studies have shown that Na can cause inhibition in AD systems. Only a few studies have been conducted to identify the threshold of K inhibition, and most of these were carried out in batch systems rather than in continuous digesters which are more relevant in practice. Little previous research works has discussed on antagonistic effect (the role of certain ion to reduce the inhibition of other ion) which also required further investigation. Therefore, this study was conducted to investigate the adaptation level and capability of CSTR to withstand at different salinity level in domestic wastewater.

METHODOLOGY

Experimental set up

The continuously stirred tank reactors (CSTR) each had a 5-L capacity and a 4-L working volume as shown in Figure 1. They were constructed of PVC tube with gas-tight top and bottom plates. The top plate was fitted with a gas outlet, a feed port sealed with a rubber bung, and a draught tube liquid seal through which an asymmetric bar stirrer was inserted with a 40 rpm motor mounted directly on the top plate. Temperature was controlled at 35^{o}C by circulating water from a thermostatically-controlled bath through a heating coil around the digesters. Biogas was measured using tipping bucket gas counters with continuous datalogging (Walker et al., 2009) and all gas volumes reported are corrected to standard temperature and pressure (STP) of 0 ^{o}C and 101.325 kPa. Semi-continuous operation was achieved by removing digestate through an outlet port in the base before adding feed via the feed port. During this process, a small amount atmospheric air enters the headspace but in insufficient quantities to affect the redox conditions in the digester: any nitrogen detected in the gas composition is corrected for, as this is not normally produced as a result of the digestion process. Feed was added on a daily basis with regular checks to ensure a constant level was maintained in the digesters. The CSTR digesters were initially seeded with digestate from a digester treating municipal wastewater biosolids at Millbrook Wastewater Treatment plant, Southampton, UK and then acclimated to DW until they achieved stable specific methane production of 0.32 l CH_4 g^{-1} COD $_{added.}$

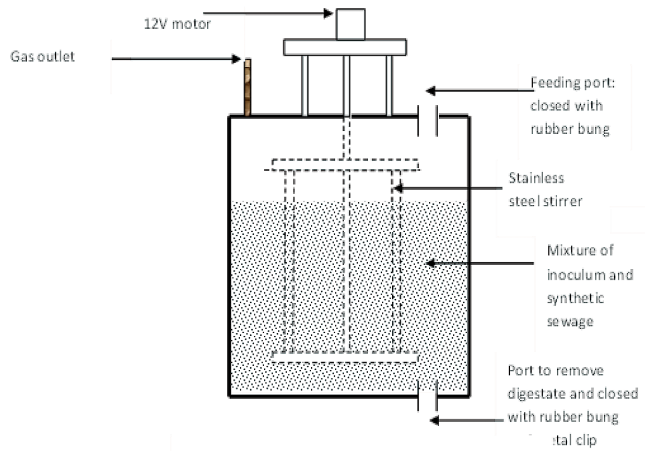


Figure 1. Continuously stirred tank reactors

Analytical Method

Biogas composition (CH_4 and CO_2) was determined using a Varian star 3400 CX gas chromatograph (GC), with a standard of 65.12% (v/v) CH_4 and 34.88% (v/v) CO_2 for calibration. Conductivity was measured using a conductivity meter (LF330, WTW GmbH, Germany); pH was measured using a Jenway 3010 pH meter (Bibby Scientific Ltd, UK) with a combination glass electrode calibrated in buffers at pH 4, 7 and 9 (Fisher Scientific, UK). Suspended solids were determined according to Standard Method 2540 D (APHA, 2005), using glass fibre filter paper with pore size 0.4 μm (Whatman, UK). The COD was measured using a closed tube digestion and titration (APHA, 2005). Total Organic Carbon (TOC) was measured using a Dohrmann TOC (DC-190) based on Standard Method 5310 (APHA, 2005). Ammonia was determined using a Kjeltech block digestion and steam distillation unit according to the manufacturer's instructions (Foss Ltd, Warrington, UK). Potassium (K), Magnesium (Mg), Sodium (Na) and Calcium (Ca) in leachate samples were analysed by first filtering the sample (Whatman No. 1) and then diluting it into 12.5% of nitric acid (HNO_3). The acidified samples were analysed using a Varian Spectra AA-200 atomic absorption spectrometer (Varian Ltd, UK), according to the manufacturer's instructions. The VFA concentrations were quantified in a Shimazdu 2010 gas chromatograph using a flame ionisation detector and a capillary column type SGE BP 21 with helium as the carrier gas.

RESULTS AND DISCUSSION

Two digesters were used, C1 was the control digester and fed only on domestic wastewater (DW) while C2 was fed on DW supplemented with NaCl, C3 on DW with Potassium Chloride (KCl) and Sodium Chloride (NaCl) in a 1:1 ratio by weight and C4 on DW and KCl. The DW was used at a concentration of 48 g COD l^{-1} and the applied organic loading rate was increased during the experiment as shown in Figure 2. The control and digester C3 fed on KCl and NaCl received the same Organic Loading Rate (OLR). Digester C2 was inhibited by

NaCl and required a temporary reduction in loading. Digester C4 with KCl addition also had a slightly reduced load compared to the control. From day 11 salts was added to the DW to give a concentration of 10 g TS l^{-1} of either the individual salt or the mixed NaCl and KCl. No salt additions were made to the inoculum and the increase in salt concentration in the digester therefore occurred by dilution into the reactor. From day 65, additional trace element supplementation was carried out by adding 0.132 ml of oxyanion and cation solutions twice a week for 3 weeks.

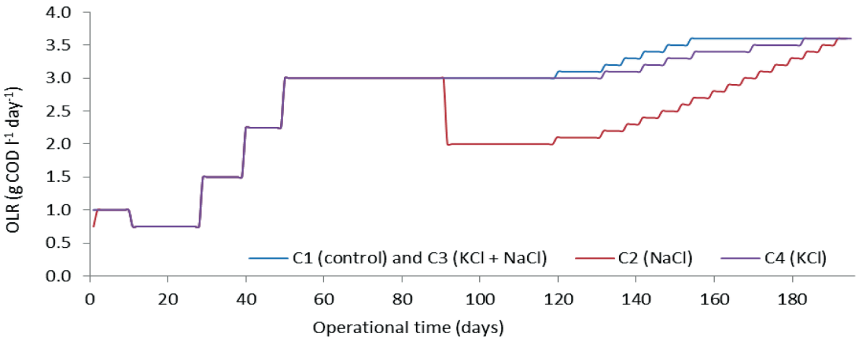


Figure 2. OLR applied to digesters C1-4 during CSTR trial.

On starting to feed the digesters with DW, there were some initial fluctuations in specific biogas production (SBP) and SMP, probably due to the change in feedstock and the gradual depletion of old substrate. By day 25 the SMP had reached a typical value for the DW of around 0.32 l CH_4 $g^{-1.}$ Subsequently, on each occasion when the OLR was increased, there was a fall in SBP and SMP followed by recovery over the next 5-10 days (Figure 3). Until day 45, the behaviour of all four digesters was similar in terms of gas production: by this point, the salt concentration had increased to around 6.5 g l^{-1}, a concentration already higher than that in the UASB trials. From this time, however, the digesters receiving the salt-supplemented feedstock began to show some signs of stress, with a reduction in SBP and SMP relative to the control.

Signs of stress were also seen in the pH, which fell first and most sharply in C2 (receiving NaCl), then less sharply in C4 (KCl) and C3 (KCl + NaCl). The drop in pH was associated with a rise in the alkalinity ratio and in the total VFA concentration (Figure 4). This increase in VFA also occurred in the control digester, and has previously been observed in the laboratory at Southampton following a change of substrate (Zhang et al., 2012, and unpublished data). The change in VFA concentration could therefore not be attributed to salt addition alone: however, the increase in concentration started earlier, lasted longer and gave a higher peak in the following order: C2 (NaCl) >C4 (KCl) > C3 (KCl + NaCl) > C1 (control).

The VFA profiles for each digester are shown in Figure 5. In the control digester C1 the VFA peak was mainly composed of acetic acid, which peaked around day 72. A smaller peak in propionic acid was seen between days ~60 to 90. In C2 (NaCl) propionic acid rose much

more rapidly than acetic, reaching a peak around day 72 that declined only slowly in the next 20 days, before falling rapidly. In C3 (KCl + NaCl) acetic and propionic acid increased at similar rates and the propionic peak was both higher and more prolonged than the acetic, but fell rapidly from day 78 onwards. In C4 (KCl) the acetic acid concentration increased first, with a very sharp rise in propionic after day 61 by which time the salt concentration had reached around 9.3 g l^{-1}. The appearance of propionic acid and of other longer-chain VFA (Figure 5 and 6) thus also suggested that the digesters adapted more easily to KCl + NaCl or KCl alone than to NaCl. Addition of Trace Element twice a week for 3 weeks starting from day 65 may also have helped to reduce the VFA concentration and thus improve biogas production particularly in the digester fed on DW+NaCl alone. In terms of inhibition, digester fed on DW + NaCl was badly affected in term of biogas and methane production as compared to digester fed on DW + KCl. This finding showed that Na can affect cell of microbes at lower concentration as compared to K.

Ammonia concentrations in the digesters showed good agreement throughout the experimental period (Figure 4), and in general, remained between 2.5 - 3.0 g N l^{-1}, a level likely to provide some buffering without a strong risk of inhibition.

Because of the particularly high alkalinity ratio and VFA concentration in C2, the OLR on this digester was reduced from day 92 to promote recovery. This lower loading was maintained until day 120, by which time the alkalinity ratio in all digesters was less than 0.3 and the total VFA concentration was below 1000 mg l^{-1}. SBP and SMP values were converging and remained stable from day 140 onwards, during which time the OLR on all digesters was increased to the final value of 3.6 g COD l^{-1} day^{-1}. During the last 30 days of operation (Table 1), all digesters achieved stable operation with methane production of 0.30 l CH_4 g^{-1} COD_{added} and VFA of 100 mg l^{-1}.

Average performance and operating parameters for last 30 days of run are shown in Table 1. From these results, it appears that the inoculum adapted more readily to the mixture of KCl + NaCl than to KCl alone, and to KCl alone better than to NaCl. This circumstance can be explained by the presence of excessive salts alters many cellular properties which include cell volume (dehydration of bacterial cell), turgor pressure and the ionic strength in the cytoplasmic membrane. The alteration of intracellular metabolites concentration can lead to the inhibition of a variety of cellular processes. The phenomenon is known as cell lysis (dehydration under high-osmolarity growth conditions) (Kraegeloh et al., 2005). Despite this difference, in all cases, the digesters were able to adapt successfully to the added salt concentrations.

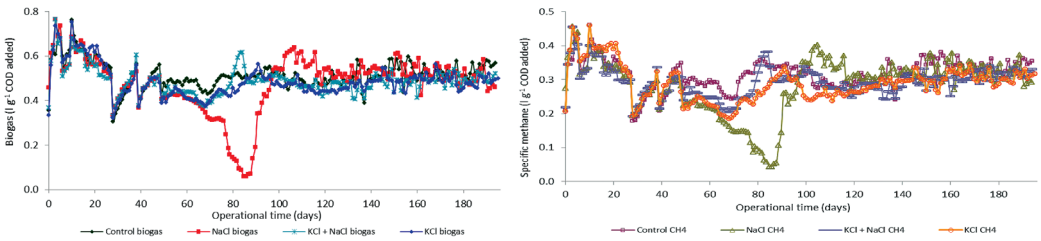


Figure 3. Specific biogas and methane productions for digesters C1-4 during CSTR testing of salt addition

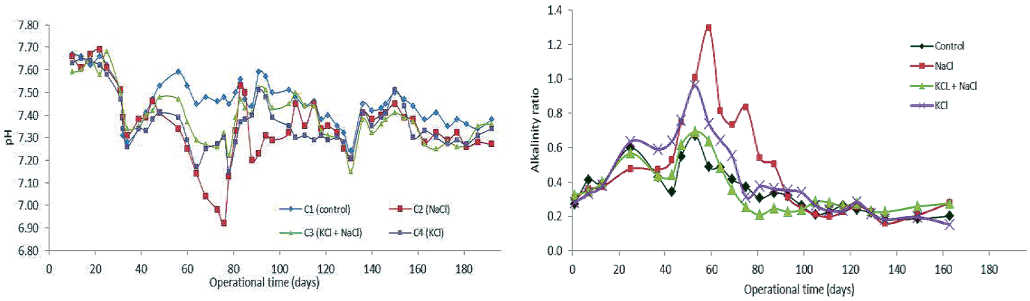


Figure 4. pH and alkalinity in digesters C1-4 during CSTR testing of salt addition

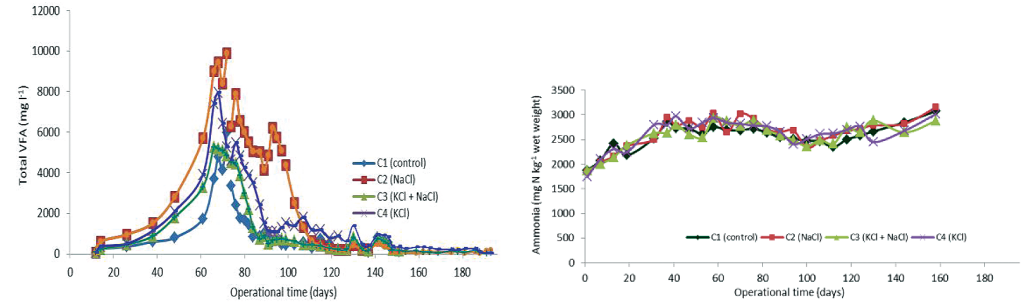


Figure 5. Total VFA and ammonia concentrations in digesters C1-4 during CSTR testing of salt addition

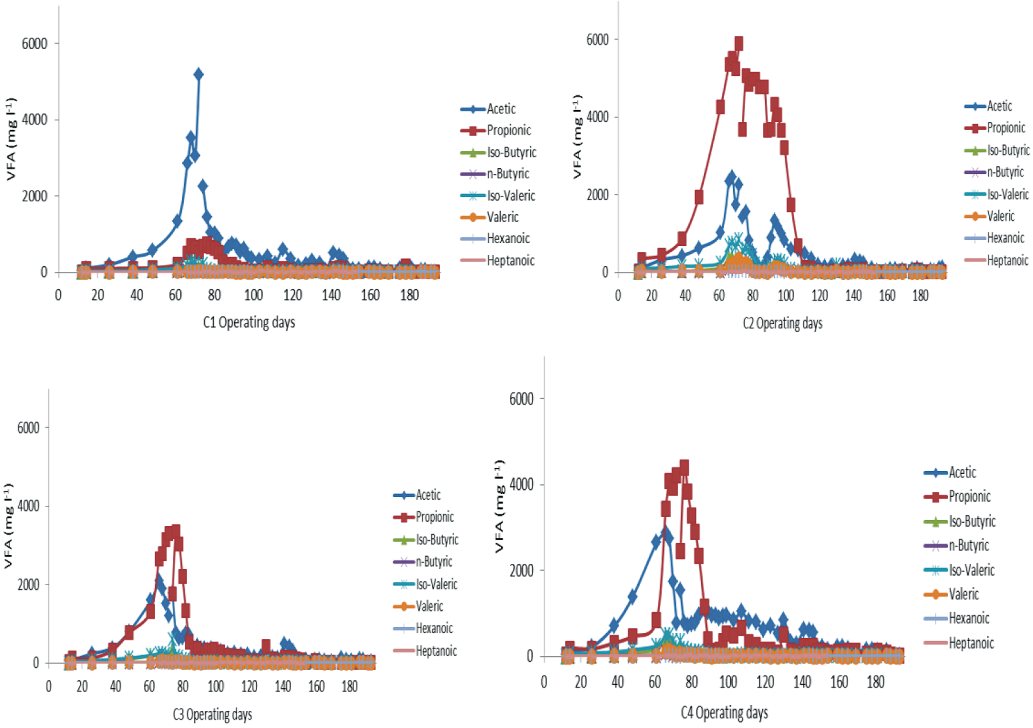


Figure 6. VFA profiles in digesters C1-4 during CSTR testing of salt addition

Table 1
*Average performance over the last 30 days of the experimental run**

	Unit	C1 Control	C2 NaCl	C3 KCl + NaCl	C4 KCl
SBP	l biogas g^{-1} COD $_{added}$	0.531	0.509	0.498	0.490
SMP	l CH_4 g^{-1} COD $_{added}$	0.335	0.323	0.316	0.308
pH	-	7.37	7.29	7.29	7.30
VFA	mg l^{-1}	117	107	91	169

* NB Only C1 control and C4 KCl are at full steady state conditions at this point

CONCLUSION

A CSTR digestion study revealed that that digesters seeded with an inoculum from a conventional mesophilic digester treating municipal wastewater biosolids and fed on DW plus salts were able to acclimate successfully to a final salt concentration of 10 g l^{-1}. This indicated that this source of inoculum would be suitable for start-up of a large-scale CSTR or anaerobic filter plant. The digesters showed some disturbances during the acclimatisation period as indicated by reductions in SBP, SMP and pH and increases in IA/PA ratio and VFA concentration. The order of severity of disturbance was NaCl > KCl > KCl + NaCl. Average values for SBP and SMP after stabilisation were below in the controlled digester, at 0.335 (controlled), 0.323 (NaCl), 0.316 (KCl + NaCl) and 0.308 l CH_4 g^{-1} COD added (KCl). In addition, the role of KCL to reduce the toxicity of NaCl was clearly identified where C3 produced less disturbance compared with C2 at similar amount of final salt concentration.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions of Prof. Dr. Charles Banks, Dr. Sonia Heaven and Bioenergy and Organic Resources Group, Faculty of Engineering and the Environment, University of Southampton, UK for their contributions.

REFERENCES

Abbott, T., & Eskicioglu, C. (2015). Effects of metal salt addition on odor and process stability during the anaerobic digestion of municipal waste sludge. *Waste Management*, *46*, 449-458.

APHA. (2005). *Standard Methods for the Examination of Water and Wastewater* (21st Ed.). American Public Health Association/American Water Works Association/Water Environment Federation, Washington, DC, USA.

Astals, S., Nolla-Ardèvol, V., & Mata-Alvarez, J. (2013). Thermophilic co-digestion of pig manure and crude glycerol: process performance and digestate stability. *Journal of biotechnology*, *166*(3), 97-104.

Azman, S., Khadem, A. F., Zeeman, G., van Lier, J. B., & Plugge, C. M., (2015). Mitigation of Humic Acid Inhibition in Anaerobic Digestion of Cellulose by Addition of Various Salts. *Bioengineering*, *2*(2), 54-65.

Boardman, G. D., Tisinger, J. L., & Gallagher, D. L. (1995). Treatment of clam processing wastewaters by means of upflow anaerobic sludge blanket technology. *Water Research, 29*(6), 1483-1490.

Feijoo, G., Soto, M., Méndez, R., & Lema, J. M., (1995). Sodium inhibition in the anaerobic digestion process: antagonism and adaptation phenomena. *Enzyme and Microbial Technology, 17*(2), 180-188.

Fernandez, N., & Forster, C. (1994). The anaerobic digestion of a simulated coffee waste using thermophilic and mesophilic upflow filters. *Process safety and environmental protection, 72*(1), 15-20.

Gebauer, R. (2004). Mesophilic anaerobic treatment of sludge from saline fish farm effluents with biogas production. *Bioresource Technology, 93*(2), 155-167.

Guerrero, L., Omil, F., Méndez, R., & Lema, J. M. (1997). Treatment of saline wastewaters from fish meal factories in an anaerobic filter under extreme ammonia concentrations. *Bioresource Technology, 61*(1), 69-78.

Kugelman, I. J., & McCarty, P. L. (1965). Cation toxicity and stimulation in anaerobic waste treatment. *Journal of Water Pollution Control Federation, 37*(1), 97-116.

Kumar, S., Das, A., Srinivas, G. L. K., Dhar, H., Ojha, V. K., & Wong, J. (2016). Effect of calcium chloride on abating inhibition due to volatile fatty acids during the start-up period in anaerobic digestion of municipal solid waste. *Environmental technology*, *37*(12), 1501-1509.

Kraegeloh, A., Amendt, B., & Kunte, H. J. (2005). Potassium transport in a halophilic member of the bacteria domain: identification and characterization of the K+ uptake systems TrkH and TrkI from Halomonas elongata DSM 2581T. *Journal of bacteriology, 187*(3), 1036-1043.

Lai, M. C., & Gunsalus, R. P. (1992). Glycine betaine and potassium ion are the major compatible solutes in the extremely halophilic methanogen Methanohalophilus strain Z7302. *Journal of bacteriology, 174*(22), 7474-7477.

Mouneimne, A. H., Carrere, H., Bernet, N., & Delgenes, J. P. (2003). Effect of saponification on the anaerobic digestion of solid fatty residues. *Bioresource technology, 90*(1), 89-94.

Reid, E., Liu, X., & Judd, S. J. (2006). Effect of high salinity on activated sludge characteristics and membrane permeability in an immersed membrane bioreactor. *Journal of Membrane Science, 283*(1), 164-171.

Rovirosa, N., Sánchez, E., Cruz, M., Veiga, M. C., & Borja, R. (2004). Coliform concentration reduction and related performance evaluation of a down-flow anaerobic fixed bed reactor treating low-strength saline wastewater. *Bioresource technology, 94*(2), 119-127.

Walker, M., Zhang, Y., Heaven, S., & Banks, C. J. (2009), Potential errors in the quantitative evaluation of biogas production in anaerobic digestion processes. *Bioresource Technology, 100*(24), 6339–6346

Wood, J. M. (2006). Osmosensing by bacteria. *Science's STKE, 2006*(357), pe43.

Zhang, Y., Banks, C. J., & Heaven, S. (2012) Co-digestion of source segregated domestic food waste to improve process stability. *Bioresource Technology, 114*(0), 168-178.

Pertanika J. Sci. & Technol. 25 (1): 221 - 246 (2017)

SCIENCE & TECHNOLOGY

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

Experimental and Numerical Study of the Heat Transfer Characteristics of Aluminium Metal Foam (with/without channels) Subjected to Steady Water Flow

Bayomy, A. M.* and Saghir, M. Z.

Department of Mechanical and Industrial Engineering, Ryerson University, 350 Victoria St, Toronto, ON, Canada

ABSTRACT

Rapid developments in the design of chips and electronic devices for high-performance computers have necessitated new and more effective methods of chip cooling. The purpose of this study was to investigate the heat transfer characteristics of aluminium foam heat sink for the Intel core i7 processor. Three aluminium foam heat sink models were used in the study: without channels (A), with two channels (B), and with three channels (C). The aluminium foam heat sink was subjected to a steady flow of water covering the non-Darcy flow regime (541 to 1353 Reynolds numbers). The bottom side of the heat sink was heated with a heat flux between 8.5 and 13.8 W/cm^2. The distributions of the local surface temperature and the local Nusselt number were measured and compared with numerical data obtained using the finite element method for all three models. The average Nusselt number was obtained for the specified range of Reynolds numbers, and an empirical correlation of the average Nusselt number as a function of the Reynolds number was derived for each model. The results revealed that the local surface temperature increases as the heat flux increases, decreasing the Reynolds number and increasing the flow direction axis for all three models. Model (A) achieved a lower local temperature than models (B) and (C). Compared with model (A), models (B) and (C) reduced the average Nusselt number by 10% and 25%, respectively. The pressure drop across the foam was also measured. The thermal efficiency index was defined in this study in order to obtain the optimal design condition for the aluminium foam heat sink models. The results revealed that the optimum design condition is model (B) at Re=1353. The numerical results were in good agreement with the local Nusselt number and the local experimental temperature with a maximum relative error of 3% and 2% respectively.

Keywords: Electronic cooling, forced convection, porous media, heat transfer, heat sink, metal foam

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
ayman.bayomy@ryerson.ca (Bayomy, A. M.),
zsaghir@ryerson.ca (Saghir, M. Z.),
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Electronic components depend on electrical currents passing through resistance. This is accompanied by heat flux dissipation. The rapid developments in the design of electronic chips have resulted in an increase in the amount of heat generated per unit of volume. The international technology road map for semiconductors reported that the heat flux of chips rose by 200 W/cm^2 (Lee, 2009). Conventional convective cooling methods are only capable of removing small amounts of heat flux, making it important to search for new methods of cooling high-speed electronic components. The performance reliability and lifetime of electronic components are inversely related to the temperature of the component. In silicon semi-conductor devices, reductions in temperature increase their performance reliability and lifetime. Gochman et al. (2003) reported that the heat dissipation of desktop and mobile processors are 100W and 30W respectively.

In order to enhance the heat transfer rate of modern high speed electronic devices, researchers have conducted extensive investigations using different shapes and arrangements (single or multiple square, rectangular and circular modules (rods)) mounted on the heated surface in order to increase the surface to volume ratio of heat sinks (Buller & Kilburn, 1981; Sparrow, Niethammer, & Chaboki, 1982; Sparrow, Yanezmoreno, & Otis, 1984; Jubran, Swiety, & Hamdan, 1996; Iwasaki & Ishizuka, 2000; Igarashi, Nakamura, & Fukuoka, 2004)

Many studies have been conducted using various types of heat sinks for electronic cooling, either extending the surface area or increasing the fluid flow. Despite this research, there still exists a demand for more effective electronic cooling methods.

The use of porous media as a heat sink subjected to forced cooling fluid is a new technique used to enhance heat transfer from the surface of electronics. Metal foam is a porous material with a low density and novel thermal, electrical, mechanical and acoustic properties. This material provides good heat transfer due to its large surface area to volume ratio (roughly 1000-3000 m^2/m^3) (Zhao, 2012). The heat transfer characteristics of metal foam are directly affected by microstructural properties such as porosity, relative density, pore density, pore size, ligament diameter and permeability (Hwang, Hwang, Yeh, and Chao, 2002; Zhao, Kim, and Hodson, 2004; Lu, Zhao, and Tassou, 2006 ; Seyf and Layeghi, 2010; Mancin, Zilio, Diani, and Rossetto, 2012).

The effective thermal conductivity of ERG aluminium foams was studied by Boomsma and Poulikakos (2001). They demonstrated that in high porosity foam, the solid phase thermal conductivity controls the overall effective thermal conductivity of the aluminium foam. Boomsma and Poulikakos (2001) also derived an expression of effective thermal conductivity. In addition, Zhao, Lu, & Tassou (2006) and Klett et al. (2001) observed that the metal foam heat exchanger achieves higher heat transfer than the conventional finned tube heat exchanger. Studies have also been conducted on the effects of using metal foam (Sung, Kim, and Hyun, 1995; Fu, Leong, Huang, and Liu, 2001; Rachedi and Chikh, 2001; Mahajan & Bhattachatya, 2002; Nield, Kuznetsov, and Xiong, 2003; Kim, Paek, and Kang, 2003; Tzeng, Jeng, and Wang, 2006; Hooman and Ejlali, 2007a; Hooman and Haji-Sheikh, 2007b; Bai and Chung, 2011; Ding, Lu, Chen, He, & Ou, 2011) which demonstrated the positive effect caused by employing the metal foam.

A small number of studies have used water as a working coolant fluid through metal foams. Boomsma, Poulikakos, and Zwick (2003) conducted an experimental comparison of the thermal resistance of a compressed metal foam heat exchanger and commercially available heat exchangers using water as a coolant. They found that the thermal resistance of the compressed foam heat exchanger is two to three times lower than that of other heat exchangers. Noh, Lee, and Lee (2006) conducted an experimental study of a non-Darcy water flow in an annulus filled with high porosity aluminium foam and presented the correlations of the average Nusselt number and friction factor. Hetsroni, Gurevich, and Rozenblit (2005) conducted an experimental study of the transmission window cooling technique of an accelerator using aluminium foam. Dukhan, Bağci, and özdemir (2015) conducted an experimental study of thermal development in open cell metal foam subjected to a constant heat flux. The flow rates used covered both the Darcy and non-Darcy regimes.

This paper presents an experimental and numerical study of aluminium foam models used as heat sinks in the cooling of an Intel core i7 processor (electronic cooling). Three aluminium foam heat sink models were used in the study: without channels (A), with two channels (B), and with three channels (C). The aluminium foam was subjected to a water flow covering the non-Darcy flow regime. Local temperature distributions were measured for different heat fluxes and Reynolds numbers and the local Nusselt number was calculated based on the local surface temperature of each heat sink model. The average Nusselt number was obtained for the entire range of Reynolds numbers. An empirical correlation of the average Nusselt number was developed based on the Reynolds number of each model. The pressure drop across each heat sink model was also measured and the thermal performance of the aluminium foam heat sink was evaluated based on the average Nusselt number and the pumping power required for each model. The thermal efficiency index was defined in order to obtain the optimum design condition for the use of aluminium foam as a heat sink. Lastly, a numerical approach was used (finite element method) (COMSOL Multiphysics, 2015) and the results were compared with those obtained experimentally.

EXPERIMENTAL APPARATUS AND PROCEDURES

An experimental setup was developed in order to examine the heat transfer characteristics of aluminium foam (with and without channels) as a heat sink in the cooling of an Intel core i7 processor.

Test Section and Experimental Facility

The experimental setup was developed before by Bayomy, Saghir, and Yousife (2016). It consisted of a pump with a control valve, a tank, flow meter and pressure transducer (see Figure 1). The flow meter range was from 0 to 3 gpm with corresponding output signals between 4 and 20 mA.

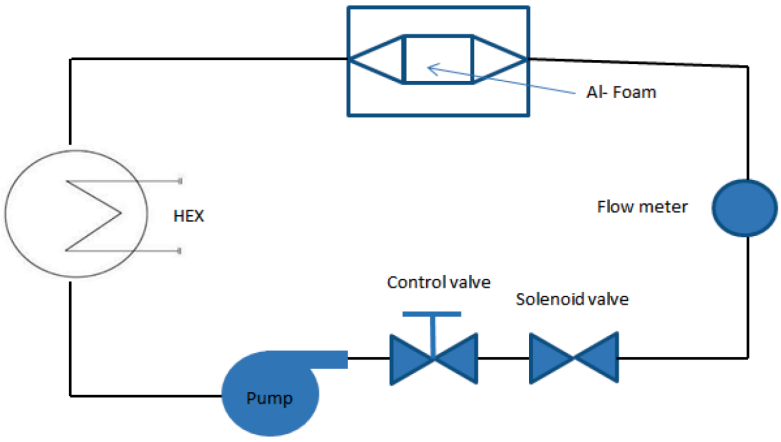


Figure 1. Experimental schematic diagram (Bayomy et al. 2016)

The pressure differential transducer range was from 0 to 30 psi with corresponding output signals between 0 and 10 vdc. The test section consisted of high temperature Teflon insulation attached to a 37.5mm 37.5mm heater (corresponding to the size of an Intel core i7 processor).

The heater contained an adjustable current input in order to control the heat flux, as shown in Figure 2. Eight type (T) thermocouples were attached to the surface of the heater to measure the surface temperature and two thermocouples were used to measure the inlet and outlet water temperatures. The positioning of the thermocouples can be seen in Figure 3(a). Because of their large head connections, the thermocouples were inserted through the Teflon insulation in a staggered arrangement, as shown in Figure 3(b).

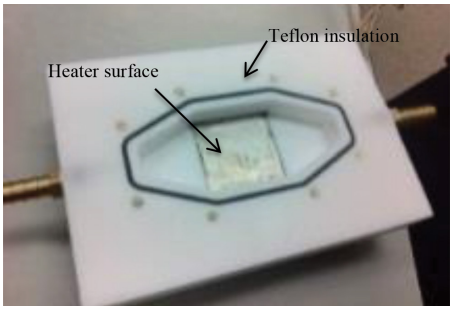


Figure 2. Test section heater (Bayomy et al. 2016)

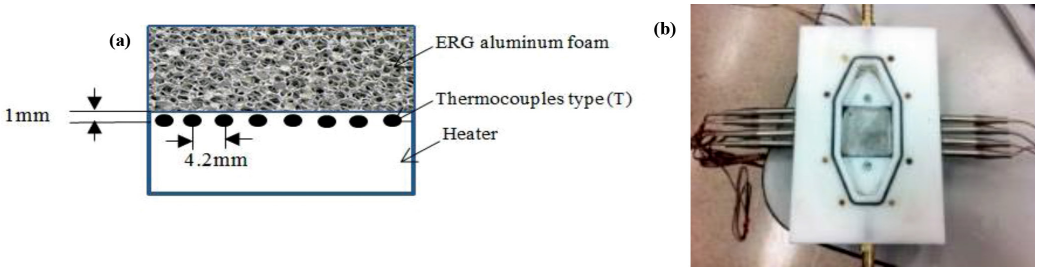


Figure 3. Thermocouples positions and arrangements (Bayomy et al. 2016)

All of the signals from the thermocouples, the flow metre and the pressure transducer were connected to a data acquisition system in order to monitor and save the experimental data. The ERG aluminium metal foam (alloy 6101-T6) was cut to match the size of the heater (37.5mm37.5mm). Three aluminium foam heat sink models were used in the study: one without channels (A), one with two channels (B) and one with three channels (C), as shown in Figure 4. The dimensions of the channels for models (B) and (C) are shown in Figure 5. The physical and geometric properties of the ERG test section foam are as follows: relative density (ρ_r of 9%, pore density of 40PPI, permeability (K) of 3.38e-8 m^2, porosity (ε) of 0.9 and thermal conductivity of solid material (K_s) of 218 W/m.K. It is important to note that these properties are not all independent of one another. The relationships between them are presented by Calmidi and Mahajan (2000).

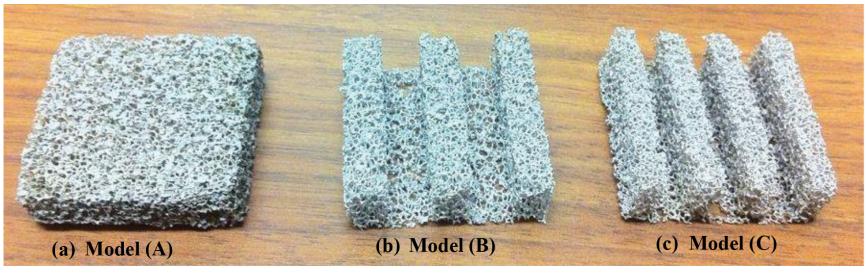


Figure 4. ERG aluminium foam models

The Plexiglas cover has two openings, just before and after the aluminium foam, to allow for connections to the pressure differential transducer ports, as shown in Figure 6. In order to obtain experimental steady state conditions, the surface temperature, inlet and outlet temperatures, pressure drop across the foam, and water flow rate were monitored using the data acquisition system.

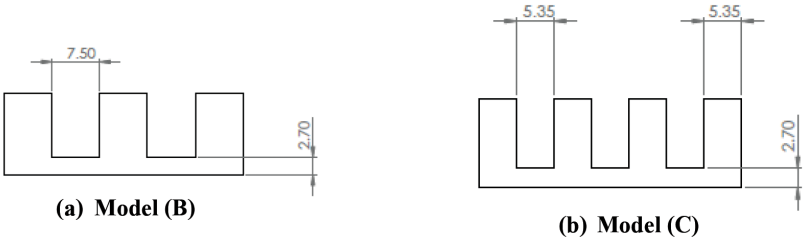


Figure 5. Models B & C dimensions

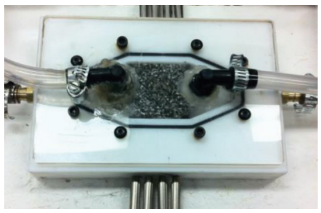


Figure 6. Test section (Bayomy et al. 2016)

Uncertainty Analysis

The uncertainties of the temperature, flow rate and pressure drop were 0.75% (°C), 0.44% (gpm) and 0.03% (psi), respectively. These values were obtained from the calibration process for each instrument based on slandered and random error. The uncertainties of (Nu_x) and (Re) were obtained using the Taylor method (Taylor, 1995). If x, y, z measure quantities such as temperature, flow rate and pressure difference, and those measurement quantities have uncertainties of $\delta x, \delta y, \delta z$ and are used to calculate parameters such as the local Nusselt number Nu_x (x, y, z), then the uncertainty of the local Nusselt number is:

$$\delta Nu_x = \sqrt{\left(\frac{\partial Nu_x}{\partial x}.\delta x\right)^2 + \cdots + \left(\frac{\partial Nu_x}{\partial y}.\delta y\right)^2 + \cdots \left(\frac{\partial Nu_x}{\partial z}.\delta z\right)^2} \quad (1)$$

The local Nusselt number and the Reynolds number are calculated as follows:

$$h_x = \frac{q''}{(T_x - T_{in})} \quad (2)$$

$$Nu_x = \frac{h_x D_e}{k_{eff}} \quad (3)$$

$$Re = \frac{U.D_e}{\upsilon_f} \quad (4)$$

Where h_x represents the local heat transfer coefficient over the heater surface, T_x represents the local surface temperature, T_{in} represents the inlet water temperature, D_e represents the hydraulic diameter of the channel, U represents the water velocity throughout the test section, υ_f represents the kinematic viscosity of the fluid, and represents the effective thermal conductivity of the metal foam filled with water.

The maximum value of the uncertainty of the local Nusselt number was ± 2.4 % and the uncertainty of the Reynolds number was ± 0.44%. The energy balance was checked by comparing the quantity of heat absorbed by the water with the actual heat input of the heater. The heat loss was obtained by subtracting the actual heat input from the quantity of heat absorbed by the water. The result was 0.2% heat loss.

NUMERICAL MODEL DESCRIPTION

Numerical results were obtained using the finite element technique in order to allow for a comparison with the experimental results.

Governing Equations

It is important to state the assumptions upon which the equations are based prior to the formulation of the model. Those assumptions are:

1. The fluid passing through the channel is Newtonian and incompressible.
2. The porous medium is homogenous and has a uniform porosity.

3. Local thermal equilibrium between the fluid and the solid phase of metal foam.

4. No heat generation occurs inside the porous medium.

5. Variation in the thermo-physical properties of the solid phase can be neglected.

The set of governing equation were solved based on the numerical model assumptions. These governing equation consists of the Brinkman- Forchheimer equation and energy equation which describe the fluid flow and heat transfer inside porous media respectively, as follows:

$$\frac{\rho}{\varepsilon}\left((U.\nabla)\frac{U}{\varepsilon}\right) = \nabla.(-pI + \frac{\mu}{\varepsilon}(\nabla U + (\nabla U)^T)) - (\frac{\mu}{K} + \beta_f|U|)U + F \quad (5)$$

$$\nabla.(\rho U) = 0 \quad (6)$$

$$\rho c_p U.\nabla T = \nabla.(k_{eff}.\nabla T) \quad (7)$$

Where ρ represents the water density, c_p represents the water specific heat, ε represents the porosity of the aluminium metal foam, p represents the pressure, U represents the velocity field vector, β_f represents the Forchheimer coefficient, T represents the temperature, μ represents the dynamic viscosity of water, K represents the permeability of the aluminium foam and k_{eff} represents the effective thermal conductivity of the aluminium metal foam when filled with water. The effective thermal conductivity of the porous medium filled with fluid can be described by the porosity (ε) and thermal conductivities of the solid phase (k_s) and fluid phase (k_f) (Kaviany, 1995) using the following equation:

$$k_{eff} = \varepsilon.kf + (1-\varepsilon).ks \quad (8)$$

However, an accurate representation of the structural parameters of the metal foam is important for the estimation of effective thermal conductivity. Calmidi and Mahajan (1999) obtained the effective thermal conductivity of ERG aluminium foam based on one-dimensional heat conduction through two-dimensional foam structures. Boomsma and Pouliakos (2001) extended the previous research by investigating a three-dimensional analytical model of effective thermal conductivity of ERG foams. This model was used to evaluate the effective thermal conductivity of the ERG aluminium foam (alloy 6101-T6) used in the present study.

Boundary Conditions

The boundary conditions can be concluded as an inlet temperature and velocity (assuming a flat profile), at the inlet portion, open boundary at the outlet portion, heat flux from the bottom of the heater and adiabatic walls at the remainder of the surface (see Figure 7).

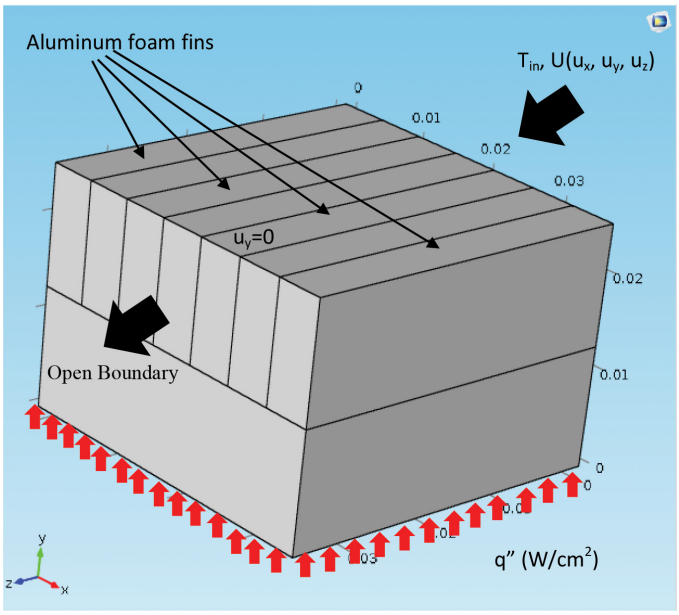


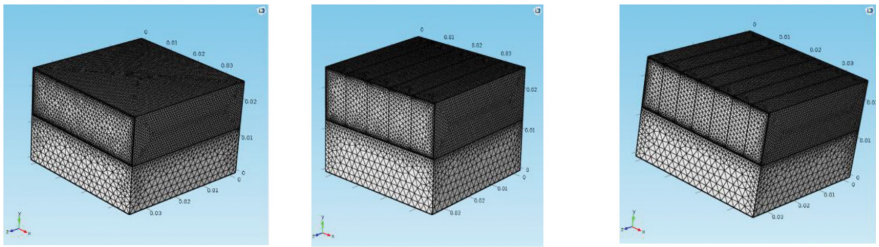
Figure 7. Boundary conditions

Mesh Sensitivity Analysis

A tetrahedral element was used to describe the numerical model. In order to perform mesh sensitivity analysis, calculations of the maximum temperature and local Nusselt number was performed for different numbers of domain elements, as shown in Figure 8. The number of elements used was 1847400 and the variation was less than 0.001, as shown in Figure 9. At every time step, the finite element technique obtained the errors in all independent parameters such as U, P, and T as follows:

$$R = \frac{1}{n.m}\sum_{i=1}^{i=m}\sum_{j=1}^{j=n}\left|\frac{\left(F_{i,j}^{s+1}-F_{i,j}^{s}\right)}{F_{i,j}^{s+1}}\right| \tag{9}$$

Where F represents one of the independent parameters, s represents the number of iterations, and (i, j) represent the x and y coordinates. The solution reached convergence when R was below 1×10^{-6} for each independent parameter.



(a) Model A **(b) Model B** **(c) Model C**

Figure 8. Finite element model

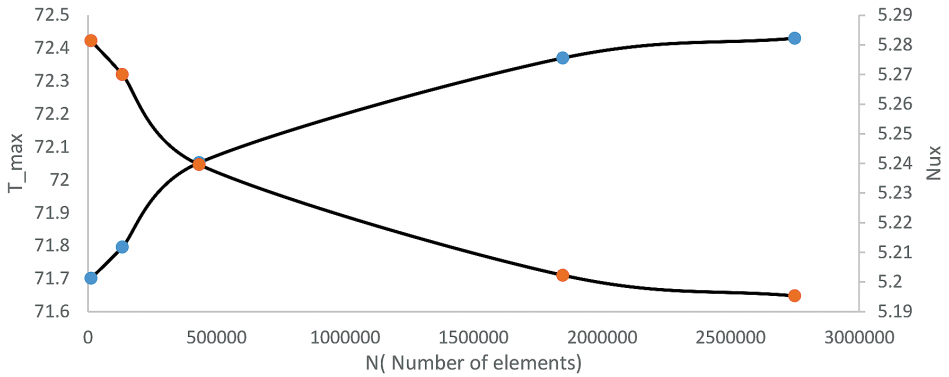


Figure 9. Mesh independent

RESULTS AND DISCUSSION

In this experimental study, three aluminium foam heat sink models (A, B, and C) were subjected to a uniform heat flux ranging from 13.8 to 8.5W/cm^2. The water flow covered the Forchheimer regime (non-Darcy regime). In addition, a numerical model was developed using the finite element technique and the numerical results compared with the experimental results.

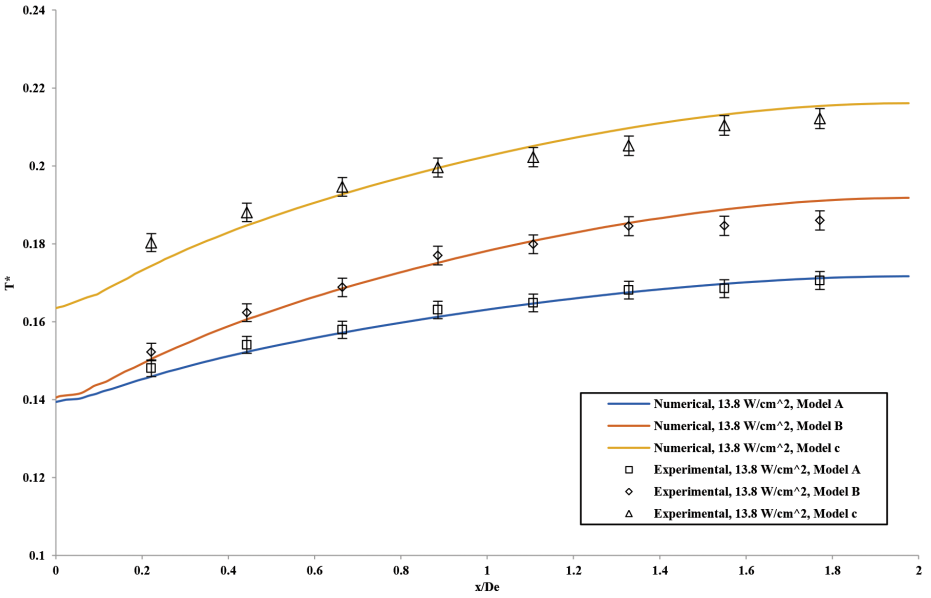
Local Dimensionless Temperature Distributions Over the Surface

Figures 10(a, b) illustrate the surface temperature distributions along the dimensionless flow direction axis (x/De) at q" = 13.8 and 10.6 W/cm^2 respectively, and Re= 1353 for models (A), (B) and (C). As we can see, the surface temperature showed an increase in the flow direction. This trend was observed in previous experimental studies (Fu et al., 2001; Noh et al., 2006). The results revealed that model (A) achieved a lower temperature distribution when compared with models (B) and (C). The results also revealed that the dimensionless local surface temperatures of models (A) and (B) were almost identical at the beginning of the channel (x/De= 0 - 0.3) and began to diverge from each other along with the axial flow direction distant. The same trend can be observed in Figure 11(a, b) which illustrates the temperature distributions at Re=902 and q" = 13.8 and 8.5 W/cm^2 respectively. Figure 12 shows the temperature distributions at Re= 541 and q" =13.8 W/cm^2. We can also see that the surface temperature increases along with decreases in the Reynolds number and increases in the heat flux.

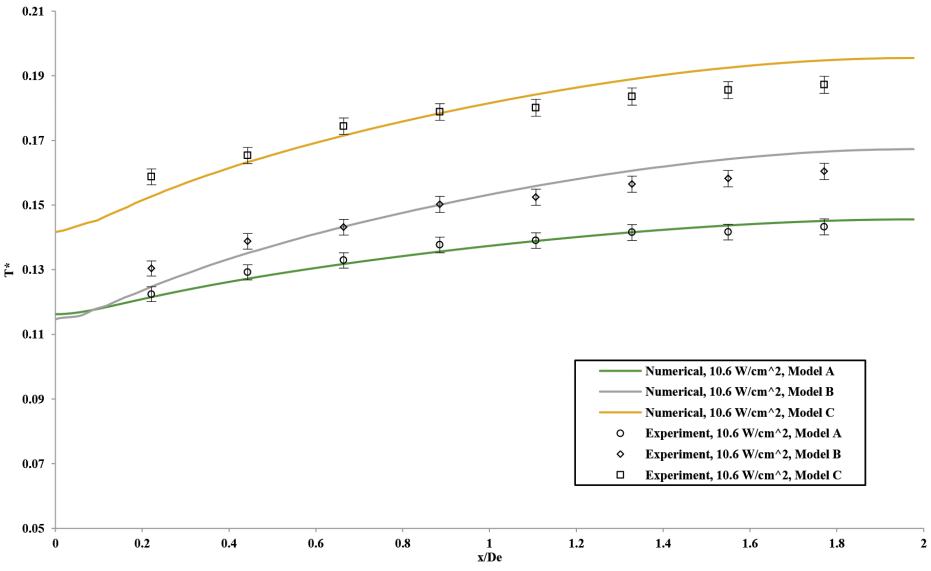
In order to complement the experimental results, a numerical model was developed using the finite element technique. Figures 10, 11 and 12 show the experimental and numerical surface temperature results for different heat flux and Reynolds numbers (Re=1353, 902 and 541). The numerical results are in good agreement with the experimental results, with a maximum relative error of 2%.

In addition, Figures 13 and 14 show the temperature contours of each model at q" =13.8 W/cm^2 and Re= 1353 and 902 respectively. The results revealed that the temperature of the heated surface (electronic surface) increases along with the flow direction axis. In addition, the use of channels in the aluminium foam (models B and C) creates a non-uniform temperature

distribution along the line which is perpendicular to the flow direction. As shown in Figure 15, the temperature distribution along the line perpendicular to the flow direction of model (A) is perfectly uniform and lower than that of models (B) and (C) which suffer from non- uniform temperature.

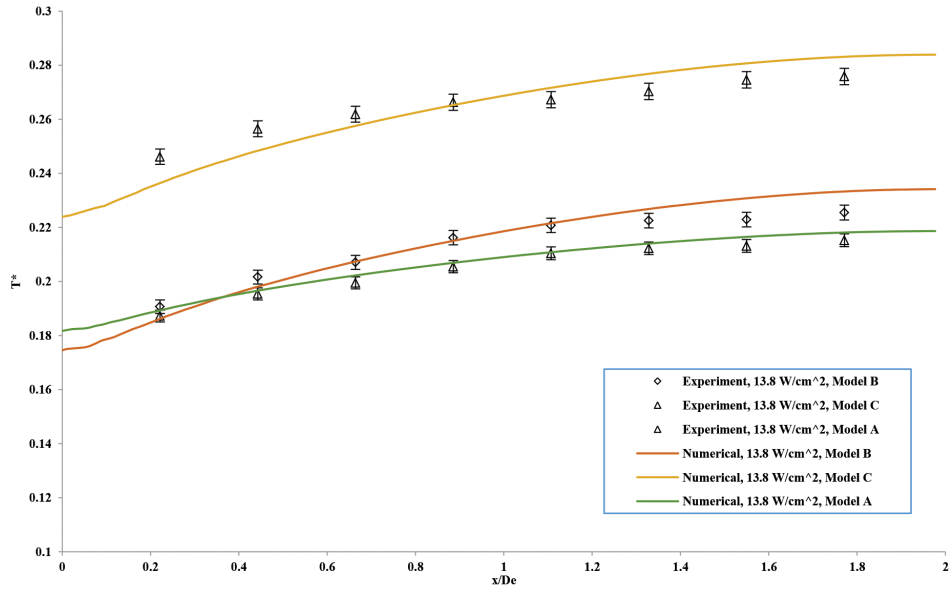


(a) Temperature distributions at q" =13.8W/cm^2

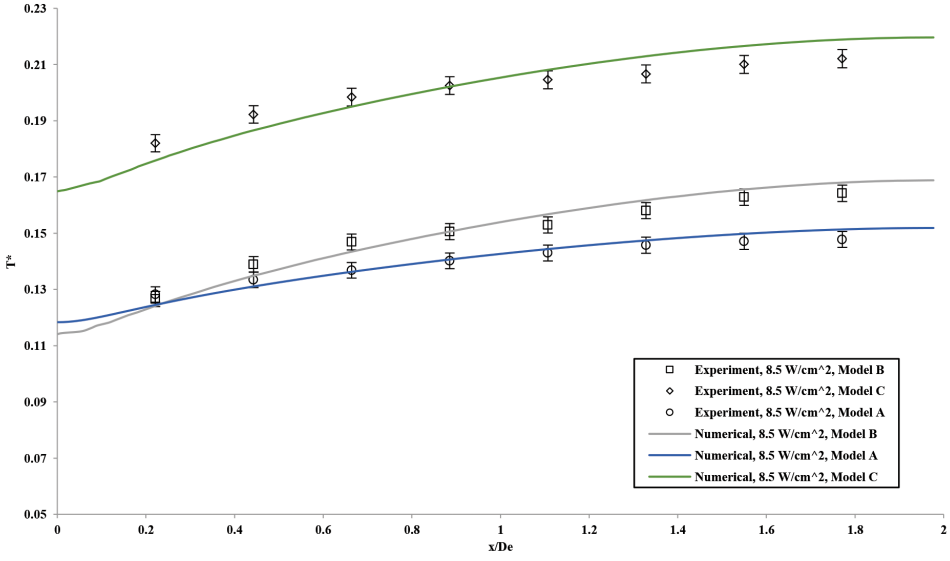


(b) Temperature distributions at q" =10.6W/cm^2

Figure 10. Surface temperature distributions at Re= 1353



(a) Temperature distributions at q" =13.8W/cm^2



(b) Temperature distributions at q" =8.5W/cm^2

Figure 11. Surface temperature distributions at Re= 902

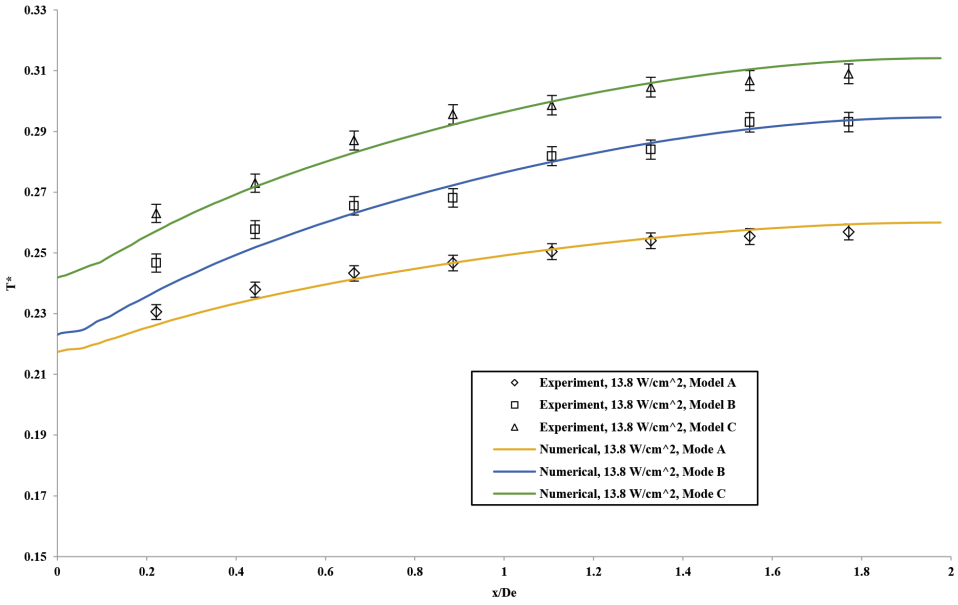
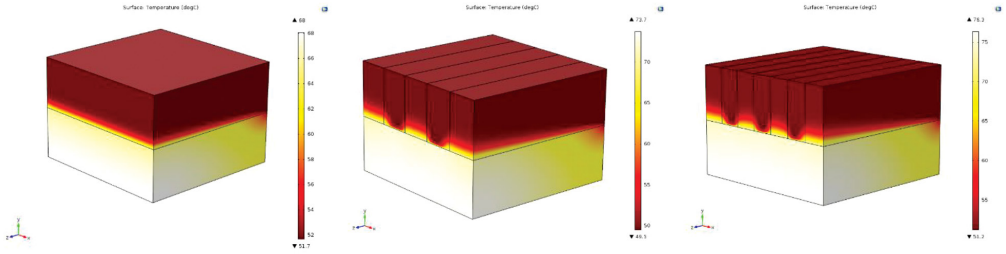
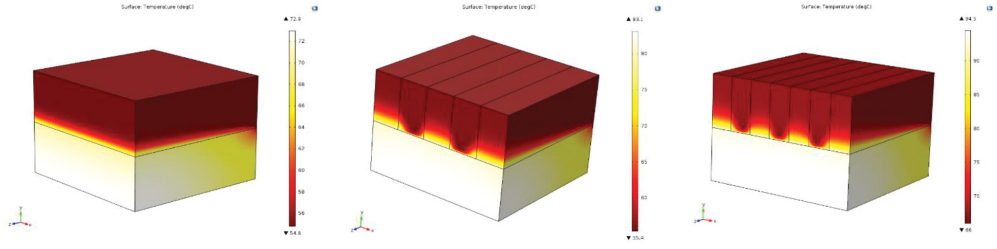


Figure 12. Surface temperature distributions at Re= 541



(a) Model (A) **(b) Model (B)** **(c) Model (C)**

Figure 13. Temperature contours at q" =13.8 W/cm^2 and Re=1353



(a) Model (A) **(b) Model (B)** **(c) Model (C)**

Figure 14. Temperature contours at q" =13.8 W/cm^2 and Re=902

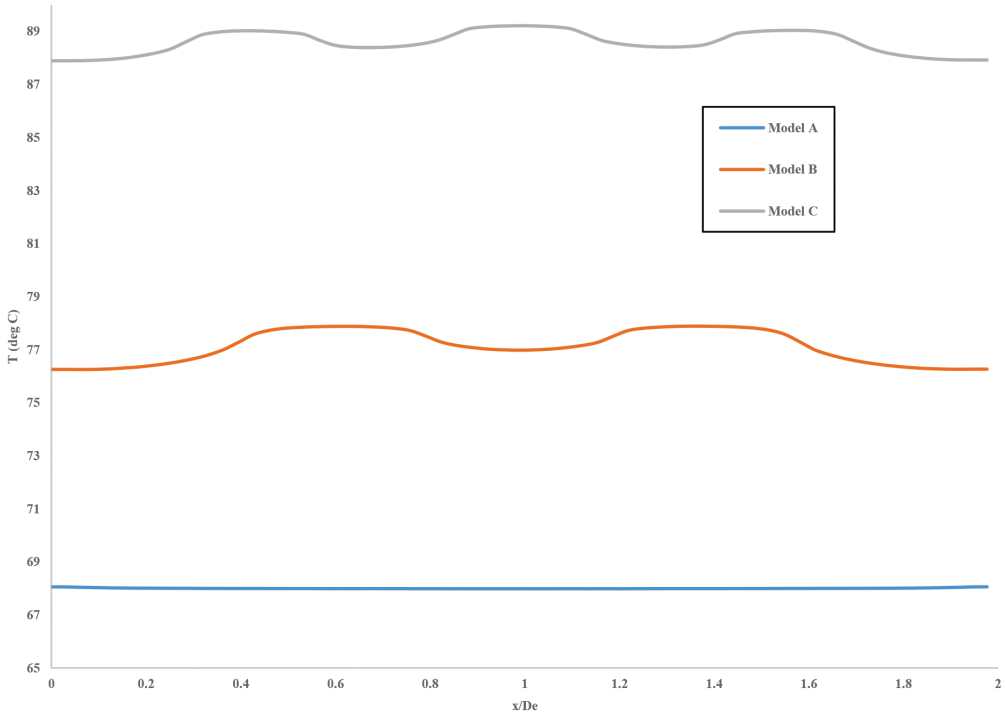
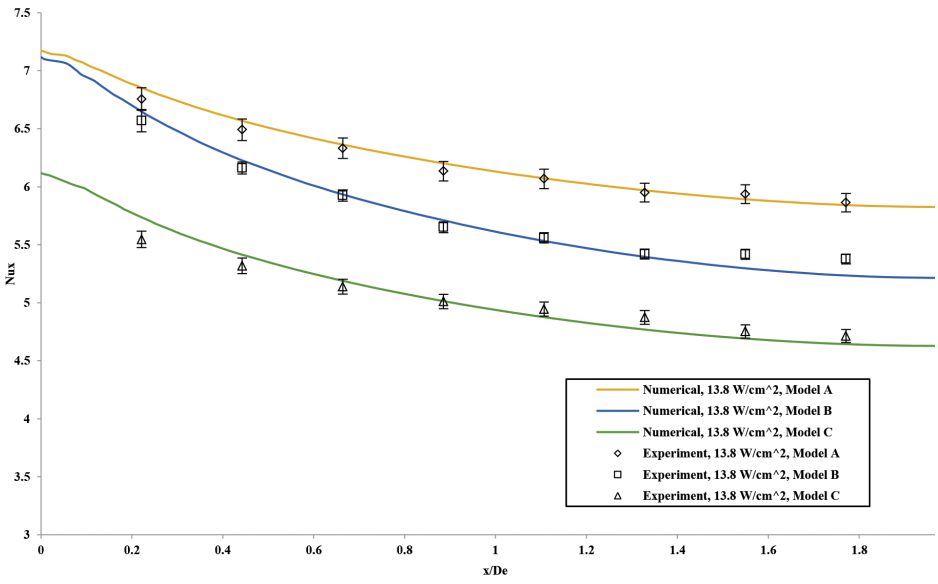


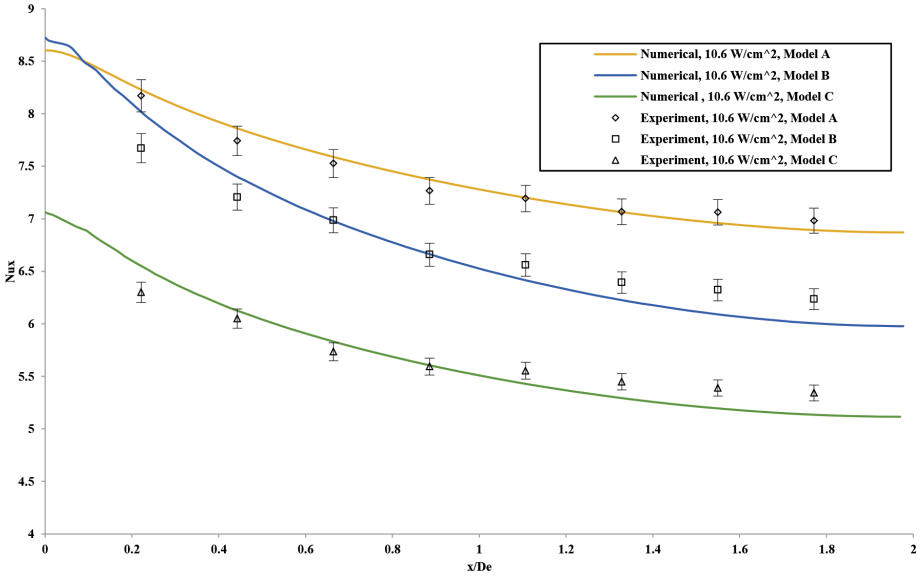
Figure 15. Temperature distribution at q" =13.8 W/cm^2 and Re=902

Local and Average Nusselt Number Distributions

Figures 16 and 17 show the calculated local Nusselt number variation with the dimensionless flow direction axis at Re= 1353 and 902 respectively. As we can see, the local Nusselt number is high in the entry region and begins to decrease until a constant value is reached when the flow and temperatures become thermally fully developed. This means that in the entry regions, the local Nusselt number is inversely proportional to the boundary layer thickness. The results revealed that the Nusselt number was higher for model (A) compared with models (B) and (C). The numerical results were in good agreement with the experimental results with a maximum relative error of 3%.

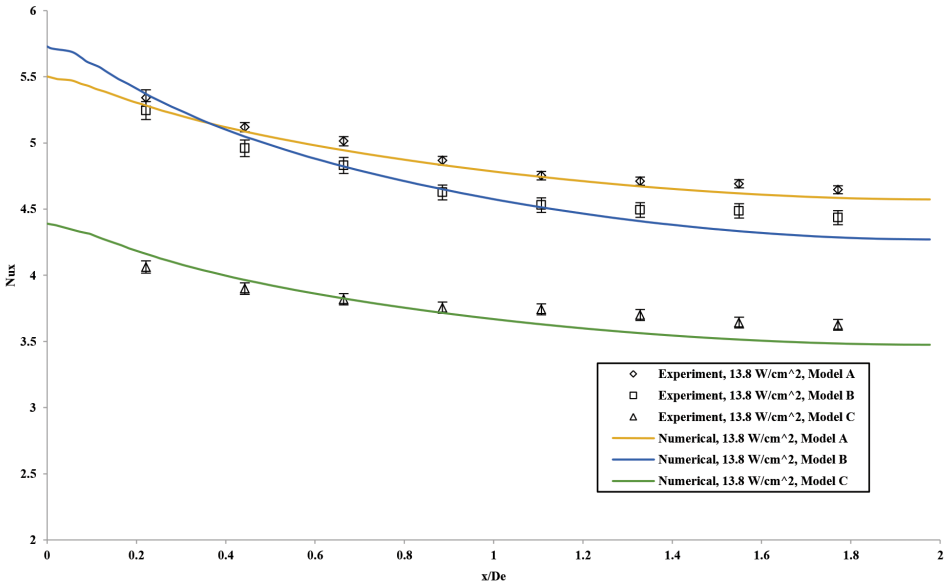


(a) Nusselt number distributions at q" =13.8 W/cm^2

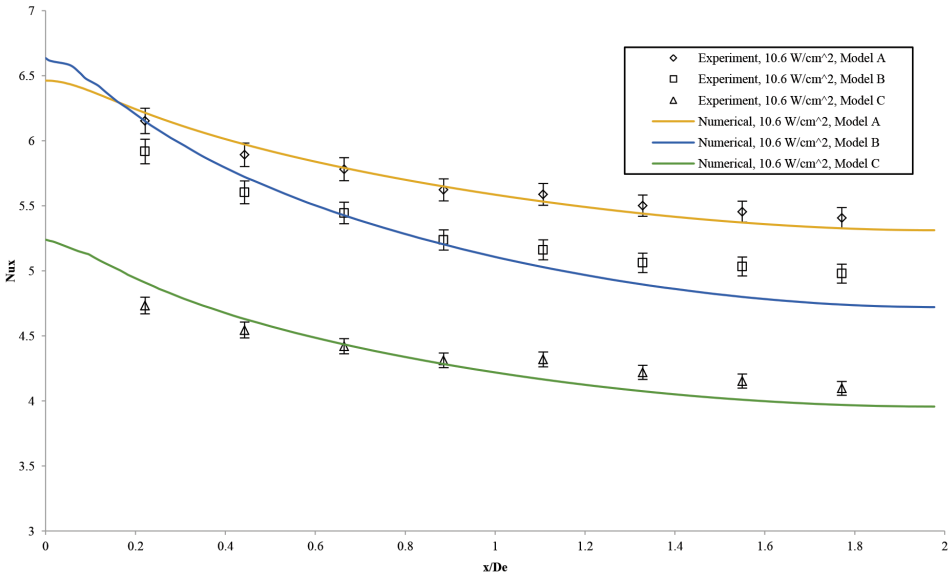


(b) Nusselt number distributions at q" =10.6 W/cm^2

Figure 16. Experimental and numerical results of Nusselt number at Re=1353



(a) Nusselt number distributions at q" =13.8 W/cm^2



(b) Nusselt number distributions at q" =10.6 W/cm^2

Figure 17. Experimental and numerical results of Nusselt number at Re=902

In order to determine the heat transfer performance of aluminium foam models used as a heat sink, the average Nusselt number was obtained using the following equation:

$$\mathrm{Nu}_{,\mathrm{avg}} = \frac{1}{\mathrm{L}} \int_0^{\mathrm{L}} \mathrm{Nu}_{\mathrm{x}}\, \mathrm{dx} \tag{16}$$

Where L represents the channel length and Nu_x represents the local Nusselt number. Figure 18 shows the relationship between the average Nusselt number of models (A), (B) and (C) and the Reynolds number. This figure clearly illustrates that the average Nusselt number increases with increase in the Reynolds number for different models. The results also revealed that the average Nusselt number was higher for model (A) compared with models (B) and (C).

The relationship between the average Nusselt number and the Reynolds number was obtained using the following equation:

$$Nu_{,avg} = CRe^{m} \tag{17}$$

Where C and m are constants listed in Table 1 and obtained using the present experimental data of steady water flow through different aluminium foam heat sink models for a given range of Reynolds numbers (Forchheimer regime).

Table 1
Average Nusselt Number Constants

Heat Sink Models	C	m
Model (A)	**0.342**	**0.41**
Model (B)	**0.121**	**0.54**
Model (c)	**0.1**	**0.56**

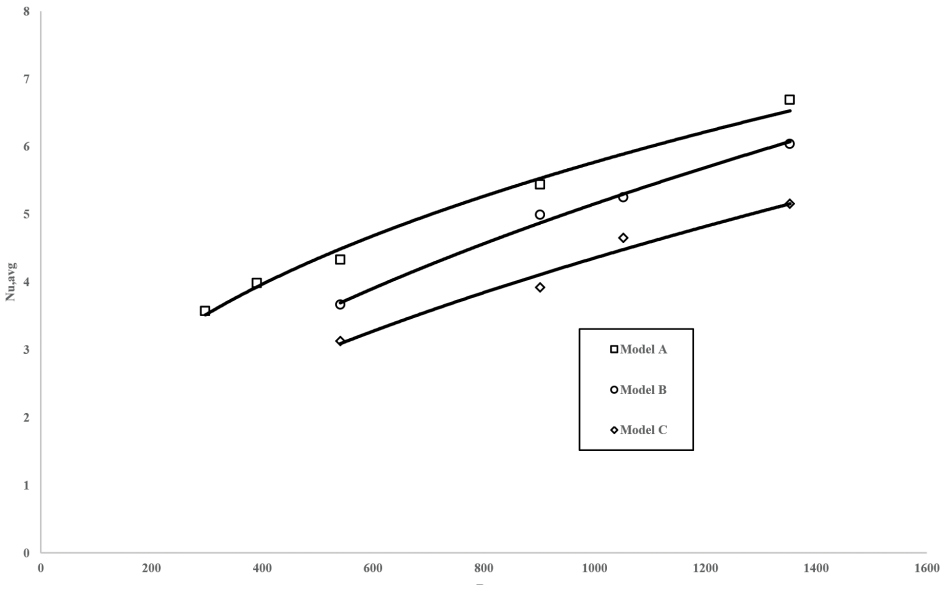


Figure 18. Average Nusselt number verses Reynolds number

As shown in Figure 19, the ratio between the average Nusselt number of model (B) ($Nu_{,avg,B}$) and model (A) ($Nu_{,avg,A}$) is about 0.9 and the ratio between the average Nusselt number of model (C) ($Nu_{,avg,C}$) and model (A) ($Nu_{,avg,A}$) is about 0.75. This means that model (B) and model (C) reduced the heat transfer by 10% and 25% respectively.

Introducing channels in the heat sink (models B and C) is supposed to increase the surface area to volume ratio, causing increases in the heat transfer rate. This is in direct contrast to the results obtained in the experiment which revealed that models (B) and (C) reduced the heat transfer rate by 10% and 25%, respectively. The velocity contours obtained from the finite element will explain this phenomenon.

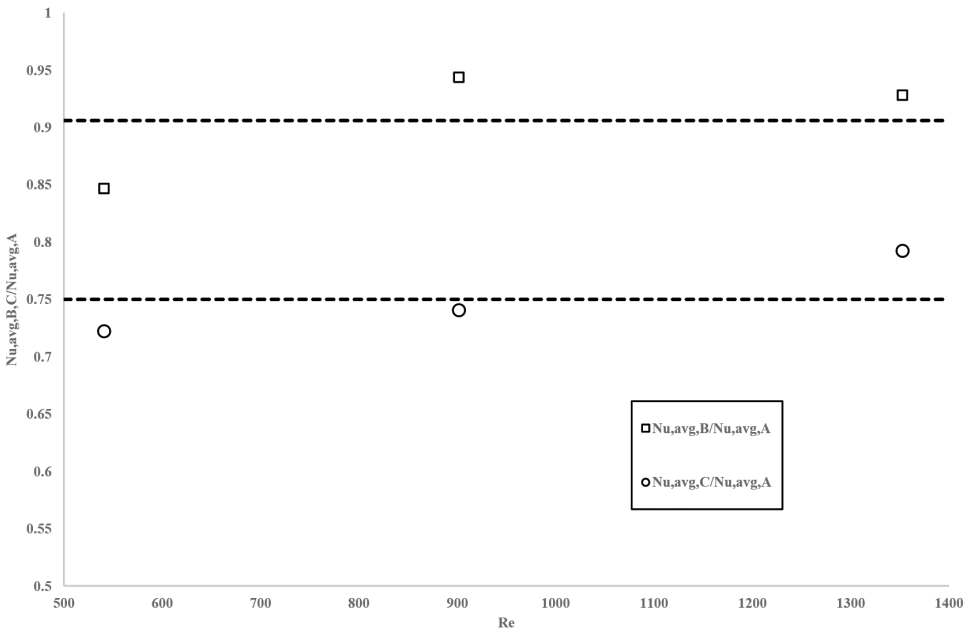
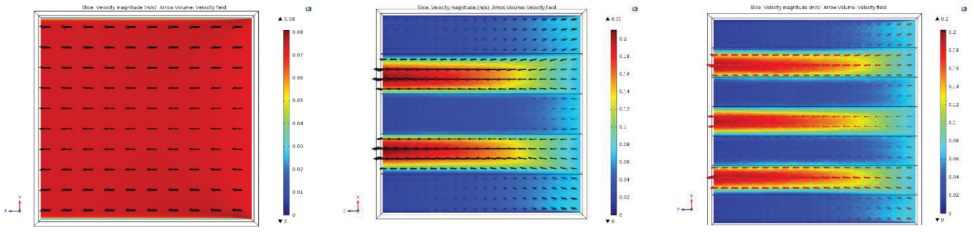


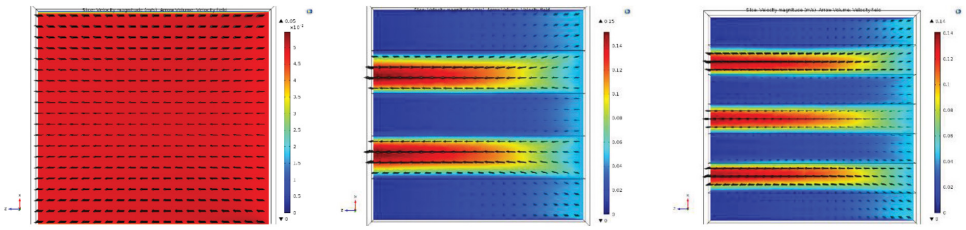
Figure 19. Average Nusselt number ratio of model (B) and (C) to model (A)

Figures 20 and 21 show the velocity contours for models (A), (B) and (C) at q" =13.8 W/cm^2 and Re=1353 and 902 respectively. The results revealed that model (A) has a uniform velocity throughout the aluminium foam heat sink. On the other hand, the velocity contours of models (B) and (C) are different due to the acceleration of the flow inside the channels and a reduction of the water flow rate inside the aluminium foam portion (aluminium foam fin) because of high flow resistance caused by the aluminium foam.



(a) Model (A) **(b) Model (B)** **(c) Model (C)**

Figure 20. Velocity contours at q" =13.8 W/cm2, Re=1353



(a) Model (A) **(b) Model (B)** **(c) Model (C)**

Figure 21. Velocity contours at q" =13.8 W/cm^2, Re=902

Figure 22 shows the velocity profile at Re=1353 in order to provide a better picture of the velocity profiles of the models. The figure illustrates that model (A) achieved a perfect uniform flat velocity profile around 0.07 m/s. The figure also illustrates that models (B) and (C) achieved large parabolic velocity profiles through the channels and lower flat velocity profiles in aluminium foam portion. The same trend can be observed in Figure 23 at Re=902.

This means that due to the high flow resistance caused by the aluminium foam, around 70% of water flow goes into the empty channel and around 30% of the flow goes into the aluminium foam portion, which causes decreases in the positive effect of the aluminium foam on the heat transfer rate. The velocity profiles of models (B) and (C) also revealed that a larger percentage of the water flow entered the aluminium foam portion of model (B). That is why model (B) achieved a higher heat transfer rate than model (C).

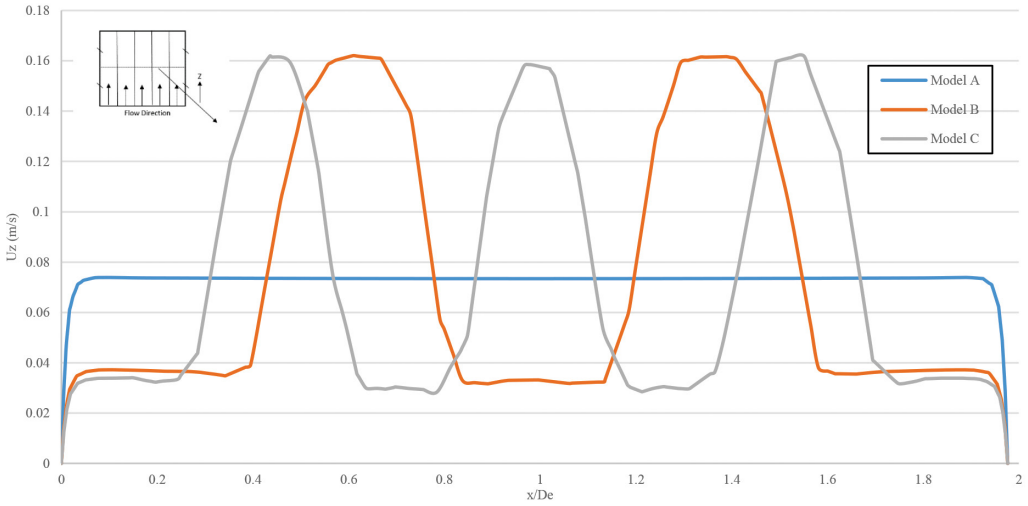


Figure 22. Velocity profile at Re= 1353

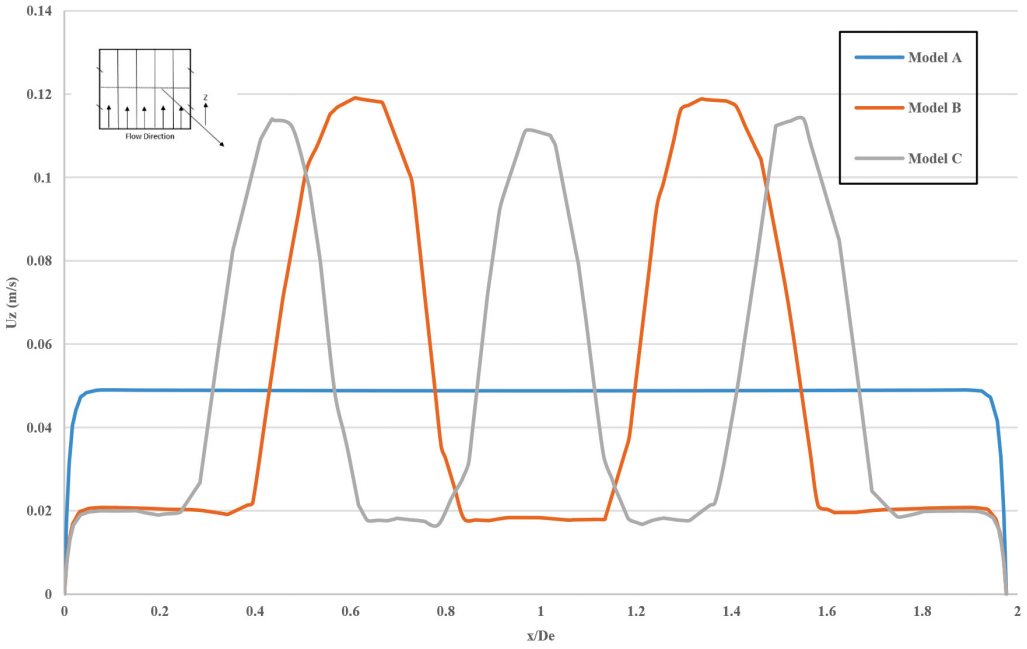


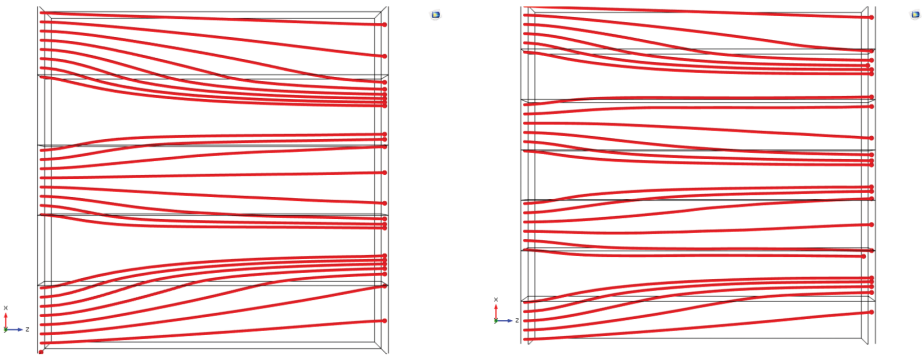
Figure 23. Velocity profile at Re=902

On the other hand, the results showed that the percentage of the flow which goes into the aluminium foam portion decreases along with the flow direction. This phenomenon occurs because the fluid particles turn from the aluminium foam portion (high flow resistance) to the empty channel. In order to take a closer look at this phenomenon, a particle tracing study was conducted using the finite element method. The results are displayed in Figure 24. As we can see, a portion of the fluid particles (red point) turn from the aluminium fin to the empty channel.

Thermal Performance of the Aluminium Foam Heat Sink Models

In order to evaluate the thermal performance of the aluminium foam heat sink models, the pressure drop across the aluminium foam was measured. Figure 25 illustrates the pressure drop verses the Reynolds number for models A, B and C. As we can see in this figure, the pressure drop increases along with increases in the Reynolds number and model (A) achieved a higher pressure drop than models (B) and (C). There was good agreement between the numerical and experimental pressure drop results.

The Fanning friction factor (f) is commonly used to provide information regarding the required pressure drop of the heat exchanger. This is done to ensure that the pressure drop across the foam is non-dimensional. The friction factor of aluminium foam is obtained using the following equation:



(a) Model (B) **(b) Model (C)**

Figure 24. Fluid flow particle tracing at Re=1353

$$f_{friction} = \frac{\Delta p}{4\left(\frac{L}{De}\right)\left(\frac{\rho U^2}{2}\right)} \tag{18}$$

Where L represents the channel length, De represents the hydraulic diameter of the channel, ρ represents the water density, and U represents the velocity of the fluid. Figure 26 illustrates the friction factor of the foam verses the Reynolds number.

As we can see from the analysis of the heat transfer characteristics of water flow through the aluminium foam models, the average Nusselt number for all three models increased along with increases in the Reynolds number. This increase is accompanied by an increase in the required pumping power which is due to an increase in the pressure drop across the aluminium foam. In order to combine the heat transfer rate with the pressure drop, the thermal efficiency index is calculated using the following equation:

$$I_{efficiency} = \frac{Nu_{,avg} * L}{f_{friction} * H} \tag{19}$$

Where $f_{friction}$ represents the Fanning friction factor of the aluminium foam, L represents the channel length, H represents the channel height, and $Nu_{,avg}$ represents the average Nusselt number. The thermal efficiency index combines the heat transfer with the pressure drop across the aluminium foam in order to find the optimal design condition that achieves higher heat transfer with lower pumping power. Figure 27 shows the thermal efficiency index verses the Reynolds number. As we can see in this figure, Model (B) achieved a higher thermal efficiency index over the Reynolds number range compared with models (A) and (C) based on the heat transfer rate and required pumping power. Model (B) achieved a thermal efficiency index of 6.1 at Re=1353, which represents the optimal design condition.

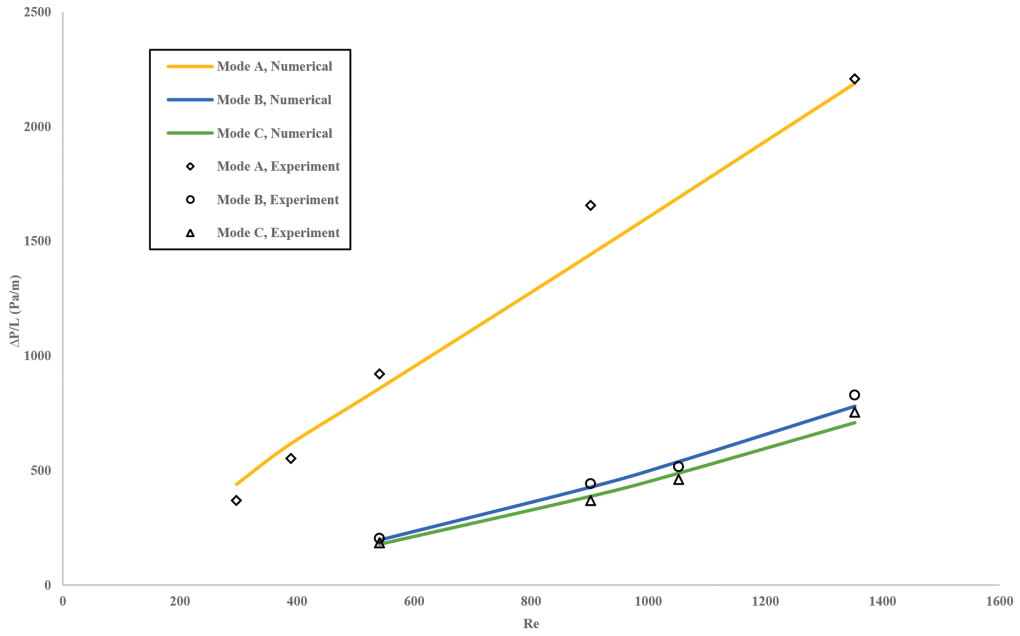


Figure 25. Pressure drop verses Reynold number

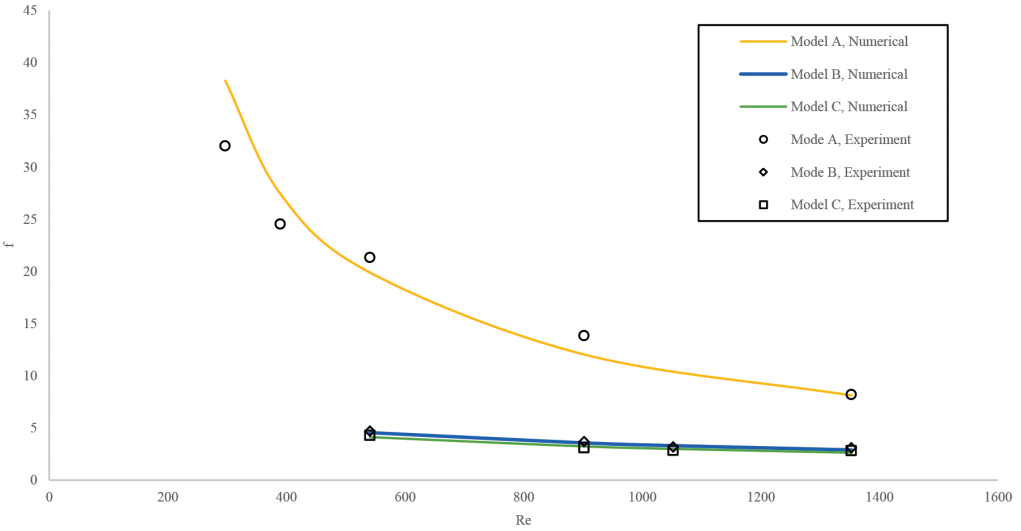


Figure 26. Fanning friction factor verses Reynolds number

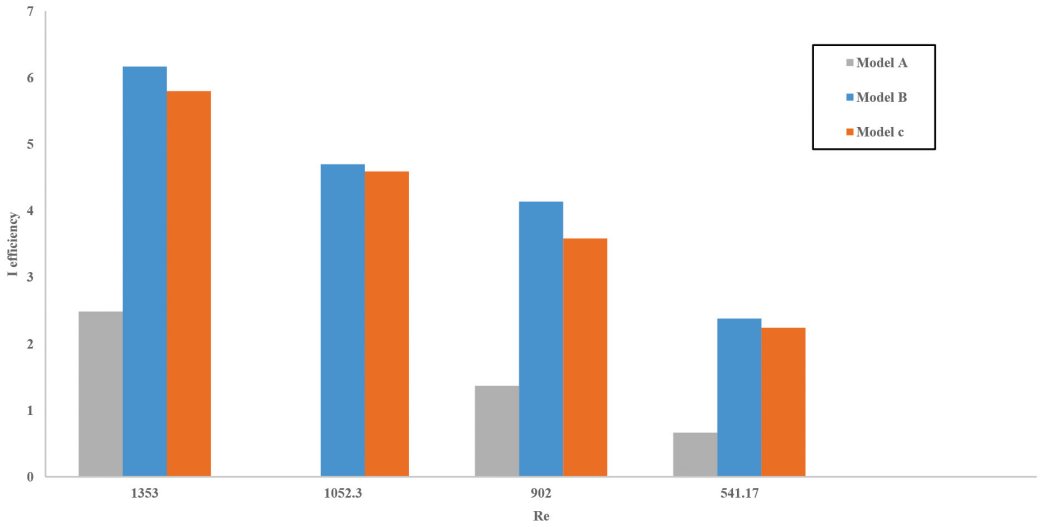


Figure 27. Thermal efficiency index verses Reynold number

CONCLUSIONS

This paper presents an experimental and numerical study of three different aluminium foam heat sink models in the cooling of an Intel core i7 processor (electronic cooling). The aluminium foam models were model (A) (without channels), model (B) with two channels and model (C) with three channels. The aluminium foam was subjected to a water flow covering the non-Darcy flow regime. The following conclusions were drawn based on the findings of the study:

- The local temperature distributions for models A, B and C increase along with increases in the dimensionless flow direction axis, decreasing the Reynolds number and increasing the heat flux.
- Model (A) achieved a lower local temperature than models (B) and (C).
- The numerical local surface temperature results were in good agreement with the experimental results, with a maximum relative error of 2%.
- At any given Reynolds number, the local Nusselt number is inversely proportional to the boundary layer thickness and reaches a constant value at the fully developed region.
- Model (A) achieved a higher local Nusselt number than model (B), and model (B) achieved a higher local Nusselt number than model (C).
- The numerical results of the local Nusselt number were in good agreement with the experimental results, with a maximum relative error of 3%.
- The average Nusselt number of models A, B and C increase along with increases in the Reynolds number.

- Models (B) and (C) reduced the average Nusselt number by 10% and 25% respectively, compared with model (A).
- The pressure drop across the foam was measured. The results revealed that the pressure drop increases as the Reynolds number increases.
- The pressure drop across model (A) was higher than the pressure drop across models (B) and (C).
- The thermal efficiency index combines heat transfer with the pressure drop across the aluminium foam models in order to find the optimal design condition that achieves higher heat transfer with lower pumping power. Model (B) achieved the optimal design condition with a thermal efficiency index of 6.1 at Re= 1353.

ACKNOWLEDGEMENTS

The authors acknowledge the full financial support of the National Science and Engineering Research Council (NSERC).

REFERENCES

Bai, M., & Chung, J. (2011). Analytical and numerical prediction of heat transfer and pressure drop in open-cell metal foams. *International Journal of Thermal Sciences, 50*(6), 869-880.

Bayomy, A., Saghir, M., & Yousefi, T. (2016). Electronic Cooling Using Water Flow in Aluminium Metal Foam Heat Sink: Experimental and Numerical Approach. *International Journal of Thermal Sciences, 109*, 182-200.

Bhattacharya, A., & Mahajan, R. (2002). Finned metal foam heat sinks for electronics cooling in forced convection. *Journal of Electronic Packaging, 124*(3), 155-163.

Boomsma, K., & Poulikakos, D. (2001). On the Effective Thermal Conductivity of a Three-Dimensionally Structured Fluid-Saturated Metal Foam. *International Journal of Heat Mass Transfer, 44*(4), 827–836.

Boomsma, K., Poulikakos, D., & Zwick, F. (2003). Metal foams as compact high performance heat exchangers. *Mechanics of materials, 35*(12), 1161-1176.

Buller, M., & Kilburn, R. (1981). Calculation of surface heat transfer coefficient for electronic module. *Heat Transfer in Electronic Equipment, HTD, ASME Winter Annual Meeting, 20*, 15-29

Calmidi, V. V., & Mahajan, R. L. (2000). Forced Convection in High Porosity Metal Foams. *Journal of Heat Transfer, 122*(3), 557-565.

Calmidi, V., & Mahajan, R. (1999). The effective thermal conductivity of high porosity fibrous metal foams. *Journal of Heat Transfer, 121*(2), 466-471.

COMSOL Multiphysics. (2015, March). *COMSOL Multiphysics*. Retrieved from http://www.comsol.com/comsol-multiphysics

Darcy, H. (1856). *Les Fontaines Publiques de la Ville de Dijon*. Victor Dalmont, Paris.

Ding, X., Lu, L., Chen, C., He, Z., & Ou, D. (2011). Heat transfer enhancement by using four kinds of porous structures in a heat exchanger. In *Applied Mechanics and Materials* (Vol. 52, pp. 1632-1637). Trans Tech Publications.

Dukhan, N., Bağci, Ö., & Özdemir, M. (2015). Thermal development in open cell metal foam: an experiment with constant heat flux. *Journal of Heat and Mass Transfer, 85*, 852-859.

Fu, H., Leong, K., Huang, X., & Liu, C. (2001). An experimental study of heat transfer of a porous channel subjected to oscillating flow. *Journal of heat transfer, 123*(1), 162-170.

Gochman, S., Ronen, R., Anati, I., Berkovits, A., Kurts, T., Naveh, A., … Valentine, R. C. (2003). The Intel Pentium M processor: microarchitecture and performance. *Intel Technology Journal, 7*(2), 21-36.

Hetsroni, G., Gurevich, M., & Rozenblit, R. (2005). Metal foam heat sink for transmission window. *International journal of heat and mass transfer, 48*(18), 3793-3803.

Hooman, K., & Ejlali, A. (2007). Entropy generation for forced convection in a porous saturated circular tube with uniform wall temperature. *International communications in heat and mass transfer, 34*(4), 408-419.

Hwang, J., Hwang, G., Yeh, R., & Chao, C. (2002). Measurement of interstitial convective heat transfer and frictional drag for flow across metal foams. *Journal of Heat Transfer, 124*(1), 120-129.

Igarashi, T., Nakamura, H., & Fukuoka, T. (2004). Pressure drop and heat transfer of arrays of in-line circular blocks on the wall of parallel channel. *International journal of heat and mass transfer, 47*(21), 4547-4557.

Iwasaki, H., & Ishizuka, M. (2000). Forced convection air cooling characteristics of plate fins for notebook personal computers. In *The Seventh Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, 2000. ITHERM 2000.* (Vol. 2, pp. 21-26). IEEE.

Jubran, B., Swiety, S., & Hamdan, M. (1996). Convective heat transfer and pressure drop characteristics of various array configurations to simulate the cooling of electronic modules. *International Journal of Heat and Mass Transfer, 39*(16), 3519-3529.

Kaviany, M. (1995). *Principles of Heat Transfer in Porous Media* (2en Ed.). Springer-Verlag, New York.

Kim, S. Y., Paek, J. W., & Kang, B. H. (2003). Thermal Performance of aluminium foam heat sinks by forced air cooling. *IEEE transactions on components and packaging technologies, 26*(1), 262-267.

Klett, J., Stinton, D., Ott, R., Walls, C., Smith, R., & Conway, B. (2001). *Heat exchangers/radiators utilizing graphite foams*. Oak Ridge National Laboratory, US Department of Energy.

Lee, J. (2009). *Convection performance of nanofluids for electronics cooling*. (Doctoral Dissertation). Stanford University.

Lu, W., Zhao, C., & Tassou, S. (2006). Thermal analysis on metal foam filled heat exchangers Part I: Metal-foam filled pipes. *International journal of heat and mass transfer, 49*(15), 2751-2761.

Mancin, S., Zilio, C., Diani, A., & Rossetto, L. (2012). Experimental air heat transfer and pressure drop through copper foams. *Experimental thermal and fluid science, 36*, 224-232.

Nield, D., Kuznetsov, A., & Xiong, M. (2003). Thermally developing forced convection in a porous medium: Parallel-plate channel or circular tube with walls at constant heat flux. *Journal of Porous Media, 6*(3), 203–212.

Noh, J., Lee, K., & Lee, C. (2006). Pressure loss and forced convective heat transfer in an annulus filled with aluminium foam. *International communications in heat and mass transfer, 33*(4), 434-444.

Rachedi, R., & Chikh, S. (2001). Enhancement of electronic cooling by insertion of foam materials. *Heat and Mass Transfer, 37*(4-5), 371-378.

Seyf, H., & Layeghi, M. (2010). Numerical analysis of convective heat transfer from an elliptic pin fin heat sink with and without metal foam insert. *Journal of Heat Transfer, 132*(7), 1-9.

Sparrow, E., Niethammer, J., & Chaboki, A. (1982). Heat transfer and pressure drop characteristics of array of rectangular modules in electronic equipment. *International Journal of Heat and Mass Transfer, 25*(7), 961-973.

Sparrow, E., Yanezmoreno, A., & Otis, D. (1984). Convective heat transfer response to height differences in an array of block-like electronic components. *International journal of heat and mass transfer, 27*(3), 469-473.

Sung, H., Kim, S., & Hyun, J. (1995). Forced convection from an isolated heat source in a channel with a porous medium. *International Journal of Heat and Fluid Flow, 16*(6), 527-535.

Taylor, J. (1995). *An Introduction to Error Analysis-Study of Uncertainty in Physical Measurements*. Oxford University Press.

Tzeng, S. C., Jeng, T. M., & Wang, Y. C. (2006). Experimental study of forced convection in asymmetrically heated sintered porous channel with/without periodic baffles. *International Journal of Heat and Mass Transfer, 49*(1), 78-88.

Zhao, C., Kim, T., & Hodson, H. (2004). Thermal Transport in High Porosity Cellular Metal Foams. *Journal of Thermophysics and Heat Transfer, 18*(3), 309-317.

Zhao, C., Lu, W., & Tassou, S. (2006). Thermal analysis on metal-foam filled heat exchangers. Part II: Tube heat exchangers. *International journal of heat and mass transfer, 49*(15), 2751-2761.

Zhao, C. (2012). Review on thermal transport in high porosity cellular metal foams with open cells. *International Journal of Heat and Mass Transfer, 55*(13), 3618-3632.

NOMENCLATURE

q”	Heat flux (W/cm^2)		
Nu_x	Local Nusselt number	F	Body force (N)
$Nu_{,avg}$	Average Nusselt number	D_e	Hydraulic diameter (m), $D_e = \frac{4 * \text{Cross section area}}{\text{Wetted perimeter}}$
Re	Reynolds number	T^*	Dimensionless temperature $\left[T^* = \frac{T - T_{in}}{q'' * \frac{D_e}{K_{eff}}}\right]$
U	Velocity fields (m/s)	x^*	Dimensionless flow direction axis, $x^* = \frac{x}{D_e}$
K_b	Permeability (m^2)	ΔP	Pressure difference (Pa)
k_f	Thermal conductivity of fluid	$I_{efficiency}$	Thermal efficiency index
d_f	Ligament diameter (m)	**Greek symbols**	
k_s	Thermal conductivity of solid (W/m.K)	ρ	Density (kg/m^3)
K_{eff}	Effective thermal conductivity (W/m.K)	C_p	Specific heat (J/kg.℃)
T_x	Local surface temperature (℃)	ε	Porosity
$f_{friction}$	Fanning friction factor	υ	Kinematic viscosity (m^2/s)
T_{in}	Inlet water bulk temperature (℃)	β_f	Forchheimer coefficient
h_x	Local heat transfer coefficient (W/m^{2}K)	μ	Dynamic viscosity (Pa. s)

Pertanika J. Sci. & Technol. 25 (1): 247 - 254 (2017)

SCIENCE & TECHNOLOGY

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

Molecular dynamics study of the diffusion behaviour of Li in $Li_{10}GeP_2S_{12}$

Kawano, S. and Iikubo, S.*

Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu 808-0196, Japan

ABSTRACT

In this study, we investigated the anisotropic Li diffusion in $Li_{10}GeP_2S_{12}$ using a molecular dynamics method. Although the calculated ion conductivity is slightly lower in comparison to the calculated and experimental ion conductivity of previous studies, our results support anisotropic Li diffusion. These findings show that the anisotropic Li diffusion estimated from the activation energy is an important property of $Li_{10}GeP_2S_{12}$, and a potential factor for good ion conductivity. Further, a high diffusion coefficient was observed in $Li_{10}P_3S_{12}$, which is a good Li-ion conductor.

Keywords: Solid electrolytes, $Li_{10}GeP_2S_{12}$, molecular dynamics simulation, Li-ion battery, activation energy, ion conductivity

INTRODUCTION

A Li-ion battery is nowadays widely used in electric vehicles and portable devices. Despite their usefulness, the liquid electrolytes used in them are potentially a fire risk. Therefore, solid electrolytes have been studied with regards to their safety and stability (Hayashi et al., 2001; Kanno et al., 2001; Homma et al., 2010; Holzwarth et al., 2011; Homma et al., 2011; Bron et al., 2013; Lepley et al., 2013; Hori et al., 2015). In recent studies, Kamaya et al. (2012) reported a new solid electrolyte, $Li_{10}GeP_2S_{12}$, which shows the highest Li-ion conductivity. The discovery of this material has led to intensive studies based on experimental techniques (Kuhn et al., 2013; Kuhn et al., 2014), molecular dynamics simulation (Adams et al., 2012), and first-principles calculations (Mo et al., 2012; Ong et al., 2013; Du et al., 2014). The anisotropic nature of Li diffusion is an interesting property of this material, and may explain its high conductivity. A number of previous studies supported the anisotropic Li diffusion (Mo et al., 2012; Ong et al., 2013; Du et al., 2014). On the other hand, Kuhn et

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
kawano-shoya@edu.life.kyutech.ac.jp (Kawano, S.),
iikubo@life.kyutech.ac.jp (Iikubo, S.)
*Corresponding Author

ISSN: 0128-7680

al. (2013) reported the nearly isotropic Li hopping process in the bulk lattice of $Li_{10}GeP_2S_{12}$. In this study, the anisotropy of Li diffusivity has been investigated using a molecular dynamics method. Our results support the anisotropic ion conductivity in the $Li_{10}GeP_2S_{12}$.

METHODOLOGY

Molecular dynamics simulations were conducted in order to investigate the Li-ion diffusivity. A $3 \times 3 \times 2$ supercell was used in this simulation for $Li_{10}GeP_2S_{12}$ and $Li_{10}P_3S_{12}$. These simulations were carried out at 5, 50, 100, 150, 200, 250, 300, 350, and 400 K; at each temperature, the system was kept for 400 ps for structure relaxation, and then, the data were collected. The collected data between 250 K and 400 K were used for the estimation of Li diffusivity. The potential parameters that control interaction between atoms are the values determined by Adams et al. (2012). In our simulations, we used an orthorhombic unit cell, which is different from that used in Adams et al. which used a tetragonal unit cell.

From the simulation data a number of physical properties were derived. The diffusion coefficient D is defined as,

$$D = (1/2dt)\langle[r(t)]^2\rangle \quad (1)$$

where $<[r(t)]^2>$ and t denote the mean square displacement and time, respectively. Here, $d = 3$, which is the dimension of the lattice, in which diffusion takes place. The value of D is obtained by performing a linear fitting to the relationship of mean square displacement against $2dt$. The ion conductivity σ was derived from diffusion coefficient according to the following relation:

$$\sigma = Nq^2D/k_BT \quad (2)$$

where, the number density N of the mobile Li-ions of charge q was calculated from the crystal structure. The activation energy E for Li-ion diffusion is obtained by fitting the Arrhenius equation to the diffusion coefficient:

$$D = D_0\exp(-E/k_BT) \quad (3)$$

where, D_0, k_B, and T denote a constant, the Boltzmann's constant, and the temperature, respectively.

RESULTS AND DISCUSSION

The potential parameters used in this study were confirmed to reproduce a number of experimental results. Figure 1 shows the dependence of the unit cell volume of $Li_{10}GeP_2S_{12}$on the temperature. The experimental data at room temperature, which is reported by Kamaya et al. (2012), is also denoted by filled square in Figure 1. The calculated values in this study are 1.7% larger than the results obtained by Kamaya et al. (2012); however, the difference between the calculated and experimental values is in a permissible range. Further, the calculated data are compared with the work of Adams et al. (2012). Our values are lower than that obtained by Adams et al. (2012), but both follow similar trends.

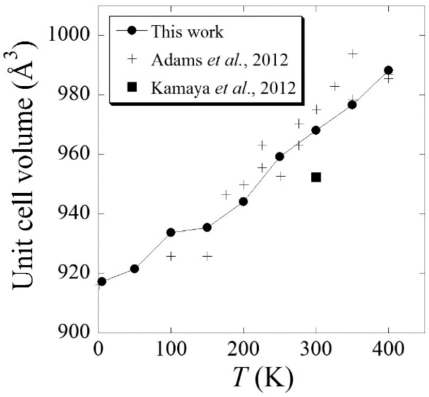


Figure 1. Temperature dependence of unit cell volume

Figure 2 shows the temperature dependence of the lattice parameters of $Li_{10}GeP_2S_{12}$. The calculated data is also compared with the studies of Adams et al. (2012) and Kamaya et al. (2012). Our results along the *a*- and *b*- axes is similar to that of Adams et al. (2012). For the *c*-axis, a minor difference is observed between our results and that of Adams et al. (2012). This can be due to the difference in the constrain condition of the unit cell. As mentioned above, Adams et al. (2012) has used a tetragonal unit cell while our simulations were performed using an orthorhombic unit cell. We think that the Li ions in this material are distributed inhomogeneously. The inhomogeneous distribution causes local lattice distortion, which derives low symmetry structure, such as orthorhombic or monoclinic structure. Therefore, it is unreasonable to impose tetragonal symmetry on this system. Instead, the lower symmetry orthorhombic condition seems to be suitable to reveal realistic Li diffusivity. Regardless of a minor quantitative difference between our results and previous reports, they show an almost similar temperature dependence. Therefore, the simulations in the present study are reliable to discuss about microscopic Li diffusivity.

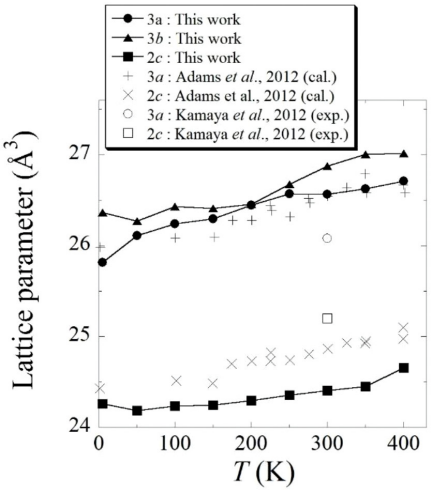


Figure 2. Temperature dependence of lattice parameters

Moreover, an estimation of several physical properties was presented from the molecular dynamics simulation. Figure 3 shows the temperature dependence of the Li-ion diffusion coefficient in $Li_{10}GeP_2S_{12}$. This graph also illustrates the results of Adams et al. (2012), Kamaya et al. (2012), and Mo et al. (2012). It was found that when the temperature decreases, the diffusion coefficient decreases monotonically. The calculated values are slightly smaller than the previous reports; however, they follow a similar trend towards temperature variation. The difference in the values of the diffusion coefficient is due to the difference in the constrain condition of unit cell shape. Similar feature can be seen in ion conductivity.

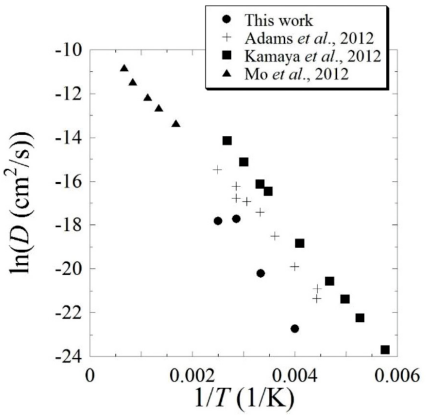


Figure 3. Temperature dependence of the Li-ion diffusion coefficient in $Li_{10}GeP_2S_{12}$

Figure 4 shows the temperature dependence of Li-ion conductivity in $Li_{10}GeP_2S_{12}$ which is compared with Adams et al. (2012) Kamaya et al. (2012) and Mo et al. (2012) Calculated diffusion constant of Li ion and Li-ion conductivity are slightly smaller than experimental values. A similar underestimation is observed in Adams et al. (2012) and Mo et al. (2012). Ong et al. (2013) has pointed out the relationship between lattice parameters and activation energy or ion conductivity. As the lattice parameters increase, the activation energy decreases and the ion conductivity increases. Compared to the previous work, we have obtained a smaller value for ion conductivity, which is mostly because of the short lattice parameter of the *c*-axis, as shown in Figure 2.

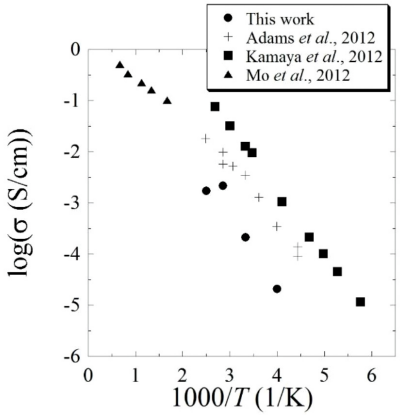


Figure 4. Temperature dependence of Li-ion conductivity in $Li_{10}GeP_2S_{12}$

To investigate the anisotropy of Li-ion conductivity, it is better to examine the direction of decomposed diffusion coefficients D_a, D_b, and D_c. First, the mean square displacement was decomposed along the *a*-, *b*-, and *c*-axes, and then D_a, D_b, and D_c were estimated by performing a linear fitting using Equation (1). Then, the activation energy for Li diffusivity was obtained by fitting Equation (3) to the temperature dependence of D_x ($x = a, b, c$). Table 1 lists the activation energy of Li diffusion, which was calculated on the three crystalline axes. The values we obtained for activation energy are 0.29 eV, 0.29 eV, and 0.13 eV along the *a*-, *b*-, and *c*-directions, respectively. These data support the anisotropic Li diffusion, which is in better agreement with the reports of Adams et al. (2012) and Mo et al. (2012), although they have used the tetragonal unit cell. Therefore, the anisotropic Li diffusion is one of the important properties of this material, suggesting potential factors for better ion conductivity.

Table 1
Activation energy of Li diffusion for three crystalline axes

Activation energy of Li (eV)			
a direction	*b* direction	*c* direction	References
0.29	0.29	0.13	This work
0.30		0.19	Adams et al., 2012 (cal.)
0.28		0.17	Mo et al., 2012 (cal.)
0.22		0.22	Kuhn et al., 2013 (exp.)

Finally, the calculated results for Ge-free $Li_{10}P_3S_{12}$ are presented in Figure 5. For the practical use of this solid electrolyte in batteries the Ge content needs to be as low as possible. The Ge free crystal structure was constructed by the substitution of Ge with P, and the molecular dynamics simulation was performed using the same procedure as we use for $Li_{10}GeP_2S_{12}$. Figure 5 depicts the temperature dependence of the Li-ion diffusion coefficient in $Li_{10}P_3S_{12}$, and the values are compared to the results of $Li_{10}GeP_2S_{12}$ and the first-principle molecular dynamics study of $Li_9P_3S_{12}$ reported by Ong et al. (2013). Note that the calculation on $Li_9P_3S_{12}$ was performed, but the crystal structure of this atomic composition was not recorded during the simulation. The result obtained in $Li_{10}P_3S_{12}$ is compared with the previous report of $Li_9P_3S_{12}$ by Ong et al. (2013). Accordingly, the results of $Li_{10}P_3S_{12}$ follow a similar behaviour to that of $Li_{10}GeP_2S_{12}$ and $Li_9P_3S_{12}$ calculated by Ong et al. (2013). Note that diffusion constant is dependent on the temperature, and the difference between Ong et al. (2013) and our calculation can be attributed to the effect of temperature. This confirms that $Li_{10}P_3S_{12}$ also demonstrates a good Li-ion conductivity. To the best of our knowledge, $Li_{10}P_3S_{12}$ has not been synthesized experimentally, but our study suggests that it has a good conductivity for solid electrolytes used in Li-ion battery.

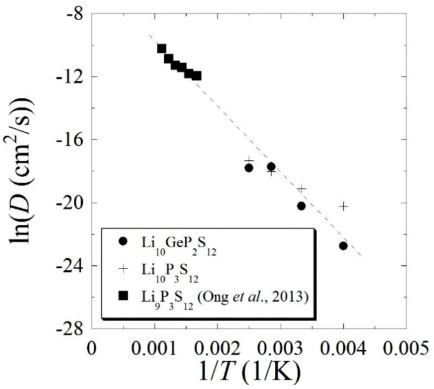


Figure 5. Temperature dependence of the Li ion diffusion coefficient in $Li_{10}P_3S_{12}$

CONCLUSION

We have investigated the anisotropy of Li diffusivity in $Li_{10}GeP_2S_{12}$ using the molecular dynamics method. The results obtained in our study support anisotropic Li diffusion. These findings confirm that anisotropic Li diffusion as an important property of this material, suggesting one of the potential factors for good ion conductivity. Furthermore, our calculations showed that Ge-free $Li_{10}P_3S_{12}$ is expected to have high Li-ion conductivity.

REFERENCES

Adams, S., & Rao, R. P. (2012) Structural requirements for fast lithium ion migration in $Li_{10}GeP_2S_{12}$. *Journal of Materials Chemistry, 22*(16), 7687-7691.

Bron, P., Johansson, S., Zick, K., Schmedt auf der Gunne, J., Dehnen, S., & Roling, B. (2013). $Li_{10}SnP_2S_{12}$: an affordable lithium superionic conductor. *Journal of the American Chemical Society, 135*(42), 15694-15697.

Du, F., Ren, X., Yang, J., Liu, J., & Zhang, W. (2014). Structures, Thermodynamics, and Li^+ Mobility of $Li_{10}GeP_2S_{12}$: A First Principles Analysis. *The Journal of Physical Chemistry C, 118*(20), 10590-10595.

Hayashi, A., Hama, S., Morimoto, H., Tatsumisago, M., & Minami, T. (2001). Preparation of Li_2S–P_2S_5 Amorphous Solid Electrolytes by Mechanical Milling. *Journal of the American Ceramic Society, 84*(2) 477–479.

Holzwarth, N. A. W., Lepley, N. D., & Du, Y. A. (2011). Computer modeling of lithium phosphate and thiophosphate. *Journal of Power Sources, 196*(16), 6870-6876.

Homma, K., Yonemura, M., Kobayashi, T., Nagao, M., Hirayama, M., & Kanno, R. (2011). Crystal structure and phase transitions of the lithium ionic conductor Li_3PS_4. *Solid State Ionics*, *182*, 53-58.

Homma K., Yonemura, M., Nagao, M., Hirayama, M., & Kanno, R. (2010). Crystal Structure of High-Temperature Phase of Lithium Ionic Conductor, Li_3PS_4. *Journal of the Physical Society of Japan, 79*(Suppl. A), 90-93.

Hori, S., Kato, M., Suzuki, K., Hirayama, M., Kato, Y., Kanno, R., & Sprenkle, V. (2015). Phase Diagram of the Li_4GeS_4-Li_3PS_4 Quasi-Binary System Containing the Superionic Conductor Li10GeP2S12. *Journal of the American Ceramic Society, 98*(10), 3352-3360.

Kamaya, N., Homma, K., Yamakawa, Y., Hirayama, M., Kanno, R., Yonemura, M., … Mitsui, A. (2011). A lithium superionic conductor. *Nature materials, 10*(10), 682-686.

Kanno, R., & Murayama, M., (2001). Lithium Ionic Conductor Thio-LISICON The Li_2S-GeS_2-P_2S_5 System. *Journal of the Electrochemical Society, 148*(7), A742-A746.

Kuhn, A., Duppel, V., & Lotsch, B. V. (2013). Tetragonal $Li_{10}GeP_2S_{12}$ and Li_7GePS_8 – exploring the Li ion dynamics in LGPS Li electrolytes. *Energy & Environmental Science, 6*(12)*,* 3548-3552.

Kuhn, A., Gerbig, O., Zhu, C., Falkenberg, F., Maier, J., & Lotsch, B. V. (2014). A new ultrafast superionic Li-conductor: ion dynamics in $Li_{11}Si_2PS_{12}$ and comparison with other tetragonal LGPS-type electrolytes. *Physical Chemistry Chemical Physics, 16*(28), 14669-14674.

Lepley, N. D., Holzwarth, N. A. W., & Du, Y. A. (2013). Structures, Li^+ mobilities, and interfacial properties of solid electrolytes Li_3PS_4 and Li_3PO_4 from first principles. *Physical Review B, 88*(10), 104103.

Mo, Y., Ong, S. P., & Ceder, G. (2012). First Principles Study of the $Li_{10}GeP_2S_{12}$ Lithium Super Ionic Conductor Material. *Chemistry of Materials, 24*(1), 15-17.

Ong, S. P., Mo, Y., Richards, W. D., Miara, L., Lee, H. S., & Ceder, G. (2013). Phase stability, electrochemical stability and ionic conductivity of the $Li_{10\pm1}MP_2X_{12}$ (M = Ge, Si, Sn, Al or P, and X = O, S or Se) family of superionic conductors. *Energy & Environmental Science, 6*(1), 148-156.

Pertanika J. Sci. & Technol. 25 (1): 255 - 262 (2017)

SCIENCE & TECHNOLOGY

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

Temperature Measurement and Optimisation in Machining Magnesium Alloy Using RSM and ANOVA

Viswanathan R[1]*, Ramesh S[2], Elango N[3] and Kamesh Kumar D[2]

[1]*Mechanical Engineering Department, Kongunadu College of Engineering and Technology, Thottiam, Tamilnadu, 621215, India*

[2]*Mechanical Engineering Department, KCG College of Technology, Chennai, Tamilnadu, 600097, India*

[3]*Faculty of Engineering, Technology and Built Environment, UCSI University (North campus), 56000 Cheras, Kuala Lumpur, Malaysia*

ABSTRACT

This paper focuses on examining the 'cutting zone temperature' while performing turning operation on AZ91Mg alloy using cemented carbide tools. The regression model is developed by using the RSM techniques based on experimental results. It is revealed that the cutting speed (v) is the most dominant factor affecting cutting zone temperature. The developed models of cutting zone temperature sufficiently map within the range of the turning conditions considered. The adequacy and accuracy of the regression equation is justified through ANOVA. It is found that the optimal combinations of machining parameters minimize the cutting temperature.

Keywords: ANOVA, Cutting temperature, Magnesium, Optimization, Turning and Response surface methodology

INTRODUCTION

Several studies (Kleiner et al., 2003; Thein at al., 2009; Hirsch & Al-Samman, 2013) proposed that the main advantage of magnesium alloy is its low density of 1.8 g/cm$^{3.}$ Since it possesses a high stiffness- to-weight ratio and a high strength-to-weight, it has been considered the best choice for utilisation automobile and aerospace industry.

The main problem in the machining of magnesium alloy is auto ignition risk and the particles size generated in the machining process. In the event of a fire accident it is important that a water based coolant is not

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
vissona2005@yahoo.co.in (Viswanathan R),
ramesh_1968in@yahoo.com (Ramesh S),
cad.elango.n@gmail.com (Elango N),
kameshmech111@gmail.com (Kamesh Kumar D)
*Corresponding Author

ISSN: 0128-7680

used to extinguish a magnesium fire due to the formation of highly explosive hydrogen when water reacts with magnesium (Tomac & Tonnessen, 1991; Weinert et al., 2004; Kulekci, 2008).

In the machining process, majority of the energy is transformed into heat (Peloubet, 1965; Diniz & José de Oliveira, 2004; Park et al., 2009). The heat generation during machining process increases the cutting zone temperature. An Investigation was carried out on the AISI 52100 alloy steel with multilayer coated carbide insert. The effect of cutting parameters were determined using ANOVA. The quadratic model equation was used to predict the output response. Response surface methodology (RSM) was utilized to find out the optimal machining parameters. It is concluded the cutting temperature is highly influenced by cutting speed and feed rate (Shihab et al., 2014).

In machining operations, cutting forces, tool life and workpiece surface integrity, are strongly affected by cutting temperature. The yield strength of the workpiece material was decreased at higher cutting temperatures. Tool life was reduced due increased workpiece surface temperatures (Ghani et al., 2008).

Montgomery (1997) exposed that, Design of experiments (DoE) is widely used in machining investigations because it refers to the process of planning the experiments. Design methods such as taguchi methods, response surface methodology and factorial designs are now widely used in the experimental approach (Chomsamutr & Jongprasithporn, 2012).

The objective of this study is to select the most influential factors and to consequently select the optimal turning conditions that generate minimum cutting zone temperature.

EXPERIMENTAL PROCEDURE

In this experiment, magnesium alloy (AZ91D) is employed as workpiece in the form of a cylinder. The experiments were conducted based on 3 levels i.e. 3 factor designs of experiment. In turning of AZ91D magnesium alloy, the three main influential process parameters - Cutting speed (V), Feed rate (f) and Depth of cut(d) at three different levels are considered. The input parameters and their levels are designated as shown in table 1. In this work, experiments were conducted based on Taguchi's L9 orthogonal array measured output response are shown in table 2. The turning operations were conducted in a Kirloskar Turn master 35 lathe in dry environment using cemented carbide tools. Cutting zone temperature is measured using METRAVI MT-9 made compact IR thermometer with dual laser targeting with specification of IR range of -50°C to 1000°C.

Table 1
Turning parameters with machining conditions

Control parameters	**Levels**		
	1	**2**	**3**
Cutting Speed (V), m/min	40	80	120
Feed (f), mm/rev	0.10	0.15	0.20
Depth of cut (d), mm	0.50	0.75	1.00

Table 2
Experimental result based on L9 orthogonal array

Trial	Designation	Actual factor			Cutting Zone Temperature	S/N ratio
		V (m/min)	f (mm/rev)	d (mm)	T (°C)	
1	$A_1B_1C_1$	40	0.10	0.5	62	-35.85
2	$A_1B_2C_2$	40	0.15	0.75	65	-36.26
3	$A_1B_3C_3$	40	0.20	1	68	-36.65
4	$A_2B_1C_2$	80	0.10	0.75	79	-37.95
5	$A_2B_2C_3$	80	0.15	1	82	-38.28
6	$A_2B_3C_1$	80	0.20	0.5	75	-37.50
7	$A_3B_1C_3$	120	0.10	1	98	-39.82
8	$A_3B_2C_1$	120	0.15	0.5	92	-39.28
9	$A_3B_3C_2$	120	0.20	0.75	90	-39.08

OPTIMIZATION OF PROCESS PARAMETERS

The tools used in this study are Taguchi method of optimization, Response surface methodology and Regression analysis for developing the empirical model.

Taguchi method

Ramesh et al. (2012) state that, taguchi method of robust design has been extensively used in the manufacturing processes for single performance/quality characteristics. The aim may be to minimize surface roughness of the machined parts, maximize metal removal rate, minimize cutting force, and minimize tool wear rate. Taguchi algorithm aids in converting the quantitative evaluation of particular objective as signal to noise (S/N) ratio. The target value of the objective is achieved by minimizing the variation with ability of S/N ratio.

In present work, smaller the better type S/N quality was utilized to obtain optimal parameter for cutting zone temperature. Hence smaller the better type S/N ratio is expressed in eqn (1).

$$S/N = -10\log\frac{1}{n}\left(\sum_{i=1}^{n} y_i^2\right) \quad (1)$$

Where *S/N* is the signal-to-noise ratio, *n* is the no. of measured observations, and *y* is the experimentally measured data.

Regression analysis

Regression analysis is mainly used to determine the relationship between input factors and experimental results. Almost all engineering fields and technologies are utilizing this statistical tool for developing a model for output response (Ashvin & Nanavati, 2013). The estimated

regression coefficient of cutting zone temperature is shown in table 3. The p value for the output response model is obtained as less than 0.05 which indicates that the model terms are significant. In addition, the R^2 value is 0.9909 and the Adj. R^2 is 0.9854. The predicted R^2 value 0.9666 is very close to Adj. R^2 value. The obtained R^2 value is close to the desired 1.

Table 3
Estimated regression coefficient for cutting zone temperature (T)

Predictor	Coef	SE Coef	T	P
Constant	44.167	2.969	14.88	0.000
Cutting Speed (v)	0.35417	0.01559	22.72	0.000
Feed (f)	-20.00	12.47	-1.60	0.170
Depth of cut (d)	12.667	2.494	5.08	0.004

Response surface methodology was utilized to develop model from the experimental observation with the view of understanding the turning process. In this study, MINITAB 16 software package was used for the optimization work. The following regression equation was developed using RSM method. The cutting zone temperature is predicted using the Eq. (2).

The regression equation is

Cutting Zone Temperature (T) = 44.2 + 0.354 Cutting Speed (V) - 20.0 Feed (f) + 12.7 Depth of cut (d) (2)

From table 4 on the cutting zone temperature, the optimal parameter is identified as cutting speed - 40m/min, feed rate - 0.2mm/rev and depth of cut - 0.5mm. Hence the optimum condition is represented as $A_1B_3C_1$. Figure 1 shows the main effect plot for cutting zone temperature. It reveals that cutting temperature increases as cutting speed increases and feed rate does not impact on the cutting temperature.

Table 4
Response table for cutting zone temperature

Level	Cutting Speed (v)	Feed (f)	Depth of cut (d)
1	-36.25	-37.87	-37.54
2	-37.91	-37.94	-37.77
3	-39.40	-37.75	-38.25
Delta	3.14	0.19	0.71
Rank	1	3	2

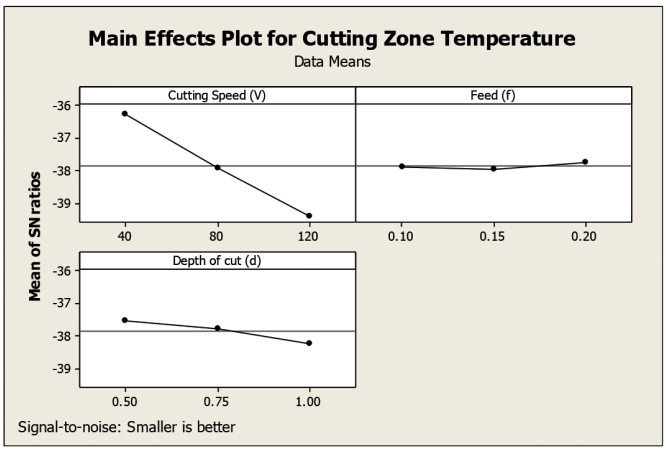


Figure 1. Main effect plot for Cutting Zone Temperature

To determine the control parameter that has a major effect on cutting zone temperature, ANOVA is performed and the results are given in the table 5. From the ANOVA table it is seen that cutting speed (v) is the most critical control factor significantly contributing 94.61% followed by depth of cut by 4.72% whereas the contribution of feed rate is negligible.

Table 5
ANOVA for cutting zone temperature (T)

Source	DF	Seq SS	Adj MS	F	P	% Contribution
Cutting Speed (V)	2	1204.17	1204.17	516.07	0.000	94.61
Feed (f)	2	6.00	6.00	2.57	0.170	0.47
Depth of Cut (d)	2	60.17	60.17	25.79	0.004	4.72
Residual Error	2	11.67	2.33			0.18
Total	8	1282.00				

The model's adequacy has been investigated by the assessment of residuals. In Figure 2 the normal probability plot of the residuals follows a straight line in which all the points are very nearer to the straight line and distributed evenly on both sides of the straight line. In residual versus fit plot, the residual appears to be randomly scattered around zero. The histogram of residual shows the distribution of residual for cutting zone temperature, distributed evenly on both sides zero residual. Residual versus order graph shows that the residual is higher for the observation order 8 and for majority of the observation, the residual values are situated zero line. The Figure 3 represents the similarity of experimental and predicted values of cutting zone temperature. It is revealed that the predicted values are agreed with the measured values.

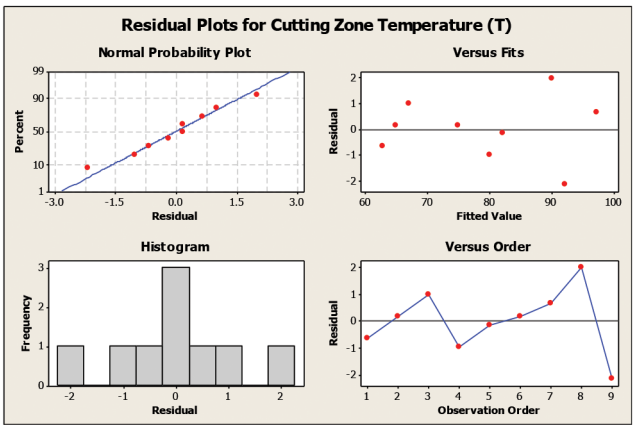


Figure 2. Residual plot of cutting zone temperature during regression analysis

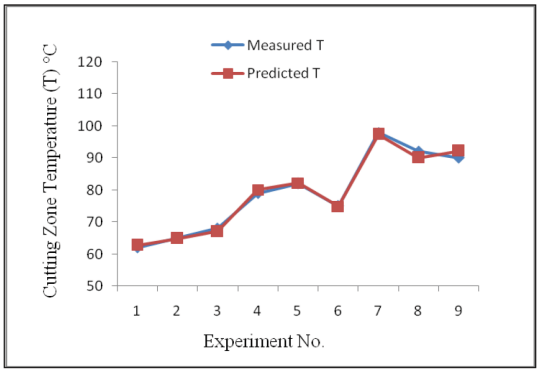


Figure 3. Comparison of experimental and predicted responses

Response Surface Methodology

The main aim of the work associated with machining magnesium alloy is to attain the minimum cutting temperature of the optimal control factors. The response surface optimization technique is used to find out optimal cutting parameters. Here, the objective is to minimize cutting temperature (T). RSM optimization results for cutting zone temperature are shown in Figure 4 and Table 6. From Figure 4, it can be seen that the cutting speed is the main impact factor on the cutting zone temperature. Feed have no significant effect on the output response. Optimal machining parameters are found to be cutting speed (V) at 40 m/min, feed (f) at 0.2 mm/rev and depth of cut (d) at 0.5 mm. The optimized cutting zone temperature parameter is T = 60.667°C.

Table 6
Response optimization for cutting zone temperature

Parameters	Goal	Optimum condition			Lower	Target	Upper	Pre.res	Desirability
		v (m/min)	F (mm/rev)	d (mm)					
T (°C)	Minimum	40	0.20	0.5	62	62	98	60.667	1

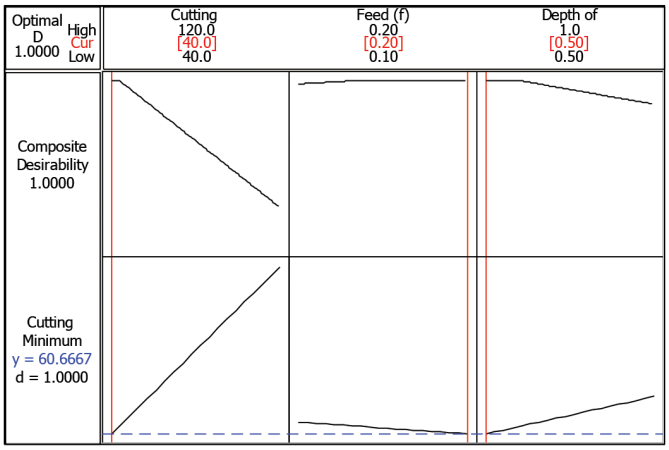


Figure 4. Response optimization for cutting zone temperature parameters

Confirmation test

With the identified optimal process parameter, the confirmation test is conducted to validate the analysis. In the confirmation test, an experiment has been conducted with optimal process parameter settings. Table 7 shows the comparison of experimentally obtained temperature value with RSM predicted value. Hence the model equation for the cutting zone temperature developed using RSM can be used to effectively predict the cutting zone temperature.

Table 7
Conformation test values

optimal process parameters			Experimental	Predicted	Error
v (m/min)	f (mm/rev)	d (mm)	T (°C)	T (°C)	%
40	0.20	0.50	63	60.667	3.70

CONCLUSION

The following conclusions have been derived on turning of AZ91D Magnesium alloy. The results are as follows:

(a) The cutting speed (V) is the main dominant factor on the cutting zone temperature (T). Increasing the cutting speed increases cutting zone temperature.

(b) From regression analysis, a prediction model has been developed for the cutting zone temperature in conditions of control factors. Predicted values are in good agreement with measured output responses.

(c) From RSM optimization, the minimum cutting zone temperature is found to be turning cutting speed at 40m/min, feed at 0.2mm/rev and depth of cut at 0.50mm.

(d) A conformation test result proves that the developed regression model can be used for turning of Mg alloy less than 4% error.

ACKNOWLEDGEMENTS

The authors would like to convey their gratefulness to the KCG College of Technology, Chennai, India for having given its fullest assistance for carrying out this work.

REFERENCES

Chomsamutr, K., & Jongprasithporn, S. (2012). Optimization Parameters of tool life Model Using the Taguchi Approach and Response Surface Methodology. *International Journal of Computer Science Issues*, *9*(1)*,* 120-125.

Diniz, A. E., & José de Oliveira, A. (2004). Optimizing the Use of Dry Cutting in Rough Turning Steel Operations. *International Journal of Machine Tools and Manufacture*, *44*(10), 1061-1067.

Ghani, M. U., Abukhshim, N. A., & Sheikh, M. A. (2008). An investigation of heat partition and tool wear in hard turning of H13 tool steel with CBN cutting tools. *The International Journal of Advanced Manufacturing Technology, 39*(9-10), 874-888.

Hirsch, J., & Al-Samman, T. (2013). Superior light metals by texture engineering: Optimized aluminium and magnesium alloys for automotive applications. *Acta Materialia*, *61*(3), 818–843.

Kleiner, M., Geiger, M., & Klaus, A. (2003). Manufacturing of Lightweight Components by Metal Forming. *CIRP Annals –Manufacturing Technology*, *52*(2), 521-542.

Kulekci, M. K. (2008). Magnesium and its alloys applications in automotive industry. *International Journal of Advanced Manufacturing Technology*, *39(9–10),* 851–865.

Makadia, A. J., & Nanavati, J. I. (2013). Optimisation of machining parameters for turning operations based on response surface methodology. *Measurement, 46*(4), 1521-1529.

Montgomery, D. C. (1997). *Design and Analysis of Experiments* (4th Ed). John Wiley & sons Inc.

Peloubet, J. A. (1965). Machining Magnesium-a Study of Ignition Factors. *Fire Technology*, *1*(1), 5-14.

Park, C. W., Kwon, K. S., Kim, W. B., Min, B. K., & Park, S. J. (2009). Energy Consumption Reduction Technology in Manufacturing- A Selective Review of Policies, Standards, and Research. *International Journal of Precision Engineering Manufacturing, 10*(5), 151-173.

Ramesh, S., Karunamoorthy, L., & Palanikumar, K. (2012). Measurement and analysis of surface roughness in turning of aerospace titanium alloy (gr5). *Measurement*, *45*(5), 1266–1276.

Shihab, S. K., Khan, Z. A., Mohammad, A., & Siddiqueed, A. N. (2014). RSM based study of cutting temperature during hard turning with multilayer coated carbide insert. *Procedia Materials Science, 6*, 1233-1242.

Thein, M. A., Lu, L., & Lai, M. O. (2009). Effect of milling and reinforcement on mechanical properties of nanostructured magnesium composite. *Journal of Materials Processing Technology*, *209*(9), 4439– 4443.

Tomac, N., & Tonnessen, K. (1991). Formation of flank build-up in cutting magnesium alloys. *CIRP Ann-Manuf Technol, 40*(1), 79–82.

Weinert, K., Inasaki, I., & Sutherland, J. W. (2004). Wakabayashi T Dry machining and minimum quantity lubrication. *CIRP Ann-Manuf Technol, 53*(2), 511–537.

Pertanika J. Sci. & Technol. 25 (1): 263 - 270 (2017)

SCIENCE & TECHNOLOGY

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

Clique Regular Graphs

Bhat, R. S.[1]*, Bhat, Surekha, R.[2], Bhat, Smitha, G.[1] and Udupa, Sayinath.[1]

[1]*Department of Mathematics, Manipal Institute of Technology, Manipal University, Manipal, 576104, INDIA*

[2]*Department of Mathematics, Milagres College, Kallainapur, Udupi, 576105, INDIA*

ABSTRACT

A maximal complete subgraph of G is a clique. The *minimum* (maximum) *clique number* $\vartheta = \vartheta(G)$ $(\omega = \omega(G))$ is the order of a minimum (maximum) clique of G. A graph G is clique regular if every clique is of the same order. Two vertices are said to dominate each other if they are adjacent. A set S is a dominating set if every vertex in V- S is dominated by a vertex in S. Two vertices are independent if they are not adjacent. The independent domination number $i = i(G)$ is the order of a minimum independent dominating set of G. The order of a maximum independent set is the independence number $\beta_0 = \beta_0(G)$. A graph G is well covered if $i(G) = \beta_0(G)$. In this paper it is proved that a graph G is well covered if and only if $\bar{G}$ is clique regular. We also show that $\vartheta(\bar{G}) = i(G)$.

Keywords: Clique, Minimum clique number, Maximum clique number, Domination number, Well covered graphs and clique regular graphs

INTRODUCTION

All the graphs considered in this paper are finite, simple and undirected. For any undefined terminologies and notations refer to Harary (1969). If a graph G is isomorphic to r copies of a graph H, then we write it as $G = r\,H$. Two vertices are said to dominate each other if they are adjacent. A set $S \subseteq V$ is a dominating set if every vertex in V- S is dominated by a vertex in S. The *domination number* $\gamma = \gamma(G)$ is the order of a minimum dominating set of G. The *upper domination number* $\Gamma = \Gamma(G)$ is the maximum order of a minimal dominating set. These concepts of domination are well studied in (Cockayane & Hedetniemi, 1977; Walikar et al., 1979; Haynes et al., 1998; Kamath &Bhat, 2006; Kamath &Bhat, 2007; Bhat et al., 2011; Bhat, Surekha and Bhat, 2011; Bhat et al., 2013; Bhat et al., 2014). The vertex covering

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
rs.bhat@manipal.edu; ravishankar.bhats@gmail.com (Bhat, R. S.),
surekharbhat@gmail.com (Bhat, Surekha, R.),
smitha.holla@manipal.edu (Bhat, Smitha, G.),
sayinath.udupa@manipal.edu (Udupa, Sayinath.)
*Corresponding Author

ISSN: 0128-7680

number $\alpha_0 = \alpha_0(G)$ is the minimum number of vertices needed to cover all the edges of a graph while independence number $\beta_0 = \beta_0(G)$ is the maximum number of vertices in an independent set of *G*. These two numbers are related by classical Gallai's Theorem: $\alpha_0(G) + \beta_0(G) = p$. The *upper vertex covering number* $\epsilon = \epsilon(G)$ is the maximum order of a minimal covering of *G*. The *independent domination number* $i = i(G)$ is the minimum order of an independent dominating set of G. Naturally, we have an extension of Gallai's theorem to these numbers as: $\epsilon(G) + i(G) = p$. A maximal complete subgraph is a *clique.* The minimum number of cliques (not necessarily maximal) that cover all the vertices of a graph is well known in graph theory as *partition number* $\theta_0 = \theta_0(G)$ introduced by (Berge, 1962) and has been celebrated in Berge's conjecture on perfect graphs. Choudam et al. (1975) studied its edge analogue line *clique covering number* $\theta_l(G)$ defined as the minimum number of cliques that cover all the lines of a graph. The minimum number of colours needed to properly colour the vertices of *G* is the *chromatic number* $\chi = \chi(G)$. Since independent sets and cliques exchange their properties on complementation $\theta_0(G) = \chi(\bar{G})$. Bhat et al. (2013) defined block domination parameters and studied their relationship between other domination parameters. In this paper we obtain few bounds on minimum clique number and characterized well covered graphs using clique regular graphs.

MINIMUM CLIQUE NUMBER

The *minimum clique number* $\vartheta(G)$ is the order of a minimum clique of *G while* the *maximum clique number* $\omega(G)$ is the order of a maximum clique of *G.* It is immediate that $\vartheta(G) \leq \omega(G)$. Even though these two parameters are well studied in literature, the first parameter *minimum clique number* $\vartheta(G)$ received less attention and we are interested in it than the later. If *G* has an isolated vertex, then $\vartheta(G) = 1$. If *G* is a triangle free graph without isolates, then $\vartheta(G) = 2$. The girth $g(G)$ of a graph is the length of the shortest cycle in *G*. Girth of a graph is defined if *G* has a cycle otherwise we define $g(G) = \infty$. Since girth of any graph is at least 3, $\vartheta(G) \leq g(G)$ if $\vartheta(G) \leq 3$. Moreover, $\vartheta(G) \geq 4$ if then every minimum clique contains a triangle and hence $g(G) = 3 < 4 \leq \vartheta(G)$. It is well known that $\omega(G) = \beta_0(\bar{G})$. A similar result for minimum clique number is obtained in the next proposition.

Proposition 1 *For any graph G,*

$$\vartheta(G) = i(\bar{G})$$

Proof. Let $\vartheta(G) = k$, and S be the set of vertices of a minimum clique of G. Since independent sets and cliques exchange their properties on complementation, S forms a maximal independent set with minimum number of vertices in $\bar{G}$. Then by Ore's theorem (Ore, 1962), we have every maximal independent set is a minimal dominating set. Therefore, S is a minimum independent dominating set of $\bar{G}$. Hence $\vartheta(G) = k = |S| = i(\bar{G})$.

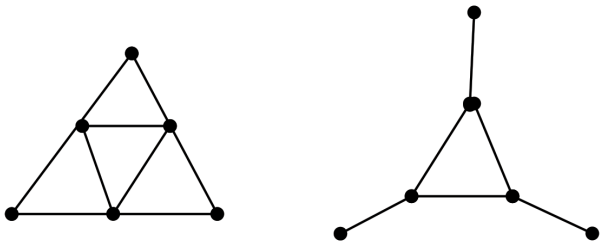


Figure 1. Hajo's Graph G and its compliment (Haynes et al.,1998)

We can observe that for the Hajo's graph *G* in Figure 1, $\vartheta(\mathrm{G}) = 3$ and for the compliment of Hajo's graph $\mathrm{i}(\overline{\mathrm{G}}) = 3$.

The doma*tic number* (*G*) is the maximum order of a partition of vertex set in to dominating sets. As $\omega(G)$ served a best lower bound for chromatic number is evident from the known inequality, $\omega(G) \leq \chi(G) \leq 1 + \Delta(G)$ (Cockayane &Hedetniemi, 1977, p. 250). A clique graph $K_G(G)$ of *G* is a graph with vertex set as cliques of *G* and any two vertices in $K_G(G)$ are adjacent if and only if the corresponding cliques in *G* have a vertex in common. Independence graph $I(G)$ is a graph with vertex set as set of all maximal independent sets of *G* and any two vertices in $I(G)$ are adjacent if they have a vertex in common. We observe that any maximal independent set in *G* is a clique in $\bar{G}$ and vice versa. Hence $K_G(\bar{G}) \cong I(G)$. Cockayne &Hedetniemi (1977, p. 257) proved that if $K_G(G)$ is an even cycle, then $\vartheta(G) \leq d(G)$. Zelinca (1981), constructively shown that the analogous assertion is false if $K_G(G)$ is an odd cycle. Hence $\vartheta(G)$ can exceed the domatic number. Thus $\vartheta(G)$ and $d(G)$ are incomparable. We now provide an upper bound for minimum clique number in terms of minimum degree and order of *G*. We use the following notations. Let $N(v) = \{u \epsilon V | u \text{ is adjacent to } v\}$ and $N[v] = N(v) \cup \{v\}$. Then $\langle N[v] \rangle$ denote the subgraph induced by the set N[v]. Let $\delta(G)$ and $\Delta(G)$ denote the minimum and maximum degree of G while $\bar{\delta} = \delta(\bar{G})$and $\bar{\Delta} = \Delta(\bar{G})$. Let $V_\delta = \{v \epsilon V | \deg(v) = \delta\}$.

Proposition 2 *For any graph G with minimum degree* $\delta(G)$*,*

$$\vartheta(G) \leq 1 + \delta(G).$$

Further, the equality holds if and only if $\langle N[v] \rangle$ *is a minimum clique of G for every* $v \epsilon V_\delta$

Proof. We first note that $\Delta + \bar{\delta} = \bar{\Delta} + \delta = p - 1$. It is well known that $i(G) \leq p - \Delta(G)$ (Haynes et al., 1998, p. 312). Therefore $\vartheta(G) = i(\bar{G}) \leq p - \bar{\Delta} = 1 + \delta(G)$.

Suppose that $\vartheta(G) = 1 + \delta(G)$. Then if $\langle N[v] \rangle$ is not a minimum clique of *G* for some $v \epsilon V_\delta$ then $\vartheta(G) < |\langle N[v] \rangle| = 1 + \delta(G)$ a contradiction.

Converse is straight forward and we omit the proof.

The bound is sharp for the complete graph K_n and star graph $K_{1,n.}$

The following results relate the different graph parameters which appears in (Haynes et al., 1998, p.374).

Proposition 3. *For any graph G*

$$\frac{p}{1+\Delta} \leq \gamma(G) \leq i(G) \leq \beta_0(G) \leq \Gamma(G) \quad [1]$$

On complementing the result [1], we get the next corollary and one can see that the $\vartheta(G)$ fits best in between the known graph parameters.

Corollary 3.1 *For any graph G,*

$$\frac{p}{p-\delta} \leq \gamma(\bar{G}) \leq \vartheta(G) \leq \omega(G) \leq \Gamma(\bar{G}). \quad [2]$$

The idiomatic number $d_i = d_i(G)$ is the maximum order of partition of vertex set in to independent dominating sets. The idiomatic number does not exist for all graphs. A graph *G* is indominable if *G* admits an independent dominating set partition. The maximal clique partition number $\theta_m = \theta_m(G)$ is the maximum order of partition of vertex set in to cliques of *G*. A graph which admits a clique partition is called clique partitionable. Hence $\theta_m(G) = d_i(\bar{G})$. If *G* is indominable then $\bar{G}$ is clique partitionable. If both *G* and $\bar{G}$ are indominable then *G* is called strongly indominable. We now provide an upper bound to domination number of an indominable graph in terms of minimum clique number.

Proposition 4. *If G or* $\bar{G}$ *is indominable*

Proof. If *G* or $\bar{G}$ is indominable, it is proved that $\vartheta(G) \leq d_i(G) \leq d(G)$ (Walikar et al., 1979, p.109). Therefore $\gamma(G)\vartheta(G) \leq i(G)\vartheta(G) \leq i(G)\, d_i(G) \leq p$. This yields the desired inequality.

The bound is attained for any even cycle, regular bipartite graph or complete graph.

Corollary 4.1 *If G is clique partitionable then* $i(G) \leq d_i(\bar{G}) = \theta_m(G)$- *is the partition vertex set in to maximal cliques of G.*

Walikar et al. (1979) has proved that for any cubic graph, if there exists a maximal clique of order 2 then $\gamma(\bar{G}) = 2$. We now prove a stronger result with much more ease and the above result is a corollary to the next proposition.

Proposition 5. *For any graph G with,* $\vartheta(G) = 2$ *then* $\gamma(\bar{G}) = i(\bar{G}) = 2$

Proof. From Proposition 1, we have $2 = \vartheta(G) = i(\bar{G})$. As every independent dominating set is a dominating set we have $\gamma(\bar{G}) \leq i(\bar{G})$. Suppose $\gamma(\bar{G}) < i(\bar{G})$then $\gamma(\bar{G}) = 1$. As any singleton set is independent we then have $i(\bar{G})$. This is a contradiction to the statement that $i(\bar{G}) = 2$. Therefore $\gamma(\bar{G}) = i(\bar{G})$.

Corollary 5.1. *If G is a cubic graph with a maximal clique of order 2*, *then* $\gamma(\bar{G}) = 2$.

Proof. Since *G* is cubic and there exists a maximal clique of order 2 together implies that . Then the result follows by Proposition 2.5.

CLIQUE REGULAR GRAPHS

The concept of well covered graphs is studied in (Plummer, 1970; Plummer, 1993; Dean &

Zeto, 1994; Ravindra, 1997). A graph *G* is *well covered* if every maximal independent set is of same order. In other words, *G* is well covered if and only if $i(G) = \beta_0(G)$. Equivalently, $\epsilon(G) = \alpha_0(G)$.

The above definition motivated the description of another special class of graphs called clique regular graphs. A graph *G* is *clique regular* if every clique is of same order. Thus *G* is k- *clique regular* graph if $\omega(G) = \vartheta(G) = k$. For example, 3-clique regular graph and 5-clique regular graphs are shown in the Figure 2.

Remark 1. The maximum number of vertices in a minimal vertex cover is called maximum vertex covering number $\varepsilon(G)$. It is proved that for any graph *G,* $\varepsilon(G) + i(G) = p$ (Haynes, et al.,1998, p. 524). Using Proposition 5, this result can now be restated as $\varepsilon(G) + \vartheta(\bar{G}) = p$ or equivalently, $\varepsilon(\bar{G}) + \vartheta(G) = p$.

Remark 2. Similarly, from the above Remark 1, we may write $\alpha_0(G) + \beta_0(G) = p$ (Gallai's Theorem) as $\alpha_0(G) + \omega(\bar{G}) = p$ or equivalently, $\alpha_0(\bar{G}) + \omega(G) = p$.

Example 1.

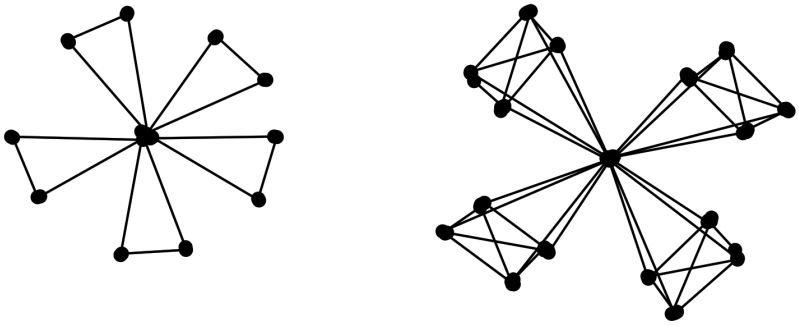


Figure 2. A 3-clique regular and 5-clique regular graphs (Haynes et al.,1998)

The advantage of knowing $\vartheta(G)$ and $\omega(G)$ is that one can easily determine the independent domination number and independence number of $\bar{G}$. Using this technique $i(\bar{G})$ and $\beta_0(\bar{G})$ for some standard graphs are obtained for some standard graph. A double star is a tree $T = K_{1,n} * K_{1,m}$ obtained by joining the two nonpendant vertices of $K_{1,n}$ and $K_{1,m}$.

Proposition 6

(i) *For any double star* $T = K_{1.n} * K_{1.m}$,
$i(\bar{T}) = 2 = \beta_0(\bar{T})$; $\vartheta(\bar{T}) = \min(m+1, n+1)$ and $\omega(\bar{T}) = m + n$

(ii) *For any tree T*, $i(\bar{T}) = 2 = \beta_0(\bar{T})$.

(iii) *For any triangle free graph*, $i(\bar{G}) = 2 = \beta_0(\bar{G})$.

Proof. Note that for any double star, $T = K_{1,n} * K_{1,m}$, $i(T) = \min(m+1, n+1)$, $\beta_0(T) = m + n$ *and* $\vartheta(T) = \omega(T) = 2$. Then the result (i) follows from Proposition 5 and Remark 1. The rest of the results can be proved similarly.

As any Cube graph Q_n, Petersens graph, Hexagonal hub graph in Figure 3 are triangle free graphs without isolates and hence the next corollary.

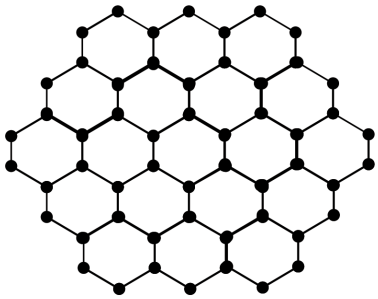


Figure 3. Hexagonal hub graph H (Haynes et al.,1998)

Corollary 6.1

(i) If G is a Petersen's graph, then $i(\bar{G}) = 2 = \beta_0(\bar{G})$

(ii) *For any cube graph* Q_n, $i(\bar{Q}_n) = 2 = \beta_0(\bar{Q}_n)$

(iii) *For the hexagonal hub graph H,* $i(\bar{H}) = 2 = \beta_0(\bar{H})$

(iv) *For any grid graph G,* $i(\bar{G}) = 2 = \beta_0(\bar{G})$

(v) *For any wheel graph*, W_n, $i(\bar{W}_n) = \beta_0(\bar{W}_n) = 4, if\ n = 4$ $and i(\bar{W}_n) = \beta_0(\bar{W}_n) = 3,\ if\ n \geq 5$

Proposition 7 *For any graph G*,

(i) $\beta_0\big(\overline{B_G(G)}\big) = \Delta_{vb}(G) = \Delta_{bv}\big(B_G(G)\big)$

(ii) $i\big(\overline{B_G(G)}\big) = \delta_{cvb}(G) = \delta_{bv}\big(B_G(G)\big)$

(iii) $i\big(\overline{C_G(G)}\big) = \delta_{NPc}(G) = \delta_{cvb}(B_G(G))$

(iv) $\beta_0(\overline{C_G(G)}) = \Delta_c(G) = \Delta_{cvb}(B_G(G))$

Proof. For any graph *G*, every block of the block graph $B_G(G)$ is a clique. Since all the blocks incident on a cut vertex of *G* are mutually adjacent, these blocks form a clique in $B_G(G)$. Therefore, number of blocks incident on a cut vertex $= d_{vb}(G) =$ number of vertices in the corresponding block in $B_G(G) = d_{bv}\big(B_G(G)\big)$. Hence $\Delta_{vb}(G) = \Delta_{bv}\big(B_G(G)\big) = \omega\big(B_G(G)\big)$. Similarly, $\delta_{cvb}(G) = \delta_{bv}\big(B_G(G)\big) = \vartheta\big(B_G(G)\big)$. Then the results (i) and (ii) follow by Proposition 5 and Remark 1.

Again, for any graph *G*, every block of the cutvertex graph $C_G(G)$ is a clique. Since all cutvertices in a nonpendant block are mutually adjacent, these cutvertices form a clique in $C_G(G)$. Therefore, number of cutvertices incident on a block $= d_c(G) =$ number of vertices in

the corresponding block in $C_G(G) = d_{bv}\big(C_G(G)\big)$. Hence $\Delta_c(G) = \Delta_{bv}\big(C_G(G)\big) = \omega\big(C_G(G)\big)$. Similarly, $\delta_{NPc}(G) = \delta_{bv}\big(C_G(G)\big) = \vartheta\big(C_G(G)\big)$. Then the results (iii) and (iv) follow by Proposition 5 and Remark 1.

Proposition 8 *For any graph G*, with maximum degree $\Delta(G)$ *and minimum degree* $\delta(G)$,

(i) $\beta_0\big(\overline{L(G)}\big) = \Delta(G)$

(ii) $i\big(\overline{L(G)}\big) = \delta(G)$

Proof. Let *v* be a vertex of maximum degree $\Delta(G)$ and x be an edge containing the vertex *v*. Then all the $\Delta(G)$ edges incident on v are mutually adjacent and hence form a maximum clique of order $\Delta(G)$ in *L*(*G*). Hence $\omega\big(L(G)\big) = \Delta(G)$. The result (ii) follows by Proposition 5 and Remark 1.

The following corollaries are immediate from the above proposition.

Corollary 8.1 *If G is regular then L*(*G*) *is clique regular*

Corollary 8.2 *If G is regular then complement of L*(*G*) *is well covered*

DISCUSSION AND CONCLUSION

A graph is regular if every vertex is of same degree. This class of graphs are well studied in literature. Here we have introduced and studied a new class of graphs called clique regular graphs. It is observed that every regular graph need not be clique regular and every clique regular graph need not be regular. The properties of clique regular graphs can be studied in depth as future work. The effect of regular cliques in *G* can be extended to Line graph, Block graph and clique graphs.

REFERENCES

Berge, C., (1962). *Theory of graphs and its applications*. Methuen, London: North Holland.

Bhat, R. S., (2007). *A study of strong (weak) domination and related parameters in graphs*. (Doctoral Dissertation). National Institute of Technology, Karnataka, Surathkal.

Bhat, R. S., Kamath, S. S., & Bhat, S. R., (2011). A bound on weak domination number using strong (weak) degree concepts in graphs. *Journal of International Academy of Physical Science*, *15*(11), 1- 15.

Bhat, R. S., Bhat, S., R., & Bhat, P. G., (2014). Sum of block degrees. In *Proceedings of National Conference on Graph Networks* (pp. 207-215). University of Kerala, Trivandrum.

Bhat, P. G., Bhat, R. S., & Bhat, S. R., (2013). Relationship between block domination parameters of a graph. *Discrete Mathematics Algorithms and Applications*, *5*(3), 181-191.

Bhat, S. R., & Bhat, P. G., (2011). Mixed block domination in graphs. *Journal of International Academy of Physical Science*, *15*(3), 345-357.

Choudum, S. A., Parthasarathy, K. R., & Ravindra, G. (1975). Line-clique cover number of a graph. *Proceedings of the Indian National Science Academy, 41*(3 Part A), 289-293.

Cockayane, E. J., & Hedetniemi, S. T., (1977). Towards theory of domination in graphs. *Networks*, *7*(3), 247-261.

Dean, N., & Zeto, J. (1994). Well covered graphs and extendibility. *Discrete Mathematics, 126*(1), 67-80.

Harary, F. (1969). *Graph Theory*. London: Addison Wisley.

Haynes, T. W., Hedetniemi, S. T., & Slater, P. J. (1998). *Fundamentals of domination in graphs*. New York, NY: Marcel Dekker, Inc.

Kamath, S. S., & Bhat, R. S. (2006). Some new degree concepts in graphs. In *Proceedings of International Conference on Discrete Mathematics* (pp. 237-243).

Kamath, S. S., & Bhat, R. S. (2007). Strong (weak) independence and covering numbers of a graph. *Discrete Mathematics*, *307*(9), 1136 - 1145.

Ore, O. (1962). *Theory of Graphs*. United States of America, USA: American Mathematical Society Publications.

Plummer, M. (1970). Some covering concepts in graphs. *Journal of Combinatorial Theory*, *8*(1), 91-98.

Plummer, M. (1993). Well covered graphs. *Questiones Mathematica*, *16*(3), 253 - 287.

Ravindra, G. (1997). Well covered graphs. *Journal of Combinatorial Information System Sciences, 2*(1), 20-21.

Walikar, H. B., Acharya, B. D., & Sampathkumar, E. (1979). *Recent developments in theory of domination in graphs*. Allahabad, India: MRI Lecture notes, Mehta Research Institute.

Zelinka, B. (1981). On domatic numbers of graphs. *Mathematica Solvaca*, *31*(1), 91-95.

Pertanika J. Sci. & Technol. 25 (1): 271 - 282 (2017)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

A Review Article of Multi–Band, Multi–Mode Microstrip Filters for RF, WLAN, WiMAX, and Wireless Communication by Using Stepped Impedance Resonator (SIR)

Mohammed K. A.* and Ratna K. Z. S.

Center of Excellence of wireless and Photonics Network, Department of Computer and Communication system Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Filters are the basic part in wired, and wireless telecommunications and radar system circuits and they play an important role in determining the cost and performance of a system. The increasing demand for high performance in the fields of RF, WLAN, WiMAX and other wireless communications led to the great revolution in the advancement of the development of a compact microstrip resonator filter design. All these have made a vital contribution to both the required performance specifications for filters and other commercial requirements in terms of low cost, large storage capacity and high-speed performance. This review paper presents several design examples for multi-band, multi - mode microstrip filter resonators to satisfy RF, WLAN, WiMAX, UWB and other wireless communication frequency bands. To analyse the resonant frequencies odd - mode and even -modes can be used for the symmetrical structure. In general, the multi-mode resonators can be designed by using different methods like cross-coupling resonators Structure, and the allocation of the fundamental resonant frequencies of the resonator as stated by the Chebyshev's insertion loss function.

Keywords: Wireless telecommunication, WLAN, WiMAX, (SIR) filter, multi – mode resonators, UWB (Ultra-Wide Band), cross-coupling resonators

INTRODUCTION

A network that provides full transmission of electrical and communication signals with the specific passband region and unlimited attenuation at the stopband regions is called a filter. Filters are main components in an electronic and communication circuits,

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
mkks67@yahoo.com (Mohammed K. A.),
ratna@upm.edu.my (Ratna K. Z. S.)
*Corresponding Author

ISSN: 0128-7680

especially microstrip filters which are used in the RF/Microwave applications because of their simple structure, ease to fabricate by using simple printed circuit technology, low loss, low cost, and light weight. For reducing the size of the bandpass filter (BPF), a big value substrate of (ε_r) should be utilized in spite of it is expensive. To obtain a reduced size (Yoon et al., 2013) and a minor cost, certain types of microstrip resonators can be used for example a U - shape resonator, (SIR) (Stepped Impedance Resonator) (Mahyuddin, Ain, Hassan, & Singh, 2006), and stubs elements, have introduced to design various types of bandpass filters (Cassivi & Wu, 2003)...etc. Surrounded by these public filters, a bandpass filter (BPF) with half-wavelength open stubs appear very wide bandwidth. To achieve a narrow bandwidth (<10%) characteristic of a stub (BPF) design, the stub impedances would be excessively low and the stubs can be very complicated to realize. A narrow bandwidth half - wavelength open stubs bandpass filter can be achieved, it has introduced via fine-tuning the place of stubs by utilizing the (Q_e) where (Q_e) is the external quality factor without any change in the stub impedance (Zhang & Chen, 2006). For further reduction in the horizontal transmission line a narrow bandpass filter with a half – wave length open stubs are proposed. Stepped impedance resonator in many cases can be utilized instead of the transmission line to make a reduction in the physical dimensions of the bandpass filter on the dielectric substrate with low (ε_r). This paper offers an easy conventional dual – mode resonator such as a hairpin single – mode resonators. The shape of the resonator degenerates condition with respect to the multi - mode loop microstrip resonator design has rapidly stated in (Tyurnev & Serzhantov, 2011). The split microstrip resonators can be used to design the compact, high selective narrow (BPF). The edge of the O-ring resonator is split by making the O-ring into an open –loop microstrip resonator with the similar resonant frequency (Tyurnev, 2010). The size of the open - loop resonator can be more minimized into a quarter-wavelength to produce a dual - mode resonator by inserting a grounding at the midpoint of the microstrip open – loop resonator (N.-W. Chen & Fang, 2007). The benefit of the quarter-wavelength microstrip resonator is that the even harmonics of the resonator are repressed. The physical dimensions of the quarter–wavelength microstrip resonator has reduced by using a stepped impedance line (Dovbysh & Tyurnev, 2010). A centrally loaded triple – mode resonator (Kunlim, Pinghui, Zhiyuan, & Lin, 2011) is modified to design a third order and a sixth order bandpass filter. Microstrip dual-band bandpass filter represents an important part in an RF/Microwave transceivers for the development of the applications of wireless communication which, working in multi-band, exclusively the recent advanced Wireless Local Area Networks (WLANs) standards like IEEE 802.11b/g (2.4 GHz) and IEEE 802.11a (5.2/5.7 GHz) specifications (Xiao & Huang, 2010). One of the main tasks for the layout of dual-band bandpass filter is to obtain high selectivity in the passband, the small size, in addition, the easy design technique at the same time. In previous, the approaches of many designs have introduced for dual - band bandpass filters [(Xiao & Huang, 2010), (Chu & Chen, 2008)]. The stepped impedance resonators are the best technique has used to transfer the spurious frequencies of the Stepped Impedance Resonators (SIRs) to generate one more passband [(Xiao & Huang, 2010), (Chu & Chen, 2008)]. While there is a difficulty of controlling more than one passband separately by using the (SIR), so the dual passband responses are generated by the two resonator responses at the same time. Lately, we have described a dual - band (BPF) with broad band and narrow band simultaneously (Guo, Yu, & Zhang, 2010).

Conversely, the design technique is complex, meanwhile, the multi-layered technology has been used to invent the four (SIRs). A reduced size of a planar dual – band bandpass filter exploitation a dual structure feeding and embedded resonators have described in (C.-Y. Chen, Hsu, & Chuang, 2006). To realize a dual – band bandpass filter the cascading (SIR) or Stub Loaded Resonator (SLR) which is used, and two resonators, at least, are essential (Xiao & Huang, 2010) - (C.-Y. Chen et al., 2006); Last, for a compact circuit size, the dual-mode resonator develops as, significant applicant and is organized for dual-band application, which needs, at least, one or two resonators. A design process of a triple - mode quasi-planar filter has described in (Kunlim et al., 2011). It was synthesized by inserting a dual – mode microstrip square loop resonator inside a cavity that provides the third resonance (Guo et al., 2010). An incidental metal vase is used to realise the walls of the cavity that joins the top and bottom ground planes. The construction of such a filter is, still, complex and expensive.

SINGLE-BAND, MULTI-MODE MICROSTRIP BANDPASS FILTERS (BPFS)

Stepped Impedance Resonator Open Stub Bandpass Filter for Size Reduction

A half-wavelength narrow band bandpass filter using open-stubs have presented here for more decrease the horizontal transmission line (Yoon et al., 2013)-(Mahyuddin et al., 2006); (C.-Y. Chen et al., 2006). Stepped Impedance Resonator (SIR) can be used instead of transmission line to reduce the size of the bandpass filter on a low substrate. From the research outcomes, the new layout of narrow band (BPF) has reduced in size by a half-length as related to the traditional narrow band (BPF) using the transmission line.

Basic structure of (SIR). (SIR) can be formed of 2 - segments of low impedance at both sides and high impedance in the centre of the structure as illustrated in Figure1(Yoon et al., 2013).

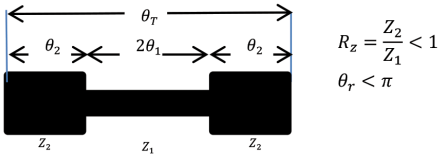


Figure 1. The basic unit structure of the Stepped Impedance Resonator (SIR).

An open-end input admittance Y_{in} is given as (Tyurnev, 2010):

$$Y_i = jY_2 \frac{2(R_z tan\theta_1 + tan\theta_2)(R_z - tan\theta_1 tan\theta_2)}{R_z(1-tan^2\theta_1)(1-tan^2\theta_2) - 2(1-R_z^2 tan\theta_1 tan\theta_2)} \quad (1)$$

The conditions of the resonant are $Y_i = 0$ and $R_z = \frac{Z_2}{Z_1} = tan\theta_1 tan\theta_2$, where $\theta = \theta_1 = \theta_2$ Eq. (1) Can be simplified as:

$$Y_i = jY_2 \frac{2(R_z+1)(R_z - tan^2\theta)tan\theta}{R_z - 2(1+R_z+R_z^2)tan\theta} \quad (2)$$

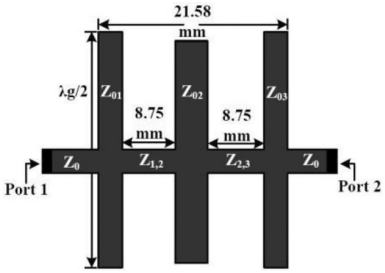


*Figure 2***.** The basic structure of the narrow bandwidth (BPF) with $\left(\frac{\lambda_g}{2}\right)$ open stubs.

The resonant condition is achieved by:

$$\theta = \theta_\circ = tan^{-1}\sqrt{R_z} \tag{3}$$

The stepped Impedance Resonator (SIR) of the open – circuited half – wavelength $(\frac{1}{2}\lambda_g)$ is estimated to act as a parallel resonant circuit in terms of magnitude of its input impedance $|Z_{in}|$ (Yoon et al., 2013).

Half – Wavelength Open Stub Narrow Bandwidth Bandpass Filter. A bandpass filter (BPF) comprises of shunt half – wavelength open stubs and series quarter wavelength transmission lines (Yoon et al., 2013).Generally, if a bandpass filter with stubs is intended to obtain ($< 10\%$) narrow bandwidth characteristic the impedances of the stubs will be excessively low and the stubs can barely be realized (Tyurnev, 2010). So, this type of bandpass filters maybe an applicant in wideband (BPF) applications (Yoon et al., 2013). The (BPF) proposed in this section, although, has resilience in bandwidth by adjusting the stubs location utilizing the external quality factor (Q_e) and the impedance of the stubs is indicated in Figure 3. Interdigital, combine, and hairpin filters are usually designed to tapped type, in which the location of tapping, (i_t) is the key factor in determining the bandwidth of filters (Mahyuddin et al., 2006). The external quality factor, (Q_e) maybe achieved by the phase response of the stubs ($\omega_\circ / \Delta\omega \mp 90$) with deviation of the position ($\boldsymbol{l_t}$) of the open stubs (Yoon et al., 2013) – (Mahyuddin et al., 2006). We can consider (Q_e) as a function of the position of stubs.

Stepped Impedance Resonator (SIR)With New Open Stub (BPF). The introduced narrow bandwidth bandpass filter with half- wavelength open stubs has been decreased in size by half (0.5) as compared with the traditional narrow bandwidth bandpass filter in the horizontal axis as it is seen in Figure 3.

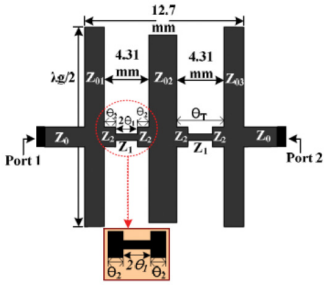


Figure3. size - reduction of the narrow bandwidth (BPF) using (SIR).

Compact Narrow Bandpass Filters Using Dual –Mode Split Microstrip Resonators

Microstrip bandpass filters (BPFs) having dual - mode resonators vary by high selectivity because every dual – mode resonator is identical to a couple of coupled single - mode resonators, i.e., the filter order describing selectivity is a double for a dual – mode filter than for a single – mode filter when they have the same number of resonators. Loop dual – mode microstrip resonators are most favoured (Tyurnev & Serzhantov, 2011)-(Dovbysh & Tyurnev, 2010). Furthermore, they are huge because their width is measurable with the length. Longitudinal and folded stepped – impedance microstrip resonators are better compact (Tyurnev & Serzhantov, 2011). However, the step -impedance technique of the resonant frequency convergence only is not adequate to realise a narrow band dual - mode bandpass filter. In this situation, the impedance step would be quite large and not passable. We propose in this section a plain dual - mode longitudinal microstrip resonator like a single – mode hairpin microstrip resonator. The resonator design is a degenerated state for the design of a multi – mode loop microstrip resonator whose specifications have obviously stated in (N.-W. Chen & Fang, 2007). The split microstrip resonator is suitable for the design of compact, high selectivity narrow - band filters. As an example of one resonator and two – resonator filters are presented.

Dual-Mode Resonator. The dual - mode split microstrip resonator strip conductor has a take shape of a arrow rectangular that is partly being divided by a longitudinal slot at one of its ends. The design and the equivalent circuit of the resonator are illustrated in Figure 4.

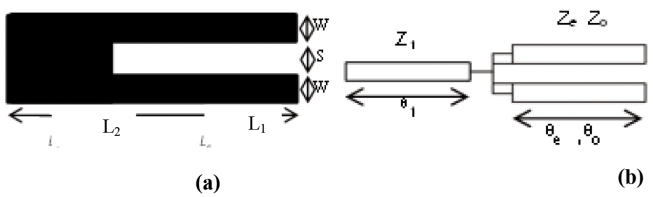


Figure 4. The dual – mode split microstrip resonator (a) the basic unit design (b) the equivalent circuit.

The even oscillation modes of the resonant frequencies are the roots of (4):

$$Z_e tan\theta_1 + 2Z_1 tan\theta_e = 0 \quad (4)$$

The odd modes of the resonant frequencies are the roots of the equation:

$$cos\theta = 0^{\circ} \quad (5)$$

The currents that pass through the divided conductors run in the same way for even – mode and in the opposite way for the odd - modes. It seems from (5) that the frequencies of odd - modes do not depend on (l_1). From (4) and (5) that the ratio f_e/f_o for the lowest resonant frequencies of the even and the odd - modes are associated with the length ratio l_1/l_2 by the equation:

$$Z_e \tan\left(\frac{\pi}{2}\frac{f_e}{f_\circ}\sqrt{\frac{\varepsilon_1 l_1}{\varepsilon_\circ l_2}}\right) + 2Z_1 tan\left(\frac{\pi}{2}\frac{f_e}{f_\circ}\sqrt{\frac{\varepsilon_e}{\varepsilon_\circ}}\right) \quad (6)$$

The coupling coefficient of a symmetrical pair of coupled single-mode resonators (k):

$$k = (f_\circ^2 - f_e^2)/(f_\circ^2 + f_e^2) \quad (7)$$

Two resonator filter. The filter of the fourth order has two dual-mode resonators. It means that its frequency response has four reflection minimums and three maximums in the passband. The design of the filter is illustrated in Figure 5. The adjustable structural parameters and its notation are given also.

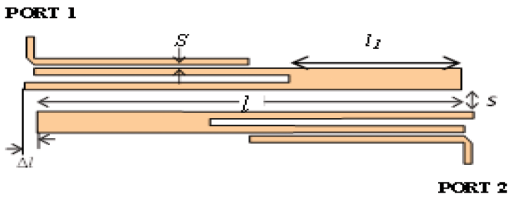


*Figure 5***.** The dual – mode dual - resonator filter design.

The instructions of intelligence, optimization have been used in designing the filter (Dovbysh & Tyurnev, 2010). The tuning of the center frequency fee was used to adjustment of the total resonator length L_t. The symmetrical dual – mode resonator in the two resonator filter is originally not tuned at the specific frequency because its upper and down portions are differently influenced by their edges.

Triple-Mode Compact Microstrip Bandpass Filter resonator Based on Open Stub-Load

A half – wavelength transmission line resonator with a pair of tapping open-ended stubs has used to develop a compact microstrip filter in (Kunlim et al., 2011). In order to repress the first spurious passband a compact microstrip filter using triple - mode stepped impedance resonator (TSIR) is used for this purpose (Xiao & Huang, 2010). A miniature dual – mode microstrip bandpass filter using compact dual – mode hexagonal open – loop resonators with an E – shaped open stub loading is proposed, with the size reduction of about 55% compared with the conventional dual – mode microstrip hexagonal loop filter. The comparison between the Traditional dual – mode microstrip hexagonal loop filter and a miniature dual – mode micro strip bandpass filter using compact dual – mode hexagonal open – loop resonators with an E – shaped open stub loading it is clear that the second type has a size reduction of about 55% (Xiao & Huang, 2010). The resonance specifications of triple-mode open stub – loaded resonator, particularly the main factors which impact the resonance frequencies and the transmission (zeros) positions have come to the fore. In order to achieve compact size, the two open stubs –loaded microstrip lines are folded into E – and T - type, then the method of extracting the coupling structure of a given triple – mode resonator filter is presented.

DUAL-BAND, MULTI-MODE MICROSTRIP BANDPASS FILTERS (BPFS)

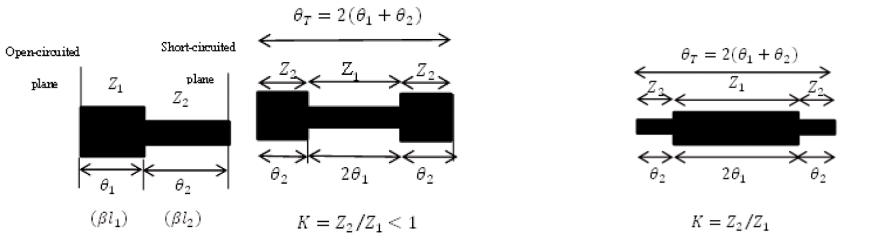
Dual-Band Compact Stepped Impedance Resonator (SIR) Bandpass Filter Structure

RF/Microwave filters with dual-band operation [(Xiao & Huang, 2010) - (Guo et al., 2010)] are drawing a lot of attention because the significant increase in desire of wireless communication applications requires RF/Microwave tranceivers operating in dualor multi – bands and therefore, the users can approach multiple services slightly with less effort [(Chu & Chen, 2008) -(C.-Y. Chen et al., 2006)]. Stepped Impedance Resonator (SIR) (Xiao & Huang, 2010) –(Guo et al., 2010) was proposed in previous years to replace of conventional half-wavelength Microstrip Parallel- Coupled Resonator (MPCR) because bandpass filters accomplished by (MPCR) have narrow stopband between the fundamental response and the first spurious response, and compared to the conventional (MPCR), stepped impedance resonator not only prevent the spurious responses, but also abridges the resonator dimensions. Dual – band and tri – bandpass filters can be designed by using stepped impedance resonator by tuning the higher order resonant modes. Reference (C.-Y. Chen et al., 2006) stated a bandpass filter using (SIR) with two-path coupling, but, only a single band was obtained. In this section, dual-band bandpass filter is achieved with split ring (SIR) and Defected Ground Structure (DGS) by using two-path coupling. Defected ground structure (DGS) has designed by engraving a defected shape on the metallic ground plane, and this shape increases the efficacious capacitance and inductance of microstrip bar, and as a result, (DGS) prevent the spurious responses by rejecting the harmonics in microwave circuits, and the performance of the filters or other microwave devices are efficiently improved. (DGS) and two-path coupling dual-band (SIR) bandpass filters are proposed in this section, and top performances of transmission zeros, low passband insertion losses, preferred dual- band other than compact structures and size reduction is realized compared with (C.-Y. Chen et al., 2006).

Resonant Characteristics Of Microstrip Stepped-Impedance Resonator (SIR).

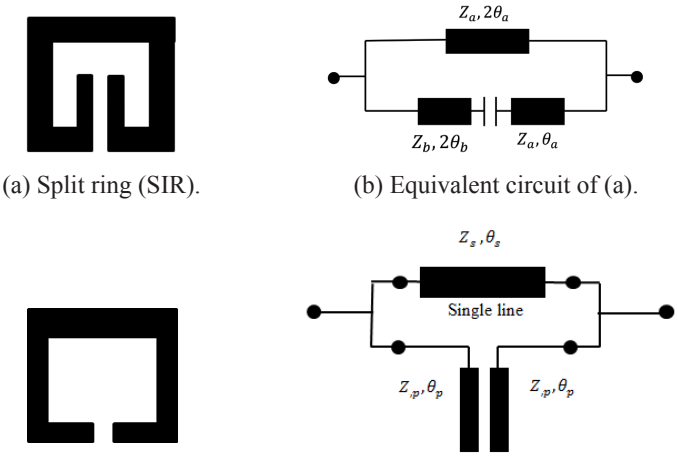
Connecting two microstrip transmission lines with different characteristic impedance Z_1 and Z_2 together have formed microstrip Stepped Impedance Resonator (SIR) unit as shown in Figure 6-a (the identical characteristic admittances are $Y_1$and Y_2), and the identical electric lengths are θ_1 and θ_2, respectively. l_1 and l_2 are physical lengths according to electric length θ_1 and θ_2, respectively. The input impedance Z_i, and the input admittance Y_i. If the intermittent of microstrip step and edge capacitance of open - circuit port are neglected, Z_i has expressed as (Xiao & Huang, 2010):

$$Z_i = jZ_2 \frac{Z_1 tan\theta_1 + Z_2 tan\theta_2}{Z_2 - Z_1 tan\theta_1 tan\theta_2} \quad (8)$$



(a) Unit (SIR) (b) Half-wavelength (SIRs)

Figure 6. Unit SI and traditional half - wavelength (SIR) structures.



(a) Split ring (SIR). (b) Equivalent circuit of (a).

(c) Split ring (SIR) with inner coupling. (d) Equivalent circuit of (c).

Figure 7. Split ring (SIR) used in the design.

The condition of the parallel resonant has obtained on the rule of:
$Y_i = 0$ as $Z_2 - Z_1 \tan\theta_1 \tan\theta_2 = 0$, and, it has been written as:

$$K = \tan\theta_1 \tan\theta_2 = Z_2/Z_1 \tag{9}$$

Where, K is impedance ratio. Figure 6-b shows half-wavelength (SIRs), and their input impedances can be explicit as:

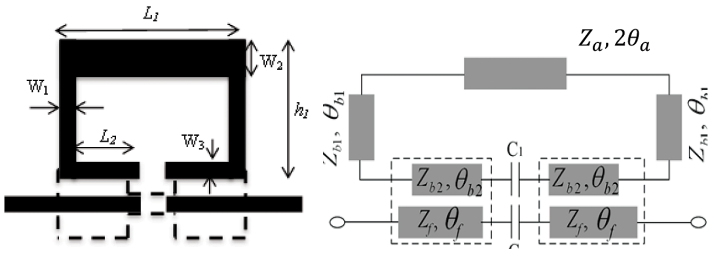
$$Z_{in} = jZ_2 \frac{2(1+K^2)tan\theta_1 tan\theta_2 - K(1-tan^2\theta_2)(1-tan^2\theta_1)}{2(K-tan\theta_1 tan\theta_2)(tan\theta_2 + Ktan\theta_1)} \tag{10}$$

The condition of the resonant achieved by $Y_i = 0$ as:

$$K = \tan\theta_1 \tan\theta_2 \tag{11}$$

Split ring (SIRs) is shown in Figures 7-a and 7-c, and their equivalent transmission line models are illustrated in Figures 7-b and 7-d respectively. Split ring (SIR) with inner coupling has formed by using a ring transmission line and end open-circuited parallel coupled lines, and when $\theta_{pe} = \theta_{po} = \theta_p$.

Dual-Band Bandpass Filters with Compact Coupling by Using Split Ring (SIRs). If the first passband has implemented by the first coupling path, and the second passband has implemented by the second coupling path, a required dual-band filter has presented by the two-path coupling. In this section, split ring stepped impedance resonators as illustrated in Figure 7 are selected for coupling in various path. In order to explain the principle of two-path coupling, hairpin (SIR) bandpass filters with 1-path coupling as illustrated in Figures 8, 11 are designed, and Figure 8-b is the equivalent transmission line sample of filter model.



(a) Topology of filter model I. (b) Equivalent transmission line model.

Figure 8. A dual - band (SIR) bandpass filter using one – path coupling.

Design of a Dual–Band, Dual - Mode Bandpass Filter Using Stepped Impedance Resonators

In this section, we present a dual – band, dual – bandpass filter by using (SIRs) with novel coupling structure. It shows that with the new coupling structure the filter has decent passbands specifications. Three transmission zeros have inserted by two stubs added in input/output lines to improve the selectivity.

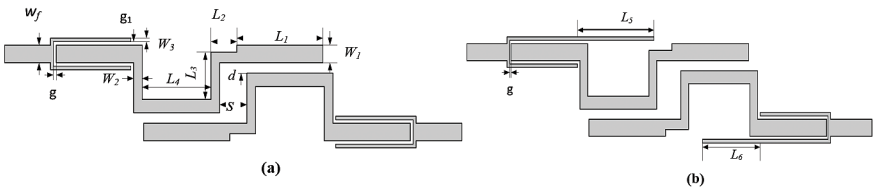


Figure 9. (a) A filter has1 -transmission zero, 9 (b) A filter have4 -transmission zeros.

CONCLUSION

In this paper, multi-band, multi-mode an open stub (BPF) with (SIR) has been introduced. The (SIR) can be utilized alternatively of transmission line which pose as an inverter in a traditional open stub (BPF) to minimize the physical dimensions of the (BPF) on the low dielectric substrate. The (BPF's) inverter had a half – length compared with one of the traditional narrow (BPF). A simple compact dual – mode microstrip resonator having a longitudinal slot

in the one end of its conventional strip conductor is presented. The resonator is perfect for compact narrow band bandpass filters. Two - resonator and three – resonator filters are well - considered. They have transmission zeros in stopbands. The filters offer a high selectivity and small size compared to the filter designs using single – mode resonators. A detailed design steps for a compact triple-mode resonator filter using open stub-loaded has presented in this paper. The design method is simple and the adjustment is effective and quick. In this article, new dual-band bandpass filters with compact structures by using two-path coupling are proposed, and split ring (SIR) and (DGS) are applied not only to reduce the circuit sizes, but also to improve the filter performances of frequency selectivity, harmonic suppression and wide stopband neighbouring the operation passband. A novel dual-mode dual-band filter with (SIRs) is proposed. The newly proposed filter has good passbands performance and is easy to control bandwidths. With introduced two stubs, another three transmissions zeros are created, and the selectivity has further improved. A novel microstrip tri- band bandpass filter has proposed and designed based on stub-loaded triple-mode resonator.

REFERENCES

Cassivi, Y., & Wu, K. (2003). Low cost microwave oscillator using substrate integrated waveguide cavity. *IEEE Microwave and Wireless Components Letters, 13*(2), 48-50.

Chen, C. Y., Hsu, C. Y., & Chuang, H. R. (2006). Design of miniature planar dual-band filter using dual-feeding structures and embedded resonators. *IEEE Microwave and Wireless Components Letters, 16*(12), 669-671.

Chen, N. W., & Fang, K. Z. (2007). An ultra-broadband coplanar-waveguide bandpass filter with sharp skirt selectivity. *IEEE microwave and wireless components letters, 17*(2), 124-126.

Chu, Q. X., & Chen, F. C. (2008). A compact dual-band bandpass filter using meandering stepped impedance resonators. *IEEE Microwave and Wireless Components Letters, 18*(5), 320-322.

Dovbysh, I. A., & Tyurnev, V. V. (2010). Synthesis and investigation of three-section microstrip filter on folded dual-mode stepped-impedance resonators. *Progress in Electromagnetics Research M, 12*, 17-28.

Guo, L., Yu, Z. Y., & Zhang, L. (2010). Design of a dual-mode dual-band filter using stepped impedance resonators. P*rogress in Electromagnetics Research Letters, 14*, 147-154.

Hong, J. S., Shaman, H., & Chun, Y. H. (2007). Dual-mode microstrip open-loop resonators and filters. *IEEE Transactions on Microwave Theory and Techniques, 55*(8), 1764-1770.

Kunlim, C., Pinghui, L., Zhiyuan, Z., & Lin, F. (2011, May). A novel compact microstrip bandpass filter based on open stub-loaded triple-mode resonator. In *Microwave Technology & Computational Electromagnetics (ICMTCE), 2011 IEEE International Conference on* (pp. 118-121). IEEE.

Mahyuddin, N. M., Ain, M. F., Hassan, S. I. S., & Singh, M. (2006, September). A 10GHz PHEMT dielectric resonator oscillator. In *2006 International RF and Microwave Conference* (pp. 26-30). IEEE.

Shen, W., Sun, X. W., & Yin, W. Y. (2009). A novel microstrip filter using three-mode stepped impedance resonator (TSIR). *IEEE Microwave and Wireless Components Letters, 19*(12), 774.

Tyurnev, V. V. (2010). Coupling coefficients of resonators in microwave filter theory. *Progress in Electromagnetics Research B, 21*, 47-67.

Tyurnev, V. V., & Serzhantov, A. M. (2011). Dual-mode split microstrip resonator for compact narrowband bandpass filters. *Progress in Electromagnetics Research C, 23*, 151-160.

Wu, H. W., Liu, S. K., Weng, M. H., & Hung, C. H. (2010). Compact microstrip bandpass filter with multispurious suppression. *Progress in Electromagnetics Research, 107*, 21-30.

Xiao, J. K., & Huang, H. F. (2010). New dual-band bandpass filter with compact SIR structure. *Progress in Electromagnetics Research Letters, 18*, 125-134.

Yoon, K. C., Lee, H., Lee, J. G., Park, J. G., Kim, T. Y., Jung, T. S., & Lee, J. C. (2013). Open Stub Band-pass Filter Using Stepped Impedance Resonator for Size Reduction. *Session 2PK*, 873.

Zhang, H., & Chen, K. J. (2006). Miniaturized coplanar waveguide bandpass filters using multisection stepped-impedance resonators. *IEEE transactions on microwave theory and techniques, 54*(3), 1090-1095.

Pertanika J. Sci. & Technol. 25 (1): 283 - 292 (2017)

SCIENCE & TECHNOLOGY

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

Wireless Structural Health Monitoring (SHM) system for damage detection using ultrasonic guided waveform response

A. H. S. Shirazi, M. K. H. Muda*, M. J. Thirumurthy and F. Mustapha

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

ABSTRACT

This paper presents an improved version of a wireless device embedded with a smart PZT sensor to detect flaws and structural defects on selected investigated structure. Smart PZT sensors were used as an actuator and sensor, coupled with two XBee's and one signal generator IC chip. Programme execution on transmitting and receiving the ultrasonic guided wave via the PZT sensor had been written in MATLAB. The developed source code is basically to receive serial data from one Xbee to another remote Xbee attached to the investigated structural system. The refined waveform response is utilised for prognosis of the true structural status. The 4-mm simulated holed into one of the aluminium structural plate is benchmarked with its pristine condition in validating the effectiveness of the developed SHM wireless module. Results showed that the wave is more even in non-defected area and disrupted in affected area. Ultrasonic waves increase continuously for non-destructive evaluation and structural health monitoring in various structural applications because the guided wave can propagate long distances and reach difficult-to-access regions; for inspecting porous and some non-porous materials ultrasonic waves attenuate fast and are very useful. Recent advances in ultrasonic wave application model and results are discussed in this paper.

Keywords: microcontroller, piezoelectric sensor, sine wave signal, smart sensor, structural health monitoring, wireless sensor

INTRODUCTION

In the field of aeronautics, reliability is important and Non-destructive test (NDT) and Structural health monitoring (SHM) concepts are used to detect structural flaws. However, SHM is different from NDT whereby the former is used to monitor the integrity of mechanical structures in a continuous and independent way. Thus, SHM helps to reduce maintenance costs.

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
hafyz801@gmail.com (A. H. S. Shirazi),
ahmad2241965@hotmail.com (M. K. H. Muda),
tkmraj@yahoo.co.in (M. J. Thirumurthy),
faizal@eng.upm.edu.my (F. Mustapha)
*Corresponding Author

ISSN: 0128-7680

Piezoelectric transducers or sensors have the capability to transform mechanical energy to electrical energy and vice versa. These sensors have been used in many applications.

The SHM involves integration of sensors, data transmission, smart materials, computational power, and handling out ability inside the structures (Chung et al., 2014). This is schematically presented in Figure 1.

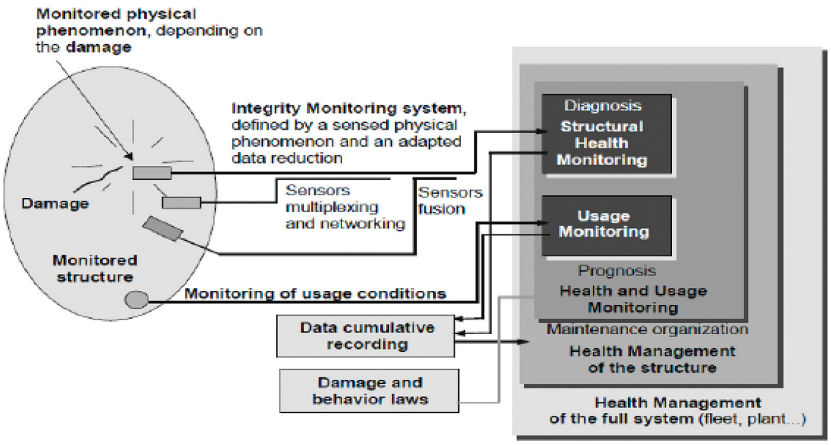


Figure 1. Organisation of SHM system (Chung et al., 2014)

Structural health monitoring has received much attention the last three decades. A brief literature review of this technology's development is presented below. Doebling et al., 1996, Inman (2003), Hoon et al. (2004) and Farrar et al. (2007) discuss extensively identification of damage and technical challenges inherent in the SHM adaptation which are common in all technology applications. Rytter proposes a categorisation system for SHM damage identification method (Mustapha et al., 2005) which had 4 levels originally and later expanded by Inman (2003) to include three more levels.

The other modal parameter technique is the mode shape method to detect damage in a structure. The MAC (Modal Assurance Criterion) is used to correlate modes, proposed by West (1984) for the first time, and is a method to locate and detect damage in a structure without using a prior FEM (Randall, 2004). The main weakness of this technique is the enormous number of sensors required to gather sufficient data.

Many SHM approaches have been used over the last decades. In many cases, there is no easy way for measuring the inputs or the external the structure. Wireless sensor can be utilised to locate, detect and assess structural damage produced by severe loading events and by progressive environmental deterioration and the economical realisation of SHM system. Cho et al. (2008) show the outcomes of cooperative international research on the smart wireless sensor. Heo and Jeon (2009) developed a monitoring smart system based on global computing method for infrastructure system. The system is designed in order to allow the usage of TCP/IP net procedure, connecting data measured by a sensing wireless unit found on Bluetooth technology. Chung (2014) investigated and developed MEMS type wireless sensor for real time seismic bridge motoring.

The wireless competence added to the established MEMS sensors offers the possibility prevent long multiple cables usage for bridge monitoring. This paper presents the use of PZT sensor and XBee module to connect remote structure and PC to detect faults in structural portions.

Sensors have many other usages and a huge potential for monitoring structures. Sensing modules consist of different types of sensors used to measure different parameters. Generally, systems can be identified by the following modules such as sensing, acquisition, wireless and output. Sensors sense changes to the environment signals before sending the signals to the acquisition module. This includes vibration analysis, damage detection and power consumption. The Acquisition module consists of controller, signal conditioning unit and analog to digital controller. In general, signal conditioning unit performs three major operations called filtering, amplifying, and isolation.

The next process is transferring data from sensing unit to server unit. A pair of Xbee's are used for this purpose, namely to detect damage. Received data can be sent to PC using XCUI software as interface for further analysis. The MATLAB is used to receive the signal wave forms. The diagram 2 below shows how the system operates .

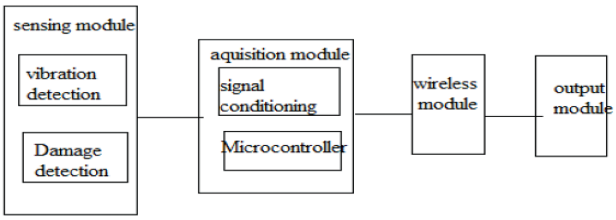


Figure 2. Modules involved in wireless system (West, 1986)

BASIC ARCHITECTURE OF THE DEVELOPED WIRELESS SHM SYSTEM

The proposed model of module detection system is shown in Figure 3. This module is mainly used to predict structural damage before it is discovered through general inspections. The overall system is combination of hardware components and software. The hardware is divided into two parts: transmitter section and receiver section. Transmitter section is developed using peripheral interface controller usually denoted by PIC and an Xbee unit coupled with a sensor. This PIC works exactly like in an acquisition module. Hardware consists of four modules as stated earlier: sensors detection, acquisition data, wireless communication and receiving data. The system reads vibration in the form pulse width. The higher the pulse width, the higher the vibrations. The PIC counts the pulse and transfers it to the wireless module section. This is seen in Figure 3.

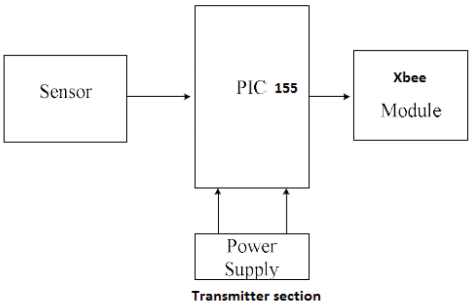


Figure 3. Detection system module (Hoon, 2004)

It is possible to use different types of accelerometer in SHM system and depends on the range of measurement. The proposed simple Interface diagram of Xbee is shown in Figure 4.

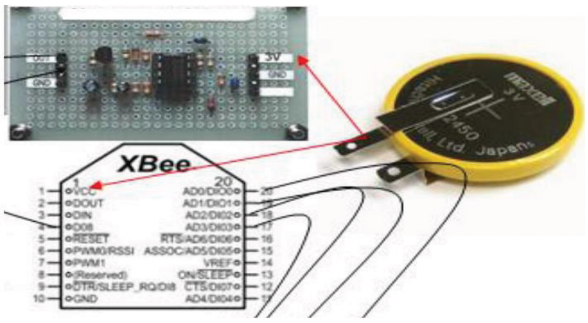


Figure 4. Interface diagram and connectivity of components

Figure 5 shows the block diagram of receiver section where the wireless Xbee module is connected to the PC and damages can be detected in the form of vibrations.

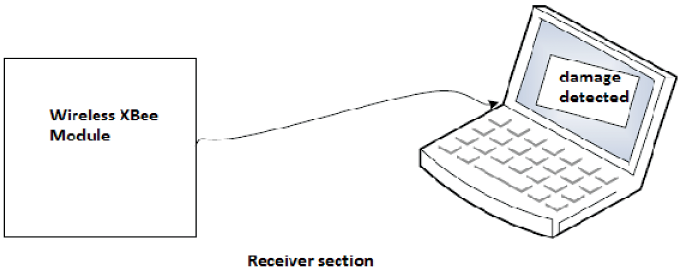


Figure 5. Receiver Section

HARDWARE

The device embeds ultrasonic Pieozo sensor and data acquisition electronics into the structure by using Xbee RF modules capabilities and radios for node communication and control. The PIC 155 has been chosen as transmitter and receiver sensor because it is categorised under soft piezo ceramic.

Defect detection, categorisation and localisation are prepared in three stages. Sensing (node stage), generation/acquisition, wireless communication stage, and base stage (Analysing). Sensing nodes are responsible for sine wave signal processing and collection and the second stage is to relay data to base station.

Sinewave Generator for piezo actuator excitation

Figure 6 shows the prototype of the sine wave board for piezo actuator. The standard amplifier is 741, with output 3V p-p and 1.5 offset. It can generate a frequency tunable between 100 Hz and 1.5 kHz. The power consumption of this circuit is less than 20 mW for 0.3% duty cycle.

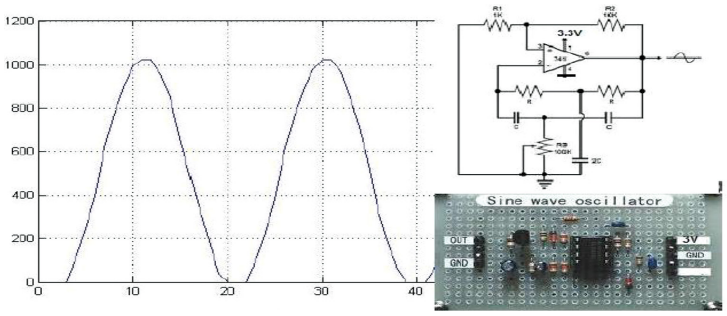


Figure 6. Sine wave generator

SYSTEM INTEGRATION

Wireless RF Modules Node Hardware

The Xbee is a wireless sensor network device which is low cost and low powered. The module operates within the ISM 2.4 GHz frequency band. Xbee radios have multiple I/O lines that can be used to collect digital or analog data and then transmit that to another Xbee for interpretation.

The analogue to digital converter (ADC) on Xbee radios are 10 bit, which will provide a resolution from 0 to 1023 (0x03FF hex). The selected XBee should not exceed 3.3 V on any pin to avoid damaging the radio module. The secondary microcontroller on this XBee is connected to the ADC pins instead of the RF processor as shown in Figure 7.

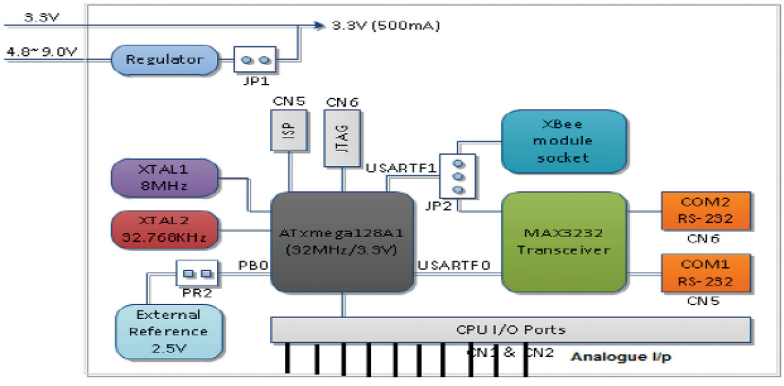


Figure 7. XBee diagram

SOFTWARE

X-CTU

The X-CTU is software used for configuring and testing XBee's data. It is like a frame work and configuration depends on the design. The software is interfaced with hardware and it is the lowest layer of software abstraction. Its role is to operate and control the unit's hardware to assist upper software layers in accomplishing their computational goals. The functionality envisioned indicates that the software could be complex and lengthy. The frame contains digital and analog mask indicating which I/O lines are configured as inputs.

EXPERIMENTAL PROCEDURES

The data acquisition method depends on several components such as selecting locations, sensors types, and the data processing that includes acquisition, storage, and data transmitting/receiving hardware. The economic consideration is quite important in choosing components and methods.

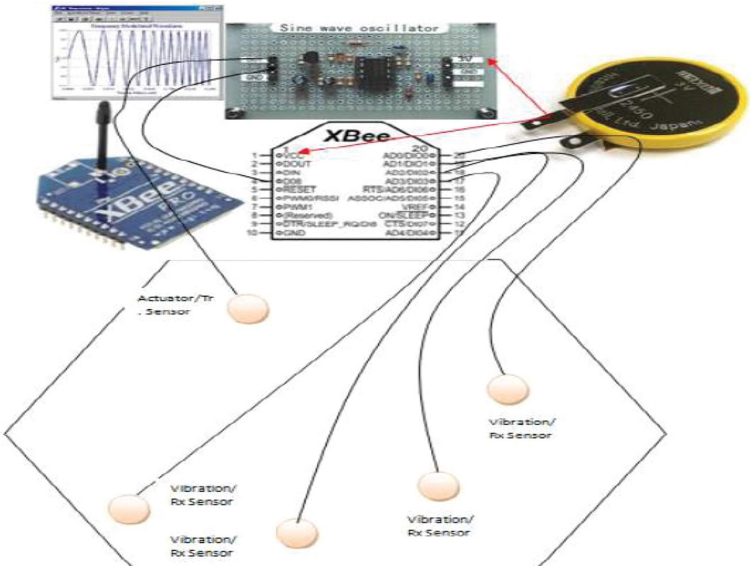


Figure 8. Single Xbee module

In this research, two PIC 155, two Xbee's and PZT sensors functioned as data acquisition components. Figure 8 shows a PCB circuit, model of single Xbee and interconnectivity of sensors in a single model space. The main objective of this experiment was to show wireless and remote communication with less power consumption and less complex compared with manual inspection. The next paragraph describes the connectivity of the basic components and basic setup of this experiment.

Two PZT sensors, PIC 155, are placed on the substrate and the distance between each sensor is 15 cm. One Xbee component is fixed on the substrate and coupled together with sensors and this Xbee is powered by a simple battery. The Xbee on substrate once powered

is able to send data wirelessly and remotely to another Xbee pair which is connected to the computer through USB port. Data is then transferred to MAT LAB for further investigations. The XCTU software was installed to connect two Xbees and gather data from the substrate. This is shown in Figure 9.

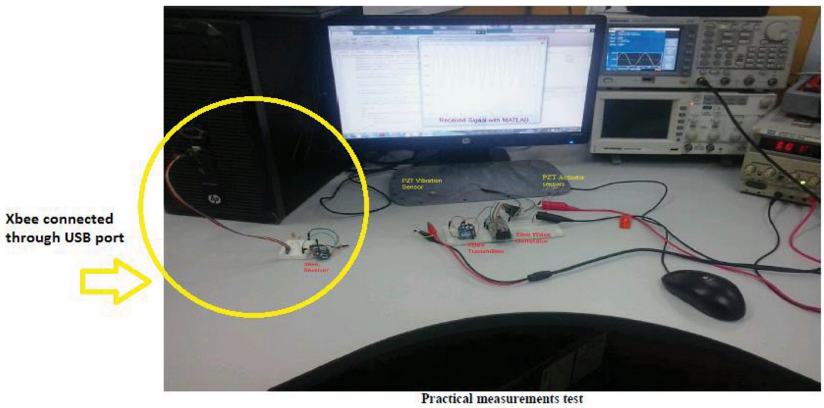


Figure 9. Connection wireless system and computer

RESULT ANALYSIS

The results of captured waveform are shown in Figure 10(a), 10(b) and 10(c). Figure 10 shows significant results of the investigated cases. There is no distortion or attenuation in Figure 10(a) and 10(b). Since the received data is similar to the PZT data, results prove that the same signals have been received with no defect or damage. This results obtained in signal amplitude value is 3V, which is equivalent to 1023 decimal code due the 10 bit ACD of the Xbee. Figure 10(c) shows significant variation in amplitude due because of the defective specimen. The signal was influenced by the damage and has distorted amplitude attenuation proportional with the position of the PZT sensor. It is found that the signal which travelled between the PZT actuator and the sensor has a value of $1.9V_{p.p}$, whereby the sensors are supposed to be distributed to cover particular area on the body for detection at any point.

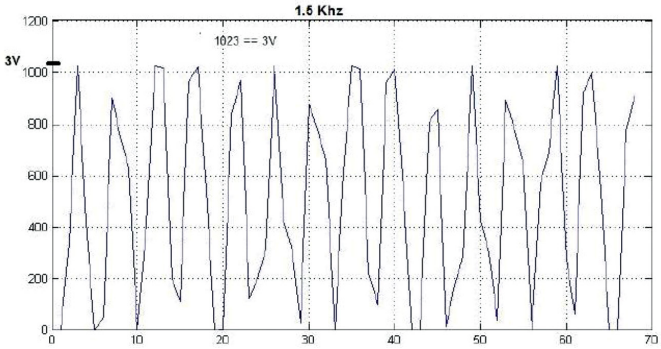


Figure 10(a). No defect plate, 1.6kHz

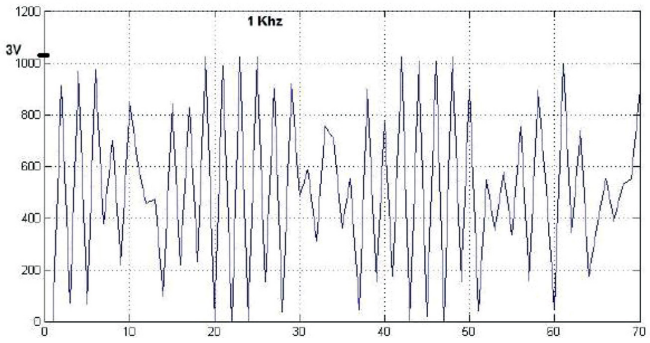


Figure 10(b). No defect plate, 1kHz

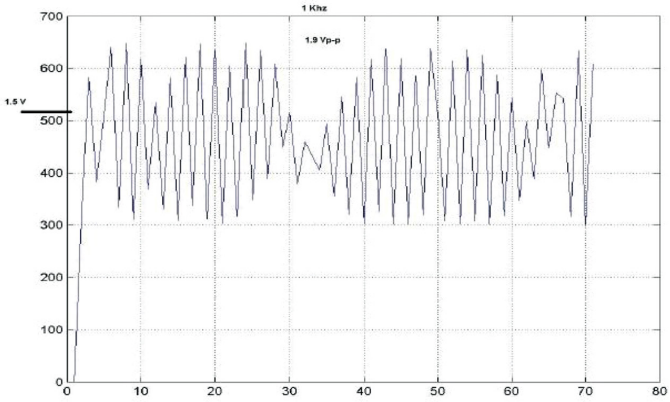


Figure 10(c). Defected plate, 1kHz

ADVANTAGES

Structural health monitoring system has a high potential to reduce cost dramatically in future. It helps us to minimise labour cost largely associated with manual inspections. In short, a wireless system reduces time cost which in turn increases profits. There is no necessity to replace components based on usage in all wireless module system. The main objective or focus of this research is to prevent or detect system damage before catastrophic failure occurs. Thus, the wireless system is promising compared with other conventional methods of detection failures. It does not lead to expensive repairs, and replacements or any periodical maintenance. This method eliminates sending humans into hazardous environments and help to extend the life of a structure.

DISADVANTAGES

In this study, a concept wireless monitoring system was developed from commercially available wireless sensing components. The monitoring system was applied to an aluminium frame structure with pre hole made in one corner. Despite the hardware limitations, the monitoring system successfully identified changes in the normal surface and affected surface.

The wireless sensor system requires steady power supply. The current power supply is limited because it is often difficult to obtain access to the system to replace the battery as well as difficulty in supplying continuous power to the system; rather, it depends purely on the characteristics of the structure where it's installed. However, in future, this could be rectified to enable large scale implementation.

CONCLUSION

This work shows the capabilities of the Xbee to achieve all required objectives of the SHM inspection with the use of low power. This device is less complex, has wireless serial communication and direct data processing and results. The resolution has been improved over the years. This is achieved by reducing conversion time when processing data. This device comes with minimal components and does not have additional microcontroller or Xbee to manage data. We investigated a particular application of wireless sensor networks. Based on the existing SHM system, the main challenges are identified first before summarising the corresponding technique. Future research should investigate further these challenges.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Putra Malaysia and MOE for providing research grant ERGS 5527089 to undertake this work.

REFERENCES

Chung H. C., Enomoto T., & Shinozouka, M. (2014). Real-time visualization of structural response with wireless MEMS sensors. In *13th World Conference on Earthquake Engineering*, Vancouver, B.C., Canada, Paper No. 121.

Doebling, S. W., Farrar, C. R., Prime, M. B., & Shevitz, D. W. (1996). *Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review*. Los Alamos National Lab., NM (United States).

Farrar, C. R. (2003). *Damage prognosis: current status and future needs*. Los Alamos National Laboratory, LA.

Farrar, C. R., & Lieven, N. A. (2007). Damage prognosis: the future of structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365*(1851), 623-632.

Heo, G., & Jeon, J. (2009). A smart monitoring system based on ubiquitous computing technique for infra-structural suspension bridge. *KSCE Journal of Civil Engineering, 13*(5), 333-337.

Hoon, S., Farrar, C. R., Hemez, F., & Czarnecki, J. (2004) *A review of structural health monitoring literature.* Los Alamos National Laboratory Los Alamos, NM.

Inman, D. J. (2003). Smart materials in damage detection and prognosis. *Key Engineering Materials, 245*, 3-18.

Mustapha, F., Manson, G., Pierce, S. G., & Worden, K. (2005). Structural health monitoring of an annular components using statistical approach. *International Journal of Strain Measurement, 41*(3), 117-127.

Randall, R. B. (2004). State of the art in monitoring rotating machinery-part 1. *Sound and vibration, 38*(3), 14-21.

Randall, R. B. (2004). State of the art in monitoring rotating machinery-part 2. *Sound and vibration, 38*(5), 10-17.

Rytter, A., & Kirkegaars, P. H. (1994). *Vibration based inspection of civil engineering structures*. Aalborg Universitetsforlag.

Soojin, C., Chung, B. Y., Jerome, P. L., Andrew, T. Z., Billie, F. S. J., & Tomonori, N. (2008). Smart wireless sensor technology for structural health monitoring of civil structures. *Steel structures, 8*(4), 267-275.

West, W. M. (1986, February). Illustration of the use of modal assurance criterion to detect structural changes in an orbiter test specimen. In *Proceedings of International Modal Analysis Conference* (Vol. 1, pp. 1-6). Los Angeles, CA.

Pertanika J. Sci. & Technol. 25 (1): 293 - 302 (2017)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Yawing Force of Electric Trimmers of a Hybrid Buoyant Aerial Vehicle

Haque, A. U.[1], Asrar, W.[1]*, Omar, A. A.[2], Sulaeman, E.[1] and Ali, J. S. M.[1]

[1] *Department of Mechanical Engineering, International Islamic University Malaysia, 50728 IIUM, Kuala Lumpur, Malaysia*

[2] *Department of Aeronautical Engineering, University Of Tripoli, 13154, Tripoli, Libya*

ABSTRACT

All buoyant and hybrid buoyant aerial vehicles have directional stability issues at low speed. Electric trimmers are one of the potential solutions for controlling the yaw motion of such vehicles in which partial lift is obtained from the wings. However, available propeller disk area of such trimmers is limited due to small surface area of the vertical tail. In the present work, maximum input power required by thin electric propellers with different pitch values are compared to obtain an optimised value of pitch for propeller selection. Analytical as well as computational techniques are employed to evaluate the moment generated by tangential thrust produced by a ducted propeller. *Motocalc*® software under predicts the thrust value when compared with the computational results under the same flow conditions. The estimated yaw force produced by the propeller is quite significant and it can also be used for creating differential thrust using twin electric motors.

Keywords: Advance Ratio, Computational Fluid Dynamics, Hybrid Buoyant Aerial Vehicle, Static Thrust, Turning, Thin Electric Propeller

INTRODUCTION

Airships are directionally unstable (Khoury & Gillett (1999; Carichner & Nicolai, 2013)) and little information is available on the stability of hybrid buoyant airships in yaw. In a recent experimental study by Andan, Asrar & Omar (2012) on a generic model of a hybrid airship, it was found that such a vehicle is marginally stable in yaw. Such response can cause poor manoeuvrability, especially at low speeds. The yaw/turning performance of a hybrid buoyant aerial vehicle can be enhanced by inserting an electric fan in the lower vertical tail or at the nose of the hull (Haque et al.,

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
anwar.haque@live.iium.edu.my (Haque, A. U.),
waqar@iium.edu.my (Asrar, W.),
aao@aerodept.edu.ly (Omar, A. A.),
esulaeman@iium.edu.my (Sulaeman, E.),
jaffar@iium.edu.my (Ali, J. S. M.)
*Corresponding Author

ISSN: 0128-7680

2014). This concept is perhaps similar to that used in traditional advertising blimps for circuit flight over a fixed location (Stockbridge et al., 2012). Direct current (*DC)* brushless motor generates the required thrust which can be manipulated as the yaw force when installed in a plane perpendicular to the incoming air.

For small blimps, the yaw is produced by using *DC* electric motors placed close to the gondola at the base of the hull. Such motors are usually ducted with the help of a metallic ring to meet safety requirements. The moment arm of such motors is quite small as the location of installation is close to the centre of gravity *(CG).* There have been many studies carried out in the area of dynamic stability of buoyant and hybrid buoyant aerial vehicles; however, information regarding the yawing response and yaw force generated by the electric thrusters is quite limited. In the present work, an analytical and computational effort is made to fill this gap. A generic model of a hybrid buoyant aerial vehicle *(HBAV)* designed by (Andan, et al*.,* 2012) is taken as a test case in which an electric propeller is used to obtain the performance parameters of an electric propeller. In order to generate the yaw force, electric trimmers are placed in a direction perpendicular to the free stream velocity.

Usually, three types of control surfaces are required to manoeuvre an aircraft i.e. *Elevators* to control the Pitch, *Rudders* to control the Yaw and *Ailerons* to control the Roll. In the case of airships as well, it is conventional to use rudders to control the yawing motion. However, rudders can be replaced by generating a differential thrust with the help of electric motors. In this way, if there is a yaw signal, the speed of the electric motors not only depends on the throttle signal but on both throttle and yaw signals. In this case, the speed of second motor set is decreased or increased compared with the first motor set, as per the requirement to either decrease or increase the yaw. This is discussed in detail in the section on result and discussion.

The term thrust is always used for the estimation of the propulsive force generated by a propeller in the forward flight, which is in fact the required yaw force to trim the vehicle as done by rudders in the case of aircraft (Schroijen, et al*.,* 2008). Similar to the helicopter rotor, the flow around such a propeller is heavily distorted and massive flow separation occurs at its tip. For hybrid buoyant aerial vehicles, such effects and losses can be minimised to some extent by placing the propeller blades inside the lower vertical tail of a+ *type* empennage arrangement. Also, similar to the wing of an aircraft, a propeller has its own airfoils, which can have an aerodynamic as well as a geometric twist. This airfoil creates pressure difference between its windward and leeward side. In order to estimate its pressure difference, the aerodynamic characteristics of the individual airfoil must be known.

In this paper, various propeller combinations are evaluated using a brushless *DC* motor as the propulsion system. The choice of a brushless DC motor is due to its better performance and comparatively high operating efficiencies compared with brushed *DC* motors. There are many different types of propellers, such as fixed-pitched, ground-adjustable, two-position, controllable, automatic, constant speed, feathering constant speed and reverse pitch; out of these, fixed pitch is selected as it does not allow for any change in blade angles and is usually available as a single piece. The available thrust is calculated using *MotoCalc*® software, which is an open source code usually used for R/C airplane's propulsion system sizing and analysis. It has a large database of electric motors, batteries and speed controllers (User's Manual, 2015).

A detailed analysis was carried out using analytical and computational techniques to evaluate their limitations for accurate prediction of thrust and hence the yawing moment produced by the electric trimmers. However, the propeller is numerically simulated for free stream flow conditions which are not the same as those of a ducted fan. These changes require corrections to the numerical data and necessary corrections can be made by performing wind tunnel tests on a generic model of such a propeller at different tangential velocities.

ANALYTICAL METHODS

A propeller looks like a simple rotating device, but the parameters representing its performance characteristics are quite complex which can analytically be represented in terms of variables by the thrust coefficient, C_T, the power coefficient, C_p, the efficiency coefficient, η and the advance ratio, J; out of which, C_T is used in the present work to estimate thrust of a thin electric propeller, (1). Whereas, J is defined as the ratio of velocity and product of n and D.

$$C_T = \frac{T}{\rho . n^2 . D^4} \tag{1}$$

In this equation, T, ρ, n and D represent the thrust, density, number of revolutions per minute and disc diameter, respectively. In order to estimate T, different propeller theories are available in open literature, some are discussed here briefly especially their pros and cons.

Momentum Theory

Rankine-Froude were among the first who proposed a theory for the prediction of propeller performance parameters (Madsen, 1996). As per this theory, "*the propeller is assumed to be working in ideal fluid and is able to absorb all the power from the engine and later on dissipate this power by causing a pressure jump in the fluid flowing through the propeller"* (Madsen, 1996). Unfortunately, by applying this theory, the actual performance of the propeller is still a question mark, as such kind of assumption does not take into account the actual geometric profile of the propeller and viscosity of the fluids. Hence, its usage for the estimation of maximum efficiency of the propeller is quite limited.

Blade Element Method

William Froude in 1878 further refined the Rankine-Froude's Momentum theory addressing its deficiencies (Friedrich & Robertson, 2014). He proposed the idea that if we divide the blade of the propeller in such a way that every element strip can be represented by its own width and chord, then torque/force produced by each such element can be integrated to define the aerodynamic properties of the blade of a thin propeller. The limitation of the linear blade angle is that it does not account for the effect of induced velocity. For example, in the case of a helicopter, the induced velocity due to rotor disk cannot be handled by Blade Element Method.

Combined Blade Element Momentum Theory:

Glauert (1926) who first propounded this theory of Combined Blade Element Momentum, which is the combination of above mentioned two theories in which the thrust generated by the individual annulus segments can be estimated by using (2):

$$dT = 4a'(1-a)\rho U_1 \Omega r^3 \pi\, dr \tag{2}$$

Where, radius r and width dr represent the local geometric parameters, U_1, a' and Ω are free stream velocity, angular induction factor and angular velocity respectively. It is well known that the vortices are rolled up on the tip of a blade and the tip chord of the wing. This is an unavoidable phenomenon due to which there is always a decrease in lift at the tip section and increase in drag, commonly known as induced drag. Unfortunately, this theory assumes that all the elements of the blades are lift producing and it does not cater for the effect of loss in lift in the tip region. Furthermore, there is no term to account for the swirl caused by the tangential forces applied by the rotor on the flow media, whether it is liquid or gas. Since this technique is already employed by *Motocalc*® software, therefore, without going into lengthy calculations, results obtained from this open source software are used for calculating thrust.

RESULTS AND DISCUSSION

The available disc area for the propellers of electric trimmers is shown in Figure 1. The length of the model of vehicle under discussion is 4.2 m and the maximum diameter of its hull is 1.6 m. Space available for trimmers on the vertical tail is quite limited and a propeller of less than one foot as maximum disc diameter can easily be fitted without any additional structural reinforcement. Complete details of the geometric and performance parameters of the configuration are defined by Farhana (2012). *Motocalc*® is used for estimation of power and performance parameters. The graph shown in Figure 2 indicates the input load power for varying combinations of propeller pitch and diameter. All these propellers have a diameter less than one foot and operate at low Reynolds numbers (ranging between 50,000 and 100,000) due to shorter chord length (Propeller, 2015). Figure 2 shows that for a single motor, an 8 inch prop consumes lowest power and hence gives lowest thrust. On the contrary, an 11 inch propeller requires highest power while 9 inch propeller gives a middle of the road solution. Therefore, by the end of analysis, the propeller choice finally converges to a 9 inch propeller. It is also obvious from Figure 2 that 9×6 inch propeller (having slightly higher pitch) also produces an increment in the input power and hence larger thrust compared to that produced by a 9×4 inch propeller. However, a *9×4 inch* propeller was not available in the market and a *9×4.5* inch propeller is the closest to the desired specifications and hence selected for further analysis.

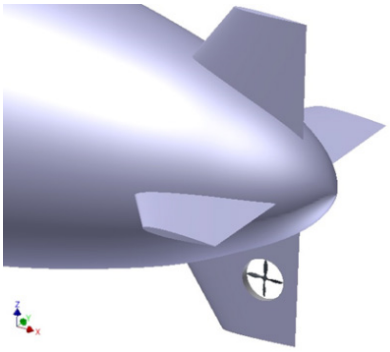


Figure 1. Location of electric trimmer on the vertical tail of the HBAV (Haque,et al., 2014)

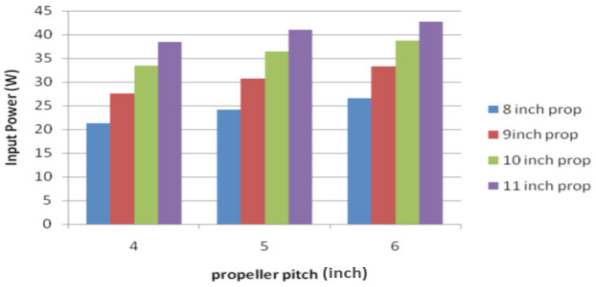


Figure 2. Input Power vs propeller pitch for different propellers

GEOMETRY DESCRIPTION

The APC propeller used in this research is described in the manual of *APC* propellers (User's Manual, 2015). Its geometry is defined in terms of diameter and pitch. It is conventional to define the diameter as twice of the radius of the circle which is swept by the tips of the blade. However, in the case of two-bladed propellers, it is the length from tip of one blade to the other blade, the value of which is 9 inch in present case. For the estimation of the yaw force, complete geometry of the selected propeller is defined by two different airfoils, NACA4412 or Clark Y airfoil, (Database of Airfoils, 2015). Both of these airfoils are initially used to evaluate the lift and drag characteristics. For the stability analysis, all the moments are computed, namely the centre of gravity *(CG)*, the location of which for our model is shown in Figure 3. The spanwise sectional views of the selected propeller are shown in Figure 4. Nomenclature used to define the geometry of the individual section of the blade, in terms of the chord and twist, is mentioned in Figure A-1 of the Appendix. The analytical results of *Motocalc* at different *RPMs* (Revolution per Minute) are also highlighted in the Appendix.

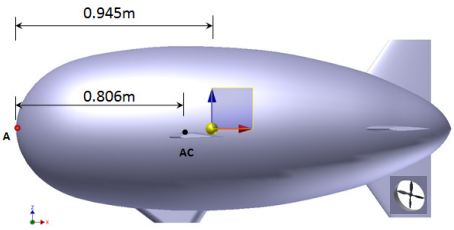


Figure 3. Location of the *CG* of the model

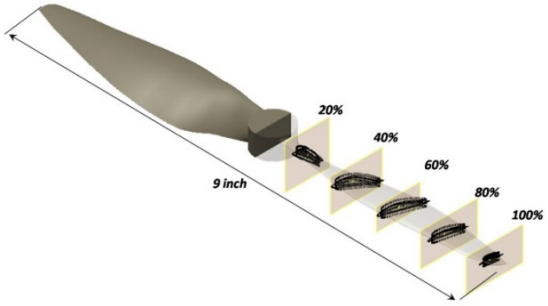
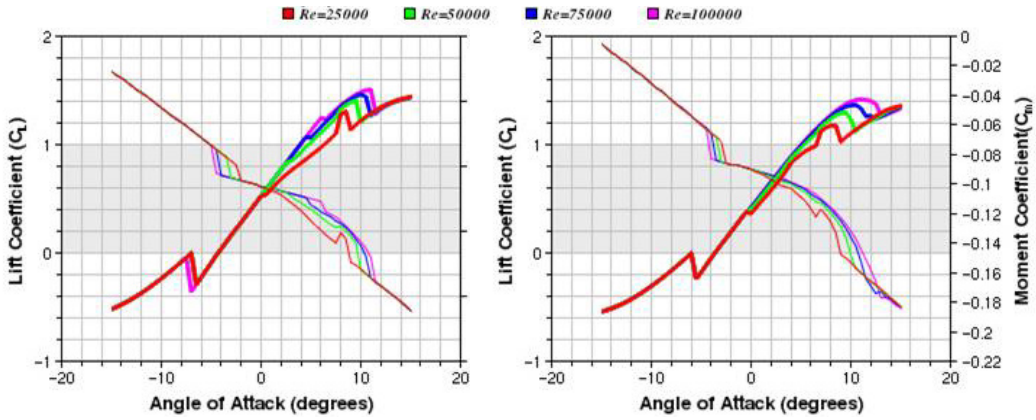


Figure 4. Sectional views of airfoil of *9 × 4.5 inch* propeller

ANALYTICAL RESULTS

Detailed geometric parameters of a *9×4.5* inch *APC* propeller have been generated in CAD and the same profile has been built in the *Motocalc* for calculating the performance parameters. The digit 4.5 in the nomenclature of the propeller represents the theoretical advancement of propeller in one revolution. Aerodynamic coefficients of the airfoils (*NACA 4412 and Clark Y* airfoil) of propeller have been estimated by using *Java-Foil*; an open source panel method, against the range of Reynolds number which have been calculated using the blade width and the resulting velocity at 75% of its radius. Plots of results taken from this software are shown in Figure 5. Combined Blade Element Momentum *(BEM)* theory is used by *Motocalc*, as the theoretical tool for the estimation of the thrust produced by the propeller. However, the accuracy of the thrust obtained by using BEM is dependent on how accurate the data points of the airfoil geometry of blade is. For the estimation of thrust force, few assumptions are made by the software. First, the airfoil has constant lift curve slope, $a_0 = 3.6897$ rad^{-1} (based on *NACA4412* C_l vs α curve) is taken as constant. Second, distribution of blade angle is linear and that the blade has negligible contribution towards its thrust in the region which is in the close vicinity of the hub of the propeller ($r < 0.012m$).



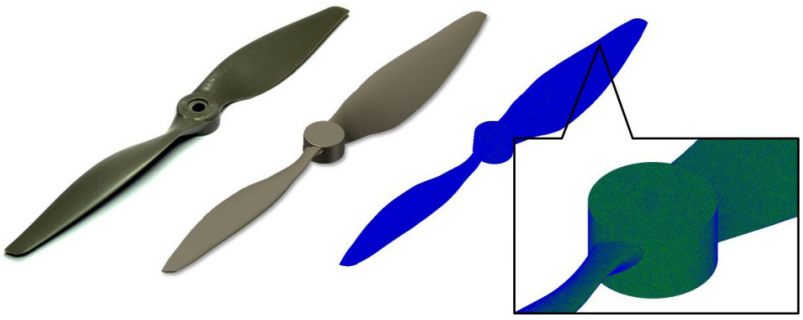
(a) NACA 4412 (b) Clark Y

Figure 5. Aerodynamic coefficients of selected airfoils obtained by *Java-Foil* at different Reynolds numbers

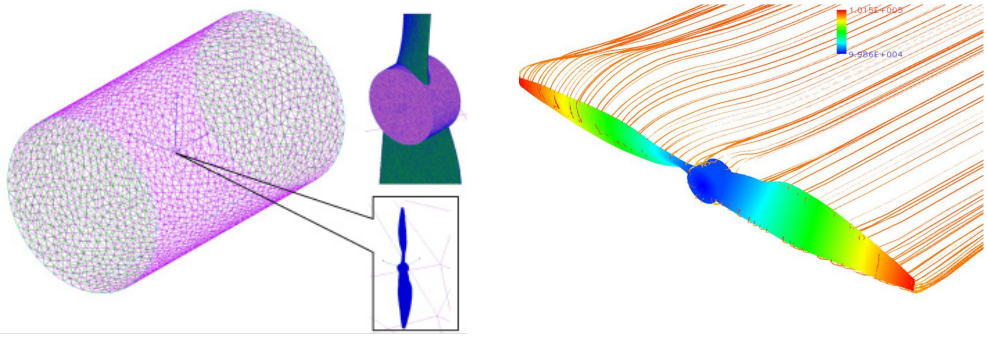
The contribution of each blade element for thrust is estimated by *Motocalc*® and the values obtained are integrated to get the total thrust value of the propeller. Performance parameters are then evaluated at different *RPMs* and all the results are presented in Appendix-A for quick reference. Based on these results shown in Figure A-2 to A-4, the value of η is found to be maximum at selected *RPM* for advance ratio equal to 0.36. It is also observed that with the increase in *RPMs,* there is an increase in C_P, C_T and η for lower range of J. However, this increase is not obvious for $J < 0.2$ as well as for $J > 0.46$. Against a fixed maximum *RPM* of 6900 with free stream velocity equal to 12 m/sec and Re (based on 75% of radius) equal to 8.114×10^4, total thrust is equal to 1.5974 N and 1.5926 N for *Clark Y and NACA 4412* airfoil respectively. Hence, no significant difference between *Motocalc*® results has been found for Clark Y airfoil and for NACA 4412 airfoil.

COMPUTATIONAL RESULTS AND ITS COMPARISON

The *CFD* analysis is carried out using commercial *CFD* software *Fluent*® to capture key flow physics in the propeller's aerodynamic characteristics. First, the geometry (Figure 6(a)) is created based on available information about the geometrical parameters and unstructured sliding mesh is employed. Such a grid allows adjacent grid to slide over it. Special attention is paid to get more grid resolution close to the wall Figure 6(b). Farfield boundary condition was applied for the computational domain shown in Figure 6(c) and the required *RPM* is incorporated in the moving boundary condition of the wall of propeller. The *SA* turbulence model was used in this research. Unsteady simulation with dual-time marching is also carried out to get a numerical solution in the form of pressure distribution on the windward and leeward side of the propeller, which finally provides its thrust value from the said pressure difference. Reynolds number and free stream velocity are kept the same as that used in the analytical work for combined BEM theory. Streamline plots show the roll-up path lines at the tip of the propeller as seen in Figure 6(d) and they show that there will be additional induced drag as well as tip loss in thrust.



(a) Existing Propeller and its CAD Model (b) Surface Mesh



(c) Computational Domain (d) Streamlines with contours of pressure

Figure 6. CFD Analysis at R_e (based on 75% of radius) = 8.11 × 10^4, V = 12 *m/s* at sea level condition

However, the computed results of *9×4.5* inch propeller have given thrust value equal to *1.89 N* and thus over-prediction is shown by *CFD* work compared to the analytical value of *1.59 N*. This over-prediction by *Motocalc* is mainly due to the limitation of combined Blade Element Momentum theory as described earlier. The thrust produced by the propeller interacts

with the tangential free stream velocity causing an obvious loss of yaw force. It is roughly assumed that *33 %* loss is expected which will result in a yaw force of *1.27 N* and the moment produced around the fixed location of the *CG* will be equal to *3.81 N.m*. This assumption might be open to question but it will give a starting point for the preliminary lateral stability analysis.

Electric trimmers studied in this work can generate thrust force in the opposite direction as well. For this purpose, the pilot has to first stop the motor for negative yaw. This additional fatigue on the operator can be minimised by employing more electric motors at the front part of the hybrid airship, pictorially shown in Figure 7.

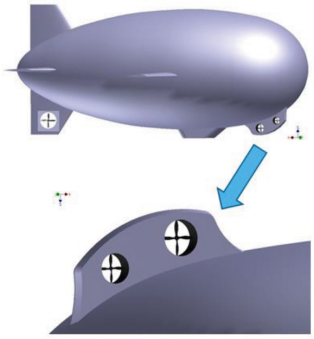


Figure 7. Twin electric motors at the nose of the hull

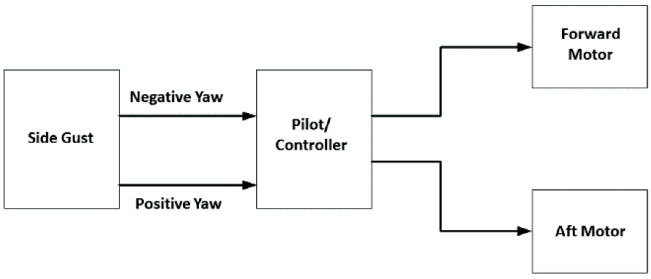


Figure 8. Block diagram for differential thrust

In the case of the electric motor failure, an alternative yaw source for the system can be introduced. For such twin electric motors, yaw is controlled by creating difference in the speed of motors i.e. difference in *revolution per minute (RPM)* of the forward and the aft fan. This is described in Figure 8 with the help of a block diagram. Speed of motors of each group can be controlled by the pilot. If no yaw manoeuvring is required, both the outputs are driven with the same pulse width signal in order to keep uniform motor speed, but in opposite direction. This concept needs thorough calculations for different flight conditions which is not covered in this study.

CONCLUSION

Based on the analytical approach, it has been found that a *9×6* inch propeller (having slightly higher pitch) produces a minute increment in the value of static as well as forward thrust, when compared with a *9×4.5* inch propeller. First order approximation of thrust produced by the propeller can be done by using *Motocalc* for which complete geometric details of the propeller and the flow conditions are a prerequisite for analysis. However, *Motocalc* cannot capture the effect of wake separation and tip losses. A first order approximation of the control force and moment generated by the electric trimmers for full load and free stream condition can be done from the *CFD* analysis. The *CFD* results can also be utilised for the dynamic stability analysis of such aerial vehicles.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Ministry of Science, Technology and Innovation (MOSTI), Malaysia, under the grant 06-01-08-SF0189 for this research work.

REFERENCES

Airfoils. (2015). *Database of Airfoils*. Retrieved September, 2105, from http://www. *airfoiltools.com*

Andan, A. D., Asrar, W., & Omar, A. A. (2012). Investigation of Aerodynamic Parameters of a Hybrid Airship. *Journal of Aircraft*, *49*(2), 658–662.

Brandt, J. B. (2005). *Small-Scale Propeller Performance at Low Speeds*. (Master dissertation). Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, Illinois.

Carichner, G. E. & Nicolai, L. M. (2013). *Fundamentals of Aircraft and Airship Design: Volume II, Airship Design*. (J. A. Schetz, Ed.). American Institute of Aeronautics and Astronautics, AIAA.

Friedrich, C., & Robertson, P. A. (2014). Hybrid-electric propulsion for aircraft. *Journal of Aircraft, 52*(1), 176-189.

Farhana, A. (2012). *The development of a mathematical model of a hybrid airship*. (Master dissertation). University of Southern California, USA.

Glauert, H. (1926). *The elements of aerofoil and airscrew theory*. Cambridge University Press, UK.

Haque, A. U., Asrar, W., Omar, A. A., Sulaeman, E., & Ali, J. S. M. (2014). Static longitudinal stability of a hybrid airship. In *Proceedings of 2014 11th International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan*, (pp.343–348). doi:10.1109/IBCAST.2014.6778167.

Khoury, G. A., & Gillett, J. D. (1999). *Airship Technology*. New York, NY: Cambridge Univ. Press.

Madsen, H. A. (1996 Dec). A CFD analysis for the actuator disc flow compared with momentum theory results". In *Proceedings of the 10th IEA Symposium of Aerodynamics and Wind Turbines*. Edinburgh, UK.

Propeller, A. P. C. (2012). *APC Propeller Performance Data*. Retrieved September, 2105, from https://www.apcprop.com/Articles.asp

Schroijen, M. J. T., Veldhuis, L. L. M., & Singerland, R. (2008). Propeller slipstream investigation using the Fokker F27 wind tunnel model with flaps deflected. In *26th International Congress of the Aeronautical Sciences, ICAS-2008*.

Stockbridge, C., Ceruti, A., & Marzocca, P. (2012). Airship Research and Development in the Areas of Design, Structures, Dynamics and Energy Systems. *International Journal of Aeronautical and Space Sciences*, *13*(2), 170–187. doi:10.5139/IJASS.2012.13.2.170

User's Manual. (2015). *MotoCalc Electric Flight Performance Prediction Software*. Retrieved July, 2015, from www.motocalc.com/

Zeune, C. O. M., & Logan, M. (2008, August). Analytical–Experimental Comparison for Small Electric Unmanned Air Vehicle Propellers. In *26th AIAA Applied Aerodynamics Conference*, (pp. 18-21).

APPENDIX

All the nomenclatures used to define the geometry in terms of the individual sections of the selected *APC* propeller are shown in Figure A-1. All the necessary calculations have been done by *Motocalc* to estimate the performance parameters of the propeller, Figure A-2 to Figure A-4. User limitations of this software are found in the User's Manual (2015). The results obtained delineate uniform trends as there is no scattering of data at different *RPM*. All these analytical results are obtained using the aerodynamic data of NACA 4412 airfoil as sectional profile of the blade. It can be observed from the figures an increase in *RPM* corresponds with an increase in C_P, C_T and η for lower range of J.

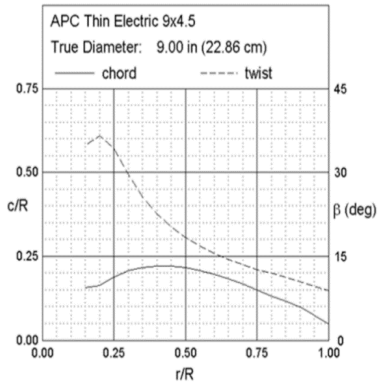


Figure A-1. Geometric parameters of the selected propeller

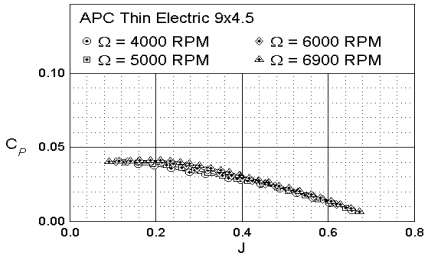


Figure A-2. Variation in C_P with J for different *RPM*

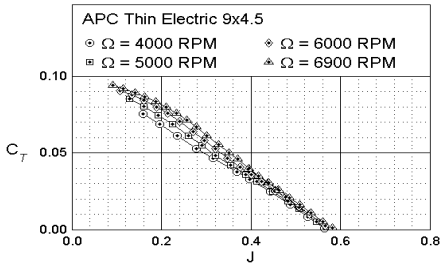


Figure A-3. Variation in C_T with J

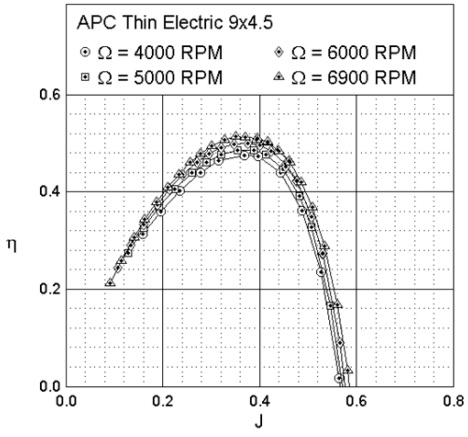


Figure A-4. Variation in η with J

However, this increase is not obvious for $J < 0.2$ as well as for $J > 0.46$. These analytical results are limited for small propellers and due to their structural anatomy, their performance results cannot be used for their larger counterparts. Moreover, similar to other thin electric propellers (Brandt, 2005; Zeune, & Logan, 2008), wind tunnel testing is recommended to investigate further the performance parameters, especially when the fan is ducted at the tail of the HBAV. For example, with the help of torque transducers, the trend plot of C_P at different RPMs can obtained.

Pertanika J. Sci. & Technol. 25 (1): 303 - 316 (2017)

SCIENCE & TECHNOLOGY

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

Impact Characterisation of Glass Fibre-Reinforced Polymer (GFRP) Type C-600 and E-800 Using a Single Stage Gas Gun (SSGG)

Syafiqah Nur Azrie, S., Mohamed Thariq, H. S.* and Francisco, C.

Aerospace Manufacturing Research Centre, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

This paper presents experimental findings derived from high velocity impact tests on woven-roving Glass Fibre-Reinforced Polymers (GFRP) Type E-800 g/m² and Type C-600 g/m². The effects on specimen thickness, projectile shape and gas gun pressure were investigated. As the gas gun pressure increases, there is a proportional increase in the projectile kinetic energy, the projectile initial velocity, the maximum force exerted on the specimens and in energy absorption upon impact. During the test, the shape of the projectile, the target thickness and the gas gun pressure significantly affected the impact performance of the GFRP. From the experiment, it was found that GFRP Type E-800 g/m² is stronger and more impact resistant compared with GFRP Type C-600 g/m², due to the fact that E-glass materials have higher fibre volume and density and overall better mechanical properties than C-glass fibres. Therefore, GFRP Type E-800 g/m² composites should be considered for applications in load and impact bearing aircraft structures.

Keywords: Energy Absorption,Glass Fiber Reinforced Polymer (GFRP), High Velocity Impact (HVI), Impact Characterisation, Polymer Matrix Composites (PMC), Single Stage Gas Gun (SSGG)

INTRODUCTION

Polymer matrix composites (PMCs) in the form of Glass Fibre-Reinforced Polymers (GFRPs) is extensively used in engineering structures. It is usually applied in the aviation industry due to their properties, such as stiffness, high strength and great fatigue resistance. It is also the most economical choice in terms of manufacturing life cycle costs. Currently, in military applications, there has been an increasing demand to reduce the weight of armour structures, which will improve the armoured vehicle mobility, fuel efficiency and transportability and thus GFRPs are good a choice as they can resist heavy loads and offer great resistance to impact.

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
snasafri@gmail.com (Syafiqah Nur Azrie, S.),
thariq@upm.edu.my (Mohamed Thariq, H. S.),
francisco.c@upm.edu.my (Francisco, C.)
*Corresponding Author

ISSN: 0128-7680

Impact loading are divided into the following categories: low, intermediate, high/ballistic and hyper velocity impact. Naik and Shrirao (2004) state that as the projectile velocity varies, there are distinct differences in damage propagations, energy transfer and energy dissipation between the target and the impacting projectile. Vaidya (2011), shows that hyper impacts occur between 2 km/s and 5 km/s, high ballistic impacts within the range of 50 m/s and 1000 m/s, intermediate is between 10 m/s and 50 m/s while low velocity impacts occur at 10 m/s.

According to Zhou (1995), Davies (1996) and Belingardi and Vadori (2002), GFRPs have higher impact damage tolerance than carbon fibre-based laminates in addition to being cheaper. Bibo and Hogg (1998) and Thanomsilp and Hogg (2003) said that GFRP laminates have greater impact resistance since they have higher energy absorption due to their higher strain to failure ratio compared with carbon fibre-reinforced composites.

The GFRPs have lower Young's modulus and density but a higher strength compared with carbon fibre laminates. The GFRP Type E fibres are commonly used in structural applications but they experience degradation in highly acidic or alkaline environment. Therefore, resistant glass was developed such as C–glass, also known as 'chemical' glass. Type C-glass is used as a surface coating in water pipes and tanks. Table 1 shows the mechanical properties of GFRP Type C and Type E, confirming that the tensile strength, density and modulus of elasticity of GFRP Type C are lower than those of GFRP Type E (Hausrath & Longobardo, 2010).

Table 1
Mechanical Properties of GFRP Type E and Type C (Hausrath & Longobardo, 2010)

Property	**E-glass**	**C-glass**
Tensile strength at 23°C (MPa)	3445	3310
Elongation percentage	4.8	4.8
Young's modulus at 23°C (GPa)	72.3	68.9
Density (g/cm3)	2.58	2.52

Type C-glass fibre with a mass of 600 g/m² is thinner compared with Type E-glass fibre with a mass of 800 g/m². The hardness of Type E fibre with a mass of 800 g/m²is higher than the value for Type C fibre with a mass of 600 g/m² since its fibre composition is greater. There are only a few studies that evaluated the performance of Type E-800 g/m² and Type C-glass/ Epoxy 600g/m²composite materials.

Factors Influencing Impact Characteristic

Influence of fibre properties. Naik and Shrirao (2004) compared twill weave T300 carbon/ epoxy composite and two-dimensional woven fabric of plain weave E-glass/epoxy to study their ballistic impact behaviour. The results indicated that E-glass/epoxy panels have higher ballistic limit than T300 carbon/epoxy laminates.

Influence of Specimen Thickness. Naik and Doshi (2008) report that during high velocity impact test, thickness of the composite materials is a significant performance factor. Gellert,

Cimpoeru and Woodward (2000) discovered that energy absorption during impact in thin Glass Reinforced Polymer (GRP) is independent of the projectile nose geometry. This study focused on three different projectile shapes and the perforation of GRP composites. Shaktivesh, Nair, Sesha Kumar and Naik (2013) studied the effect of panel thickness on the impact damage and found that energy absorbed by various mechanisms increases and is directly proportional to the target thickness and the velocity of the projectile.

Influence of Projectile Shape and mass. Projectile shapes normally tested in high velocity impact experiment are blunt, conical, or hemispherical. Corran, Shadbolt and Ruiz (1983) found that the projectile nose radius affects critical impact energy. Børvik et al. (2002) report that hemispherical projectiles affect tensile stretching after severe indentation and thinning of the specimen, while conical shape projectiles affect ductile hole expansion in thicker specimens and pealing in thinner specimens. Blunt projectiles affect failure through shear plugging. Ohte, Yoshizawa, Chiba and Shida (1982) compared three different shapes of projectile and have established that the conical projectiles required less perforation energy to penetrate the target. Wen (2000) and Wen (2001) came up with model to estimate the penetration and perforation of monolithic composite laminates impacted transversely by projectiles with different nose shapes.

Influence of Projectile Velocity. In high velocity impact testing, gas driven guns that uses high pressure can determine projectile velocity . Previous research has usually involved experiments on the impact parameter effects that affect the impact test result. Very limited research has been carried out to compare impact resistance of different types of composite materials.

Therefore, seeks to evaluate and understand the differences in impact response between Glass Fibre Reinforced Polymer (GFRP) Type C-600 g/m² and Type E-800 g/m² since both of these materials are potentially useful in aircraft applications. This research compares the impact response of Glass Fibre Reinforced Polymer (GFRP) Type C-600 g/m² and Type E-800 g/m² using high velocity impact testing and investigates specimen thickness, type of projectiles, and projectile velocity on the impact characteristics of these two types of glass-fibre reinforced composite materials.

METHODOLOGY

Specimen Preparation

The specimen was prepared at the Aerospace Material Laboratory, Faculty of Engineering, Universiti Putra Malaysia. The matrix solution is based on a 2:1 ratio portions of epoxy to the hardener. The epoxy and hardener are Zeepoxy HL002TA and Zeepoxy HL002TB respectively. The specimens were fabricated using a hot bonder machine, as previously reported by Ilcewicz et al. (Ilcewicz, Cheng, Hafenricher& Seaton, 2009). For Glass Fibre Reinforced Polymer (GFRP) Type E-800 g/m² and Type C-600 g/m², the temperature applied was 150ºF and the composite panels were hold at this temperature for 120 minutes to obtain fully cured GFRP laminates.

A hot bonder is a portable device that controls heating based on temperature feedback from the area under repair. The cure cycle is controlled using thermocouples and data can be printed out. The hot bonder machine is used together with vacuum bagging. This technique can hold

the resin in place until it fully cures using atmospheric pressure as reported by Ilcewicz et al. (2009) and Petrone et al. (2014). The high velocity impact test specimens were then cut, using a CNC Router Machine, ACM 1325, to a size of 100 mm × 100 mm (Hameed Sultan, 2007).

The stacking sequence of the glass fibre was set at 0°. Glass Fibre Reinforced Polymer (GFRP) Type E-800 g/m² and Type C-600 g/m² were fabricated into four different thicknesses, 6 mm, 8 mm, 10 mm and 12 mm, as shown in Table 2.

Table 2
Number of plies used in the laminates with TypeE-glass/epoxy 800 g/m² and Type C-glass/epoxy 600 g/m² for the high velocity impact tests

Type of GFRP	Thickness, (mm)	Number of plies
E-glass/Epoxy 800 g/m²	6	9
	8	12
	10	15
	12	18
C-glass/Epoxy 600 g/m²	6	9
	8	12
	10	15
	12	17

Experimental Setup

High energy impact tests were carried out using a single stage gas gun (SSGG) at the Faculty of Manufacturing Engineering, Universiti Malaysia Pahang. In this test, specimens with dimension of 100 mm × 100 mm were impacted with three types of projectiles: blunt, conical and hemispherical. The speed of the projectiles during the test was set up and varied using the pressure control of the gas gun. In this test, the projectile velocity range is within 70-240 m/s.

The gas gun uses compressed helium gas to fire the projectile. The velocity of the projectile is controlled by the gas pressure. Three samples for each testing condition are used to obtain average values and standard deviation values. The gas gun is connected to a ballistic data acquisition system. The software performs its calculations from the basic force-time information including projectile velocity, maximum force, kinetic energy and energy absorbed by the impacted specimen.

Impact Mechanics

In high velocity impact testing, the varying parameters are the gas gun pressure which will affect the projectile initial velocity, the target thickness and the projectile shape. The high velocity impact testing uses the single stage gas gun, which is connected to the ballistic data acquisition system. Data obtained from the system are: maximum force, kinetic energy, approximate projectile speed and absorbed energy. The approximate projectile velocity and the maximum force are obtained from the force transducer in the catch chamber. The projectile kinetic energy can be obtained from the kinetic energy, KE equation of a non-rotating object of mass, m, travelling at a speed, v:

$$KE = \frac{1}{2} mv^2 \quad [1]$$

The mass value is the projectile mass (in grams) while the velocity value refers to the projectile velocity in ms-1. The absorbed energy value can be obtained applying the (2). The work done by a constant force of magnitude, F, on a point that moves a displacement (not distance), s, in the direction of the force is the product as given in (2):

$$\text{Energy Absorbed} = F \times s \quad [2]$$

In this test, the force value, F, is obtained from the force sensor installed in the single stage gas gun (SSGG). The displacement value, s, is the horizontal distance a projectile travels from the nozzle to the specimen which is also equal to the distance from the nozzle to the specimen. Table 3 shows the pressure applied during the tests for both GFRP laminates. Three specimens will be tested for each test, to check the repeatability of the experimental results.

Table 3
Pressure tested for both GFRP laminates with each shape of projectile

Thickness	Pressure (bar)							
	8	12	15	16	20	25	30	40
6 mm	///	///		///				
8 mm			///		///	///		
10 mm					///		///	///
12 mm					///		///	///

* / indicates number of test

Material Properties of the Projectile – Mild Steel

There are three classes of projectiles: soft projectiles, semi-hard projectiles and hard projectiles. A soft projectile is the type that will experience clear deformation during impact. Grytten et al. (2009) state that a semi-hard projectile will have some deformation during the impact test, while hard projectiles undergo little or negligible deformation. The projectiles used in this experiment are made of mild steel with low hardness to avoid erosion inside the surface of the barrel. Mild steel containing a slight fraction of carbon is tough but not really hard. In this experiment, three different types of projectile are used, e.g. blunt, conical and hemispherical as shown in Figure 1.

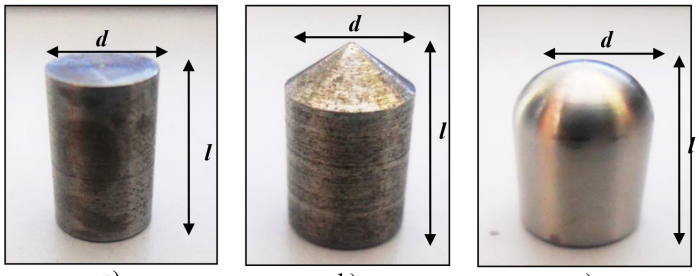


Figure 1. Types of projectile used in the actual impact testing a) blunt b) conical c) hemispherical

Table 4 shows the dimensions of the projectiles used in this experiment. Even though the dimension of these projectiles is the same, the mass is different. The differences in mass and shape of the projectile significantly affect its velocity during the impact test.

Table 4
Dimensions of projectile

Type	Length, *l* (mm)	Diameter, *d* (mm)	Mass (g)
Blunt	15	10	6
Conical	15	10	5
Hemispherical	15	10	5

RESULTS AND DISCUSSIONS

The high velocity impact test results using a Single Stage Gas Gun (SSGG) for both materials, GFRP Type E-800 g/m² and Type C-glass/Epoxy 600g/m², are presented in this section. Target plates of four different thicknesses were impacted at varying velocities by projectiles of different nose shape. For each thickness, three different gas gun pressures were used. A pilot high velocity impact test was conducted to identify the maximum gas gun pressure that can be applied before penetration occurs for each thickness value of the panels. Thus, the three best pressures were decided to test for each one of the four panels with different thickness.

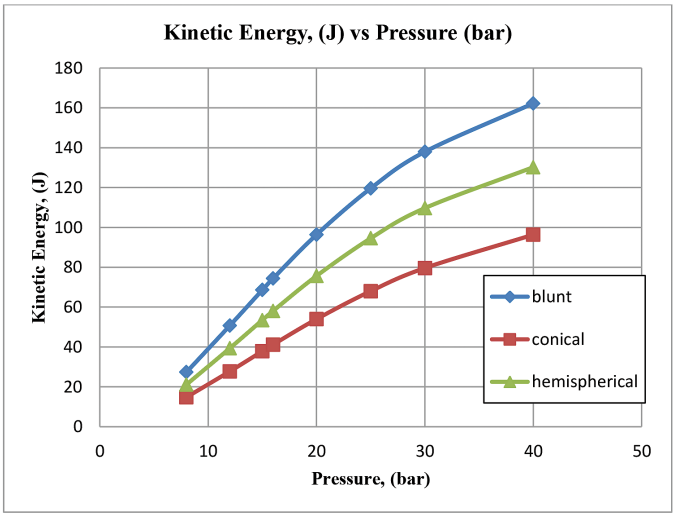
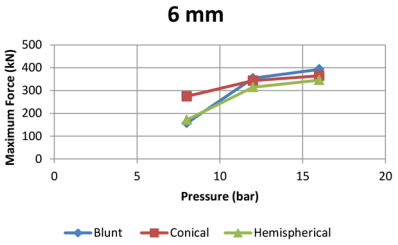
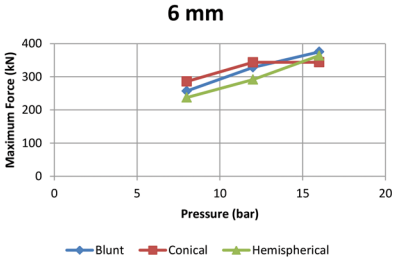
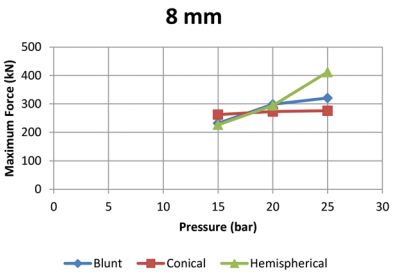
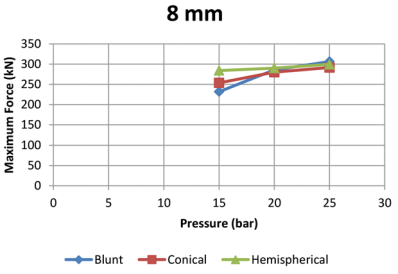
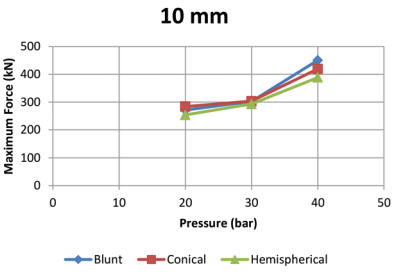


Figure 2. Kinetic Energy-Pressure curve for blunt, conical and hemispherical shapes of projectiles

Figure 2 shows the projectile kinetic energy and pressure curves for the three different projectiles experimented in this research. It shows that as the gas gun pressure increases, the projectile kinetic energy also increases. The blunt projectile has the highest kinetic energy for all tested gas pressures because the kinetic energy of a projectile is proportional to its mass and velocity. Since the mass of the blunt projectile is higher than the mass of the hemispherical and

conical projectiles, it therefore has the highest kinetic energy of all the projectiles as shown in Figure 2. Even though the hemispherical and conical projectiles have the same mass, the shape of the projectiles are different which affects their kinetic energy.

Impact Force Analysis

Type C-600 g/m²	
Type E-800 g/m²	
Type C-600 g/m²	
Type E-800 g/m²	
Type C-600 g/m²	

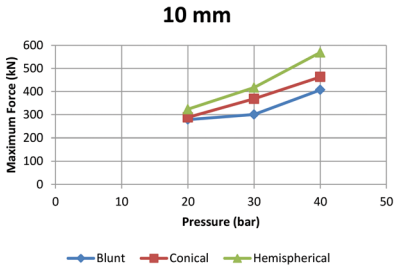
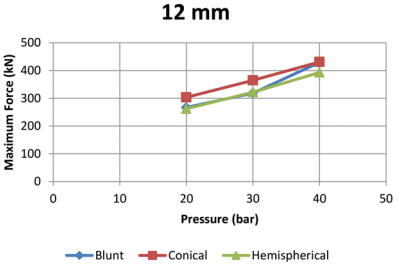
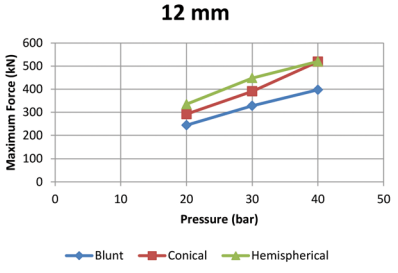
Type E-800 g/m²	
Type C-600 g/m²	
Type E-800 g/m²	

Figure 3. Maximum Force-Pressure curve for 6 mm, 8 mm, 10 mm, 12 mm thickness for Type C-glass/epoxy 600 g/m² and Type E-glass/epoxy 800 g/m²

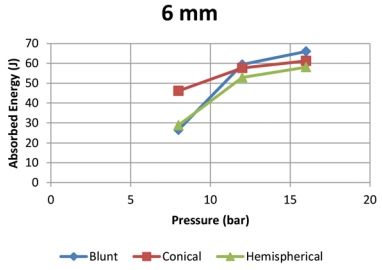
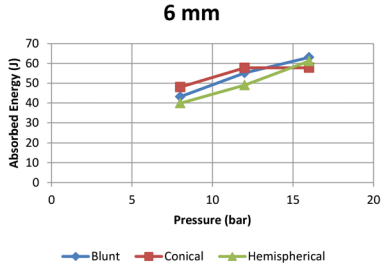
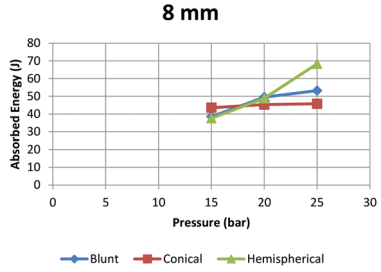
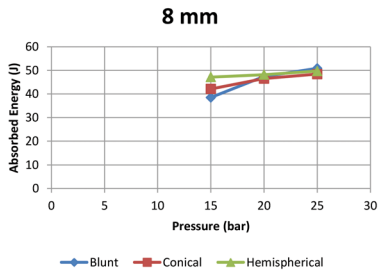
Figure 3 shows the relationship between maximum force and pressure for Type C-600 g/m² and Type E-800 g/m² composites for each of the panel thickness, 6 mm, 8 mm, 10 mm, and 12 mm. As the pressure increases, the maximum force exerted on the specimen also increases. This is because the projectile velocity increases with the pressure in the gas gun, and as a result, the maximum force exerted on the specimen also increases. Due to the different shape of the projectile the contact area between the composite and the impactor changes, affecting the force exerted onto the specimen.

From Figure 3, it can be observed that the conical projectile provides the highest maximum force followed by the hemispherical and blunt projectiles respectively. This is because the shape of the conical projectile means that it has the smallest contact area compared with the other two projectiles. Therefore, the contact area between the projectile and the tested specimen affect the maximum force exerted on the specimen during the impact. The forces acting on the projectile are the summation of the inertial force and the compressive force.

Inertial force depends on the velocity and the projectile cross sectional area, target density and shape of projectile. Awerbuch and Bodner (1974) found that inertial force is also known as work done, which can be obtained by equating the initial force on the target to the change in

the kinetic energy of the material. The compressive force is the product of compressive stress at the tip of the projectile and the cross-sectional area of the projectile. According to Udatha et al. (2012), compressive force is obtained using an interactive approach starting with a trial value of the compressive stress at the tip of the projectile. Mili (2012) says initial total contact force is the summation of inertial force and the compressive force and the impact force and central deflection are proportional to the projectile velocity.

Energy Absorbed Analysis

Type C-600 g/m²	
Type E-800 g/m²	
Type C-600 g/m²	
Type E-800 g/m²	

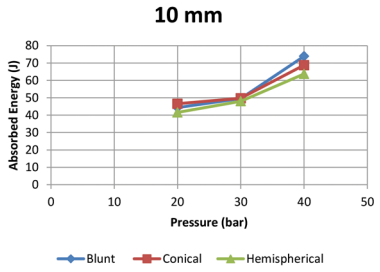
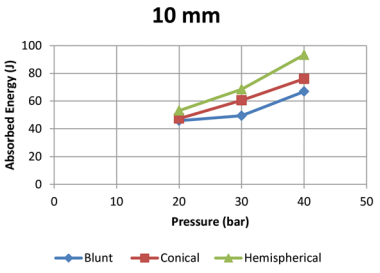
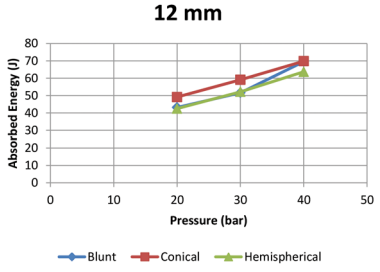
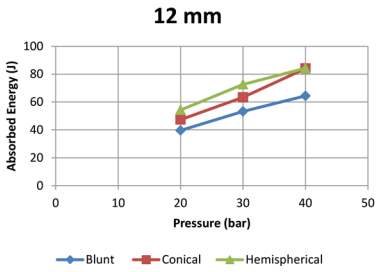
Type C-600 g/m²	
Type E-800 g/m²	
Type C-600 g/m²	
Type E-800 g/m²	

Figure 4. Absorbed Energy-Pressure curves for 6 mm, 8 mm, 10 mm and 12 mm thickness Type C-glass/epoxy 600 g/m² and Type E-glass/epoxy 800 g/m² composite panels

Figure 4 shows that as pressure increases, there is a corresponding increase in projectile velocity and absorbed energy. Increasing the initial velocity of the projectile makes penetration more localised, resulting in a proportional increase in the energy absorbed by the laminates. From the graphs, it can be seen that as the kinetic energy increased, the energy absorbed by the specimen also increased. The graphs also show that the conical projectile has the highest energy absorbed compared with hemispherical and blunt projectiles. The absorbed energy generated for impacts onto the Type E-800 g/m² are significantly larger than for Type C-600

g/m² due to its high stiffness. This shows that GFRP Type E-800 g/m² is stiffer than GFRP Type C-600 g/m². It can be surmised that the greater the peak impact forces, the stiffer the projectile-to-target interactions. As projectile contact area decreases, energy absorption also increases. This results a greater damage made by the conical projectile during impact compared with the other two types of projectiles. García-Castillo et al. (2013) noted two absorption mechanisms: energy absorbed by the tensile failure of the fibres, and the energy absorbed by the elastic deformation of the fibres.

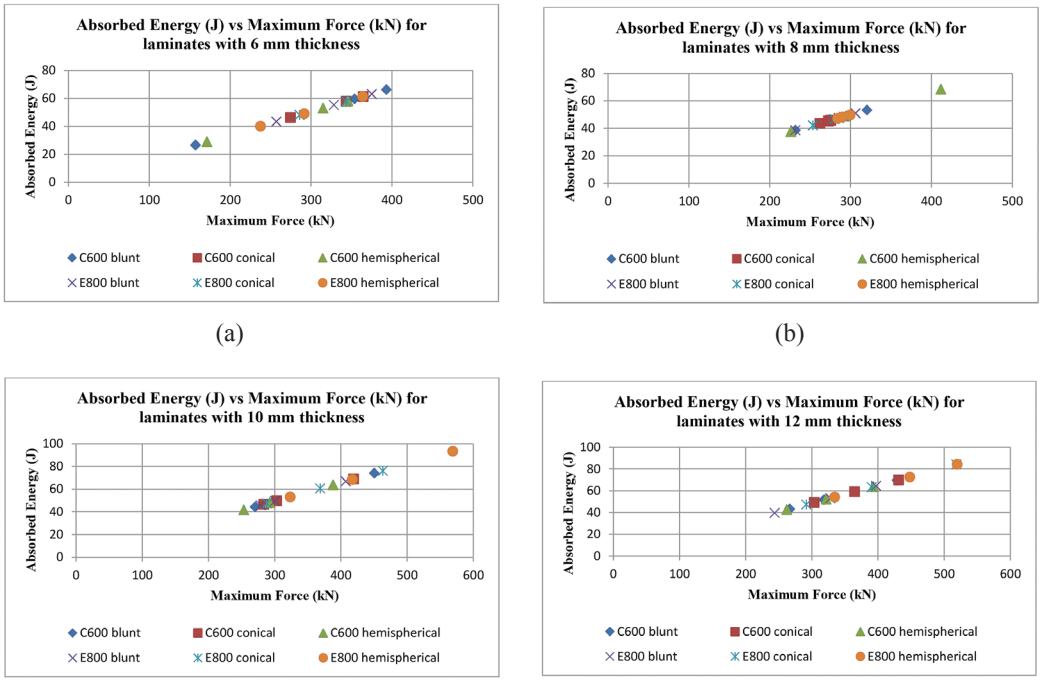


Figure 5. Absorbed Energy-Maximum Force curves for laminates (a) 6 mm (b) 8 mm (c) 10 mm (d) 12 mm thickness with Type C-Glass/Epoxy 600 g/m² and Type E-Glass/Epoxy 800 g/m²

Figure 5 shows the relationship between the maximum force and the absorbed energy for the GFRP Type C-600 g/m² and Type E-800 g/m² for the laminates with thickness of 6 mm, 8 mm, 10 mm and 12 mm. The graphs in Figure 5 show that the absorbed energy is directly proportional to the applied maximum force during impact. As the maximum force increases, there is corresponding increase in the value of the absorbed energy. This shows that maximum force has correlation with the energy absorbed. GFRP Type E-800 g/m² shows a higher maximum force compared with GFRP Type C-600 g/m² for all thicknesses as it has a greater impact resistance. The impact strength of a high impact resistant material is high thus greater force is needed to initiate damage. The laminate used in this experiment is a woven laminate. When no penetration of the laminates, the main energy-absorption mechanisms are fibre failure and elastic deformation. Therefore, the materials' initial stiffness and strength affect the impact behaviour of the laminates.

CONCLUSION

This study has measured and evaluated the effects of impact profile on parameters such as the energy absorbed, maximum force, kinetic energy, and the initial velocity of projectiles. It was found that the gas gun pressure significantly increases the maximum force and kinetic energy, which means it has a direct correlation with the projectile impact velocity. The composite specimens with the highest thickness and density experienced the greatest impact energy absorption. As the gas gun pressure increases, the projectile kinetic energy, the projectile initial velocity and the maximum force exerted on the specimen also increases. All those increasing experimental parameters result in the proportional increases of the specimen energy absorption throughout the impact tests. The impact performances of GFRPs were mainly affected by specimen thickness, projectile shape and gas gun pressure. The laminates with Type E-800 g/m² are the strongest because of E-glass fibres are stronger compared with Type C 600 g/m². The results show that GFRP with Type E-800 g/m² is better in terms of its impact performance in comparison with GFRP with Type C-600 g/m², and this is due to E-glass fibre type's higher fibre volume and density together with its superior mechanical properties. Therefore, GFRPs with Type E-800 g/m² are recommended for future aircraft structural applications.

ACKNOWLEDGEMENT

This work is supported by UPM under GP/IPB 9441503.

REFERENCES

Awerbuch, J., &Bodner, S. (1974). Analysis of the mechanics of perforation of projectiles in metallic plates. *International Journal of Solids And Structures*, *10*(6), 671-684. http://dx.doi.org/10.1016/0020-7683

Belingardi, G., &Vadori, R. (2002). Low velocity impact tests of laminate glass-fibre-epoxy matrix composite material plates. *International Journal of Impact Engineering*, *27*(2), 213-229. http://dx.doi.org/10.1016/s0734-743x(01)00040-9

Bibo, G., & Hogg, P. (1998). Influence of reinforcement architecture on damage mechanisms and residual strength of glass-fibre/epoxy composite systems. *Composites Science And Technology*, *58*(6), 803-813. http://dx.doi.org/10.1016/s0266-3538(97)00055-9

Børvik, T., Langseth, M., Hopperstad, O., & Malo, K. (2002). Perforation of 12mm thick steel plates by 20mm diameter projectiles with flat, hemispherical and conical noses. *International Journal of Impact Engineering*, *27*(1), 19-35. http://dx.doi.org/10.1016/s0734-743

Corran, R., Shadbolt, P., & Ruiz, C. (1983). Impact loading of plates — An experimental investigation. *International Journal of Impact Engineering*, *1*(1), 3-22. http://dx.doi.org/10.1016/0734

Davies, G., Hitchings, D., & Zhou, G. (1996). Impact damage and residual strengths of woven fabric glass/polyester laminates. *Composites Part A: Applied Science and Manufacturing*, *27*(12), 1147-1156. http://dx.doi.org/10.1016/1359-835x(96)00083-8

García-Castillo, S., Sánchez-Sáez, S., Santiuste, C., Navarro, C., &Barbero, E. (2013). Perforation of composite laminate subjected to dynamic loads. In S. Abrate, B. Castanié & Y. Rajapakse (Eds.), *Dynamic Failure of Composite and Sandwich Structures* (pp. 291-337). Springer Netherlands.

Gellert, E., Cimpoeru, S., & Woodward, R. (2000). A study of the effect of target thickness on the ballistic perforation of glass-fibre-reinforced plastic composites. *International Journal of Impact Engineering*, *24*(5), 445-456. http://dx.doi.org/10.1016/s0734-743x(99)00175-x

Grytten, F., Børvik, T., Hopperstad, O., &Langseth, M. (2009). Low velocity perforation of AA5083-H116 aluminium plates. *International Journal of Impact Engineering*, *36*(4), 597-610. http://dx.doi.org/10.1016/j.ijimpeng.2008.09.002

Hausrath, R., &Longobardo, A. (2010). High-strength glass fibres and markets. In F. Wallenberger & P. Bingham (Eds.), *Fibreglass and Glass Technology* (pp. 197-225). USA: Springer Science+Business Media.

Ilcewicz, L., Cheng, L., Hafenricher, J., & Seaton, C. (2009). *Guidelines for the development of a critical composite maintenance and repair issues awareness course*. Washington: U.S. Department of Transportation.

Mili, F. (2012). Effect of the Impact Velocity on the Dynamic Response of E-Glass/Epoxy Laminated Composites. *Arabian Journal for Science and Engineering, 37*(2), 413-419. http://dx.doi.org/10.1007/s13369-012-0174-9

Naik, N., &Doshi, A. (2008). Ballistic impact behaviour of thick composites: Parametric studies. *Composite Structures*, *82*(3), 447-464. http://dx.doi.org/10.1016/j.compstruct.2007.01.025

Naik, N., &Shrirao, P. (2004). Composite structures under ballistic impact. *Composite Structures*, *66*(1-4), 579-590. http://dx.doi./.05.006

Ohte, S., Yoshizawa, H., Chiba, N., & Shida, S. (1982). Impact Strength of Steel Plates Struck by Projectiles: Evaluation Formula for Critical Fracture Energy of Steel Plate. *Bulletin of JSME*, *25*(206), 1226-1231. http://dx.doi.org/10.1299/jsme1958.25.1226

Petrone, G., D'Alessandro, V., Franco, F., Mace, B., & De Rosa, S. (2014). Modal characterisation of recyclable foam sandwich panels. *Composite Structures*, *113*, 362-368. http://dx.doi.org/j.compstruct.2014.03.026

Shaktivesh, N. N., Kumar, C. S., & Naik, N. (2013). Ballistic impact performance of composite targets. *Materials and Design*, *51*, 833-846. http://dx.doi.org/10.1016/j.matdes.2013.04.093

Sultan, M. H. (2007). *High velocity impact analysis of glass epoxy laminated plates.* (Master Dissertation). Universiti Putra Malaysia, Malaysia.

Thanomsilp, C., & Hogg, P. (2003). Penetration impact resistance of hybrid composites based on commingled yarn fabrics. *Composites Science and Technology*, *63*(3-4), 467-482. http://dx.doi./s0266-3538(02)00233-6

Udatha, P., Kumar, C. S., Nair, N., & Naik, N. (2012). High velocity impact performance of three-dimensional woven composites. *The Journal of Strain Analysis For Engineering Design*, *47*(7), 419-431. http://dx.doi.org/10.1177/0309324712448578

Vaidya, U. (2011). Impact response of laminated and sandwich composites. In S. Abrate (Ed.), *Impact Engineering of Composite Structures* (pp. 97-191). New York: Springer Wien.

Wen, H. (2000). Predicting the penetration and perforation of FRP laminates struck normally by projectiles with different nose shapes. *Composite Structures*, *49*(3), 321-329. http://dx.doi.org/10.1016/s0263-8223(00)00064-7

Wen, H. (2001). Penetration and perforation of thick FRP laminates. *Composites Science and Technology*, *61*(8), 1163-1172. http://dx.doi.org/10.1016/s0266-3538(01)00020-3

Zhou, G. (1995). Prediction of impact damage thresholds of glass fibre reinforced laminates. *Composite Structures*, *31*(3), 185-193. http://dx.doi.org/10.1016/0263-8223(94)00062-x

Pertanika J. Sci. & Technol. 25 (1): 317 - 326 (2017)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Drop-Weight Impact Test on Laminated Composite Plate of Flax (Linum Usitatissimum) Using Rice Husk Ash from Paddy (Oryza Sativa) as a Natural Binder

M. S. Yaakob*, N. A. Z. Abidin, M. N. Abdullah, M. K. H. Muda and F. Mustapha

Department of Aerospace Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Composite structures and materials are used in aerospace, marine and automotive applications due to their light weight characteristics. By using rice husk ash (RHA) as natural binder to replace epoxy resin, it improves the characteristic of the composite laminated structure. Rice husk has become an important ingredient in silica, silicon carbide, and silicon nitride because of its high a silicon content. This paper evaluates the performance of laminated composite plate in a drop-weight impact testing. Two plates are attached to the test rig in the particular desired impact orientation. The main advantages of this process are the plates provide more realistic drop-weight impact and data. Data from the impact test was collected and analysed to evaluate the material properties of epoxy resin and RHA. Results show RHA's energy absorption is better and it has more deformation to prevent structure failures compared with epoxy resin. This paper aims to evaluate the application of RHA as a geopolymer binder and flax fibre as an alternative material to glass fibre in the composites industry.

Keywords: Laminated Composite Plate, Geopolymer Binder, Drop-Weight Impact Test, Rice Husk Ash (RHA), Flax Fibre, Fibre Glass

INTRODUCTION

In recent years, there has been mounting interest in environmental friendly materials among scientists and manufacturers as a result of greater environmental awareness and an increasing demand for biodegradable materials. Environmental friendly materials make optimal use of resources, produce minimum waste and are safe for the environment; new environmental laws and

Article history:
Received: 08 January 2016
Accepted: 11 November 2016

E-mail addresses:
saffuanyaakob@gmail.com (M. S. Yaakob),
mieraza90@gmail.com (N. A. Z. Abidin),
naimabdullah14@gmail.com (M. N. Abdullah),
hafyz_801@yahoo.com.my (M. K. H. Muda),
faizal@eng.upm.edu.my (F. Mustapha)
*Corresponding Author

ISSN: 0128-7680

regulations are in place to check pollution levels and thus environmental friendly materials are being promoted to achieve this. An intensive research and development over the last decade has been carried out in order to develop composites using natural fibres which offer good bio-degradability and sustainability. End-of-life after service life was a big issue involving composites structure. Compared with glass fibres, biodegradable materials undergo biodegradation at a slow pace by surrounding microorganisms, bacteria and exposure to the elements (Jinchun Zhu, 2013). In order to overcome this problem, fibres from woods, animal, grasses and other natural sources are used as reinforcement in composite materials, replacing glass fibres. These natural fibres have a potential to be utilised in the composite industry but that requires an advanced knowledge in manufacturing techniques. Compared with glass fibres, these natural fibre-reinforced biodegradable polymer d composites show equivalent performances as well as a higher fibre content. Lighter than synthetic polymer matrix, natural fibre composites contribute to increased fuel efficiency in aerospace and automotive applications. Table 1 shows the advantage and disadvantages of using natural fibre and synthetic fibres to reinforce polymers. Holbery and Houston (2004) and Avella (2009) have reviewed natural fibre composites with their classifications, properties and potential applications.

Table 1
Natural fibre and Synthetic fibre comparison

Fibre	Advantages	Disadvantages
Natural fibre	Biodegradable	non- homogeneous quality
	Low density/price	Dimensional instability
Synthetic fibre	Moisture resistant	Difficult to recycle
	Good mechanical properties	Relatively high price

Note. From Thermophysical properties of natural fibre-reinforced polyester composites by Idicula, M.; Boudenne, A.; Umadevi, L.; Ibos, L.; Candau, Y.; Thomas, S, 2006, Compo. Sci. Technol, 66, 2719–2725.

Rice husk ash

According to FAO's Liaison Office for North America (2008), 20% of the 649.7 million tonnes of rice were residue of rice husk. Utilising partially burnt rice husk from the milling plants as a supplementary cementing material safeguards the environment (Chandrasekhar, 2006). This is because the husk produces ashes when it is used as a fuel contributing to pollution. The composition of chemical in rice husk is different in each sample depending on the type of paddy, climate and geographical condition and also its crop year (Idicula, 2006). Burning the husk under controlled temperature below 800 °C can produce ash with silica mainly in amorphous form *Burning* the *husk under controlled temperature below 800°C can produce ash with silica mainly in amorphous form.* Nair (2008) conducted an investigation on the *pozzolanic activity* of *RHA* by *using various techniques* in *order to verify* the *effect* of *incineration temperature* and *burning duration.* To produce high reactivity with no significant amount of crystalline material, the RHA was burnt at 500 or 700 °C for more than 12 hours producing ashes. Because the duration of burning was short which was less than 360 minutes, it will

produce RHA with high carbon content. Mehta (1992) had reported on the physical and chemical properties of RHA, the effect of incineration conditions on the pozzolanic characteristics of the ash, and a summary of the research findings from several countries on the use of RHA as a supplementary cementing pozzolanic material.

Bio-composite Material

Derived from renewable resources, bio-composite materials are cost effective due to their large scale usage. Replacing glass fibres with bio-composite materials will have a huge impact in the composites industry. To make use of this environmental friendly material, a new way of producing bio-composite materials should be proposed. Natural composite materials, also known as green composites, are a combination of plant fibre and resin. Kenaf and flax fibres are low cost bio materials, light weighted and suitable as environmental friendly alternatives to synthetic fibres (Begum, 2013).

Flax fibre

In fibre-reinforced polymer composites, Flax fibres are considered environmental friendly along with other natural fibres and have the potential to replace synthetic fibres. A feature of natural fibres is that they have a much higher variability of mechanical properties. To produce composites from natural fibres, usually these fibres are made into some sort of a mat for easier application. Other fibre composites have been studied such as kenaf, hemp, sisal, coir and jute but the focus has been on flax fibres.

In some areas, flax fibres are competitive to glass fibres, and hence are reasonably acceptable as replacement. Except for the specific properties of flax fibres as shown in Table 2, three other reasons make the application of flax fibre more attractive: (1) cheaper than glass fibres; (2) less toxic; (3) high strength to weight ratio. Normally, flax fibres have a relatively low price compared with glass fibres. Additionally, glass fibres are suspected of causing lung cancer, but there is no such problem with regards to natural fibres (Bos, 2004). Thermal recycling of flax fibres (burning of flax fibres with few slags left) has a great advantage over glass fibres.

Table 2
Tensile properties of glass and flax fibres

Property	E-glass	Flax fibres
Specific E-modulus (Gpa/g•cm−3)	0.5–1	0.4–1.1
E-modulus (GPa)	76	50–70
Specific tensile strength (GPa/g•cm−3)	30	36–50
Tensile strength (GPa)	1.4–2.5	0.5–1.5
Density (g/cm3)	2.56	1.4
Diameter (μm)	8–14	10–80

Note. From The Potential of Flax Fiber as Reinforcement for Composite Materials. Ph.D. Thesis, Eindhoven University, Eindhoven, The Netherlands, Bos, H.L (2004).

METHODOLOGY

Specimen preparation

The tested specimens were epoxy resin and rice husk ash of flax material. In order to mix it with other chemicals, the rice husk must be fine (blending it in a blender to make it fine and smooth) so that it can dissolve into the mixture.

There are two process involved:

i. Flax Fibre with natural binder specimen

ii. Flax Fibre with epoxy specimen

Flax Fibre with natural binder specimen. First, all the apparatus and materials are prepared. Next, the materials are cut with 30cm X 30cm dimension for 3 set plate specimen. The mixer will allow the mixture to mix more evenly. It is then poured over the material. Finally, the mixture goes through the compress plate mould process.

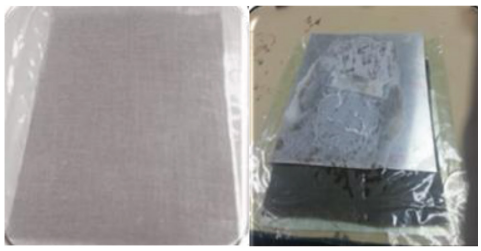


Figure 1. Specimen preparation for RHA.

Flax Fibre with epoxy. Figure 2 shows the process for plate specimen with epoxy mixture. The process is the same with the previous one but the mixture preparation and ratio are different. First, all the apparatus and the materials are prepared. Then, the materials are cut with 30X30cm dimension for 3 set plate specimen. The mixture is then over the material. The mixture then goes through the pressing process.

Figure 2. Specimen preparation for epoxy mixture.

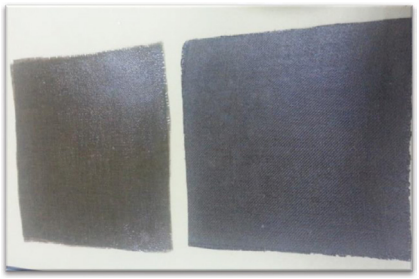


Figure 3. The specimens that will be compared.

Low velocity impact testing

The instrument used the impact test was provided by IMATEK manufacturer as seen in Figure 5. The IM1-C impact test systems consists of a manually operated, low to medium energy, impact tester primarily designed for standard test methods such as plaque penetration, compression, Charpy and Izod impact testing on polymers and advanced composite materials provide an understanding of the mechanism by which the test specimen absorbs energy or fails at rates of strain.

The base sits on four adjustable feet used for levelling. The design of the base ensures that the specimen is correctly supported on a rigid structure to prevent loss of impact energy due to movement, flexure or vibration. The specimen area (impact chamber) is enclosed for safety. The front and rear doors are fitted with impact resistant polycarbonate panels, for viewing of the impact event and to accommodate high-speed videoing of the impact event (option).

For this project, the equipment was used to identify the toughness and the energy absorbed for plate specimen with different types of material. The result will be analysed using the equipment system. This test method determines the damage resistance of a multidirectional composite laminated plate subjected to a drop-weight impact event. The test equipment utilised in this experiment meets the requirements of the American Society for Testing and Materials (ASTM D7136 / D7136M, 2005). The plate is subjected to an out-of-plane concentrated impact (perpendicular to the plane of the laminate) using a drop-weight device with a hemispherical striker tip with a diameter of 16 mm and a hardness of 60 HRC. The impact energy corresponds with the potential energy of the drop-weight, defined by the mass (2240 gr) and drop height of the impactor. The damage resistance is quantified in terms of the resulting size and type of damage in the specimen.

The plate is placed in the rigid support fixture, centred relative to the cut-out and secured using the test rig for plate as shown in Figure 4, in order to prevent the plate from rebounding during impact. Figure 5 shows the drop-weight impact-testing device. The device includes a rigid base, a cylindrical drop-weight impactor and a cylindrical guide mechanism. A rebound catcher is used to stop the impactor during its second descent. For the testing of plates that are less than 1 inch (2.42 mm) thick it has optional lightweight crossheads to meet those lower impact energies. To prevent a secondary impact of the falling weight onto the composite plate, anti - rebound device can be employed.

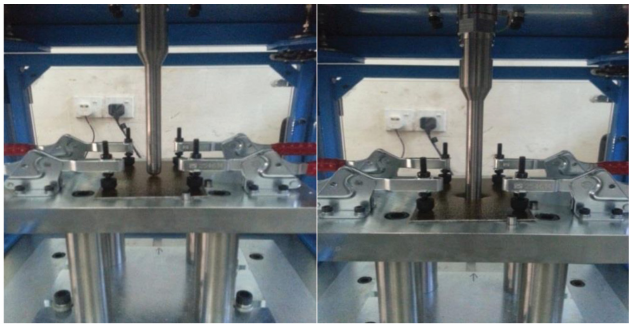


Figure 4. Plate secured in place using the test rig for impact.

Figure 5. Impact Test equipment.

RESULTS AND DISCUSSION

Visual inspections and measurements post impact show specimen failure. Complete with the Data Acquisition System, Visual Impact and a strain gauged tup failure points/modes provide instrumentation that may remain hidden under normal test conditions. The most important information that could not be found without using the instrumentation is the incipient damage point or the first crack.

Table 3
Epoxy specimen results from the impact test

	Peak force (kN)	*Impact energy (J)*	*Peak deformation (mm)*	*Absorbed energy (J)*	*Energy to peak deformation (J)*
Flax+epoxy 1	1.36	23.62	40.98	0.32	-2.22
Flax+epoxy 2	1.07	23.56	40.84	0.18	-5.96
Flax+epoxy 3	0.96	23.68	39.75	0.28	-1.70
Average	1.03	23.62	39.39	0.25	-3.32

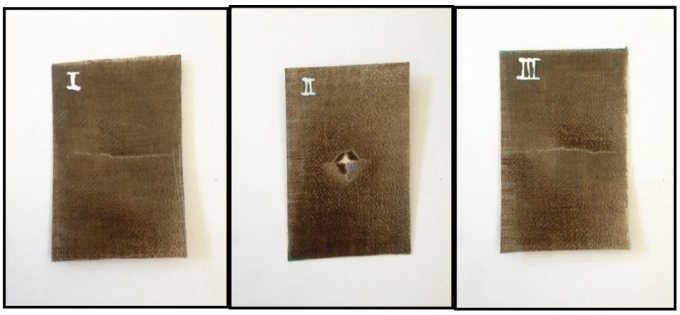


Figure 6. The epoxy specimen after impact

As shown in Figure 6, 3 specimens from epoxy were tested to the drop-weight impact event of the epoxy plate. Data from table 3 was collected and analysed. A visual inspection shows the specimen of epoxy plate could not pass the weight and height of the impactor. All the specimens indicate cracks as the impactor had punched thru the plate. Even though the plate was not impact resistant, the data could still be used to compare with the RHA plate. The energy consumed during peak deformation for the second specimen was higher and this was also the cause of the damage. Average mean of the data were compared with the RHA plate.

Table 4
RHA specimen results from the impact test.

	Peak force (kN)	*Impact energy (J)*	*Peak deformation (mm)*	*Absorbed energy (J)*	*Energy to peak deformation (J)*
Flax+RHA 1	0.98	23.58	40.54	0.27	-2.53
Flax+ RHA 2	1.13	23.55	39.99	0.32	-1.80
Flax+ RHA 3	0.95	23.56	42.33	0.27	-1.95
Average	1.01	23.56	40.17	0.29	-2.13

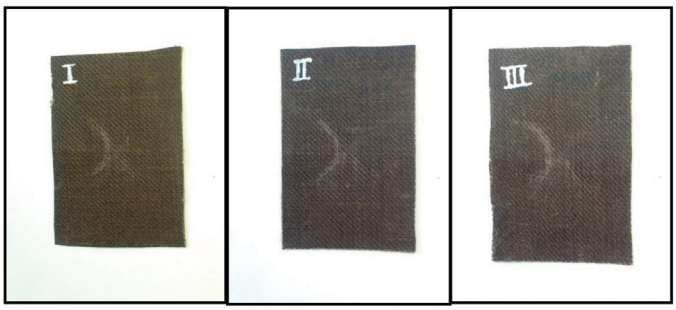


Figure 7. The RHA specimen after impact

Table 4 shows RHA specimen impact test data and the mean of dataset. Figure 7 shows the specimen post impact. A visual inspection shows the RHA plate does have any defect except the dent due to the impact. The average impact energy for this specimen is the same as the

third specimen. Even though it is a high energy absorbing material and extensive evidence of deformation because of the impact, it does not crack or damage the plate. The energy absorption during peak deformation for the first specimen was higher but did not cause any damage to it. Average data for RHA specimen was then compared with the epoxy specimen.

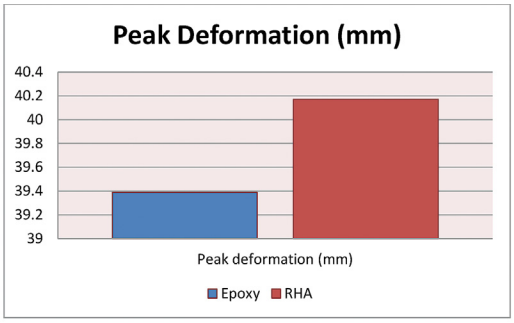


Figure 8. Comparison of Average Peak Deformation (mm) for Epoxy and RHA.

Figure 8 compares the peak deformation for epoxy plate and RHA plate. For epoxy plate, 39.39 mm deformation was measured after impact and 40.17 mm for RHA plate. Thus, RHA plate showed greater deformation compared with epoxy plate, indicating flexibility of the material. The results for the epoxy plate showed the materials are more brittle compared with the RHA plate.

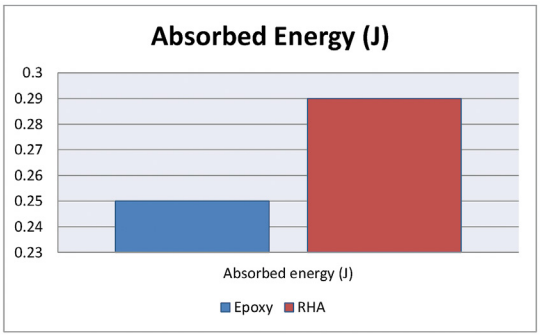


Figure 9. Comparison of Average Absorbed Energy (J) for Epoxy and RHA.

Figure 9 compares energy absorption for epoxy plate and RHA plate whereby the latter shows more deformation and greater energy absorption post impact. It shows the RHA material properties are better than epoxy materials. The energy absorption capability and deformation mechanisms of RHA plate was experimentally studied with the aim of developing a real application and also for impact analysis for future research. It also contributes to developing a comprehensive constitutive law to be implemented for industries.

CONCLUSIONS

Strength analysis under drop-weight impact was conducted for epoxy plate and RHA plate composites, results show acceptable stiffness prediction for RHA plate. The fabrication of flax composites that use natural resin are advantages in terms their recyclability allowing the reutilisation of the materials as the composites are ecologically friendly,100% biodegradable. Since the material is widely available and cheaper, there is a good future of natural fibre composites with high-performance in structure materials.

ACKNOWLEDGEMENTS

This research was supported by Universiti Putra Malaysia RUGS 9393700.

REFERENCES

ASTM D7136 / D7136M (2005, May). *Standard Test Method for Measuring the Damage Resistance of a Fiber-Reinforced Polymer Matrix Composite to a Drop-Weight Impact Event*. Retrieved August 11, 2015, from https://www.astm.org/Standards/D7136.htm

Avella, M., Buzarovska, A., Errico, M. E., Gentile, G., & Grozdanov, A. (2009). Eco-challenges of bio-based polymer composites. *Materials, 2*(3), 911–925.

Begum, K., & Islam, M. A. (2013). Natural Fiber as a substitute to Synthetic Fiber in Polymer Composites. *Research Journal of Engineering Sciences, 2*(3), 46-53.

Bos, H. L. (2004). *The Potential of Flax Fiber as Reinforcement for Composite Materials*. (Doctoral Dissertation). Eindhoven University, Eindhoven, The Netherlands.

Chandrasekhar, S., Satyanarayana, K., Pramada, P., & Majeed, J. (2006). Effect of calcinations temperature and heating rate on the optical properties and reactivity of rice husk ash. *Journal of Materials Science (Norwell), 41*(1), 7926-7933.

Chandrasekhar, S., Satyanarayan, K. G., Pramada, P. N., & Raghavan, P. (2003). Review processing, properties and applications of reactive silica from rice husk—an overview. *Journal of Materials Science (Norwell), 38*(15), 3159 – 3168.

FAO. (2008). *Food and Agriculture Organization of the United Nations. World paddy production*. Retrieved December 26, 2015 from http://www.fao.org/newsroom/en/news/2008/1000820/index.html

Idicula, M., Boudenne, A., Umadevi, L., Ibos, L., Candau, Y., & Thomas, S. (2006). Thermophysical properties of natural fiber reinforced polyester composites. *Composites Science and Technology, 66*(15), 2719–2725.

Joshi, S. V., Drzal, L. T., Mohanty, A. K., &Arora, S. (2003) Are natural fiber composites environmentally superior to glass fiber reinforced composites. *Composites Part A: Applied science and manufacturing, 35*(3), 371-376.

Mehta, P. K. (May 1992). Rice husk as: a unique supplementary cementing material. In *Proceedings of the International Symposium on Advances in Concrete Technology* (pp. 407-430). Athens, Greece. Canada: CANMET.

Nair, D., Fraaij, A., Klaassen, A. & Kentgens, A. (2008). A structural investigation relating to the pozzolanic activity of rice husk ashes. *Cement and Concrete Research* (Elmsford), 38(6), 861-869.

Pereira, R., Filho, S. P. C., Curvelo, A. A. S., & Gandini, A. (1997). Benzylated pulps from sugar cane bagasse. *Cellulose 4*(1), 21-31.

Rowell, R. M. (1998). Property Enhanced Natural Fiber Composite Material based on Chemical Modification. In P. N. Prasad, J. E. Mark, S. H. Kendil & Z. H. Kafafi (Eds.), *Science and Technology of Polymers and Advanced Materials*, (pp. 717-732). Plenum Press, New York.

Wollerdorfer, M., & Bader, H. (1998). Influence of natural fibers on the mechanical properties of biodegradable polymers. *Industrial Crops and Products, 8*(2), 105-112.

Yu, L., Dean, K., & Li, L. (2006). Polymer blends and composites from renewable resources. *Progress in polymer science, 31*(6), 576-602.

Zhu, J., Zhu, H., Njuguna, J., & Abhyankar, H. (2013). Recent Development of Flax Fibres and Their Reinforced Composites Based on Different Polymeric Matrices. *Materials, 6*(11), 5171-5198.

Pertanika J. Sci. & Technol. 25 (1): 327 - 336 (2017)

SCIENCE & TECHNOLOGY

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

Different Germanium Dopant Concentration and the Thermoluminescence Characteristics of Flat Ge-Doped Optical Fibres

Muhammad Safwan Ahmad Fadzil[1], Ung Ngie Min[2], David Andrew Bradley[3, 4] and Noramaliza Mohd Noor[1, 5]*

[1]*Department of Imaging, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

[2]*Clinical Oncology Unit, Faculty of Medicine, University of Malaya, 50603 UM, Wilayah Persekutuan, Kuala Lumpur, Malaysia*

[3]*Centre for Nuclear and Radiation Physics, Department of Physics, University of Surrey, Guildford GU2 7XH, United Kingdom*

[4]*Sunway University, Institute for Health Care Development, Jalan University, 46150 Petaling Jaya, Malaysia*

[5]*Centre for Diagnostic Nuclear Imaging, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

The influence of elevated germanium concentration on the thermoluminesence characteristics of a novel form of fabricated flat optical fibre was examined. All the samples were irradiated with two nominal photon energies (6 MV and 10 MV) and 1.25 MeV gamma energy. Flat fibres with 10 mol % Ge concentration provided the superior TL yield compared against that of 6 and 8 mol % Ge-doped optical fibres for both 6 MV and 10 MV energy. Interpretation of the results has been aided by study of the glow curves, revealing in particular new generation of defects in the flat fibres due to strain-generation at the collapsed surfaces. The strain represent deep-energy defects.

Keywords: Dosimetric characteristics, flat fibre, glow curves

Article history:
Received: 16 June 2016
Accepted: 20 December 2016

E-mail addresses:
muhammadsafwan_ahmadfadzil@yahoo.com (Muhammad Safwan Ahmad Fadzil),
nmung@ummc.edu.my (Ung Ngie Min),
d.a.bradley@surrey.ac.uk (David Andrew Bradley),
noramaliza@upm.edu.my (Noramaliza Mohd Noor)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Optical fibres are widely known in the global communications industry and have been extensively used for worldwide broadband communication such as the internet. Recent developments in optical fibre technology have led to a renewed interest in passive dosimetry

applications. The interest concerns in particular the introduction of commercially available telecommunication doped amorphous silicon dioxide (SiO_2) fibres as a thermoluminescent dosimeter (TLD), specifically because in radiation dosimetry the optical fibres offer good spatial resolution and for certain dopants concentrations there is sufficient thermoluminescent (TL) yield to serve the associated sensitivity needs (Bradley et al., 2012).

Prior to the use of any radiation dosimeter in daily clinical applications, the characteristics of the dosimeters first need to be established, including reproducibility, linearity, dose-rate and energy dependence, and for TLDs the fading effect of superficial traps. Depending on the use and characteristics of the optical fibres, the medium can be added to or modified, including change in the concentration of the dopant element in order to alter the radiation response of the fibre. Previous studies have primarily concentrated on the more usual germanium (Ge) doped cylindrical fibre shapes, showing promising dosimetric characteristic in thermoluminescene dosimetry especially in respect of radiotherapy photon beams (Hashim et al., 2009; Hashim et al., 2010; Noor et al., 2010; Abdul Rahman et al., 2011; Abdul Rahman et al., 2012; Noor et al., 2012; Noor et al., 2014).

Recent studies on Ge-doped fibres by Alawiah et al. (2013), Alawiah et al. (2015), Bradley et al. (2015) and Mahdiraji et al. (2015) have focused on the development of a new tailor-made flat fibre. Analysis and discussion of flat fibre performance have been presented in respect of several energy and dose ranges, from diagnostic applications (Ramli et al., 2015), to radiotherapy photon beams (Fadzil et al., 2014) to high dose irradiation (Noor et al., 2015). Research on the subject has been mostly restricted to limited comparisons of different sizes of optical fibre. To-date little attention has been paid to study the effect of different dopant concentration on the thermoluminescene performance of the optical fibre, with systematic understanding of how the dopant concentration contributes to TL intensity still lacking. The purpose of this paper is to study the influence of elevated concentration of Ge-doped flat fibres on the thermoluminesence's characteristics.

MATERIALS AND METHOD

Fabrication and designation of the fibre

The manufacturing of an optical fibre involves: (i) preform fabrication and; (ii) the drawing process. The preform is a glass rod hollow-tube structure with diameter of 22 ± 2 mm. A commercially used method based on the standard modified chemical vapour deposition (MCVD) technique was employed in fabricating the preform. This technique involved chemical vapour reaction at temperature ranging from 1600ºC to 2000ºC, mixing the silica tetrachloride ($SiCl_4$) with germanium tetrachloride ($GeCl_4$) and oxygen (O_2) during passage through a rotating silica glass tube (F300). These three compounds were deposited inside the inner layer of F300 and produced several layers of the core. The percentage of dopant concentration was determined by the dopant flow rate during the MCVD process. This process was carried out using facilities at the Photonics Laboratory, Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. Three Ge-doped silica dioxide optical fibres preforms with different germanium dopants concentration (6, 8 and 10 mol %) were fabricated for this study. An overview of the preform fabrication process is represented in Figure 1.

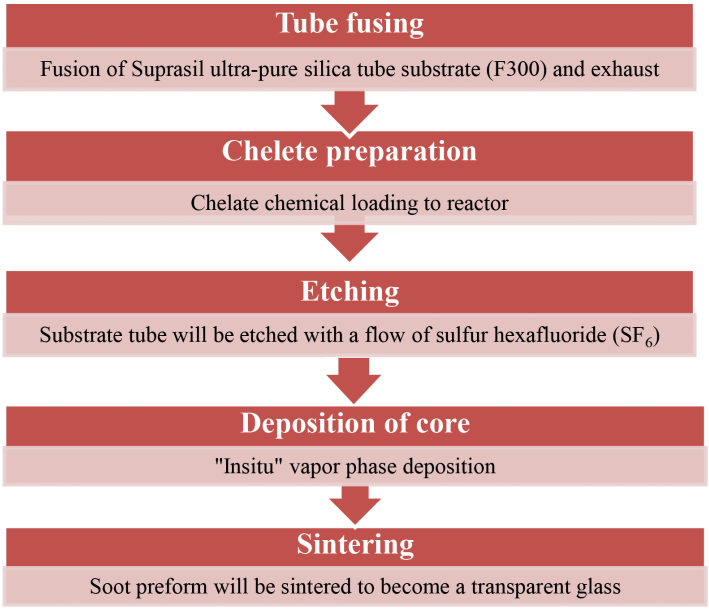
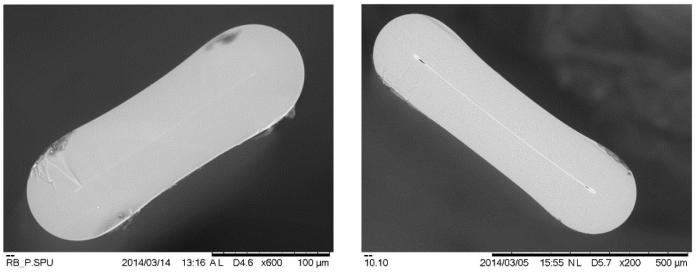


Figure 1. Preform fabrication flow chart in MCVD process.

The optical fibres then were cut and cleaned using methanol to eliminate carbon deposition. The actual flat fibre dimensions (Figure 2) and their weights are presented in Table 1 for the three Ge concentrations.



(a) (b)

Figure 2. Scanning Electron Microscope image of a small (a) and large (b) dimension of Ge-doped flat fibres.

Table 1
The flat fibre sizes and mass for the three Ge concentrations

Size		6% mole	8% mole	10% mole
Small	Size (µm)	270 ±5 x 80 ±2	255 ±5 x 50 ±2	270 ±4 x 80 ±2
	Mass (mg)	1.29 ±0.08	1.12 ±0.30	1.22 ±0.03
Large	Size (µm)	780 ±10 x 190 ±5	760 ±7 x 150 ±4	750 ±6 x 230 ±3
	Mass (mg)	3.39 ±0.16	1.90 ±0.13	3.54 ±0.28

Irradiation and Readout

The fibres were exposed to x-rays generated at 6 MV and 10 MV, with a dose rate of 600 MU/min and doses ranging from 0.5 Gy up to 10 Gy using a Varian Clinac Linear Accelerator located at the University Malaya Medical Centre (UMMC). A source to surface distance (SSD) of 100 cm was employed, with a field size of 20 × 20 cm². After the irradiation, the TL intensity of the Ge-doped fibres was obtained using a Harshaw 3500 TLD reader. The selected time-temperature profile (TTP) of the reader is shown in Table 2.

Table 2
Time-temperature profile setup for flat fibre TL assessment

Time-temperature profile parameters	
Preheat temperature	80 ºC
Acquires ramp rate	30 ºC s^{-1}
Acquisition time	13 seconds
Maximum temperature	400 ºC

Glow Curve

The shape of the glow curves obtained as a result of heating the thermoluminescent materials is important in determining the nature and TL properties of the material and also key to the media forming a suitable dosimeter. In particular, the glow curves reflect the nature of the trap parameters. The glow curve analysis was carried out herein by plotting the glow curve using Microsoft Excel software, yielding features such as the temperature at which the highest peak intensity is obtained.

RESULTS AND DISCUSSION

Here, there is increasingly complex interplay between defect generation at the collapsed fibre internal interface (so-called strain defects), the detected luminescence yield and the overlaying cladding thickness. In regard to the overall TL yield, that from the smaller dimension flat fibres doped with 6 mol % Ge is observed to produce greater TL signal intensities (Figure 3 and 4). Conversely, the TL yield of the larger dimension flat fibres shows 10% mol fibre produce greater response compared with 6 and 8 mol % fibres, both for 6 MV and 10 MV irradiations respectively (Figure 5 and 6).

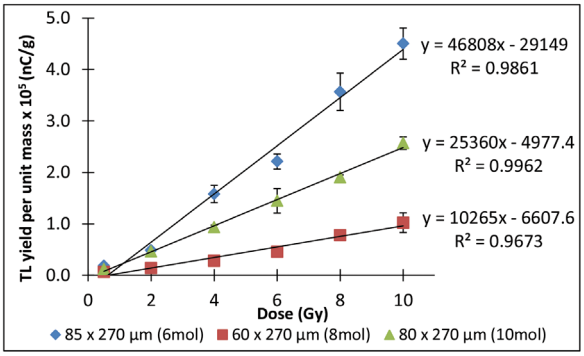


Figure 3. The TL yield of the smaller dimension flat fibres doped with 6, 8 and 10 (mol %) Ge subjected to 6 MV photon irradiation.

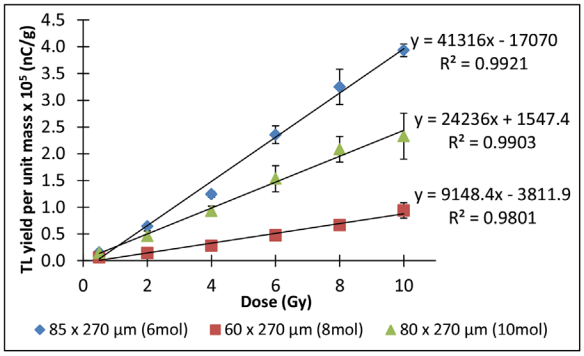


Figure 4. The TL yield of the smaller dimension flat fibres doped with 6, 8 and 10 (mol %) Ge subjected to 10 MV photon irradiation.

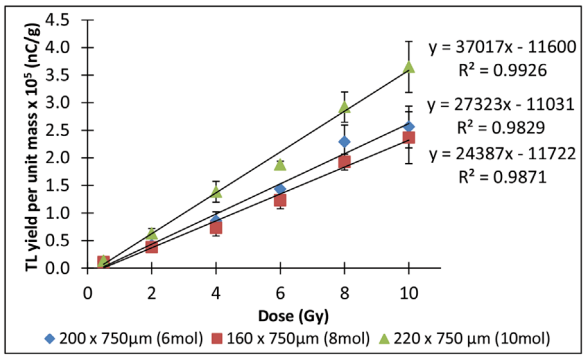


Figure 5. The TL yield of the larger dimension flat fibres doped with 6, 8 and 10 (mol %) Ge subjected to 6 MV photon irradiation.

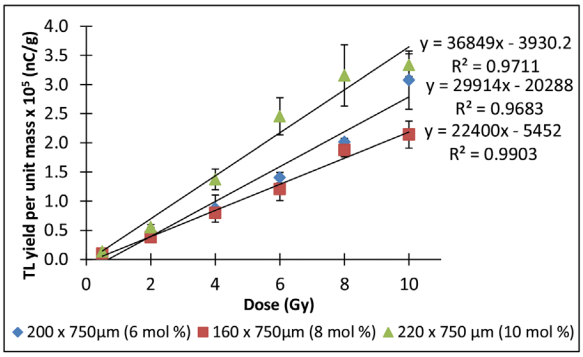


Figure 6. The TL yield of the larger dimension flat fibres doped with 6, 8 and 10 (mol %) Ge subjected to 10 MV photon irradiation.

A better understanding of the more complex situation is to be obtained from glow curve analysis, the flat fibres being represented by a two-peaks system, the first prominent peak observed for temperatures ranging from 201 ºC to 253 ºC while the second peak appears at temperatures ranging from 362 ºC to 390 ºC (Figure 7(a) and (b)), indicative of a deeper traps system resulting from strain defects generation. Glow curve peak parameters of flat fibre for the three Ge concentrations including temperature, channel and TL intensity were obtained as shown in Table 3.

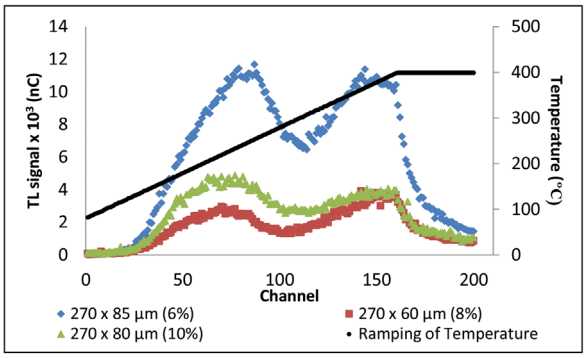


Figure 7(a). Flat fibre (small dimension) TL glow curves for the three different dopant concentrations (6, 8, 10 mol %) following 2 Gy Co-60 irradiation.

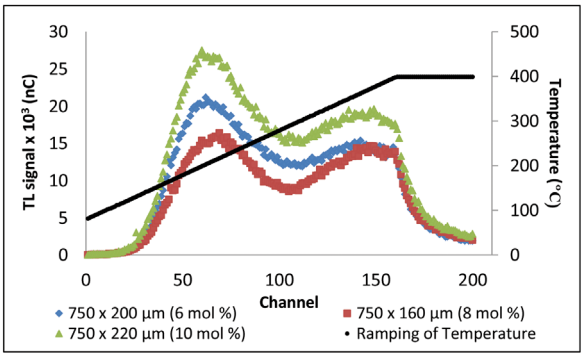


Figure 7(b). Flat fibre (largest dimension) TL glow curves for the three different dopant concentrations (6, 8, 10 mol %) following 2 Gy Co-60 irradiation.

Table 3
Glow curve peak parameters of flat fibre for the three Ge concentrations

Flat fibres (small)	First peak			Second peak		
	Temperature (ºC)	Channel number	TL signal (nC)	Temperature (ºC)	Channel number	TL signal (nC)
270 x 85 µm (6% mol)	253	87	11679	366	144	11386
270 x 60 µm (8% mol)	221	70	2944	378	150	3805
270 x 80 µm (10% mol)	233	77	4865	390	156	4057
Flat fibres (large)	First peak			Second peak		
	Temperature (ºC)	Channel number	TL signal (nC)	Temperature (ºC)	Channel number	TL signal (nC)
750 x 200 µm (6% mol)	203	62	21168	362	142	15381
750 x 160 µm (8% mol)	217	69	16383	372	147	14666
750 x 220 µm (10% mol)	201	61	26886	376	149	19624

CONCLUSIONS

The present study was designed to determine the effect of germanium dopant concentration on TL intensity of flat fibres with the intention of developing new TL dosimeters with optimum performance for radiotherapy dosimetry applications. The evidence from this study suggests that the response in the flat fibres is assisted by strain defects generation, representative of deep-energy defect levels. The findings of this study have a number of important implications for future practices in radiotherapy dosimetry, to be preceded by more detailed investigatations of the factors influencing TL yield.

ACKNOWLEDGEMENTS

This research is supported by Malaysia Fundamental Research Grant Scheme (FRGS) no: 5524789; University of Malaya High Impact Research Grants (UM-HIR) no UM.C/625/1/HIR/33; Universiti Putra Malaysia Geran Putra Incentive Putra Siswazah (UPM-GP-IPS) no 9433970 and A000007-50001 from the Ministry of Education Malaysia. Use of irradiation facilities at the Malaysian Nuclear Agency and University of Malaya Medical Centre are acknowledged. We are indebted to the staff of Telekom Malaysia Research & Development Sdn Bhd, Cyberjaya and the Pulling-Tower Laboratory, University of Malaya for their technical assistance during the sample fabrication and pulling process.

REFERENCES

Alawiah, A., Intan, A. M., Bauk, S., Rashid, H. A. A., Yusoff, Z., Mokhtar, M. R., ... Tamchek, N. (2013). Thermoluminescence characteristics of flat optical fiber in radiation dosimetry under different electron irradiation conditions. In *Proc. SPIE* (Vol. 8775, p. 87750S).

Alawiah, A., Bauk, S., Rashid, H. A. A., Gieszczyk, W., Hashim, S., Mahdiraji, G. A., ... Bradley, D. A. (2015). Potential application of pure silica optical flat fibers for radiation therapy dosimetry. *Radiation Physics and Chemistry*, *106*, 73-76.

Bradley, D. A., Hugtenburg, R. P., Nisbet, A., Rahman, A. T. A., Issa, F., Noor, N. M., & Alalawi, A. (2012). Review of doped silica glass optical fibre: their TL properties and potential applications in radiation therapy dosimetry. *Applied Radiation and Isotopes*, *71*, 2-11.

Bradley, D. A., Mahdiraji, G. A., Ghomeishi, M., Dermosesian, E., Adikan, F. R. M., Rashid, H. A., & Maah, M. J. (2015). Enhancing the radiation dose detection sensitivity of optical fibres. *Applied Radiation and Isotopes*, *100*, 43-49.

Fadzil, M. A., Ramli, N. N. H., Jusoh, M. A., Kadni, T., Bradley, D. A., Ung, N. M., ... Noor, N. M. (2014). Dosimetric characteristics of fabricated silica fibre for postal radiotherapy dose audits. In *Journal of Physics: Conference Series* (Vol. 546, No. 1, p. 012010). IOP Publishing.

Hashim, S., Al-Ahbabi, S., Bradley, D. A., Webb, M., Jeynes, C., Ramli, A. T., & Wagiran, H. (2009). The thermoluminescence response of doped SiO_2 optical fibres subjected to photon and electron irradiations. *Applied Radiation and Isotopes*, *67*(3), 423-427.

Hashim, S., Bradley, D. A., Saripan, M. I., Ramli, A. T., & Wagiran, H. (2010). The thermoluminescence response of doped SiO_2 optical fibres subjected to fast neutrons. *Applied Radiation and Isotopes*, *68*(4), 700-703.

Mahdiraji, G. A., Adikan, F. R. M., & Bradley, D. A. (2015). Collapsed optical fiber: A novel method for improving thermoluminescence response of optical fiber. *Journal of Luminescence*, *161*, 442-447.

Noor, N. M., Hussein, M., Bradley, D. A., & Nisbet, A. (2010). The potential of Ge-doped optical fibre TL dosimetry for 3D verification of high energy IMRT photon beams. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *619*(1), 157-162.

Noor, N. M., Shukor, N. A., Hussein, M., Nisbet, A., & Bradley, D. A. (2012). Comparison of the TL fading characteristics of Ge-doped optical fibres and LiF dosimeters. *Applied Radiation and Isotopes*, *70*(7), 1384-1387.

Noor, N. M., Hussein, M., Kadni, T., Bradley, D. A., & Nisbet, A. (2014). Characterization of Ge-doped optical fibres for MV radiotherapy dosimetry. *Radiation Physics and Chemistry*, *98*, 33-41.

Noor, N. M., Jusoh, M. A., Razis, A. A., Alawiah, A., & Bradley, D. A. (2015). Flat Ge-doped optical fibres for food irradiation dosimetry. *National Physics Conference 2014 (PERFIK 2014)* (p. 100007): AIP Publishing.

Rahman, A. A., Nisbet, A., & Bradley, D. A. (2011). Dose-rate and the reciprocity law: TL response of Ge-doped SiO_2 optical fibers at therapeutic radiation doses. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *652*(1), 891-895.

Rahman, A. A., Hugtenburg, R. P., Sani, S. F. A., Alalawi, A. I. M., Issa, F., Thomas, R., ... Bradley, D. A. (2012). An investigation of the thermoluminescence of Ge-doped SiO_2 optical fibres for application in interface radiation dosimetry. *Applied Radiation and Isotopes*, *70*(7), 1436-1441.

Ramli, N. N. H., Salleh, H., Mahdiraji, G. A., Zulkifli, M. I., Hashim, S., Bradley, D. A., & Noor, N. M. (2015). Characterization of amorphous thermoluminescence dosimeters for patient dose measurement in X-ray diagnostic procedures. *Radiation Physics and Chemistry*, *116*, 130-134.

Pertanika J. Sci. & Technol. 25 (1): 337 - 352 (2017)

SCIENCE & TECHNOLOGY

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

Characterization of Adnexal Masses Using Multidetector Contrast-Enhanced CT Scan – Recognising Common Pitfalls that Masquerade as Ovarian Cancer

Suppiah S[1, 2]*, Kamal SH[2], Mohd Zabid AZ[2] and Abu Hassan H[2]

[1]*Centre for Diagnostic Nuclear Imaging, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

[2]*Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

Adnexal masses are growths that form near the uterus; the majority being ovarian tumours. Although there is no established population-screening tool for detecting ovarian cancer, ultrasound and contrast-enhanced computed tomography (CECT) are useful imaging tools in the management of adnexal masses. Our study aimed to determine the characteristics of malignant adnexal masses on CECT scan and to describe common pitfalls in diagnosis of ovarian cancer when interpreting images. We also determined the sensitivity and specificity of diagnosing ovarian cancer using CECT. A retrospective study was conducted in Hospital Serdang using data from all patients who underwent CECT scan and detected with adnexal masses, and had histopathological examination correlation from January 2013 until January 2015. Out of the 64 cases analysed; the majority of malignant lesions were serous carcinoma of the ovary (40%). The CECT scan characteristics, tumour consistency of mixed type, presence of wall enhancement, septations, ascites and peritoneal nodule/omental caking were significantly associated with ovarian malignancy ($p<0.05$). The sensitivity, specificity, PPV and NPV of CECT scan was 95.45%, 71.43%, 63.63% and 96.77% respectively. Contrast-enhanced computed tomography scan is a good, non-invasive method to diagnose ovarian cancer. By using a pro-forma document as a guide, good results can be achieved to help differentiate between benign and malignant lesions. Nevertheless, caution needs to be exercised in interpreting cases that mimic features of malignancy.

Keywords: Accuracy, computed tomography, diagnosis, ovarian cancer

Article history:
Received: 16 June 2016
Accepted: 20 December 2016

E-mail addresses:
subapriya@upm.edu.my (Suppiah S),
hazirah2907@gmail.com (Kamal SH),
anis94zahirah@gmail.com (Mohd Zabid AZ),
hasyma@upm.edu.my (Abu Hassan H)
*Corresponding Author

INTRODUCTION

Adnexal masses are growths that form on the tissue near the uterus, usually in the ovary or fallopian tube. These include a spectrum of

ISSN: 0128-7680

conditions from gynaecologic and non-gynaecologic sources, which may be benign or malignant in nature. Benign gynaecologic masses include ovarian cysts and ectopic pregnancy, while ovarian carcinoma and fallopian tube carcinomas are examples of malignant gynaecologic masses. The non-gynaecologic masses may include appendicitis and metastatic tumours that originate from other parts of the body such as breast cancer and colon cancer (Givens et al., 2009).

In 2007, the National Central Report (NPC) recorded 658 cases of ovarian cancer, making it among the top four most common cancers affecting Malaysian women. The crude incidence rate (CR) among Malaysian females in 2007 was 6.9 per 100 000 general population and the age-standardised rate (ASR) was 7.8 per 100 000 general population (Ministry of Health Malaysia, 2011).

Ovarian masses are detected during pelvic examination and confirmed using screening tests such as blood test, imaging, and biopsy/surgery for histopathological examination that has been the gold standard in diagnosing ovarian cancer. Initial examination involves gynaecological assessment of the pelvic area to look for any signs of ovarian tumours such as enlarged ovary and ascites. Additional tests such as tumour markers assay and radiological diagnostic imaging may be necessary to aid in the diagnosis. Tumour marker levels e.g. CA125 have been used to predict the presence of malignancy in women with a pelvic mass, monitor response to chemotherapy, and detect possible relapse after initial response to treatment (Granato et al., 2012).

Unfortunately, even though CA125 is elevated (>35 U/ml) in more than 80% of patients with epithelial ovarian cancer, it is only 25% sensitive towards detection of early disease (Togashi, 2003). There are variable results of sensitivity of CA125 to detect ovarian cancer which ranges from 61-90%, and specificity ranging from 35-91% (American College of Obstetricians and Gynecologists, 2007). As a matter of fact, CA125 can be elevated in many other clinical conditions, including both malignant and benign, such as leiomyomas, endometriosis, pregnancy, and pelvic inflammatory diseases, and is even associated with the presence of excessive peritoneal fluid.

Furthermore, the sizes of adnexal lesions are frequently underestimated using abdominal and pelvic examination techniques alone (Padilla et al., 2000). Conversely, semi-invasive and invasive procedures such as peritoneal fluid aspiration and cytology, laparoscopic or computed tomography (CT) guided biopsy of ovarian cancer or explorative laparotomy, debulking surgery and biopsy are some of the definite ways to diagnose ovarian cancer.

Consequently, much effort has been put into developing non-invasive methods for detecting and characterising adnexal tumours. Thus, the advent of sophisticated imaging tools, such as CT scan, MRI and PET/CT scans, has enabled a non-invasive method for the assessment of adnexal masses. Hence, cross sectional diagnostic imaging can facilitate the diagnosis, and characterisation of tumours, as well as guide in disease staging, monitor treatment response and evaluate for suspected cancer recurrence. Ultrasound via trans-abdominal or trans-vaginal approach is the first line imaging of choice when ovarian cancer is suspected. Magnetic resonance imaging gives good soft tissue resolution and has good specificity in detecting and characterising soft tissue cancers, including breast cancers (Suppiah et al., 2013); however, it can be costly and time consuming.

Advantageously, computed tomography (CT) scan can act as a fast, reliable, reproducible and reasonably cost-effective means for characterising adnexal masses and diagnosing ovarian cancer. Contrast-enhanced CT scan is performed when ultrasound findings are equivocal or if the abnormality extends beyond the pelvis or field of view achievable with the endovaginal probe requiring further characterisation (Bennett et al., 2002).

The CT scan enables cross-sectional and three-dimensional imaging of internal organs and structures of the body that helps in the detection and staging of cancer. The advantages of CT scan is that it helps in diagnosing and guiding treatment for a wider range of conditions in an accurate manner. It also provides additional information that can be useful for preoperative evaluation and treatment planning, which includes information on surgical resectability (Santoso et al., 2013). The accuracy for characterising ovarian tumours using CT scan compared with ultrasound is 94% vs. 80% (Santoso et al., 2013) whereby the former also produces better quality images and an improved spatial resolution of the pelvic region anatomy compared with the latter.

Previous studies have assessed the sensitivity and specificity of abdominal CT scan in detecting ovarian cancer in correlation with histopathological examination (HPE) findings (Kubik-Huch et al., 2000 and Sebastian et al., 2008). Sensitivity as high as 100% and specificity of 67% have been reported (Kubik-Huch et al., 2000). Nevertheless, other studies have reported lower sensitivity of 89% and specificity of 59% (Sebastian et al., 2008).

The objective of this study was to determine the sensitivity and specificity of detecting ovarian cancer using contrast-enhanced computed tomography scan and to characterise ovarian cancer subtypes. By using a pro-forma document, we aimed to help reduce common pitfalls that lead to misdiagnosis of adnexal lesions.

METHODS

We conducted a retrospective cross-sectional study in Serdang Hospital, Selangor, a tertiary referral, government hospital institution, after receiving ethical approval from our institutional and national ethical committees (Ethical approval reference: FPSK(EXP15-medic) U035 and NMRR-15-497-25521). We retrospectively recruited all female patients who underwent contrast-enhanced CT scan (CECT) of the abdomen and pelvis between January 2013 and May 2015. Using Serdang Hospital's IT environment software called Total Hospital Information System (T-HIS), we selected Malaysian female patients with adnexal masses who had undergone both CECT of the abdomen and pelvis as well as biopsy with histopathological investigation. We excluded all patients with inconclusive histopathological findings on adnexal lesion and incomplete images of CECT scans.

The CECT scans were performed on a SOMATOM Definition Flash (Siemens Healthcare, Munich, Germany) 128-slice CT scanner. Initial topogram acquisition was followed by abdominal scan performed 60 seconds after intravenous injection of the water soluble low osmolar-iodinated contrast media, by means of an automatic power injector at a rate of 2 – 3 ml/s, 1.5 – 2.0 ml/kg body weight. This was done to acquire porto-venous phase of CECT scans. Image reconstruction of 1 and 5 mm slice thickness and multiplanar reformatting in sagittal and coronal planes were also performed.

Secondary data was accessed from Serdang Hospital using the Centricity Database at Radiology Department and Hospital Information System (HIS). A pro-forma was used to collect information on patients' age, race, parity, CECT findings, and biopsy results (please refer to Appendix A). In the Centricity application of hospital PACS system, present researchers retrieved CECT scan images that fulfilled the study inclusion criteria and removed image overlay that identified the patient. All patients were assigned a code number and the relevant scan images were saved onto a CD. Two experienced radiologists, SS and HAH, were blinded to the patients' information and reviewed the images separately.

Subsequently, SS and HAH filled up the designated pro-forma document, with information regarding the features of the CECT scan images of the study patient, by consensus. We then traced the histological examination (HPE) results and correlated it with CT scan findings. All information pertaining to patient identity was omitted from the pro-forma document as well as the final data sheet for the purpose of protecting patient privacy and for data protection.

Social Package for Social Science (SPSS) version 22 was used for data analysis. The study used p-value of less than 0.05 ($p<0.05$) and 95% confidence interval (CI 95%) as statistically significant. Chi-Square test or test was used to measure the association between the adnexal lesions with age groups and ethnic groups. We used Fisher's exact test to evaluate the relationship between tumour consistency, presence of septation, omental caking or peritoneal thickening, and presence of ascites with the diagnosis of malignancy. We also calculated the sensitivity, specificity, PPV and NPV of CECT in diagnosing ovarian cancer using histopathological findings as a gold standard.

RESULTS AND DISCUSSION

We accessed approximately 1450 CT scans of patients undergoing CT abdomen and pelvis during the study period, January 2013 - May 2015. We identified 121 cases with adnexal masses. However, only 64 cases were suitable for final data analysis because they had retrievable histopathological examination (HPE) results. Based on histopathological results, we identified 22 malignant and 42 benign lesions.

Socio-demographic and descriptive analysis

The subjects of this study were aged between 14 and 76 (mean age 46.80 ±15.19). Previous study found that the mean age of 104 patients with clinically or sonographically detected complex adnexal masses was 50 years (Sohaib et al., 2003). Based on Erikson's psychosocial stages (Rosenthal et al., 1981), we divided the patients into five broad age categories: adolescence (13-19years), adult (20 – 39years), middle aged (40 – 64years), and elderly (65 years above). The majority of the study population were middle-aged women (55%, n=35). The adolescence age group was the minority (3%, n=2). The patients with benign lesions ranged in age between 14 and 76 years (mean age 47.0 ±13.4). The age of patients with malignant lesions ranged between 15 and 74 years (mean age 50.55 ±17.85). Although there wasn't a statistically significant association between age and malignancy, there was evidence to support that risk of malignancy increases with age (Sohaib et al., 2003).

Our study population consisted of ethnic Malays, Chinese and Indians among whom Malays topped the list with 44 cases (68.8 %) followed by Chinese with 13 cases (20.3%), Indians with 5 cases (7.8%) and other ethnic minorities with 2 cases. Sixteen Malay patients (72.7%) had malignant lesions. There was no significant association between adnexal mass malignancy and age or race.

Majority of patients with adnexal tumours were nulliparous (25%). There was a balanced distribution of cases in parous patients and a marked reduction in the percentage of cases in grand multiparous women having more than 5 children. In addition, majority of the patients with malignancy were nulliparous women (27.3%). Majority of the patients were pre-menopausal women (59.4 %) whereby the latter had higher incidence of benign lesions (66.7%) compared with the post-menopausal group which had a higher incidence of malignant lesions (54.2%). Although it has been postulated that increased parity due to reduced ovulation, and thus reduced ovarian surface epithelium damage, has a protective role against ovarian cancer (Fleming et al., 2006), there was no significant association between parity, menstrual status and malignancy in this study.

Based on WHO classification, epithelial tumours account for about 75% of all ovarian tumours, and 90-95% of ovarian malignancies (Lee-Jones, 2011). Approximately 70% of epithelial tumours are high grade serous type and 5% are low grade serous carcinomas, while endometrioid and clear cell carcinomas follow with 10% each, and mucinous types account for 3% of cases (Santoso et al., 2013). Transitional or Brenner tumour, mixed and undifferentiated tumours are very rare and make up remainder of the cancer subtypes. Sex cord-stromal tumours account for about 5-10% of all ovarian neoplasms. Germ cell tumours account for about 15-20% of all ovarian neoplasms. Metastatic tumours, accounting for about 5% of ovarian malignancies, usually originate from breast, colon, endometrium, stomach and cervical cancers. There are also other types of neoplasms which develop from ovarian soft tissue or non-neoplastic processes.

This study identified 42 (66%) cases of benign lesions and 22 cases of (34%) malignant lesions. Among the benign adnexal lesion cases, 28 cases (67 %) were ovarian in origin, confirmed by HPE results. There were 14 cases (33%) of non-ovarian benign cases whereby 12 were uterine in origin (86%, n=12) while only one each originated from the fallopian tube (7%, n=1) and bowel (7%, n=1). These were mostly endometriosis lesions (79%, n=11) while the rest were enteric retention cyst, fallopian tube cyst and fibroma, each represented by one case. On the other hand, benign lesions of ovarian origin were categorised as physiological follicles, cysts, leiomyoma, mucinous cystadenoma, serous cystadenoma and teratoma based on HPE results. Cysts were the most common of all benign ovarian lesions encountered in this study (29%, n=8) while leiomyoma and serous cystadenoma accounted for only one case each.

There were 22 cases (34%) of malignancy based on HPE results and out of these 7 (32%) were non-ovarian lesions that included uterine in origin (23%, n=5) and metastatic lesions (9%, n=2). There were 15 cases (68%) of malignant lesions having ovarian origins. Malignant non-ovarian lesions were grouped into metastatic uterine adenocarcinoma (28.6%, n=2) and endometrioid adenocarcinoma (71.4%, n=5). There were 15 primary ovarian cancers (Table 1) that were mostly serous adenocarcinoma (40%, n=6). The two least common cases were malignant mixed germ cell tumours and adenocarcinoma each represented by one case.

Table 1
Distribution of Malignant Ovarian Lesions by HPE Results. (N=15)

Malignant ovarian HPE results	No of cases	Percentage (%)
Dysgerminoma	2	13
Clear cell carcinoma	3	20
Malignant mixed germ cell tumour	1	7
Serous carcinoma	6	40
Adenocarcinoma	1	7
Seromucinous carcinoma	2	13
Total	15	100

Common characteristics of adnexal masses detected on abdominal CECT scan

Certain types of adnexal mass characteristics were highlighted in the pro-forma document (please refer to Appendix). These were frequently reported findings in cases of ovarian malignancy, which included the consistency of the lesion, wall enhancement, loculation, septation, sizes of mass, presence of ascites and also presence of peritoneal nodule or omental caking (Jung et al., 2002).

Consistency of adnexal lesions

There was a significant association between the types of consistency of the adnexal lesions and benign or malignant results based on HPE ($p < 0.001$). The majority of cystic lesions (95.8%) were confirmed to be benign. These lesions were purely cystic, having no evidence of solid components, internal septations or papillary projections (Figure 1). Conversely, the mixed solid-cystic type of tumours was mostly malignant (55.9%). Majority of lesions identified by this study were mixed solid-cystic type which could be due to the centre receiving more complex cases as it is a tertiary referral centre. However, this finding differs from previous studies (Slanetz et al., 1997 and Chang et al., 2006) in which most of the CT scan findings were predominantly cystic in consistency. This is due to the higher prevalence of the epithelial type of ovarian tumours in the general population; which more often than not consistent of the cystic type. On the other hand, the mixed solid-cystic types of tumours are more commonly observed in metastatic diseases (Brown et al., 2001).

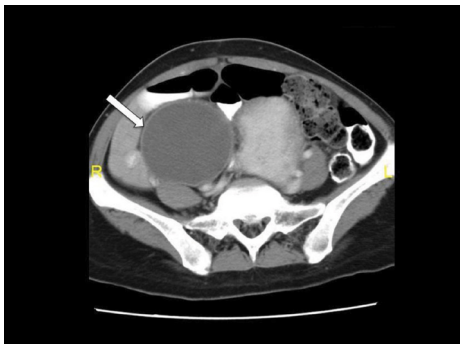


Figure 1. Simple, purely cystic, unilocular ovarian cyst, consistent with a benign cyst (white arrow). Note the absence of internal septations or solid components which are usually associated with complex lesions.

Wall enhancement of adnexal lesions

There was a significant association between lack of wall enhancement and benign lesions. Adnexal lesions that did not have wall enhancement were interpreted as benign and the researchers correctly identified 92.9% of benign cases. Only one malignant lesion lacked wall enhancement, i.e. low grade serous carcinoma of the ovary. Generally, most adnexal lesions demonstrated wall enhancement; in addition, mild wall enhancement was also seen in 49% of benign tumours. This can be attributed to physiological uptake in women of reproductive age.

Type of loculation of adnexal lesions

The study identified 30 cases of unilocular tumours in which 27 (76.7%) were benign in nature. However, there was also an almost equal distribution of benign and malignant cases that showed multiloculations. Therefore, there was no statistically significant association between types of loculation with HPE. Findings by Brown et al. (2001) are consistent with findings of the present study as they also reported an even distribution of both types of loculations.

Presence of septation in adnexal lesions

The present study found majority of the adnexal lesions had presence of septations (70.3%). This finding was supported by an earlier study by deSouza et al. (2005) where it reported majority of their cases (84%) presented with septations. They stated that septations are more prominent in invasive tumours compared with borderline tumours. The majority (93.3%) of the cases that lacked septations, were confirmed to be benign. However, in lesions that demonstrated presence of septations, 42.2% were malignant lesions and the rest were benign lesions. Fisher's exact test identified, a significant association between septation and HPE i.e. absence of septations indicated the lesion was likely to be benign.

Size of adnexal lesions

Simple, unilocular, cystic adnexal lesions that are small in size (< 5 cm), tend to run a benign course and have a low risk of undergoing malignant change (Ekerhovd et al., 2001). Therefore, the study grouped the lesions into small (less than 5cm in diameter); and large (more than 5 cm in diameter) in the pro-forma document (please refer to Appendix A). It was noted that the adnexal lesions measuring less than 5cm were mostly histopathologically confirmed benign lesions (66.7%). Contradictory to our expectations, the majority (65.3%) of the bigger lesions with sizes more than 5cm were also determined to be mostly benign by HPE. Therefore, we did not detect any significant association between sizes of adnexal masses and HPE.

Presence of peritoneal nodule or omental caking

On CT scan images, peritoneal implantation are reported when nodular, plaque-like, or infiltrative soft tissue lesions with abnormal enhancement are seen in the peritoneal fat or

on the peritoneal surface. Omental invasion is diagnosed when there is a nodular pattern; infiltrative or feathery pattern; or cake-like appearance of enhanced soft tissue in the omentum (Kitajima et al., 2008 and Coakley et al., 2002). Based on the present study, the cases that had peritoneal nodule or omental caking were mostly primary malignant (75%) and metastatic. In cases that did not demonstrate peritoneal nodule or omental caking, the majority (65.6%) of them were benign lesions. There were 3 cases (25%) of false positive findings i.e. peritoneal nodule identified in benign conditions, such as in endometriosis, as this may be caused by reactive inflammation of the peritoneum or even endometriotic deposits in the peritoneal lining (Figure 2). Nevertheless, there was a statistically significant association between presence of peritoneal nodule and malignant HPE.

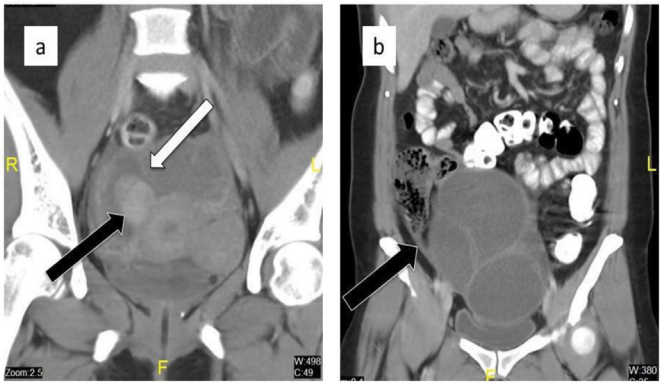


Figure 2. Commonest mimicker of ovarian cancer is ovarian endometriosis. (a) An endometrioma demonstrating mixed solid-cystic consistency (black arrow), internal septations (white arrow) and wall enhancement. (b) Presence of peritoneal nodules, which are likely due to deposits of endometriosis (black arrow).

Presence of ascites

There was an equal chance for benign and malignant lesions to have presence of ascites. Nevertheless, the majority (76.3%) of benign lesions did not have concurrent ascites whereas most malignant cases had presence of ascites. A study done Jung et al. (2002) also stated that the findings of lesion with ascites increase diagnostic confidence for malignancy. There was a significant association between ascites and HPE in the present study i.e. absence of ascites correlates with benign diagnosis.

The sensitivity, specificity, positive predictive value, negative predictive value of CT scan to detect ovarian cancer

Based on our pro-forma document, the sensitivity and specificity of contrast-enhanced CT scan in detecting ovarian cancer is 95.5% and 71.4% respectively. The positive predictive value (PPV) of CT scan is 63.6% whereas the negative predictive value (NPV) 96.8%. The slightly lower specificity was because due to a high rate of detection of false positive results

(12 cases). The commonest mimickers of malignancy were adnexal endometriosis and ovarian endometriomas (5 cases), followed by developmental follicular cysts - some having haemorrhagic component (4 cases) and the rest were mucinous cystadenoma, leiomyoma of the ovary, and infected ileum retention cyst which were represented by one case respectively (Figure 3). The high number of false positive findings was attributed to the reviewers assigning the lesions to be malignant when the lesions were indeterminate, thus having a high tendency to over diagnose malignancy. We had one false negative result that was confirmed to be an ovarian borderline seromucinous tumour with microinvasion.

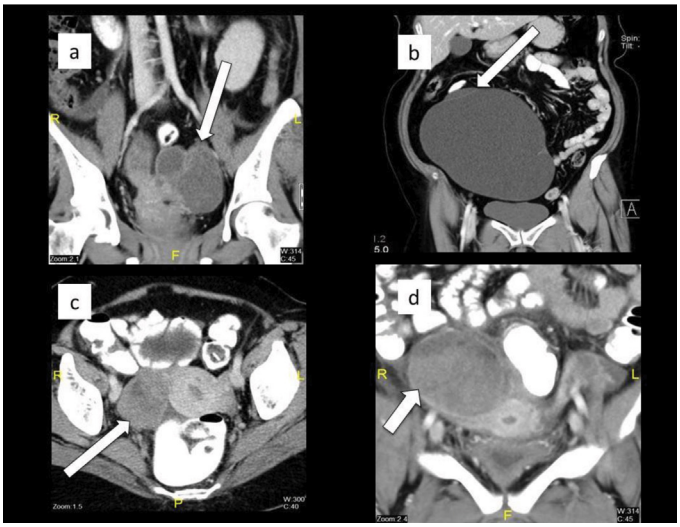


Figure 3. Adnexal lesions that masquerade as adnexal malignancy, having mixed solid-cystic consistency and internal enhancing components (white arrows). (a) Coronal view CECT Abdomen showing a haemorrhagic follicular cyst in the left adnexa. (b) Coronal view CECT Abdomen showing a large mucinous cystadenoma of the ovary. (c) Axial view CECT Abdomen showing an ovarian leiomyoma. (d)Coronal view CECT Abdomen showing an infected ileum retention cyst in the right adnexa.

Our findings on the sensitivity and specificity of CT in detecting ovarian malignancy correlated well with other studies. In general, the range for sensitivity of CT scan in detecting ovarian cancer is 73 - 85% while for specificity, it is 81 - 94% (Iyer et al., 2010). Moreover, the reported pre-operative staging accuracy of CT is 70%–90% while the reported sensitivity and specificity of CT performed before second-look surgery ranges from 59%–83% and 83%–88% respectively (Kawamoto et al., 1999). A study by Fischerova and Burgetova (2014) also stated that sensitivity of CT scan was 79% while the specificity 84%.

In the present study, the sensitivity is slightly higher compared with other studies as described earlier. This is maybe due to improved resolution in current multidetector CT scanners and the usage of pro-forma document. Classical-appearing benign and malignant adnexal lesions were successfully identified and accurately characterised using the pro-forma document as a guide (please refer to Appendix A). Nevertheless, the researchers also encountered truly negative (Figure 4), truly positive (Figure 5), falsely positive (Figure 6), falsely negative (Figure 7) as well as metastatic lesions involving bilateral ovaries (Figure 8).

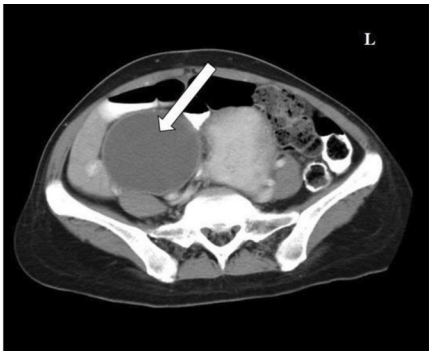


Figure 4. CECT Abdomen scan in axial view showing a benign-looking cystic lesion (white arrow) with no solid component, no septation and no concurrent ascites. HPE confirmed it to be a benign cyst in keeping with a true negative finding.

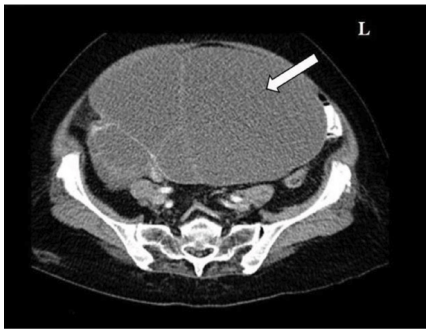


Figure 5. CECT scan showing a mixed solid cystic mass (white arrow) with septations and wall enhancement associated with peritoneal nodule that was suggestive of malignancy. HPE result confirmed that it is a malignant lesion in keeping with a true positive finding.

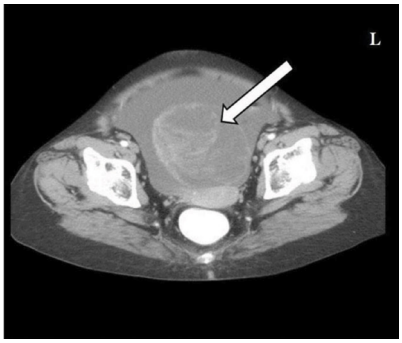


Figure 6. CECT scan showing a suspicious looking enhancing, mixed solid cystic mass (white arrow) with presence of ascites. However, HPE the result was benign endometrial cyst, in keeping with a false positive finding i.e. mimicker of malignancy.

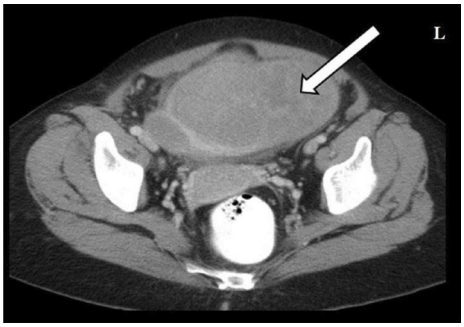


Figure 7. CECT scan showing a benign-looking cystic adnexal mass (white arrow). There was no solid component, abnormal enhancement pattern, no peritoneal nodule nor presence of ascites. However, HPE result showed it to be a borderline seromucinous tumour with microinvasion, in keeping with a false negative finding.

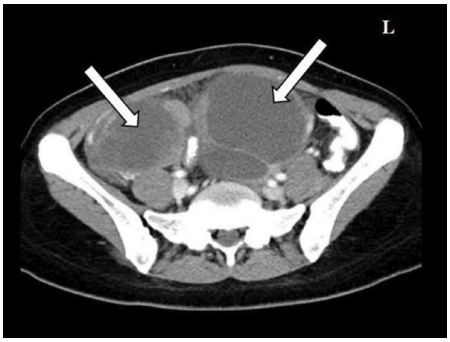


*Figure 8***.** CECT scan showing a mixed solid cystic lesion, multiloculated, septated bilateral adnexal masses (white arrows). This is an example of metastatic malignant lesions.

We noted that the commonest type of lesion masquerading as ovarian cancer is ovarian endometriomas. It is understandably a challenge to interpret these cases, as inflammatory changes occurring in endometriosis often mimic changes seen in malignancy. These changes are namely presence of solid components and septations within the cysts. Additionally, presence of concomitant peritoneal deposits of endometriosis and pelvic ascites tend to be misleading signs of potential malignancy. Therefore, we recommend correlation to be made based on clinical history, physical examination, and biochemical tests when faced with indeterminate lesions on imaging.

CONCLUSION

Contrast-enhanced computed tomography scan is a reliable, non-invasive, imaging tool which has good sensitivity but variable specificity for characterising adnexal masses and diagnosing ovarian cancer. The usage of our pro-forma document aided in the diagnosis by elucidating parameters significantly associated with ovarian cancer i.e. mixed solid-cystic consistency, presence of wall enhancement, septations, concurrent ascites and presence of peritoneal nodules.

Nevertheless, reviewers ought to be cautious of mimickers of malignancy such as ovarian endometriosis, which require further clinical correlation and may necessitate more advanced imaging for further assessment.

LIMITATION AND RECOMMENDATION

This is a retrospective, hospital-based study which may not be applicable to patients who do not come to hospital for screening of ovarian cancer although they are in the high-risk group. There may be some selection bias in terms of study population. Additionally, this research relied on secondary data and did not have control over data quality.

We recommend the use of our pro-forma document to correlate the CT scan characteristics with tumour marker levels. In this way, we can improve upon the specificity of the CT scan thus, avoiding much more invasive procedures and aid in accurate and early detection of malignancy.

Nevertheless, we note that although it has high sensitivity for detecting ovarian cancer, the interpretation of CT scan characteristics lack specificity due to similar features noted in benign conditions such as ovarian endometriosis that masquerade as malignant lesions. Thus, the utility of advanced imaging tools such as positron emission tomography / computed tomography (PET/CT) and positron emission tomography/ magnetic resonance (PET/MR) imaging may be considered in indeterminate cases.

ACKNOWLEDGEMENTS

We would like to thank the Dean and Deputy Dean of Academic, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia for facilitating this study. We also thank the Director General of Health, Ministry of Health Malaysia and the Director of Serdang Hospital, Serdang for giving us permission to conduct our study and access the hospital database.

REFERENCES

ACOG. (2007). American College of Obstetricians and Gynecologists: Practice Bulletin no. 83: Management of Adnexal Masses. *Obstet Gynecol, 110*(1), 201-214. Retrieved January 5, 2016, from http://www.ncbi.nlm.nih.gov/pubmed/17601923

Bennett, G., Slywotzky, C., & Giovanniello, G. (2002). Gynecologic Causes of Acute Pelvic Pain: Spectrum of CT Findings. *Radiographics*, *22*(4), 785-801.

Brown, D., Zou, K., Tempany, C., Frates, M., Silverman, S., McNeil, B., & Kurtz, A. (2001). Primary versus Secondary Ovarian Malignancy: Imaging Findings of Adnexal Masses in the Radiology Diagnostic Oncology Group Study. *Radiology, 219*(1), 213-218.

Chang, W., Meux, M., Yeh, B., Qayyum, A., Joe, B., Chen, L., & Coakley, F. (2006). CT and MRI of Adnexal Masses in Patients with Primary Non-ovarian Malignancy. *American Journal of Roentgenology, 186*(4), 1039-1045.

Coakley, V., Choi, F. H., Gougoutas, P. A., Pothuri, C., Venkatraman, B., & Chi, D. (2002). Peritoneal Metastases: Detection with Spiral CT in Patients with Ovarian Cancer. *Radiology, 223*(2), 495-499.

DeSouza, N., O'Neill, R., McIndoe, G., Dina, R., & Soutter, W. (2005). Borderline Tumors of the Ovary: CT and MRI Features and Tumor Markers in Differentiation from Stage I Disease. *American Journal of Roentgenology, 184*(3), 999-1003.

Ekerhovd, E., Wienerroith, H., Staudach, A., & Granberg, S. (2001) Preoperative assessment of unilocular adnexal cysts by transvaginal ultrasonography: a comparison between ultrasonographic morphologic imaging and histopathologic diagnosis. *American Journal of Obstetrics & Gynecology, 184*(2), 48–54.

Fischerova, D., & Burgetova, A. (2014) Imaging Techniques for the Evaluation of Ovarian Cancer. *Best Practic & Research Clinical Obstetrics & Gynaecology, 28*(5), 697-720.

Fleming, J., Beaugié, C., Haviv, I., Trench, G. C., & Tan, O. (2006). Incessant ovulation, inflammation and epithelial ovarian carcinogenesis: Revisiting old hypotheses. *Molecular and Cellular Endocrinology, 247*(1-2), 4-21.

Givens, V., Mitchell, G., Smith, C. H., Reddy, A., & Maness, D. (2009). Diagnosis and Management of Adnexal Masses. *American Family Physician*. Retrieved April 12, 2015, from http://www.aafp.org/afp/2009/1015/p815.html

Granato, T., Midulla, C., Longo, F., Colaprisca, B., Frati, L., & Anastasi, E. (2012). Role of HE4, CA72.4, and CA125 in monitoring ovarian cancer. *Tumour Biology, 33*(5), 1335-1339.

Iyer, V. R., & Lee, S. I. (2010). MRI, CT, and PET/CT for ovarian cancer detection and adnexal lesion characterization. *American Journal of Roentgenology, 194*(2), 311–321.

Jones, L. L. (2011). Ovarian tumours: An overview. *Atlas of Genetics and Cytogenetics in Oncology and Haematology*. Retrieved April 12, 2015, from http://hdl.handle.net/2042/38082. doi:10.4267/2042/38082

Jung, S., Lee, J., Rha, S., Byun, J., Jung, J., & Hahn, S. (2002). CT and MR Imaging of Ovarian Tumours with Emphasis on Differential Diagnosis. *Radiographics, 22*(6), 1305-1325.

Kawamoto, S., Urban, B. A., & Fishman, E. K. (1999) CT of Epithelial Ovarian Tumors. *Radiographics, 19*(suppl_1), S85-102.

Kitajima, K., Murakami, K., Yamasaki, E., Kaji, Y., Fukasawa, I., Inaba, N., & Sugimura, K. (2008). Diagnostic accuracy of integrated FDG-PET/contrast-enhanced CT in staging ovarian cancer: Comparison with enhanced CT. *European Journal of Nuclear Medicine and Molecular Imaging, 35*(10), 1912-1920.

Kubik-Huch, R. A., Dörffler, W., Von Schulthess, G. K., Marincek, B., Köchli, O. R., Seifert, B., ... Steinert, H. C. (2000). Value of (18F)-FDG positron emission tomography, computed tomography, and magnetic resonance imaging in diagnosing primary and recurrent ovarian carcinoma. *European radiology, 10*(5), 761-767.

Ministry of Health Malaysia. (2011). *National Cancer Registry Report* (p. 48). Putrajaya: National Cancer Registry, Ministry of Health, Malaysia. Retrieved April 12, 2015, from http://www.care.upm.edu.my/dokumen/13603_NCR2007.pdf]

Padilla, L. A., Radosevich, D. M., & Milad, M. P. (2000) Accuracy of the pelvic examination in detecting adnexal masses. *Obstetrics & Gynecology, 96*(4), 593-598.

Rosenthal, D. A., Gurney, R. M., & Moore, S. M. (1981). From trust on intimacy: A new inventory for examining Erikson's stages of psychosocial development. *Journal of Youth and Adolescence, 10*(6), 525 -537.

Santoso, J., Robinson, A., Suganda, S., Praservit, S., Wan, J., & Ueland, F. (2013). Computed tomography adnexal mass score to estimate risk for ovarian cancer. *Archives of Gynecology and Obstetrics, 289*(3), 595-600.

Sebastian, S., Lee, S., Horowitz, N., Scott, J., Fischman, A., Simeone, J., & Han, P. (2007). PET–CT vs. CT alone in ovarian cancer recurrence. *Abdominal Imaging, 33*(1), 112-118.

Slanetz, P., Hahn, P., Hall, D., & Mueller, P. (1997). The frequency and significance of adnexal lesions incidentally revealed by CT. *American Journal of Roentgenology, 168*(3), 647-650.

Sohaib, S., Sahdev, A., Trappen, P., Jacobs, I., & Reznek, R. (2003). Characterization of Adnexal Mass Lesions on MR Imaging. *American Journal of Roentgenology, 180*(5), 1297-1304.

Suppiah, S., Rahmat, K., Shah, M. N. M., Azlan, C. A., Tan, L. K., Aziz, Y. F., & Yip, C. H. (2013) Improved diagnostic accuracy in differentiating malignant and benign lesions using single-voxel proton MRS of the breast at 3 T MRI. *Clinical Radiology, 68*(9), e502-510.

Togashi, K. (2003). Ovarian cancer: The Clinical Role of US, CT, and MRI. *European Radiology, 3*(4), L87-104.

APPENDIX A

Pro-forma Document

Assigned ID : _____

Age : _____

Race : _____

Date of birth : _____

CA125 value : _____

Imaging date- Computed tomography (CT) : _____

Biopsy/Operation date : _____

Diagnosis of histopathological examination (HPE) : _____

Item	Imaging features	Yes	No
1.	Consistency		
	Solid component		
	Cystic		
	Mixed solid cystic		
2.	Wall enhancement		
3.	Loculation		
	Uniloculated		
	Multiloculated		
4.	Large size of mass (greater than 5cm)		
5.	Septations		
6.	Peritoneal nodule / omental caking		
7.	Ascites		

*please tick the relevant option (✓)

Pertanika J. Sci. & Technol. 25 (1): 353 - 358 (2017)

SCIENCE & TECHNOLOGY

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

Botox Injections to the Face: A Mimic of Richter's Transformation

Shafik-Eid, R[1]*, Nandurkar, D[2] and Ramdave, S[2]

[1]*St. Vincent's Hospital, 41 Victoria Parade, Fitzroy, 3065 Melbourne, Australia*

[2]*Monash Imaging, Monash Medical Centre, 246 Clayton Road, Clayton, 3168 Melbourne, Australia*

ABSTRACT

Chronic Lymphocytic Lymphoma (CLL) may transform into an aggressive high grade variant in approximately 4-6% of cases known as Richter's Transformation (RT). Cutaneous lymphoma as well as other high-grade lymphomas may be a rare presentation of RT. As minimally invasive cosmetic procedures increase in popularity, we present an important mimicker of cutaneous uptake of the lower face simulating RT in a patient with CLL, after the injection of collagen fillers and botulinum toxin A (Botox) to the lower face. To our knowledge there are no previous studies on this important observation. Reading physicians should be cognisant of this masquerade to prevent misdiagnosis.

Keywords: Botox, mimicker, PET/CT, Richter's Transformation

INTRODUCTION

Minimally invasive cosmetic procedures that include injection of Botox and soft tissue fillers (including collagen based fillers and Hyaluronic acid) for rejuvenation of the face are increasing in popularity.

The PET/CT scan remains the most sensitive method for detecting high-grade transformation of chronic low-grade lymphoma. We present a unique case of uptake within the face on PET/CT in a 58 year old woman with a one-year history of Chronic Lymphocytic Leukaemia (CLL) who recently underwent cosmetic injections into the face mimicking Richter's Transformation (RT). To our knowledge this is the first case of facial cosmetic injections mimicking high-grade transformation of lymphoma. This study describes the findings and presents features to aid readers to differentiate between inflammatory changes related to cosmetic injections and those related to true high-grade transformation of lymphoma to prevent misdiagnosis that may result in inappropriate treatments.

Article history:
Received: 16 June 2016
Accepted: 20 December 2016

E-mail addresses:
rayshafikeid@gmail.com (Shafik-Eid, R),
Nandurkar.Dee@monashhealth.org (Nandurkar, D),
Shakher.Ramdave@monashhealth.org (Ramdave, S)
*Corresponding Author

ISSN: 0128-7680

CASE REPORT

A 58-year old woman with a one-year history of CLL was referred as an outpatient to our institution for PET/CT. Significant past history voluntarily disclosed by the patient included coronary bypass 10 years prior to presentation and L4-L5 disc prolapse treated conservatively, nil else. The request for PET/CT scan was to investigate a recent onset of unexplained fatigue and specifically exclude high-grade transformation of CLL. Clinical examination of the patient revealed prominent lymph nodes in the left posterior triangle of the neck, which the patient self-reported as remaining unchanged over the last 12 months. No new lymphadenopathy was identified. Fasting blood sugar level was 3.8 mmol/L. The patient was injected with 305 MBq of F-18 FDG and underwent imaging after 81 minutes of uptake. Imaging was obtained on a Siemens Biograph PET/CT over 5 bed positions, with 3-minute acquisitions per position. Interrogation of the images demonstrated moderate to intense (SUV Max 7.00) symmetrical uptake within the soft tissue and skin of the face at the infra-orbital ridges, and symmetric uptake within the soft tissues along the mandible bilaterally extending to the soft tissues anterior to the symphysis mentis. These findings were on a background of low-grade uptake (average SUV Max 2.4) within bilateral deep cervical lymph nodes, bilateral axillary, mediastinal, subcarinal, and bilateral inguinal lymph node stations. Remaining visceral and osseous uptake was within physiologic limits. Due to the isolated symmetric nature of the uptake within the face and in light of the previous cosmetic surgery (breast implantation evidenced by CT images), the patient was subsequently questioned regarding any recent cosmetic procedures or instrumentation to the face. The patient then disclosed having had injections of soft tissue fillers and Botox seven days earlier in the exact distribution as seen on the metabolic images. This was discussed with the referring physician who also confirmed the patient's surgical history.

Table 1
Features Common to RT and Mimic (Surgical Inflammation)

Features common to both RT and Mimic (Surgical Inflammation)
SUV Max > 5.0
Involve extra-nodal soft tissues

Table 2
Differences between RT and Mimic (Surgical Inflammation)

High-grade Transformation	**Mimic (Surgical Inflammation)**
Asymmetric Uptake	Symmetric uptake within the face
High grade uptake within high-volume nodal tissue	Predictable distribution within the soft tissues of the face in low volume nodal areas.
Possible visceral or osseous extra-nodal uptake	No visceral or extra nodal uptake (uptake isolated to the face)

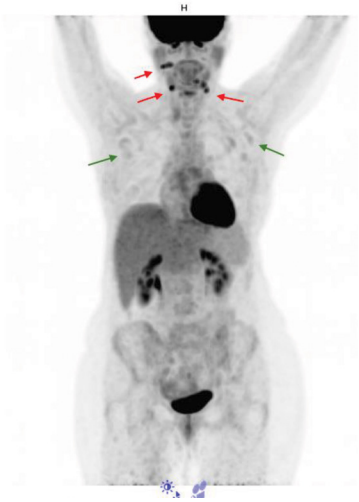


Figure 1. Antero-posterior and lateral whole Body Maximum Intensity Projection (MIP) images of a 56 year old women with CLL, being investigated for RT. MIP image demonstrate multiple foci of moderate to intense FDG avidity identified at the infra-orbital ridges, right greater than left (red arrows) and along the mandible and symphysis mentis (pre-jowl distribution)[1]. Low grade FDG uptake is seen within the deep cervical and axillary nodal stations in keeping with the patient's CLL (green arrows)

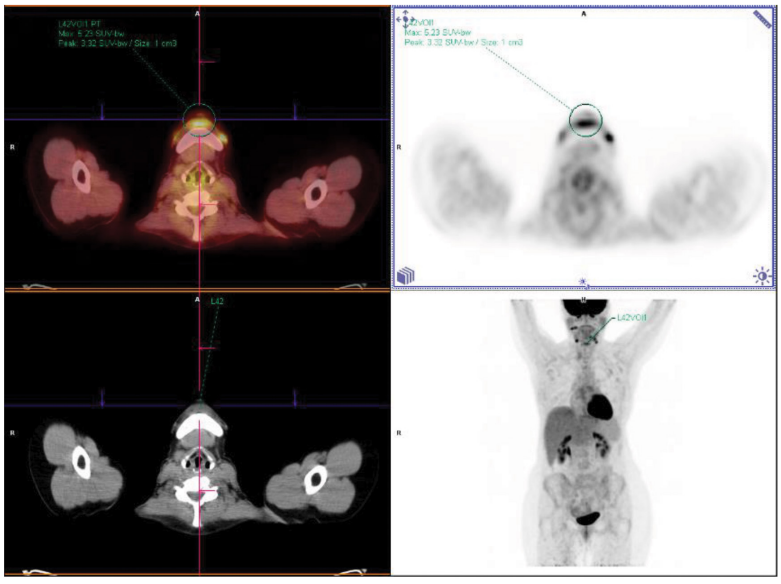


Figure 2. Fused PET/CT images of the lower face. Further symmetrical focal moderate to intense uptake within the soft tissue anterior to the ramus and body of the mandible. Focal uptake is also seen anterior to the symphysis mentis. This region is also a target for injectable therapy to counteract the pre-jowl sag which contributes to aging of the face[1-5]. The area demonstrates similar moderate to intense FDG avidity to the infra-orbital ridges, again exceeding an SUV max of 5.0. All areas were confirmed to be the result of injections performed seven days prior to the study after obtaining a thorough history of thepatient.

DISCUSSION

The number of minimally invasive cosmetic procedures are increasing worldwide. In the United States Botox and soft tissue filler injections (STFI) are the two most common procedures for facial rejuvenation with approximately 9 million procedures performed by board-certified plastic surgeons in 2014 (Plastics Surgery Statistics, 2015). In light of this trend, radiologists and nuclear medicine physicians need to be aware and remain cognisant of the potential for misdiagnosis to prevent inappropriate treatments. The PET/CT scan remains the most accurate method of diagnosing transformation of lymphoma from indolent, low-grade subtypes to high-grade malignancies[2]. The RT is seen in 4-6% of all patients with CLL (Robak et al., 2005; Bruzzi et al., 2006; Yilmaz et al., 2014). Using PET/CT images alone, RT is diagnosed when intense metabolic activity is seen within nodal or extra-nodal tissue, with an SUV Max of greater than 5 relatively specific for high grade transformation (Bruzzi et al., 2006). The diagnosis of RT has important implications for patient management in that more aggressive chemotherapy regiments are needed. Biopsy is generally mandatory for diagnosis when it is highly suspect on PET/CT images (Bruzzi et al., 2006). In our specific case, biopsy (a follow-up of PET/CT) was deemed unethical once the patient disclosed a clear surgical history and the uptake following a distribution consistent with the procedure (Kane, 2005; Wise and Greco, 2006; Fattabi, 2007; Niamtu, 2009; Schierle and Casas, 2011). Nodal and extra nodal uptake in cases of RT can be highly varied in terms of distribution. Typically, intense uptake is seen within pre-existing low grade nodal or extra nodal disease (Robak et al., 2005; Bruzzi et al., 2006; Yilmaz et al., 2014). As seen in our case however, a symmetric pattern of isolated uptake, especially in areas with low volume lymphoid tissue such as the soft tissues anterior to the inferior and supra-orbital ridge, forehead, lateral canthus of the eyes and along the jaw-line, generally point to inflammatory changes typical of the procedure which should prompt the reporting physician or radiologist of this important mimicker and prevent misdiagnosis.

REFERENCES

ASPS. (2015). Plastics Surgery Statistics. *American Society of Plastic Surgeons*. Retrieved November, 2015, from http://www.plasticsurgery.org/news/plastic-surgery-statistics/2014-statistics.html

Bruzzi, J. F., Macapinlac, H., Tsimberidou, A. M., Truong, M. T., Keating, M. J., Marom, E. M. & Munden, R. F. (2006). Detection of Richter's Transformation of Chronic Lymphocytic Leukemia by PET/CT. *Journal of Nuclear Medicine, 47*(8), 1267-1273.

Fattahi, T. (2008). The Prejowl Sulcus: An Important Consideration in Lower Face Rejuvenation. *Journal Oral of Maxillofacial Surgery, 66*(2), 355-358.

Kane, M. A. (2005). Botox injections for lower facial rejuvenation. *Oral and Maxillofacial Surgery Clinics of North America, 17*(1), 41-49.

Niamtu, J. (2009). Complications in fillers and Botox. *Oral and Maxillofacial Surgery Clinics of North America, 21*(1), 13-21.

Robak. E., Tybor, J. G., Kordek, R., Wawrzyniak, E., Bartkowiak, J., Bednarek, A., … Robak, T. (2005). Richter syndrome first manifesting as cutaneous B-cell lymphoma clonally distinct from primary B-cell chronic lymphocytic leukaemia. *British Journal of Dermatology, 153*(4), 833-7.

Schierle, C. F., & Casas, L. A. (2011). Non-surgical rejuvenation of the ageing face with injectable Poly-L-Lactic acid for restoration of soft tissue volume. *Aesthetic Surgery Journal*, *31*(1), 95-109.

Wise, J. B., & Greco, T. (2006). Injectable treatments for the ageing face. *Facial Plastic Surgery*, *22*(2), 140 – 146.

Yılmaz, S., Ozhan, M., Asa, S., Sağer, M. S., Biricik, F. S., Halaç, M., & Sonmezoğlu, K. (2014). Detection of Hodgkin's Transformation in a case of Chronic Lymphocytic Leukemia by PET/CT. *Molecular Imaging & Radionuclide Therapy*, *23*(2), 67-69.

Pertanika J. Sci. & Technol. 25 (1): 359 - 370 (2017)

SCIENCE & TECHNOLOGY

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

One-Time Prolonged Ultrasound Exposure during Early Pregnancy Affects Bone Strength in Young Aged *Oryctolagus Cuniculus*

Che Isa, I. N.*, Md Dom, S., Abdul Razak, H. R. and Hashim, U. F.

Department of Medical Imaging, Faculty of Health Sciences, Universiti Teknologi MARA, 42300 UiTM, Puncak Alam, Selangor, Malaysia

ABSTRACT

The use of prenatal ultrasound has become controversial as it is increasingly being performed for business and social interests rather than for medical use. This nonmedical use of the modality has violated the US FDA guideline. Ultrasound scans have been proven to increase temperature in insonated tissue and their effects have been investigated via phantom and animal experiments. Absorption coefficient of the bone is the highest compared with any other structure. Thus, exposure to ultrasound, especially during osteogenesis, can cause significant damage to developing foetus. Twenty-two pregnant does of known gestation were enrolled in the control and experimental groups. No exposure was given to the control group while the experimental groups were exposed accordingly to the prenatal ultrasound in the 1st, 2nd and 3rd stage for 30, 60 and 90 minutes respectively. A total of 142 subjects aged between 1 and 5 months were analysed for bone strength. The Tb.Th of the experimental group was reduced significantly as compared to the control group. Po, TMD and empty lacunae were higher in the experimental group. It is thus concluded that one-time prenatal ultrasound can affect bone strength in young subjects.

Keywords: Bone histology, bone morphology, bone strength, prenatal ultrasound, tissue mineral density, young age

INTRODUCTION

In recent times, it has become a trend for expectant ladies to go for ultrasound scans even during normal pregnancy to fullfil their social needs. Scans make them feel reassured about their pregnancy and knowing sex of their child-to-be gives them satisfaction (Lumley, 1990; Beck Black, 1992; Molander et al., 2010). The images and videos are kept as mementoes. Modern ultrasound imaging for diagnostic purposes has a wide range of

Article history:
Received: 16 June 2016
Accepted: 20 December 2016

E-mail addresses:
zawaniisa@yahoo.com (Che Isa, I. N.),
sulaimanmd@salam.uitm.edu.my (Md Dom, S.),
hairil@salam.uitm.edu.my (Abdul Razak, H. R.),
ummifaihananhashim@yahoo.com (Hashim, U. F.)
*Corresponding Author

ISSN: 0128-7680

applications. However, private doctors often recommend parents to perform scans to boost their healthcare business purpose. Hence, most pregnant women undergo up to 20 scans per pregnancy (Bashour et al., 2005; Gammeltoft & Nguyen, 2007).

According to The United States Food and Drug Administration (US FDA), diagnostic ultrasound for pregnancy is considered safe when it is performed for valid medical reasons whereby its benefits outweigh risks. The FDA prohibits any institutions to sell, promote or apply the ultrasound to make keepsake images or videos (US Food and Drug Administration, 2013). Several laboratory studies have proven that ultrasound exposure during pregnancy can have adverse effects on the tissue. Energy is deposited in the tissue by ultrasound absorption and thus diagnostic ultrasound cannot be considered completely safe. The effects of rising temperature during ultrasound scanning in animal model are well established, in particular, its effects on birth weight, organ weight, brain and bone. Vella et al. (2003) and Wu et al. (1995) have exposed phantoms that mimic biological tissue in diagnostic ultrasound. They found that ultrasound exposure can lead to temperature increase, particularly in the bone. Bone is sensitive to heat as it has high absorption coefficient (10 dB/cm/MHz) (Barnett et al., 1997) which is 30 times higher than any other structure.

According to Gent (1997), an adult bone is able to absorb 60% or more of ultrasound energy while the absorption is less in foetal bone. However, the absorption coefficient of the foetal bone differs depending on gestational age where there are changes in heat capacity, mineralisation and density (Drewniak et al., 1989; Shankar & Pagel, 2011). Ultrasound intensity and duration of exposure are two main parameters that lead to a linear rise in tissue temperature. In addition, beam angle to the bone, location of transducer, and the orientation of thermocouples in the tissue are able to indirectly increase the tissue temperature.

Our aim in this study was to investigate the effects of prolonged ultrasound exposure during pregnancy on bone status in young rabbits.

MATERIALS AND METHODS

Prior to the investigation, all procedures performed on the animals were approved by Universiti Teknologi MARA Animal Research and Ethics (UiTM CARE) Committee. Female (doe) and male (buck) Malaysian breed New Zealand white rabbits (*Oryctolagus cuniculus)* aged between five to eight months were enrolled in this study. In order to allow the use of 22 pregnant does of known gestation, the rabbits were time-mated. Four does were used as control and 18 were assigned in experimental groups. The control group was allowed to have full term delivery without exposure to the ultrasound while the experimental group was given a one-time exposure to ultrasound. The full term gestational period for a doe is between 30 and 33 days (Kaplan & Timmons, 1979) and divided into three stages namely as 1st, 2nd and 3rd gestational periods. The rabbits were given ultrasound exposure of 30, 60 and 90 minutes respectively using a 2-D B-mode Philips HD3 ultrasound system with a 5-9 MHz broadband high-resolution linear array transducer (L9-5, Philips Electronics E.V., Germany). The transducer had a focal depth of approximately 5.5 cm, and directed to where the foetuses were located. During the exposure, the recorded mechanical index (MI) and thermal index (TI) were 1.0 and 0.2 respectively. The output power and spatial peak temporal average (I_{SPTA}) were varied from 0.4 W to 0.7

W and 0.13 to 0.19 W/cm^2 respectively as calculated based on the previous characterisation of the transducer (Ahmad Zaiki et al., 2013). The exposure factors remained constant for all exposures. 'My Rabbit Burrow,' a rabbit restrainer designed by Dom (Md. Dom et al., 2012) was utilised to keep the does calm and cooperative during the scanning.

The offsprings were taken as a subject when they reached 1 and 5 month old. All rabbits (bucks, does and offsprings) were kept under the same laboratory environment. They were given ad lib water supply, and pelleted feed which was measured at 5 % of their body weight. The animal house was set for 16 hours: 8-hour light and dark cycle with temperature between 14ºC – 28°C (Matics et al., 2013). BioGS air purifier was utilised to provide proper ventilation from the harmful gases released by the rabbits such as carbon dioxide and ammonia (Lebas, 1997). A total of 142 subjects were analysed for bone strength.

The subjects were euthanised by administering a dose of ketamine hydrochloride (50 mg/kg body weight) and xylazine hydrochloride (10 mg/kg body weight). Femoral bone was dissected using a scalpel. The Skyscan 1176 (Skyscan 1176, SkyScan bvba, Aartselaar, Belgium) was used to provide data on bone morphology in terms of trabecular thickness (Tb. Th), trabecular separation (Tb.Sp), tissue mineral density (TMD) and porosity (Po) (Che Isa et al., 2015). After micro-CT scanning, the femurs were processed for histological examination and subsequently decalcified using hydrochloric acid- (HCL) based decalcifying agent. The paraffin-embedded specimens were sectioned 5 µm (Smith et al., 2001) in a longitudinal orientation and stained with hematoxylin and eosin (H&E). The slide was observed for empty lacunae using a light microscope (Primo Star iLED™, Carl Zeiss Microimaging, Oberkochen, Germany). An illumination system of a commercial video-microscope (AM413T5 Dino-Lite Pro, AnMo Electronics Corporation, Hsinchu, Taiwan) was utilised to obtain microscopic pictures. The percentage of empty lacunae was calculated from the number of empty lacunae compared with the total number of lacunae counted in each slide.

Data from both the experimental group and control group were compared. Analysis of variance (ANOVA) was used to analyse the data using Statistical Package for the Social Sciences (SPSS) version 21.0. All differences were assumed statistically significant at $p \leq 0.05$.

RESULT AND DISCUSSION

Morphological and histological data of 1 and 5 months subjects is presented in Table 1. Ultrasound exposure has the potential to raise temperatures in insonated tissues. As osteogenesis is very sensitive to heat, a temperature rise can lead to denaturation of the enzymatic and membrane proteins (Augustin et al., 2012), microcirculation blockage, bone tissue necrosis and activation of bone marrow macrophages (Yoshida et al., 2009). Prenatal ultrasound exposure in the 1st stage has caused Tb.Th to reduce significantly from the control ($p<0.05$) among 1 month subjects (Figure 1). However, in 5 months subjects, Tb.Th was reduced in the 2nd stage (Figure 2). In terms of Tb.Sp, no difference was noted in 1 month subjects (Figure 3) but in 5 months subjects, Tb.Sp increased in the 2nd stage after 90 minutes of exposure (Figure 4). Cortical bone region was likely affected by the ultrasound exposure. A significant increase in porosity was found in both 1 month (Figure 5) and 5 month subjects (Figure 6) especially in 1st and 2nd stage. In addition, 1 (Figure 7) and 5 month old subjects

(Figure 8) in the experimental group showed higher TMD compared with the control group. Dom et al. (2012) observed the heat-induced changes in rabbit foetal mineral density. The bone strength can be altered by poor status of the cortical bone and likely lead to impaired resistance to fractures.

Table 1
In-house normal range for metaphyseal bone analysis based on control group data

Group	Parameters	($\bar{x}$)	SD
1 month old	Tb.Th (μm)	1.44	0.2
	Tb.Sp (μm)	3.42	0.49
	Po (%)	0.83	0.16
	TMD ($g.cm^{-3}$)	16.42	3.01
	EL (%)	6.09	1.01
5 months old	Tb.Th (μm)	2.00	0.15
	Tb.Sp (μm)	4.23	0.72
	Po (%)	0.06	0.02
	TMD ($g.cm^{-3}$)	40.16	7.84
	EL (%)	6.17	1.04

Abbreviations: $\bar{x}$, mean; SD, standard deviation; S.E.M, standard error mean; Tb.Th, trabecular thickness; Tb.Sp, trabecular separation; Po, porosity; TMD, tissue mineral density; EL, empty lacuna

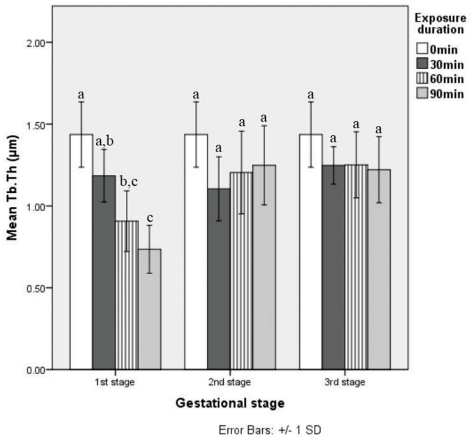


Figure 1. Graph of mean trabecular thickness (Tb.Th) of 1 month old subjects. Data in each stage sharing same alphabets is not significant ($p>0.05$)

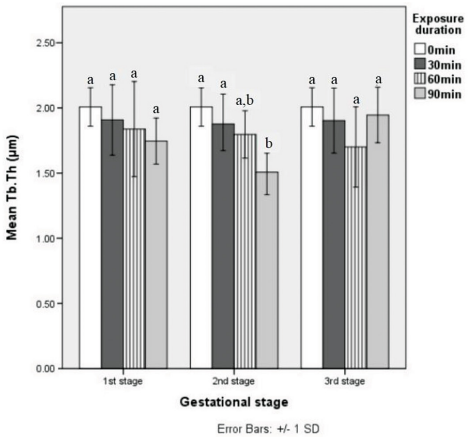


Figure 2. Graph of mean trabecular thickness (Tb.Th) of 5 month old subjects. Data in each stage sharing same alphabets is not significant (p>0.05)

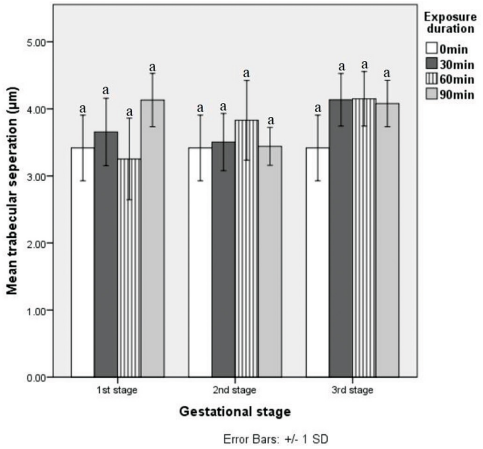


Figure 3. Graph of mean trabecular separation (Tb.Sp) of 1 month old subjects. Data in each stage sharing same alphabets is not significant (p>0.05)

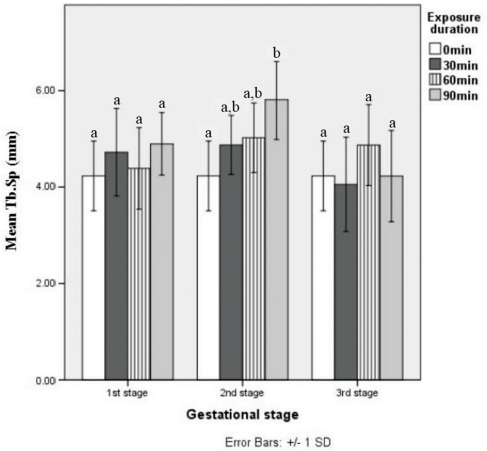


Figure 4. Graph of mean trabecular separation (Tb.Sp) of 5 month old subjects. Data in each stage sharing same alphabets is not significant (p>0.05)

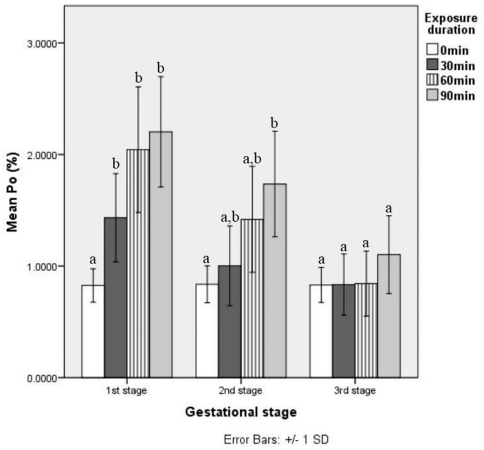


Figure 5. Graph of mean total porosity (Po) of 1 month old subjects. Data in each stage sharing same alphabets is not significant (p>0.05)

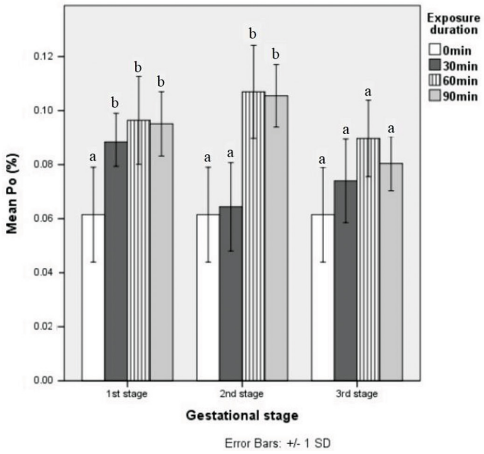


Figure 6. Graph of mean total porosity (Po) of 5 month old subjects. Data in each stage sharing same alphabets is not significant (p>0.05)

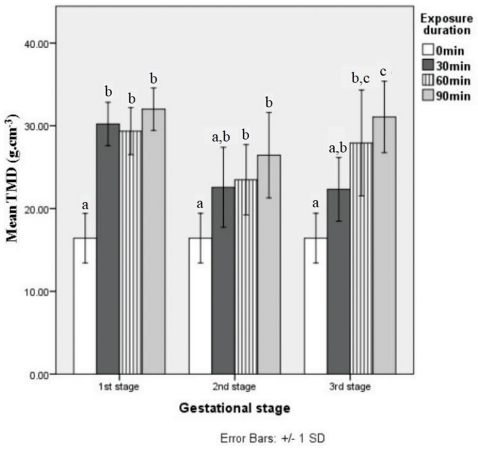


Figure 7. Graph of tissue mineral density (TMD) of 1 month old subjects. Data in each stage sharing same alphabets is not significant (p>0.05)

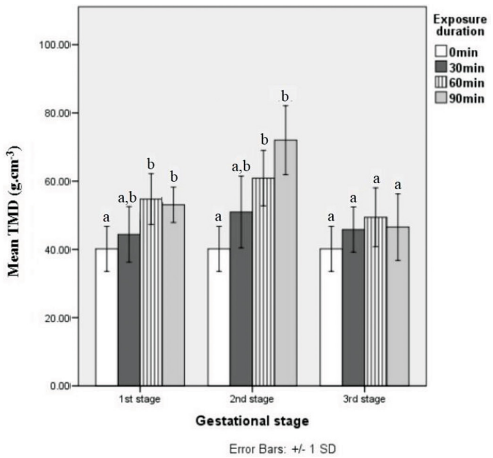


Figure 8. Graph of tissue mineral density (TMD) of 5 month old subjects. Data in each stage sharing same alphabets is not significant ($p>0.05$)

In order to determine the temperature rise expected in an ultrasound examination during pregnancy, it is important to consider bone density in the foetus. In this study, the formation of a major structure, including the bone, was disrupted as the exposures were given during organogenesis period. Bones in the early stage of pregnancy (1st stage of the gestational period) has a high ability to absorb heat even though it is in osteoid form and not yet ossified. This is mainly because the osteoid tissue is rich in collagen (National Council on Radiation Protection Measurements, 1992). On the other hand, ultrasound-induced heating has a direct effect on bone density during 2nd and 3rd stage of the gestational period as it is thicker and denser.

The anatomic and amount distribution of cortical and trabecular bone in the femoral region might be a major indication in determining resistance to fracture. This anatomic distribution is crucial for the mechanical performance of the bone as a whole. As indicated by decreased Tb.Th and, increased Tb.Sp, Po and TMD in 1 and 5 month subjects, the ultrasound exposure has led to poor bone strength.

The empty lacunae was found higher in 2nd and 3rd stage after a 90-minute exposure in 1 month (Figure 9) and 5 month old subjects (Figure 10). No histological damage attributable to ultrasound was noted in a previous investigation based on thermal damage caused by low-intensity pulsed ultrasound (Shimazaki et al., 2010). However, another study demonstrated osteocyte damage and necrosis characterised by empty lacunae and pyknotic cells following focused ultrasound energy (Smith et al., 2001). The authors observed that acoustic power levels of 39, 52 and 65 W had caused osteocytes loss to be pronounced and had suggested that longer ultrasound exposure could possibly cause more severe and irreparable damage.

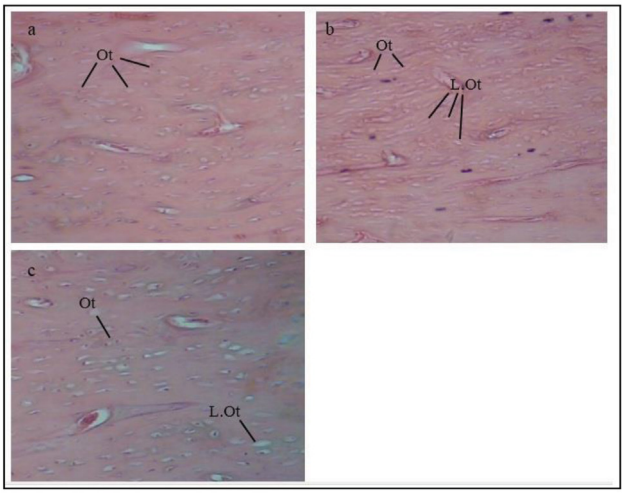


Figure 9. Histological results in 1 month old subjects. (a) 90 min of exposure in the 1st gestational stage, (b) 90 min of exposure in the 2nd gestational stage, (c) 90 min of exposure in the 3rd stage. (Ot=lacunae filled with osteocytes, L.Ot=empty lacunae)

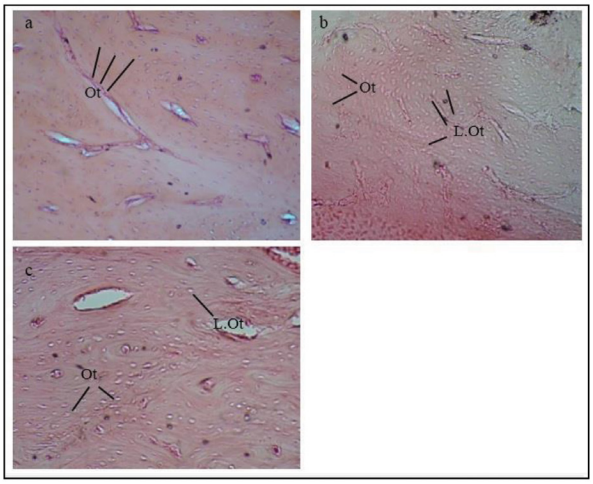


Figure 10. Bone histological images of 5 month old subjects. (a) 90 min of exposure in the 1st gestational stage, (b) 90 min of exposure in the 2nd gestational stage, (c) 90 min of exposure in the 3rd stage. (Ot=lacunae filled with osteocytes, L.Ot=empty lacunae)

As osteocytes are potential regulators of bone modelling and remodelling (Lanyon, 1993), bone fragility may be increased due to reduction in osteocyte cells viability. As a result of minimal short-term effects of radiation-induced cell death, loss of osteocytes is likely to affect bone strength in short term, and could cause fractures in the long term (Power et al., 2001). Loss of osteocytes in the femoral neck might include local vasculature disruption thus affecting the nutrient supply to the adjacent bone cell population (Power et al., 2001). Low rate of remodelling could also be one of the effects of osteocytes loss in bone tissue (Wand et al., 1992).

Bone strength is dependent on quality, quantity, and balanced remodelling of the bone tissue. In a normal condition, mammalian bone undergoes continuous remodelling process where formation of new bones takes place replace older ones. This remodeling process is balanced in young adults and becomes imbalanced as a result of ageing. While it is known that bone architecture changes according to aging, there are other factors, such as social activities, nutrition, and co-morbidities, that affect bone metabolism, independent of direct mechanical stimuli (Parkinson & Fazzalari, 2013). In this study, disruption during osteogenesis caused by ultrasound exposure has affected bone metabolism of the subjects. Even though bone can repair the damage, it was not observed in this study. The damage in the subjects' bone was at a faster rate than can be repaired by normal bone remodelling. As a result, bone becomes more fragile and loses its mass.

CONCLUSION

This study has proven that prolonged exposure to ultrasound during early intrauterine life can cause disturbance in osteogenesis. The effects last until after birth leading to poor bone strength and increasing bone fragility.

ACKNOWLEDGEMENTS

We wish to thank product specialists from Crest NanoSolutions for their dedicated work, and the staff of Medical Lab Technology Department, Universiti Teknologi MARA (UiTM), Puncak Alam Campus, in providing the and facilities for this study.

REFERENCES

Augustin, G., Zigman, T., Davila, S., Udilljak, T., Staroveski, T., Brezak, D., & Babic, S. (2012). Cortical bone drilling and thermal osteonecrosis. *Clinical Biomechanics, 27*(4), 313-325.

Barnett, S. B., Rott, H. D., Haar, G. R. T., Ziskin, M. C., & Maeda, K. (1997). The sensitivity of biological tissue to ultrasound. *Ultrasound in Medicine and Biology, 23*(6), 805-812.

Bashour, H., Hafez, R., & Abdulsalam, A. (2005). Syrian women's perceptions and experiences of ultrasound screening in pregnancy: implications for antenatal policy. *Reprod Health Matters, 13*(25), 147-154.

Black, R. B. (1992). Seeing the baby: The impact of ultrasound technology. *Journal of Genetic Counseling, 1*(1), 45-54.

Dom, S. M., Salikin, M. S., Hassan, H. F., & Yusoff, N. M. (2012). The effect of B-mode diagnostic ultrasound exposure on rabbit foetal bone mineral density (BMD). *Radiography, 18*(3), 197-200.

Drewniak, J. L., Carnes, K. I., & Dunn, F. (1989). In vitro ultrasonic heating of fetal bone. *Journal of the Acoustical Society of America, 86*(4), 1254-1258.

Gammeltoft, T., & Nguyen, H. T. (2007). The commodification of obstetric ultrasound scanning in Hanoi, Viet Nam. *Reprod Health Matters, 15*(29), 163-171.

Gent, R. (1997). *Applied Physics and Technology of Diagnostic Ultrasound*. Milner Publishing.

Isa, I. N. C., Dom, S. M., & Hashim, U. F. (2015). Bone Analysis of Young Adult Rabbit Femur via Micro-Computed Tomography. *Research Updates in medical Sciences, 3*(3), 3-9.

Kaplan, H. M., & Timmons, E. H. (1979). *The rabbit: a model for the principles of mammalian physiology and surgery*. Academic Press.

Lanyon, L. E. (1993). Osteocytes, strain detection, bone modeling and remodeling. *Calcified Tissue International, 53*(Suppl 1), S102-106; discussion S106-107.

Lebas, F. (1997). *The Rabbit: husbandry, health, and production*. Food and Agriculture Organization of the United Nations.

Lumley, J. (1990). Through a Glass Darkly: Ultrasound and Prenatal Bonding. *Birth, 17*(4), 214-217.

Matics, Z., Gerencsér, Z., Radnai, I., Zotte, A. D., Palumbo, M., Mikó, A., ... Szendrő, Z. (2013). Effect of different lighting schedules (16L:8D or 12L:6D) on reproductive performance and nursing behaviour of rabbit does. *Livestock Science, 157*(2–3), 545-551.

Molander, E., Alehagen, S., & Bertero, C. M. (2010). Routine ultrasound examination during pregnancy: a world of possibilities. *Midwifery, 26*(1), 18-26.

National Council on Radiation Protection Measurements. (1992). *Heat Generation by Ultrasound in Mammalian Tissues Exposure Criteria for Medical Diagnostic Ultrasound: I, Criteria Based on Thermal Mechanisms : Recommendations of the National Council on Radiation Protection and Measurements*. The Council.

Parkinson, I. H., & Fazzalari, N. L. (2013). Characterisation of Trabecular Bone Structure. In J. M. Silva (Ed.), *Skeletal Aging and Osteoporosis: Biomechanics and Mechanobiology* (pp. 31-51). Berlin, Heidelberg: Springer Berlin Heidelberg.

Power, J., Noble, B. S., Loveridge, N., Bell, K. L., Rushton, N., & Reeve, J. (2001). Osteocyte lacunar occupancy in the femoral neck cortex: an association with cortical remodeling in hip fracture cases and controls. *Calcified Tissue International, 69*(1), 13-19.

Shankar, H., & Pagel, P. S. (2011). Potential adverse ultrasound-related biological effects: a critical review. *Anesthesiology, 115*(5), 1109-1124.

Shimazaki, A., Inui, K., Azuma, Y., Nishimura, N., & Yamano, Y. (2000). Low-intensity pulsed ultrasound accelerates bone maturation in distraction osteogenesis in rabbits. *Journal of Bone & Joint Surgery, British Volume, 82-B*(7), 1077-1082.

Smith, N. B., Temkin, J. M., Shapiro, F., & Hynynen, K. (2001). Thermal effects of focused ultrasound energy on bone tissue. *Ultrasound in Medicine and Biology, 27*(10), 1427-1433.

US Food and Drug Administration. (2013). *Fetal Keepsake Videos*. Retrieved April 23, 2014, from http://www.fda.gov/medicaldevices/safety/alertsandnotices/patientalerts/ucm064756.htm

Vella, G. J., Humphrey, V. F., Duck, F. A., & Barnett, S. B. (2003). Ultrasound-induced heating in a foetal skull bone phantom and its dependence on beam width and perfusion. *Ultrasound in Medicine and Biology, 29*(6), 779-788.

Wand, J. S., Smith, T., Green, J. R., Hesp, R., Bradbeer, J. N., & Reeve, J. (1992). Whole-body and site-specific bone remodelling in patients with previous femoral fractures: relationships between reduced physical activity, reduced bone mass and increased bone resorption. *Clinical Science (London, England: 1979), 83*(6), 665-675.

Wu, J., Cubberley, F., Gormley, G., & Szabo, T. L. (1995). Temperature rise generated by diagnostic ultrasound in a transcranial phantom. *Ultrasound in Medicine and Biology, 21*(4), 561-568.

Yoshida, K., Uoshima, K., Oda, K., & Maeda, T. (2009). Influence of heat stress to matrix on bone formation. *Clinical Oral Implants Research, 20*(8), 782-790.

Zaiki, F. W. A., Dom, S. M., Razak, H. R. A., & Hassan, H. F. (2013). Prenatal ultrasound heating impacts on fluctuations in haematological analysis of Oryctolagus cuniculus. *Quantitative Imaging in Medicine and Surgery, 3*(5), 262-268.

Pertanika J. Sci. & Technol. 25 (1): 371 - 378 (2017)

SCIENCE & TECHNOLOGY

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

Popliteal Artery to Tibial Plateau Distance at the Knee Level: A Radiological Study to Assess Injury Risks in Osteoarthritic Knees Using Dual Source Dual Energy CT Scan

Ezamin, A. R.[1]*, Hasyma, A. H.[1], Suppiah, S.[1], Suraini, M. S.[1], Arifaizad, A.[2], Paisal, Hussin.[2], Nasir, M. Nizlan.[2], Sidique, S. F.[3] and Hariati, J.[4]

[1]*Department of Imaging, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43600 UPM, Serdang, Selangor, Malaysia*

[2]*Department of Orthopaedic, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43600 UPM, Serdang, Selangor, Malaysia*

[3]*Faculty of Economics and Management, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

[4]*Putrajaya Hospital, Precinct 7, 62250 Putrajaya, Malaysia*

ABSTRACT

Popliteal artery injury is the most disastrous intraoperative complication during total knee replacement. This study aims to determine the mean distance between the popliteal artery (PA) and the tibial plateau in normal and osteoarthritic patients who underwent Dual Energy CT Angiography (CTA) of the lower limb. Materials and Methods: All CTA lower limb examinations from January 2013 to October 2014 were retrospectively reviewed. The distance between the PA the tibial plateau distance and the thickness of popliteus muscle were electronically measured. We used modified Kellgren and Lawrence's Classification to grade the osteoarthritis in patients who underwent CT examinations regardless of symptoms. Results: There were a total of 126 patients who underwent CTA (93 males and 33 females). 54 of them were Malays, 47 Indians, and 24 Chinese. The mean age of patients was 58 years (range 16 to 92). The mean PA-to-tibial plateau distance was 9.9 mm for the right lower limb (range 2.5 mm to 17.2 mm) and 10.24 mm for the left (range 5.5 mm to 15.4 mm). There were no significant correlations between PA-to-tibial plateau distance with osteoarthritis grade, age, gender, and racial origin ($P > 0.05$); however, there was a positive correlation between PA-to-tibial plateau distances with popliteus muscle thickness ($P = 0.000$). Conclusion: Osteoarthritic condition in the knee does not reduce the popliteal artery to the tibial plateau distance. Hence, a higher osteoarthritic grade does not impose additional risks with regards to popliteal artery to tibial plateau distance, with relatively similar arterial injury risks compared to normal knees.

Keywords: Radiology, popliteal artery, osteoarthritic

Article history:
Received: 16 June 2016
Accepted: 20 December 2016

E-mail addresses:
drezahar@gmail.com (Ezamin, A. R.),
hasyma@upm.edu.my (Hasyma, A. H.),
subapriya@upm.edu.my (Suppiah, S.),
surainimd@yahoo.com.my (Suraini, M. S.),
aripmd@gmail.com (Arifaizad, A.),
shaufique@gmail.com (Sidique, S. F.),
hariati@gmail.com (Hariati, J.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Popliteal artery injury is a disastrous intraoperative complication when doing total knee replacement (TKR). The incidence rate of popliteal artery injury following TKR is rare; up to 0.57% based on a recent study using Nationwide Inpatient Sample (NIS) of 1,120,508 in the United States from 1998 to 2009 (Ko et al., 2014). This study also stated that the rate of popliteal artery injury was greatly increased during the revision of the surgery; however, the causes for the popliteal artery injury were not revealed.

Another smaller scale study suggested that of arterial injury patients post-total hip and knee arthroplasty, 50% tend to take legal action later. This should be an eye opener to the operating surgeon dealing with such cases (Parvizi et al., 2008). Several mechanisms have been proposed as the cause of popliteal artery injury perioperatıvely: atheromatous plaque embolism, thrombosis as a consequence of intimal tears and slow flow induced by tourniquet release of contracture or traction. Thus leading to intimal tear or compression of the artery against bony or hard structure, and lastly iatrogenic injury caused by instruments, traction, or implants (Butt et al., 2010). Some authors suggested that the artery moves further posteriorly by knee flexion (Shiomi et al., 2001; Yoo et al., 2009; Mitsuhiro et al., 2011). It has become a popular practice for the surgeons to flex the knee 90 degrees, presumably to avoid arterial complications. However, it is still debatable whether flexion of the knee will be a protection from direct arterial complication during the knee surgery. Other authors suggested popliteal artery is not always away from the tibia during flexion (Zaidi et al., 1995; Smith et al., 1999; Shetty et al., 2003; Eriksson et al., 2010). CT angiography (CTA) allows precise mapping of the arterial component and soft tissue of the bilateral lower limbs eloquently. Our study has measured the distance of the popliteal artery and the tibial plateau of normal and osteoarthritic (OA) patients who underwent routine CTA of the lower limb.

MATERIAL AND METHODS

All lower limb CTA images from January 2013 to October 2014 were retrospectively reviewed. Incomplete data or poor images were excluded in this study. Standard local CTA of the lower limb protocol was used in this study, using CTA with 128-DECT (dual energy) CT scanner (Somatom Sensation Flash, Siemens Medical Solutions). Consent for CTA was taken in each contrast study in our facility prior to the CTA examination. CT scanning was done in supine position with the knee extended. Images were retrieved from the local database server (PACS) and viewed using BARCO diagnostic display monitor and using GE- centricity version 2.0-2004 software (General Electric software).

We measured the shortest distance of PA to tibial plateau and the popliteus muscle thickness in axial 5 x 5 mm slice thickness with knee at 0-degree flexion using electronic callipers GE Centricity version 2.0 (Fig 1). The measurements at each parameter were taken three times.

The average measurement was taken as the final measurement. For osteoarthritis grading, we used the coronal view 5 mm slice perpendicular to the axial plane and also AP projection of the topography. We used Kellgren and Lawrence's grading scale for osteoarthritis as the basis of Osteoarthritis (OA) grading. Grade 0: no radiographic features of OA are present; grade 1: doubtful joint space narrowing (JSN) and possible osteophytic lipping; grade 2: definite osteophytes and possible JSN on anteroposterior weight-bearing radiograph; grade 3: multiple osteophytes, definite JSN, sclerosis, possible bony deformity; grade 4: large osteophytes, marked JSN, severe sclerosis and definite bony deformity (Kellgren & Lawrence, 1957). Poor quality images with poor visualisation of the popliteal artery due to any other reasons were excluded in our study. These data were then compiled into the Microsoft excel spreadsheet and later analysed using SPSS version 20. Spearman correlation P = 0.000 is considered as significant.

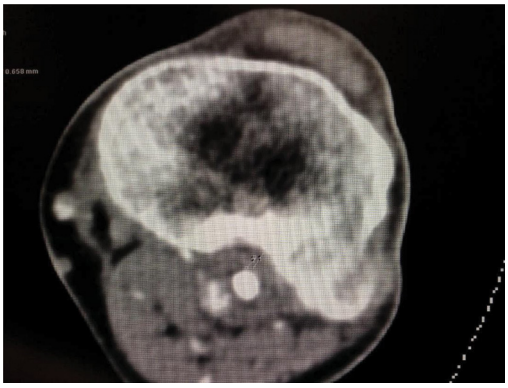


Figure 1. The axial view at the level of tibial plateau. Line number 1 shows the shortest distance of PA to tibial plateau and line 2 shows the popliteus muscle thickness.

RESULTS

A total of 126 patients underwent CTA (93 males and 33 females). 54 patients in this study were Malays, 47 Indians, and 24 Chinese. The mean age of the patients was 58 years old (range 16 to 92). The mean PA-to-tibial plateau distance is 9.9 mm for the right lower limb (range 2.5 mm to 17.2 mm) and 10.24 mm for the left (range 5.5 mm to 15.4 mm). Table 1 further illustrates the results. There were no significant correlations between PA-to-tibial plateau distance with osteoarthritis grade, age, gender, and racial origin (P > 0.000); however, there were positive correlations between PA-to-tibial plateau distances with popliteus muscle thickness (P = 0.000). There was also significant correlation between the right PA to tibial plateau distance with the left PA to tibial plateau distance (P = 0.000).

Table 1
Results obtained in the study. rgrad = right OA grade, rdis = right PA to tibial plateau distance, rmuscle = right popliteus muscle thickness, lgrad = left OA grade, ldis = left PA to tibial plateau distance, lmusc = left popliteus muscle thickness.

Variable	Obs	Mean	Std deviation	Min	Max
rgrad	*126*	2.31746	1.306297	*0*	*5*
lgrad	*126*	2.198413	1.344737	0	*5*
rdis	*124*	9.933871	2.660579	*2.5*	*17.2*
ldis	*119*	10.24118	2.329996	*5.5*	*15.4*
rmusc	*124*	7.614516	1.929897	*1.5*	*13.8*
lmusc	*119*	8.097479	2.053953	4	13
Age	*126*	58.19048	15.14924	*16*	*92*
Gender	*126*	.7380952	.4414263	0	1

DISCUSSION

To our knowledge, based on the scientific literature written in English e in PUBMED and Google Scholar. The study finds a correlation between mean popliteal artery and tibial plateau distance. . It is known that Computed Tomography (CT) is sensitive and accurate in estimating osteophytes formation compared with plain radiography (Chan et al., 1991). Overestimation or underestimation was expected when determining the joint space especially if the CTA was done on a non-weight bearing supine position. We tried to limit over/underestimation of the osteoarthritis grading, based on Kellgren Lawrence's grading system, by correlating the reconstructive coronal view of CT images and the topography images acquired during CTA.

There are multiple ways to study the popliteal artery distance to the proximal tibia. In-vivo ultrasound (Farrington et al., 1999; Eriksson et al., 2010) and MRI knee images (Smith et al., 1999; Shiomi et al., 2001; Sanz-Pérez M et al., 2015) were the popular methods of investigation; until recently, cadaveric anatomical study with CT angiographic analysis (Bisicchia et al., 2014) was also being used.

CTA has the advantage of reproducibility, and reliability when measuring the popliteal artery distance without external compression, unlike using ultrasound. The ultrasound probe may compress the popliteal area structure, thus causing the popliteal artery to be nearer to the bony prominence during the examination, especially if it is done in prone position, thus skewing the actual measurement. The biggest disadvantages of CTA imaging modality are often due to contrast media complications and radiation burden concern.

There was no significant correlation between PA-to-tibial plateau distances with osteoarthritis grades in our study. This finding may suggest that osteoarthritic condition in the knee does not reduce the popliteal artery to the tibial plateau distance. Hence, a higher osteoarthritic grade does not impose additional risks with regards to popliteal artery to tibial plateau distance, with relatively similar arterial injury risks compared to normal knees.

There was no significant correlation seen in the PA to tibial plateau distances across h age, gender, and racial origin in our study. Contrary to our finding, Sanz-Pérez (2015) has suggested that the distance of the posterior of the external meniscus to the popliteal artery is significantly greater in male gender (P = 0.006). In the same study, it was also stated that there was no correlation between the distance from the posterior wall of the external meniscus to the popliteal artery with height, weight, BMI, diameter of tibial plateau, and AP diameter of the external meniscus (Sanz-Pérez et al., 2015).

A positive correlation between PA-to-tibial plateau distances with popliteus muscle thickness (P = 0.000) was seen in our study probably can be explained that the muscle is sandwiched in between the posterior tibia and the popliteal artery. The larger popliteus muscle mass thus will increase the distance of the mean PA-to-tibial plateau distance.

Our study showed that the mean PA-to-tibial plateau distance is 9.9 mm for the right lower limb and 10.24 mm for the left knee. This is not very different from the findings of other studies. Eriksson (2010) showed that the PA to the tibial plateau distance in osteoarthritic patients using ultrasound bilaterally in full knee extension is 8 mm. Sanz-Pérez (2015) suggested that the minimum distance from the posterior wall of the external meniscus to the popliteal artery was 10.01 mm. A study by Farrington (1999), using ultrasound doppler in 0 degree in osteoarthritic patient preoperatively, found that the position of PA in the arthritic knee is less than 1 cm. Bisicchia (2015) suggested the median as the shortest distance PA to the proximal tibial osteotomy zone is 9.6 mm. Yang (2011) suggested the average distance between the tibial cortex and the neurovascular structure was less than 10 mm. The shortest distance found in the literature was seen by Mitsuhiro (2011), using ultrasound as the mode of examination. The mean distance in the study control knees was 4.9 ± 0.3 mm at 0° flexion. Kim (2010) in a cadaveric anatomical study and lateral plain X-ray findings suggested that at 2 cm below the joint line, the mean distance at 0° was 6.2 ± 4.2 mm.

We did not find any presence of aberrant anterior tibial artery in our series although the incidence in a study by Klecker (2008) had reported the incidence to be as high as 2.1%.

The study population does not represent the real normal population because most of our patients already suffer from lower limb arterial disease. Concern over the effects of radiation exposure during CTA examination limits the feasibility of doing exploratory prospective study for normal subjects.

ACKNOWLEDGEMENTS

The authors would like to thank hospital Serdang radiology staff for their cooperation in data acquisition and mentoring.

Disclosure:

The authors did not receive any payments, services from a third party for any aspect in this study.

REFERENCES

Bisicchia, S., Rosso, F., Pizzimenti, M. A., Rungprai, C., Goetz, J. E., & Amendola, A. (2015). Injury risk to extraosseous knee vasculature during osteotomies: a cadaveric study with CT and dissection analysis. *Clinical Orthopaedics and Related Research*, *473*(3), 1030-1039.

Butt, U., Samuel, R., Sahu, A., Butt, I. S., Johnson, D. S., & Turner, P. G. (2010). Arterial injury in total knee arthroplasty. *Journal of arthroplasty, 25*(8), 1311-1318.

Chan, W. P., Lang, P., Stevens, M. P., Sack, K., Majumdar, S., Stoller, D. W., & Genant, H. K. (1991). Osteoarthritis of the knee: comparison of radiography, CT, and MR imaging to assess extent and severity. *American journal of roentgenology*, *157*(4), 799-806.

Eriksson, K., & Bartlett, J. (2010). Popliteal artery–tibial plateau relationship before and after total knee replacement: a prospective ultrasound study. *Knee Surgery, Sports Traumatology, Arthroscopy*, *18*(7), 967-970.

Farrington, W. J., Charnley, G. J., Harries, S. R., Fox, B. M., Sharp, R., & Hughes, P. M. (1999). The position of the popliteal artery in the arthritic knee. *The Journal of arthroplasty*, *14*(7), 800-802.

Kellgren, J. H., & Lawrence, J. S. (1957). Radiological assessment of osteo-arthrosis. *Annals of the rheumatic diseases, 16*(4), 494–502.

Kim, J., Allaire, R., & Harner, C. D. (2010). Vascular Safety During High Tibial Osteotomy A Cadaveric Angiographic Study. *The American journal of sports medicine*, *38*(4), 810-815.

Klecker, R. J., Winalski, C. S., Aliabadi, P., & Minas, T. (2008). The Aberrant Anterior Tibial Artery Magnetic Resonance Appearance, Prevalence, and Surgical Implications. *The American journal of sports medicine*, *36*(4), 720-727.

Ko, L. J. M., DeHart, M. L., Yoo, J. U., & Huff, T. W. (2014). Popliteal artery injury associated with total knee arthroplasty: trends, costs and risk factors. *The Journal of arthroplasty*, *29*(6), 1181-1184.

Parvizi, J., Pulido, L., Slenker, N., Macgibeny, M., Purtill, J. J., & Rothman, R. H. (2008). Vascular injuries after total joint arthroplasty. *The Journal of arthroplasty*, *23*(8), 1115-1121.

Sanz-Pérez, M., García-Germán, D., Ruiz-Díaz, J., Navas-Pernía, I., & Campo-Loarte, J. (2015). Localización de la arteria poplítea y su relación con el riesgo vascular en la sutura del cuerno posterior del menisco externo. *Revista Española de Cirugía Ortopédica y Traumatología*, *59*(3), 165-171.

Shetty, A. A., Tindall, A. J., Qureshi, F., Divekar, M., & Fernando, K. W. K. (2003). The effect of knee flexion on the popliteal artery and its surgical significance. *Bone and Joint Journal*, *85*(2), 218-222.

Shiomi, J., Takahashi, T., Imazato, S., & Yamamoto, H. (2001). Flexion of the knee increases the distance between the popliteal artery and the proximal tibia: MRI measurements in 15 volunteers. *Acta Orthopaedica Scandinavica*, *72*(6), 626-628.

Smith, P. N., Gelinas, J., Kennedy, K., Thain, L., Rorabeck, C. H., & Bourne, R. B. (1999). Popliteal Vessels in Knee Surgery A Magnetic Resonance Imaging Study. *Clinical orthopaedics and related research*, *367*, 158-164.

Takeda, M., Ishii, Y., Noguchi, H., & Sato, J. (2011). Change in the Position of the Popliteal Artery with Knee Flexion after Total Knee Arthroplasty. *Journal of Bone and Joint Surgery*, *93*(21), e123.

Yang, D., Zhou, Y., Tang, Q., Xu, H., & Yang, X. (2011). Anatomical relationship between the proximal tibia and posterior neurovascular structures: a safe zone for surgeries involving the proximal tibia. *The Journal of arthroplasty*, *26*(7), 1123-1127.

Yoo, J. H., & Chang, C. B. (2009). The location of the popliteal artery in extension and 90-degree knee flexion measured on MRI. *The Knee*, *16*(2), 143-148.

Zaidi, S. H., Cobb, A. G., & Bentley, G. (1995). Danger to the popliteal artery in high tibial osteotomy. *Bone and Joint Journal*, *77*(3), 384-386.

REFEREES FOR THE PERTANIKA
JOURNAL OF SCIENCE AND TECHNOLOGY

VOL. 25(1) JAN. 2017

The Editorial Board of the Journal of Science and Technology wishes to thank the following for acting as referees for manuscripts published in this issue of JST.

Abdul Aziz Abdul Raman
(UM, Malaysia)

Abdul Aziz Hairuddin
(UPM, Malaysia)

Abu Abdullah
(UTeM, Malaysia)

Ahmad Farhan Mohd Sadullah
(USM, Malaysia)

Ahmad Kamal Ariffin Mohd Ihsan
(UKM, Malaysia)

Ali Yeon Md Shakaff
(UNIMAP, Malaysia)

Ambarish Mohapatra
(SIT, India)

Aminuddin Abd. Ghani
(USM, Malaysia)

Amir Akramin Shafie
(IIUM, Malaysia)

Azah Mohamed
(UKM, Malaysia)

Azinoor Azida Abu Bakar
(UiTM, Malaysia)

Azlina Harun
(USM, Malaysia)

Baba Md. Deros
(UKM, Malaysia)

Chonticha Kaewanuchit
(ARU, Thailand)

Desa Ahmad
(UPM, Malaysia)

Dian Daruis
(UPNM, Malaysia)

Fadzilah Siraj
(UUM, Malaysia)

Garry Kuan
(USM, Malaysia)

Ghazali Osman
(UiTM, Malaysia)

Hanapi Mat
(UTM, Malaysia)

Hasnida Ab Samat
(USM, Malaysia)

Heng Swee Huay
(MMU, Malaysia)

Jacek Uziak
(UB, Botswana)

Jagannath Korody
(MIT, India)

Jethro Henry Adam
(UNIMAS, Malaysia)

K. Kannan
(SASTRA University, India)

Kartini Kamaruddin
(UiTM, Malaysia)

Kek Sie Long
(UTHM, Malaysia)

Khairuddin Omar
(UKM, Malaysia)

Krishnamurthy Nayak
(MIT, India)

Leonard I Wiebe
(UAlberta, Canada)

Mariam Mohamad
(UiTM, Malaysia)

Marlinda Abdul Malek
(UNITEN, Malaysia)

Mohamad Kamarol Mohd Jamil
(USM, Malaysia)

Mohamad Rushdan Md. Said
(UPM, Malaysia)

Mohamed Abd. Rahman
(UPM, Malaysia)

Mohd Zaid Akop
(UTeM, Malaysia)

Mohd. Wajid Shamsi
(JUIT, India)

Nagaraj S Nayak
(CCE, Oman)

Noor Atinah Ahmad
(USM, Malaysia)

Norzila Othman
(UTHM, Malaysia)

Ong Hong Choon
(USM, Malaysia)

Puvanasvaran A. Perumal
(UTeM, Malaysia)

Ravikrian Kadoli
(NIT, India)

S M Abdul Khader
(MU, Malaysia)

S. Sendhil Kumar
(KIT, India)

Sabariah Jemali
(MISI, Malaysia)

Shahrin Samat
(KPJ, Malaysia)

Shatirah Mohamed Akib
(UM, Malaysia)

Siow Chun Lim
(Taylor's University, Malaysia)

Siti Fauziah Toha
(IIUM, Malaysia)

Srikanth Prabhu
(MIT, India)

Srinivasa Pai P
(NMAM Institute Technology, India)

Vijay G S
(MIT, India)

Wan Tat Chee
(USM, Malaysia)

Yap Tze Chuen
(HWU, Malaysia)

Zakiah Ahmad
(UiTM, Malaysia)

ARU - Phranakhon Si Ayutthaya Rajabhat University
CCE - Caledonian College of Engineering
HWU - Heriot-Watt University
IIUM - International Islamic University Malaysia
JUIT - Jaypee University of Information Technology
KIT - Karpagam Institute of Technology
KPJ - Kumpulan Perubatan Johor
MISI - Malaysia Institute for Supply Chain Innovation
MIT - Manipal Institute of Technology
MMU - Multimedia University
MU - Manipal University
NIT - National Institute of Technology
NMAM - Nitte Mahalinga Adyanthaya Memorial Institute of Technology
SASTRA -Shanmugha Arts, Science, Technology & Research Academy
SIT - Silicon Institute of Technology
UAlberta - University of Alberta
UB - University of Botswana
UiTM - Universiti Teknologi MARA
UKM - Universiti Kebangsaan Malaysia
UM - Universiti Malaya
UNIMAP - Universiti Malaysia Perlis
UNIMAS - Universiti Malaysia Sarawak
UNITEN - Universiti Tenaga Nasional
UPM - Universiti Putra Malaysia
UPNM - Universiti Pertahanan Nasional Malaysia
USM - Universiti Sains Malaysia
UTeM - Universiti Teknikal Malaysia Melaka
UTHM - Universiti Tun Hussein Onn Malaysia
UTM - Universiti Technologi, Malaysia
UUM - Universiti Utara 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*

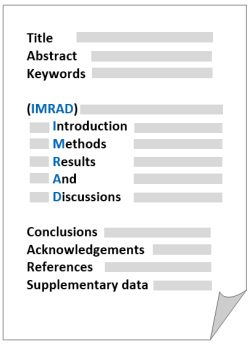
This page should **repeat** the **full title** of your paper with only the **Abstract** (the abstract should be less than 250 words for a Regular Paper and up to 100 words for a Short Communication), and **Keywords**.

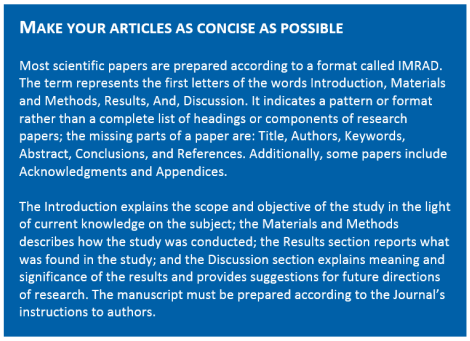
Keywords: Not more than eight keywords in alphabetical order must be provided to describe the contents of the manuscript.

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. 25 (1) Jan. 2017

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