Pertanika Journal of

SCIENCE & TECHNOLOGY

JST

VOL. 28 (4) OCT. 2020

PERTANIKA JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

Journal of Science & Technology

Journal of Science & Technology

Journal of Science & Technology

About the Journal

Overview

PPertanika Journal of Science & Technology is the official journal of Universiti Putra Malaysia. It is an open-access online scientific journal. It publishes original scientific outputs. It neither accepts nor commissions third party content.

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

Pertanika Journal of Science & Technology 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 for submission by authors from all over the world.

Aims and scope

Pertanika Journal of Science & 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 3 journals as Pertanika Journal of Tropical Agricultural Science, Pertanika Journal of Science & Technology, and Pertanika Journal of Social Sciences & Humanities to meet the need for specialised journals in areas of study aligned with the interdisciplinary strengths of the university.

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

Vision

To publish journals of international repute.

Mission

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 can expect to receive a decision within 90 days. The elapsed time from submission to publication for the articles averages 180 days. We are working towards decreasing the processing time with the help of our editors and the reviewers

Abstracting and indexing of *Pertanika*

Pertanika is over 42 years old; this accumulated knowledge and experience has resulted in Pertanika Journal of Science & Technology being abstracted and indexed in SCOPUS (Elsevier), Clarivate Web of Science [ESCI], EBSCO, DOAJ, Agricola, ASEAN CITATION INDEX, ISC, Microsoft Academic, Google Scholar, National Agricultural Science (NAL), and MyCite.

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 take 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

Originality

The author must ensure that when a manuscript is submitted to *Pertanika*, the manuscript must be an original work. The author should check the manuscript for any possible plagiarism using any program such as Turn-It-In or any other software before submitting the manuscripts to the Pertanika Editorial Office, Journal Division.

All submitted manuscripts must be in the journal's acceptable **similarity index range**:
≤ 20% – PASS; > 20% – REJECT.

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.

Pertanika 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 90 days (average). The elapsed time from submission to publication for the articles averages 180 days.

Authorship

AAuthors 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

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words *Introduction, **M**aterials and Methods, **R**esults, And **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, References, and Acknowledgement. Additionally, some papers include 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 Instruction to Authors (http://www.pertanika.upm.edu.my/Resources/regular_issues/Regular_Issues_Instructions_to_Authors.pdf).

Editorial process

Authors who complete any submission are notified with an acknowledgement containing a manuscript ID on receipt of a manuscript, and upon the editorial decision regarding publication.

Pertanika follows a **double-blind peer-review** process. Manuscripts deemed suitable for publication are sent to reviewers. Authors are encouraged to suggest names of at least 3 potential reviewers at the time of submission of their manuscripts to Pertanika, but the editors will make the final selection and are not, however, bound by these suggestions.

Notification of the editorial decision is usually provided within 90 days 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 may identify authorship of the paper should be placed only on page 2 as described in the first-4-page format in *Pertanika's* **Instruction to Authors** (http://www.pertanika.upm.edu.my/Resources/regular_issues/Regular_Issues_Instructions_to_Authors.pdf).

The Journal's peer-review

In the peer-review process, 2 to 3 referees independently evaluate the scientific quality of the submitted manuscripts. At least 2 referee reports are required to help make a decision.

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 7 steps to the editorial review process:

1. The journal's Chief Executive Editor and the Editor-in-Chief examine the paper todetermine whether it is relevance to journal needs in terms of novelty, impact, design,procedure, language as well as presentation and allow it to proceed to the reviewingprocess. If not appropriate, the manuscript is rejected outright and the author is informed.

2. The Chief Executive Editor sends the article-identifying information having beenremoved, to 2 to 3 reviewers. They are specialists in the subject matter of the article.The Chief Executive Editor requests that they complete the review within 3 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 research field.The Editor-in-Chief (EiC) examines the review reports and decides whether to accept or reject the manuscript, invites the author(s) to revise and resubmit the manuscript, or seek additional review reports. Final acceptance or rejection rests with the CEE and EiC, who reserve 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 to the authors for attending to the reviewers' suggestions and perhaps additional advice about revising the manuscript.

3. The Editor-in-Chief, examines the review reports and decides whether to accept orreject the manuscript, invite the authors to revise and resubmit the manuscript, orseek additional review reports. In rare instances, the manuscript is accepted withalmost no revision. Almost without exception, reviewers' comments (to the authors) are forwarded to the authors. If a revision is indicated, the editor provides guidelines for attending to the reviewers' suggestions and perhaps additional advice about revising the manuscript. The CEE sends the revised paper out for re-review. Typically, at least 1 of the original reviewers will be asked to examine the article.

4. The authors decide whether and how to address the reviewers' comments andcriticisms 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 addressed' the concerns of the reviewers and the editor, usually in a tabular form. The authors may also submit a rebuttal if there is a need especially when the authors disagree with certain comments provided by reviewers.

5. The Chief Executive Editor sends the revised manuscript out for re-review. Typically, at least 1 of the original reviewers will be asked to examine the article.

6. When the reviewers have completed their work, the Editor-in-Chief examines their comments and decides whether the manuscript is ready to be published, needs another round of revisions, or should be rejected. If the decision is to accept, the Chief Executive Editor is notified.

7. The Chief Executive Editor reserves the final right to accept or reject any material for publication, if the processing of a particular manuscript is deemed not to be in compliance with the S.O.P. of *Pertanika*. An acceptance letter is sent to all the authors.

   The editorial office ensures that the manuscript 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 editorial office. 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 manuscript appears in the pages of the journal and is posted on-line.

Pertanika Journal of

SCIENCE & TECHNOLOGY

Vol. 28 (4) Oct. 2020

A scientific journal published by Universiti Putra Malaysia Press

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

EDITORIAL OFFICE
EDITORIAL OFFICE
JOURNAL DIVISION
Putra Science Park
1st Floor, IDEA Tower II
UPM-MTDC Technology Centre
Universiti Putra Malaysia
43400 Serdang, Selangor Malaysia.
Gen Enq.: +603 9769 1622 | 1616
E-mail:
executive_editor.pertanika@upm.edu.my
URL: www.journals-jd.upm.edu.my

PUBLISHER
UPM Press
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia.
Tel: +603 9769 8855
E-mail: penerbit@putra.upm.edu.my
URL: http://penerbit.upm.edu.my

ABSTRACTING AND INDEXING OF PERTANIKA JOURNALS

Pertanika has reached 40 years old; this accumulated knowledge has resulted in the journals being abstracted and indexed in SCOPUS (Elsevier), Clarivate-Emerging Sources Citation Index [ESCI (Web of Science)], BIOSIS, National Agricultural Science (NAL), Google Scholar, MyCite and ISC.

Pertanika Journal of Science & Technology
Vol. 28 (4) Oct. 2020

Contents

Foreword

Welcome to the fourth issue of 2020 for 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 for the benefit of the world-wide science community.

This issue contains 22 articles; a review article and the rest are regular articles. The authors of these articles come from different countries namely India, Indonesia, Iraq, Malaysia, Nigeria, Pakistan, Thailand, and UK.

A regular article titled "Predictive performance of logistic regression for imbalanced data with categorical covariate" was written by Hezlin Aryani Abd Rahman and co-researchers from Universiti Teknologi MARA and UCSI University, Malaysia. Logistic regression is often used for the classification of a binary categorical dependent variable using various types of covariates which is continuous or categorical. They investigated the effect of imbalanced data measured by imbalanced ratio on the parameter estimate of the binary logistic regression with a categorical covariate. The simulation results provided evidence that the effect of imbalance ratio on the parameter estimate of the covariate decreased as sample size increased. Detailed information on this study can be found on page 1141.

The next article is on fast and robust diagnostic technique for the detection of high leverage points by Habshah Midi et al. from Universiti Putra Malaysia and a co-author from University of Al-Qadisiyah, Iraq. The existing Diagnostic Robust Generalized Potential based on Minimum Volume Ellipsoid is quite successful in identifying high leverage points, but its running time is very slow. The Diagnostic Robust Generalized Potential based on Index Set Equality running time was much faster than the Diagnostic Robust Generalized Potential based on Minimum Volume Ellipsoid. However, the Diagnostic Robust Generalized Potential based on Index Set Equality was not computationally stable and still possessed masking and swamping effect. Contrarily, the proposed Diagnostic Robust Generalized Potential based on Reweighted Fast Consistent and High Breakdown was very successful in detecting HLPs with negligible swamping effect. Full information on this study is presented on page 1203.

Athraa Jasim Mohammed and Khalil Ibrahim Ghathwan from University of Technology, Baghdad, Iraq had proposed a new method for color image-based segmentation using whale optimization algorithm in clustering. Comparison of the proposed method was performed with the wolf color image-based segmentation, cuckoo color image-based segmentation, bat color image-based segmentation, and k-means color image-based

segmentation over four benchmark color images. Experimental results demonstrated that the proposed whale color image-based segmentation had higher value of Peak Signal to Noise Ratio and lower value of Root Mean Squared Error in most cases compared to other methods. Further details of the article are available on page 1389.

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

All the papers published in this edition underwent Pertanika's stringent peer-review process involving a minimum of two reviewers comprising internal as well as external referees. This was to ensure that the quality of the papers justified the high ranking of the journal, which is renowned as a heavily-cited journal not only by authors and researchers in Malaysia but by those in other countries around the world as well.

We would also like to express our gratitude to all the contributors, namely the authors, reviewers, Editor-in-Chief and Editorial Board Members of JST, who have made this issue possible. 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
Dato' Dr. Abu Bakar Salleh
executive_editor.pertanika@upm.edu.my

Pertanika J. Sci. & Technol. 28 (4): 1117 - 1140 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Review Article

A Review: Autodissemination of Pyriproxyfen as Novel Strategy to Control Dengue Outbreaks

Ahmad Mohiddin Mohd Ngesom[1], David Greenhalgh[2], Asmalia Md Lasim[3], Mazrura Sahani[1], Rozita Hod[4] and Hidayatulfathi Othman[1]*

[1]Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 UKM, Kuala Lumpur, Malaysia
[2]Department of Mathematics and Statistics, University of Strathclyde, 26, Rivhmond Street, Glasgow, G11XH, UK
[3]Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, 43600 UKM, Selangor, Malaysia
[4]Department of Community Health, Faculty of Medicine, Universiti Kebangsaan Malaysia, 56000 UKM, Cheras, Kuala Lumpur, Malaysia

ABSTRACT

The new emergence and re-emergence of arbovirus infections transmitted by *Aedes* mosquitoes have been spreading across Southeast Asia, Central Africa, United States, tropical Oceania and has become a major of public health concern. These arbovirus diseases were found to have a similar vector, symptoms, and environments. The situation is complex due to no specific vaccine or treatments being available for the diseases. Therefore, vector control is currently the best defense against arbovirus diseases, but with its own challenges such as the difficulty in controlling scattered breeding sites and biological behavior. Herein, we present a literature review of studies on current techniques proposed to combat dengue transmission that can fill a crucial gap in vector control programs, which is the inability of conventional control methods to eliminate and destroy cryptic breeding sites. In particular, we focused on the concept of autodissemination, which is a self-delivery technique by manipulating the behavior of mosquitoes, carrying the

ARTICLE INFO

Article history:
Received: 31 March 2020
Accepted: 27 July 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.01

E-mail addresses:
aksmohiddin@yahoo.com (Ahmad Mohiddin Mohd Ngesom)
david.greenhalgh@strath.ac.uk (David Greenhalgh)
asmaliaccb@gmail.com (Asmalia Md Lasim)
mazrura@ukm.edu.my (Mazrura Sahani)
rozita.hod@ppukm.ukm.edu.my (Rozita Hod)
hida@ukm.edu.my (Hidayatulfathi Othman)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

insecticide and disseminating it to cryptic breeding sites. This technique has shown promising results in some countries and can be considered as an additional tool in a vector control program. Therefore, we conducted Boolean searches in several electronic databases including Google Scholar, PubMed, SciELO and ScienceDirect to identify relevant published data regarding dengue and autodissemination techniques.

Keywords: Aedes, autodissemination, dengue, pyriproxyfen, vector control

INTRODUCTION

To date, there are more than 300 known species of mosquitoes in the world but with only a small number of species responsible for vector-borne diseases (Caraballo & King, 2014). *Aedes aegypti* and *Aedes albopictus* are the most important vectors for the transmission of dengue, Zika, and chikungunya virus. According to the World Health Organization (WHO), mosquito infections affected more than one third of the world's population and caused about one million deaths annually (WHO, 2016).

Aedes aegypti is the most important disease vector worldwide. It is the principal vector of dengue, chikungunya, yellow fever and Zika viruses. *Ae. aegypti* is extremely anthropophilic, has frequent blood-feeding behavior, a long lifespan and is closely associated with humans, because of their ability to breed in human dwellings, and preferring to feed on humans even in the presence of other mammals (Faraji et al., 2014). The blood-feeding behavior becomes the primary concern because it is the major component of the dengue virus transmission where an infected patient gets bitten by a mosquito which then carries the virus to another person. Most mosquito species have their preferred distributions, with the *Ae. aegypti* being the primary vector of the dengue virus, but it is also an efficient vector for other viruses that cause human diseases such as the Zika virus and chikungunya virus.

Meanwhile, *Aedes albopictus* was found to be able to transmit over more than 20 arboviruses, increasing the risk of humans to mosquito-borne diseases (Wong et al., 2013) and was considered as a secondary vector in importance according to their zoophilic preferences (Delatte et al., 2010). In certain areas around the world, it has become one of the most invasive mosquito species. *Ae. albopictus* is considered a greater threat compared to *Ae. aegypti* due to the former's higher survival rate in larval densities (Camara et al., 2016), adaptation to extreme temperatures (Brady et al., 2013) and showing greater survivorship in food-limited competitions (Banerjee et al., 2017; Hopperstad & Reiskind, 2016).

SEARCH STRATEGY

We searched PubMed, ScienceDirect, SciELO and Google Scholar with the following keywords: dengue fever; dengue control; vector control; autodissemination; pyriproxyfen; pyriproxyfen autodissemination; horizontal transfer; autodissemination AND dengue or

malaria; emergence inhibition; and pyriproxyfen (PPF) in several combinations using Boolean searches. Non-English language articles were also included if there was an available English translation. Due to the limited data published in scientific journals, we also explored the databases of leading organization, such as World Health Organization (WHO) and Ministry of Health Malaysia (MOH) and Centers for Disease Control and Prevention (CDC).

Dengue

Although dengue fever is among the most common mosquito-borne infections in the world, yet it has long been categorized as a "neglected tropical disease" (Faraji et al., 2014). The Chinese medical encyclopedia from the Jin Dynasty (265–420 AD) was probably the first record of which diseases were associated with flying insects. The first recognized dengue epidemics occurred almost simultaneously in Asia, Africa, and North America in the 1780s. The first confirmed case report of dengue hemorrhagic fever in Southeast Asia occurred in Manila in 1953 and during the second outbreaks in 1956. Today, dengue fever has become one of the major public health issues and is endemic in more than 125 countries globally such as in Malaysia (Othman et al., 2019), China (Xiang et al., 2017), Pakistan (Yousaf et al., 2018) and Sri Lanka (Sun et al., 2017).

There are an estimated 120 million travelers to the subtropical and tropical regions with around two billion people living in developing countries (Banu et al., 2014), of which 40% are at risk of contracting the dengue virus annually, while the mortality rate is 2.5% with more cases of classical dengue with severe symptoms (Rasgon, 2011). There were an estimated 390 million cases of dengue infections in 128 countries, with approximately half a million with dengue fever needing hospitalization (WHO, 2020). In Malaysia, a total of 130,101 cases and 182 deaths were reported in 2019, compared to 80,615 cases and 147 deaths in 2018. Most of the cases were reported in the state of Selangor, with 72,543 cases which comprised 55.76% of the total cases reported in Malaysia (KKM, 2020). The escalation in the number of cases was attributed to poor environmental cleanliness, changes in serotypes and high population density in urban areas.

Control of Mosquitos

Currently, there are no curative treatments or vaccines available. Therefore, control of the dengue vector is crucial to reduce the transmission of dengue fever. A combination of multidisciplinary approach involving local authorities, mobilization of the community and integration of vector control must be done together to ensure the success of the strategy. However, there are many challenges in the response against dengue outbreaks that include unmanaged and rapid urbanization, poor sanitation systems and bad disposal management in the communities (Chang et al., 2011).

One of the most effective vector control methods is to control the level of vector populations (Ahmad-Azri et al., 2019; Mazrura et al., 2012). However, most of the vector control strategies nowadays are solely chemical-based, thus without careful monitoring, it will increase the insecticide resistance phenomenon (Besar et al., 2019). This situation has led the WHO and other agencies to demand for more sustainable and reliable strategies (Bourtzis et al., 2016). Thus, it is important to find novel surveillance and prevention strategies for dengue management (Huong et al., 2004). The occurrence of insecticide resistance may vary by geographical distribution due to high usage of pesticides by households and in agriculture. The development of resistance may also become faster and more severe than expected, thus it is essential to control *Aedes* mosquito populations by eliminating breeding sites through self-initiatives for source reduction in homes and the community (Mazrura et al., 2010).

There are several ways to control a vector population, which are through source reduction, biological control, genetic manipulation, and chemical control (Buhler et al., 2019; Evans et al., 2019; Arham et al., 2018; Awang et al., 2012; Sulaiman et al., 2007). However, the best vector control strategy is to combine source reduction methods, law enforcement and controlling the use of insecticides, which are crucial for the success of any vector control program (Boubidi et al., 2016; Schorkopf et al., 2016). The responsible agencies should prepare several strategies on mosquito management such as performing location assessment, insecticide treatment, educating the public, establishing the policies and regulations, conducting mosquito surveillance and utilizing new technologies (Zaini et al., 2019; Gerding et al., 2016).

The WHO has announced the concept of Integrated Vector Management (IVM) to optimize the use of resources in the management of vector programs that is a rational decision-making for the best management of vector control resources (Sukumaran et al., 2019). The IVM aims to more effectively prevent the transmission of vector-borne diseases such as dengue, chikungunya, malaria and Zika virus, while being more ecologically sound and sustainable (Niang et al., 2018). This strategy implies the collaboration between the public sectors, health sectors and the empowerment of communities by making insecticide application as the last resort. There are five key elements of IVM guidelines: (i) advocacy, social mobilisation and legislation which is enabling the embedding of IVM in policy-design across relevant organisations, (ii) collaboration within the health sector and other agencies, (iii) integrated approach by ensuring the rational use of available tools and resources, (iv) evidence-based decision-making based on local ecology, epidemiology and resources and, (v) capacity-building at national and local level (WHO, 2012). However, without a proper direction, the efficacy of the IVM would not be able to be fully utilized.

Insect Growth Regulators as an Alternative?

Insect Growth Regulators (IGR) are the third generation of insecticides that differ from earlier insecticides on their mechanism through influencing insect development, disrupting the activities of the insect endocrine system and metamorphosis (Palli, 2016). In comparison, the first generation of insecticides is characterized by chemicals known throughout human history that include oils and arsenic. The second generation of insecticides includes carbamates, organophosphates, and organochlorides that were synthesized after the discovery of dichlorodiphenyltrichloroethane DDT.

Insect Growth Regulators (IGR) have a selective mechanism in their mode of action that specifically inhibit the development of insects into adults. The mechanisms of IGR delay the transformation and inhibit the cuticle formation in an insect's immature stage (Darabi et al., 2011). IGRs will be applied during the early life stage of the insects and kill the insect before they become adults, and thus unable to reproduce. IGR is a form of "birth control" and can be described as control of insects through the release of hormones/IGR into the populations. It is not necessary for the IGR to have high toxicity against the target sites but enough that it may lead to a change and abnormalities against insect survival. Two compounds are important in regulating the development of insects, namely: (i) ecdysone (also known as molting hormone, MH), and (ii) juvenile hormone (JH) which directly interfere with insect metamorphosis, embryogenesis, or reproduction.

Pyriproxyfen is a powerful JH agonist which affects the physiology of metamorphosis, embryogenesis, and reproductivity in insects with a unique mode of action. In general, pyriproxyfen is used in larvicidal activities to control mosquitoes, an example being the Sumilarv™ brand of IGR. Pyriproxyfen is useful in inhibiting the emergence of adult *Aedes* spp. and is used in the range of only 0.000048 ppm (Seccacini et al., 2008) to 0.33 ppm (Suman et al., 2017) and could inhibit the emergence of *Ae. albopictus* with a LC_{50} and LC_{90} of 0.0012 ppm and 0.021 ppm respectively (Lau et al., 2015), while the LC_{50} and LC_{90} for *Ae. aegypti* were 0.012 ppb and 0.61 ppb, respectively (Sihuincha et al., 2005). These studies proved that pyriproxyfen is one of the most effective larvicides against *Aedes* species in laboratory settings.

In addition to the larvicidal impact, pyriproxyfen has been reported to reduce the longevity, fertility, and fecundity in exposed mosquitoes. This is a novel strategy as the mosquitoes exposed to the pyriproxyfen are laying unviable eggs. Earlier studies on *Ae. aegypti* using sublethal pyriproxyfen doses demonstrated effects on mosquito fecundity and fertility (Harburguer et al., 2014). Due to its environmental persistence, specific activity against insect stages and being non-toxic to mammals, pyriproxyfen is considered as an attractive alternative to conventional insecticides. Furthermore, pyriproxyfen has been evaluated as a safe insecticide for application in drinking water with minimal impacts on non-target aquatic insects and the environment.

Pull and Push Concept: Autodissemination Potential Against *Aedes* Spp. Mosquitoes

Autodissemination is a novel technology to control mosquito populations by exploiting the ability of female mosquitoes to find suitable breeding sites, and on the skip-oviposition, the mosquito will lay their eggs in different containers during a single gonotrophic cycle. The contaminated mosquito is carrying a small particle of insecticide and will transfer it to other mosquitoes via mating, and oviposition in the hidden and cryptic breeding sites that humans may not be able to assess during inspections (Mains et al., 2015). The mechanism of autodissemination approaches is presented in Figure 1. Autodissemination can be considered as pull and push technologies with the attraction and dispersal concept which are beneficial and cost-effective to the vector control program (Unlu et al., 2017). For the autodissemination technique to be successful, specific prerequisites have to be met which are: 1) The ability of the autodissemination station to attract mosquitoes, 2) the autodissemination station is able to transfer the chemical to the mosquitoes and 3) the insecticide distribution from the exposed mosquitoes to the breeding sites (Caputo et al., 2012; Gaugler et al., 2012). In addition, another requirement is that the chemical used can work at a low concentration and persists in the environment for a long time (Mains et al., 2015).

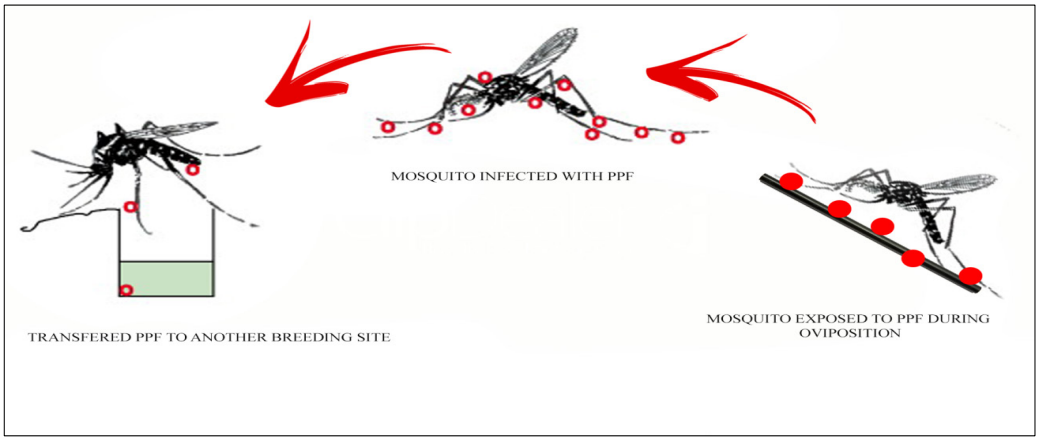


Figure 1. The mechanism of autodissemination of pyriproxyfen by female mosquitoes from exposure containers to other breeding sites.

History of Autodissemination Methods

The concept of autodissemination devices has been tested on other weeds and pests (Pell et al., 1993; Vega et al., 2000). One example is an automatic device using a fungus as biocontrol agents (BCAs) working as a lure. Once the insect is attracted and enters the

device, it becomes contaminated and transfers the BCA into populations after leaving the device. The use of BCA dissemination by honey bees was significantly more efficient than conventional sprayers to spread the inoculums against a pest-infested flower (Gross et al., 1994). To enhance the impact of BCAs, the "push-pull" strategy was devised by stimulating the insects out from the population ("push") with feeding the deterrents and extracted ("pull") into the trap. The deterrent lures must be more favorable and attractive than the crop to encourage the insect to enter the traps. In several studies, autodissemination has shown promising results within the pest population when using an attractant device as the initial source of infection (Moslim et al., 2011; Scholte et al., 2004).

Autodissemination has widely been used against diamondback moth (DBM), *Plutella xylostella* using *Zoophthora radican* and rhinoceros beetle, *Oryctes rhinoceros* using *Metarhizium anisopliae.* The manipulation of DBM behavior using semiochemicals such as pheromones has been used to attract and developed epizootic populations before the critical level was achieved (Pell et al., 1993). Other studies in Klang and Selangor, Malaysia found that the mortality rate of *Oryctes rhinoceros* was 75% to 90% when infected by *Metarhizium anisoliae* with the percentage of trapped adults leaving the traps ranging from 85% to 95% (Moslim et al., 2011).

Autodissemination of Pyriproxyfen on Mosquitoes

Although there is still a lack of autodissemination field trials, more studies are being carried out on the effectiveness of pyriproxyfen, with some researchers making favourable general conclusions. Most of these studies found that autodissemination did occur and successfully killed the larvae and reduced *Aedes* spp. populations. Table 1 summarizes the autodissemination studies conducted under semi-field and field settings. The first trial was done in a laboratory (Itoh et al., 1994), and then the concept was furthered by others (Chism & Apperson, 2003). These studies operated under the assumption that mosquitoes were better at carrying insecticides to their preferred breeding sites compared to human operators. The key to the success of the technique is to use pyriproxyfen in small particles to contaminate mosquitoes which can affect control to the whole area (Gaugler et al., 2012)

A novel method was developed based on the skip oviposition behavior of mosquitoes where female mosquitoes transfer small particles of insecticides to their cryptic breeding sites, thus interrupting the development of the larvae and subsequently killing and reducing the mosquito population in target areas. In 2009, a significant trial for an autodissemination concept with pyriproxyfen was conducted in Iquitos, Peru. The placement of the autodissemination was 5% from the available resting areas, which achieved a 42-98% reduction in adult mosquito emergence with a maximum mortality rate of 98% (Devine et al., 2009). Commercially prepared pyriproxyfen in an emulsifiable concentrate (EC) form was shown to be transferred as far as 200 m from the treated areas.

Table 1

Overview of autodissemination trials conducted under semi-field and open field trials

Settings	Target species	Methods	Result	Conclusion	References
SF	*Ae. aegypti*	A dissemination of pyriproxyfen was conducted in a house (6 x 10 m) in Bangkok, Thailand. Four resting cups (source of pyriproxyfen) were placed in the house against ten oviposition cups. The study was maintained for 16 days and all oviposition cups were monitored every four days. All oviposition cups were brought to the laboratory for larva bioassays and monitoring of their development	Most of the samples collected from oviposition cups showed a high emergence inhibition, comfirming the autodissemination of pyriproxyfen had occurred. The male mosquitoes may also play an important role as a transportation of the pyriproxyfen.	This small study was the first to demonstrate autodissemination activity from the resting place to the oviposition sites	Itoh (1993)
OF	*Ae. albopictus*	Field evaluations of autodissemination were conducted in two isolated sites in Rome with a high density of *Aedes* sp. Two sites were selected: i) oldest cemetery (22 x 2 m) and, ii) enclosed garden (1 ha) Ten DS and ten SS were deployed in both sites placed 90 cm to two meters from each other. Two replicates were conducted using 0.5% and 5% pyriproxyfen formulation in Site I; while 5% formulation was used in Site II. Samples collected from all DS and SS together with *Ae. aegypti* larvae were monitored for seven to 12 days.	Two sets of trials showed a significant increase in mortality ranging from 20.8 to 71.2% compared to control. However, a higher concentration of pyriproxyfen (5%) showed better dissemination with higher mortalities in sentinel sites.	The application using 5% pyriproxyfen showed > 50% larvae mortality. This suggested that the higher concentration of pyriproxyfen was able to improved the efficacy of the autodissemination approaches.	Caputo et al. (2012)

Note: SF, semi-field; OF, open field; PPF, Pyriproxyfen; EI, emergence inhibitions; d, days; w, weeks; m, months; DS, Dissemination Station; SS, Sentinel sites

Table 1 *(Continued)*

Settings	Target species	Methods	Result	Conclusion	References
SF	*Ae. albopictus*	Six bed nets (height, 227 cm; width, 270 cm; length, 270 cm) were used in a large greenhouse (16 m x 10 m) at Nagasaki, Japan. Each bed net was provided with artificial and natural resting sites, sugar water, ten ovitraps (3 mg of 1:1 mixture of mouse pellet and yeast as food larvae and 0.2 g hay infusion). A total of 100 pairs of *Aedes albopictus* were released into each bed net and were monitored for 20 (Exp I) to 44 (Exp II) days. Different concentrations of pyriproxyfen were used in Exp I (1 %) and Exp II (0.1 %) Number of eggs laid, hatching rate and females on rats were monitored. The activity of autodissemination were then assessed via larva bioassays obtained from the water samples collected from each ovitrap.	All monitored parameters (number of eggs laid, pupae and egg hatch) showed a significant reduction after pyriproxyfen treatment compared to the control. Moreover, larva bioassays showed significantly increased larva mortality. The pyriproxyfen-treated bed net could be transferred on mosquitoes and subsequently suppressed the mosquito density.	*Aedes albopictus* were exposed to pyriproxyfen by direct contact with the bed nets. Significant reductions (fecundity and egg hatchability) in mosquito population were observed during the trials	Ohba et al. (2013)
SF	*An. arabiensis*	The semi-field trial was conducted in a semi-field system (SFS) (9.6 m x 9.6 m, 3.9 m high) in rural Southern Tanzania. A small hut was provided for tethered cow bait and posible resting sites for mosquitoes. Four identical chambers were used to assess the autodissemination activity. Unfed mosquitoes were released into the SFS and the experiment was conducted for 25 days. All pupae were removed from clay pots and the emergence rates were recorded. The impact of pyriproxyfen were compared between treatment and control sites.	The mean number of adult emergence rates was significantly higher in treatment compared to the control. A higher number of autodissemination clay pots point (8 autodissemination devices) have shown a better reduction in pupae collection and adult emergence rather than using one clay pot.	The first study conducted against the malaria vector in Africa. The application of autodissemination approaches can be considered as one of the additional tools in the malaria program.	Lwetoijera et al. (2014)

Note: SF, semi-field; OF, open field; PPF, Pyriproxyfen; EI, emergence inhibitions; d, days; w, weeks; m, months

Table 1 *(Continued)*

Settings	Target species	Methods	Result	Conclusion	References
OF	*Ae. albopictus*	Two types of experiments were conducted in NJ, USA. Point source application This study was conducted using a tire pile that was treated with pyriproxyfen. Oviposition containers were placed in four opposite compass directions up to 200 m (2010) and 400 m (2011). Larva bioassays were conducted against water samples collected from each oviposition container for six weeks after treatment Areawide treatment In 2010, approximately 25% were treated in scattered checkboard pattern using backpack sprayers to treat the foliage-covered areas. While in 2011, truck-mounted ULV was sprayed around 24.8% of the areas. BG traps were used to monitor adult mosquitoes and pyriproxyfen autodissemination was assessed via larva bioassays	A direct treatment using point source treatment showed a high number of larval mortalities in 2010 compared to 2011. The mosquitoes were found able to disseminate pyriproxyfen up to 200 meters. However, the area-wide treatment was unable to reduce mosquito populations and rainfall may be playing an important role in reducing the effectiveness of pyriproxyfen in the field.	This proved that autodissemination activities had occur from the direct treatment. The implementation under area wide assessment showed no signifcant impact against the control site.	Suman et al. (2014)
OF	*Ae. aegypti* *Ae. albopictus*	A total of 100 dissemination stations were deployed in 7 ha of sub-areas in Amazonas, Brazil for four months coinciding with the wet season. The pre-treatment took 10 months, intervention (4 months) and post treatment (6 months). The oviposition were collected after six days before further investigation in the laboratory. The percentage of larvae mortality, and contamination of oviposition were assessed seperately.	The coverage of pyriproxyfen contamination was > 85% with sentinel ovitrap distances up to 397 m. Larval bioassay showed 75.1% (after treatment) compared to 4.2% (before treatment) in all ovitraps against all species collected. The mortality decreased over the distance between oviposition sites and dissemination stations.	This study found that high numbers of *Aedes* sp. and mosquitoes would facilitate the efficacy of the autodissemination technique	Abad-Franch et al. (2015)

Note: SF, semi-field; OF, open field; PPF, Pyriproxyfen; EI, emergence inhibitions; d, days; w, weeks; m, months

Table 1 *(Continued)*

Settings	Target species	Methods	Result	Conclusion	References
OF	*Ae. aegypti* *Ae. albopictus*	The study was conducted in two sites in Lexington, USA. Field study 1 – mosquito populations were monitored using BG Sentinel traps in both treatment and control areas. About 4500 males treated with pyriproxyfen were released for 20 weeks from the release point. Oviposition sites were placed around 150 m from the release point. Larva bioassays were conducted against water sampled from oviposition sites to assess emergence inhibition of mosquitoes.	A significant contamination result was observed in larval mortality ranging from 40% to 70% up to 150 m from the release station points. Bioassays demonstrated a significant reduction in mosquito survival compared with control. This evidence showed that the released mosquitoes can transfer the pyriproxyfen to other breeding sites	This study revealed the feasibility of male mosquitoes for autodissemination approaches	Mains et al. (2015)
OF	*Ae. aegypti*	Field evaluation was conducted in a public cemetery, Iquitos Peru. Two treatment sites (Site A: 53 x 4 m; Site B: 52 x 3 m) were treated with 10 dissemination stations and 40 oviposition sites in each site. Oviposition sites consisted of 25 uncontaminted laboratory larvae and 200 ml of water. All oviposition and dessimination stations were located between 1.05 m to 37.5 m from each other.	The use of 3 -5 % breeding site treated with pyriproxyfen caused > 95% contamination in sentinel sites. The overall number of mosquito reduction (EI) was 42 – 98% during the trials.	This is the first study reported in an open field setting with 49 – 84% larval mortality in sentinel sites. However, the locations were close to the dissemination station and oviposition station, which would be unreasonable for an operational program.	Devine et al. (2009)

Note: SF, semi-field; OF, open field; PPF, Pyriproxyfen; EI, emergence inhibitions; d, days; w, weeks; m, months

Table 1 *(Continued)*

Settings	Target species	Methods	Result	Conclusion	References
OF	*Ae. aegypti*	The treatment was conducted in two villages in Rayong Province, Thailand. Dissemination stations were deployed around a 150 m radius circle. Each house was given four dessimination stations. Oviposition sites were deployed to monitor mosquito populations. The autodissemination activity was assessed using bioassays on water samples collected from oviposition sites.	A significant difference was recorded against eggs laid ($p < 0.05$) and larva mortalities ($p < 0.01$). However, there is no evidence of pyriproxyfen transferred between treated devices to water-holding containers.	Even though the BG Trap data showed a reduction in *Aedes* sp. population, there was no clear evidence of autodissemination activity having occurred in the study sites.	Ponlawat et al. (2013)
SF	*Ae. albopictus*	Two different studies were conducted in NJ, USA under semi-field and open field settings. SF – mosquito proof tunnel cage (50 m x 2 m x 3 m) with open and cryptic breeding sites were introduced with gravid females. The number collected from each ovipostion site was used to assess the prefences of oviposition OP-	SF: most of the mosquitoes demonstrated a strong preference for cryptic breeding sites even after exposure to pyriproxyfen.	The exposure of pyriproxyfen during oviposition does not effect the preferences of female mosquitoes to lay their eggs in the cryptic sites. A higher preference for cryptic sites was observed during the study periods.	Chandel et al. (2016)
OF	*Ae. aegypti* *Ae. albopictus*	Three oviposition cups consisting of 100 ml of water and 3rd instar larvae were deployed around the treatment sites. One of the oviposition cup was covered in mesh to prevent any contamination from other mosquitoes and served as control site. All oviposition were deployed for four to five days and were brought back to the laboratory to assess larva mortality and development.	A significant larvae mortality (45%) was observed in the treatment cup (Site 1), however, no significant difference was detected in Site 2. The abundance of ovisites, climatic conditions, and population dynamics may affect the efficacy of ADS.	Local mosquito was found to transfer the pyriproxyfen particles. However, the resultsvaried and were not significant in study 2.	Kartzinel et al. (2016)

Note: SF, semi-field; OF, open field; PPF, Pyriproxyfen; EI, emergence inhibitions; d, days; w, weeks; m, months

Table 1 *(Continued)*

Settings	Target species	Methods	Result	Conclusion	References
OF	*Culex quinquefasciatus*	The study was conducted in CA, USA using a treated catch basin. The inhibition emergence was assessed up to eight weeks by comparing the result from treatment and control catch basin.	A high number of emergence inhibition (100%) was recorded against *Culex quinquefasciatus* for the first 3 weeks. However, the efficacy of the insecticide was reduced to 69% (pyriproxyfen) and 68% (methoprene) at 4 weeks. The result was also obstructed by the high mortality in control sites.	A high mortality in control catch basin. Although the study was not specifically designed for autodissemination approach, there is some evidence showing that autodissemination occurred during the trials	Mian et al. (2017)
OF	*Ae. albopictus*	This study was conducted in NJ, USA. Six hotspots were selected and 26 to 28 autodissemination stations were deployed in the treatment plot. All treatments and control sites were assessed using BG Sentinel, oviposition cups and SBS. Larva bioassays were conducted from water samples collected in each SBS. The trials were conducted for nine weeks following the installation of autodissemination stations.	A significant difference was reported in pupa mortality which was 12.4% compared to control (0.5%) The mean number of egg/ovicup and mean number of larvae in ovitraps were also reduced and were significantly different compared to control sites.	This study demonstrated a clear finding regarding the egg population reduction and larvae mortality obtained from the SBS.	Unlu et al. (2017)
SF	*Ae. aegypti/ Ae. albopictus*	The evaluation was conducted in semi-field settings using 4.6 x 4.6 x 4.6 m adjoining screening rooms. The larvicidal activity was assessed by comparing the mortality of larvae between treatment and control stations. The activity of autodissemination was assessed based on emergence inhibition.	A significant inhibition of emergence was observed i) direct trap (100%), ii) in ovitrap (81% - 94%). The exposure against in2Care for 48 h significantly reduced the longevity of adult mosquitoes.	This commercial device was found to effectively transfer pyriproxyfen to other breeding sites and showed higher mortality rate in every stage of larvae growth.	Buckner et al. (2017)

Note: SF, semi-field; OF, open field; PPF, Pyriproxyfen; EI, emergence inhibitions; d, days; w, weeks; m, months

Table 1 *(Continued)*

Settings	Target species	Methods	Result	Conclusion	References
OF	*Ae. aegypti*	Evaluation of BGST as an autodissemination station was conducted in Madeira, Portugal. The BGST capture bag was treated with pyriproxyfen. Seven adapted BGST and 37 artificial breeding sites (ABS) were distributed randomly throughout the study areas. Each ABS contains 250 ml de-chlorinated water, cat food and 20 third instar larvae. The activity of autodissemination was assessed using larval bioassays.	The autodissemination approach consistently showed an impact against the number of mosquitoes. The effectiveness of the autodissemination strategy was also dependent on the distance, local density of adult mosquitoes and time of the implementations.	BGST was found to be effective as an autodissemination station.	Seixas et al. (2019)
SF	*An. arabiensis*	This study was conducted using the semi-field system (SFS) in Southern Tanzania. Four clay pots served as pyriproxyfen sources and 20 aquatic habitats were introduced into the treatment chambers. Self-propagating populations were established inside the SPS. The emergence larvae were fed on tethered cattle. The activity of autodissemination were assessed by larval bioassays and mortality were compared between control and treatment chambers.	A significant reduction in *An. arabiensis* population compared to the control chamber was observed. Before treatment, the emergence of mosquito was similar between control and treament chambers. Three months after treatment, the mean number of emergence was 5.22 ± 0.42 (control) and 0.14 ± 0.04 (treatment) with > 97% suppression was reported in treatment chambers.	This study demonstrated the abilities of a malaria vector to disseminate pyriproxyfen and subsequently reduced the populations.	Lwetoijera et al. (2019)

Note: SF, semi-field; OF, open field; PPF, Pyriproxyfen; EI, emergence inhibitions; d, days; w, weeks; m, months

Table 1 *(Continued)*

Setting	Target Species	Methods	Result	Conclusion	Result
OF	*Ae. aegypti/ Ae. albopictus*	A total of 80 autodissemination stations with 120 conventional ovitraps were deployed throughout the study areas. the autodissemination stations and conventional ovitraps were placed less than five meters from each other. Pre-intervention study was conducted for two weeks, intervention (3 months) and post intervention study (four weeks).	There was a significant reduction in egg collection after 3 months of intervention by using autodissemination stations ($p < 0.05$). All eggs directly collected from the autodissemination station showed a significant emergence inhibition compared to the conventional ovitrap.	The autodissemination station (Aedestech Mosquito HomeSystem) was found to effectively control *Aedes* sp. population	Yazan et al. (2020)
OF	*Ae. albopictus*	A three years study of autodissemination tehnique was conducted in NJ, USA from 2014 to 2016. A total of 26-28 autodissemination stations, five oviposition cups and ten sentinel cups were deployed in each city block. Both oviposition cups and sentinel cups were sampled and re-deployed at weekly or biweekly intervals. All water samples collected from ovitraps or sentinel cups were then filtered to remove any debris or dead insects. Autodissemination activity in the field were assessed using larval bioassays. Pupal and larval mortality were recorded for statistical analysis.	The larva and pupal mortalities ranged from 12-19% in the treatment areas. Even though the number of eggs was significantly reduced in 2014, there was no significant difference observed in 2015 and 2016. The impact of autodissemination against mosquito populations was lower compared to previous trials. Several problems and limitations were observed due to the reduction in autodissemination effectiveness such as logistical and technical problems (blockage and exit size of ADS) and limited access to individual properties	The autodissemination station successfully transferred pyriproxyfen to other breeding sites. However, the impact of pyriproxyfen against mosquito populations was low in 2015 - 2016 and it remains unclear and need further investigations.	Unlu et al. (2020)

Note: SF, semi-field; OF, open field; PPF, Pyriproxyfen; EI, emergence inhibitions; d, days; w, weeks; m, months

More recently, the effectiveness of autodissemination methods using pyriproxyfen has been studied in Trenton, New Jersey, which was reported to have a high *Ae. albopictus* population (Unlu et al., 2017). The study showed a reduction in the number of dengue cases, fewer numbers of eggs and residual effects of pyriproxyfen within the treatment areas. Autodissemination has also been studied under different conditions in small-scale field trials against *Ae. albopictus* and *Ae. aegypti* (Buckner et al., 2017; Caputo et al., 2012). Moreover, in 2014, the first study against *Anopheles arabiensis* was conducted in Tanzania, Africa, which observed a 82% inhibition of adult mosquitoes by using pyriproxyfen autodissemination and similar results were observed against *Culex quinquefasciatus* and *Anopheles gambiae* (Mbare et al., 2014). This strategy has shown promising results and can be considered as an alternative vector control against malaria in the future (Lwetoijera et al., 2014). However, a large-scale application of the autodissemination approach needs further investigation.

Recent Advances in Autodissemination on Mosquitoes

A recent advance in autodissemination is the boosted sterile insect technique (SIT) which is a combination of two methods: releasing sterile mosquitoes and contaminating them with pyriproxyfen. Despite the failure of mating, the percentage of pyriproxyfen contamination between the populations is high. During the mating, male mosquitoes will contaminate females and at the same time can contaminate the breeding sites too. Currently, ongoing research has been conducted to explore the effects of pyriproxyfen against mosquito competitiveness, the amount of pyriproxyfen needed from males to females and then, from females to breeding sites and the design of the smart machines to release the mosquitoes in a large scale (Bouyer & Lefrancois, 2014).

One study enhanced the persistence of the insecticide and attraction using a dual treatment system involving an EC formulation and granule formulation for the autodissemination stations and showed that the mosquito can carry the insecticide up to 200 m along the residential areas in the study as detected by residue analysis (Suman et al., 2017). Other studies using a combination of *Beauveria bassina* (entomopathogenic fungus) and pyriproxyfen also gave 100% larval mortality in the laboratory; however, the study did not include any field trials (Buckner et al., 2017; Snetselaar et al., 2014). Another approach used a combination of oil and pyriproxyfen to enhance the attachment of the particle on female mosquitoes and increase the transfer of pyriproxyfen to the oviposition sites (Wang et al., 2014).

An Innovation of Autodissemination Device

An autodissemination station is a new device that has shown promising results in different settings. This device significantly reduced *Aedes* mosquito populations and subsequently inhibited the transmission of dengue cases. In general, most autodissemination stations

use a similar concept of attracting mosquitoes to lay their eggs into the traps. Female mosquitoes were then contaminated with pyriproxyfen and transferred the particles to other breeding containers. Some autodissemination stations were developed with their own unique features while still maintaining the basic concept of autodissemination. It is important to determine the effectiveness of prototype autodissemination stations based on their ability to serve as a lethal ovitrap, high mosquito attraction, low cost, biodegradability, and ease of maintenance (Gaugler et al., 2012). These stations are described below and may become widely available with a significant impact on dengue control programs. Thus, a comparative study of autodissemination devices is warranted.

Auto-Dissemination Augmented by Males (ADAM)

Auto-Dissemination Augmented by Males (ADAM) was developed by using a combination of autodissemination and autocidal concepts, and by enhancing the abilities of male *Aedes* mosquitoes (Mains et al., 2015). This technique uses mosquitoes as a transporter to transfer pyriproxyfen via mating attempts, and subsequently, it will transfer the pyriproxyfen to cryptic breeding sites during the oviposition. Corbel et al., (2017) found a significant cross-contamination between males and females who were suitable to release the ADAM in areas of low density *Aedes* populations. Although the density of adult mosquitoes declined after four weeks of treatment, further studies on the mosquito behavior, study area with a different ecological context, as well as the size of the field trials are needed to validate the efficacy of ADAM in the vector control program (Corbel et al., 2017).

In2Care Mosquito Trap

The In2Care mosquito trap has been approved by the United States Environmental Protection Agency (EPA) for use by professionals against *Aedes* mosquitoes. It is the first commercial trap that controls both larvae and adult mosquitoes using pyriproxyfen and *B. bassiana,* a fungus with slow-killing abilities on adult mosquitoes, which can be disseminated by affected female mosquitoes. During the oviposition, female mosquitoes will touch their tarsal on a gauze contaminated with pyriproxyfen and *B. bassiana*. Both agents will stick to the mosquito bodies and be able to disseminate pyriproxyfen to other breeding sites within their flight range (Buckner et al., 2017). The In2Care mosquito traps manipulate the behavior of female mosquitoes by "skip oviposition" and finding cryptic breeding sites. At the same time, the *B. bassiana* spores will grow hyphae, feed upon the bodies, reproduce and kill the mosquitoes within 18 hours (Ragavendran et al., 2017).

AedesTech Mosquito Home (AHM)

AedesTech Mosquito Home (AHM) works by using a "pull" and "push" control strategy combined with the lure-kill technology to reduce *Aedes* populations and interrupting dengue transmission. It has the potential to target larval breeding sites that cannot be reached by

traditional larvicidal applications (cryptic breeding sites). The "pull" component uses a formulation with an attractant ingredient lure against female mosquitoes to breed inside the AHM while the "push" component refers to the contaminated mosquitoes dispersing after ovipositing their eggs. In this method, female mosquitoes are used as a carrier for the transfer of insecticide to other containers, and thus disseminating the insecticide to the larval habitats (Liang et. at., 2019). The AHM devices have been designed with a sustainable feature in which gravity is used to control the level of water in the casing. During evaporation, water is released in the casing, filling it back to the pre-set level and the evaporation rate will always be kept at the optimum level which is sufficient for the development of *Aedes* from egg to pupae.

STATUS AND ONGOING STUDIES IN MALAYSIA

Studies on autodissemination using pyriproxyfen are not well established in Malaysia. Most of the studies conducted were only to evaluate the effectiveness of pyriproxyfen against *Aedes* spp. mosquitoes. In a study by Nazni et al. (2015), a total of 350 autodissemination devices were deployed in three apartment blocks consisting of 27 storeys with ten-unit houses per floor. Four units of autodissemination were placed at each level at a distance of 5 m from apartment units. After the exposure periods, there were no viable larvae in the traps, and a significant reduction in number of dengue cases was observed from 53 cases (2013) to 13 cases in 2014. Another study was carried out to explore the additive impact of targeted outdoor residual spraying (TORS) and deployment of autodissemination devices (ADD) in Johor Bharu, Malaysia. Unfortunately, there was no significant effect on the mosquito populations following the combination of TROS and ADD (Hamid et al., 2020).

CONCLUSION

Vector control programs for arboviruses transmitted by mosquitoes are facing many obstacles. However, with new advances and technologies, it may be viable in the foreseeable future. One of these advances is autodissemination which is a self-delivery technique that manipulates the behavior of mosquitoes to carry insecticides and disseminate it to cryptic breeding sites. Autodissemination has shown promising results in several countries and can be considered as additional tools in a vector control program.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the Dengue Tech Challenge 2016 (Application Reference DTC 16022), funded through the High-Impact Programme (HIP2) of PlaTCOM Ventures and the Newton-Ungku Omar Fund of the British Council. We thank One Team Networks Sdn Bhd and Prof Dr. Srijit Das for helping with the valuable comment and useful suggestions.

REFERENCES

Abad-Franch, F., Zamora-Perea, E., Ferraz, G., Padilla-Torres, S. D., & Luz, S. L. B. (2015). Mosquito-disseminated pyriproxyfen yields high breeding-site coverage and boosts juvenile mosquito mortality at the neighborhood scale. *PLoS Neglected Tropical Diseases, 9*(4), 1-17.

Ahmad-Azri, M., Syamsa, R. A., Ahmad-Firdaus, M. S., & Aishah-Hani, A. (2019). A comparison of different types of ovitraps for outdoor monitoring of Aedes mosquitoes in Kuala Lumpur. *Tropical Biomedicine, 36*(2): 335-347.

Arham, A. F., Razman, M. R., Amin,L., & Mahadi, Z. (2018). Dengue review: Issues, challenges and public attitudes. *International Journal of Academic Research in Business and Social Sciences, 8*(4), 956-972.

Awang, N., Kosnon, N. A., Othman, H., & Kamaludin, N. F. (2012). The effectiveness of organotin (IV) benzylisopropyldithiocarbamate compounds as insecticides against *Aedes aegypti* Linn (Diptera: Culicidae) in laboratory. *American Journal of Applied Sciences, 9*(8), 1214-1218.

Besar, N. A. U. A., Sulaiman, A., Asri, L. N., & Khairuddin, K. (2019). Resistance status of *Aedes aegypti* towards different insecticides in selected dengue outbreak area in Petaling District (Diptera: Culicinae). *Serangga, 24*(2), 41-48.

Banerjee, S., Mohan, S., Pramanik, S., Banerjee, S., Saha, G. K., & Aditya, G. (2017). Effect of food types on competitive interaction between *Aedes aegypti* (Linnaeus, 1762) and *Ae. albopictus* (Skuse, 1894) (Diptera: Culicidae): A proximate level appraisal. *Polish Journal of Entomology, 86*, 99-118.

Banu, S., Hu, W., Guo, Y., Naish, S., & Tong, S. (2014). Dynamic spatiotemporal trends of dengue transmission in the Asia-Pacific region, 1955-2004. *PLoS One*, *9*(2), 1-7.

Boubidi, S. C., Roiz, D., Rossignol, M., Chandre, F., & Benoit, R. (2016). Efficacy of ULV and thermal aerosols of deltamethrin for control of *Aedes albopictus* in Nice, France. *Parasites and Vectors*, *9*(1), 1-8.

Bourtzis, K., Lees, R. S., Hendrichs, J., & Vreysen, M. J. B. (2016). More than one rabbit out of the hat: Radiation, transgenic and symbiont-based approaches for sustainable management of mosquito and tsetse fly populations. *Acta Tropica*, *157*, 115-130.

Bouyer, J., & Lefrancois, T. (2014). Boosting the sterile insect technique to control mosquitoes. *Trends in Parasitology*, *30*(6), 271-273.

Brady, O. J., Johansson, M. A., Guerra, C. A., Bhatt, S., Golding, N., Pigott, D. M., … & Styer, L. M. (2013). Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasites and Vectors, 6*(1), 1-12.

Buckner, E. A., Williams, K. F., Marsicano, A. L., Latham, M. D., & Lesser, C. R. (2017). Evaluating the vector control potential of the in2care® mosquito trap against *Aedes aegypti* and *Aedes albopictus* under semifield condition. *Journal of American Mosquito Control Association*, *33*(3), 193-199.

Buhler, C., Winkler, V., Runge-Razinger, S., Boyce, R., & Horstick, O. (2019). Environmental methods for dengue vector control – A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases, 13*(7), 1-15.

Camara, D. C. P., Codeco, C. T., Juliano, S. A., Lounibos, L. P., Riback, T, I, S., & Honorio, N. A. (2016). Seasonal differences in density but similar competitive impact of *Aedes albopictus* (Skuse) on *Aedes aegypti* (L.) in Rio de Janeiro, Brazil. *PloS ONE, 11*(6), 1-15.

Caputo, B., Lenco, A., Cianci, D., Pombi, M., Petrarca, V., Baseggio, A., … & della Torre, A. (2012). The autodissemination approach: A novel concept to fight *Aedes albopictus* in urban areas. *PLoS Neglected Tropical Disease*, *6*(8), 1-8.

Caraballo, H., & King, K. (2014). Emergency department management of mosquito-borne illness: Malaria, dengue and west nile virus. *Emergency Medicine Practice, 16*(5), 1-23.

Chandel, K., Suman, D. S., Wang, Y., Unlu, I., Wiliges, E., Williams, G. M., & Gaugler, R. (2016). Targeting a hidden enemy: Pyriproxyfen autodissemination strategy for the control of the container mosquito *Aedes albopictus* in cryptic habitats. *PLoS Neglected Tropical Diseases, 10*(12), 1-15.

Chang, M. S., Christophel, E. M., Gopinath, D., & MdAbdur, R. (2011). Challenges and future perspective for dengue vector control in the Western Pacific Region. *Western Pacific Surveillance and Response Journal, 2*(2), 9-16.

Chism, B. D., & Apperson, C. S. (2003). Horizontal transfer to the insect growth regulator pyriproxyfen to larval microcosms by gravid *Aedes albopictus* and *Ochlerotus triseriatus* mosquitoes in the laboratory. *Medical and Verterinary Entomology*, *17*, 211-220.

Corbel, V., Fonseca, D. M., Weetman, D., Pinto, J., Achee, N. L., Chandre, F., … & David, J. P. (2017). International workshop on insecticide resistance in vectors of arboviruses, December 2016, Rio de Janeiro, Brazil. *Parasites and Vectors*, *10*, 1-16.

Darabi, H., Vatandoost, H., Abaei, M. R., Gharibi, O., & Pakbaz, F. (2011). Effectiveness of methoprene, an insect growth regulator, against malaria vectors in Fars, Iran: A field study. *Pakistan Journal of Biology Sciences, 14*, 69-73.

Delatte, H., Desvars, A., Boutard, A., Bord, S., Gimonneau, G., Vourc'h, G., & Fontenille, D. (2010). Blood-feeding behaviour of *Aedes albopictus* a vector of chikungunya on La Reunion. *Vector Borne and Zoonotic Diseases*, *10*(3), 249-258.

Devine, G. J., Perea, E. Z., Killeen, G. F., Stancil, J. D., Clark, S. J., & Morrison, A. C. (2009). Using adult mosquitoes to transfer insecticides to *Aedes aegypti* larval habitats. *Proceedings of the National Academy of Sciences*, *106*(28), 11530-11534.

Evans, B. R., Kotsakiozi, P., Costa-da-Silva, A. L., Ioshino, R. S., Garziera, L., Pedrosa, M. C., … & Powell, J. R. (2019). Transgenic *Aedes aegypti* mosquitoes transfer gene into a natural population. *Scientific Reports, 9*(1), 1-6.

Faraji, A., Egizi, A., Fonseca, D. M., Unlu, I., Crepeau, T., Healy, S. P., & Gaugler, R. (2014). Comparative host pattern of the tiger asian tiger mosquito, *Aedes albopictus*, in urban and suburban Northeastern USA and implication for disease transmission. *PLoS Neglected Tropical Diseases*, *8*(8), 1-11.

Gaugler, R., Suman, D., & Wang, Y. (2012). An autodissemination station for the transfer of an insect growth regulator to mosquito oviposition sites. *Medical and Veterinary Entomology*, *26*, 37-45.

Gerding, J., Kirshy, M., Moran, J. W., Bialek, R., Lamers, V., & Sarisky, J. (2016). A performance management initiative for local health department vector control programs. *Environmental Health Insight*, *10*, 113-118.

Gross, H. R., Hamm, J., & Carpenter, J. E. (1994). Design and application of a hive-maintained device uses honey bee (Hymenoptera: Apidae) to disseminate *Heliothis nuclear polyhedrosis* virus. *Environmental Entomology*, *23*, 492-501.

Hamid, N. A., Alexander, N., Suer, R., Ahmed, N. W., Mudin, R, N., Omar, T., ... & Saadatian-Elahi, M. (2020). Targeted outdoor residual devices and their combination against *Aedes* mosquitoes: Field implementation in a Malaysia urban settings. *Bulletin of Entomological Research, 110*(4)*,* 1-8.

Harburguer, L., Zerba, E., & Licastro, S. (2014). Sublethal effect of pyriproxyfen released from a fumigant formulation on fecundity, fertility and ovicidal action in *Aedes aegypti* (Diptera: Culicidae). *Journal of Medical Entomology*, *51*(2), 436-443.

Hopperstad, K. A., & Reiskind, M. H. (2016). Recent changes in the local distribution of *Aedes aegypti* (Diptera: Culicidae) in South Florida, USA. *Journal of Medical Entomology*, *53*(4), 836-842.

Huong, V. D., Ngoc, N. T. B., Hein, D. T., & Lien, N. T. B. (2004). Susceptibility of *Aedes aegypti* to insecticide in Viet Nam. *Dengue Bulletin, 28,* 179-183.

Itoh, T. (1993). Control of DF/ DHF vectors, *Aedes* mosquito, with insecticides. *Tropical Medicine,35*(4), 259-267.

Itoh, T., Kawada, H., Abe, A., Eshita, Y., Rongsriyam, Y., & Igarashi, A. (1994). Utilization of bloodfed females of *Aedes aegypti* as vehicle for the transfer of the insect growth, pyriproxyfen to larval habitats. *Journal of American Mosquito Control Association*, *10*(3), 344-347.

Kartzinel, M. A., Alto, B. W., Deblasio, M. W., & Burkertt-Cadena, N. D. (2016). Testing of visual and chemical attractants in correlation with the development and field evaluation of an autodissemination station for the suppression of *Aedes aegypti* and *Aedes albopictus* in Florida. *Journal of American Mosquito Control Association*, *32*(3), 194-202.

KKM. (2020). *Kenyataan akhbar ketua pengarah kesihatan Malaysia mengenai situasi semasa demam denggi, zika dan chikungunya di Malaysia -ME 23.2020* [Press Statement Minister of Health Malaysia – Current Situation of Dengue, Chikungunya and Zika in Malaysia – EW 23. 2020]. Retrieved July 17, 2020, from https://www.moh.gov.my/index.php/database_stores/store_view_page/21/1622

Lau, K. W., Chen, C. D., Lee, H. L., & Sofian-Azirun, M. (2015). Evaluation of insect growth regulators, temephos and *Bacillus thuringiensis israelensis* against *Aedes aegypti* (L) in plastic containers. *Tropical Biomedicine*, *32*(4), 684-692.

Liang, Y., Ahmad Mohiddin, M. N., Bahauddin, R., Hidayatul, F. O., Nazni, W. A., Lee, H. L., & Greenhalgh, D. (2019). Modelling the effect of a novel autodissemination trap on the spread of dengue in Shah Alam and Malaysia. *Computational and Mathematical Methods in Medicine, 2019*(1923479), 1-15.

Lwetoijera, D., Harris, C., Kiware, S., Dongus, S., Devine, G. J., McCall, P. J., & Majambere. (2014). Effective autodissemination of pyriproxyfen to breeding sites by the exophilic malaria vector *Anopheles arabiensis* in semi-field setting in Tanzania. *Malaria Journal*, *13*(1), 1-10.

Lwetoijera, D., Kiware, S., Okumu, F., Devine, G. J., & Majambere, S. (2019). Autodissemination of pyriproxyfen suppress stable populations of *Anopheles arabiensis* under semi-controlled settings. *Malaria Journal, 18*(1), 1-10.

Mains, J. W., Brelsfoard, C. L., & Dobson, S. L (2015). Male mosquitoes as vehicle for inscticide. *PLoS Neglected Tropical Diseases*, *9*(1), 1-17.

Mazrura, S., Rozita, H., Hidayatulfathi, O., Zainudin, M. A., Naim, M. R., Nadia, A. M. N., ... & Joy, J. P. (2010). Community vulnerability on dengue and its association with climate variability in Malaysia: A public health approach. *Malaysia Journal of Public Health Medicine, 10*(2), 25-34.

Mazrura, S., Othman, H., Nor, N. A. M., Hod, R., Ali, Z. M., Rasidi, M. N. M., … & Choy, E. A. (2012). Kajian ekologi nyamuk *Aedes* di Senawang Negeri Sembilan, Malaysia [Ecology Survey on *Aedes* Mosquito in Senawang, Negeri Sembilan]. *Sains Malaysiana, 41*(2), 261-269.

Mbare, O., Lindsay, S. W., & Filinger, U. (2014). Pyriproxyfen for mosquito control: Female sterilization or horizontal to oviposition substrate by *Anopheles gambiae* sensu stricto and *Culex quinquefasciatus*. *Parasites and Vectors*, *7*(1), 1-12.

Mian, L. S., Dhillon, M. S., & Dodson, L. (2017). Field evaluation of pyriproxyfen against mosquitoes in catch basins in Southern California. *Journal of American Mosquito Control Association, 33*(2), 145-147.

Moslim, R., Kamarudin, N., & Wahid, M. B. (2011). Trap for autodissemination of *Metarhizium anisopliae* in the management of rhinoceros beetle*, Oryctes rhinoceros*. *Journal of Oil Palm Research*, *23*, 1011-1017.

Nazni, W. A., Teoh, G. N., Farah, H., Suhana, O., Sakinah, A., Chandru, A., … & Lee, H. L. (2015, March 3-4). Field evaluation of auto-dissemination of dengue in high-rise condominium in Selangir. In *51st Annual Scientific Seminar of the Malaysian Society of Parasitology and Tropical Medicine.* Kuala Lumpur, Malaysia.

Niang, E. H. A., Bassene, H., Fenollar, F., & Mediannikov, O. (2018). Biological control of mosquito-borne diseases: the potential of Wolbachia-based interventions in an IVM framework. *Journal of Tropical Medicine, 2018*, 1-15.

Ohba, S. Y., Ohashi, K., Pujiyati, E., Higa, Y., Kawada, H., Nobuaki, M., & Masairo, T. (2013). The effect of pyriproxyfen as a "population growth regulator" against *Aedes albopictus* under semi-field condition. *PLoS One*, *8*(7), 1-10.

Othman, H., Zaini, Z. I., Karim, N., Rashid, N. A. A., Abas, M. B. H., Sahani, M., ... & Nor, N. A. M. (2019). Applying health belief model for the assessment of community knowledge, attitude and prevention practices following a dengue epidemic in a township in Selangor, Malaysia. *International Journal of Community Medicine and Public Health, 6*(3): 958-970.

Palli, S. R. (2016). Hormonal regulation of development and reproduction. In H. Czosnek & M. Ghanim (Eds.), *Management of insect pest to agriculture* (pp. 99-114). Cham, Switzerland: Springer.

Pell, J. K., Macaulay, E. D. M., & Wilding, N. (1993). A pheromone trap for dispersal of the pathogen *Zoophthora radicans* Brefeld (Zygomycetes: Entomopthorales) amongst populations of the diamondback moth, *Plutella xylostella* L. (Lepidoptera: Yponomeutidae). *Biocontrol Science and Technology, 3,* 315-320.

Ponlawat, A., Fansiri, T., Kurusarttra, S., Pongsiri, S., McCardle, P. W., Evan, B. P., & Ricardson, J. H. (2013). Development and evaluation of a pyriproxyfen-treated device to control the dengue vector, *Aedes aegypti* (L.) (Diptera: Culicidae). *The Southeast Asian Journal of Tropical Medicine and Public Health*, *44*(2), 167-178.

Ragavendran, C., Dubey, N. K., & Natarajan, D. (2017). *Beauveria bassiana* (Clavicipitaceae): A potent fungal agent for controlling mosquito vectors of *Anopheles stephensi*, *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae). *RSC Advances*, *7*, 3838-3851.

Rasgon, J. L. (2011). Mosquitoes attacked from within. *Nature*, *476*, 407-408.

Scholte, E. J., Knols, B. G. J., & Takken, W. (2004). Autodissemination of the entomopathogenic fungus *Metarhizium anisopliae* amongst adult of the malaria vector *Anopheles gambiae* s.s. *Malaria Journal*, *3*(1), 1-6.

Schorkopf, D. L. P., Spanoudis, C. G., Mboera, L. E. G., Mafra-Neto, A., Ignell, R., & Dekker, T. (2016). Combining attractants and larvicides in biodegradebale matrices for sustainable mosquito vector control. *PLoS Neglected Tropical Diseases, 10*(10), 1-22.

Seccacini, E., Lucia, A., Harburguer, L., Zerba, E., Licastro, S., & Masuh, H. (2008). Effectiveness of pyriproxyfen and diflubenzuron formulations as larvicides against Aedes aegypti. *Journal of the American Mosquito Control Association, 24*(3), 398-403.

Seixas, G., Paul, R. E., Pires, B., Alves, G., de Jesus, A., Silva, A. C., ... & Sousa, C. A. (2019). An evaluation of efficacy of the auto-dissemination technique as a tool for Aedes aegypti control in Madeira, Portugal. *Parasites and Vectors, 12*(1), 1-13.

Sihuincha, M., Zamora-Perea, E., Orellana-Rios, W., Stancil, J. D., Sifuentes, V. L., Vidal-Ore, C., & Devine, G. J. (2005). Potential use of pyriproxyfen for control of *Aedes aegypti* (Diptera: Culicidae) in Iquitos, Peru. *Journal of Medical Entomology, 42*(4), 620-630.

Snetselaar, J., Andriessen, R., Suer, R. A., Osinga, A. J., Knol, B. G. J., & Farenhost, M. (2014). Development and evaluation of a novel contamination device that targets multiple life-stages of *Aedes aegypti*. *Parasite and Vectors*, *7*(1), 1-10.

Sukumaran, B., Kandasamy, K., Ramanathan, S., Tom, A., & Rajagopal, S. S. (2019). Review on current status of dengue and its prevention in India. *International Journal of Research in Pharmaceutical Sciences, 10*(4), 2748-2754.

Sulaiman, S., Fadhlina, K., & Othman, H. (2007). Evaluation of pyrethtin formulations on dengue/dengue haemorrhagic fever vectors in the laboratory and sublethal effects. *Iranian Journal Athropod-Borne Diseases, 1*(2), 1-6.

Suman, D. S., Farajollahi, A., Healy, S., Williams, G. M., Wang, Y., Schoeler, G., & Gaugler, R. (2014). Point-source and area wide field studies of pyriproxyfen autodissemination against urban container-inhabiting mosquitoes. *Acta Tropica*, *135*, 96-103.

Suman, D. S., Wang, Y., Faraji, A., Williams, G. M., Wiliges, E., & Gaugler, R. (2017). Seasonal field efficacy of pyriproxyfen autodissemination stations against container-inhabiting mosquito *Aedes albopictus* under different habitat conditions. *Pest Management Science*, *74*(4), 885-895.

Sun, W., Xue, L., & Xie, X. (2017). Spatial-temporal distribution of dengue and climate characteristics for two clusters in Sri Lanka from 2012 to 2016. *Scientific Reports, 7*(1), 1-12.

Unlu, I., Rochlin, I., Suman, D. S., Wang, Y., Chandel, K., & Gaugler, R. (2020). Large-scale operational pyriproxyfen autodissemination deployment to suppress the immatureasian tiger mosquito (Diptera: Culicidae) populations. *Journal of Medical Entomology, 57*(4), 1120-1130.

Unlu, I., Suman, D. S., Klinger, K., Faraji, A., & Gaugler, R. (2017). Effectiveness of autodissemination stations containing pyriproxyfen in reducing immature *Aedes albopictus* populations. *Parasites and Vectors*, *10*(1), 1-10.

Vega, F. E., Dowd, P. F., Lacey, L. A., Pell, K., Jackson, D. M., & Klein, M. (2000). Trapping for autodissemination of entomopathogens. In L. A. Lacey & H. Kaya (Eds.), *Field manual techniques in insect pathology* (pp. 153-176). Dordrecht, The Netherlands: Kluwer Academic Publisher.

Wang, Y., Suman, D. S., Bertrand, J., Dong, L., & Gaugler, R. (2014). Dual-treatment autodissemination station with enhances transfer of an insect growth regulator to mosquito oviposition sites. *Pest Management Sciences*, *70(8),* 1299-304.

WHO. (2012). *Handbook for integrated vector management.* Retreived June 18, 2020, from https://apps.who.int/iris/bitstream/handle/10665/44768/9789241502801_eng.pdf;jsessionid=B4DDA992C5E2E68501690DE62289B785?sequence=1

WHO. (2016). *Fact sheet on vector-borne disease*. Retreived March 30, 2020, from https://www.who.int/en/news-room/fact-sheets/detail/vector-borne-diseases

WHO. (2020). *Global strategy for dengue prevention and control 2012-2020*. Retreived March 30, 2020, from https://apps.who.int/iris/bitstream/handle/10665/75303/9789241504034_eng.pdf?sequence=1

Wong, P. S., Li, M. Z. I., Chong, C. S., Ng, L. C. & Tan, C. H. (2013). *Aedes* (Stegomyia) *albopictus* (Skuse): A potential vector of zika virus in Singapore. *PLoS Neglected Tropical Diseases, 7*(8), 1-5.

Xiang, J., Hansen, A., Liu, Q., Liu, X., Tong, M. X., Sun, Y., … & Bi, P. (2017). Association between dengue fever incidence and meteorological factors in Guangzhou, China, 2005-2014. *Environmental Research*, *153*, 17-26.

Yazan, L. S., Paskaran, K., Gopalsamy, B., & Majid, R. A. (2020). Aedestech mosquito home system prevents the hatch of *Aedes* mosquito eggs and reduces its population. *Pertanika Journal Science and Technology, 28*(1), 263-278.

Yousaf, M. Z., Siddique, A., Ashfaq, U. A., & Ali, M. (2018). Scenario of dengue infection & its control in Pakistan: an update and way forward. *Asian Pacific Journal of Tropical Medicine*, *11*(1), 15-23.

Zaini, Z. I. I., Othman, H., Karim, N., Rashid, N. A. A., Abas, M. B. H., Sahani, M., … & Nordin, S. A. (2019). Knowledge and practices regarding *Aedes* control amongst residents of dengue hotspot area in Selangor: A cross-sectional study. *Sains Malaysiana, 48*(4), 841-849.

Pertanika J. Sci. & Technol. 28 (4): 1141 - 1161 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Predictive Performance of Logistic Regression for Imbalanced Data with Categorical Covariate

Hezlin Aryani Abd Rahman[1]*, Yap Bee Wah[1,2] and Ong Seng Huat[3]

[1]Centre of Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia
[2]Advanced Analytics Engineering Centre, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia
[3]Department of Actuarial Science and Applied Statistics, UCSI University, 56000, Kuala Lumpur, Malaysia

ABSTRACT

Logistic regression is often used for the classification of a binary categorical dependent variable using various types of covariates (continuous or categorical). Imbalanced data will lead to biased parameter estimates and classification performance of the logistic regression model. Imbalanced data occurs when the number of cases in one category of the binary dependent variable is very much smaller than the other category. This simulation study investigates the effect of imbalanced data measured by imbalanced ratio on the parameter estimate of the binary logistic regression with a categorical covariate. Datasets were simulated with controlled different percentages of imbalance ratio (IR), from 1% to 50%, and for various sample sizes. The simulated datasets were then modeled using binary logistic regression. The bias in the estimates was measured using Mean Square Error (MSE). The simulation results provided evidence that the effect of imbalance ratio on the parameter estimate of the covariate decreased as sample size increased. The bias of the estimated depends on sample size whereby for sample size 100, 500, 1000 – 2000 and 2500 – 3500, the estimated were biased for IR below 30%, 10%, 5% and 2% respectively. Results also showed that

ARTICLE INFO

Article history:
Received: 13 February 2020
Accepted: 27 July 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.02

E-mail addresses:
hezlin@tmsk.uitm.edu.my (Hezlin Aryani Abd Rahman)
beewah@tmsk.uitm.edu.my (Yap Bee Wah)
ongsh@ucsiuniversity.edu.my (Ong Seng Huat)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

parameter estimates were all biased at IR 1% for all sample size. An application using a real dataset supported the simulation results.

Keywords: Categorical covariate, imbalanced data, logistic regression, parameter estimates, predictive analytics, simulation

INTRODUCTION

Imbalanced data are a condition where the dependent variable contains one class which has more observations than the other. Imbalanced data will have prominent effect on the classification performance of classifiers such as logistic regression, decision trees, support vector machine (SVM) and artificial neural network (ANN). Imbalanced data also affects the classification "power" of various classifiers. The effect of imbalanced data has been reported by researchers through the application of real data sets (Blagus & Lusa, 2010 ; Longadge et al., 2013; Ramyachitra & Manikandan, 2014).

Logistic regression (LR) is frequently used in predictive modeling as a benchmark model when other classifiers' performances were evaluated. It is a conventional statistical model used widely in business, engineering, and social science research (Hamid, 2016 ; Hamid et al., 2018 ; Ahmad et al., 2011 ; Shariff et al., 2016 ; Yap et al., 2014), and medical and healthcare studies (Longadge et al., 2013; Mena & Gonzalez, 2006; Oztekin et al., 2009; Pourahmad et al., 2011; Rothstein, 2015; Roumani et al., 2013; Srinivasan & Arunasalam, 2013; Uyar et al., 2010). However, the presence of imbalanced data challenges LR's ability to classify, whereby majority of classifiers normally focus in the prediction without consideration on the relative distribution between the classes (Dong et al., 2014). Normally, when class imbalance are present, classification results from standard classifiers are biased towards the majority class. As a result, if the event of interest is the minority class, the sensitivity of the classifier will be zero and the specificity will be 100%. The real dataset in reality often suffers from some imbalance problem (Goel et al., 2013) and the minority class is often misclassified (Chawla et al., 2004; He & Garcia, 2009; Weiss & Provost, 2003). Thus, whenever class imbalance is found in healthcare and medical datasets, the credibility of the models generated by the classifiers are often misleading.

Imbalanced problem affects standard classifiers (Chawla, 2003; Cohen et al., 2006; Galar et al., 2011) and logistic regression based on application to real datasets studies (Blagus & Lusa, 2010; Burez & Van den Poel, 2009; Mena & Gonzalez, 2006; Van Hulse et al., 2007). In our previous study, we performed simulation to study the impact of imbalanced ratio (IR) on LR parameter (β) estimates and the odds ratio (e^{β}) of the LR model using a continuous covariate (Rahman & Yap, 2016). The results provided enough evidence to conclude that extreme imbalanced ratio (IR = 1%, 2%, 5%) and small sample

size had more serious effect on parameter estimates of LR model. Imbalanced ratio is the ratio of the number of cases in minority class to the majority class. For example, if the response variable is the presence of cancer and has two categories Cancer or No Cancer, the imbalance ratio is n_1/n_0, where n_1 is the number of patients diagnosed with cancer while n_0 is the number of patients who do not have cancer.

The effect of imbalanced data on the performance of the classifiers can be determined through simulation studies. In addition, the various types (categorical or continuous) of variables in a set of data might show different effects. In this simulation study, we focus on the logistic regression model, a useful statistical model for classification problem and investigate the imbalanced effects on the parameter estimate of the model with a single categorical covariate.

The aim of this study was to determine the effects of different IR on the logistic regression parameter estimate via simulation and an application to real dataset. The results of this study will guide practitioners on the severity of bias in estimates as a result imbalanced data.

MATERIALS & METHOD

Review on Methods

Machine learning techniques i.e. LR, DT, ANN and SVM, may have great classification performance if they involve balanced data. However, these techniques performs poorly when imbalanced problem arises (Anand et al., 2010).

Most studies concluded that there was an effect of IR towards the performance of standard classifiers (Rahman et al., 2012; Chawla, 2003; Lemnaru et al., 2012 ; Mena & Gonzalez, 2006; Prati et al., 2014; Van Hulse et al., 2007; Yap et al., 2014). A study by Mena and Gonzalez (2006) introduced a 3-step algorithm using simple LR called Rule Extraction Medical Diagnosis (REMED) which enabled users to select attributes for the model and improved the accuracy of the model by adjusting the percentage of the partition. Although REMED's algorithm claimed to improve the prediction accuracy, it was limited to medical diagnostics. Lemnaru et al. (2012) reported that IR, size and complexity of the dataset affected the predictive performance of different classifiers (k-nearest neighbor (KNN), C4.5, SVM, multi-layered perceptron (MLP), Naïve Bayes (NB), and Adaboost (AB). In their extensive study, the IR was categorized into three categories (balance, small, large), four categories of dataset size (very small, small, medium, and large) and four categories of complexity of the dataset (small, medium, large and very large). They concluded that the performance of the classifiers was lower when the IR was high. Another extensive experiment performed by Van Hulse et al. (2007), using different sampling strategies (random oversampling (ROS), random undersampling (RUS), one-sided selection (OSS), cluster-based oversampling (CBOS), Wilson's editing (WE), SMOTE (SM), and borderline-

SMOTE (BSM) on different classifiers (NB, DT C4.5, LR, random forest (RF), and SVM) on 35 real datasets with different ratio of imbalance (1.33% - 34.90%), concluded that sampling strategy improved the performance of the chosen classifiers. However, their study also concluded that there was no one universal sampling strategy that worked best for all classifiers. Chawla (2003) experimented on five real datasets using C4.5 as the classifier and reported that their synthetic sampling method, SMOTE, improved the performance of the classifier better than other sampling strategies. He also concluded that RUS was better than ROS with replication. Prati et al. (2014) also experimented on 22 real datasets with different IR on different classifiers (C4.5, C4.5Rules, CN2 and RIPPER, Back-propagation Neural Network, NB and SVM) and by using different sampling strategies (ROS, SMOTE, borderline-SMOTE, AdaSyn, and MetaCost). They concluded that in terms of accuracy (AUC), the rule-based algorithm (C4.5Rule, RIPPER) was the most affected while Support Vector Machine (SVM) was least affected by imbalanced data. However, the authors also stated that severe imbalanced class distributions would have a strong influence on SVM and any classifier for that matter.

Thus, in a nutshell, we can conclude that the predictive performance of different standard classifiers compared by the mentioned studies arrived at different conclusions as to which classifier and sampling strategies performed better (Blagus & Lusa, 2010; Lemnaru et al., 2012; Mena & Gonzalez, 2006; Prati et al., 2014; Sarmanova & Albayrak, 2013).

In classification and predictive analytics, LR is normally considered a very informative classifier as it provides important information about the effect of an independent variable (IV) on the dependent variable (DV) through the odds ratio (Hosmer & Lemeshow, 2004). However, the presence of imbalance problem hinders the predictive "power" of LR (Wallace & Dahabreh, 2012). Blagus & Lusa (2010) performed a simulation study to evaluate the performance of six types of classifiers [ANN, Linear Discriminant Analysis (LDA), RF, SVM and penalized logistic regression (PLR)] on highly imbalanced data. However, their results showed that the PLR with ROS method, failed to remove the biasness towards the majority class.

A simulation study by Hamid et al. (2015) discovered that when sample size was large (at least 500) the parameter estimates accuracy for LR improved. In addition, the estimation of LR parameters was severely affected by types of covariates; either continuous, categorical, or count data. Simulation studies, usually, enables us to provide a more conclusive evidence on the effect of IR, as the simulated datasets were mold perfectly to cater specific problem types. In our previous study (Rahman & Yap, 2016), our results were consistent with the study by Hamid et al. (2015), which reported that the performance of LR was affected by sample size. However, Hamid et al. (2015) did not consider imbalanced data. Simulation studies are important to obtain empirical evidence on the impacts of IR on the estimate of logistic regression parameter, β-value and the odds ratio of the LR model.

Simulation Methods

This study considered a simple binary logistic regression (LR). In the LR model, two unknown parameters, β_0 and β_1, are estimated using the maximum likelihood method. Assuming observations to be independent, the likelihood function is given by the following Equation 1 (Hosmer & Lemeshow, 2004):

$$L(\beta_0,\beta_1)=\prod_{i=1}^{n}\pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i} \quad \textbf{(1)}$$

To estimate β_0 and β_1, the maximization of the likelihood function is required. Therefore, the maximization of the natural logarithm of the likelihood function is denoted by the following Equation 2:

$$\log\left[L(\beta_{0,}\beta_1)\right]=\sum_{i=1}^{n}\{y_i\log\left[\pi(x_i)\right]+(1-y_i)\log\left[1-\pi(x_i)\right]\} \quad \textbf{(2)}$$

By referring to the simple LR Equation 1 , the Equation 2 can also be expressed as Equation 3 (Hosmer & Lemeshow, 2004):

$$\log\left[L(\beta_{0,}\beta_1)\right]=\sum_{i=1}^{n}y_i(\beta_0+\beta_1x_i)-\sum_{i=1}^{n}\log\left[1+\exp(\beta_0+\beta_1x_i)\right] \quad \textbf{(3)}$$

By differentiating $\log\left[L(\beta_{0,}\beta_1)\right]$ with respect to β_0 and β_1 and setting the resulting Equation 4 to zero, we can obtain β that maximizes Equation 3.

$$\sum_{i=1}^{n}\left[y_i-\pi(x_i)\right]=0 \text{ and } \sum_{i=1}^{n}x_i\left[y_i-\pi(x_i)\right]=0 \quad \textbf{(4)}$$

The maximum likelihood estimates of β_0 and β_1, are denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$ and is obtained using Newton -Raphson method. The probability that the event occurs, $\pi(x_i)$ for case *i* is then obtained as follows:

$$\hat{\pi}(x_i)=\frac{e^{\hat{\beta}_0+\hat{\beta}_1x_i}}{1+e^{\hat{\beta}_0+\hat{\beta}_1x_i}} \quad \textbf{(5)}$$

In addition, $\hat{\pi}(x_i)$ is also known as fitted or predicted value and the sum of $\hat{\pi}(x_i)$ is equal to the sum of the observed values:

$$\sum_{i=1}^{n}y_i=\sum_{i=1}^{n}\hat{\pi}(x_i) \quad \textbf{(6)}$$

The final estimated simple logistic regression model is written as:

$$\log\left[\frac{\hat{\pi}(x_i)}{1-\hat{\pi}(x_i)}\right]=\hat{\beta}_0+\hat{\beta}_1x_i \quad \textbf{(7)}$$

We assessed the effect of various percentages of IR and sample size on estimation of the parameter coefficient, β for binary LR model with one categorical independent variable. The estimate, $\hat{\beta}_1$ were compared with the true β_1 value. The simulations were performed using R-Studio. The value of the regression coefficient (β_1) for the logistic model was set at 2.08 which gave a significant odds ratio (*OR*) of 8.004 for *X* ($OR = e^{2.08} = 8.004$). The R

code developed for this simulation is available at https://github.com/hezlin/simulationx1cat.git. It is also provided in the Appendix.

Odds-ratio provide important information of the effect of the covariate on the event (dependent variable). Given a binary Y(1=Died, 0=Survived) and a categorical covariate X(Hypertension-HPT) with two categories (1=Yes and 0=No), an odds-ratio of 1 will indicate both patients with or without HPT has equal chance of Y=1 (Died). Meanwhile, an odds-ratio greater than 1 will indicate that patients with HPT are more likely to die, and if odds-ratio is less than 1, patients with no HPT are more likely to die.

Eight imbalance ratios were considered for this simulation study: 1%, 2%, 5%, 10%, 20%, 30%, 40%, and 50%. Imbalance ratio (IR) is the percentage of occurrence of minority class between the two predictor classes. For example, in this simulation, if we generate a dataset N=100, if the IR = 1% means that 1 out of 100 has y=1 and the rest 99 out of 100 has y=0. The IR 5% or less represents high IR in the response variable. However, due to the complexity of generating the simulated dataset, especially for fixing definite percentages of IR, the simulation model requires β_0 values to be flexible for different IR ratio. Thus, the full LR model used for this study is denoted as the following:

$$\log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_{0k} + 2.08x_{ik} \qquad \textbf{(8)}$$

where β_{0k} is determined by the IR and is not fixed at one value.

The data for the covariate (*X*) considered in this study were generated using a binomial distribution, *Bin (n= sample size, p=0.5).* We considered sample size of 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, and 5000. This simulation study involved 10,000 replications. The simulation algorithm is as follows:

Step 1: Generate random data for the categorical covariate X*, for* sample size, *n* and imbalance ratio, *IR.*

Step 2: Set β at 2.08 and obtain $f(x) = \beta_{0k} + 2.08x_{1k}$, where k = 1, 2, … 10,000. β_{0k} is not fix to create a fix percentage of imbalance accordingly, whereby the confidence interval of β_{0k} is set within the range of (-2, 10).

Step 3: Fit binary logistic regression to the generated data in Step 2.

Step 4: Obtain the parameter estimate, $\hat{\beta}$.

Step 5: Repeat Steps 1-4 for 10,000 replications.

Step 6: Calculate the MSE where $MSE = \frac{\sum_{i=1}^{10000}(\beta - \hat{\beta})^2}{10{,}000}$

Repeat Steps 1 – 6 for different sample size and imbalanced ratio.

RESULTS AND DISCUSSION

Simulation Results

Table 1 presents the simulation results for the LR parameter estimates for various sample sizes and IR. %.). The effect of IR reduced when sample size increased. The results showed that the estimates for β_0 *and* β_1 were very far from the true parameter values for smaller sample size (n=100) and for IR 1%, 2%, 5%, 10%, 20% and 30%. The bias in estimate was clearly seen for IR 20% or less for n=500. Meanwhile, for n=1000, the bias was seen for IR 10% or less. However, the effect of IR was less for sample size more than 3000 and above was only affected by IR of 1% and 2%. Table 2 summarizes the findings.

Figure 1 presents the effect of sample size and IR on the parameter estimate values. It clearly shows the parameter estimates is biased for IR 30% and below for n=100. The figure also shows that for all sample sizes, the estimate is close to the true parameter values at IR=30% and above. In Figure 2, we focus on high IR, 1% to 10% and omitting 20% to 50% so that visualization of the effect is clearer. The figure shows threshold of effect of IR decreases as sample size increases. For example, estimates are biased for n=500 for IR 5% and below, while for n=1000, estimates are biased at IR 5% and below. When estimates are biased the MSE will be larger. Figure 3 illustrates the effect of IR and sample size through the MSE and Figure 4 further emphasizes results in Figure 3 by focusing on the IR of 1% to 10%, by omitting the 20% to 50% ratios. In Figure 3, the effect of imbalance is less (lower MSE) at IR=30%, similar to the illustration in Figure 1. Further focusing on highly imbalanced ratios, Figure 4 illustrates that the MSE values are the largest for small sample sizes (n=100 and n=500).

Figure 5 and 6 illustrates the effects of imbalanced using a clustered boxplot. As shown in Figure 5, the effect of imbalanced data is obvious for sample size n=500 (IR=1% and 2%) and n=1000 (IR=1%). In Figure 6, we omit the imbalanced ratio 1% and 2%, and now there are no huge spikes in the boxplots. These figures clearly showed the effect of IR for various sample sizes, whereby the patterns show that the effect of IR on the bias of parameter estimates depends on sample size. The estimates get closer to the true value when the sample size and IR increases. The dispersion (standard deviation) of $\hat{\beta}_1$ also improves as sample size and IR increases.

Hence, by referring to all the figures (Figure 1 to 6), it can be concluded that the effect of the imbalance problem on the categorical covariate's parameter estimation is most severe for smaller sample sizes ($n \leq 500$) and for highly imbalanced ratios ($IR \leq 5\%$). The severity of the imbalanced problem is identified by the difference between the parameter estimated values and the fixed true beta value ($\beta_1 = 2.\theta$), as well as larger value of MSE. MSE is a good indicator of the bias in parameter estimates of the model. A larger MSE will indicate estimates are biased.

Table 1

Parameter Estimates for Categorical Covariate (β = 2.08) for model with different n and IR

Size	IR	$\hat{\beta}_1$	C.I (lower)	C.I (upper)	β_0	$\hat{\beta}_0$
100	**1**	**13.8056**	13.5891	14.0222	-6.5778	-19.6347
	2	**14.1244**	13.9647	14.2841	-5.6621	-17.6550
	5	**10.3026**	10.1344	10.4707	-4.5099	-12.6521
	10	**6.0332**	5.8896	6.1768	-3.6729	-7.5887
	20	**2.8190**	2.7535	2.8845	-2.7598	-3.4784
	30	**2.2406**	2.2164	2.2648	-2.1098	-2.2586
	40	2.1501	2.1394	2.1607	-1.5605	-1.6111
	50	2.1370	2.1275	2.1464	-1.0423	-1.0704
500	**1**	**10.2400**	10.0788	10.4011	-6.1980	-14.2681
	2	**6.7500**	6.6006	6.8993	-5.4361	-10.0587
	5	**2.8327**	2.7680	2.8975	-4.4340	-5.1721
	10	**2.1722**	2.1581	2.1863	-3.6419	-3.7273
	20	2.1168	2.1107	2.1229	-2.7487	-2.7787
	30	2.0902	2.0854	2.0950	-2.1131	-2.1219
	40	2.0912	2.0870	2.0954	-1.5615	-1.5692
	50	2.0928	2.0887	2.0968	-1.0390	-1.0458
1000	**1**	**6.4870**	6.3443	6.6298	-6.1395	-10.5075
	2	**3.5189**	3.4285	3.6094	-5.4078	-6.8257
	5	**2.1924**	2.1767	2.2081	-4.4262	-4.5303
	10	2.1163	2.1099	2.1227	-3.6365	-3.6712
	20	2.0976	2.0933	2.1018	-2.7460	-2.7619
	30	2.0883	2.0850	2.0917	-2.1118	-2.1192
	40	2.0861	2.0832	2.0891	-1.5620	-1.5665
	50	2.0844	2.0816	2.0872	-1.0401	-1.0418
1500	**1**	**4.6278**	4.5124	4.7431	-6.1250	-8.6452
	2	**2.5863**	2.5351	2.6376	-5.3983	-5.8921
	5	2.1322	2.1246	2.1322	-4.4223	-4.4704
	10	2.1034	2.0984	2.1085	-3.6382	-3.6587
	20	2.0894	2.0860	2.0929	-2.7467	-2.7541
	30	2.0855	2.0827	2.0883	-2.1102	-2.1162
	40	2.0835	2.0810	2.0859	-1.5618	-1.5649
	50	2.0841	2.0818	2.0864	-1.0405	-1.0424

Table 1 *(Continued)*

Size	IR	$\hat{\beta}_1$	C.I (lower)	C.I (upper)	β_0	$\hat{\beta}_0$
2000	**1**	**3.5815**	3.4914	3.6717	-6.1199	-7.5977
	2	**2.3260**	2.2945	2.3574	-5.3962	-5.6309
	5	2.1204	2.1140	2.1269	-4.4212	-4.4580
	10	2.0970	2.0927	2.1013	-3.6369	-3.6524
	20	2.0864	2.0835	2.0894	-2.7453	-2.7514
	30	2.0836	2.0812	2.0860	-2.1130	-2.1151
	40	2.0837	2.0816	2.0858	-1.5612	-1.5645
	50	2.0822	2.0802	2.0842	-1.0402	-1.0411
2500	**1**	**3.0220**	2.9515	3.0924	-6.1132	-7.0367
	2	**2.1933**	2.1766	2.2100	-5.3953	-5.4988
	5	2.1088	2.1029	2.1146	-4.4198	-4.4469
	10	2.0933	2.0895	2.0972	-3.6371	-3.6486
	20	2.0842	2.0816	2.0869	-β2.7455	-2.7489
	30	2.0821	2.0800	2.0842	-2.1111	-2.1137
	40	2.0838	2.0819	2.0856	-1.5616	-1.5647
	50	2.0817	2.0799	2.0835	-1.0393	-1.0409
3000	**1**	**2.6591**	2.6051	2.7132	-6.1100	-6.6739
	2	**2.1619**	2.1493	2.1746	-5.3902	-5.4671
	5	2.1020	2.0968	2.1072	-4.4196	-4.4398
	10	2.0920	2.0885	2.0955	-3.6365	-3.6474
	20	2.0867	2.0843	2.0891	-2.7457	-2.7511
	30	2.0833	2.0814	2.0852	-2.1121	-2.1146
	40	2.0814	2.0797	2.0831	-1.5626	-1.5634
	50	2.0809	2.0793	2.0825	-1.0399	-1.0402
3500	**1**	**2.4233**	2.3839	2.4628	-6.1072	-6.4387
	2	2.1460	2.1361	2.1559	-5.3877	-5.4511
	5	2.1011	2.0963	2.1059	-4.4195	-4.4386
	10	2.0913	2.0881	2.0946	-3.6369	-3.6466
	20	2.0858	2.0835	2.0880	-2.7457	-2.7501
	30	2.0817	2.0799	2.0834	-2.1118	-2.1134
	40	2.0811	2.0796	2.0827	-1.5612	-1.5626
	50	2.0808	2.0793	2.0823	-1.0400	-1.0404

Table 1 *(Continued)*

Size	IR	$\hat{\beta}_1$	C.I (lower)	C.I (upper)	β_0	$\hat{\beta}_0$
4000	**1**	**2.3202**	2.2895	2.3508	-6.1038	-6.3349
	2	2.1316	2.1241	2.1391	-5.3910	-5.4365
	5	2.0982	2.0938	2.1026	-4.4184	-4.4357
	10	2.0874	2.0844	2.0905	-3.6363	-3.6427
	20	2.0850	2.0830	2.0871	-2.7445	-2.7494
	30	2.0823	2.0806	2.0839	-2.1116	-2.1138
	40	2.0810	2.0796	2.0825	-1.5623	-1.5631
	50	2.0801	2.0787	2.0815	-1.0402	-1.0399
4500	**1**	**2.2362**	2.2144	2.2579	-6.1051	-6.2509
	2	2.1231	2.1161	2.1301	-5.3879	-5.4281
	5	2.0964	2.0922	2.1006	-4.4192	-4.4340
	10	2.0878	2.0850	2.0906	-3.6362	-3.6431
	20	2.0824	2.0805	2.0843	-2.7454	-2.7474
	30	2.0825	2.0809	2.0841	-2.1112	-2.1139
	40	2.0812	2.0798	2.0826	-1.5619	-1.5630
	50	2.0804	2.0791	2.0817	-1.0400	-1.0405
5000	**1**	**2.2074**	2.1893	2.2254	-6.1037	-6.2219
	2	2.1176	2.1111	2.1242	-5.3905	-5.4227
	5	2.0928	2.0888	2.0968	-4.4193	-4.4305
	10	2.0874	2.0847	2.0901	-3.6372	-3.6429
	20	2.0841	2.0823	2.0860	-2.7450	-2.7487
	30	2.0825	2.0809	2.0841	-2.1112	-2.1139
	40	2.0817	2.0803	2.0832	-2.1118	-2.1133
	50	2.0804	2.0791	2.0817	-1.0400	-1.0402

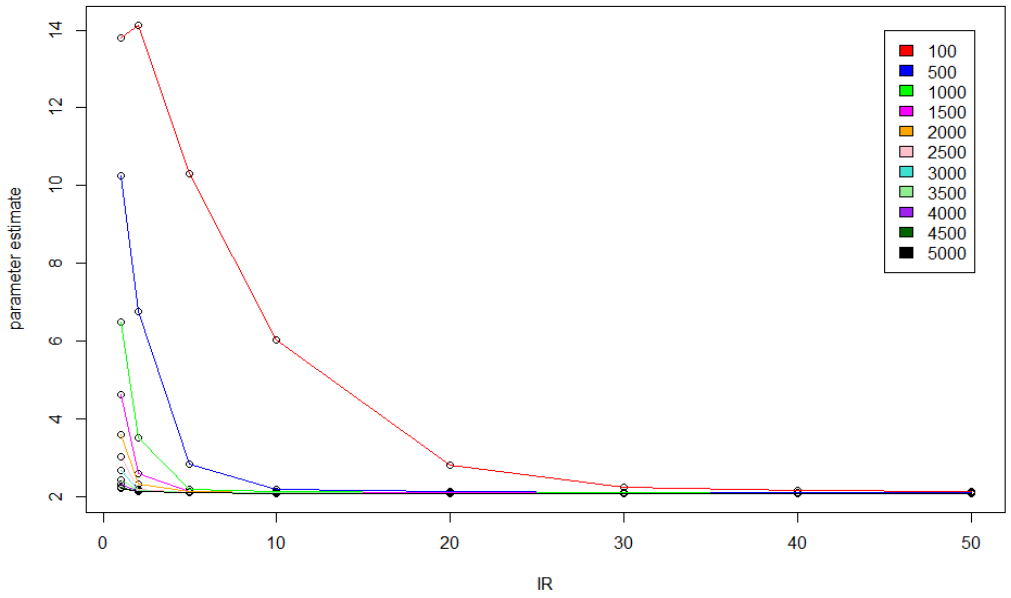


Figure 1. Categorical covariate's parameter estimates, $\hat{\beta}_1$, for different sample size and imbalanced ratio (Imbalance Ratio (IR): 1% to 50%).

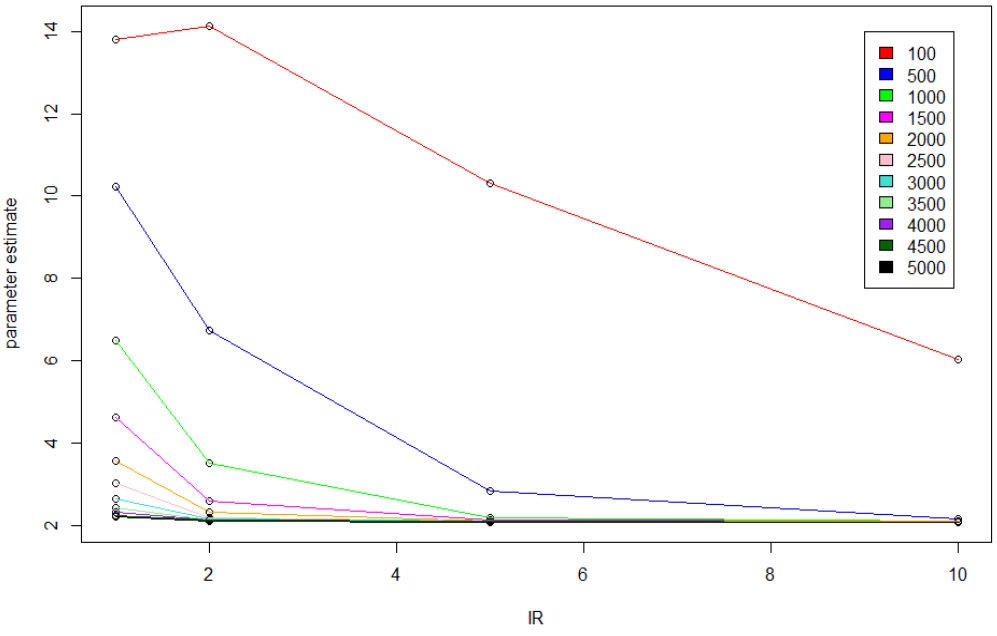


Figure 2. Categorical covariate's parameter estimates, $\hat{\beta}_1$, for different sample size and highly imbalanced ratio (IR : 1-10%).

Table 2

Summary of findings on the effect of IR and associated sample size

Sample Size	Estimates biased if IR is
100	30% and below
500	10% and below
1000 – 2000	5% and below
2500 – 3500	2% and below
4000 and above	1% and below

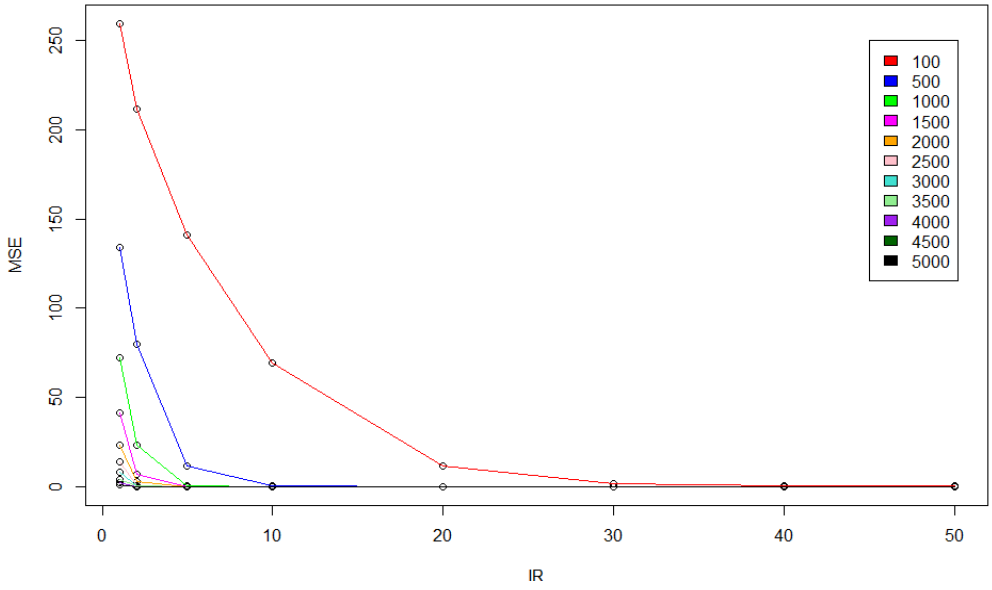


Figure 3. Mean square error (MSE) of categorical covariate's parameter estimates, $\hat{\beta}_1$, for different sample sizes and imbalanced ratio

From this simulation results, the effect of IR for small sample size is very prominent with large MSE. Even for larger sample size (n=1000 and 1500), the effect of imbalance towards the parameter estimation is still apparent. For small sample size, n=100, only at IR = 30% onwards the value of the estimates becomes closer to the actual parameter value. Sample size n=500, the estimates improve at IR = 10% onwards. For other sample sizes $1000 \leq n \leq 2000$, $2500 \leq n \leq 3500$ and $n \geq 4000$, the parameter estimation improves at IR = 5%, 2% and 1% onwards. The summary of these findings is shown in Table 2.

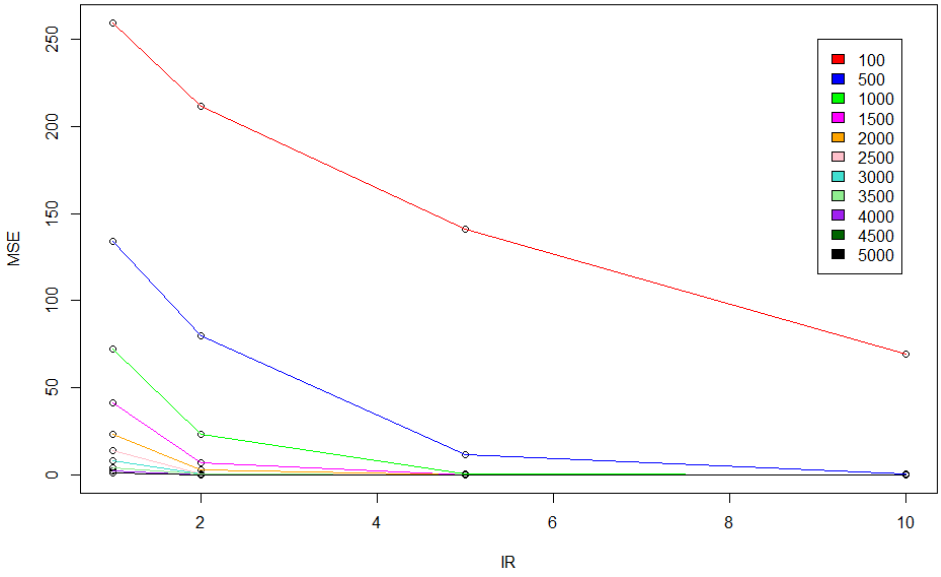


Figure 4. Mean square error (MSE) categorical covariate's parameter estimates, $\hat{\beta}_1$, for different sample size and highly imbalanced ratio (IR : 1-10%).

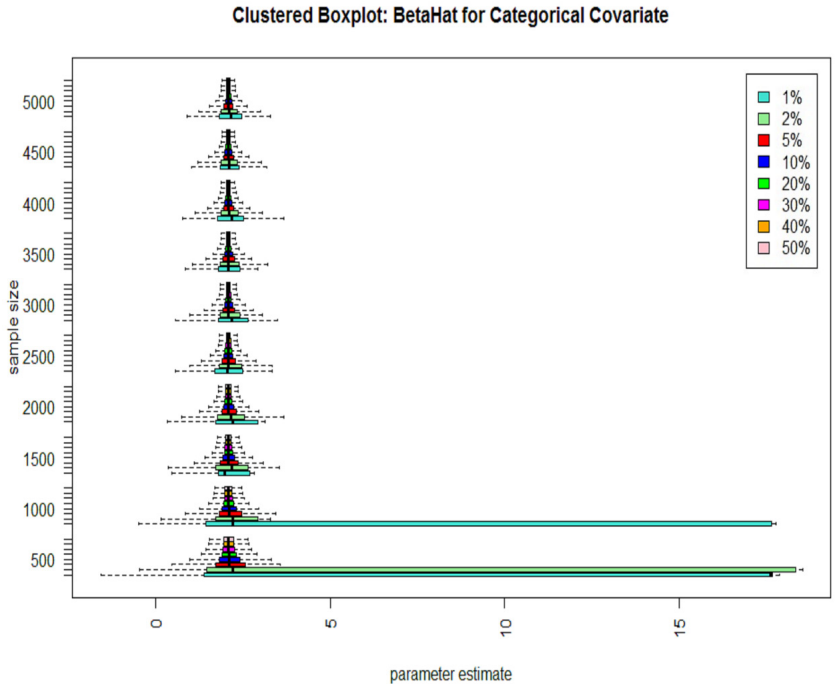


Figure 5. Clustered boxplots of $\hat{\beta}_1$ for a categorical covariate

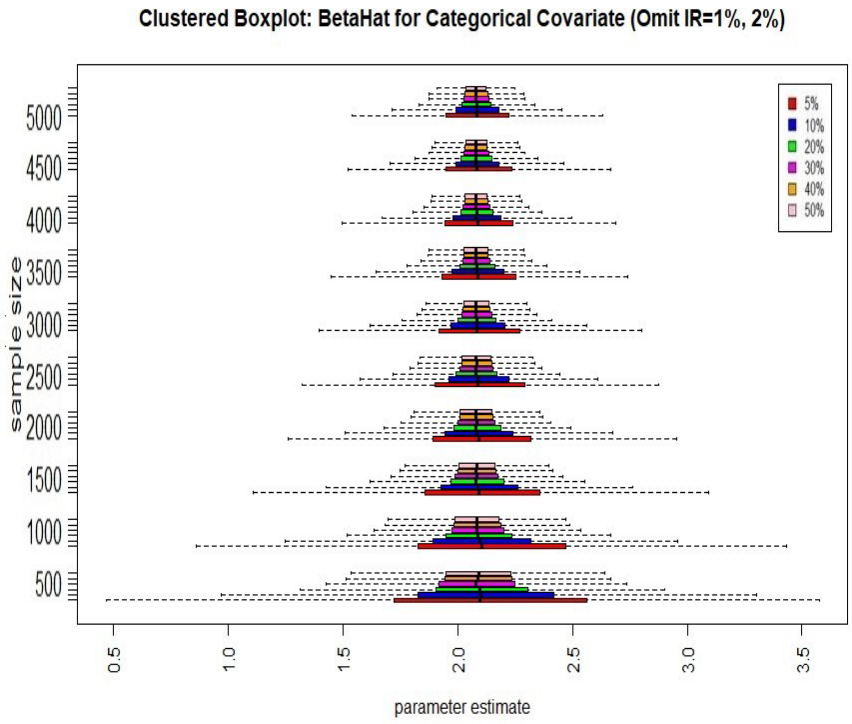


Figure 6. Clustered boxplots for $\hat{\beta}_1$ for a categorical covariate (omit IR=1%, 2%)

Table 3

Effect of imbalanced with application to real dataset (Diabetes Messidor)

Independent Variable	**Data/ IR**	$\hat{\beta}_0$ **, [p-value] C.I (lower, upper)**	$\hat{\beta}_1$ **, [p-value] C.I (lower, upper)**	**Odds-Ratio (OR) C.I (lower, upper)**
Retinal Abnormality (1 = yes, 0 = no)	**Original (540:611)**	0.6614, [0.002] (0.6613, 0.6614)	-0.5838, [0.010] (-0.5837, -0.5838)	0.5578 (0.5577, 0.5578)
	40% (407:611)	0.9477, [0.000] (0.9414, 0.9539)	-0.5872, [0.026] (-0.5938, -0.5805)	0.5591 (0.5554, 0.5626)
	30% (261:611)	1.4151,[0.000] (1.4033, 1.427)	--0.6110, [0.075] (-0.6236, -0.5983)	0.5537 (0.5470, 0.5604)

Table 3 *(Continued)*

Independent Variable	Data/ IR	$\hat{\beta}_0$, [p-value] C.I (lower, upper)	$\hat{\beta}_1$, [p-value] C.I (lower, upper)	Odds-Ratio (OR) C.I (lower, upper)
	20% (152:611)	1.9717,[0.000] (1.9527, 1.9906)	-0.6262, [0.178] (-0.6462, -0.6062)	0.5609 (0.5506, 0.5712)
	10% (68:611)	2.9467,[0.007] (2.8593, 3.0342)	-0.7962, [0.345] -0.8846, -0.7078)	0.5700 (0.5528, 0.5871)
	5% (35:611)	4.6431,[0.080] (4.3845, 4.9016)	-1.8264, [0.502] -2.0863, -1.5664)	0.5920 (0.5680, 0.6159)
	2% (13:611)	10.5126, [0.418] (10.0310, 10.9941)	-6.7028, [0.806] (-7.1881, -6.2176)	0.6347 (0.5902, 0.6792)
	1% (7:611)	13.9315, [0.631] (13.4669, 14.3960)	-9.5009, [0.856] (-9.9711, -9.0306))	0.6799 (0.6137, 0.7461)
AMFM status (1 = FM, 0 = AM)	**Original (540:611)**	0.1837, [0.011] (0.1836, 0.1837)	-0.1785, [0.153] (-0.1784, -0.1785)	0.8364 (0.8364, 0.8365)
	40% (407:611)	0.4669, [0.000] (0.4657, 0.4680)	-0.1787, [0.219] (-0.1820, -0.1754)	0.8375 (0.8348, 0.8403)
	30% (261:611)	0.9124, [0.000] (0.9104, 0.9143)	-0.1804, [0.307] (-0.1860, -0.1750)	0.8381 (0.8335, 0.8428)
	20% (152:611)	1.4515, [0.000] (1.4483, 1.4548)	-0.1721, [0.404] (-0.1811, -0.1630)	0.8510 (0.8432, 0.8587)
	10% (68:611)	2.2623, [0.000] (2.2570, 2.2677)	-0.1785, [0.451] (-0.1934, -0.1635)	0.8615 (0.8480, 0.8749)
	5% (35:611)	2.9268, [0.000] (2.9193, 2.9343)	-0.1601, [0.480] (-0.1816, -0.1387)	0.9061 (0.8850, 0.9272)
	2% (13:611)	3.9382, [0.000] (3.9240, 3.9525)	-0.0850, [0.500] (-0.1557, -0.0143)	79921.66 (-10534.34, 170377.67)
	1% (7:611)	4.5989, [2.0000e-03] (4.5452, 4.6526)	0.5738, [5.4100e-01] (0.3403, 0.8074)	1676700 (1185601, 2167798)

Application to Real Data Results

This section illustrates an application using a real medical dataset (Diabetes Messidor dataset) from the UCI repository which has 16 covariates and known as "The Diabetes Messidor" dataset (Antal & Hajdu, 2014), consists of 1151 observations. This dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not (DR status). All features represent either a detected lesion, a descriptive feature of a anatomical part or an image-level descriptor. The two categorical covariate selected for this illustration are the *retinal abnormality* and *AMFM status*. We modeled the binary dependent variable, *DR status* (1=with DR (53%) and 0=without DR (47%)). We used retinal abnormality (1 = yes, 0 = no) and AMFM status (0 = AM, 1 = FM) as the independent variable in Model 1 and Model 2 respectively. Using stratified sampling on the original dataset, we obtained the IR percentage as shown in Table 3.

Results in Table 3 show that the estimate $\hat{\beta}_1$ in Model 1 is affected for IR 5% and below. The p-values for $\hat{\beta}_1$ increases (leading to independent variable becoming insignificant) as imbalance becomes more severe thus leading to misleading results. Results of Model 2 shows the effect of imbalance on odds-ratio. The odds-ratios were extremely large at IR 1% and 2%. This application to real dataset confirmed the results of the simulation study, which strengthened the conclusion that imbalanced problem will be misleading on the effect of the independent variable on the response variable.

CONCLUSIONS

Imbalanced data have effect on the parameter estimates and classification performance of binary logistic regression model with a categorical covariate. The optimal IR for different sample size for less biased estimates was determined via a simulation study. It was concluded that all samples were affected by imbalanced even for larger sample sizes. The effect of imbalanced data on parameter estimates was reduced as sample size increased. The imbalanced ratio in the response variable would not only affect the parameter estimates, but the p-value and odds- ratio for the covariate as well. Hence, imbalanced data can lead to inaccurate findings. There are approaches recommended for handling imbalanced problem such as resampling strategies (ROS (Random Oversampling), RUS, (Random Undersampling) and SMOTE (Synthetic Minority Oversampling Technique). Future simulation studies can investigate which sampling techniques can improve the parameter estimates and predictive performance of the binary logistic regression when data is highly imbalanced.

ACKNOWLEDGEMENT

Our gratitude goes to Universiti Teknologi MARA for the indirect contribution of this publication. We also thank Prof. Dr. Haibo He (Rhodes Island University), Prof. Dr. Ronaldo Prati (Universidade Federal do ABC), Dr. Pam Davey and Dr. Carolle Birrell (University of Wollongong) for sharing their knowledge and providing valuable comments for this study. Not to forget Dr. Hamzah Abdul Hamid (Univeristi Malaysia Perlis) for his indirect contribution to the simulation studies.

REFERENCES

Ahmad, S., Midi, H., & Ramli, N. M. (2011). Diagnostics for residual outliers using deviance component in binary logistic regression. *World Applied Sciences Journal, 14*(8), 1125-1130.

Anand, A., Pugalenthi, G., Fogel, G. B., & Suganthan, P. N. (2010). An approach for classification of highly imbalanced data using weighting and undersampling. *Amino Acids*, *39*(5), 1385-1391.

Antal, B., & Hajdu, A. (2014). An ensemble-based system for automatic screening of diabetic retinopathy. *Knowledge-Based Systems*, *60*, 20-27.

Blagus, R., & Lusa, L. (2010). Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *11*(1), 1-17.

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, *36*(3), 4626-4636.

Chawla, N. V. (2003, August 21). C4. 5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure. In *Proceedings of the International Conference on Machine Learning, Workshop Learning from Imbalanced Data Set II* (pp. 1-8). Washington, DC.

Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Editorial : Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, *6*(1), 1-6.

Cohen, G., Hilario, M., Sax, H., Hugonnet, S., & Geissbuhler, A. (2006). Learning from imbalanced data in surveillance of nosocomial infection. *Artificial Intelligence in Medicine*, *37*(1), 7-18.

Dong, Y., Guo, H., Zhi, W., & Fan, M. (2014, October 13-15). Class imbalance oriented logistic regression. In *2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (pp. 187-192). Shanghai, China.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42*(4), 463-484.

Goel, G., Maguire, L., Li, Y., & McLoone, S. (2013). Evaluation of sampling methods for learning from imbalanced data. *Intelligent Computing Theories*, *7995*, 392-401.

Hamid, H. A. (2016). Effects of different type of covariates and sample size on parameter estimation for multinomial logistic regression model. *Jurnal Teknologi*, *78*(12-3), 155-161.

Hamid, H. A., Yap, B. W., Xie, X. J., & Rahman, H. A. A. (2015). Assessing the effects of different types of covariates for binary logistic regression. In *AIP Conference Proceedings 1643* (Vol. 425, pp. 425-430). College Park, Maryland: American Institute of Physics Inc.

Hamid, H. A., Yap, B. W., Xie, X. J., & Ong, S. H. (2018). Investigating the power of goodness-of-fit tests for multinomial logistic regression. *Communications in Statistics: Simulation and Computation*, *47*(4), 1039-1055.

He, H., & Garcia, E. E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263-1284.

Hosmer, D. W., & Lemeshow, S. (2004). *Applied logistic regression* (2nd Ed.). Hoboken, NJ: John Wiley & Sons.

Lemnaru, C., Potolea, R., Lenmaru, C., & Potolea, R. (2012). Imbalanced classification problems: Systematic study, issues and best practices. *Enterprise Information Systems: Lecture Notes in Business Information Processing*, *102*, 35-50.

Longadge, R., Dongre, S. S., & Malik, L. (2013a). Class imbalance problem in data mining: Review. *International Journal of Computer Science and Network*, *2*(1), 83-87.

Mena, L., & Gonzalez, J. A. (2006). Machine learning for imbalanced datasets: Application in medical diagnostic. In *Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2006)* (pp. 574-579). Florida, USA: AAAI Press.

Oztekin, A., Delen, D., & Kong, Z. J. (2009). Predicting the Graft Survival for Heart-Lung Transplantation Patients: An Integrated Data Mining Methodology. *International Journal of Medical Informatics*, *78*(12), e84-e96.

Pourahmad, S., Ayatollahi, S. M. T., & Taheri, S. M. (2011). Fuzzy logistic regression: A new possibilistic model and its application in clinical vague status. *Iranian Journal of Fuzzy Systems*, *8*(1), 1-17.

Prati, R. C., Batista, G. E. A. P. A., & Silva, D. F. (2014). Class imbalance revisited: A new experimental setup to assess the performance of treatment methods. *Knowledge and Information Systems*, *45*(1), 247-270.

Rahman, H. A. A., & Yap, B. W. (2016). Imbalance effects on classification using binary logistic regression. In *International Conference on Soft Computing in Data Science* (pp. 136-147). Singapore: Springer.

Rahman, H. A. A., Yap, B. W., Khairudin, Z., & Abdullah, N. N. (2012, September 10-12). Comparison of predictive models to predict survival of cardiac surgery patients. In *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)* (pp. 1-5). Langkawi, Malaysia.

Ramyachitra, D., & Manikandan, P. (2014). Imbalanced dataset classification and solutions: A review. *International Journal of Computing and Business Research (IJCBR)*, *5*(4), 1-29.

Rothstein, M. A. (2015). Ethical issues in big data health research: Currents in contemporary bioethics. *The Journal of Law, Medicine & Ethics*, *43*(2), 425-9. DOI: https://doi.org/10.1111/jlme.12258

Roumani, Y. F., May, J. H., Strum, D. P., & Vargas, L. G. (2013). Classifying highly imbalanced ICU data. *Health Care Management Science*, *16*(2), 119-128.

Sarmanova, A., & Albayrak, S. (2013, April 24-26). Alleviating class imbalance problem in data mining. In *2013 21st Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). Haspolat, Turkey.

Shariff, S. S. R., Rodzi, N. A. M., Rahman, K. A., Zahari, S. M., & Deni, S. M. (2016, August 16-18). Predicting the "graduate on time (GOT)" of PhD students using binary logistics regression model. In *AIP Conference Proceedings* (Vol. 1782, No. 1, p. 050015). Selangor, Malaysia.

Srinivasan, U., & Arunasalam, B. (2013). Leveraging big data analytics to reduce healthcare costs. *IT Professional*, *15*(6), 21-28.

Uyar, A., Bener, A., Ciray, H., & Bahceci, M. (2010). Handling the imbalance problem of IVF implantation prediction. *IAENG International Journal of Computer Science*, *37*(2), 164-170.

Van Hulse, J., Khoshgoftaar, T. M., & Napolitano, A. (2007). Experimental Perspectives on Learning from Imbalanced Data. In *Proceedings of the 24th International Conference on Machine Learning* (pp. 935-942). New York, USA: Association for Computing Machinery.

Wallace, B. C., & Dahabreh, I. J. (2012, December 10-13). Class probability estimates are unreliable for imbalanced data (and how to fix them). In *2012 IEEE 12th International Conference on Data Mining* (pp. 695-704). Brussels, Belgium.

Weiss, G. M., & Provost, F. (2003). Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, *19*, 315-354.

Yap, B. W., Rani, K. A., Rahman, H. A. A., Fong, S., Khairudin, Z., & Abdullah, N. N. (2014). An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)* (Vol. 285, pp. 13-23). Singapore: Springer.

APPENDIX

```
#fitting the model
set.seed(54321)
ndata <- 100
nrep <- 10000 #set the number of replications
start <- -10 #set initial value of bnot
end <- 10 #set end value for bnot
n <- 1 #set initial value for the loop
perc <- 40 #set the percentage of imbalance

#replication setup
beta0Hat<-rep(NA,nrep)
beta1Hat<-rep(NA,nrep)
betanot<-rep(NA,nrep)
betaone <- 2.08

while(n<=nrep)
{
 #set bnot value
 for(i in seq(start,end,0.001))
 {
  x <- rbinom(ndata,1,1/2)
  rx <-chartr("01", "AB", x)
  dummy(x)
  k <-dummy(x)
  linpred <- cbind(1,dummy(x)[,-1])%*% c(i,betaone) #(b)
  pi<-exp(linpred)/(1+exp(linpred))
  ru <- runif(ndata,0,1)
  u<-as.vector(ru)
  ry <- ifelse((u<=pi),1,0)
  m_y <- (mean(ry)*100)
  if(m_y == perc && n <=nrep)
  {
   dt <-data.frame(x=rx, y=ry) #fit the logistic model
   #print(dt)
   betanot[n]<-i
   mod <- glm(y~x, family="binomial", data=dt)
```

```
    beta0Hat[n]<-mod$coef[1]
    beta1Hat[n]<-mod$coef[2]
    n <- n + 1
   }
  }
}

Round1<-round(c(beta0=mean(beta0Hat),beta1Hat=mean(beta1Hat)),3)
mean(beta1Hat)
ci.b1 <- CI(beta1Hat,ci=0.95)
MSEbeta1Hat <- round(sum((beta1Hat-2.08)^2/nrep),3)
meanb0 <- mean(betanot)
mean(beta0Hat)
```

Pertanika J. Sci. & Technol. 28 (4): 1163 - 1177 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Classification of Familial Hypercholesterolaemia Using Ordinal Logistic Regression

Muhammad Hamizan Jamaludin[1], Yap Bee Wah[2]*, Hapizah Mohd Nawawi[3], Chua Yung-An[3], Marshima Mohd Rosli[1] and Muthukkaruppan Annamalai[1]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia
[2]Advanced Analytics Engineering Centre, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia
[3]Institute for Pathology, Laboratory and Forensic Medicine (I-PPerForM), University Teknologi MARA, 47000 UiTM, Sungai Buloh, Selangor, Malaysia

ABSTRACT

Familial hypercholesterolaemia (FH) is a genetic disease that causes the elevation of low-density lipoprotein cholesterol (LDL-C), which subsequently leads to premature coronary heart disease (CHD). Features which have been reported to be associated with FH include lipids level, tendon xanthomata, and history of CHD. The Ordinal Logistic Regression model using the classification of FH patients with the Dutch Lipid Clinic Network Criteria (DLCN) as the dependent variable (where 1=Possible, 2=Probable, 3=Definite) was developed and evaluated for different types of link functions. The FH patients (n = 449) were recruited from health screening programmes conducted in hospitals and clinics in Malaysia from 2010 to 2018. Results indicate there is a significant association between FH categories with demographic factors (ethnicity and smoking) and physical symptoms (corneal arcus and xanthomata). The Ordinal Logistic Regression using Cauchit link function has lower Akaike Information Criterion (AIC) value, higher Nagelkerke's R-Square and classification accuracy compared to Probit and Logit

ARTICLE INFO

Article history:
Received: 6 March 2020
Accepted: 12 June 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.03

E-mail addresses:
mizanjamaludin93@gmail.com (Muhammad Hamizan Jamaludin)
beewah@tmsk.uitm.edu.my (Yap Bee Wah)
hapizah.nawawi@gmail.com (Hapizah Mohd Nawawi)
yungan.chua@gmail.com (Chua Yung-An)
marshima@tmsk.uitm.edu.my (Marshima Mohd Rosli)
mk@tmsk.uitm.edu.my (Muthukkaruppan Annamalai)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

link function, diastolic blood pressure, corneal arcus and xanthomata were found to be significant covariates of FH.

Keywords: Classification, dutch lipid clinic network criteria, familial hypercholesterolaemia, ordinal logistic regression

INTRODUCTION

Familial hypercholesterolemia (FH) is an inherited dyslipidaemia that causes abnormal elevation of low-density lipoprotein cholesterol (LDL-C) since birth and leads to premature cardiovascular diseases as early as in the second decade of life (Mundal et al., 2018). It is an autosomal dominant genetic disorder that usually affects LDL receptor (*LDLR*), apolipoprotein B (*APOB*) and proprotein convertase subtilisin/kexin type 9 (*PCSK9*) genes. The genetic defects impair the liver capability in metabolising or removing excess LDL-C, resulting in very high plasma LDL-C levels which can lead to atherosclerotic cardiovascular diseases (ASCVD), such as coronary heart disease (CHD), strokes, and peripheral vascular diseases (Wong et al., 2016). With the global prevalence of 1:250 (Akioyamen et al., 2017), FH is not a rare disease and will significantly contribute to the global premature ASCVD-related mortality if not treated.

There is a need to improve the identification of FH in general population. Using a validated tool developed from data available in primary care records may provide a high potential solution to the problem of under diagnosing diagnostic criteria of FH (Weng et al., 2018). Clinicians should give priority to the individuals or patients with highest chances of suffering FH to be assessed clinically using widely known diagnostic toolssuch as Simon-Broome (SB) or Dutch Lipid Clinic Network Criteria (DLCN)(Varghese, 2014).

A good classification model will be useful to classify FH patients in Malaysia using some significant clinical and genetic factors related to FH. Several studies have used chi-square test of association to identify the factors (age, gender, ethnic and medical history) that are associated with FH (Abdul-Razak et al., 2017; Khoo et al., 2000). Lye et al. (2013) applied the logistic regression to determine the variations in *LDLR, APOB*, PCSK9 and other lipid-related genes inclinically diagnosed FH patients. Shi et al.(2014) also applied logistic regression and found work that needed physical skills and living in rural areas were significantly associated with FH. Most studies applied binary logistic regression model where the dependent variable is a dichotomous variable (FH or non-FH) (Norusis, 2005; Warner, 2008). However, the Dutch Lipid Clinic Network Criteria has four categories (Definite, Probable, Possible and Unlikely FH), while the Simon-Broome has three categories (Definite, Possible and Unlikely FH). All of the categories are sorted in ranks which perceives the Definite FH as the highest rank followed by “Probable FH” and the lowest rank is the Unlikely FH.

Multinomial Logistic Regression is useful when the dependent variable of study has more than two categories. However, when the categories of the dependent variable are ordinal in nature, the Ordinal Logistic Regression (OLR) model should be used because the model takes into consideration the ordinal nature of the data. The OLR model involves the proportional odds assumption or parallel lines assumption which allows a set of parallel line for each category of the dependent variable. If the test of parallel line is accepted, we can determine the effect of the covariate on the odds of moving to a higher-order category of the dependent variable. There are five different link functions that relates the expected value of the response to the linear predictors. The aim of this study was to identify the most suitable Ordinal logistic regression model using different link functions for classification of FH using DLCN for FH patients in Malaysia. The OLR models were evaluated based on AIC (Akaike Information Criterion), Nagelkerke's R-Square and classification accuracy.

MATERIALS AND METHODS

Data Source

Volunteers (n= 5171) were recruited from community health screening programmes and Specialist Lipid Clinics in Malaysia from 2010 to 2018. Familial hypercholesterolaemia was diagnosed using DLCN. Individuals with DLCN category of Possible, Probable or Definite FH were considered as FH patients. Irrespective of the DLCN categories, we excluded individuals with LDL-C<4.0 mmol/L. After excluding non-FH individuals, a total of 449 patients were recruited for this study.

Theoretical Framework

The demographic factos are age, gender, ethnicity, smoking status, and history of coronary artery disease (CHD). Other factors data consist of systolic blood pressure, diastolic blood pressure, total cholesterol (TC), triglyceride (TG), LDL-C, high-density lipoprotein (HDL), xanthomata, and corneal arcus. The theoretical framework proposed in this study is presented in Figure 1 where the dependent variable is FH Category (1=Possible, 2=Probable, 3=Definite).

Descriptive Analysis

Based on the frequency distribution of patients shown in Table 1, the distribution of gender is slightly higher for female (240 or 53.5% female , 209 or 46.5% male). About 78% were Malay (78%) more than half of the patients do not smoke (64.6%). About 78% did not suffer from hypertension and 72.4% (n = 325) were without diabetes while 80% did not

have CHD. About 31.8% (or 143) patients did not have signs of corneal arcus and 30.3% did not suffer from xanthomata. As shown in Figure 2, 65% (n = 290) were diagnosed as Possible, 11% (n = 48) were Probable, while 25% (111) were Definite FH.

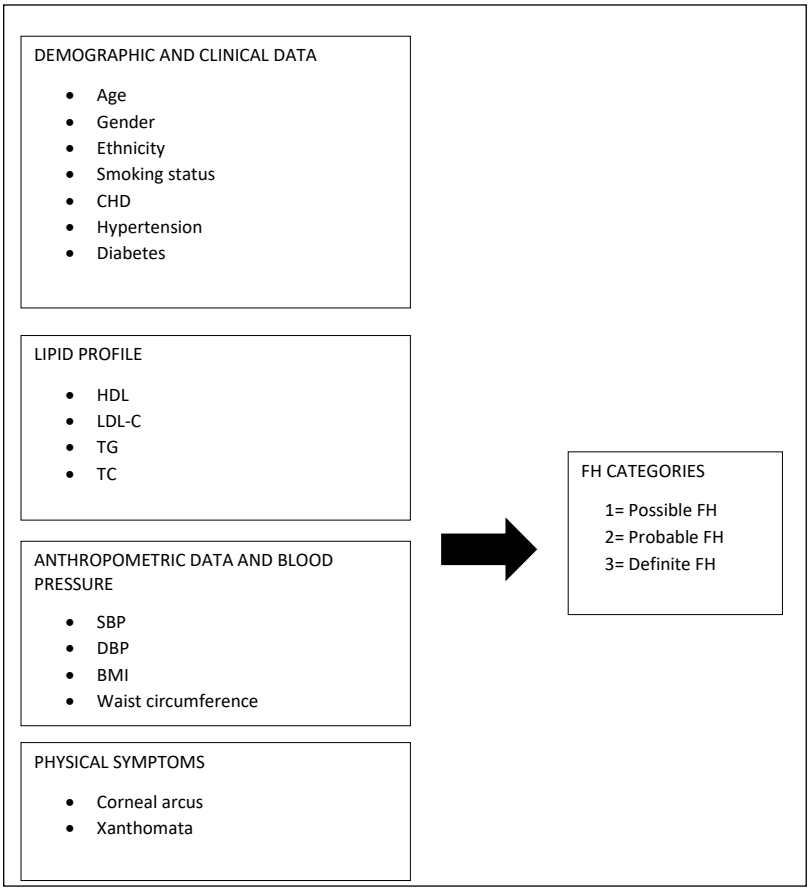


Figure 1. Theoretical framework

Notes. CHD = Coronary heart disease (CHD); HDL = High-density lipoprotein; LDL-C = Low-density lipoprotein cholesterol; TG = Triglyceride; TC = Total cholesterol; SBP = Systolic blood pressure; DBP = Diastolic blood pressure; BMI = Body mass index.

Table 1

Frequency distribution for demographic and clinical data, and physical symptoms variables

Type of Variable	Variable	Description of Variable	Frequency	Percentage (%)
Demographic and clinical data	Gender	Male	209	46.5
		Female	240	53.5

Table 1 *(Continued)*

Type of Variable	Variable	Description of Variable	Frequency	Percentage (%)
	Ethnicity	Malay	350	78.0
		Non-Malay	99	22.0
	Smoking	No	290	64.6
		Yes	144	32.1
	Hypertension	No	349	77.7
		Yes	77	17.1
	Diabetes	No	325	72.4
		Yes	25	5.6
	Coronary heartdisease (CHD)	No	401	89.3
		Yes	32	7.1
Physical symptoms	Corneal arcus (CA)	No	143	31.8
		Yes	84	18.7
	Xanthomata	No	136	30.3
		Yes	93	20.7

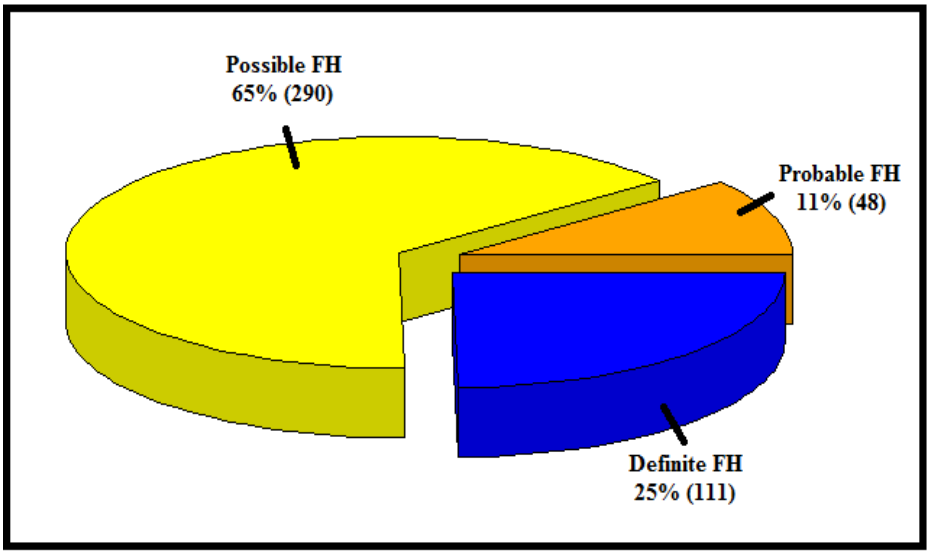


Figure 2. Distribution of FH category based on DLCN

Ordinal Logistic Regression

The ordinal logistic regression is appropriate to use when the dependent variable has ordinal categories. In this study, the dependent variable consists of three categories of FH. The ordinal categorical variables can have ordering from low to high or from high to low (Agresti, 2010). The model assumptions include no existence of multicollinearity and proportional odds to predict ordinal outcomes.

General Linear Model for Ordinal Data

The general linear model for ordinal data is based on the assumption that the dependent variableis in j[th] ordered categories according to link functions chosen (Hilbe, 2009; Norusis, 2005) (Equation 1):

$$link\left(\gamma_j\right) = \frac{\theta_j - \left[\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k\right]}{\exp\left(\tau_1 z_1 + \tau_2 z_2 + ... + \tau_m z_m\right)} \quad [1]$$

where γ_j is the cumulative probability for the j[th] category, θ_j is the threshold for the j[th] category, $\beta_1, \beta_2, ..., \beta_k$ are the regression coefficients, $x_1, x_2, ..., x_k$ are the predictor variables, k is the number of predictors, $\tau_1, \tau_2, ..., \tau_m$ are the coefficients for the scale component, $z_1, z_2, ..., z_m$ are m predictor variables chosen from the same set variables as the x's.

Build-in Link Functions

The ordinal regression model assumes that the slopes of several response levels are proportional (Hilbe, 2009). The different link functions for ordinal logistic regression is shown in Table 2. This study uses logit, probit and cauchit link only as complementary log-log link can only be applied if the chances of event is very small or very large (Allison, 1999). Negative log-log link can explain the data when the cumulative probability for lower group is high and it slowly approaches 1 (Norusis, 2005).

The cumulative probability of Casei for category *j* can be obtained as follows (Kutner et al., 2005) (Equation 2):

$$P\left(Y_i \le j\right) = \frac{\exp(\alpha_i + (\beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k)}{1 + \exp(\alpha_i + (\beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k)} \quad [2]$$

RESULTS

Results in Table 3 showed that a higher percentage of non-Malays were found to be Definite FH. Among patients with corneal arcus, 53.6% were Definite FH while 95.7%

Table 2

Link Function for Ordinal Logistic Regression model

No	Binomial Link	Function	Typical application
1	Logit	$\ln\left(\frac{\gamma}{1-\gamma}\right)$	Evenly distributed categories
2	Complementary log-log	$\ln(-\ln(1-\gamma))$	Higher categories more probable
3	Negative log-log	$-\ln(-\ln(\gamma))$	Lower categories more probable
4	Probit	$\Phi^{-1}(\gamma)$	Analyses with explicit normally distributed latent variable
5	Cauchit	$\tan(\pi(\gamma-0.5))$	Outcome with many extreme values

Note. γ is probability of event *Y*.

with xanthomata are Definite FH. The results of Chi Square and Fisher's Exact Test in Table 4 showed there was a significant association between FH category (using DLCN) with smoking, corneal arcus and xanthomata at 5% significance level. Ethicity is significant only at 10% significance level.

The ordinal regression model analysis was carried out to determine the best classification model and to identify the demographic, lipid, medical and physical symptoms factors that were significantly associated with the dependent variable (FH Category). The models were evaluated using Akaike Information Criterion (AIC), NagelkerkeR-Square and classification accuracy. Table 5 presents the value for AIC, Nagelkerke R-Square and accuracy value for OLR (Logit), OLR (Probit) and OLR (Cauchit). Results in Table 5 showed that OLR (Cauchit) was the best model as it has the lowest AIC, highest NagelkerkeR-Square and classification accuracy value compared to Probit and Logit link function.

Next, we performed data imputation for missing values using mean imputation and treating of extreme values using the "Winsorisation process". Winsorization set all outliers (extreme values) to a specified percentile of the data. We performed 90% winsorization where all data below the 5th percentile were set to the 5th percentile, and data above the 95th percentile were set to the 95th percentile. The results for the treated dataset were compared with the original one to check whether there were any insights that could be obtained from untreated and treated datasets. The model compasrison results are presented in Table 6.

The results in Table 6 showed the OLR model was significant ($p < 0.05$). Additionally, the model fitted the data with Pearson chi-squared of 0.510 (C1) and 0.912 (C2). C1

Table 3

Frequency distribution of patients based on demographic & clinical data, and physical symptoms variables with respect to FH category

Type of Variable	Variable	Description of Variable	DLCN Diagnosis, n(%)			Total (n)
			Possible FH	Probable FH	Definite FH	
Demographic and clinical data	Gender	Male	144(68.9)	21(10)	44(21.1)	209
		Female	146(60.8)	27(11.3)	67(27.9)	240
	Ethnicity	Malay	232(66.3)	40(11.4)	78(22.3)	350
		Non-Malay	58(58.6)	8(8.1)	33(33.3)	99
	Smoking	No	171(59)	37(12.8)	82(28.3)	290
		Yes	110(76.4)	10(6.9)	24(16.7)	144
	Hypertension	No	240(68.8)	34(9.7)	75(21.5)	349
		Yes	44(57.1)	10(13)	23(29.9)	77
	Diabetes	No	242(74.5)	25(7.7)	58(17.8)	325
		Yes	19(76)	4(16)	2(8)	25
	Coronary heartdisease (CHD)	No	259(64.6)	44(11)	98(24.4)	401
		Yes	16(50)	4(12.5)	12(37.5)	32
Physical Symptoms	Corneal arcus (CA)	No	72(50.3)	8(5.6)	63(44.1)	143
		Yes	4(4.8)	35(41.7)	45(53.6)	84
	Xanthomata	No	76(55.9)	39(28.7)	21(15.4)	136
		Yes	0(0)	4(4.3)	89(95.7)	93

Table 4

Chi-Square test of association between FH category and covariates

Test used	Variable	Test value	P-value (α=0.05)	Association
Chi-squared Test	Gender	3.405	0.182	Not Significant
	Ethnicity	5.327*	0.070	**Significant**
	Smoking	12.825***	0.002	**Significant**
	Hypertension	3.848	0.146	Not Significant
	Corneal arcus (CA)	70.203***	<0.001	**Significant**
	Xanthomata	143.510***	<0.001	**Significant**
Fisher's exact test	Diabetes	3.178	0.194	Not Significant
	Coronary heartdisease (CHD)	3.179	0.189	Not Significant

p<0.10: **p-value<0.05;*p-value<0.01*

Table 5

Model comparisons: Ordinal Logistic Regression with different link functions

FH Instrument	DLCN		
OLR (Link)	Nagelkerke R^2	AIC	Accuracy
OLR (Logit)	0.577	497.2866	84.54%
OLR (Probit)	0.753	504.5605	84.54%
OLR (Cauchit)	**0.797**	**465.0818**	**85.01%**

was Cauchit Model for untreated data while C2 was Cauchit Model for treated data. The assumption of parallelism can be accepted as test of the parallel lines is not significant (C1: p-value =0.603, C2: p-value=0.655). The Nagelkerke's pseudo R^2 is higher for model C1 (0.797) compared with C2 (0.401). The significant variables based on C1 model are Diastolic Blood Pressure (DBP), Corneal Arcus (CA) and Xanthomata for Cauchit model (C1). Meanwhile, the Cauchit model (C2) consists of 11 significant variable: Total Cholesterol (TC), Triglycerides (TG), Low Density Lipoprotein (LDL), Diastolic Blood Pressure (DBP), Body Mass Index (BMI), Age, Corneal Arcus (CA), Gender , Ethnicity, Hypertension and CHD.

Ordinal Logistic Regression using Cauchitlink function for untreated data (C1) has higher Nagelkerke R-Square (0.797) compared to the model for treated data C2 (0.401). The model classification accuracy for C1(85.01%) is also higher compared to C2 (73.05%). It

can be concluded that Cauchit model for untreated data (C1) performed better in classifying FH category. The final model with significant covariates is shown in Table 7 and some findings are as follows:

a) Patients who has high diastolic blood pressure is more likely to be classified into Definite FH [odds-ratio: exp(0.015) = 1.0151].

b) Patient who has no Corneal Arcus is less likely to be classified as Definite FH [odds ratio:exp(-8.648) = 0.0002]

c) Patients who do not suffer from xanthomatais less likely to be classified as Definite FH [odds ratio: exp(-13.063) = 0.000002].

Table 6

Ordinal logistic regression using Cauchit link function for untreated and treated data

Dataset	Untreated dataset	Treated dataset
Model	Cauchit Final Model (C1)	Cauchit Final Model (C2)
Model fitting information, χ^2	Chi-squared = 247.396 p-value <0.000	Chi-squared = 180.071 p-value <0.000
Pearson, χ^2	Chi-squared = 460.639 p-value = 0.510	Chi-squared = 822.933 p-value = 0.912
Deviance, χ^2	Chi-squared = 175.012 p-value = 0.378	Chi-squared = 598.353 p-value = 1.000
Test of parallel line, χ^2	Chi-squared = 3.639 p-value = 0.603	Chi-square = 56.178 p-value = 0.655
Cox and Snell	0.697	0.330
Nagelkerke	0.797	0.401
McFadden	0.574	0.231
Variable significant	Diastolic blood pressure (DBP) Corneal arcus (CA) Xanthomata	TC TG LDL-C DBP BMI Age CA

Table 6 *(Continued)*

Dataset	Untreated dataset	Treated dataset
Model	Cauchit Final Model (C1)	Cauchit Final Model (C2)
Variable significant	Diastolic blood pressure (DBP) Corneal arcus (CA) Xanthomata	CA Gender (Male) Ethnicity(Malay) Hypertension CHD
Total number of patients(n)	208	449

Notes. C1 (model for untreated data); C2 (model for treated data); TC= Total cholesterol; TG = Triglycerides; LDL-C = Low-density lipoprotein cholesterol; DBP = Diastolic blood pressure; BMI = Body mass index; CHD = Coronary heart disease.

Table 7

Parameter Estimates for Final OLR Model

Variables	Cauchit link function				
	Estimate	Standard Error	Wald	Degree of Freedom	P-Value
Constant1 (Possible)	-18.385	5.489	11.217	1	0.001
Constant2 (Probable)	-12.969	4.084	10.085	1	0.001
Diastolic blood pressure	0.015	0.029	0.263	1	0.008
Corneal arcus (No vs. Yes)	-8.648	2.648	10.669	1	0.001
Xanthomata (No vs. Yes)	-13.063	3.217	16.492	1	<0.001

*Note.*Since there are 3 categories there are 2 constant values (or intercepts).

DISCUSSION

This study seeked to classify FH patients in Malaysia using significant clinical and genetic factors related to FH. We used 449 patient data which were collected through health screening events, hospitals and clinics in Malaysia from 2010 to 2018. Results showed ethnicity smoking, and physical symptoms (corneal arcus and xanthomata) had significant association with FH. This finding show consistency with prior findings that indicated

xanthomata and corneal arcus were significant factors of FH (Khoo et al., 2000; Michaelides et al., 2004; Scicali et al., 2018). This indicates that patients who have physical symptoms such as corneal arcus and xanthomata more likely (have higher likelihood) to to be classified as FH patient. Smoking by itself is not the cause of FH, but may reduce HDL level and increase the risk of CHD (Gepner et al., 2011). Significant univariate association between smoking and FH and smoking was probably due to the increased scoring of DLCN among smokers due to history of CHD.

Next, ordinal logistic regression models were developed for the classification of FH patients using DLCN categories as the dependent variable. The ordinal logistic regression with Cauchit link function performed better (highest Nagelkerke R-square and smaller AIC) compared to models with logit and probit link functions. The ordinal logistic regression with Cauchit link function had an accuracy of 85%. The model was also evaluated by splitting data into training (70%, n_{train}=146) and testing sample (30%; $_{ntest}$=62). The accuracy was 88.3% and 83.9% for training and testing sample respectively. This shows consistency in the model classification accuracy. The significant factors were only diastolic blood pressure, corneal arcus and xanthomata. These findings indicate that patients who have high diastolic blood pressure are more likely to be classified into Definite FH, while patients who do not suffer from corneal arcus are less likely to be classified as Definite FH.

The overall findings of the study are that patients with associated clinical FH factors such as diastolic blood pressure, corneal arcus and xanthomata have higher chances to be classified as Definite FH. Therefore, there is a need to create health awareness in the community focusing on FH risk factors. Awareness of FH hereditary disease can be educated to the public in primary care or specialist lipid clinics. Upon clinical diagnosis, family cascade screening should be recommended to the FH patients. General practitioners or even Lipid Specialists should be sufficiently trained to deliver effective consultation and management to FH patients. FH patients should be advised to undergo therapeutic lifestyle changes, including adjusting their dietand involve in physical activities.Educated FH patients who are aware of the CHD risk tend to practise healthy lifestyle, such as cease smoking (Razali et al., 2019). The establishment of Familial Hypercholesterolaemia Support Group, a community-level voluntary organisation, may provide education to the FH patients and their family members on how to live with FH, and morally support the FH patients to practise healthy lifestyle. The support group is being implemented in Australia, Canada and some European Countries (Payne et al., 2018; Watts et al., 2012), but currently,it is still not practised in Malaysia.

Treatment of FH incurs large financial burden to the patient and nation. Direct medical costs may include preventive, diagnostic, and treatment services which is related to FH. Meanwhile, indirect costs relate to morbidity and mortality costs including productivity. Productivity measures include ‘absenteeism’ (costs due to employees being absent from

work for health-related reasons) and 'presentism' (decreased productivity of employees while at work) as well as premature mortality and disability. The developed OLR model highlighted the major observable clinical symptoms that significantly associated with FH in Malaysian population , thus allowing the clinicians to confidently screen the FH patients based on the symptoms, and treat the primary outcome of FH, which is CHD death, by advocating practise of healthy lifestyle.Thus, it is important if the trend or major groups who suffers this disease can be determined so that, they can benefit through the effortstaken by healthcare stakeholders to lower the risk of premature CHD-related death.An efficient statistical model allows clinicians to assess the risk (probability) of the patient in each FH category. Thus, patients with higher risk of being Possible FH should be advised to further clinical tests and healthcare treatments.

The limitation of this study is due to some numbers of missing data, an inherent problem suffered by secondary community data. This study found that the model fitted better with complete cases (i.e without data imputation or removing outliers). Future studies can also consider the new discrimination procedure proposed by Hamid and Hamid et al. (2018, 2018a) or treatement of outliers with new location model through the integration of Winsorization and smoothing approach (Hamid, 2018b). A larger sample with complete cases should be employed for future studies to confirm the findings of this study. Future studies will compare the models using different FH instruments for classification of FH.

CONCLUSION

This study successfully developed an OLR model for classification of Malaysian patients using a sample of secondary data. Ordinal logistic regression models with Cauchit link function for complete case was found to have higher Nagelkerke R-Square and classification accuracy. The final model results revealed that FH was found to be associated with diastolic blood pressure, corneal arcus and xanthomata. Ordinal logistic regression model is more appropriate when the dependent variable has ordinal categories. Taking into account the ordinal nature of the data, the probability obtained will be more accurate in the classification of patients into the three possible categories. The model is more informative than just modelling FH as a binary Yes or No category. It is imperative that effective early detection and timely control of FH must be strengthened in Malaysia to reduce disease burden. Future research can validate the statistical model by involving a larger sample of patients. Research should be ongoing to develop an efficient prediction model to assist clinicians in classification of FH patients. Early and timely detection is important for prevention of premature cardiovascular disease. Health education, intervention and preventive strategies of FH are highly important and more population-based disease screening for early detection and treatment should be carried out.

ACKNOWLEDGEMENT

This study was funded by UiTM MITRA Grants [(600-IRMI/MYRA 5/3/MITRA (003/2017)-1 and 600-IRMI/MYRA 5/3/MITRA (003/2017)-3].

REFERENCES

Abdul-Razak, S., Rahmat, R., Kasim, A. M., Rahman, T. A., Muid, S., Nasir, N. M., … & Ghani, R. A. (2017). Diagnostic performance of various familial hypercholesterolaemia diagnostic criteria compared to Dutch lipid clinic criteria in an Asian population. *BMC Cardiovascular Disorders*, *17*(1), 1-8.

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd Ed.).Hoboken, New Jersey; John Wiley & Sons.

Akioyamen, L. E., Genest, J., Shan, S. D., Reel, R. L., Albaum, J. M., Chu, A., & Tu, J. V. (2017). Estimating the prevalence of heterozygous familial hypercholesterolaemia: A systematic review and meta-analysis. *BMJ Open*, *7*(9), 1-13.

Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and application*. Cary, NC: SAS Institute. Inc. and John Wiley and Sons.

Gepner, A. D., Piper, M. E., Johnson, H. M., Fiore, M. C., Baker, T. B., & Stein, J. H. (2011). Effects of smoking and smoking cessation on lipids and lipoproteins: outcomes from a randomized clinical trial. *American Heart Journal*, *161*(1), 145-151.

Hamid, H. (2018a). New location model based on automatic trimming and smoothing approaches. *Journal of Computational and Theoretical Nanoscience, 15*(2), 493-499.

Hamid, H. (2018b). Winsorized and smoothed estimation of the location model in mixed variables discrimination. *Applied Mathematics and Information Sciences, 12*(1), 133-138.

Hamid, H., Ngu, P. A. H., & Alipiah, F. M. (2018). New smoothed location models integrated with PCA and two types of MCA for handling large number of mixed continuous and binary variables. *Pertanika Journal of Science and Technology, 26*(1), 247-260.

Hilbe, J. M. (2009). *Logistic regression models*. Boca Raton, Florida: Chapman and hall/CRC.

Khoo, K., Van Acker, P., Defesche, J., Tan, H., Van de Kerkhof, L., Heijnen-van Eijk, S., … & Deslypere, J. (2000). Low-density lipoprotein receptor gene mutations in a Southeast Asian population with familial hypercholesterolemia. *Clinical Genetics*, *58*(2), 98-105.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (Vol. 5). New York, NY: McGraw-Hill Irwin.

Lye, S. H., Chahil, J. K., Bagali, P., Alex, L., Vadivelu, J., Ahmad, W. A. W., … & Mohamed, R. (2013). Genetic polymorphisms in LDLR, APOB, PCSK9 and other lipid related genes associated with familial hypercholesterolemia in Malaysia. *PloS One*, 8(4), 1-8.

Michaelides, A. P., Fourlas, C. A., Pitsavos, C., Andrikopoulos, G. K., Skoumas, I., Kartalis, A., … & Stefanadis, C. I. (2004). Exercise testing in asymptomatic patients with heterozygous familial hypercholesterolaemia. *Coronary Artery Disease*, *15*(8), 461-465.

Mundal, L. J., Igland, J., Veierød, M. B., Holven, K. B., Ose, L., Selmer, R. M., … & Leren, T. P. (2018). Impact of age on excess risk of coronary heart disease in patients with familial hypercholesterolaemia. *Heart*, *104*(19), 1600-1607.

Norusis, M. J. (2005). Ordinal regression. In M. J. Norusis (Ed.), *SPSS 13.0 Advanced Statistical Procedures Companion* (pp. 69-89). New Jersey, USA: Prentice Hall.

Payne, J., Williams, S., Maxwell, D., Pariente, M. T., Olivares, R. A., ten Haaf, M. J., … & Wong-Rieger, H. (2018). Familial hypercholesterolaemia patient support groups and advocacy: A multinational perspective. *Atherosclerosis*, *277*, 377-382.

Razali, S., Ismail, Z., Chua, Y. A., & Nawawi, H. M. (2019). Quality of life and its contributing factors in patients with familial hypercholesterolaemia in Malaysia. *Environment-Behaviour Proceedings Journal*, *4*(10), 86-92.

Scicali, R., Di Pino, A., Platania, R., Purrazzo, G., Ferrara, V., Giannone, A., … & Farruggia, E. (2018). Detecting familial hypercholesterolemia by serum lipid profile screening in a hospital setting: Clinical, genetic and atherosclerotic burden profile. *Nutrition, Metabolism and Cardiovascular Diseases*, *28*(1), 35-43.

Shi, Z., Yuan, B., Zhao, D., Taylor, A. W., Lin, J., & Watts, G. F. (2014). Familial hypercholesterolemia in China: Prevalence and evidence of underdetection and undertreatment in a community population. *International Journal of Cardiology*, *174*(3), 834-836.

Varghese, M. J. (2014). Familial hypercholesterolemia: A review. *Annals of Pediatric Cardiology*, *7*(2), 107-117.

Warner, P. (2008). Ordinal logistic regression. *Journal of Family Planning and Reproductive Health Care, 34*(3), 169-170.

Watts, G. F., Sullivan, D. R., van Bockxmeer, F. M., Poplawski, N., Hamilton-Craig, I., Clifton, P. M., … & Burnett, J. R. (2012). A model of care for familial hypercholesterolaemia: Key role for clinical biochemistry. *The Clinical Biochemist Reviews*, *33*(1), 25-31.

Weng, S., Kai, J., Tranter, J., Leonardi-Bee, J., & Qureshi, N. (2018). Improving identification and management of familial hypercholesterolaemia in primary care: Pre-and post-intervention study. *Atherosclerosis*, *274*, 54-60.

Wong, B., Kruse, G., Kutikova, L., Ray, K. K., Mata, P., & Bruckert, E. (2016). Cardiovascular disease risk associated with familial hypercholesterolemia: A systematic review of the literature. *Clinical Therapeutics*, *38*(7), 1696-1709.

Pertanika J. Sci. & Technol. 28 (4): 1179 - 1202 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Ridge Regression as Efficient Model Selection and Forecasting of Fish Drying Using V-Groove Hybrid Solar Drier

Hui Yin Lim[1], Pei Shan Fam[1]*, Anam Javaid[1,2] and Majid Khan Majahar Ali[1]

[1]School of Mathematical Sciences, Universiti Sains Malaysia 11800 USM, Penang, Malaysia
[2]Department of Statistics, The Women University, Multan, Pakistan

ABSTRACT

Application of the Internet of things (IoT) for data collection in solar drying can be very efficient in collecting big data of drying parameters. There are many variables involved so it is hard to find a model to predict the moisture content of the food product during drying. In model building, interaction terms should be incorporated because they also contribute to the model. Eight selection criteria (8SC) is a very useful method in model building. This study applied ordinary least squares (OLS) regression and ridge regression with 8SC in model building to predict the moisture content of drying fish. A total of eighty models were considered in this study. One best model was chosen each from OLS regression and ridge regression. M78.7.3 with a total of eleven independent variables was the best OLS model after conducting multicollinearity and coefficient test. Next, the best ridge model M56.0.0 was obtained after the coefficient test. The mean absolute percentage error (MAPE) was used to measure the accuracy of the prediction model. For OLS model M78.7.3, the MAPE value was 15.7342. The MAPE value for ridge model M56.0.0 was 17.4054. From the MAPE value, OLS model M78.7.3 provided a better estimation than the ridge model M56.0.0. However, OLS model M78.7.3 violated the normality assumptions of residuals. This is highly caused by the outlier problem. So, due to non- normality of the residuals and presence of outliers in the dataset, ridge regression is preferred for the best forecast model.

Keywords: Eight selection criteria, IoT, model selection, ordinary least squares, ridge regression

ARTICLE INFO

Article history:
Received: 27 March 2020
Accepted: 9 June 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.04

E-mail addresses:
limhuiyincorrine@student.usm.my (Hui Yin Lim)
fpeishan@usm.my (Pei Shan Fam)
anamjavaid7860@yahoo.com (Anam Javaid)
majidkhanmajaharali@usm.my (Majid Khan Majahar Ali)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

The global food demand grows rapidly due to the increase of world population (Bodirsky et al., 2015). Hence, the rising of the food demand brings to food insecurity issues. Food security is defined as "all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for a healthy and active life" (FAO, 1996). Therefore, to deal with food insecurity, substantial improvements in food processing are required to satisfy the increased food demand.

Drying is one of the food post-processing techniques, which plays a vital role in the preservation of agriculture crops and marine harvest (Silva et al., 2017 and Ali et al., 2017b). It reduces the moisture content of food to inhibit the growth of microorganisms. The advantages of drying include longer shelf life, smaller size for storage purpose and lighter weight for transportation (Ertekin & Yaldiz, 2004). Traditional drying involves the process of drying agriculture crops or marine harvest under the direct sun exposition (Tiwari, 2016).

However, dehydrated food products will be contaminated easily due to the exposure of direct sunlight in open space. Besides, non-uniform sun-drying under open space increases the chance of fungal attack and the growth of microorganisms (Tiwari, 2016). Open sun drying also cannot control the drying parameter due to weather uncertainties. Furthermore, this conventional drying method is very time-consuming. The conventional method of fish drying that is still being used is shown in Figure 1.

Therefore, the effort to improve sun drying has led to the usage of renewable energy, specifically solar drying. For instance, Ali et al. (2017a), Stiling et al. (2012), Hossain and Bala (2007), Alfiya et al. (2018) and many other researchers applied solar drying by

Figure 1. Traditional method of fish drying under direct sunlight

using solar drier in their study. Furthermore, the Internet of things (IoT) based solar drying system using v-Groove Hybrid Solar Drier (v-GHSD) by Ali et al. (2017a, 2017b) was more effective in monitoring the drying behavior.

Since the development of drier, especially v-GHSD provides more benefits in terms of quality and hygienic aspects, all the important factors involved with the solar drying system should be investigated.

Drying parameters play an important role in the drying process. Tiwari (2016) stated that temperature, air humidity, area of exposed surface and pressure had effects on the removal of the moisture content. Besides, Silva et al. (2017) found out air temperature was a very important factor that would affect the drying process. Furthermore, Krokida et al. (2003) found out drying temperature had more influence than the air velocity and air humidity during the drying process. Hence, all of these drying parameters may contribute to the fish drying process. However, there is a very limited research study on the effect of important drying parameters and its interaction terms for fish drying using solar drier towards the fish drying model.

Furthermore, Javaid et al. (2020) found that there were significant interactions among variables in the drying seaweed process. Hence, regression analysis is one of the existing methods to investigate the relationship between variables in a data set and a continuous response variable with the interaction terms.

Ordinary Least Squares (OLS) is one of the popular estimation methods for the linear regression model. OLS regression estimates the functional relationship by minimizing the sum of squares differences between the observed and predicted response variable. It produces unbiased estimates with the smallest standard errors and provides the best linear unbiased estimator (BLUE) if all the model assumptions are satisfied (Wen et al., 2013). However, real data always suffer from multicollinearity. The application of least squares method in parameter estimation in the presence of multicollinearity may cause the estimates becoming unstable (Mahajan et al., 1977).

Apart from multicollinearity, the outlier is also one of the problems in regression analysis. Rajarathinam and Vinoth (2014) stated that outliers were commonly present in agriculture production data due to uncontrolled factors. Outliers will inflate the error variance as well as the standard errors. OLS estimator is extremely sensitive to outliers in linear regression analysis. However, agriculture and marine production data always suffer from multicollinearity and outlier problems. Hence, a suitable method should be done to solve these problems in the fish drying data. The initial moisture content of fish is between eighty-two percent, and the moisture content needs to be reduced to thirty-five percent after drying in the solar drier to achieve Equilibrium Moisture Content (EMC).

To overcome the limitations of the OLS estimator, researchers implemented a few methods. Regularization is one of the most common approaches to solve multicollinearity.

Regularization methods can be applied to control the instability of OLS estimates. Ridge regression is one of the regularization methods that shrinks the coefficients towards zero by minimizing the mean square error of the estimates (Ullah et al., 2018).

Furthermore, Steece (1986) concluded that ridge estimation was able to curb outliers in regressor space by downweighting their influence. Besides, Chatterjee and Hadi (2015) also stated that ridge estimators were stable as they were not affected by slight variations in the estimation data. Hence, ridge regression provides estimates that are more robust as compared to least squares estimates for small perturbations in the data.

Many researchers such as Delaney and Chatterjee (1986), Golub et al. (1979) and Kennard (1971) studied on estimation of biasing parameter in the ridge regression. There are also many proposed methods in selecting the biasing parameter but it does not have a general agreement on the best way to choose an optimal value of the biasing parameter (Khalaf, 2012). Besides, Zhang and Ibrahim (2005) stated that it was uncertain if ridge regression provided better estimates than OLS regression during different applications. Therefore, a more thorough approach is using the *lmridge* package in R developed by Ullah et al. (2018) to estimate the biasing parameter because ridge regression is a multiple regression with no penalty. Ullah et al. (2018) stated that the *lmridge* package in R provided suitable tools for ridge regression analysis in R as compared to other packages.

During model building, most of the researchers in the agriculture field only consider the individual term without considering the interaction term between the variables. For example, Jamal and Rind (2007) did not include interaction terms in developing the forecast models for acreage and production of the wheat crop in their study. However, interaction terms should be included during model building to avoid bias. Therefore, Javaid et al. (2019a) also addressed the interaction terms in their regression model to examine the main factors with their interaction terms affecting the collector efficiency, and they found that the interaction terms had a significant effect in the best final model.

Eight selection criteria (8SC) are always used for model selection purpose. For instance, in the study of Abdullah et al. (2015), they found that the application of multiple regression with 8SC was able to model and forecast biomass and biofuel production. Besides, Abdullah et al. (2011) used the polynomial regression technique with 8SC to find out the best model to estimate the volumetric stem biomass. Javaid et al. (2019b) applied multiple regression with 8SC in their study on forecasting the moisture ratio removal during the seaweed drying process. Yahaya et al. (2012) selected the best model in estimating the electrical conductivity levels by using 8SC.

Fish drying data were fitted to the thin layer drying model by many researchers. For example, Guan et al. (2013) applied nine thin layer models and found out the Page model was able to predict and describe the drying process more accurately. Kituu et al. (2010) also applied a thin layer model in drying fish. However, the thin layer model is used to

understand the drying behavior and does not involve model building. Besides, the thin layer drying model does not incorporate the interaction term in the drying model.

Furthermore, there is limited research conducted on the moisture content of drying fish and the factors affecting it with its interaction terms by using ridge regression with 8SC. Besides, in different applications, the performance of OLS regression and ridge regression may vary. Hence, OLS regression and ridge regression were conducted in this study. From all possible models, 8SC was applied for the model selection purpose to choose the best model to forecast the moisture content of drying fish.

MATERIALS AND METHODS

v-GHSD

The v-GHSD used in the fish drying process in this study consists of fans that back powered by solar panels. Besides, it also consists of a drying chamber, solar collector, v-aluminum roof, solar panel, and sensors using IoT for data collection every thirty minutes. The sensors are placed to measure the inlet and outlet temperature, inlet and outlet humidity, wind speed, and solar radiation. For this study, we looked at the effect of some factors and their interaction. Figure 2 shows the v-GHSD used in this study. Figure 3 shows the Chemical Fluid Dynamic (CFD) analysis using original data collected by using IoT and the parameters involved in this study.

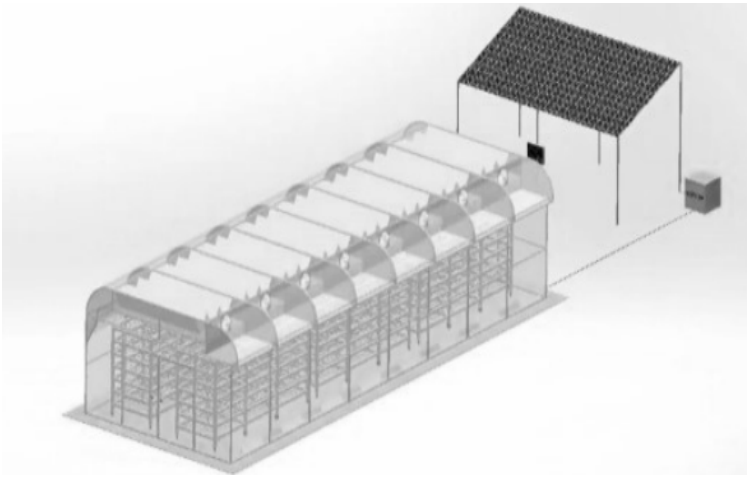


Figure 2. Simulation diagram of v-GHSD

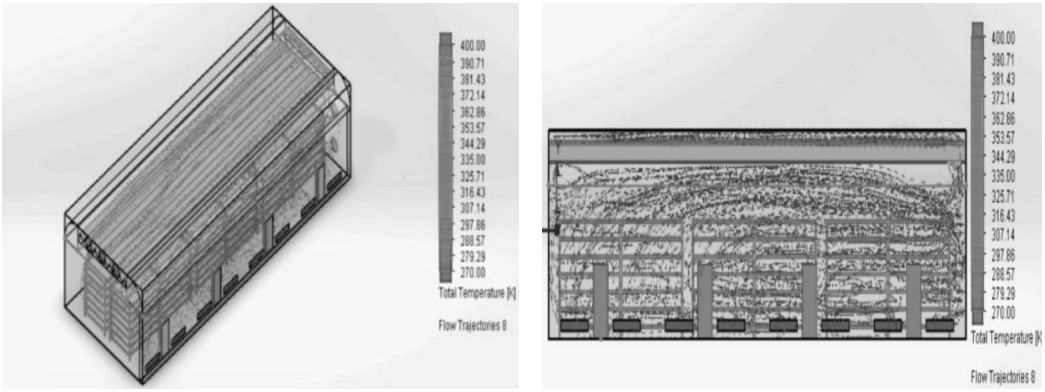


Figure 3. CFD Simulation diagram of v-GHSD

Model Development

Consider a multiple regression model (Equation 1),

$$y = X\beta + \varepsilon, \quad (1)$$

where y is a n x 1 vector of response variables, X is known as the design matrix of order n x p, β is a p x 1 vector of unknown parameters and ε is a n x 1 vector of identically and independent distributed errors.

According to Gujarati (2004), the OLS estimator of β is obtained as in Equation 2

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (2)$$

In Equation 2, if the regressors are nearly dependent, matrix $X'X$ becomes ill conditioned. Hence, Hoerl and Kennard (1970) suggested ridge estimator as in Equation 3,

$$\hat{\beta}^{ridge} = (X'X + \lambda I)^{-1}X'y, \quad (3)$$

where λ is a ridge parameter and I is an identity matrix. The ridge parameter, $\lambda > 0$ indicates the degree of shrinkage. Note that a value $\lambda = 0$ gives rise to OLS estimates.

Golub et al. (1979) proposed generalized cross-validation (GCV) as a method for choosing the ridge parameter (Equation 4).

$$GCV = \frac{SS_e}{\left(n - tr(H)\right)^2} \quad (4)$$

where SS_e refers to the residual sum of squares of a model using the ridge coefficients and H refers to an augmented hat matrix (Equation 5),

$$H = X(X'X + \lambda I)^{-1}X'. \quad (5)$$

We look for λ value that minimizes Equation 4. The ridge regression is carried out if the λ obtained is greater than zero for minimum GCV. If λ obtained is equal to zero, then ridge regression will be automatically equal to the OLS regression analysis. The *lmridge* package in R software was used in this study.

Phase 1– All Possible Models

Phase 1 involves computations of all possible models for the best model selection. According to Ali et al. (2017a), the formulae to compute the total number of all possible models are shown in Equation 6:

$$N = \sum_{j=1}^{k} j \left(k_{C_j} \right) \quad (6)$$

where N indicates the number of possible models, k indicates the total number of independent variables and j is 1, 2, ..., k. C shows the combinations for all possible models.

By using Equation 6, all possible models are computed.

Phase 2- Selected Models

Multicollinearity is checked among the variables by obtaining the correlation matrix for all factors. Only one highly correlated variable is removed from the analysis at a time. This procedure is performed until there is no collinear variable left in the model. However, for ridge regression, there is no need to check the problem of multicollinearity as it has the ability to deal with this problem.

Once the multicollinearity is checked among the variables in all possible models, a coefficient test is conducted for the OLS regression model after the model is free from the multicollinearity issue. For the ridge regression model, the coefficient test is conducted directly without checking the multicollinearity. The coefficient test is conducted in this phase to check the significance of the individual regression coefficient, β_j at the 5% level of significance. Adding an unimportant variable may make the model worse. The hypothesis statement of the coefficient test is shown as below:

$$H_0: \beta_j = 0,$$

$$H_1: \beta_j \neq 0.$$

where β_j is the coefficient of variable in the model for $j = 1,2, ..., k$. The test statistics of this test is (Equation 7)

$$t_0 = \frac{\hat{\beta}_j}{s(e\hat{\beta}_j)} \quad (7)$$

where $\hat{\beta}_j$ is the estimated regression coefficient of β_j and $s(e\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$.

Note that the null hypothesis is rejected if $|t_0| > t_{\frac{\alpha}{2},n-k-1}$. If the null hypothesis is rejected, then the selected parameter will be eliminated from the regression model. The selected model will be renamed as shown in Figure 4, where M denotes the model.

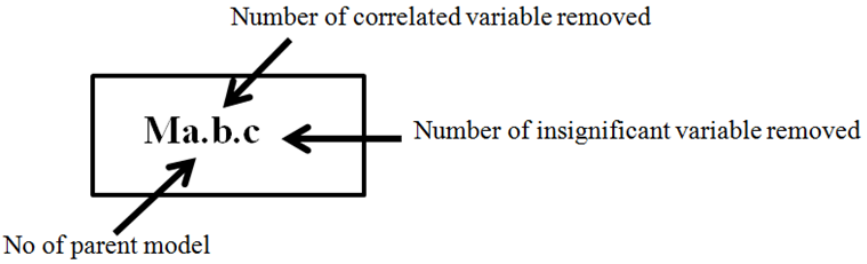


Figure 4. Model labeling in regression model

Phase 3 - The Best Model

Next, the selection of the best model from every selected model is conducted by using 8SC. According to Ali et al. (2017a, 2017b), the 8SC includes Akaike information criterion (AIC), RICE, Final prediction error (FPE), SCHWARZ, generalized cross-validation (GCV), sigma square (SGMASQ), Hannan-Quinn information criterion (HQ) and SHIBATA. The formulae of all of the model selection criteria are listed in Table 1. The most efficient model is selected based on the most number of the minimum value of the selection criteria.

Where SSE indicates the sum of squares error, $k+1$ indicates the number of estimated parameters and n indicates the sample size. According to Hajijubok and Gopal (2008), the condition that needs to be fulfilled when doing evaluation by using these model selection criteria is $2(k+1) < n$.

Phase 4 - Goodness of Fit

Five percent of the dataset reserved previously was used as test data to fit into the final best model chosen from phase 3. Then, residual analysis was conducted. The residual analysis is very important to check the randomness and normality of the residuals. In this study, a run test was used to check the randomness of the residuals, while the Kolmogorov-Smirnov test is used to check the normality assumption of the residuals. However, if the best model obtained from phase 3 is ridge regression model, then the normality of the residuals is not required because ridge regression does not require the residuals normality assumptions. Scatter plot and box plot of the residuals are used as supporting evidence of the goodness of fit test. Besides, the mean absolute percentage error (MAPE) is calculated as a measure of prediction accuracy (Ali et al., 2017a). The smaller the MAPE value the better, the higher the prediction accuracy. The formula of MAPE is shown in Equation 8:

$$MAPE = \frac{100}{N}\left(\frac{\sum_{i=1}^{j}|A_i - E_i|}{A_i}\right) \quad \text{for } i = 1,2,\ldots,j \tag{8}$$

where
A = Actual value of dependent variable (y)
E = Expected value $(\hat{y})$
N = Number of reserved data.

Table 1
Formula used for 8SC

AIC: $\left(\frac{SSE}{n}\right)(e)^{2(k+1)/n}$ Akaike (1969)	RICE: $\left(\frac{SSE}{n}\right)\left[1-\left(\frac{2(k+1)}{n}\right)\right]^{-1}$ (Rice, 1984)
FPE: $\left(\frac{SSE}{n}\right)\frac{n+(k+1)}{n-(k+1)}$ (Akaike, 1974)	SCHWARZ: $\left(\frac{SSE}{n}\right)n^{(k+1)/n}$ (Schwarz, 1978)
GCV: $\left(\frac{SSE}{n}\right)\left[1-\left(\frac{k+1}{n}\right)\right]^{-2}$ (Golub et al., 1979)	SGMASQ: $\left(\frac{SSE}{n}\right)\left[1-\left(\frac{k+1}{n}\right)\right]^{-1}$ (Ramanatam, 2002)
HQ: $\left(\frac{SSE}{n}\right)(\ln n)^{2(k+1)/n}$ (Hannan & Quinn, 1979)	SHIBATA: $\left(\frac{SSE}{n}\right)\frac{n+2(k+1)}{n}$ (Shibata, 1981)

All the four phases are summarized in Figure 5.

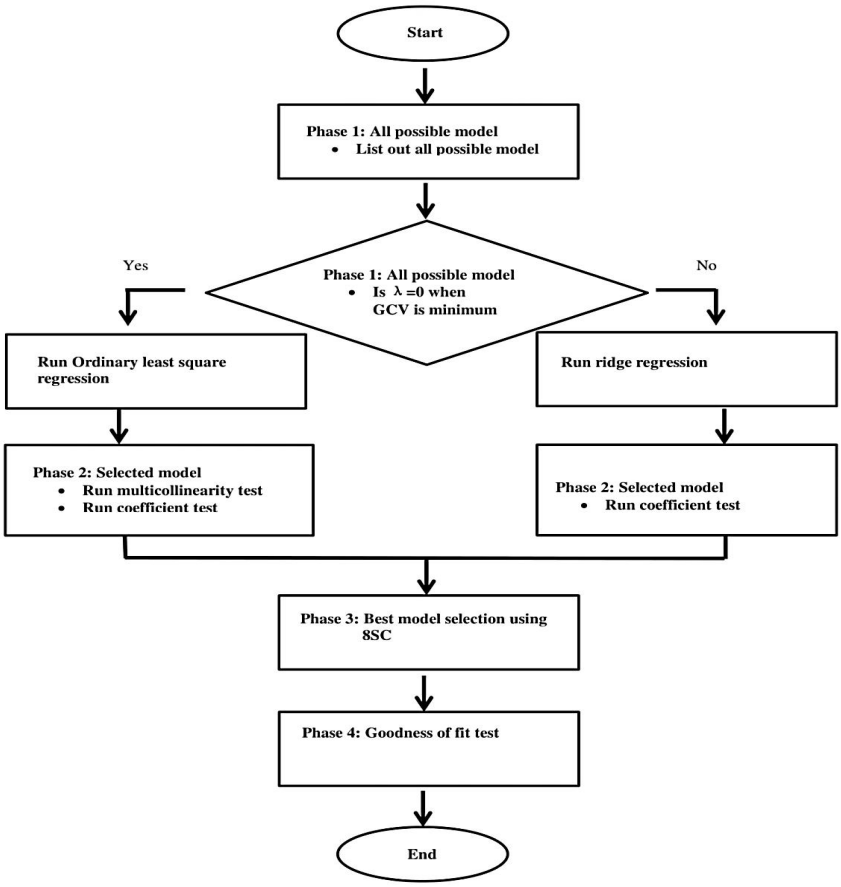


Figure 5. Flow Chart on the Procedures in Getting Best Model

RESULTS AND DISCUSSIONS

Data Collection and Procedure

In this study, the data were taken during the experiment drying process for drying fish by using v-GHSD at Selakan Island, Semporna. The fish was dried to thirty-five percent moisture content until it reached the EMC. The data collection started from 8th to 12th October 2019. The total number of data collected was 1914 and there were no missing data. Five percent of the dataset which is 96 data was reserved as test data. In this study, moisture content of fish (y)is the dependent variable, whereas the inlet temperature chamber (X_1), outlet temperature chamber (X_2), outlet humidity chamber (X_3), inlet humidity chamber (X_4) and solar radiation (X_5) are the independent variables. The five days drying data was collected for every thirty minutes.

Since five independent variables were used in this study, there were total 80 possible models until fourth order of interaction as shown in Table 2.

Table 2
All possible models

No of variables	Single	Interact				Total	Model Label
		1st Order	2nd Order	3rd Order	4th Order		
1	5	-	-	-	-	5	**M1-5**
2	10	10	-	-	-	20	**M6-25**
3	10	10	10	-	-	30	**M26-55**
4	5	5	5	5	-	20	**M56-75**
5	1	1	1	1	1	5	**M76-80**
Total Models	31	26	16	6	1	80	

The coefficient test is conducted, and a list of selected models with its ridge parameter λ and Error Sum of Squares (SSE) are obtained. Where *k* denotes the number of variables left in the model. The models with the same number of variables are kept in a single group. After grouping, the 69 models are left out of 80 possible models, and results are shown in Table 3. For example, M21.0.0 represents the original model. One variable is removed during the multicollinearity test so the model becomes M21.1.0 while no variable is removed from the coefficient test. So, the final model remains as M21.1.0.

Table 3
Selected Models by using OLS or Ridge Regression

Sr. NO	Selected models using OLS/Ridge	*k*	*λ*	SSE
1	*M1.0.0*	1	0.00000	407251.8445
2	*M2.0.0*	1	0.00000	382866.6136
3	*M3.0.0*	1	0.00000	415428.5487
4	*M4.0.0*	1	0.00000	496042.5512
5	*M5.0.0*	1	0.00000	346497.159
6	*M6.0.0=M16.1.0*	2	0.00800	381262.7479
7	*M7.0.0=M17.0.1*	2	0.00200	322025.8196
8	*M8.0.0*	2	0.00100	358834.0995

Table 3 *(Continued)*

Sr. NO	Selected models using OLS/Ridge	k	λ	SSE
9	M9.0.0	2	0.00500	342096.8842
10	M10.0.0=M20.1.0	2	0.00200	312301.581
11	*M11.0.0*	2	0.00000	303981.4905
12	*M12.0.0=M22.0.1*	2	0.00500	318511.0194
13	*M13.0.0*	2	0.00500	414295.4192
14	*M14.0.0*	2	0.00500	334854.2714
15	*M15.0.0*	2	0.00200	343805.3683
16	*M16.1.0*	2	0.00000	379272.0951
17	*M18.0.0*	3	0.00300	348307.4055
18	*M19.0.0*	3	0.01800	342107.8585
19	*M21.1.0*	2	0.00000	301316.8478
20	*M23.0.0*	3	0.00100	395358.3048
21	*M24.0.0*	3	0.01400	333336.4064
22	*M25.0.0*	3	0.00500	334033.4548
23	*M26.0.0*	3	0.01000	306848.4051
24	*M27.0.0*	3	0.00100	296660.9027
25	*M28.0.0*	3	0.00100	310557.9255
26	*M29.0.0=M59.0.1*	3	0.00100	288289.5733
27	*M30.0.0*	3	0.01000	315115.4032
28	*M31.0.0*	3	0.00600	325462.092
29	*M32.0.0*	3	0.00100	258597.1048
30	*M33.0.0=M57.0.1*	3	0.00900	294813.4643
31	*M34.0.0=M58.0.1*	3	0.00100	270406.179
32	*M35.0.0*	3	0.00600	333986.6613
33	*M36.3.0*	3	0.00000	304165.6933
34	*M37.2.1*	3	0.00000	289651.0102
35	*M38.3.0*	3	0.00000	315425.3665
36	*M39.0.1*	5	0.00900	287071.6748
37	*M40.2.0*	4	0.00000	294634.3668
38	*M41.1.1*	4	0.00000	319472.1287
39	*M42.2.0=M52.3.0*	4	0.00000	257707.6509
40	*M43.2.0*	4	0.00000	260872.4566
41	*M44.0.2*	4	0.00800	268781.3893
42	*M45.0.3*	3	0.00400	320402.8475

Table 3 *(Continued)*

Sr. NO	Selected models using OLS/Ridge	k	λ	SSE
43	M46.4.0	3	0.00000	304165.6933
44	M47.2.2	3	0.00000	289072.1514
45	M48.4.0	3	0.00000	314537.9494
46	M49.2.2	3	0.00000	287679.0882
47	M50.2.1	4	0.00000	291784.7212
48	M51.1.1	5	0.00000	318219.7034
49	M53.3.0	4	0.00000	264752.4059
50	M54.3.1	2	0.00000	281515.2994
51	M55.0.2	5	0.00600	317817.7528
52	M56.0.0/M76.0.1	4	0.00200	247253.6408
53	M60.0.0	4	0.00100	251008.3619
54	M61.4.2=M66.7.3=M71.8.3	4	0.00000	246598.7709
55	M62.5.0	5	0.00000	257482.0627
56	M63.4.2	4	0.00000	268756.3648
57	M64.2.2	6	0.00000	283819.5595
58	M65.3.3	4	0.00000	249022.3153
59	M67.9.1	4	0.00000	263008.9208
60	M68.8.2	4	0.00000	266633.687
61	M69.4.2	8	0.00000	280462.7962
62	M70.6.2	6	0.00000	247520.6769
63	M72.10.0	4	0.00000	266930.9018
64	M73.9.2	4	0.00000	268766.6487
65	M74.4.2	9	0.00000	276344.7077
66	M75.7.2	6	0.00000	248298.1572
67	M77.6.1	8	0.00000	244108.2359
68	M78.7.3=M79.17.4	11	0.00000	230561.7746
69	M80.18.3	9	0.00000	236260.0805

After the coefficient test, all of the best selected models, as shown in Table 4 are evaluated by using 8SC.

From the results in Table 4, M78.7.3 provides the minimum of all the 8SC value. Hence, M78.7.3 is obtained as the best model among all the selected models. Since M78.7.3 is with λ equal to 0, hence, this model is an OLS regression model. Furthermore, M56.0.0 with λ equal to 0.002 provides the minimum 8SC value for the ridge regression model. The best model M78.7.3 for OLS and M56.0.0 for ridge are shown as in Equation 9 and Equation 10 respectively. The coefficients are obtained using R software.

$$M78.7.3 = \hat{Y} = -105.3 + 5.007x_2 - 0.0515^{x_3} + 0.03444x_{14} + 0.0186x_{24} - 0.0007453x_{25} - 0.00279600x_{45} - 0.00133600x_{124} + 0.00004118x_{135} + 0.00003168x_{145} - 0.00011440x_{234} + 0.00002246x_{345} \quad (9)$$

Table 4

8SC for OLS/Ridge Selected Models

Selected models from OLS/Ridge	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M1.0.0	224.5043	224.5043	224.5046	225.0066	224.5049	225.8682	224.2576	224.5038
M2.0.0	211.0616	211.0616	211.0618	211.5337	211.0621	212.3438	210.8296	211.0611
M3.0.0	229.0119	229.0119	229.0122	229.5242	229.0124	230.4031	228.7602	229.0113
M4.0.0	273.4517	273.4517	273.452	274.0634	273.4523	275.1129	273.1512	273.451
M5.0.0	191.0123	191.0123	191.0125	191.4396	191.0128	192.1727	190.8024	191.0118
M6.0.0=M16.1.0	210.4088	210.4088	210.4093	211.1152	210.4099	212.329	210.0621	210.4076
M7.0.0=M17.0.1	177.7175	177.7175	177.718	178.3142	177.7184	179.3394	177.4247	177.7165
M8.0.0	198.031	198.031	198.0315	198.6959	198.0321	199.8383	197.7047	198.0299
M9.0.0	188.7942	188.7942	188.7947	189.428	188.7952	190.5172	188.4831	188.7931
M10.0.0=M20.1.0	172.3509	172.3509	172.3514	172.9296	172.3519	173.9239	172.067	172.35
M11.0.0	167.7593	167.7593	167.7597	168.3225	167.7602	169.2903	167.4829	167.7584
M12.0.0=M22.0.1	175.7777	175.7777	175.7782	176.3679	175.7787	177.382	175.4882	175.7768
M13.0.0	228.6386	228.6386	228.6392	229.4063	228.6399	230.7252	228.2619	228.6374
M14.0.0	184.7972	184.7972	184.7977	185.4176	184.7982	186.4837	184.4927	184.7962
M15.0.0	189.737	189.737	189.7375	190.3741	189.7381	191.4686	189.4244	189.736
M16.1.0	209.3102	209.3102	209.3107	210.0129	209.3113	211.2204	208.9653	209.309
M18.0.0	192.4332	192.4332	192.4341	193.2951	192.435	194.7783	192.0107	192.4313
M19.0.0	189.008	189.008	189.0089	189.8546	189.0099	191.3115	188.5931	189.0062
M21.1.0	166.2887	166.2887	166.2892	166.8471	166.2897	167.8064	166.0148	166.2878

Table 4 *(Continued)*

Selected models from OLS/Ridge	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M23.0.0	218.4279	218.4279	218.4289	219.4063	218.43	221.0899	217.9483	218.4258
M24.0.0	184.162	184.162	184.1629	184.9869	184.1638	186.4063	183.7577	184.1602
M25.0.0	184.5471	184.5471	184.548	185.3737	184.5489	186.7961	184.1419	184.5453
M26.0.0	169.5279	169.5279	169.5287	170.2872	169.5295	171.5939	169.1557	169.5262
M27.0.0	163.8995	163.8995	163.9003	164.6336	163.9011	165.8969	163.5396	163.8979
M28.0.0	171.5773	171.5773	171.5781	172.3458	171.579	173.6683	171.2006	171.5756
M29.0.0=M59.0.1	159.2745	159.2745	159.2752	159.9879	159.276	161.2155	158.9248	159.2729
M30.0.0	174.0952	174.0952	174.0961	174.875	174.0969	176.2169	173.713	174.0935
M31.0.0	179.8116	179.8116	179.8124	180.617	179.8133	182.0029	179.4168	179.8098
M32.0.0	142.8699	142.8699	142.8706	143.5099	142.8713	144.6111	142.5563	142.8686
M33.0.0=M57.0.1	162.8788	162.8788	162.8796	163.6084	162.8804	164.8638	162.5212	162.8772
M34.0.0=M58.0.1	149.3942	149.3942	149.395	150.0634	149.3957	151.2149	149.0663	149.3928
M35.0.0	184.5212	184.5212	184.5221	185.3477	184.523	186.77	184.1161	184.5194
M36.3.0	168.0457	168.0457	168.0465	168.7984	168.0473	170.0937	167.6768	168.0441
M37.2.1	160.0266	160.0266	160.0274	160.7434	160.0282	161.9769	159.6753	160.0251
M38.3.0	174.2665	174.2665	174.2673	175.047	174.2682	176.3902	173.8839	174.2648
M39.0.1	158.9509	158.9509	158.9527	160.0201	158.9544	161.8655	158.4281	158.9475
M40.2.0	162.959	162.959	162.9602	163.8719	162.9615	165.4453	162.5121	162.9566
M41.1.1	176.6965	176.6965	176.6978	177.6864	176.6992	179.3923	176.2119	176.6938
M42.2.0=M52.3.0	142.5353	142.5353	142.5363	143.3338	142.5374	144.7099	142.1443	142.5331

Table 4 *(Continued)*

Selected models from OLS/ Ridge	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M43.2.0	144.2857	144.2857	144.2868	145.094	144.2879	146.487	143.8899	144.2835
M44.0.2	148.66	148.66	148.6611	149.4928	148.6623	150.9281	148.2523	148.6578
M45.0.3	177.0164	177.0164	177.0173	177.8093	177.0181	179.1737	176.6278	177.0147
M46.4.0	168.0457	168.0457	168.0465	168.7984	168.0473	170.0937	167.6768	168.0441
M47.2.2	159.7068	159.7068	159.7076	160.4222	159.7084	161.6532	159.3562	159.7053
M48.4.0	173.7762	173.7762	173.777	174.5546	173.7779	175.894	173.3947	173.7745
M49.2.2	158.9372	158.9372	158.9379	159.6491	158.9387	160.8741	158.5883	158.9356
M50.2.1	161.3829	161.3829	161.3841	162.287	161.3854	163.8451	160.9403	161.3805
M51.1.1	176.1975	176.1975	176.1995	177.3827	176.2014	179.4283	175.6179	176.1937
M53.3.0	146.4316	146.4316	146.4327	147.252	146.4339	148.6657	146.03	146.4294
M54.3.1	155.3608	155.3608	155.3612	155.8824	155.3616	156.7787	155.1048	155.36
M55.2.1	175.975	175.975	175.9769	177.1586	175.9788	179.2017	175.3961	175.9712
M56.0.0/M76.0.1	136.7533	136.7533	136.7543	137.5194	136.7553	138.8397	136.3782	136.7512
M60.0.0	138.83	138.83	138.831	139.6077	138.8321	140.9481	138.4492	138.8279
M61.4.2=M66.7.3=M71.8.3	136.3911	136.3911	136.3921	137.1551	136.3931	138.472	136.017	136.389
M62.5.0	142.5672	142.5672	142.5688	143.5262	142.5704	145.1814	142.0983	142.5641
M63.4.2	148.6462	148.6462	148.6473	149.4789	148.6484	150.914	148.2385	148.6439

Table 4 *(Continued)*

Selected models from OLS/Ridge	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M64.2.2	157.3232	157.3232	157.3256	158.5585	157.3279	160.6938	156.7198	157.3186
M65.3.3	137.7315	137.7315	137.7325	138.5031	137.7336	139.8328	137.3537	137.7294
M67.9.1	145.4673	145.4673	145.4684	146.2823	145.4695	147.6867	145.0684	145.4651
M68.8.2	147.4721	147.4721	147.4733	148.2983	147.4744	149.7221	147.0677	147.4699
M69.4.2	155.805	155.805	155.8088	157.3796	155.8127	160.1098	155.0375	155.7974
M70.6.2	137.2025	137.2025	137.2045	138.2798	137.2066	140.142	136.6762	137.1984
M72.10.0	147.6365	147.6365	147.6377	148.4636	147.6388	149.889	147.2316	147.6343
M73.9.2	148.6519	148.6519	148.653	149.4846	148.6541	150.9198	148.2442	148.6496
M74.4.2	153.6862	153.6863	153.6909	155.413	153.6956	158.4115	152.8455	153.677
M75.7.2	137.6335	137.6335	137.6355	138.7141	137.6376	140.5822	137.1056	137.6294
M77.6.1	135.609	135.609	135.6123	136.9795	135.6157	139.3558	134.941	135.6024
M78.7.3=M79.17.4	128.507	128.507	128.5126	130.2416	128.5183	133.2628	127.6643	128.4959
M80.18.3	131.3936	131.3936	131.3976	132.8699	131.4016	135.4335	130.6748	131.3857

$$M56.0 = -66.2806 + 0.3273x_1 + 2.8945\,x_2 - 0.0317x_3 + 0.4450x_4 \quad (10)$$

For model M78.7.3, eleven variables were retained in the model, including the interaction terms. The signs of the coefficient show the type of relationship of the independent variable with the dependent factor. The coefficients that are far away from the zero mean that they are the strongest factors in the analysis. From the results, the significance of the variables with the interaction term shows that the interaction terms are very important and cannot be ignored. For model M56.0.0, four variables are remained in the model without including the interaction term. For both of the models, MAPE was computed by using formulae as stated in Equation 8. The MAPE value for M78.7.3 is 15.7342. The MAPE value for M56.0.0 is 17.4054. Both of the MAPE value is less than 20 and indicates both models can be used to forecast the moisture content of the fish.

```
        Runs Test

data:  std_res$stdres
statistic = 0.59445, runs = 971, n1 = 957, n2 = 957, n = 1914, p-value
= 0.5522
alternative hypothesis: nonrandomness
```

Figure 6. Run test for standardized residuals M78.7.3

```
        Runs Test

data:  std_res$stdres
statistic = 1.2346, runs = 985, n1 = 957, n2 = 957, n = 1914, p-value = 0.217
alternative hypothesis: nonrandomness
```

Figure 7. Run test for standardized residuals M56.0.0

```
        One-sample Kolmogorov-Smirnov test

data:  std_res$stdres
D = 0.070452, p-value = 1.121e-08
alternative hypothesis: two-sided
```

Figure 8. Kolmogorov-Smirnov test for standardized residuals M78.7.3

To test the randomness of the standardized residuals, a run test was conducted. From the results as shown in Figure 6, the run test p-value was equal to 0.5522 for M78.7.3. From the results as shown in Figure 7, the run test p-value was equal to 0.217 for M56.0.0. Since the p-value of the run test of both models is more than 0.05, hence, the standardized residuals are random. Furthermore, Kolmogorov-Smirnov test was conducted for M78.7.3 to test the normality assumptions of residuals. The results are shown in Figure 8. The p value obtained from the Kolmogorov-Smirnov test for M78.7.3 is less than 0.05. Therefore, the residuals are not normally distributed.

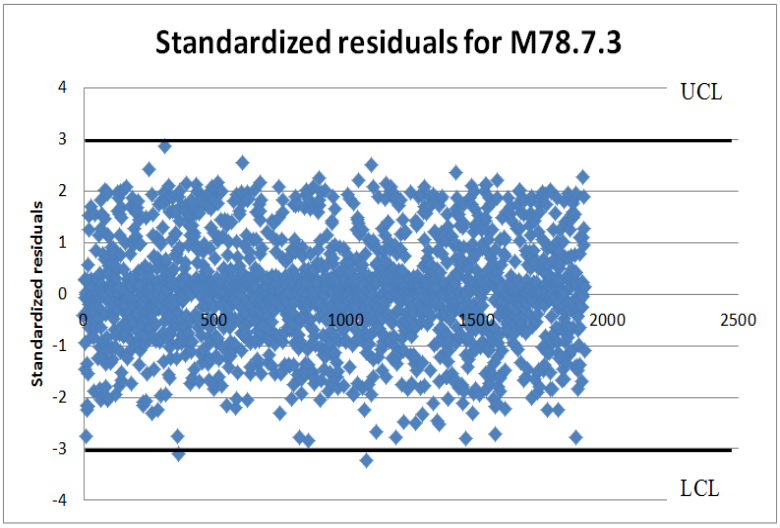


Figure 9. Scatterplot of standardized residuals for OLS regression M78.7.3

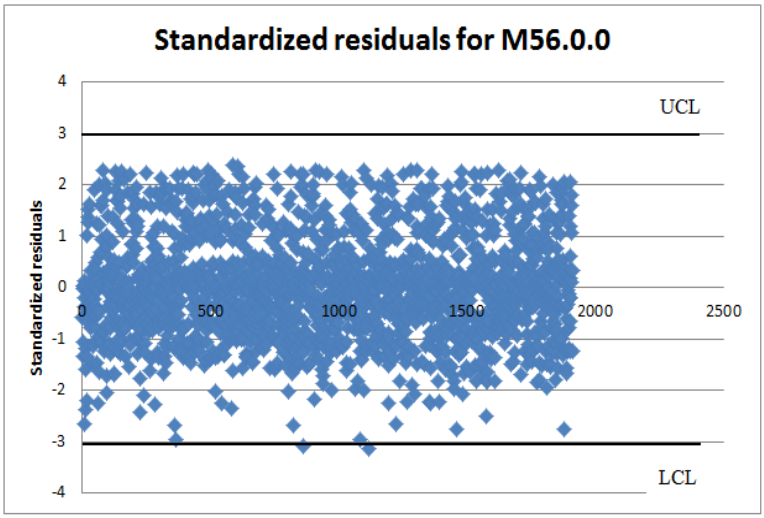


Figure 10. Scatterplot of standardized residual for Ridge regression M56.0.0

Outliers outside the 3-sigma limit can be observed from Figure 9 and 10. UCL and LCL represent the upper-class limit and lower-class limit respectively. The percentage of outliers is obtained based on the number of observations outside the 3-sigma limit. Table 5 shows the percentage of outliers outside 3-sigma limit for M78.7.3 and M56.0.0.

Table 5

Percentage of outliers outside 3-sigma limits

Selected model	Method	$\mu \pm 3\sigma$
M78.7.3	OLS	0.11%
M56.0.0	Ridge	0.11%

There are a total of 0.11% of outliers for both of the OLS and the ridge model. Apart from standardized residual plots, a box plot is able to provide a clear graphical representation by labeling outliers (Ramachandran & Tsokos, 2014). Hence, box plot of both models are observed as shown in Figure 11 and 12.

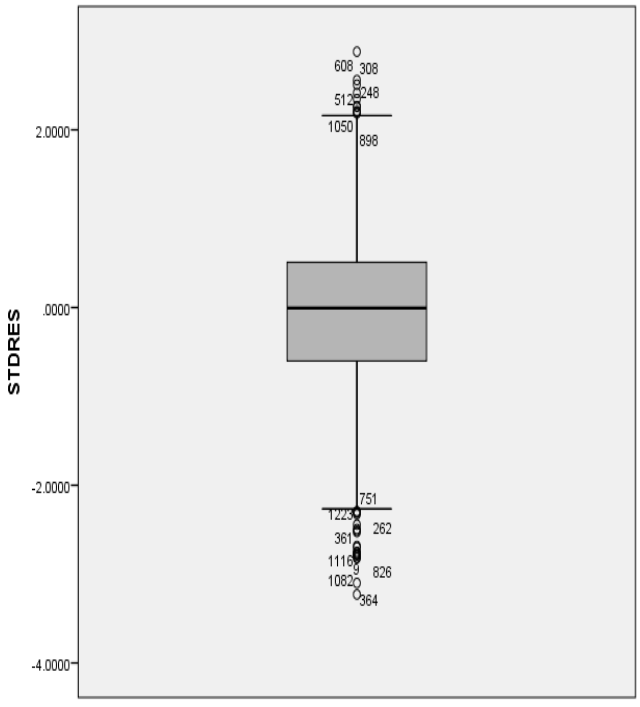


Figure 11. Box plot for OLS regression M78.7.3

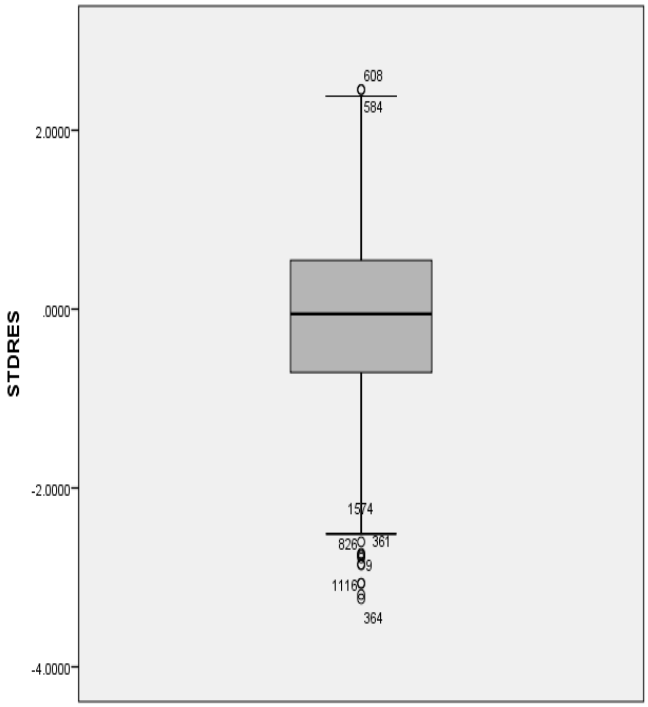


Figure 12. Box plot for Ridge regression M56.0.0

From Figure 11 and 12, the outliers in the dataset can be observed. There are more outliers for M78.7.3 as compared to M56.0.0. Deleting the outliers is not always the best option in the real life dataset. So, the results obtained from OLS cannot be trusted for a better forecast in the presence of outliers. On the other hand, ridge regression has the ability to deal in the presence of outliers (Steece, 1986). So, the ridge regression can be trusted to forecast the moisture content of the fish. Although the MAPE for OLS regression is less than the MAPE for the ridge regression, but due to the non- normality of the residuals and the presence of outliers, OLS cannot be trusted for a better forecast. On the other hand, ridge regression does not need any kind of normality assumptions.

CONCLUSIONS

In a nutshell, the best OLS model obtained to forecast the moisture content of fish was M78.7.3 with a total of 11 independent variables in this model after checking the multicollinearity and conducting a coefficient test. Furthermore, the best ridge model obtained to forecast the moisture content of fish was M56.0.0 with ridge parameter 0.002 and a total of 4 independent variables in this model after the conduct coefficient test. However, more outliers are detected for OLS model M78.7.3 as compared to the ridge

model M56.0.0. The MAPE value of both of the models shows satisfying results. For OLS model M78.7.3, the MAPE value is 15.7342. The MAPE value for ridge model M56.0.0 is 17.4054. Due to non-normality of the residuals, and presence of outliers in the dataset, ridge regression is preferred for the best forecast with the MAPE of 17.4054. So, the moisture content of fish can forecast with the crucial factors as inlet temperature chamber, outlet temperature chamber, outlet humidity chamber and inlet humidity chamber. Since the MAPE is less than 20, so it will provide a good forecast. This paper only addressed multicollinearity and outliers by assuming no autocorrelated errors. We will consider autocorrelated errors in future study.

ACKNOWLEDGEMENT

The authors would like to extend their greatest gratitude to Universiti Sains Malaysia for funding this study under the Short term Grant Scheme (304.PMATHS.6315132).

REFERENCES

Abdullah, N., Jubok, Z. H., & Ahmed, A. (2011). Improved stem volume estimation using P-Value approach in polynomial regression models. *Research Journal of Forestry, 5*(2), 50-65.

Abdullah, N., Lee, C. L., & Jubok, Z. H. (2015). Factors on palm oil fruit bunches production volume for biomass fuel and biofuel during cogeneration processes. *Journal of the Japan Institute of Energy, 94*(12), 1428-1439.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics, 21*(1), 243-247.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Alfiya, P., Murali, S., Delfiya, D. A., & Samuel, M. P. (2018). Empirical modelling of drying characteristics of elongate glassy perchlet (Chanda nama)(Hamilton, 1822) in solar hybrid dryer. *Fishery Technology, 55*(2), 138-142.

Ali, M. K. M., Fudholi, A., Muthuvalu, M., Sulaiman, J., & Yasir, S. M. (2017a, December 4-7). Implications of drying temperature and humidity on the drying kinetics of seaweed. In *Proceedings of the 13th IMT-GT International Conference on Mathematics, Statistics and their Applications (ICMSA2017)*. Kedah, Malaysia.

Ali, M. K. M., Fudholi, A., Muthuvalu, M., Sulaiman, J., Yasir, S. M., & Hurtado, A. Q. (2017b). Post-harvest handling of eucheumatoid seaweeds. In *Tropical seaweed farming trends, problems and opportunities* (pp. 131-145). Cham, Switzerland: Springer.

Bodirsky, B. L., Rolinski, S., Biewald, A., Weindl, I., Popp, A., & Lotze-Campen, H. (2015). Global food demand scenarios for the 21st century. *PLoS One, 10*(11), 1-27.

Chatterjee, S., & Hadi, A. S. (2015). *Regression analysis by example*. Hoboken, New Jersey: John Wiley & Sons.

Delaney, N. J., & Chatterjee, S. (1986). Use of the bootstrap and cross-validation in ridge regression. *Journal of Business and Economic Statistics, 4*(2), 255-262.

Ertekin, C., & Yaldiz, O. (2004). Drying of eggplant and selection of a suitable thin layer drying model. *Journal of Food Engineering, 63*(3), 349-359.

FAO. (1996). *The state of food and agriculture 1996*. Rome, Italy: Food & Agriculture Org.

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics, 21*(2), 215-223.

Guan, Z., Wang, X., Li, M., & Jiang, X. (2013). Mathematical modeling on hot air drying of thin layer fresh tilapia fillets. *Polish Journal of Food and Nutrition Sciences, 63*(1), 25-33.

Gujarati, D. N. (2004). *Basic econometrics* (4th Ed.). New York, USA: The McGraw-Hill Companies.

Hajijubok, Z., & Gopal , P. K. (2008). Procedure in getting best model using multiple regression. *Journal of Borneo Science, 23*, 47-63.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological), 41*(2), 190-195.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55-67.

Hossain, M., & Bala, B. (2007). Drying of hot chilli using solar tunnel drier. *Solar Energy, 81*(1), 85-92.

Jamal, N., & Rind, M. Q. (2007). Ridge regression: A tool to forecast wheat area and production. *Pakistan Journal of Statistics and Operation Research, 3*(2), 125-134.

Javaid, A., Ismail, M., & Ali, M. K. M. (2020). Efficient model selection of collector efficiency in solar dryer using hybrid of LASSO and robust regression. *Pertanika Journal of Science and Technology, 28*(1), 193-210.

Javaid, A., Ismail, M. T., & Ali, M. K. M. (2019a). Model selection for collector efficiency of seaweed drier by using LASSO and multiple regression analysis using 8sc. In *Proceedings of the International Conference on Mathematical Sciences and Technology 2018 (MATHTECH2018)* (pp. 1-9). New York, NY: AIP Publishing LLC.

Javaid, A., Muthuvalu, M. S., Sulaiman, J., Ismail, M. T., & Ali, M. K. M. (2019b). Forecast the moisture ratio removal during seaweed drying process using solar drier. In *Proceedings of the International Conference on Mathematical Sciences and Technology 2018 (MATHTECH2018)* (pp. 1-8). New York, NY: AIP Publishing LLC.

Kennard, R. W. (1971). A note on the Cp statistic. *Technometrics, 13*(4), 899-900.

Khalaf, G. (2012). A proposed ridge parameter to improve the least square estimator. *Journal of Modern Applied Statistical Methods, 11*(2), 443-449.

Kituu, G. M., Shitanda, D., Kanali, C., Mailutha, J., Njoroge, C., Wainaina, J., & Silayo, V. (2010). Thin layer drying model for simulating the drying of Tilapia fish (Oreochromis niloticus) in a solar tunnel dryer. *Journal of Food Engineering, 98*(3), 325-331.

Krokida, M. K., Karathanos, V., Maroulis, Z., & Marinos-Kouris, D. (2003). Drying kinetics of some vegetables. *Journal of Food Engineering, 59*(4), 391-403.

Mahajan, V., Jain, A. K., & Bergier, M. (1977). Parameter estimation in marketing models in the presence of multicollinearity: An application of ridge regression. *Journal of Marketing Research, 14*(4), 586-591.

Rajarathinam, A., & Vinoth, B. (2014). Outlier detection in simple linear regression models and robust regression–A case study on wheat production data. *International Journal of Scientific Research, 3*(2), 531-536.

Ramachandran, K. M., & Tsokos, C. P. (2014). *Mathematical statistics with applications in R* (2nd Ed.). Oxford, UK: Elsevier.

Ramanatam, R. (2002). *Introductory econometrics with application* (5th Ed.). South Western, USA: Harcourt College Publishers.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics, 12*(4), 1215-1230.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika, 68*(1), 45-54.

Silva, B. G., Fileti, A. M. F., Foglio, M. A., Rosa, P. D. T. V., & Taranto, O. P. (2017). Effects of different drying conditions on key quality parameters of pink peppercorns (*Schinus terebinthifolius* Raddi). *Journal of Food Quality, 2017*, 1-12.

Steece, B. M. (1986). Regressor space outliers in ridge regression. *Journal of Communications in Statistics*, *15*(12), 3599-3605.

Stiling, J., Li, S., Stroeve, P., Thompson, J., Mjawa, B., Kornbluth, K., & Barrett, D. M. (2012). Performance evaluation of an enhanced fruit solar dryer using concentrating panels. *Energy for Sustainable Development, 16*(2), 224-230.

Tiwari, A. (2016). A review on solar drying of agricultural produce. *Journal of Food Processing and Technology, 7*(9), 1-12.

Ullah, M. I., Aslam, M., & Altaf, S. (2018). lmridge: A comprehensive R package for ridge regression. *The R Journal, 10*(2), 326-346.

Wen, Y. W., Tsai, Y. W., Wu, D. B. C., & Chen, P. F. (2013). The impact of outliers on net-benefit regression model in cost-effectiveness analysis. *PLoS One, 8*(6), 1-9.

Yahaya, A. H., Abdullah, N., & Zainodin, H. (2012). Multiple regression models up to first-order interaction on hydrochemistry properties. *Asian Journal of Mathematics and Statistics, 5*(4), 121-131.

Zhang, J., & Ibrahim, M. (2005). A simulation study on SPSS ridge regression and ordinary least squares regression procedures for multicollinearity data. *Journal of Applied Statistics, 32*(6), 571-588.

Pertanika J. Sci. & Technol. 28 (4): 1203 - 1220 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Fast and Robust Diagnostic Technique for the Detection of High Leverage Points

Habshah Midi[1,2]*, Hasan Talib Hendi[1], Jayanthi Arasan[2] and Hassan Uraibi[3]

[1]*Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[2]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*
[3]*Department of Statistics, University of Al-Qadisiyah, 88 -Al-Qadisiyah -Al-Diwaniyah, Iraq*

ABSTRACT

High Leverage Points (HLPs) are outlying observations in the X -directions. It is very imperative to detect HLPs because the computed values of various estimates are affected by their presence. It is now evident that Diagnostic Robust Generalized Potential which is based on the Minimum Volume Ellipsoid (DRGP(MVE)) is capable of detecting multiple HLPs. However, it takes very long computational running times. Another diagnostic measure which is based on Index Set Equality denoted as DRGP(ISE) is put forward with the main aim of reducing its running time. Nonetheless, it is computationally not stable and still suffers from masking and swamping effects. Hence, in this paper, we propose another version of diagnostic measure which is based on $\sqrt{n}$ Reweighted Fast Consistent and High Breakdown (RFCH) estimators. We call this measure Diagnostic Robust Generalized Potential based on RFCH and it is denoted by DRGP(RFCH). The results of simulation study and real data indicate that our proposed method outperformed the other two methods in term of having the least computing time, highest percentage of correct detection of HLPs and smallest percentage of swamping and masking effects compared to the DRGP(MVE) and DRGP (ISE).

Keywords: Diagnostic robust generalized potentials, high leverage points, mahalanobis distance, outliers

ARTICLE INFO

Article history:
Received: 19 April 2020
Accepted: 27 July 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.05

E-mail addresses:
habshah@upm.edu.my (Habshah Midi)
h.applied.t88@gmail.com (Hasan Talib Hendi)
jayanthi@upm.edu.my (Jayanthi Arasan)
hssn.sami1@gmail.com (Hassan Uraibi)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

The Ordinary Least Squares (OLS) is the most popular technique in regression analysis because of tradition and ease of computation. Moreover, the OLS is easy to use as it is available in most of the statistical software like SPSS, SAS and MINITAB. The OLS technique has many attractive features under normality assumption of regression errors, not only in parameters estimation but also in testing of hypothesis. However, many are not aware that the one immediate consequence of the presence of outliers especially outlying observations in the X- direction which is call High Leverage Points (HLPs) may cause apparent non-normality (Huber, 1973). Since most of the statistical analysis are based on normality assumption, the violation of this assumption may lead to invalid inferential statements and inaccurate predictions. Evidences are now available in the literatures that the presence of HLPs have an adverse effect on the computed values of various estimates (Rousseeuw, 1985; Imon & Khan, 2003; Midi et al., 2009; Riazoshams et al., 2010; Bagheri et al., 2012). As such it is very crucial to search for a very effective method of detecting HLPs. The HLPs can easily be spotted from a plot of response variable against the predictor variable for simple linear regression model. Nonetheless it is hard to identify multiple HLPs for more than one independent variable due to swamping and masking effect (Peña & Yohai, 1995).

There are many papers that deal with the diagnostic tools for the identification of HLPs (Rousseeuw, 1985; Rousseeuw & Driessen, 1999; Midi et al., 2009). Midi et al. (2009) had shown that their method was very successful for the detection of HLPs compared to hat matrix approach of (Hoaglin & Welsh, 1978) and Hadis' potential (Hadi,1992). Even though some of them are able to correctly identify multiple HLPs, their running times are very long due to using Minimum Volume Ellipsoid (MVE) or Minimum Covariance Determinant (MCD) for obtaining the final estimator of location and scatter. Lim and Midi (2016) exemplified that Index Set Equality (ISE) had tremendously sped up the computation of location and scatter estimator, even much faster than fast MCD (Rousseeuw & Driessen, 1999). The only shortcoming of this method is that it is not very stable because its computation depends on the selected initial subset, h. According to Salleh (2013), the final estimator of location and scatter of (ISE) is equivalent to MCD if the same initial subset is utilized, otherwise the results will be quite different. In order to obtain more efficient and much faster location and scatter estimators, we propose employing Reweighted Fast Consistent and High Breakdown (RFCH) estimators (Olive & Hawkins, 2010; Alkenani & Yu, 2013). By employing the results of Lopuhaa (1999), Olive and Hawkin (2010) proved that the location and scatter estimators of RFCH were $\sqrt{n}$ consistent estimators. For this reason, Uraibi et al. (2017) developed a robust forward selection method by formulating a correlation matrix based on RFCH estimators which produced very appealing results. Hence in this paper, we propose using the $\sqrt{n}$ RFCH consistent estimators of location and scatter in the establishment of Robust Mahalanobis Distance (RMD).

The paper is organized as follows. Section 2 discusses the importance of detection of high leverage points. Section 3 reviews a few methods of detection of high leverage points. The proposed Diagnostic Robust Generalized Potential based on RFCH estimators is presented in Section 4. Section 5 discusses the results of the simulation and numerical example. The concluding remarks are given in Section 6.

MATERIALS AND METHODS

Real Data to Show Why it is Very Important to Detect HLPs

As already mentioned in the preceding section, various estimates can be affected by HLPs. That is why it is very important to first check their existence before making any inferences to avoid misleading conclusion. In this section, we want to show that HLPs can cause multicollinearity and heteroscedasticity by using real examples. Let us first focus on Hawkins Bradu Kass Data (Hawkins et al., 1984). This artificial data set consists of 75 observations and 3 independent variables. Hawkins et al. (1984) claimed that this data set had 14 HLPs. Bagheri et al. (2012) exemplified that the first 14 observations (included in the original data) as displayed in Table 1, caused multicollinearity evident by showing maximum value of Variance Inflation Factor (VIF=33.342 for $X3$) greater than 10. On the contrary, no multicollinearity was observed in their absence.

Table 1
Multicollinearity Diagnostics (VIF) for Hawkin Bradu Kass Data

Status	$X1$	$X2$	$X3$
Original Data	13.432	23.853	33.432
Without observations 1-14	1.012	1.017	1.027

Education Expenditure Data taken from Chatterjee and Hadi (2006) will be our second example to illustrate that HLPs can caused heteroscedasticity. Many authors frequently used this data (n = 50 observations with three independent variables) to deal with heteroscedasticity (Chatterjee & Hadi, 2006; Imon, 2002; Midi et al., 2014). This data consists of 50 observations where per capita income on education project for 1975 is the dependent variable and three explanatory variables namely per capita income in 1973, number of residents per thousand under 18 years of age, and number of residents per thousand under 18 years of age in 1974. According to Midi et al. (2014), observation 49, i.e. Alaska (AK) is HLP and influences the heteroscedasticity pattern of the data. We investigated this data and apply the White test (WT) which is a Lagrange Multiplier (LM) test statistic proposed by White (1980) to test the presence of heteroscedasticity in linear regression model. The white test is defined as $LM = nR^2$, where R^2 is the coefficient of multiple determination and n is size of sample. The LM test statistics is distributed

as Chi-Squared χ_p^2 , where p is the number of predictors. Table 2 exhibits the values of *LM* test statistics with their corresponding *p*-values. We can see from Table 2 that in the absence of HLP the WT shows no heteroscedasticity but in their presence the WT shows heteroscedasticity.

Table 2
Heteroscedasticity Diagnostics (White test)

Status	$LM = nR^2$	*p*-values
Without AK (HLP)	5.7978	0.1219
With AK (HLP)	22.7817	4.48e-05

We have seen the effect that the HLPs had on the heteroscedasticity and multicollinearity pattern of a data and it is crucial to detect them before any further analysis to be carried out. This is the reason why we need to find the more reliable method for detecting their existence.

Review of Some Methods of Identifications of HLPs

In this section, some methods of identification of HLPs are reviewed.

Diagnostics Robust Generalized Potential (DRGP)

Mahalanobis (1936) defined Mahalanobis Distance (MD) as a measure of deviation of a data point from its center. Let us write the i^{th} vector of predictor variables as:

$$X_i' = (1, X_1, X_2, \dots, X_P) = (1, t_i),$$

where t_i is a p-dimensional row vector. The mean vector and the variance covariance matrix are calculated as:

$$\bar{t} = {}^1\!/_n \sum_{i=1}^n t_i \text{ and } C = \left(\frac{1}{n-1}\right) \sum_{i=1}^n (t_i - \bar{t})\,(t_i - \bar{t})', \text{ respectively.}$$

Subsequently, the *(MD)* for each observation is written as Equation 1:

$$MD_i = \sqrt{(t_i - T(X))' C(X)^{-1} (t_i - T(X))} \qquad i = 1,2, \dots, n, \qquad [1]$$

where *T(X)* is the mean vector $(\bar{t})$ and *C(X)* is the variance covariance matrix *(C)*. Rousseeuw and Leroy (1987) suggested using Robust Mahalanobis Distance (RMD) as a diagnostic tool for detection of HLPs by replacing the classical mean vector *T(X)*, and classical covariance matrix, *C(X)* of MD_i in Equation 1 by robust estimators such as Minimum Volume Ellipsoid (MVE) or Minimum Covariance Determinant (MCD)

(Rousseeuw & Yohai, 1984), because the former estimators are not robust. They considered observation as HLPs if its corresponding (RMD value) exceeds the cutoff points $\sqrt{x^2_{p,\alpha}}$. Midi et al. (2009) noted that the RMD was not very successful in the identification of HLPs and established Diagnostics Robust Generalized Potential (DRGP) whereby its algorithm consisted of two steps. The suspected HLPs were detected using RMD based on MVE and on the second steps, generalized potentials, denoted as (p_{ii}), were employed to confirm the suspected HLPs. Since the distribution of generalized potentials was intractable, they suggested a confidence bound of cutoff points as follows:

$$cutoff.p_{ii} = median\,(p_{ii}) + 3 * MAD(p_{ii}),$$

where

$$MAD(p_{ii}) = median\{|\,p_{ii} - median(p_{ii})|\}/0.6745.$$

Even though the DRGP is very successful in identifying HLPs, its running time is very slow due to using MVE in the first step. Lim and Midi (2016) improvised the DRGP to speed up the computation of location and scatter estimator by using Index Set Equality. They showed that the DRGP based on ISE was much faster than the DRGP based on MVE.

Index Set Equality (ISE)

Salleh (2013) established Index Set Equality (ISE) where it is an innovation from fast MCD. The following steps illustrate the computation of ISE.

Step 1. Choose arbitrarily *h* observations from a dataset to be included in the subsample denoted as H_{old}, where $h = \frac{n+p+1}{2}$ and *p* is the number of independent variables (Rousseeuw & Driessen, 1999).

Let $I_{old} = \{\pi^{old}_{(1)}, \pi^{old}_{(2)}, \ldots, \pi^{old}_{(h)}\}$ be the index set for H_{old}.

Step 2. Compute the *p*-dimensional mean vector $\bar{T}_{H_{old}}$ and the (*p x p*) covariance matrix of C_{old} from the subset H_{old}.

Step 3. Compute the squared Mahalanobis Distance for each observation, as

$$d^2_{old}(i) = (t_i - \bar{T}_{Hold})'\, C^{-1}_{Hold}(t_i - \bar{T}_{Hold}) \quad \text{for } i = 1, 2, \ldots, n.$$

Step 4. Arrange $d^2{}_{old}(i)$ in increasing order,

$$d^2_{old}(\pi(1)) \le d^2_{old}(\pi(2)) \le \cdots \le d^2_{old}(\pi(n)),$$

where π is permutation equal to {1, 2, ..., n}.

Step 5. The first *h* items that correspond to the smallest $d^2_{old}(i)$ will be placed in set $H_{New} = \{t_{\pi(1)}, t_{\pi(2)}, \ldots, t_{\pi(h)}\}$. Then list the new Index Set, as

$$I_{New} = \{\pi^{New}_{(1)}, \pi^{New}_{(2)}, \ldots, \pi^{New}_{(h)}\}.$$

Step 6. If $I_{New} = I_{old}$. Stop the process, then the location vector $H_{old} = H_{New}$ and covariance matrix $C_{H_{old}} = C_{H_{New}}$, if $I_{New} \neq I_{old}$ let $H_{old} = H_{New}$, then recompute $\bar{T}_{H_{new}}$ and $C_{H_{New}}$, repeat steps 3 to 6, until $I_{New} = I_{old}$. Then the corresponding $\bar{T}_{H_{new}}$ and are the location and scatter estimates for ISE.

Reweighted Fast Consistent and High Breakdown (RFCH)

Olive and Hawkins (2010) developed Reweighted Fast Consistent and High breakdown (RFCH) estimators of location and scatter which was faster than the fast MCD developed by Rousseeuw and Driessen (1999). The attractive feature of RFCH technique is that not only its computation is very fast which is even faster than Fast MCD (Zhang et al., 2012), but it is $\sqrt{n}$ consistent estimators. The RFCH utilizes the $\sqrt{n}$ consistent DGK (Devlin et al., 1981) estimator and high breakdown Median Ball (MB) (Olive & Hawkins, 2008) estimators as attractors. The RFCH algorithms can be summarized as follows:

The DGK Algorithm Steps

Step 1. Compute the *p*-dimensional row vector of location and (*pxp*) the $(T(X), C(X))$ covariance matrix, () of the original data and use it as the initial or starting point $(T_{0,Start}, C_{0,Start})$, for calculating the initial Mahalanobis Distance (Equation 2).

$$MD_{i0,DGK} = \sqrt{(t_i - T_{0,Start})'(C_{0,Start})^{-1}(t_i - T_{0,Start})}, \quad i = 1, 2, ..., n. \qquad [2]$$

Step 2. Sort the $MD_{i0,DGK}$ in increasing order. Then calculate its median, $MED = median(MD_{i0,DGK})$. The observation corresponding to the Mahalanobis Distance less than the median will be in the remaining half dataset (*m* observations), defined as Equation 3

$$\tilde{X}_{1,DGK} = \{X_{jl}: MD_{i0,DGK} \leq MED\},\ j = 1,2, \dots, k,\ l = 1,2, \dots, m, \qquad [3]$$

where *k* is the number of predictor variables.

Step 3. Consider $C_{0,DGK} = C_{0,Start}$ where $C_{0,Start}$ is the original dataset's scatter matrix, then recompute the location and scatter estimators for the $\tilde{X}_{1,DGK}$ dataset to obtain the first attractors $(T_{1,DGK}, C_{1,DGK})$.

Step 4. Stop the process if the diagonal elements of $C_{1,DGK} = C_{0,Start}$,otherwise repeat Steps 1 to 3 until convergence where at convergence the final location and scatter estimates $(T_{K,DGK}, C_{K,DGK})$ is acquired from the $\tilde{X}_{K,DGK}$, where *K* is final step at which convergence takes place.

The MB Algorithm Steps

Step 1. Let an identity matrix be the scatter matrix, denoted as $C = I_p$. Then compute Mahalanobis Distance based on the median vector, median (X) and C as Equation 4:

$$MD_i = \sqrt{(t_i - Med(X))' \ (C)^{-1}(t_i - Med(X))} \ , \ i = 1,2, \dots, n, \quad [4]$$

where $Med(X) = \text{median } (X)$.

Let the median of MD_i be the cut-off point, which is denoted by $Lcut$ (Equation 5),

$Lcut$=median (MD_i), [5]

where $Lcut \neq 0.5$. Determine the $\tilde{X}_0$ for half of the dataset *(m)* whose MD_i is less than or equal to the *Lcut*, such that (Equation 6)

$$\tilde{X}_0 = \{X_{jl} : MD_i \leq Lcut\}, j = 1,2, \dots, k, l = 1,2, \dots, m. \quad [6]$$

Step 2. Compute the *p*-dimensional row vector of location and the *(pxp)* covariance matrix of scatter estimators of the $\tilde{X}_0$ and use it as the initial or starting point $(T_{0,Start}, C_{0,Start})$, for calculating the initial Mahalanobis Distance (Equation 7).

$$MD_{0i,MB} = \sqrt{(t_i - T_{0,Start})' \ (C_{0,Start})^{-1}(t_i - T_{0,Start})} \ , i = 1,2, \dots, n, \quad [7]$$

determined the remaining half dataset by using new cut-off point as Equation 8:

$$\tilde{X}_{1,MB} = \{X_{jl} : MD_{0i,MB} \leq Lcut0\}, j = 1,2, \dots, k, \ \ l = 1,2, \dots, m, \quad [8]$$

where $Lcut0 = \text{median } (MD_{0i,MB})$.

Step 3. Based on the $\tilde{X}_{1,MB}$, calculate the attractor $(T_{1,,MB}, C_{1,MB})$.

Step 4. If the diagonal elements of $C_{1,MB} = C_{0,Start}$ stop the process, otherwise recalculate the $MD_{1,MB}$ based on attractor $(T_{1,MB}, C_{1,MB})$ and iterate the Steps 2 to 3, until the convergence is achieved at final attractor $(T_{K,MB}, C_{K,MB})$ and final remaining set $\tilde{X}_{K,MB}$.

The RFCH Algorithm Steps

The RFCH consists of three steps where in the first step the Fast Consistent and High breakdown (FCH) attractors of Olive and Hawkins (2010) is determined based on the final attractors of DGK and MB estimators that adhere the following rules:

Step 1. The T_{FCH} and C_{FCH} are determined as Equation 9:

$$T_{FCH} = \begin{Bmatrix} T_{K,DGK} \ \ if \ \ \sqrt{|C_{K,DGK}|} < \sqrt{|C_{K,MB}|} \\ T_{K,MB} \qquad\qquad Otherwise \end{Bmatrix}, \quad [9]$$

And Equation 10

$$C_{FCH} = \begin{cases} \dfrac{Med\left(MD_i(T_{K,DGK}, C_{K,DGK})\right)}{x^2{}_{(p,0.5)}} \times C_{K,DGK}, & if \quad \sqrt{|C_{K,DGK}|} < \sqrt{|C_{K,MB}|} \\ \dfrac{Med\left(MD_i(T_{K,MB}, C_{K,MB})\right)}{x^2{}_{(p,0.5)}} \times C_{K,MB}, & Otherwise \end{cases}, \quad [10]$$

where $x^2_{(p,0.5)}$ is chi-square distribution with p degrees of freedom and significance level 0.5. The (T_{FCH}, C^*_{FCH}) are the consistent estimators of the FCH attractors according to Theorem 1 of Olive and Hawkins (2010),

where $C^*_{FCH} = \frac{Med(MD_i(T_{FCH}, C_{FCH}))}{x^2{}_{(p,0.5)}} * C_{FCH}$.

Step 2. Construct a new set of data, $\tilde{X}_{FCH}$ by using the following Equation 11,

$$\tilde{X}_{FCH} = \{X_{jl}: MD_i(T_{FCH}, C^*_{FCH}) \leq x^2{}_{(p,1-\alpha)}\}, \quad [11]$$

$j = 1, 2, \ldots, k,\ l = 1, 2, \ldots, m,$

where $MD_i(T_{FCH}, C^*_{FCH})$ is the Mahalanobis Distance based on the location and scatter of FCH estimators in Step 1. Then compute the location and scatter estimators for the $\tilde{X}_{FCH}$ dataset to obtain the RFCH attractors, $(T_{1,RFCH}, C_{1,RFCH})$. Again, following Theorem 1 of Olive and Hawkins (2010),
$C^*_{1,RFCH}$ is defined as Equation 12

$$C^*_{1,RFCH} = \frac{Med\left(MD_i(T_{1,RFCH}, C_{1,RFCH})\right)}{x^2{}_{(p,0.5)}} * C_{1,RFCH}. \quad [12]$$

Subsequently the Mahalanobis Distance based on is computed and a new set of data is constructed using the following Equation 13;

$$\tilde{X}_{2,RFCH} = \{X_{jl}: MD_i(T_{1,RFCH}, C^*_{1,RFCH}) \leq x^2{}_{(p,1-\alpha)}\}, \quad [13]$$

$$j = 1,2, \ldots, k,\ l = 1,2, \ldots, m.$$

Following the same process, $(T_{2,RFCH}, C_{2,RFCH})$ estimators are calculated based on the $\tilde{X}_{2,RFCH}$ dataset. Afterwards, $C^*_{2,RFCH}$ is defined as in Equation 14 by applying Theorem 1 of Olive and Hawkins (2010),

$$C^*_{2,RFCH} = \frac{Med\left(MD_i(T_{2,RFCH}, C_{2,RFCH})\right)}{x^2{}_{(p,0.5)}} * C_{2,RFCH}. \quad [14]$$

Step 3. Step 1 to 2 is repeated K-times until convergence. Convergence is achieved if the number of detected outliers or HLPs is the same for $MD_i(T_{K,RFCH}, C_{K,RFCH})$ and $MD_i(T_{K-1,RFCH}, C^*_{K-1,RFCH})$.

As stated by Olive and Hawkins (2010), on convergence, the final estimators of RFCH, i.e. $(T_{K,RFCH}, C_{K,RFCH})$ are High Breakdown (HB) $\sqrt{n}$ consistent estimators (see Olive and Hawkins (2010) for description of $\sqrt{n}$ consistent estimator).

The Proposed Diagnostic Robust Generalized Potential based on Reweighted Fast Consistent and High Breakdown Estimators (DRGP(RFCH))

Midi et al. (2009) proposed Diagnostic Robust Generalized Potential based on Minimum Volume Ellipsoid (DRGP(MVE)) for detecting HLPs. The DRGP algorithm comprises two steps where in the first step, Robust Mahalanobis Distance (RMD) based on MVE is used to detect the suspected HLPs and on the second step, the generalized potential is used to confirm whether or not the suspected HLPs is a genuine HLPs. Although the DRGP is proven to be very successful in detecting HLPs, its computation running time is very slow since it uses the location and scatter estimators obtained from the MVE. As such, Lim and Midi (2016) proposed another diagnostic method, Diagnostic Robust Generalized Potential based on Index Set Equality (DRGP(ISE) to identify HLPs by incorporating the location and scatter estimators based on ISE. However, through our investigation, the DRGP (ISE) is not very stable and we anticipate that it still suffers from small percentage of swamping and masking effect. We also expect that the running time of the DRGP (ISE) can be improved. In this regard, we attempt to improvise the existing DRGP by integrating the location and scatter estimators obtained from the Reweighted Fast Consistent and High breakdown (RFCH) estimators (Olive & Hawkins, 2010). The attractive feature of this estimator is that it is High Breakdown $\sqrt{n}$ consistent estimator as noted by Olive and Hawkins (2010). Our improvised DRGP is denoted as DRGP (RFCH).

The Proposed DRGP (RFCH) Technique is Summarized as Follows

Step 1. Identify the suspected HLPs by using RMD_i for each i^{th} observation based on RFCH (Equation 15)

$$RMD_i = \sqrt{(t_i - T_{RFCH})'(C_{RFCH})^{-1}(t_i - T_{RFCH})} \ , i = 1,2, \dots, n. \quad [15]$$

Step 2. As per Midi et al. (2009) the cut-off point is defined as follows;

$$cut\_off = median\ (RMD_i) + 3 * MAD(RMD_i),$$

where

$$M(A(D(RMD_i(= \ (\ (m(edian\ abs\ RMD_i - \ median\ \ RMD_i/0.6745\ .)$$

We declare that any i^{th} case with Robust $RMD_i > cut\_off$ point, is the suspected HLPs and include them in a deletion group, denoted as *D* Group, while the rest of the observations are kept in the *R* group.

Step 3. Following Midi et al. (2009), we employ generalized potential, P_{ii}, to confirm the suspected HLPs whether are not they still can be considered as HLPs (Equation 16).

$$p_{ii} = \begin{cases} z_{ii}^{(-D)} & for\ i \in D \\ \dfrac{z_{ii}^{(-D)}}{1 - z_{ii}^{(-D)}} & for\ i \in R \end{cases} \quad [16]$$

Where (Equation 17)

$$z_{ii}{}^{(-D)} = x_i^T (X_R^T X_R)^{-1} x_i, \quad i = 1,2, \dots, n\ . \quad [17]$$

Step 4. Compute the cut-off point p_{ii}, for i.e. cut-off $p_{ii} = Median(p_{ii}) + 3Q_n(p_{ii})$. Rousseeuw and Croux (1993) defined $Q_n = C\{|x_i - x_j|; i < j\}_{(k)}$ as a pairwise order statistic of whole distance where $k = \binom{h}{2} \approx \binom{n}{2} \approx \binom{n}{2}/4$ and $h = \left[\frac{n}{2}\right] + 1$. Rousseeuw and Croux (1993) noted that to make Qn a consistent estimator for Gaussian data the value of C should be chosen equals to 2.2219.

We declare that all members of the D group as HLPs if P_{ii} > cut-off P_{ii}, otherwise place those observations back into the estimation subset R sequentially begin with the least p_{ii} value.

RESULTS AND DISCUSSION

Monte Carlo Simulation Study

Monte Carlo simulation study was carried out to assess the performance of our proposed DRGP(RFCH) compared with DRGP(MVE) and DRGP(ISE). As per Lim and Midi (2016), we consider a general linear regression model with p explanatory variables as Equation 18

$$Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2}, \dots, B_p X_{ip} + e_i, \quad \mathrm{i} = 1,2, \dots, \mathrm{n}, \quad [18]$$

where each of the explanatory variable is generated from Uniform Distribution (0, 10), e_i is generated from standard normal distribution with varying sample of sizes, n = 20, 40, 60, 80, 100 and 200. We consider various proportion of High Leverage Points (α = 0.05, 0.10 and 0.15) and p = 2and p = 4. For p = 2, we set B_0= 1, B_1 = 2 and B_2 = 3 as the true parameter values and set B_0 = 1, B_1 = 2, B_2 = 3, B_3 = 4 and B_4 = 5 as the true parameter values for p = 4. The HLPs are created by replacing the first 100 α% observations of the original good data for p = 2 and p = 4 with values of X_1 and X_2 and values of X_1, X_2, X_3, and X_4, respectively, generated from Uniform Distribution U (15, 20), without changing their y values. The simulation was repeated 10,000 times.

Table 3 and 4 exhibit the percentage of correct detection, masking, and swamping of HLPs for p = 2 and p = 4. It can be observed from Table 3 and 4 that for p =2, p = 4, n = 20 and at 5% HLPs, the performance of the three methods are fairly closed, but DRGP(ISE)

Table 3

Percentage of Correct detection, Masking and Swamping, p = 2

% of HLPs		% of Correct detection			% of Masking			% of Swamping		
	n	DRGP (MVE)	DRGP (ISE)	DRGP (RFCH)	DRGP (MVE)	DRGP (ISE)	DRGP (RFCH)	DRGP (MVE)	DRGP (ISE)	DRGP (RFCH)
5	20	100	100	100	0	0	0	7.73	6.875	7.726
	40	100	100	100	0	0	0	1.419	1.527	1.193
	60	100	100	100	0	0	0	0.685	0.824	0.617
	80	100	100	100	0	0	0	0.426	0.539	0.389
	100	100	100	100	0	0	0	0.289	0.379	0.281
	200	100	100	100	0	0	0	0.085	0.104	0.084
10	20	99.98	99.89	99.99	0.02	0.11	0.01	3.318	3.145	2.854
	40	100	99.99	100	0	0.01	0	0.847	1.048	0.731
	60	100	100	100	0	0	0	0.372	0.572	0.369
	80	100	100	100	0	0	0	0.206	0.368	0.199
	100	100	100	100	0	0	0	0.135	0.29	0.132
	200	100	100	100	0	0	0	0.028	0.1	0.024
15	20	99.86	99.55	99.85	0.14	0.45	0.15	2.222	2.242	1.935
	40	100	99.84	100	0	0.16	0	0.505	0.685	0.451
	60	100	99.9	100	0	0.1	0	0.187	0.408	0.176
	80	100	99.98	100	0	0.02	0	0.095	0.256	0.079
	100	100	99.95	100	0	0.05	0	0.048	0.209	0.048
	200	100	100	100	0	0	0	0.006	0.071	0.004

Table 4

Percentage of Correct detection, Masking and Swamping, p = 4

% of HLPs		% of Correct detection			% of Masking			% of Swamping		
	n	DRGP (MVE)	DRGP (ISE)	DRGP (RFCH)	DRGP (MVE)	DRGP (ISE)	DRGP (RFCH)	DRGP (MVE)	DRGP (ISE)	DRGP (RFCH)
5	20	100	100	100	0	0	0	7.397	5.348	6.810
	40	100	100	100	0	0	0	2.195	1.821	1.544
	60	100	100	100	0	0	0	1.118	1.011	0.995
	80	100	100	100	0	0	0	0.720	0.700	0.675
	100	100	100	100	0	0	0	0.531	0.534	0.506
	200	100	100	100	0	0	0	0.262	0.271	0.257
10	20	99.99	99.24	99.99	0.01	0.76	0.01	4.789	3.690	4.582
	40	100	100	100	0	0	0	1.362	1.477	1.188
	60	100	100	100	0	0	0	0.623	0.590	0.59
	80	100	100	100	0	0	0	0.401	0.382	0.374
	100	100	100	100	0	0	0	0.274	0.273	0.283
	200	100	100	100	0	0	0	0.117	0.117	0.115
15	20	99.84	89.16	99.92	0.16	10.84	0.08	3.092	2.553	3.306
	40	100	99.71	100	0	0.29	0	0.788	0.707	0.651
	60	100	99.99	100	0	0.01	0	0.337	0.309	0.295
	80	100	100	100	0	0	0	0.201	0.19	0.185
	100	100	100	100	0	0	0	0.132	0.124	0.120
	200	100	100	100	0	0	0	0.039	0.038	0.025

Table 5

Running time by (Seconds) and Average of HLPs Detection of DRGP(MVE), DRGP(ISE) and DRGP(RFCH).

			DRGP(MVE)		DRGP(ISE)		DRGP(RFCH)	
%HLPs	Sample Size	Actual HLPs	Average of HLPs	Running time	Average of HLPs	Running time	Average of HLPs	Running time
5	20	1	2.456	60.81	2.375	40.61	2.545	13.25
	40	2	2.676	100.6	2.611	48.28	2.477	13.34
	60	3	3.411	129.2	3.494	54.48	3.370	13.35
	80	4	4.341	152.9	4.431	60.45	4.311	13.44
	100	5	5.289	178.6	5.379	65.21	5.281	13.51
	200	10	10.17	310.2	10.252	91.28	10.168	13.78
10	20	2	2.663	60.79	2.629	40.45	2.571	13.28
	40	4	4.339	100	4.419	47.51	4.292	13.35
	60	6	6.223	126.3	6.343	53.70	6.221	13.34
	80	8	8.164	152.5	8.290	59.58	8.159	13.39
	100	10	10.135	178.7	10.29	64.33	10.132	13.39
	200	20	20.056	308.2	20.20	89.47	20.048	13.41
15	20	3	3.445	60.72	3.45	40.34	3.387	13.22
	40	6	6.202	100.2	6.274	47.84	6.200	13.25
	60	9	9.112	126.8	9.245	52.95	9.106	13.48
	80	12	12.076	152.4	12.205	58.56	12.062	13.49
	100	15	15.048	178.1	15.209	63.05	15.048	13.58
	200	30	30.01	307.01	30.143	86.95	30.009	13.72

is slightly better than the other two methods in terms of having the smallest swamping effect. However, other than at 5% of HLPs and $n = 20$, the DRGP(RFCH) outperforms other methods regardless of sample size and percentage of HLPs, followed by DRGP(MVE) and DRGP(ISE). In this situation, the DRGP(RFCH) consistently having the highest percentage of correct detection of HLPs and the least percentage of swamping and masking effects. Similar conclusion can be made for $p = 4$ where again the DRGP(RFCH) shows the best result followed by DRGP(MVE) and DRGP(ISE). Let us now focus on the computer running time for our DRGP(RFCH) compared to DRGP(ISE) and DRGP(MVE) as displayed in Table 5 and Figure 1. Table 5 presents the computer running times in seconds, the average HLPs detected by DRGP(RFCH), DRGP(ISE) and DRGP(MVE) and the actual number of HLPs planted in the dataset. At 5% of HLPs, $n = 20$, DRGP(RFCH) slightly over detected the HLPs because the average of HLPs detected is slightly larger than the actual HLPs planted in the data. Nonetheless, the computer running time for the DRGP(RFCH) is much smaller than the DRGP(ISE) and DRGP(MVE). On other scenarios, the average of HLPs detected by DRGP(RFCH) consistently the nearest to the actual HLPs planted in the data, followed by the DRGP(MVE) and DRGP(ISE). It is also interesting to see that the computer running times for the DRGP(RFCH) was consistently the least, followed by the DRGP(ISE) and DRGP(MVE). The results are depicted in Figure 1 for clear and quick visualization. The results for $p = 4$ and greater than $p = 4$ are consistent and not reported here due to space limitation.

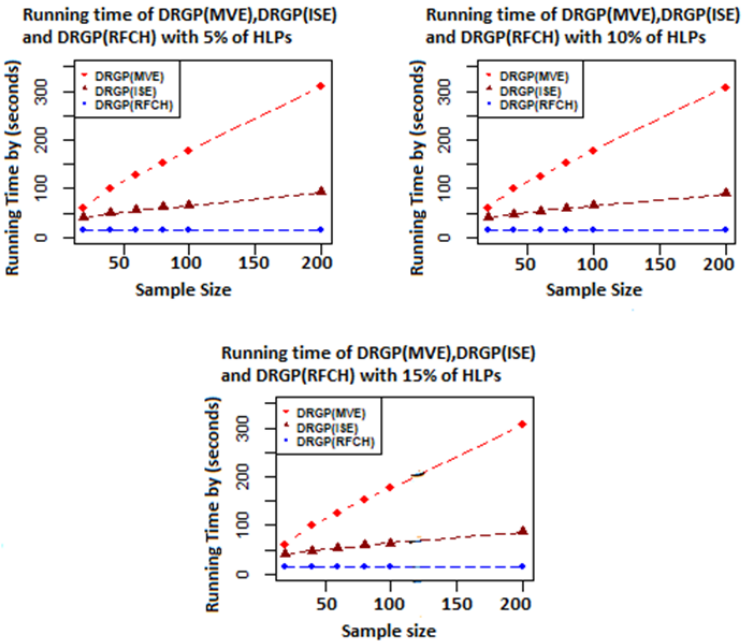


Figure 1. The running time for DRGP(MVE), DRGP(ISE) and DRGP(RFCH), at various sample size

Real Example

A real Hand Grip Strength dataset (Hossain et al., 2012) was used to evaluate the performance of our proposed DRGP(RFCH) method. In this study, we considered a sample of size 196 men, comprising of healthy staff, medical students and visitors of University of Malaya Medical Center between January and April. Four explanatory variables (Age, Height, Weight and BMI) were considered in this study and the dependent variable is the right-hand grip strength. The DRGP(RFCH), DRGP(ISE) and DRGP(MVE) were then applied to the data. The number of HLPs detected by each method is displayed in Figure 2. It is interesting to see from the graph of Figure 2(a) and 2(c) that both DRGP(RFCH) and DRGP(MVE), having the same cut-off points, detected the same observations as HLPs (cases 24, 45, 91, 107, 137, 140, 183). On the other hand, as expected the DRGP(ISE) with cut-off point 0.0615 does not detect the same number of observations. It detects only six observations as HLPs (cases 24, 45, 91, 107, 137, 140, 183) where its masked case 107. The value of DRGP(ISE) which corresponds to case 107 is less than the cut-off point 0.0615. The results of real data are consistent with the results of simulation study where the DRGP(ISE) suffers from swamping and masking effect.

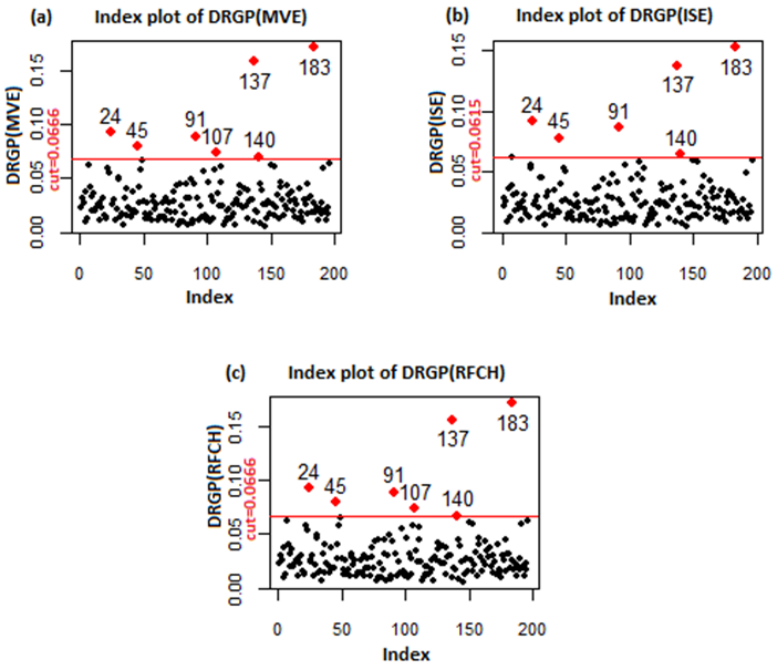


Figure 2. The number of detected HLPs by DRGP(MVE), DRGP(ISE) and DRGP(RFCH)

CONCLUSION

The main aim of this paper is to propose another diagnostic method of detecting HLPs that we call DRGP(RFCH). The existing DRGP(MVE) is quite successful in identifying HLPs but its running time is very slow. The DRGP(ISE) running time is much faster than the DRGP(MVE). However, the DRGP(ISE) was not computationally stable and still possessed masking and swamping effect. Contrarily, the propose DRGP(RFCH) is very successful in detecting HLPs with negligible swamping effect. Moreover, it is based on $\sqrt{n}$ RFCH consistent estimators of location and scatter. The numerical study also signifies that the DRGP(RFCH) needs much lesser computer running time and computationally very stable in the sense of having consistent estimated values. The results of this study appear to recommend that the DRGP(RFCH) may give the most appealing diagnostic method for the identifying HLPs in multiple linear regression model.

ACKNOWLEDGMENTS

This article was partially supported by the Fundamental Research Grant Scheme (FRGS) under Ministry of Education with project number FRGS/1/2019/STG06/UPM/01/1

REFERENCES

Alkenani, A., & Yu, K. (2013). A comparative study for robust canonical correlation methods. *Journal of Statistical Computation and Simulation*, *83*(4), 692-720.

Bagheri, A., Midi, H., & Imon, A. H. M. R. (2012). A novel collinearity-influential observation diagnostic measure based on a group deletion approach. *Communications in Statistics-Simulation and Computation, 41*(8), 1379-1396.

Chatterjee, S., & Hadi, A. S. (2006). *Regression analysis by example*. New Jersey, USA: John Willey & Sons.

Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association*, *76*(374), 354-362.

Hadi, A. S. (1992). A new measure of overall potential influence in linear regression, *Computational Statistics and Data Analysis, 14*(1), 1-27.

Hawkins, D. M, Bradu, D., & Kass, G. V. (1984). Location of several outliers in multiple regression data using elemental sets. *Technometrics 26*(3), 197-208.

Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician, 32*(1), 17-22.

Hossain, M. G., Zyroul, R., Pereira, B. P., & Kamarul, T. (2012). Multiple regression analysis of factors influencing dominant hand grip strength in an adult Malaysian population. *Journal of Hand Surgery (European Volume), 37*(1), 65-70.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics, 1*(5), 799-821.

Imon, A. H. M. R. (2002). Identifying multiple high leverage points in linear regression. *Journal of Statistical Studies*, *3*, 207-218.

Imon, A. H. M. R., & Khan, M. A. I. (2003). A solution to the problem of multicollinearity caused by the presence of multiple high leverage points. *International Journal of Statistical Sciences*, *2*, 37-50.

Lim, H. A., & Midi, H. (2016). Diagnostic robust generalized potential based on index set equality (DRGP (ISE)) for the identification of high leverage points in linear model. *Computational Statistics 3*(31),859-877.

Lopuhaa, H. P. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *Annals of Statistics*, *27*(5), 1638-1665.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *National Institute of Science of India, 2*(1), 49-55.

Midi, H., Ramli, N. M., Imon, A. H. M. R. (2009). The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics, 36* (5),507-520.

Midi, H., Rana, S., & Imon, A. H. M. (2014). Tow-step robust estimator in heteroscedastic regression model in the presence of outliers. *Economic Computation and Economic Cybernetics Studies and Research, 48*(3), 255-272.

Olive, D. J., & Hawkins, D. M. (2008). *High breakdown multivariate estimators*. Retrieved September 5, 2019, from https://www.researchgate.net/profile/David_Olive2/publication/240737720_High_Breakdown_Multivariate_Estimators/links/0a85e53234b7db7f90000000.pdf.

Olive, D. J., & Hawkins, D. M. (2010). *Robust multivariate location and dispersion*. Retrieved September 5, 2019, from https://www.researchgate.net/profile/David_Olive2/publication/228434748_Robust_multivariate_location_and_dispersion/links/02bfe51015be5c88ca000000.pdf.

Peña, D., & Yohai, V. J. (1995). The detection of influential subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 145-156.

Riazoshams, H., Midi, H., & Sharipov, O. S. (2010). The performance of robust two-stage estimator in nonlinear regression with autocorrelated error. *Communications in Statistics-Simulation and Computation*, *39*(6), 1251-1268.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications, 37*(8), 283-297.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, *88*(424), 1273-1283.

Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics, 41*(3), 212-223.

Rousseeuw, P., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York, USA: Wiley Series in Probability and Mathematical Statistics.

Rousseeuw, P., & Yohai, V. (1984). Robust regression by means of S-estimators. In J. Franke, W. Härdle & D. Martin (Eds.), *Robust and nonlinear time series analysis* (pp. 256-272). New York, NY: Springer.

Salleh, R. (2013). *A robust estimation method of location and scale with application in monitoring process variability* (Doctoral dissertation). Universiti Teknologi Malaysia, Malaysia.

Uraibi, H. S., Midi, H., & Rana, S. (2017). Selective overview of forward selection in terms of robust correlations. *Communications in Statistics-Simulation and Computation, 46*(7), 5479-5503.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, *48*, 817-838.

Zhang, J., Olive, D., & Ye, P. (2012). Robust covariance matrix estimation with canonical correlation analysis. *International Journal of Statistics and Probability, 1*(2), 119-136.

Pertanika J. Sci. & Technol. 28 (4): 1221 - 1265 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

On the Formulation of Metaheuristic Algorithm-Based Approximation Approach for Nonlinear Ordinary Differential Equations with Application to Heat Exchanger Problem

Ee Soong Low and Pauline Ong*

Faculty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), 86400 Parit Raja, Batu Pahat, Johor, Malaysia

ABSTRACT

The problems that arise in multitudinous fields often involve solving complex nonlinear ordinary differential equations (ODE), and it remains challenging since the actual solutions to these problems are hard to obtain. In this regard, the solution strategy with the formulation of Fourier series expansion, calculus of variation and metaheuristic algorithm, was introduced to determine the approximate solution of the nonlinear ODE. The nonlinear ODE was formulated as an optimization problem, specifically, the moth-flame optimization (MFO) algorithm and flower pollination algorithm (FPA) were utilized to find the coefficients of the Fourier series. This paper aimed to determine the feasibility of the proposed method to solve the ODEs with different characteristics and compare the obtained results with other optimization algorithms. Moreover, the suitable number of terms (NT) of Fourier series were determined for different test problems for MFO and FPA. The quantitative analysis in terms of the generational distance (GD) metric demonstrated that the approximate solutions were reasonably accurate, with the low GD within the range of 1E-03 to 1E-05 for all test problems. The comparative analysis showed that the approximate performances of MFO and FPA were superior to or comparable with the genetic algorithm, particle swarm optimization and water cycle algorithm.

ARTICLE INFO

Article history:
Received: 6 May 2020
Accepted: 5 June 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.06

Keywords: Flower pollination algorithm, Fourier series, metaheuristic algorithm, moth-flame optimization algorithm, ordinary differential equations

E-mail addresses:
ongp@uthm.edu.my (Pauline Ong)
low_es@hotmail.com (Ee Soong Low)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Ordinary differential equations (ODE) have found a widespread application in modelling various dynamic systems, for instance, the membrane distillation process (Perfilov et al., 2018) and Human Immunodeficiency Virus model (Tian et al., 2019), to name a few. A *n*-th order nonlinear ODE can be expressed as Equation 1:

$$F(x, y, y', y'', ..., y^{(n)}) = 0 \quad (1)$$

where y is a function of a single independent variable x and $y^{(n)}$ is the *n*-th order derivative with respect to x .

Solving the nonlinear ODE is a challenging task, especially using analytical approaches, in which the solutions of the complex nonlinear ODE may be too complicated to obtain. Therefore, numerical methods, such as Euler and Runge-Kutta, are applied to determine the approximate solutions of the nonlinear ODE. Numerical methods, however, have their operational limitations. For example, Euler's method is numerically unstable, especially for stiff equations (Lambert, 1986). A finer discretization is required for smaller approximation error, and unavoidably, leads to higher computation load (Hussain et al., 2016). Moreover, most of the numerical methods have large cumulative errors due to the rounding-off error, which increases quickly with finer discretization (Butcher, 2016).

In addition to numerical methods, variant of analytical methods, exemplarily, variational iteration method (VIM) (He, 1999), differential transformation method (DTM) (Hussin et al., 2016), homotopy perturbation method (HPM) (He, 2003), and Adomian decomposition method (ADM) (Adomian, 1988), have been utilized to determine the approximate solutions of the ODE. Each of these methods has its limitations, albeit it gains prominence in solving the modeling problems in various fields (Das & Kundu, 2019; El-Sayed & El-Mongy, 2019; Turkyilmazoglu, 2018).

VIM – a modified Lagrange multiplier method based on the variational theory, was initially introduced by He to solve the wave equation (He, 1999). This method is of high calculation complexity due to the repetitive calculation of unnecessary terms (Abassy et al., 2007) and subjected to inherent inaccuracy in solving the Lagrange multiplier (Mohammadi & Hosseini, 2011). DTM, though based on the Taylor series, is distinct in the sense that it does not require the explicit calculation of high-order derivatives as in the Taylor series, which may not always formidable (Chen & Liu, 1998). The DTM, instead, iteratively calculates the derivatives using the differential transformation, but some drawbacks arise, such as the inefficiency in handling the oscillatory systems (Ebaid, 2011) and differential equation with complex nonlinearities (Ebaid, 2013). HPM is an improved classic perturbation method based on the idea of homotopy in topology (He, 2003). The HPM is limited in the way that the initial parameter has to be properly assigned to achieve a good approximation (Babolian et al., 2009). The salient feature of the ADM lies in its ability

to solve nonlinear differential equations without system linearization due to the utilization of the Adomian polynomials (Adomian, 1988). This method is somehow computationally complicated due to the calculation of the Adomian polynomials (Mohammadi & Hosseini, 2011), and proned to divergence when solving ODEs of Emden–Fowler type (Biazar & Hosseini, 2017).

In heat exchanger problems, finned surfaces are often used to increase the rate of heat transfer. In one-dimensional fin analysis, the energy transfer process can be modelled as an ODE if there is no heat variation in the fin cross-section. Otherwise, a partial differential equation will be formed (Torabi et al., 2013). The study of thermal conductivity of fins has been conducted by many researchers. Solving the nonlinear fin problems using numerical approach was studied by Bouaziz et al. (2001). In the analysis of thermal conductivity of convective radial and straight fins, VIM was used and compared with finite element analysis (Coşkun & Atay, 2007; Coşkun & Atay, 2008). Straight fins with different heat transfer coefficients problems were approximately solved using the VIM (Fouladi et al., 2010). To assess the temperature distribution of fins, Domairry and Fazeli (2009) applied the homotopy analysis method (HAM) to solve the differential equations formed by the nonlinear straight fins. Besides, HAM was used to find the approximate solutions for fin efficiency and temperature distribution in the study conducted by Khani et al. (2009). Other than rectangular fins, concave parabolic and trapezoidal fins were studied for space and airborne applications (Torabi et al., 2013).

The limitations of numerical methods and analytical methods prompt the investigation on the suitability to apply the metaheuristic optimization technique to solve the nonlinear ODE. Literature review reveals the credibility of using genetic algorithm (GA) (Mateescu, 2006), particle swarm optimization (PSO) (Babaei, 2013; Nemati et al., 2015), water cycle algorithm (WCA) (Sadollah et al., 2015a), harmony search algorithm (Sadollah et al., 2017), imperialist competitive algorithm (Almasi et al., 2016) and to the more recent, orthogonal colliding bodies optimization (OCBO) (Panda & Pani, 2017) in this regard.

Solving the ODE using the metaheuristic approach, generally, requires the formulation of the ODE as an optimization problem. A base approximate function is used as the approximate solution of the ODE. Subsequently, a population of self-organized agents collaboratively works to find the approximate solution of the ODE based on some rules. Various functions can be applied as the base function of the approximate solution, such as Runge-Kutta method (Mateescu, 2006), method of bilaterally bounded (Lee, 2006), finite element method (Mastorakis, 2006), constrained integration approach (Rudd & Ferrari, 2015), collocation method (Beidokhti & Malek, 2009) and Fourier series (Babaei, 2013). In this study, the Fourier series is chosen as it can decompose any periodic function into the sum of sine and cosine functions. Hence, it can approximate any form of ODE using the same procedure. Moreover, in comparison with the *n*-th order polynomial base function

which can only be differentiated up to *n*-th times, the Fourier series can be infinitely differentiated since it only involves the sine and cosine functions.

The Fourier series is made up of the sum of sine and cosine functions along with the unknown coefficients. In this study, the unknown coefficients of the Fourier series are optimized using the moth-flame optimization (MFO) algorithm and flower pollination algorithm (FPA), in which minimizing the weight error function of the ODE is formed as the objective function of the metaheuristic algorithms. The obtained approximate solutions in solving the well-known ODEs and the longitudinal heat transfer fins problem using the MFO and FPA are compared with the results from other optimizers, namely, PSO, WCA, GA and OCBO, in terms of the generational distance (GD) metric and mean squared error (MSE).

The paper is organized as follows. A brief introduction of the metaheuristic algorithms used in this study – MFO and FPA, is presented in metaheuristic algorithma section. Methods section 3 describes the solution strategy of formulating the approximate solution of ODE using the Fourier series. In results and discussion section, the experimental simulations on solving the well-known ODEs and the longitudinal heat transfer fins problem are presented. The relationship of using the different number of terms in Fourier series to the approximation accuracy is also analyzed. Lastly, some conclusions are drawn.

METAHEURISTIC ALGORITHM

Flower Pollination Algorithm

The flower pollination process inspires the development of FPA (Yang, 2012). There exist two types of pollination process: self-pollination, which does not require a pollinator to transfer the pollen; and cross-pollination, which needs the pollinator such as bees and birds to travel for a long-distance while obeying the Lévy flight behaviour (Dash et al., 2016). To simplify the search process in the FPA, the following four rules are assumed:

(i) The cross-pollination corresponds to the global search using Lévy flight, which is modelled as Equation 2:

$$x_i^{t+1} = x_i^t + \gamma L(\lambda)\left(x_i^t - g_*\right) \tag{2}$$

where x_i^t is pollen *i* at iteration t, g_* is the best solution at current iteration, γ is a scaling factor to control the step size and $L(\lambda)$ is a parameter corresponding to the strength of pollination. The Lévy distribution is described as Equation 3:

$$L \sim \frac{\lambda\Gamma(\lambda)\sin(\pi\lambda/2)}{\pi}\frac{1}{s^{1+\lambda}}, (s \gg s_o > 0) \tag{3}$$

where $\Gamma(\lambda)$ is a standard gamma function.

(ii) The self-pollination acts as the local search, which may involve the same flower or different flowers of the same plant. This exploitation process is expressed as Equation 4:

$$x_i^{t+1} = x_i^t + \varepsilon\left(x_j^t - x_k^t\right) \tag{4}$$

where x_j^t and ε are pollens from different flowers and ε is a random number from uniform distribution [0,1].

(iii) The degree of similarity between two flowers is proportional to the reproduction ratio, which is considered as the flower constancy.

(iv) The switching probability $p \in (0,1)$ is used to control the probability of local and global search.

The pseudocode of FPA is summarized in Figure 1.

Flower Pollination Algorithm

Determine the objective function, $f(x), x = (x_1, x_2, ..., x_d)$

Define population size, n and switch probability, p

Define maximum iteration, *max_iter*

Initialize a population of n flowers randomly

Identify the global best solution, g_*

Define iteration, $t = 1$

while ($t <$ *max_iteration*)

  for i = 1:n (all n flowers in the population)

    if rand > p

      Draw a (d-dimensional) step vector L which obeys a Lévy distribution

      Do global search using $x_i^{t+1} = x_i^t + L\left(x_i^t - g_*\right)$

    else

      Pick ε between 0 and 1 randomly

      Randomly choose j and k among all the solutions

      Do local search using $x_i^{t+1} = x_i^t + \varepsilon\left(x_j^t - x_k^t\right)$

    end if

    Evaluate fitness of new solution (fmin)

    If fmin is better than previous solution, update the i solution and fmin in the population

    If fmin is better than global best solution, update global best solution and its fitness

  end for

  Record the current global best solution

  $t = t + 1$

end while

Figure 1. The pseudocode of FPA

Moth Flame Optimization

The moth performs a spiral path around a light source in the night. Inspired by the transverse orientation of such navigation method, the MFO was proposed (Mirjalili, 2015).

In the MFO, the moths are the search agents (candidates of the solutions), the positions of moths represent the variables of the underlying problem, while the flame is the best position found by the moth. Through constantly changing the position vector, the moths can fly in multiple dimensional spaces. After the best position is identified, it acts as a flame and moth will start searching around the flame to search for a better solution.

The MFO algorithm consists of three main functions: I, P and T. The function I is used to generate a random position within the search space. The function P moves the moths in a spiral pattern, while the function T terminates the operation when the stopping conditions are achieved.

The logarithmic spiral within the function P is modelled as Equation 5:

$$S\left(M_i, F_j\right) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j \quad (5)$$

where D_i is the distance between the *i*-th moth and the j -th flame, *b* is a constant used to define the shape of the logarithmic spiral and *t* is a random number within the range [-1,1]. The spiral or the motion path of the moth should start from the initial position of a moth and end at the position of a flame. The distance between both flame and moth can be determined by manipulating the parameter *t* in Equation 5, where *t* varies from -1 (closest) to 1 (farthest).

For each iteration, the flame list is sorted from the best to the worst solution. The moth is confined to update its position by referring to only one of the flames. This is to avoid the solution from getting trapped in the local optima. Besides, to preserve better convergence towards the end of the search process, the number of flames decreases as the number of iteration increases, according to Equation 6:

$$Flame\_no = round\left(N - I * \frac{N-1}{T}\right) \quad (6)$$

Here, *I* is the current iteration number, N is the maximum number of flames and T is the maximum number of iterations. The flow chart of MFO algorithm is illustrated in Figure 2.

METHODS

Solving the Ordinary Differential Equation using Fourier Series

Consider an ODE of order *n* as in Equation 1 with initial conditions (Equation 7):

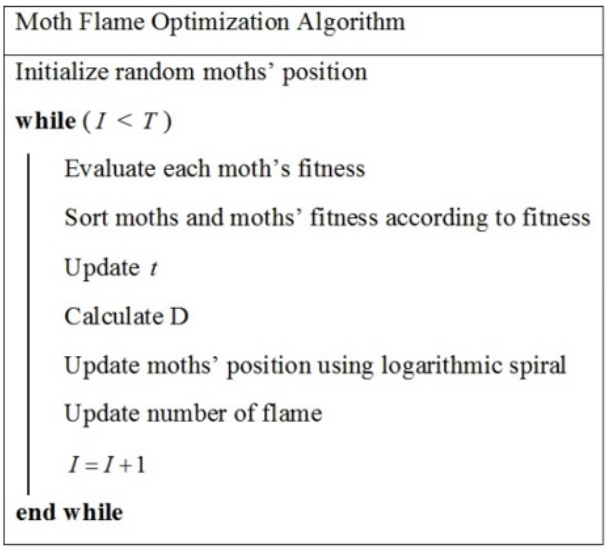


Figure 2. The pseudocode of MFO

$$\begin{aligned} y(x_o) &= y_o \\ y'(x_0) &= y_o' \\ &\vdots \\ y^{(n)}(x_o) &= y_o{}^{(n)} \end{aligned} \tag{7}$$

or boundary conditions (Equation 8):

$$\begin{aligned} y(x_o) = y_o; y(x_n) &= y_n \\ y'(x_0) = y_o'; y'(x_n) &= y_n{}' \\ &\vdots \\ y^{(n)}(x_o) = y_o{}^{(n)}; y^{(n)}(x_n) &= y_n{}^{(n)} \end{aligned} \tag{8}$$

Let the approximate solution of the ODE represented by the Fourier series and centered at x_0 to be expressed in the general form of in Equation 9:

$$y(x) \approx Y_{app}(x) = a_o + \sum_{m=1}^{NT}[a_m \cos(\frac{m\pi(x-x_o)}{L}) + b_m \sin(\frac{m\pi(x-x_o)}{L})] \tag{9}$$

where a_o, a_m and b_m are the Fourier coefficients, and NT is number of terms. *L* represents the length of interval span x_0 to x_n.

The derivatives of Equation 9 are thus given by Equation 10:

$$\begin{aligned} y'(x) &\approx Y_{app}\,'(x) \\ &= \sum_{m=1}^{NT}\left[-\frac{m\pi}{L}a_m \sin\left(\frac{m\pi(x-x_o)}{L}\right)+\frac{m\pi}{L}b_m \cos\left(\frac{m\pi(x-x_o)}{L}\right)\right] \\ y''(x) &\approx Y_{app}\,''(x) \\ &= \sum_{m=1}^{NT}\left[-\left(\frac{m\pi}{L}\right)^2 a_m \cos\left(\frac{m\pi(x-x_o)}{L}\right)-\left(\frac{m\pi}{L}\right)^2 b_m \sin\left(\frac{m\pi(x-x_o)}{L}\right)\right] \end{aligned} \tag{10}$$

The approximate solution (Equation 9) and its derivatives (Equation 10) are then substituted into Equation 1, creating an error function of $E(x)$ as Equation 11:

$$E(x) = F(x, Y_{app}, Y_{app}\,', Y_{app}\,'', ..., Y_{app}{}^{(n)}) \tag{11}$$

Subsequently, the weighted error function (WEF) given as Equation 12:

$$WEF = \int_D |W(x)| \times |E(x)|\, dx \tag{12}$$

is formulated, where $W(x)$ is named as the weight function. The WEF is then utilized as the objective function of the metaheuristic algorithm, in which the metaheuristic algorithm attempts to find the optimal combination of Fourier coefficients such that the error is minimized. In this study, the evaluation of integration in Equation 12 is solved using the Trapezoidal integration method. The weight function is assumed to be 1 for simplicity. Intuitively, the smaller the value of WEF, the better the approximation accuracy is.

The approximate solution should satisfy the ODE on one hand, and the initial or boundary conditions on another hand. Since solving the ODE using the metaheuristic algorithm is formulated as an optimization problem, the initial or boundary conditions are formed as the constraints of the optimization problem. For homogeneous conditions, the constraints for initial conditions are expressed as Equation 13:

$$\begin{aligned} &y(x_0) = 0 \Rightarrow g_1(x_0) = |\, Y_{app}(x_0)\,| \\ &y'(x_0) = 0 \Rightarrow g_2(x_0) = |\, Y_{app}\,'(x_0)\,| \\ &\vdots \\ &y^{(n)}(x_0) = 0 \Rightarrow g_{n+1}(x_0) = |\, Y_{app}{}^{(n)}(x_0)\,| \end{aligned} \tag{13}$$

or the constraints for boundary conditions are expressed as Equation 14:

$$\begin{aligned} &y(x_0) = 0 \Rightarrow g_1(x_0) = |\, Y_{app}(x_0)\,|, y(x_n) = 0 \Rightarrow g_1(x_n) = |\, Y_{app}(x_n)\,| \\ &y'(x_0) = 0 \Rightarrow g_2(x_0) = |\, Y_{app}\,'(x_0)\,|, y'(x_n) = 0 \Rightarrow g_2(x_n) = |\, Y_{app}\,'(x_n)\,| \\ &\vdots \\ &y^{(n)}(x_0) = 0 \Rightarrow g_{n+1}(x_0) = |\, Y_{app}{}^{(n)}(x_0)\,|, y^{(n)}(x_n) = 0 \Rightarrow g_{n+1}(x_n) = |\, Y_{app}{}^{(n)}(x_n)\,| \end{aligned} \tag{14}$$

For nonhomogeneous conditions, the constraints for initial conditions are written as Equation 15:

$$\begin{aligned} &y(x_0) = y_0 \Rightarrow g_1(x_0) = \mid Y_{app}(x_0) - y_0 \mid \\ &y'(x_0) = y_0' \Rightarrow g_2(x_0) = \mid Y_{app}'(x_0) - y_0' \mid \\ &\vdots \\ &y^{(n)}(x_0) = y_0^{(n)} \Rightarrow g_{n+1}(x_0) = \mid Y_{app}^{(n)}(x_0) - y_0^{(n)} \mid \end{aligned} \tag{15}$$

or the constraints for boundary conditions are written as Equation 16:

$$\begin{aligned} &y(x_0) = y_0 \Rightarrow g_1(x_0) = \mid Y_{app}(x_0) - y_0 \mid, \\ &y(x_n) = y_n \Rightarrow g_1(x_n) = \mid Y_{app}(x_n) - y_n \mid \\ &y'(x_0) = y_0' \Rightarrow g_2(x_0) = \mid Y_{app}'(x_0) - y_0' \mid, \\ &y'(x_n) = y_n' \Rightarrow g_2(x_n) = \mid Y_{app}'(x_n) - y_n' \mid \\ &\vdots \\ &y^{(n)}(x_0) = y_0^{(n)} \Rightarrow g_{n+1}(x_0) = \mid Y_{app}^{(n)}(x_0) - y_0^{(n)} \mid, \\ &y^{(n)}(x_n) = y_n^{(n)} \Rightarrow g_{n+1}(x_n) = \mid Y_{app}^{(n)}(x_n) - y_n^{(n)} \mid \end{aligned} \tag{16}$$

The penalty approach is employed for constraint handling in this study. The penalty function value (PFV) of the violated solution is multiplied with a penalty coefficient, λ, and added to the WEF as a penalty function. The fitness function value (FFV) of the optimization problem is thus given as Equation 17 and 18:

$$FFV = WEF + PFV \tag{17}$$

$$PFV = \lambda \sum_{m=1}^{nIC+nBC} g_m^2 \tag{18}$$

where *nIC* and *nBC* are number of initial conditions (IC) and boundary conditions (BC), respectively. The penalty coefficient, λ, is assigned as 10 for all considered ODEs in this study.

Problem Solution Strategy

The study aims to solve the nonlinear ODE using Fourier series approximation approach, in which the metaheuristic algorithm is used to optimize the Fourier coefficients such that the FFV is minimized. Hence, solving the ODE can be formulated as Equation 19:

Minimize $FFV = WRF + PFV$ (19)

subjected to the homogeneous and/or nonhomogeneous initial and/or boundary constraints as in Equation 20 and 21:

$$g_i^{IVP}(x_0) = \begin{cases} 0, & i = 1,2,...,nIC \\ y_0, & i = 1,2,...,nIC \end{cases} \quad (20)$$

$$g_i^{BVP}(x_n) = \begin{cases} 0, & i = 1,2,...,nBC \\ y_n, & i = 1,2,...,nBC \end{cases} \quad (21)$$

This proposed solution strategy is implemented as follows:

Step 1: Rewrite the ODE in an implicit form, as in Equation 1.

Step 2: Rewrite the initial and/or boundary condition in the form of Equation 13 and 15.

Step 3: Select an appropriate value of NT for Fourier series, as in Equation 9.

Step 4: Assign a variable to each Fourier coefficients.

Step 5: Apply the metaheuristic algorithm as the optimizer to determine the optimal combination of Fourier coefficients in Equation 9 such that the FFV in Equation 17 is minimized.

Step 6: Calculate the MSE and GD metric to assess the fitness of the approximate solution in comparison with the exact solution. The GD metric is given as Equation 22 (Sadollah et al., 2015c):

$$GD = \frac{\left(\sum_{i=1}^{n} d_i^2\right)^2}{n} \quad (22)$$

where n is the number of points and d is the Euclidean distance between point i in the approximate solution and the nearest point in the exact solution. In this study, n is chosen as 200 for all case studies.

The flowchart of the proposed strategy in solving the nonlinear ODE using the Fourier series and metaheuristic algorithm is summarized in Figure 3.

RESULTS AND DISCUSSION

Numerical Simulation

In this section, the well-known ODEs with different characteristics were used to evaluate the applicability of the approximate solutions obtained by the FPA and MFO algorithm. Specifically, test problems 1 to 3 are the well-known ODEs in science and engineering, test problem 4 is an ODE with Neumann boundary condition, while two mechanical vibration problems (mass-spring system) without (test problem 5) and with damping (test problem 6) were used to evaluate the effectiveness of the proposed approach as well. Also, the ODE of the longitudinal fins was studied in test problem 7. The illustrative comparison between the obtained solution and the actual solution was presented in addition to the obtained GD metric and MSE. Additionally, the relationship of using different NTs to the approximation accuracy was explored.

All simulations were repeated for 30 runs. For each run, the simulation was continued until the maximum number of iteration was reached. Next, the best results of all algorithms among 30 runs were reported. Except for test problem 2, the lower and upper bounds of the unknown Fourier coefficients were set between -1 and 1, while the search space of the unknown Fourier coefficients in test problem 2 was restricted to -2 and 2.

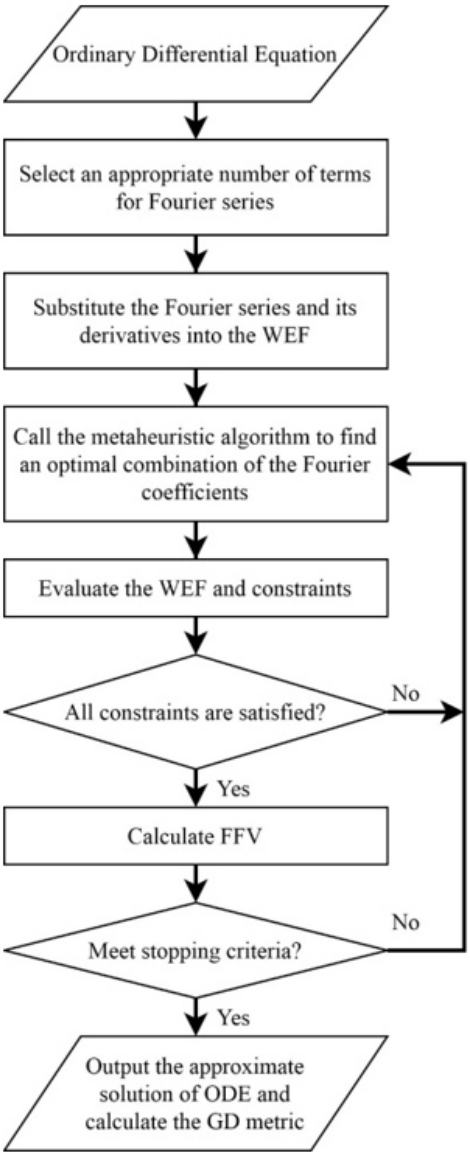


Figure 3. The flowchart of the proposed solution strategy

Performance Comparison between Different Optimization Algorithms

In this study, the FPA, MFO and GA were used to optimize the coefficients of the Fourier series, with the WEF formulated as the cost function of the optimization algorithm. Since finding the approximate solutions of the aforementioned test problems using PSO, WCA and OCBO has been conducted in (Babaei, 2013; Panda & Pani, 2018; Sadollah et al., 2015b; Sadollah et al., 2015d) and hence, the results were included for comparative purpose. The NT of Fourier series in Equation 9 was assigned in accordance with the reported studies. Table 1 summarizes the NT, population size (n) and the maximum number of iteration used in all optimization algorithms. The value was left as NA if the corresponding information was not available (NA) in the respective study. Table 2 lists the parameters setting used in all metaheuristic algorithms. It is pertinent to note that the parameters of OCBO are not provided in this table as they are not stated in the respective study (Panda & Pani, 2018).

Table 1

Initial parameter of all optimization algorithms for comparative analysis

Test Problem	NT	WCA		PSO		OCBO		FPA		MFO		GA	
		n	Max Iter	n	Max Iter	n	Max Iter	n	Max Iter	n	Max Iter	n	Max Iter
1	6	50	500	200	1000	50	500	50	700	300	1000	50	700
2	3	200	200	200	200	200	200	200	500	500	500	200	500
3	3	50	500	300	1000	50	500	50	800	50	500	50	800
4	6	200	100	NA	NA	NA	NA	200	500	200	300	200	500
5	6	NA	NA	200	200	NA	NA	200	500	200	400	200	500
6	8	NA	NA	200	200	NA	NA	200	1000	200	600	200	1000
7	3	NA	NA	300	1000	NA	NA	50	1000	50	1000	50	1000

Table 2

The parameter setting used in all optimization algorithms

Algorithm	Parameter
WCA	Number of sea and river, $N_{sr} = 8$
PSO	Inertia weight, $\omega = 0.9$
	Cognition component, $c_1 = 0.5$
	Social component, $c_2 = 1.5$
FPA	Switching probability, $p = 0.5$

Table 2 *(Continued)*

Algorithm	Parameter
FPA	λ = 1.5
MFO	b = 1
GA	Crossover percentage, pc = 0.7
	Mutation percentage, pm = 0.3
	Mutation rate, mu = 0.1
	Random selection
Euler	Number of steps, N = 200

The performance indicator of GD metric, which has been used in several studies to evaluate the optimization performance of each algorithm quantitatively, was utilized in this study (Panda & Pani, 2018; Sadollah et al., 2015b; Sadollah et al., 2015c). The GD measures the total Euclidean distance between the approximate solution and the exact solution. The approximate and the exact solutions were divided into 200 points and the Euclidean distance between all the points of the approximate solution and the exact solution was calculated. The lower the GD, the closer the approximate solution to the exact solution. Besides GD metric, the MSE was also used to evaluate the performance of all optimization algorithms.

Test Problem 1: Integro-Differential Equation. The integro-differential equation is given by Equation 23 (Sadollah et al., 2015c):

$$y' + 2y + 5\int_0^x y(t)\,dt = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{23}$$

with initial condition $y(0) = 0$. The solution interval of x is chosen from 0 to π. The exact solution of this Initial Value Problem (IVP) is given by Equation 24:

$$y(x) = \frac{1}{2} e^{-x} \sin(2x) \tag{24}$$

The NT of the Fourier series was fixed as 6. Hence, the WCA, PSO, OCBO, GA, MFO and FPA optimization approaches have to determine the best combination of a_0, a_1, b_1, a_2, b_2, a_3, a_4, a_4, b_4, a_5, b_5, a_6 and b_6 of Equation 25, such that the residual error is minimized.

$$\begin{aligned} Y_{app}(x) = {} & a_o + a_1 \cos(x) + b_1 \sin(x) + a_2 \cos(2x) + b_2 \sin(2x) \\ & + a_3 \cos(3x) + b_3 \sin(3x) + a_4 \cos(4x) + b_4 \sin(4x) \\ & + a_5 \cos(5x) + b_5 \sin(5x) + a_6 \cos(6x) + b_6 \sin(6x) \end{aligned} \tag{25}$$

Table 3 summarizes the best approximate solutions obtained by the WCA, PSO, OCBO, GA, MFO and FPA, while Figure 4 presents the comparison between the actual solution and the obtained best approximate solution of each algorithm. As shown in Figure 4, all the approximate solutions were close to the exact solutions. However, by visual inspection, the approximate solutions of MFO, FPA and GA deviate from the actual curve from $x = 2.5$ onward. Table 4 summarizes the values of the GD metric for all the considered algorithms. It can be observed that Euler's method can approximate the solution more accurately than the others, indicated by the lowest GD, while the GA gave the lowest accuracy. The same finding was reported in Table 5, where the performance evaluation in terms of MSE was made.

Table 3

Test Problem 1 - The best approximate solution obtained by various algorithms

Fourier Coefficient	Algorithm					
	WCA	PSO	OCBO	MFO	FPA	GA
a_0	1.9996E-02	3.74E-02	1.9998E-02	5.3744E-01	-1.8435E-01	1.0338E-01
a_1	1.1002E-02	4.67E-03	1.1000E-02	1.2014E-01	-3.6568E-01	2.8827E-01
b_1	3.8363E-02	5.31E-03	3.8360E-02	-1	3.5412E-01	-1.7517E-01
a_2	-7.9942E-04	-3.87E-02	-7.9942E-04	-1	1.0558E-01	-3.1155E-01
b_2	1.1888E-01	1.36E-01	1.1888E-01	-2.2236E-02	6.6087E-01	-2.5309E-01
a_3	-3.6247E-02	-1.49E-02	-3.6247E-02	-1.2050E-01	4.2466E-01	-2.9862E-01
b_3	4.7980E-02	8.24E-02	4.7975E-02	8.7854E-01	1.1973E-01	3.8416E-01
a_4	-4.2281E-02	-1.11E-02	-4.2281E-02	4.9884E-01	7.9053E-02	2.2051E-01
b_4	6.5364E-02	4.14E-02	6.5365E-02	8.3412E-02	-1.9287E-01	1.5877E-01
a_5	2.7610E-02	1.21E-02	2.7610E-02	1.5567E-02	-6.0481E-02	2.6291E-02
b_5	4.9891E-02	2.81E-02	4.9890E-02	-1.9518E-01	-3.0025E-02	-8.6205E-02
a_6	2.0754E-02	1.06E-02	2.0755E-02	-4.1044E-02	-4.5552E-03	-1.7893E-02
b_6	-4.1469E-03	1.01E-03	-4.1470E-03	4.6750E-03	5.7165E-03	7.5348E-03

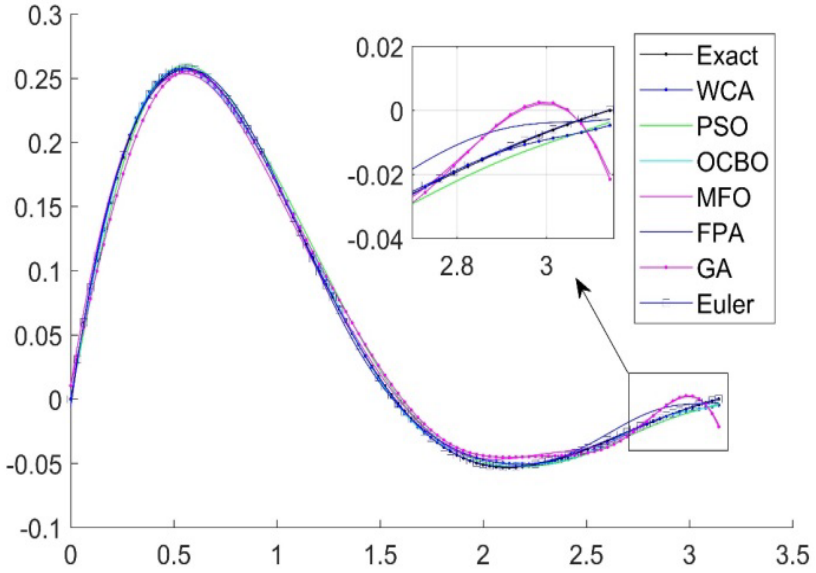


Figure 4. Test Problem 1 – Comparison of the best approximate solutions with the exact solution for WCA, PSO, OCBO, MFO, FPA, GA and Euler

Table 4

The obtained best GD metric of all algorithms for all test problems

Test Problem	Best GD							Best	Worst
	WCA	PSO	MFO	FPA	OCBO	GA	Euler		
1	1.63E-04	3.29E-04	3.66E-04	2.58E-04	1.63E-04	4.76E-04	1.11E-04	Euler	GA
2	5.17E-03	5.15E-03	3.77E-04	1.06E-03	5.15E-04	1.41E-03	8.84E-04	MFO	WCA
3	1.53E-05	2.16E-05	1.47E-05	1.08E-05	6.74E-06	2.40E-04	4.58E-03	OCBO	Euler
4	1.09E-03	-	5.15E-04	8.18E-04	-	2.70E-04	4.45E-03	GA	Euler
5	-	1.77E-03	5.48E-04	1.10E-03	-	5.70E-04	1.37E-02	MFO	Euler
6	-	1.16E-04	4.74E-04	4.77E-04	-	3.02E-04	5.27E-04	PSO	Euler
7	-	3.95e−04	2.81E-04	4.27E-04	-	2.66E-04	4.40E-05	Euler	FPA

Note: There were some blank spaces due to some best approximate solutions were not provided in the respective references

Table 5

The obtained best MSE of all algorithms for all test problems

Test Problem	Best MSE							Best	Worst
	WCA	PSO	MFO	FPA	OCBO	GA	Euler		
1	5.32E-06	2.17E-05	2.67E-05	1.33E-05	5.32E-06	4.53E-05	2.44E-06	Euler	GA
2	2.04E-04	6.94E-03	1.22E-04	4.12E-03	2.00E-04	1.47E-04	4.49E-05	Euler	PSO
3	4.68E-08	9.35E-08	4.32E-08	2.34E-08	9.08E-09	1.15E-05	5.82E-03	OCBO	Euler
4	2.36E-04	-	5.30E-05	1.34E-04	-	1.46E-05	3.97E-03	GA	Euler
5	-	2.22E-03	6.39E-05	1.12E-03	-	7.40E-05	5.37E-02	MFO	Euler
6	-	2.67E-06	6.10E-05	1.21E-04	-	1.82E-05	5.82E-05	PSO	FPA
7	-	5.22E-06	1.58E-05	3.64E-05	-	1.42E-05	3.88E-07	Euler	FPA

Note: There were some blank spaces due to some best approximate solutions were not provided in the respective references

Test Problem 2: Simple Electrical Circuit. In this problem, a system of ODEs for a simple circuit was considered (Sadollah et al., 2015c). The ODEs are expressed as Equation 26:

$$\frac{dI}{dt} = -I - V$$
$$\frac{dV}{dt} = 2I - V \tag{26}$$

with initial conditions $I(0) = 2$ and $V(0) = 2$

Here, I and V represent the current and voltage, respectively. The solution interval of time, t varies from 0 to 1.5. The exact solutions of current and voltage are given by Equation 27:

$$I(t) = 2e^{-t}\cos(\sqrt{2}t) - \sqrt{2}e^{-t}\sin(\sqrt{2}t)$$
$$V(t) = 2\sqrt{2}e^{-t}\sin(\sqrt{2}t) + 2e^{-t}\cos(\sqrt{2}t) \tag{27}$$

The test problem is unique in which two ODEs are solved simultaneously.

The NT of Fourier series was assigned to 3. Since the test problem is a system of two ODEs, hence, a total of 14 variables needs to be optimized, given as Equation 28 and 29:

$$I_{app}(t) = a_0 + a_1 \cos\left(\frac{\pi t}{1.5}\right) + b_1 \sin\left(\frac{\pi t}{1.5}\right) + a_2 \cos\left(\frac{2\pi t}{1.5}\right) + b_2 \sin\left(\frac{2\pi t}{1.5}\right) + a_3 \cos\left(\frac{3\pi t}{1.5}\right) + b_3 \sin\left(\frac{3\pi t}{1.5}\right) \quad (28)$$

$$V_{app}(t) = c_o + c_1 \cos\left(\frac{\pi t}{1.5}\right) + d_1 \sin\left(\frac{\pi t}{1.5}\right) + c_2 \cos\left(\frac{2\pi t}{1.5}\right) + d_2 \sin\left(\frac{2\pi t}{1.5}\right) + c_3 \cos\left(\frac{3\pi t}{1.5}\right) + d_3 \sin\left(\frac{3\pi t}{1.5}\right) \quad (29)$$

Table 6 and 7 summarize the best approximate solutions for the current function, I_{app} and the voltage function, V_{app} obtained by all algorithms. Figure 5 presents the comparison of the obtained best approximate solutions from each algorithm to the actual solution in Equation 27. It can be observed that MFO, GA and Euler's method can approximate both solutions almost perfectly. The approximate solutions given by the WCA, PSO, OCBO and FPA were only fitted well for one solution (I_{app} for WCA, PSO and OCBO while V_{app} for FPA) while having another solution overshoot (such as for WCA, PSO and OCBO while I_{app} for FPA). Since this test problem needs to solve a system of ODEs and thus, the algorithm which can optimize two functions simultaneously is desirable. In this regard, the MFO outperformed the others, which was further corroborated by the lowest GD of 3.7683E-04 in Table 4. The WCA gave the worst approximate solution, with the highest GD of 5.1703E-03. The unsatisfactory performance of WCA was due to it's inability to approximate the V_{app} precisely, as observed in Figure 5. Though this test problem has 2 curves, the GD values of both curves will be added up.

In terms of MSE, Table 5 reports different finding where the Euler's method gave the highest accuracy while the PSO was ranked last. This is attributed that for the GD, the Euclidean distance between point *i* in the approximate solution with the nearest point in the exact solution was considered, as illustrated in Figure 6. As such, the performance evaluation is different from the MSE in which the same point *i* in both the approximate and exact solutions is compared. Referring to Figure 5, it can be seen that comparison in terms of GD metric was more suited to visual assessment where the approximate solution given by the WCA was more significantly deviated from the actual solution.

Table 6

Test Problem 2 - The best approximate solution obtained by various algorithms for I_{app}

Fourier Coefficient	Algorithm					
	WCA	PSO	OCBO	MFO	FPA	GA
a_0	0.8381	0.8385	0.8385	9.6502E-01	9.2523E-01	8.3108E-01
a_1	1.3621	1.3626	1.3625	1.2481	1.4493	1.2764
b_1	−1.0557	-1.0563	−1.0562	-1.2309	-1.2998	-1.0187
a_2	−9.8533E-02	-9.8530E-02	−9.8535E-02	-1.8188E-01	-2.9654E-01	-6.8844E-02
b_2	−0.4597	-0.4600	−0.4600	-3.1873E-01	-4.7508E-01	-3.4941E-01
a_3	−0.1021	-0.1025	−0.1025	-3.6749E-02	-8.8318E-02	-4.8406E-02
b_3	1.1915E-02	1.1919E-02	1.1920E-02	2.7173E-02	9.3774E-02	2.9935E-03

Table 7

Test Problem 2 - The best approximate solution obtained by various algorithms for V_{app}

Fourier Coefficient	Algorithm					
	WCA	PSO	OCBO	MFO	FPA	GA
c_0	0.6618	0.6622	0.6620	1.0349	1.2179	5.3369E-01
c_1	0.8179	0.8184	0.8185	8.0697E-01	7.6768E-01	8.6274E-01
d_1	1.3265	1.326	1.3270	7.1052E-01	4.2777E-01	1.5600
c_2	0.4821	0.4818	0.4820	1.2034E-01	-3.5610E-02	5.8976E-01
d_2	8.5909E-02	-8.5912E-02	8.5915E-02	8.0109E-02	9.6923E-02	3.3360E-02
c_3	3.7436E-02	3.7431E-02	3.7430E-02	3.1987E-02	5.1374E-02	6.6653E-03
d_3	−0.1137	-0.1136	−0.1135	-8.2506E-04	5.4181E-02	-1.3986E-01

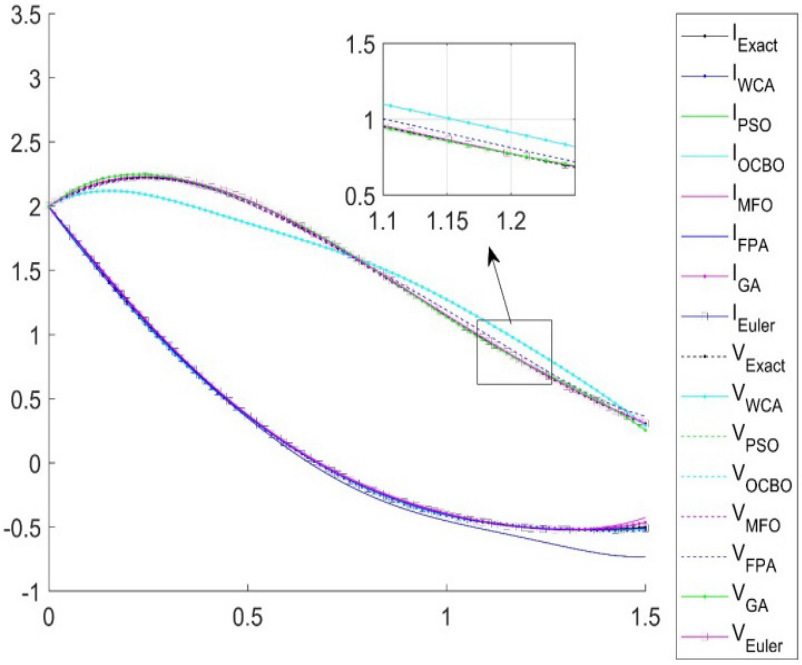


Figure 5. Test Problem 2 – Comparison of the best approximate solutions with the exact solution for WCA, PSO, OCBO, MFO, FPA, GA and Euler

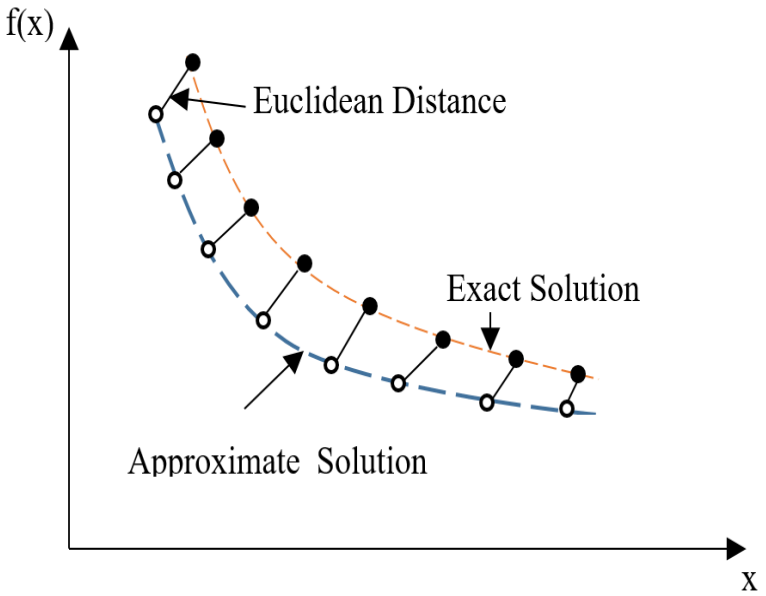


Figure 6. The schematic diagram of GD metric in evaluating the Euclidean distance between the approximate solution and the exact solution

Test Problem 3: Bernoulli Equation. The famous Bernoulli equation is considered in this test problem, which is formulated as Equation 30 (Sadollah et al., 2015c):

$$y'' + \left(y'\right)^2 - 2e^{-y} = 0 \tag{30}$$

with boundary conditions of $y\left(0\right) = 0$ and $y\left(1\right) = 0$. The exact solution of this second order nonlinear ODE is given by Equation 31:

$$y\left(x\right) = \ln\left(\left(x - \frac{1}{2}\right)^2 + \frac{3}{4}\right) \tag{31}$$

with x lies between 0 and 1.

The NT of the Fourier series was assigned as 3 and thus, the approximate function that needs to be optimized by all algorithms is as in Equation 32. The obtained best approximate solutions by all algorithms are shown in Table 8

$$\begin{aligned} Y_{app}\left(x\right) = a_o + a_1\cos\left(\pi x\right) + b_1\sin\left(\pi x\right) + a_2\cos\left(2\pi x\right) \\ + b_2\sin\left(2\pi x\right) + a_3\cos\left(3\pi x\right) + b_3\sin\left(3\pi x\right) \end{aligned} \tag{32}$$

Table 8

Test Problem 3 - The best approximate solution obtained by various algorithms

Fourier Coefficient	Algorithm					
	WCA	PSO	OCBO	MFO	FPA	GA
a_0	2.25E-02	2.40E-02	2.2512E-02	2.1661E-02	2.1926E-02	2.8790E-02
a_1	8.31E-05	-1.48E-05	8.3125E-05	-2.6965E-04	3.1164E-05	1.9494E-03
b_1	-0.3288	-3.32E-01	−0.3290	-3.2742E-01	-3.2808E-01	-3.4499E-01
a_2	-2.25E-02	-2.40E-02	−2.2490E-02	-2.1572E-02	-2.2097E-02	-3.1143E-02
b_2	1.67E-04	4.09E-05	−1.6735E-04	6.5889E-05	8.9764E-05	-2.8985E-04
a_3	-8.31E-05	1.45E-05	−8.3096E-05	1.5966E-05	5.8905E-05	-1.5094E-05
b_3	3.69E-03	4.08E-03	3.6882E-03	3.3740E-03	3.5244E-03	7.6365E-03

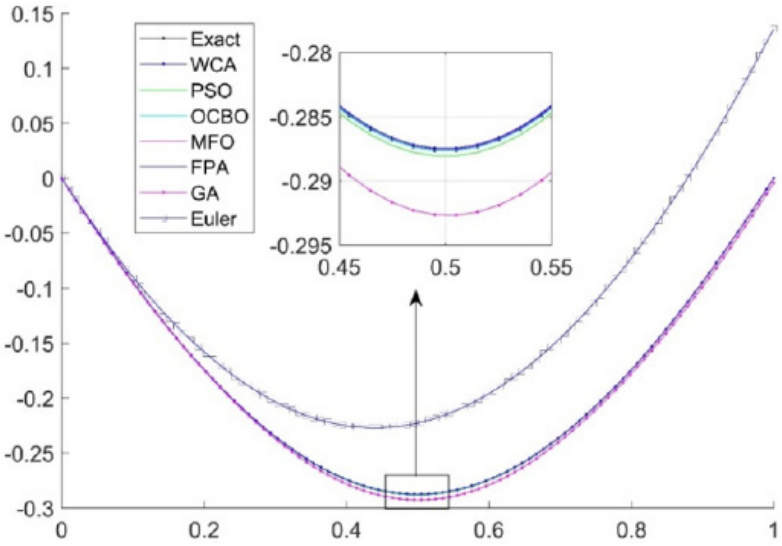


Figure 7. Test Problem 3 – Comparison of the best approximate solutions with the exact solution for WCA, PSO, OCBO, MFO, FPA, GA and Euler

The comparison of the approximate solution to the exact solution (Equation 31) for each algorithm is presented in Figure 7. All algorithms were able to approximate the solution satisfactorily, where the approximate solutions fitted the exact solution well except for the Euler's method. The approximate solution of the Euler's method deviated from the exact solution as the x increases. In terms of the GD metric (Table 4), the GD of the OCBO was the lowest, followed closely by the FPA, MFO, WCA, PSO and GA with differences of 4.0885E-6, 7.9635E-06, 8.5635E-06, 1.4885E-05 and 2.3317E-04. Euler's method gave the highest GD of 4.5792E-03. The same findings were reported in terms of MSE, where the Euler's gave the lowest approximation accuracy while OCBO was ranked first.

Test Problem 4: Nonlinear ODE with Neumann Boundary Condition. A more complex ODE is examined in this test problem, in which the Neumann boundary condition is considered. Moreover, the difficulty of solving the ODE increases as this test problem has a wide range of interval solution, *i.e.* x varies from 0 to 15. The ODE is expressed as Equation 33 (Sadollah et al., 2015 c)

$$y'' + 0.3y' + y - 1 = 0 \tag{33}$$

with boundary conditions of $y(0) = 0$ and $y'(0) = 0$

The exact solution of this test problem is given by Equation 34:

$$y(x)=\left(\frac{\frac{-3\sqrt{391}}{391}\sin\left(\frac{\sqrt{391}x}{20}\right)}{e^{\left(\frac{3x}{20}\right)}}\right)-\left(\frac{\cos\left(\frac{\sqrt{391}x}{20}\right)}{e^{\left(\frac{3x}{20}\right)}}\right)+1 \tag{34}$$

The obtained Fourier series with NT=6 is expressed as Equation 35:

$$\begin{aligned}Y_{approx}(x)=a_o&+a_1\cos\left(\frac{\pi x}{15}\right)+b_1\sin\left(\frac{\pi x}{15}\right)+a_2\cos\left(\frac{2\pi x}{15}\right)\\&+b_2\sin\left(\frac{2\pi x}{15}\right)+a_3\cos\left(\frac{3\pi x}{15}\right)+b_3\sin\left(\frac{3\pi x}{15}\right)\\&+a_4\cos\left(\frac{4\pi x}{15}\right)+b_4\sin\left(\frac{4\pi x}{15}\right)+a_5\cos\left(\frac{5\pi x}{15}\right)\\&+b_5\sin\left(\frac{5\pi x}{15}\right)+a_6\cos\left(\frac{6\pi x}{15}\right)+b_6\sin\left(\frac{6\pi x}{15}\right)\end{aligned} \tag{35}$$

Table 9 shows the best approximate solutions given by the WCA, MFO and FPA. The results of PSO and OCBO were not presented due to this problem was not simulated in Babaei (2013) and Panda and Pani (2017). The best approximate solutions were compared against the exact solution in Figure 8. As observed in Figure 8, all the obtained approximate solutions exhibited a slight deviation from the exact solution, but the given approximate solutions were able to capture the behaviour of the exact solution. However, the approximate solution given by the Euler's method was unsatisfactory at the regions with local minima and local maxima. The comparison in terms of the GD metric in Table 4 shows that the lowest GD was obtained by the GA, and followed by MFO, FPA and WCA, while the highest GD was obtained by the Euler. In terms of MSE, it can be seen in Table 5 that Euler's method also produced the lowest approximation accuracy.

Table 9

Test Problem 4 - The best approximate solution obtained by various algorithms

Fourier Coefficient	Algorithm			
	WCA	MFO	FPA	GA
a_0	0.8861	9.5814E-01	9.6552E-01	7.4148E-01
a_1	-0.1652	-6.9514E-01	3.3187E-01	-9.9751E-02
b_1	0.2533	2.1481E-01	6.1323E-02	5.4197E-01
a_2	0.3278	5.1538E-01	5.6435E-02	6.0468E-01

Table 9 *(Continued)*

Fourier Coefficient	Algorithm			
	WCA	MFO	FPA	GA
b_2	0.2068	1	-5.3821E-01	1.0262E-01
a_3	0.1606	8.4533E-01	-5.4024E-01	2.5716E-02
b_3	-0.4637	-8.0719E-01	-1.5700E-01	-6.9240E-01
a_4	-0.6637	-1	-4.0653E-01	-8.1705E-01
b_4	-0.1965	-6.1628E-01	2.4779E-01	-1.0015E-01
a_5	-0.5218	-6.7200E-01	-3.0362E-01	-4.5059E-01
b_5	0.1909	3.8418E-01	5.0291E-02	2.6097E-01
a_6	-0.0238	4.0047E-02	-6.2747E-02	-2.1383E-04
b_6	0.0926	1.2095E-01	3.6468E-02	7.1086E-02

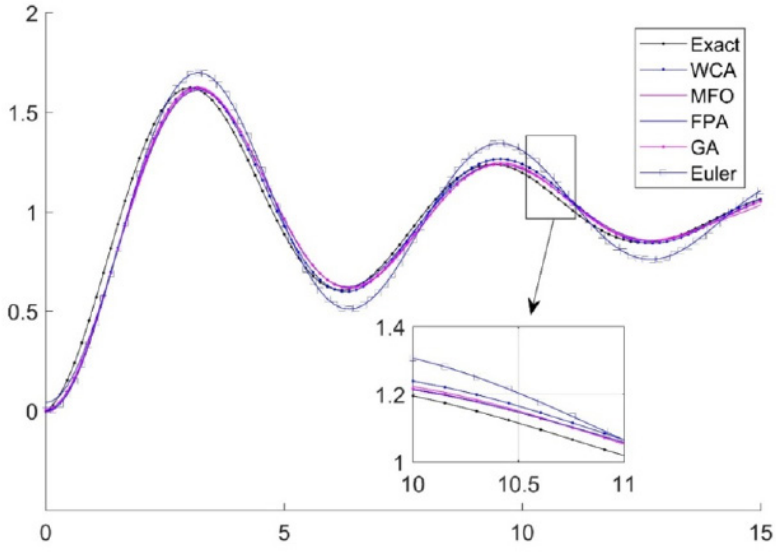


Figure 8. Test Problem 4 – Comparison of the obtained best approximate solutions with the exact solution for WCA, MFO, FPA, GA and Euler

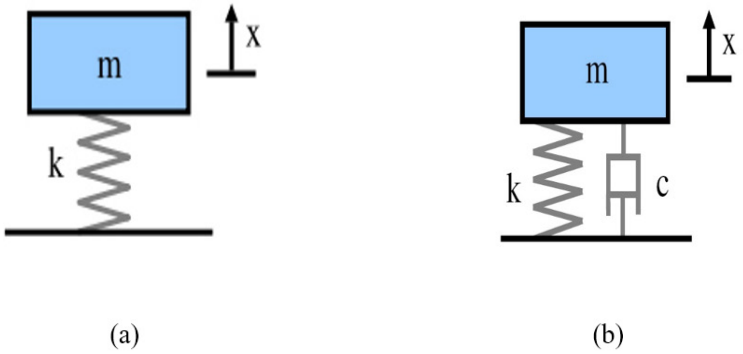
Test Problem 5: Mass-Spring System without Damping Force. A mechanical mass-spring system without damping force as illustrated in Figure 9(a) is considered in this test problem. The spring is stretched with a length of 27cm by a 7.25kg object. At the initial position, the spring is displaced 182.88cm upwards from equilibrium, together with an initial downward velocity of 30.48 cm/s. No external and damping forces are involved. By solving this mass-spring system, the ODE as in Equation 36 (Sadollah et al., 2015c):

$$\frac{1}{2}u'' + 18u = 0 \tag{36}$$

with initial conditions of $u(0) = -1/2$ and $u'(0) = 1$ is obtained. The analytical solution of this mechanical problem is given by Equation 37:

$$u(t) = 0.52705\cos(6t - 2.81984) \tag{37}$$

with solution interval of t varies from 0 to 3.



(a) (b)

Figure 9 Mass-Spring System (a) without damping force, (b) with damping force

The NT of the Fourier series was chosen as 6. The formulated Fourier series to be optimized by the PSO, MFO and FPA algorithms is as in Equation 38:

$$\begin{aligned} Y_{app}(x) = a_o &+ a_1\cos\left(\frac{\pi x}{3}\right) + b_1\sin\left(\frac{\pi x}{3}\right) + a_2\cos\left(\frac{2\pi x}{3}\right) + b_2\sin\left(\frac{2\pi x}{3}\right) \\ &+ a_3\cos\left(\frac{3\pi x}{3}\right) + b_3\sin\left(\frac{3\pi x}{3}\right) + a_4\cos\left(\frac{4\pi x}{3}\right) + b_4\sin\left(\frac{4\pi x}{3}\right) \\ &+ a_5\cos\left(\frac{5\pi x}{3}\right) + b_5\sin\left(\frac{5\pi x}{3}\right) + a_6\cos\left(\frac{6\pi x}{3}\right) + b_6\sin\left(\frac{6\pi x}{3}\right) \end{aligned} \tag{38}$$

Table 10 summarizes the best approximate solutions obtained by all algorithms. The comparison between the best approximate solutions and the analytical solution in Equation 36 (Figure 10) shows that all algorithms can capture the changing trend of the exact solution. The best fit of the MFO with the actual solution can be observed, with the lowest GD value as presented in Table 4. The approximate solutions obtained by the GA, PSO

and FPA demonstrated a slight deviation from the exact solution. The worst approximate solution was given by Euler's method, with the highest GD value of 1.3702E-02. The approximate solution of Euler's method deviated gradually from the exact solution on each local minimum and maximum. The lowest approximation accuracy in terms of MSE was also reported by Euler's method, as shown in Table 5.

Table 10

Test Problem 5 - The best approximate solution obtained by various algorithms

Fourier Coefficient	Algorithm			
	PSO	MFO	FPA	GA
a_0	0.1867	-4.1899E-01	4.9163E-01	-2.2289E-01
a_1	0.1356	2.9464E-01	-6.0758E-02	-1.6153E-01
b_1	-0.2727	8.9775E-01	-8.3639E-01	4.3788E-01
a_2	-3.3241E-02	1	-4.6687E-01	4.1815E-01
b_2	-0.2059	-4.5991E-01	1.0118E-01	2.6953E-01
a_3	-0.1890	-4.4598E-01	1.0218E-01	2.9498E-01
b_3	-0.2059	-1	5.1262E-02	-3.9466E-01
a_4	-0.3420	-8.3637E-01	-2.4337E-01	-3.7587E-01
b_4	0.1170	3.0496E-01	-9.4811E-02	-2.4499E-01
a_5	-3.4034E-02	1.3113E-01	-9.1927E-02	-1.6475E-01
b_5	0.4219	6.4957E-01	4.1332E-01	4.1686E-01
a_6	-0.2922	-2.3638E-01	-2.7505E-01	-2.9000E-01
b_6	-5.3461E-02	-8.1045E-02	-4.9702E-02	8.5314E-03

Test Problem 6: Mass-Spring System with Damping Force. For test problem 5, when the velocity is 60.96 cm/s, a damper which imposes a force of 2.26 kg is added to the mass-spring system (Figure 9 (b)). The second-order ODE derived from this system is given by Equation 39 (Sadollah et al., 2015 c)

$$\frac{1}{2}u'' + \frac{5}{2}u' + 18u = 0 \tag{39}$$

with initial conditions of $u(0) = -1/2$ and $u'(0) = 1$. The analytical solution of this mass-spring system is given by Equation 40:

$$u(t) = 0.502096e^{\frac{5}{2}t}\cos\left(\frac{\sqrt{119}}{2}t - 3.2321\right) \tag{40}$$

with the interval of t is between 0 and 3.

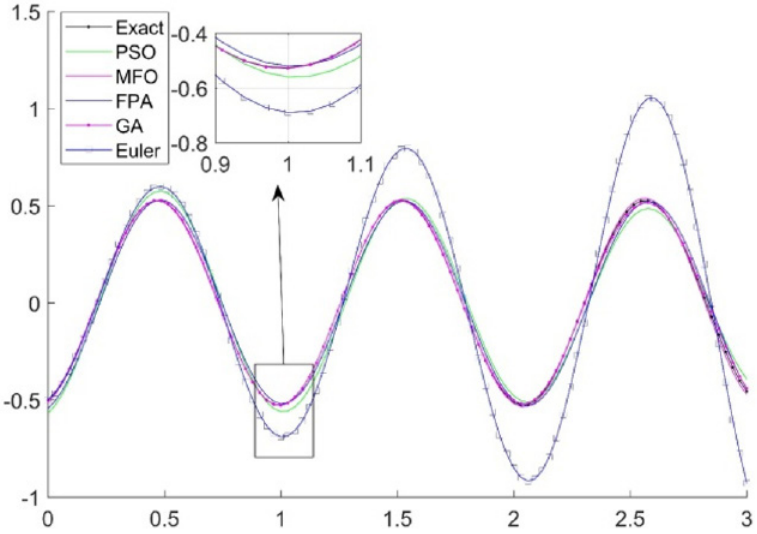


Figure 10. Test Problem 5 – Comparison of the best approximate solutions with the exact solution for PSO, MFO, FPA, GA and Euler

Following the work in (Sadollah et al., 2015c), the NT of 8 was chosen. The Fourier coefficients in Equation 41 were then optimized using the PSO, MFO and FPA algorithms. Table 11 presents the best approximate solutions obtained by each algorithm.

$$\begin{aligned} Y_{app}(x) = a_o &+ a_1\cos\left(\frac{\pi x}{3}\right) + b_1\sin\left(\frac{\pi x}{3}\right) + a_2\cos\left(\frac{2\pi x}{3}\right) + b_2\sin\left(\frac{2\pi x}{3}\right) \\ &+ a_3\cos\left(\frac{3\pi x}{3}\right) + b_3\sin\left(\frac{3\pi x}{3}\right) + a_4\cos\left(\frac{4\pi x}{3}\right) + b_4\sin\left(\frac{4\pi x}{3}\right) \\ &+ a_5\cos\left(\frac{5\pi x}{3}\right) + b_5\sin\left(\frac{5\pi x}{3}\right) + a_6\cos\left(\frac{6\pi x}{3}\right) + b_6\sin\left(\frac{6\pi x}{3}\right) \\ &+ a_7\cos\left(\frac{7\pi x}{3}\right) + b_7\sin\left(\frac{7\pi x}{3}\right) + a_8\cos\left(\frac{8\pi x}{3}\right) + b_8\sin\left(\frac{8\pi x}{3}\right) \end{aligned} \tag{41}$$

Table 11

Test Problem 6 - The best approximate solution obtained by various algorithms

Fourier Coefficient	Algorithm			
	PSO	MFO	FPA	GA
a_0	4.2427E-02	1.6649E-01	2.0069E-02	-2.4357E-01
a_1	0.2692	6.3516E-01	-2.9983E-02	-2.2733E-01
b_1	-8.1634E-02	-1.8759E-01	-2.0882E-01	3.7190E-01
a_2	-3.8633E-02	1.4685E-01	-5.3383E-01	1.4775E-01
b_2	-0.4001	-9.2785E-01	-6.1953E-03	2.3527E-01
a_3	-0.2950	-7.1377E-01	-2.3839E-02	7.8729E-02
b_3	-7.6912E-02	-4.4013E-01	6.8551E-01	-2.1613E-02
a_4	-0.1595	-4.6111E-01	6.0477E-01	-8.5710E-04
b_4	-3.2617E-02	1.1574E-01	-9.8299E-02	-1.8529E-03
a_5	-0.3102	-3.9711E-01	-2.3365E-01	-2.5997E-02
b_5	3.6654E-02	1.4303E-01	-5.0980E-01	-9.4478E-02
a_6	-0.1159	-1.5763E-01	-3.8550E-01	-1.6933E-01
b_6	0.2508	4.1643E-01	1.4217E-01	-3.8009E-02
a_7	8.5309E-02	1.9945E-01	1.9615E-02	-7.2358E-02
b_7	8.8175E-02	1.5156E-01	1.6755E-01	9.1919E-02
a_8	2.2355E-02	4.7600E-02	2.9148E-02	1.7476E-02
b_8	-1.3398E-02	-5.1783E-02	6.6004E-03	2.9737E-02

Figure 11 presents the comparison of the best approximate solutions found by all algorithms to the exact solution. It can be observed that the approximate solution obtained by the PSO fitted the exact solution well for all data points. On the other hand, a noticeable deviation from the exact solution after $x = 2$ can be observed for the approximate solution obtained by the MFO, while the approximate solutions of the FPA and GA deviated from the exact solution before $x = 0.5$ and at $x = 1$. All the algorithms, in general, were able to approximate the exact solution satisfactorily, except significant deviation from the exact solution at $x = 0.5$ and $x = 1$ was observed for the Euler's method. Considering the values of GD metric in Table 4 the PSO gave better approximate solution than others, corroborated by the lowest GD metric in addition to the visual inspection. Euler's method, on the other hand, gave the worst approximate solution.

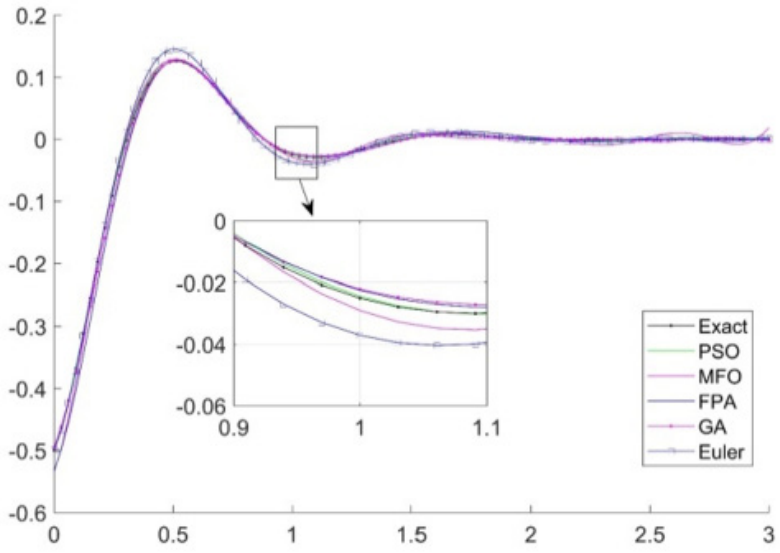


Figure 11. Test Problem 6 – Comparison of the best approximate solutions with the exact solution for PSO, MFO, FPA, GA and Euler

Test Problem 7: Longitudinal Heat Transfer Fins. Longitudinal fins are widely utilized in heat exchanger application to increase the rate of heat dissipation to or from the environment. Consider a longitudinal fin with constant rectangular profile as shown in Figure 12, the fin with length L and semi-base thickness t_b is attached to a surface with a constant temperature T_b. The convective heat transfer coefficient, h , fin thermal conductivity, k and fin surface emissivity, ε are assumed as Equation 42, 43 and 44:

$$h = h_b \left[\frac{T - T_a}{T_b - T_a} \right]^m \tag{42}$$

$$k = k_a \left[1 + \alpha \left(T - T_a \right) \right] \tag{43}$$

$$\varepsilon = \varepsilon_s \left[1 + \beta \left(T - T_s \right) \right] \tag{44}$$

where h_b is convection heat transfer coefficient according to temperature difference T_a and T_b , k_a is thermal conductivity at convection sink temperature T_a , ε_s is fin surface emissivity at radiation sink temp T_s and constant α and β are variation of thermal conductivity and surface emissivity with temperature, respectively.

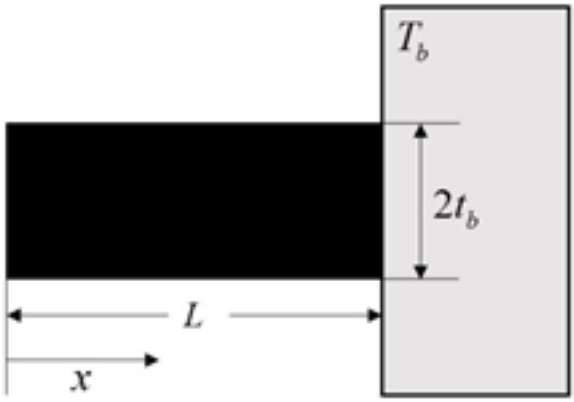


Figure 12. A longitudinal fin with constant rectangular profile

The energy balance of longitudinal fin per unit width based on one-dimensional heat conduction is as in Equation 45 (Torabi et al., 2013)

$$k_a \frac{d}{dx}\left[\left[1+\alpha\left(T-T_a\right)\right]t\left(x\right)\frac{dT}{dx}\right]-h_b\left[\frac{T-T_a}{T_b-T_a}\right]^m\left(T-T_a\right) \\ -\sigma\varepsilon_s\left[1+\beta\left(T-T_s\right)\right]\left(T^4-T_s^4\right)=0 \tag{45}$$

where $t\left(x\right)$ is the local semi-fin thicknesses (Equation 46):

$$t\left(x\right)=t_b+\delta\left[\left(x/L\right)^n-1\right] \tag{46}$$

where exponent n = 0, 1 and 2 for rectangular, trapezoidal and concave parabolic fins, respectively and δ is the fin taper. The boundary conditions of adiabatic tip and constant base temperature are (Equation 47 & 48):

$$\left.\frac{dT}{dx}\right|_{x=0}=0 \tag{47}$$

$$T\left(L\right)=T_b \tag{48}$$

By assuming these dimensionless parameters (Equation 49, 50, 51 & 52):

$$\theta=\frac{T}{T_b},\theta_a=\frac{T_a}{T_b},\theta_s=\frac{T_s}{T_b},X=\frac{x}{L},C=\frac{\delta}{t_b},Nc=\frac{h_bL^2T_b^m}{k_at_b\left(T_b-T_a\right)^m}, \\ A=\alpha T_b,B=\beta T_b,Nr=\frac{\sigma\varepsilon_sL^2T_b^3}{k_at_b},\psi=\frac{t_b}{L} \tag{49}$$

and hence, the fin profile forms the following ODE:

$$\frac{d}{dx}\left[\left[1+A\left(\theta-\theta_a\right)\right]\left[1+C\left(X^n-1\right)\right]\frac{d\theta}{dX}\right]-Nc\left(\theta-\theta_a\right)^{m+1} \\ -Nr\left[1+B\left(\theta-\theta_s\right)\right]\left(\theta^4-\theta_s^4\right)=0 \tag{50}$$

with boundary conditions of:

$$\theta\left(1\right)=1 \tag{51}$$

$$\left.\frac{d\theta}{dX}\right|_{X=0}=0 \tag{52}$$

where parameter A represents thermal conductivity, B represents emissivity, m is associated with convective heat transfer coefficient, Nc represents convection-conduction parameter, Nr represents radiation-conduction parameter, θ_a represents temperature of convection, θ_s represents temperature of radiation sinks and C represents fin taper ratio. In a longitudinal fin with constant cross-sectional area, A_c, length, L and perimeter of cross-section, P, the following dimensionless parameters in Equation 53 are introduced into Equation 45:

$$\theta=\frac{T}{T_b},\theta_a=\frac{T_a}{T_b},\theta_s=\frac{T_s}{T_b},X=\frac{x}{L},Nc=\frac{h_bL^2P}{k_aA_c}, \\ A=\alpha T_b,B=\beta T_b,Nr=\frac{\sigma\varepsilon_sL^2T_b^3P}{k_aA_c} \tag{53}$$

The obtained ODE is as in Equation 54:

$$\frac{d}{dX}\left[\left[1+A\left(\theta-\theta_a\right)\right]\frac{d\theta}{dx}\right]-Nc\left(\theta-\theta_a\right)\left[\frac{\theta-\theta_a}{1-\theta_a}\right]^m \\ -Nr\left[1+B\left(\theta-\theta_s\right)\right]\left(\theta^4-\theta_s^4\right)=0 \tag{54}$$

with boundary conditions in Equation 55 and 56:

$$\theta\left(0\right)=1 \tag{55}$$

$$\left.\frac{d\theta}{dX}\right|_{X=1}=0 \tag{56}$$

Assuming that the heat conduction occurs only in the longitudinal direction (x-direction) for a pure convection fin (*i.e.* radiation conduction, $Nr = 0$) with constant thermal conductivity ($A = 0$) operating in a constant convective heat transfer coefficient, m is assigned to 0,

Nc is assumed to be 1, C is equal to zero for fin with constant cross-sectional area and θ_a is 0.5, the governing ODE can be expressed as Equation 57 (Sadollah et al, 2015a):

$$\frac{d^2\theta}{dx^2} - \theta + 0.5 = 0 \tag{57}$$

with the length of fin L is assigned to be 1. The analytical solution of this ODE is given by Equation 58:

$$\theta(x) = 0.162e^{x} + 0.162e^{-x} + 0.5 \tag{58}$$

with x is within interval 0 to 1.

The NT was chosen as 3 in this test problem. Table 12 summarizes the obtained best approximate solutions for each algorithm, where it can be observed that the obtained coefficients for the Fourier series are similar for all algorithms, except for GA. As shown in Table 4, Euler's method has surpassed GA, MFO, PSO and FPA with the lowest GD value, albeit in Figure 13 the obtained approximated solutions by all algorithms are close to the actual solution. The performance of all algorithms were considered satisfactorily with the GDs within the resolution of 1E-04. As shown in Table 5, the same performance can be observed in which the Euler's method gave the highest accuracy while the FPA was ranked last.

Table 12

Test Problem 7 - The best approximate solution obtained by various algorithms

Fourier Coefficient	Algorithm			
	PSO	MFO	FPA	GA
a_0	0.917338	9.1900E-01	9.1917E-01	9.0374E-01
a_1	-0.093953	-8.6519E-02	-8.3184E-02	-8.5112E-02
b_1	-0.054890	-5.9048E-02	-5.7407E-02	-3.8041E-02
a_2	-0.005338	-7.5052E-03	-6.6807E-03	3.9307E-03
b_2	0.026293	2.5749E-02	2.4814E-02	2.2851E-02
a_2	0.005953	5.1650E-03	4.9853E-03	3.6390E-03
b_3	7.67E-04	7.7181E-04	5.9018E-04	-3.0567E-03

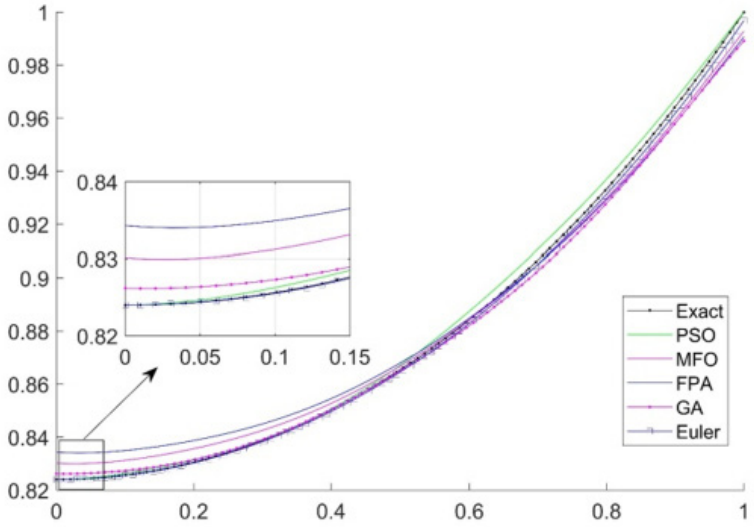


Figure 13. Test Problem 7 – Comparison of the best approximate solutions with the exact solution for PSO, MFO, FPA, GA and Euler

Comparison of Optimization Performance and Discussion

From the obtained results in Table 4, each algorithm was ranked based on its optimization performance in terms of the GD metric for all considered test problems. The results are summarized in Table 13. The OCBO was ranked first among others, followed by the MFO, FPA and GA with close competition.

Also, since this study attempts to develop an approximate solution of ODE with the formulation of Fourier series expansion based on the MFO and FPA as the optimizers, the results of the average, worst and standard deviation (SD) of GD assessment metric for these optimizers are presented in Table 14. Interestingly, though MFO and GA outperformed the FPA in most of the test problems if considering only the best GD values (Table 4), their average GD values in all test problems were found to be higher than the FPA in Table 14. This implies that the FPA produces more consistent approximate solutions than the MFO and GA optimizers, indicated by the lower SD values in all test problems.

Table 13

The ranking of all considered algorithms based on the obtained results in Table 4

Test Problem	Algorithm						
	PSO	WCA	OCBO	MFO	FPA	GA	Euler
1	5	2	3	6	4	7	1
2	6	7	2	1	4	5	3
3	5	4	1	3	2	6	7
4	-	4	-	2	3	1	5
5	4	-	-	1	3	2	5
6	1	-	-	3	4	2	5
7	4	-	-	3	5	2	1
Average Rank	4.17	4.25	2	2.71	3.57	3.57	3.86
Total Rank	6^{th}	7^{th}	1^{st}	2^{nd}	3^{rd} /4^{th}	3^{rd} /4^{th}	5^{th}

Table 14

Results of the average, worst and standard deviation (SD) of GDs for MFO, FPA and GA algorithms for all test problems

Test Problem	GD value	Algorithm		
		MFO	FPA	GA
1	Best	3.6552E-04	2.5799E-04	4.7591E-04
	Average	3.8010E-03	5.6645E-04	2.3227E-03
	Worst	2.7739E-2	8.8413E-04	6.9962E-03
	SD	5.6027E-03	1.5857E-04	1.4704E-03
2	Best	3.7638E-04	1.0593E-03	8.3174E-04
	Average	1.5327E-02	4.2612E-03	2.3910E-03
	Worst	6.1616E-02	7.6770E-03	4.3628E-03
	SD	1.5193E-02	1.7392E-03	9.4657E-04
3	Best	1.4702E-05	1.0827E-05	2.3991E-04
	Average	1.5521E-03	1.5522E-04	5.6107E-04
	Worst	4.5131E-02	3.2881E-04	1.3685E-03
	SD	8.0924E-03	7.9581E-05	2.8355E-04
4	Best	5.1476E-04	8.1790E-04	2.7037E-04
	Average	6.3881E-03	1.8513E-03	2.4832E-03

Table 14 *(Continued)*

Test Problem	GD value	Algorithm		
		MFO	FPA	GA
4	Worst	3.1902E-02	3.0593E-03	5.1182E-03
	SD	7.1452E-03	5.3710E-04	1.1382E-03
5	Best	5.4773E-04	1.1004E-03	5.7002E-04
	Average	1.3354E-02	2.0702E-03	2.4800E-03
	Worst	6.8411E-02	5.3404E-03	4.4761E-03
	SD	1.6160E-02	9.1795E-04	9.1920E-04
6	Best	4.7393E-04	4.7710E-04	3.0154E-04
	Average	2.5800E-03	7.2895E-04	9.9495E-04
	Worst	1.4132E-02	1.0160E-03	2.7513E-03
	SD	3.0067E-03	1.3648E-04	5.1470E-04
7	Best	2.8087E-04	4.2669E-04	2.6636E-04
	Average	1.5787E-02	4.4573E-04	6.2556E-03
	Worst	8.8986E-02	4.6948E-04	2.0060E-02
	SD	2.7739E-02	1.1897E-05	5.5743E-03

Table 15 presents the average iterations used by MFO, FPA and GA to converge for all problems. The simulation of each algorithm was repeated for 30 runs. For each run, the required number of iteration to reach the stopping condition was recorded, and subsequently summed up and averaged. Hence, the average iterations shown in Table 15 are in decimal number instead of natural number. For test problem 1 and 2, in comparison with FPA, more iterations are needed by the MFO algorithm and GA in order to converge. For other test problems, MFO outperformed both FPA and GA in terms of convergence speed. The significant difference can be observed in test problem 6 in which the MFO can converge to the optimal solutions with 400 iterations lesser, as compared to FPA and GA.

Table 15

Average iterations used to reach convergence by MFO, FPA and GA algorithms for all test problems

Test Problem	Iteration	Algorithm		
		MFO	FPA	GA
1	Average Iteration	999.4	677.9	699.4
	Maximum Iteration	1000	700	700

Table 15 *(Continued)*

Test Problem	Iteration	Algorithm		
		MFO	FPA	GA
2	Average Iteration	499.6	488.8	499.5
	Maximum Iteration	500	500	500
3	Average Iteration	499.5	791.8	798.7
	Maximum Iteration	500	800	800
4	Average Iteration	299.6	489.9	499.1
	Maximum Iteration	300	500	500
5	Average Iteration	399.7	481.6	499.3
	Maximum Iteration	400	500	500
6	Average Iteration	599.6	971.6	999.5
	Maximum Iteration	600	1000	1000
7	Average Iteration	936.4	987.2	999.6
	Maximum Iteration	1000	1000	1000

Relationship of Number of Terms to the Approximation Accuracy

In the previous section, the performances of the optimization algorithms were evaluated in terms of the GD metric and MSE. The NT used in the test problems was chosen according to the studies in Babaei (2013), Sadollah et al. (2015a), and Sadollah et al. (2015c). In this section, the relationship between using different NTs to the accuracy of the approximate solutions was explored. The NTs were varied from 1 to 9, and the obtained Fourier series was then optimized using the MFO, FPA and GA. The number of population and the maximum number of iteration of these algorithms were fixed as in Table 1. All simulations were repeated for 30 trials. For each run, the simulation continues until the maximum iteration was satisfied.

By varying the NTs from 1 to 9, the comparisons of the approximate solutions obtained by the FPA to the actual solutions for test problems 4, 5 and 6 are depicted in Figure 14 to 16. Others are omitted for brevity. A common trend can be observed that for the first few NTs, the algorithms were unable to predict the approximate solutions due to insufficient NTs. When sufficient NT was employed, the algorithms were able to approximate the exact solutions satisfactorily. However, the further increase in NT led to a decline in the approximate accuracy, due to more unknown coefficients were involved. Thus, more iterations and bigger population size were needed such that the MFO, FPA or GA can converge to the optimal solution. However, in this study, the population size and the

maximum number of iteration were fixed. Therefore, these algorithms could not reach the optimal solution within the allowable iterations for NT = 7 and onwards for test problem 4.

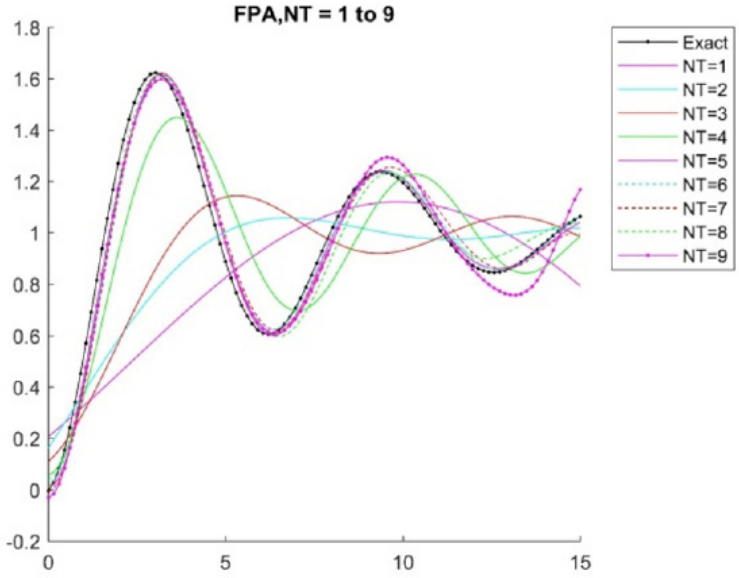


Figure 14. Test Problem 4 – Comparison of the obtained best approximate solutions with the exact solution for FPA by varying the NTs from 1 to 9

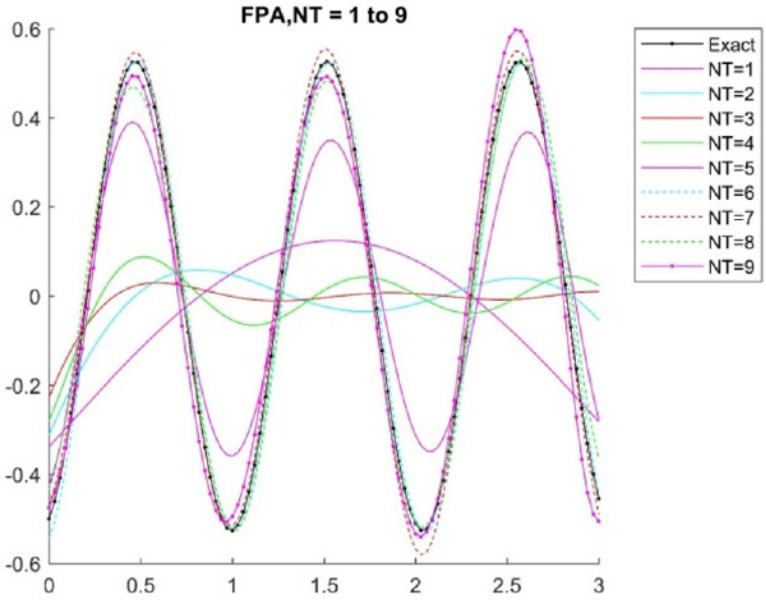


Figure 15. Test Problem 5 – Comparison of the obtained best approximate solutions with the exact solution for FPA by varying the NTs from 1 to 9

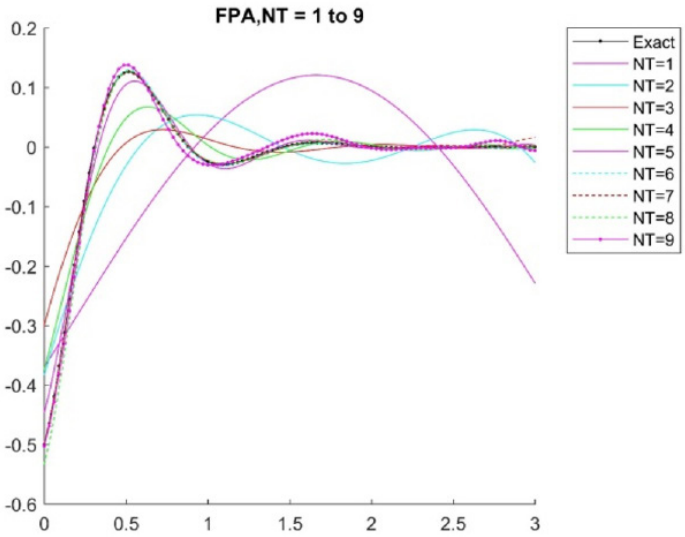


Figure 16. Test Problem 6 – Comparison of the obtained best approximate solutions with the exact solution for FPA by varying the NTs from 1 to 9

Table 16 summarizes the GD values of the best approximate solutions obtained by the MFO, FPA and GA optimizers for all test problems when different NTs were considered and depicted further in Figure 17. It can be observed that similar decreasing and increasing behaviours were shown by the optimizers. Initially, the GD value decreases with increasing NTs until it reaches the minimum. After this point, a further increase in NT was no longer proportional to the obtained approximation accuracies. It can be inferred that neither low nor high NTs guarantee high approximation accuracy. The NT = 3, NT = 5 or 6 and NT = 4 or 7 were recommended for MFO, FPA and GA, respectively, based on the obtained results.

Table 16

The best GD values obtained by MFO, FPA and GA optimizers for all test problems using different NTs

Algorithm	NT	Test Problem			
		1	2	3	4
MFO	1	5.0502E-03	7.7745E-03	3.9283E-04	2.6548E-02
	2	1.2662E-03	5.7775E-04	1.8552E-04	2.2889E-02
	3	2.9339E-04	**3.7683E-04**	**1.4702E-05**	2.2890E-02
	4	3.3058E-04	5.4313E-04	1.7945E-04	9.9603E-03

Table 16 *(Continued)*

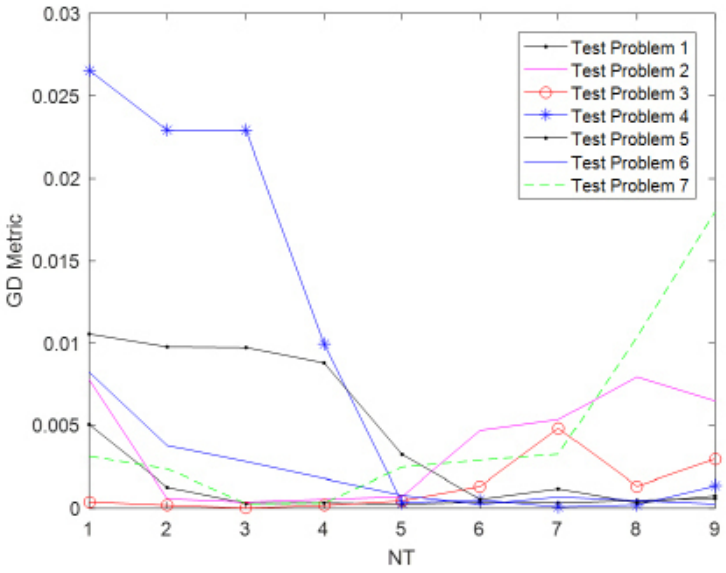
Algorithm	NT	Test Problem			
		1	2	3	4
	5	2.4864E-04	7.0202E-04	4.5420E-04	3.4455E-04
	6	3.6552E-04	4.6924E-03	1.3182E-03	5.1476E-04
	7	3.2673E-04	5.3539E-03	4.8136E-03	**8.1944E-05**
	8	4.8252E-04	7.9415E-03	1.3306E-03	2.0955E-04
	9	5.9491E-04	6.4883E-03	2.9632E-03	1.3481E-03
FPA	1	5.0648E-03	7.5922E-03	3.9793E-04	2.6556E-02
	2	1.4192E-03	**4.6520E-04**	1.8879E-04	2.2934E-02
	3	2.3567E-04	1.0593E-03	**1.0827E-05**	2.2727E-02
	4	2.0434E-04	2.8545E-03	1.8994E-04	9.4469E-03
	5	**1.8545E-04**	6.5862E-03	2.4402E-04	**6.4572E-04**
	6	2.5799E-04	4.9055E-03	8.1734E-04	8.1790E-04
	7	4.0502E-04	1.1536E-02	6.3980E-04	1.4909E-03
	8	3.8971E-04	1.3597E-02	6.0018E-04	2.3740E-03
	9	7.4868E-04	1.9429E-02	2.5486E-03	2.9649E-03
GA	1	4.1139E-03	7.2000E-03	3.7000E-04	2.6540E-02
	2	1.1776E-03	6.3031E-04	1.2644E-04	2.2604E-02
	3	4.2029E-04	8.3174E-04	2.3991E-04	2.2937E-02
	4	**2.9718E-04**	5.3528E-04	**1.1964E-04**	1.2308E-02
	5	8.6889E-04	5.5544E-04	1.7728E-04	4.2365E-03
	6	4.7591E-04	**3.0247E-04**	1.9681E-04	2.7037E-04
	7	5.0370E-04	1.8568E-03	3.3357E-04	1.2700E-04
	8	3.5526E-04	1.0719E-03	4.0010E-04	**1.2113E-04**
	9	1.0009E-03	2.2051E-03	2.0830E-04	1.9080E-04

Algorithm	NT	Test Problem		
		5	6	7
MFO	1	1.0556E-02	8.2317E-03	3.1265E-03
	2	9.7771E-03	3.7754E-03	2.3641E-03
	3	9.7323E-03	2.8035E-03	**2.8087E-04**
	4	8.7953E-03	1.7855E-03	3.1540E-04
	5	3.2337E-03	7.8302E-04	2.4732E-03
	6	5.4773E-04	**2.3517E-04**	2.8963E-03

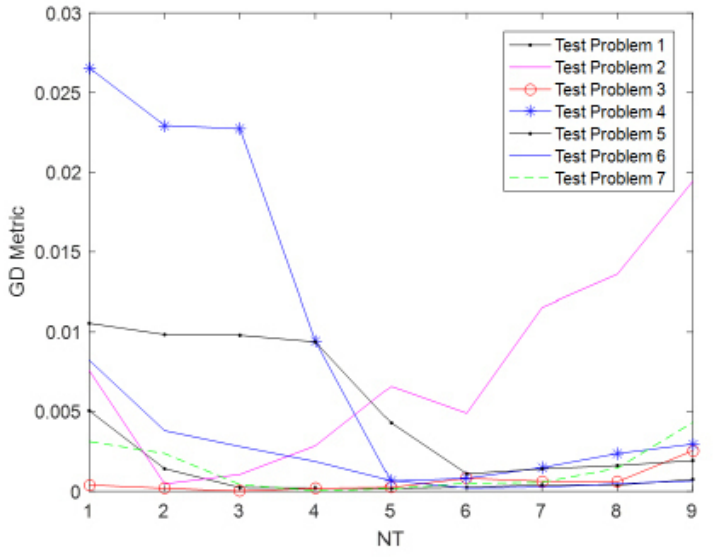
Table 16 *(Continued)*

Algorithm	NT	Test Problem		
		5	6	7
	7	1.1805E-03	6.9160E-04	3.2579E-03
	8	**3.3571E-04**	4.7393E-04	1.0370E-02
	9	7.5183E-04	2.6149E-04	1.7981E-02
FPA	1	1.0556E-02	8.2330E-03	3.1273E-03
	2	9.8460E-03	3.8064E-03	2.3766E-03
	3	9.8036E-03	2.8115E-03	4.2669E-04
	4	9.3868E-03	1.8730E-03	**3.2487E-05**
	5	4.2955E-03	6.8033E-04	1.6666E-04
	6	**1.1004E-03**	**2.0994E-04**	5.1030E-04
	7	1.4087E-03	2.5601E-04	5.6284E-04
	8	1.6190E-03	4.7710E-04	1.4553E-03
	9	1.9285E-03	6.5623E-04	4.3279E-03
GA	1	1.0556E-02	8.2328E-03	3.0958E-03
	2	9.8259E-03	3.8128E-03	2.0768E-03
	3	9.7890E-03	2.8533E-03	**2.6636E-04**
	4	9.2191E-03	2.0325E-03	4.9353E-04
	5	5.4457E-03	9.8756E-04	1.9471E-03
	6	5.7002E-04	4.5320E-04	2.2832E-03
	7	**5.5462E-04**	**2.8517E-04**	5.1471E-03
	8	7.3151E-04	3.0154E-04	3.5705E-03
	9	7.4270E-04	2.7269E-04	2.5460E-03

The bolded values are the lowest GD value for each test problem



(a)



(b)

Figure 17. Relationship of using different NTs with respect to approximation accuracies for all test problems (a) MFO; (b) FPA; (c) GA

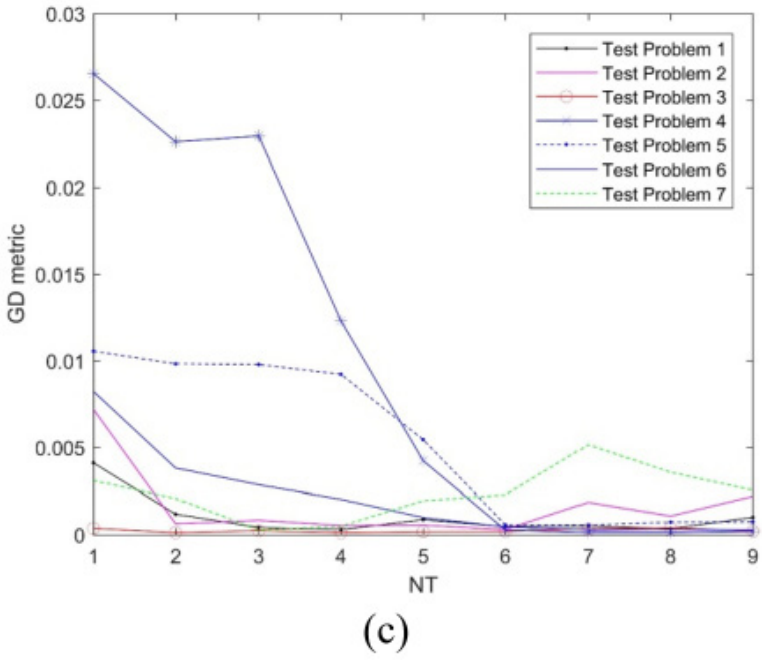


Figure 17. (Continued)

CONCLUSIONS

In this work, finding the approximate solutions of ODEs with the formulation of Fourier series expansion, calculus of variation and metaheuristic algorithm, specifically, the MFO and FPA, was addressed. The approximate approach was examined through seven test problems, including first and second-order ODEs, system of ODE, ODE with Neumann boundary condition, mechanical engineering problems and longitudinal fins heat transfer problem. It was found that all approximate solutions were able to approximate the exact solutions satisfactorily, with the GD metric within the range 1E-03 to 1E-05. Besides having reasonable accuracy, the approximate approach is promising in the sense that (i) it can be applied in a wide range of ODEs regardless of the types and orders, (ii) the procedures used to solve different ODEs can be computed using a general framework, (iii) the Fourier series can be differentiated infinitely and thus, it can be used as the approximate solution for higher-order ODEs, and (iv) the NT is adjustable to suit the personal requirement.

In terms of the dependency of approximation accuracy to the NT, it was found that the increment in NT was not proportional to the approximation accuracy. Though an increase of the NT may improve the approximation accuracy, in turn, it causes higher computational complexity due to the increasing number of Fourier coefficients. Thus, a moderate value of NT is suggested.

For future recommendations of this study, the lower and upper bounds of -1 and 1 can be increased to obtain a better approximate solution. Nonetheless, this will increase the computational time needed due to more possible coefficients to consider. Other than that, more newly invented optimization algorithms can be simulated and compared to see the feasibility of this method. Lastly, ODEs with different difficulties and fields should be tested to challenge the limits of this method.

ACKNOWLEDGEMENT

The authors would like to express the deepest appreciation to the Ministry of Education Malaysia, for funding this project through the Fundamental Research Grant Scheme (FRGS – Vot K070, Reference Code FRGS/1/2018/ICT02/UTHM/02/2). Additional support from Universiti Tun Hussein Onn Malaysia (UTHM) in the form of GPPS Vot H034 is also gratefully acknowledged.

REFERENCES

Abassy, T. A., El-Tawil, M. A., & El-Zoheiry, H. (2007). Modified variational iteration method for Boussinesq equation. *Computers and Mathematics with Applications, 54*(7), 955-965.

Adomian, G. (1988). A review of the decomposition method in applied mathematics. *Journal of Mathematical Analysis and Applications, 135*(2), 501-544.

Almasi, M. H., Sadollah, A., Kang, S., & Karim, M. R. (2016). Optimization of an improved intermodal transit model equipped with feeder bus and railway systems using metaheuristics approaches. *Sustainability, 8*(6), 1-27.

Babaei, M. (2013). A general approach to approximate solutions of nonlinear differential equations using particle swarm optimization. *Applied Soft Computing, 13*(7), 3354-3365.

Babolian, E., Azizi, A., & Saeidian, J. (2009). Some notes on using the homotopy perturbation method for solving time-dependent differential equations. *Mathematical and Computer Modelling, 50*(1), 213-224.

Beidokhti, R. S., & Malek, A. (2009). Solving initial-boundary value problems for systems of partial differential equations using neural networks and optimization techniques. *Journal of the Franklin Institute, 346*(9), 898-913.

Biazar, J., & Hosseini, K. (2017). An effective modification of adomian decomposition method for solving emden–fowler type systems. *National Academy Science Letters, 40*(4), 285-290.

Bouaziz, M., Rechak, S., Hanini, S., Bal, Y., & Bal, K. (2001). Numerical study of nonlinear heat transfer in longitudinal fins. *International Journal of Thermal Sciences, 40*(9), 843-857.

Butcher, J. C. (2016). *Numerical methods for ordinary differential equations*. Chichester, UK: John Wiley & Sons.

Chen, C. L., & Liu, Y. C. (1998). Differential transformation technique for steady nonlinear heat conduction problems. *Applied Mathematics and Computation, 95*(2), 155-164.

Coşkun, S. B., & Atay, M. T. (2007). Analysis of convective straight and radial fins with temperature-dependent thermal conductivity using variational iteration method with comparison with respect to finite element analysis. *Mathematical Problems in Engineering, 2007*, 1-15.

Coşkun, S. B., & Atay, M. T. (2008). Fin efficiency analysis of convective straight fins with temperature dependent thermal conductivity using variational iteration method. *Applied Thermal Engineering, 28*(17-18), 2345-2352.

Das, R., & Kundu, B. (2019). Forward and inverse nonlinear heat transfer analysis for optimization of a constructal T-shape fin under dry and wet conditions. *International Journal of Heat and Mass Transfer, 137*, 461-475.

Dash, P., Saikia, L. C., & Sinha, N. (2016). Flower pollination algorithm optimized pi-pd cascade controller in automatic generation control of a multi-area power system. *International Journal of Electrical Power and Energy Systems, 82*, 19-28.

Domairry, G., & Fazeli, M. (2009). Homotopy analysis method to determine the fin efficiency of convective straight fins with temperature-dependent thermal conductivity. *Communications in Nonlinear Science and Numerical Simulation, 14*(2), 489-499.

Ebaid, A. (2013). On a new differential transformation method for solving nonlinear differential equations. *Asian-European Journal of Mathematics, 6*(4), 1-12.

Ebaid, A. E. (2011). A reliable aftertreatment for improving the differential transformation method and its application to nonlinear oscillators with fractional nonlinearities. *Communications in Nonlinear Science and Numerical Simulation, 16*(1), 528-536.

El-Sayed, T. A., & El-Mongy, H. H. (2019). Free vibration and stability analysis of a multi-span pipe conveying fluid using exact and variational iteration methods combined with transfer matrix method. *Applied Mathematical Modelling, 71*, 173-193.

Fouladi, F., Hosseinzadeh, E., Barari, A., & Domairry, G. (2010). Highly nonlinear temperature-dependent fin analysis by variational iteration method. *Heat Transfer Research, 41*(2), 155-165.

He, J. H. (1999). Variational iteration method–a kind of non-linear analytical technique: some examples. *International Journal of Non-Linear Mechanics, 34*(4), 699-708.

He, J. H. (2003). Homotopy perturbation method: A new nonlinear analytical technique. *Applied Mathematics and Computation, 135*(1), 73-79.

Hussain, K., Ismail, F., & Senu, N. (2016). Solving directly special fourth-order ordinary differential equations using Runge–Kutta type method. *Journal of Computational and Applied Mathematics, 306*, 179-199.

Hussin, C. H. C., Mandangan, A., Kilicman, A., Daud, M. A., & Juhan, N. (2016). Differential transformation method for solving sixth-order boundary value problems of ordinary differential equations. *Jurnal Teknologi, 78*(6-4), 13-19.

Khani, F., Raji, M. A., & Nejad, H. H. (2009). Analytical solutions and efficiency of the nonlinear fin problem with temperature-dependent thermal conductivity and heat transfer coefficient. *Communications in Nonlinear Science and Numerical Simulation, 14*(8), 3327-3338.

Lambert, J. (1986). A stable sequence of steplengths for Euler's rule applied to stiff systems of differential equations. *Computers and Mathematics with Applications, 12*(5-6), 1141-1151.

Lee, Z. Y. (2006). Method of bilaterally bounded to solution Blasius equation using particle swarm optimization. *Applied Mathematics and Computation, 179*(2), 779-786.

Mastorakis, N. E. (2006). Unstable ordinary differential equations: Solution via genetic algorithms and the method of Nelder-Mead. *WSEAS Transactions on Mathematics, 5*(12), 1-6.

Mateescu, G. D. (2006). On the application of genetic algorithms to differential equations. *Romanian Journal of Economic Forecasting, 7*(2), 5-9.

Mirjalili, S. (2015). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems, 89*, 228-249.

Mohammadi, F., & Hosseini, M. M. (2011). A comparative study of numerical methods for solving quadratic Riccati differential equations. *Journal of the Franklin Institute, 348*(2), 156-164.

Nemati, K., Shamsuddin, S. M., & Darus, M. (2015). Solving initial and boundary value problems using learning automata particle swarm optimization. *Engineering Optimization, 47*(5), 656-673.

Panda, A., & Pani, S. (2017). Determining approximate solutions of nonlinear ordinary differential equations using orthogonal colliding bodies optimization. *Neural Processing Letters, 48*(1), 219-243.

Panda, A., & Pani, S. (2018). Determining approximate solutions of nonlinear ordinary differential equations using orthogonal colliding bodies optimization. *Neural Processing Letters, 48*(1), 219-243.

Perfilov, V., Fila, V., & Marcano, J. S. (2018). A general predictive model for sweeping gas membrane distillation. *Desalination, 443*, 285-306.

Rudd, K., & Ferrari, S. (2015). A constrained integration (CINT) approach to solving partial differential equations using artificial neural networks. *Neurocomputing, 155*, 277-285.

Sadollah, A., Choi, Y., & Kim, J. H. (2015a). Metaheuristic algorithms for approximate solution to ordinary differential equations of longitudinal fins having various profiles. *Applied Soft Computing, 33*, 360-379.

Sadollah, A., Choi, Y., Yoo, D. G., & Kim, J. H. (2015b). Metaheuristic algorithms for approximate solution to ordinary differential equations of longitudinal fins having various profiles. *Applied Soft Computing, 33*, 360-379.

Sadollah, A., Eskandar, H., & Kim, J. H. (2015c). Approximate solving of nonlinear ordinary differential equations using least square weight function and metaheuristic algorithms. *Engineering Applications of Artificial Intelligence, 40*, 117-132.

Sadollah, A., Eskandar, H., Yoo, D. G., & Kim, J. H. (2015d). Approximate solving of nonlinear ordinary differential equations using least square weight function and metaheuristic algorithms. *Engineering Applications of Artificial Intelligence, 40*, 117-132.

Sadollah, A., Yadav, N., Gao, K., & Su, R. (2017). Metaheuristic optimisation methods for approximate solving of singular boundary value problems. *Journal of Experimental and Theoretical Artificial Intelligence, 29*(4), 823-842.

Tian, X., Pang, W., Wang, Y., Guo, K., & Zhou, Y. (2019). LatinPSO: An algorithm for simultaneously inferring structure and parameters of ordinary differential equations models. *BioSystems, 182*, 8-16.

Torabi, M., Aziz, A., & Zhang, K. (2013). A comparative study of longitudinal fins of rectangular, trapezoidal and concave parabolic profiles with multiple nonlinearities. *Energy, 51*, 243-256.

Turkyilmazoglu, M. (2018). A reliable convergent Adomian decomposition method for heat transfer through extended surfaces. *International Journal of Numerical Methods for Heat and Fluid Flow, 28*(11), 2551-2566.

Yang, X. S. (2012, September 3-7). Flower pollination algorithm for global optimization. In *International Conference on Unconventional Computing and Natural Computation* (pp. 240-249). Orléans, France.

Pertanika J. Sci. & Technol. 28 (4): 1267 - 1275 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Fiber Twist-based Wavelength Tunability in Tapered Optical Fiber Filters

Maisarah Mansor[1], Afiqah Mohd Nawi[1], Nadiah Husseini Zainol Abidin[1], Muhammad Firdaus Omar[2], Mohd Adzir Mahdi[1] and Muhammad Hafiz Abu Bakar[1]*

[1]Wireless and Photonics Networks Research Center, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Department of Physics, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM, Skudai, Johor Bahru, Johor, Malaysia

ABSTRACT

This work demonstrates the tunability of tapered single mode fiber (SMF) and tapered polarization maintaining fiber (PMF) filters based on fiber twisting method. One end of the tapered fiber was twisted from 0° until 100° using the Vytran Fiber Processing System. Observation on the spectral output shows that the fiber twisting technique is a viable option to impart tunability in tapered fibers with total shift of 15 and 10 nm, respectively for SMF and PMF. Better tunability is observed in the SMF filter due to its simple physical structure and more straightforward interferometry effect but a significantly higher extinction ratio is observed for the PMF filter. Both filters exhibited region of linear wavelength shift with corresponding R^2 values of 0.9924 and 0.9294 for SMF and PMF. The simplicity and reliability of the filter may pave the way for the development of a practical and compact tunable all-fiber filter for laser systems.

Keywords: Fiber twist, optical filter, tapered fiber, wavelength tunable

ARTICLE INFO

Article history:
Received: 4 April 2020
Accepted: 27 July 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.07

E-mail addresses:
maisarahmansorkmpk@gmail.com (Maisarah Mansor)
afiqahmohdnawi@gmail.com (Afiqah Mohd Nawi)
hza_nadiah@upm.edu.my (Nadiah Husseini Zainol Abidin)
firdausomar@utm.my (Muhammad Firdaus Omar)
mam@upm.edu.my (Mohd Adzir Mahdi)
mhab@ieee.org (Muhammad Hafiz Abu Bakar)
*Corresponding author

INTRODUCTION

In the past few decades, fiber optics technology has been a major research and development area, especially in the communication field. The ultimate aim in

ISSN: 0128-7680
e-ISSN: 2231-8526

optical fiber development is to produce all-fiber devices or components to achieve coupling simplicity, ease of use, and compactness. This effort has led to the production of fiber-based component not only for communication, but also in sensing applications (Alberto et al., 2018; Correia et al., 2018; Pospíšilová et al., 2015).

One of the focus is on the wavelength-selective comb filter, which features operating simplicity and multiple wavelengths selection. A spectral output consisting of periodic resonant frequencies is conceived post-filtering process, whereby the distance between the two neighbouring frequencies is known as the free spectral range (FSR) (Jung et al., 2010). The filter is typically integrated into a laser cavity to create multiple seed signals necessary for the generation of multiwavelength laser (Martinez-Rios et al., 2014; Srivastava et al., 2014). One desired parameter in optical filter is wavelength tunability, which allows variation to the filtered wavelengths based on the intended use. Several fiber-based filters have been proposed with varying success in terms of tunability. One of them is the PMF-based Lyot filter, which unfortunately has limited tunability feature due to its dependence on the physical length difference between the two PMF sections (Fok & Ge, 2017). Tunable double Sagnac loops in a ring erbium-doped fiber (EDF) laser (Wang et al., 2012) is also an alternative wavelength-selective filter but it shows irregular output spectrum attributed to inaccurate refractive index difference of PMFs, inaccurate length of PMFs, and splicing loss. Other than that, polarization independent tunable all-fiber comb filter based on a modified dual-pass MZI can only attain discrete tunability of the spectral spacing (Zhi-Chao et al., 2009).

A simpler alternative proposed in recent times is the tapered optical fibers. Tunability in tapered fibers is acquired through its high sensitivity towards environmental changes such as stress, strain, and temperature (Da Silveira et al., 2015; Ali et al., 2014; Musa et al., 2016). The stretching of the fiber down to tens of micrometres makes it more pliable which enables bending at a much larger angle without physical damage. In Ali et al. (2014), wavelength tunability was achieved simply by varying the bend radius of the tapered region. However, a practical mechatronic mechanism to induce such bend has yet to be developed and the overall size of the device must consider the largest radius of the bend. Additionally, increasing the amount of bending on the taper leads to a higher loss in the system. Application of axial strain to achieve wavelength tunability has been tested on single taper (Kieu & Mansuripur, 2006) as well as cascaded taper (Jaddoa et al., 2016). Both studies demonstrated tuning range that was dependent on the stretching distance accorded by the linear stage thus longer motorized stage will be required for wider tuning range. In Selvas-Aguilar et al. (2014), temperature-based wavelength variation was employed for tapered fiber immersed in glycerol. Careful consideration, however, must be given to the surrounding media as glycerol may degrade at high temperature and could also generate micro-bending in the tapered fiber section.

In general, strain-based tuning is the preferred method as it is more straightforward, and variation of wavelength can be achieved faster compared to temperature-based approach. Nevertheless, device utilizing such technique must consider the range of motion of the strain-inducing mechanism. In this work, tapered fiber-based filter with wavelength tunability based on fiber twist concept is investigated. Stationary rotation stages were employed to twist the fiber-under-test. Two types of fiber; SMF and PMF were tapered, tested, and compared to observe any performance differences. The findings determined that both filters can utilize the proposed method to induce tension-based tunability. This may lead to the development of a practical all-fiber taper filter with precise and compact wavelength tunability mechanism.

MATERIALS AND METHODS

In this work, the proposed fiber twist method was deployed on Corning SMF-28 fiber to elucidate its effect towards basic tapered fiber filter while 30 cm Fujikura Panda-type PMF with corresponding birefringence and attenuation of 4.5×10^{-4} and 0.5 dB/km was included to ascertain any improvement accorded by the more complex birefringence-influenced effect in the fiber. Prior to tapering, the protective coating was removed for both the SMF and PMF. A stripper was used for the SMF while the protective coating of the PMF was removed by dipping the fiber into hydrochloric acid for about 1.5 hours. This was necessary as the the double acrylate coating of the PMF is difficult to remove using a stripper. Then, the fiber was placed onto the fiber holder block (FHB) of Vytran GPX-3400 Optical Glass Processing Workstation for the tapering process. This machine can perform precise fusion splicing and tapering with its filament furnace assembly and precision stages. Its real-time control system allows manipulation of dimensions, uniformity, and reproducibility of the fabricated taper as the pulling speed and heat are kept at a constant value of 1 mm/s and 42 W, respectively. Figure 1 illustrates the geometry of the non-adiabatic tapered optical fiber used throughout the experiment, with parameters of waist diameter, waist length, up-taper length and down-taper length set at 10 μm, 12 mm, 5 mm and 5 mm following previously reported work (Musa et al., 2016).

The tapered fiber was kept on the FHB and both ends of the tapered fiber were spliced to single-mode fiber pigtail. One end was connected to a C+L-band light source while the other end to an optical spectrum analyzer (OSA), as shown in Figure 2(a). One end of the tapered fiber was then twisted by the Vytran FHB, which can rotate up to 120°. In this experiment, Section 1, which was connected to the broadband light source was rotated anticlockwise from 10° up to 100° with a step angle of 10° while Section 2 was fixed as shown in Figure 2(b). In order to observe the correlation between tension and wavelength shift, the spectral output for each step angle rotation was acquired via the OSA while the tension on the fiber was recorded manually based on the readings provided by the Vytran

Tension Monitoring System, which observes linear tension subjected on the FHB relative to its initial condition (Thorlabs, 2015).

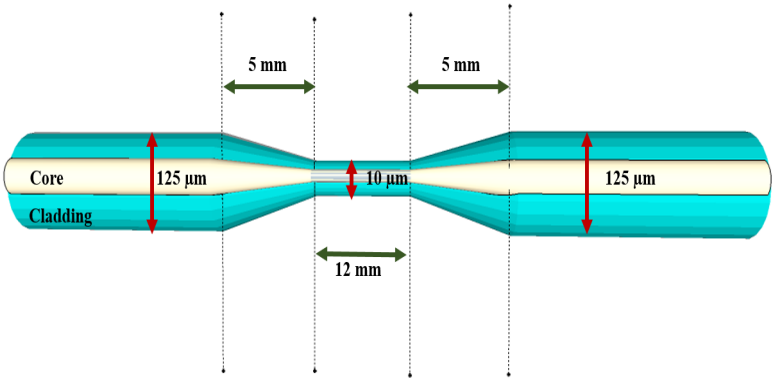


Figure 1. Structure and dimensions of fabricated tapered optical fiber

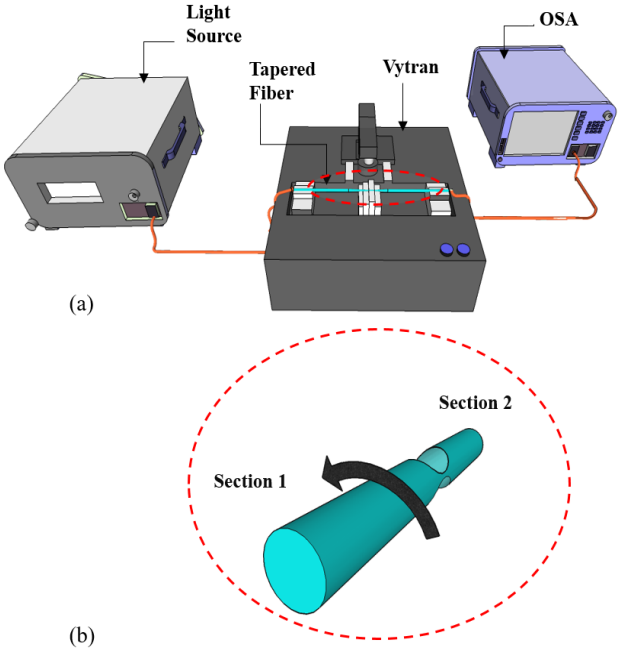


Figure 2. (a) Experimental setup and (b) illustration of fiber twist

RESULTS AND DISCUSSION

During tapering process, the fiber is heated at high temperature and stretched to form the tapered region. When stretched, the core and cladding fuse together, creating a new core and making the surrounding air as the new cladding. As light propagates from the single mode fiber to the down-taper region, the dramatic decrease in core diameter will excite higher order modes that propagate along the tapered region. At the end of the tapered region, the modes recombine, and the phases interfere. A comb-like spectrum is then produced due to the constructive and destructive interference of modes (Jung et al., 2010).

The interference yields periodic fringes as shown in Figure 3, which exhibits the insertion loss of the fabricated taper filters derived from the difference between spectral output with and without taper. The C-band insertion loss for SMF filter and PMF filter recorded average values of 6.2 dB and 7.8 dB, respectively while the free spectral range (FSR) for both filters was about 8 nm. Insertion loss of the PMF filter is higher due to the coupling mismatch between PMF fiber to the SMF pigtail (Jung et al., 2010). The 3-dB linewidths of the SMF and PMF filters are 3.3 nm and 3.9 nm, respectively with corresponding average extinction ratios of 6.5 dB and 18.0 dB, respectively. The significant difference between the extinction ratios can be attributed to the strong birefringence property of PMF and demonstrates the better capability of PMF filters to suppress unwanted signal (Piekarek et al., 2017).

Figure 4 shows the preliminary results of the spectral profile for SMF and PMF filters with twist angles of 40°, 50°, 60° and 70°. The spectra depict gradual shifting to the left or towards the shorter wavelength with larger twist angle. No significant loss of power was

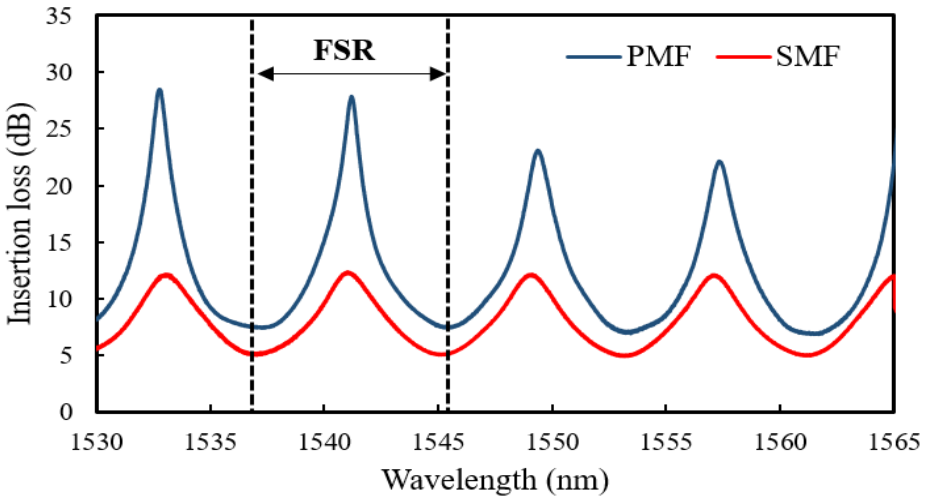


Figure 3. The insertion loss of SMF and PMF taper filters

observed during the shifts as no change was introduced to the critical angle involved in total internal reflection.

Figure 5(a) and 5(b) show the correlation between wavelength shift and tension to the angle of twist for both SMF and PMF filters. The wavelength shifts are measured by subtracting the peak or depth of the shifted spectrum against the same peak or depth of the reference spectrum. The measurement for tension was calculated by subtracting the tension reading against the reference tension when the fiber was twisted.

The wavelength shift curve in Figure 5(a) shows a minimal shift of SMF filter at 0^{o} to 20^{o} twist angle which then turns linear afterwards for 15 nm shift and eventually plateaud after reaching 80°. The reason for this trend can be observed from the tension

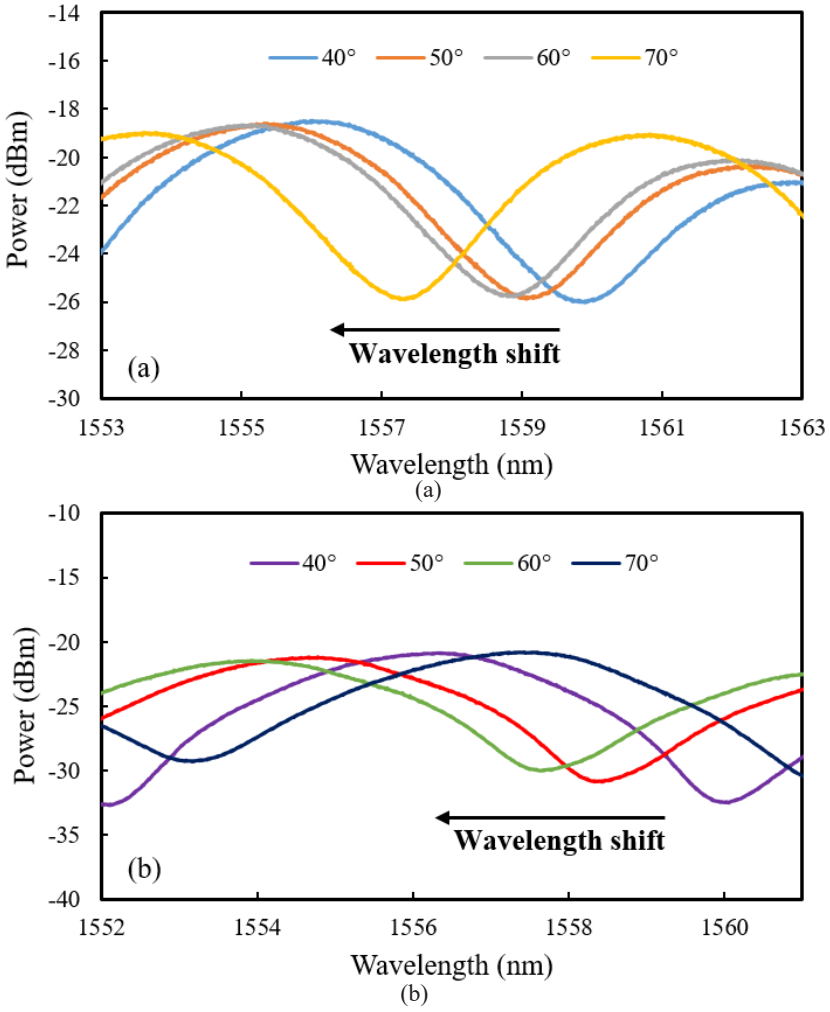


Figure 4. Spectral observation during fiber twist for (a) SMF and (b) PMF taper filters

curve, which depicts an inverted profile with similar behavior. Small twist angle produces minimal difference in tension thus explaining the lack of wavelength shift while at the larger twist angle, the wavelength shift lessens as the tension value saturates. A similar trend is observed for the PMF filter in Figure 5(b) albeit with smaller range of linear shift (20° to 60° twist angle), smaller total wavelength shift (10 nm), and higher tension, which could be attributed to the more complex PMF physical structure and light interference. The same reasons could also contribute to its lower linear region R^2 value of 0.9294 compared to the R^2 of SMF filter at 0.9924.

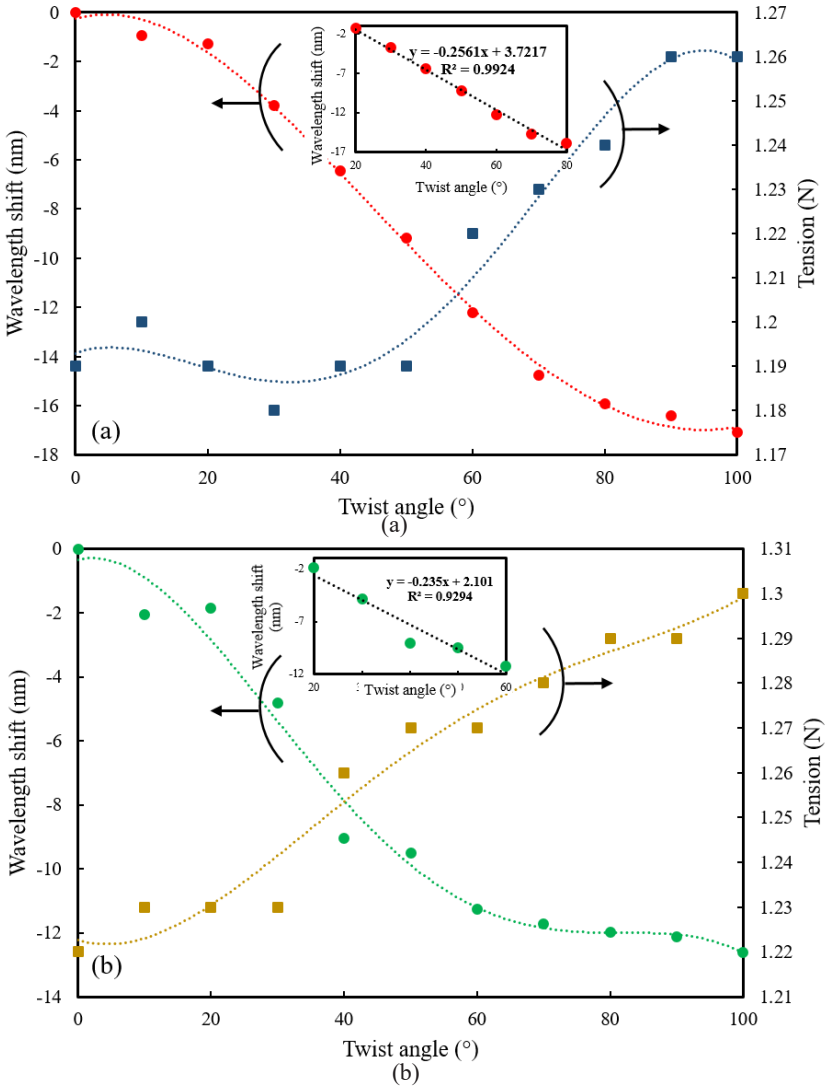


Figure 5. Wavelength shift and tension with respect to the angle of rotation for (a) SMF and (b) PMF taper filters. Insets show close-ups of linear shift region of the filters

CONCLUSIONS

In conclusion, the use of fiber twist method to achieve wavelength tunability in tapered SMF and PMF filters has been proven feasible. The shift of wavelength was correlated to the tension induced during the twisting process and region of linear shift was observed for both filters. Wider tunability range and better stability were observed for tapered SMF filter but stronger suppression capability can be obtained from the tapered PMF. The use of commercially available motorized rotation stage can be considered to develop a practical fiber filter with compact form factor and precise continuous wavelength control.

ACKNOWLEDGEMENTS

This work has been supported by Universiti Putra Malaysia under Geran Universiti Putra Malaysia #GP-IPS/2018/9607900 and Graduate Research Fellowship Scheme.

REFERENCES

Alberto, N., Domingues, M., Marques, C., André, P., & Antunes, P. (2018). Optical fiber magnetic field sensors based on magnetic fluid: A review. *Sensors, 18*(2), 1-27.

Ali, M. M., Ibrahim, S. A., Bakar, M. A., Noor, A. S. M., Anas, S. A., Zamzuri, A. K., & Mahdi, M. A. (2014). Tapered-EDF-based Mach–Zehnder interferometer for dual-wavelength fiber laser. *IEEE Photonics Journal, 6*(5), 1-9.

Correia, R., James, S., Lee, S., Morgan, S., & Korposh, S. (2018). Biomedical application of optical fibre sensors. *Journal of Optics, 20*(7), 1-26.

Da Silveira, C., Costa, J., RoccoGiraldi, M., Jorge, P., López Barbero, A., & Germano, S. (2015). Bent optical fiber taper for refractive index measurements with tunable sensitivity. *Microwave and Optical Technology Letters, 57*(4), 921-924.

Fok, M., & Ge, J. (2017). Tunable multiband microwave photonic filters. *Photonics, 4*(4), 1-20.

Jaddoa, M., Razak, M., Salim, M., Sharbirin, A., Nayan, N., Ismail, M., & Ahmad, H. (2016). Tunable single wavelength erbium-doped fiber ring laser based on in-line Mach-Zehnder strain. *Optik, 127*(20), 8326-8332.

Jung, Y., Brambilla, G., & Richardson, D. (2010). Polarization-maintaining optical microfiber. *Optics Letters, 35*(12), 2034-2036.

Kieu, K., & Mansuripur, M. (2006). Tuning of fiber lasers by use of a single-mode biconic fiber taper. *Optics Letters, 31*(16), 2435-2437.

Martinez-Rios, A., Anzueto-Sanchez, G., Monzon-Hernandez, D., Salceda-Delgado, G., & Castrellon-Uribe, J. (2014). Multiwavelength switching of an EDFL by using a fixed fiber-comb filter and a broadband tunable S-bent fiber filter. *Optics and Laser Technology, 58*, 197-201.

Musa, B., Kamil, Y. M., Bakar, M. H. A., Noor, A. S. M., Ismail, A., & Mahdi, M. A. (2016). Effects of taper parameters on free spectral range of non-adiabatic tapered optical fibers for sensing applications. *Microwave and Optical Technology Letters, 58*(4), 798-803.

Piekarek, M., Bonneau, D., Miki, S., Yamashita, T., Fujiwara, M., Sasaki, M., … & Thompson, M. (2017). High-extinction ratio integrated photonic filters for silicon quantum photonics. *Optics Letters, 42*(4), 815-818.

Pospíšilová, M., Kuncová, G., & Trögl, J. (2015). Fiber-optic chemical sensors and fiber-optic bio-sensors. *Sensors, 15*, 25208-25259.

Selvas-Aguilar, R., Martínez-Rios, A., Anzueto-Sánchez, G., Castillo-Guzmán, A., Hernández-Luna, M., & Robledo-Fava, R. (2014). Tuning of an erbium-doped fiber ring laser based on heating a tapered fiber filter. *Optical Fiber Technology, 20*(4), 391-394.

Srivastava, D., Bhatnagar, R., Kumar, A., & Parmar, V. (2014). Intensity modulation using chirped fiber bragg grating as an edge filter for temperature sensing. *Microwave and Optical Technology Letters, 56*(12), 2913-2915.

Thorlabs. (2015). *GPX3400 and GPX3600 glass processors: User guide*. New York, USA: Thorlabs.

Wang, T., Miao, X., Zhou, X., & Qian, S. (2012). Tunable multiwavelength fiber laser based on a double Sagnac HiBi fiber loop. *Applied Optics, 51*(10), C111-C116.

Zhi-Chao L., Ai-Ping L., & Wen-Cheng X. (2009). Polarization-controlled tunable all-fiber comb filter based on a modified dual-pass mach–zehnder interferometer. *IEEE Photonics Technology Letters, 21*(15), 1066-1068.

Pertanika J. Sci. & Technol. 28 (4): 1277 - 1296 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Performance of Mid-Size Combine Harvester of Grain Corn on the Field Efficiency and Energy Consumption at the Northern Johor of Malaysia

Mohamad Hairie Masroon[1], Nazmi Mat Nawi[1,2]*, Azmi Yahya[1,2], Mohamad Firdza Mohamad Shukery[1,2] and Mohamed Ezzeldien Salih Amin[1]

[1]Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Smart Farming Technology Research Centre, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

A mid-size combine harvester with 2.76 m reaping width and 103.53 hp engine output has been employed in grain corn production, especially by small-scale grain corn farmers. This study attempted to determine field performances of a typical mid-size combine harvester by measuring its effective field capacity (EFC), field efficiency (FE), fuel consumption (FC) and field machine index (FMI). Different types of energy inputs such as fuel, machinery, human, included direct, indirect, renewable and non-renewable energy involved in grain corn harvesting were also measured. The field measurements were carried out in 3 ha of grain corn farm, under similar field conditions using a typical mid-size combine harvester. The average values of EFC, FE, FC and FMI for the mid-size combine harvester were found to be 0.23 ha/h, 34.97%, 37.25 lit/ha and 0.91, respectively. The average equivalent energy values of fuel, machinery and human energy were 1780.70 MJ/ha, 587.73 MJ/ha and 8.53 MJ/ha, respectively. The average values of the direct and indirect energy were 1789.23 MJ/ha and 587.73 MJ/ha, respectively. The average values of renewable and non-renewable energy were recorded at 8.53 MJ/ha and 2368.42 MJ/ha, respectively. The mid-size combine

ARTICLE INFO

Article history:
Received: 9 April 2020
Accepted: 12 June 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.08

E-mail addresses:
hairiemasroon@yahoo.com.my (Mohamad Hairie Masroon)
nazmimat@upm.edu.my (Nazmi Mat Nawi)
azmiy@upm.edu.my (Azmi Yahya)
firdza@upm.edu.my (Mohamad Firdza Mohamad Shukery)
eezzfree@yahoo.com (Mohamed Ezzeldien Salih Amin)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

harvester investigated in this study exhibited good field performance characteristic using a reasonable amount of energy consumption as compared to harvesting operation for other grain crops. From the results, it can be concluded that good practice in harvesting operation could improve field performance, and minimise operational costs and energy consumption.

Keywords: Crop, farm, fuel consumption, machinery, speed, working width

INTRODUCTION

Corn or maize (*Zea mays* L.) is the third largest agricultural crop grown in the world after wheat and rice (Nor et al., 2019). In Southeast Asia, among the main producing countries for grain corn is Indonesia, Philippine, Thailand and Vietnam with the production values of 11.9, 8.2, 5.3 and 3.95 million tons, respectively (USDA, 2018). In Malaysia, however, there is no commercial corn production for feed that has been established. In 2017, Malaysia imported around 3.7 million tons of grain corn valued about RM3 billion (US$ 737 million) (UNC, 2019). In 2016, Malaysia imported 4.1 million tons of grain corn from Thailand, India, Australia, Brazil, Argentina, Paraguay and the United States of America (UNC, 2019). Due to this huge amount of the import value, the Malaysian government has taken initiatives to encourage local production of 30% grain corn to fulfil the demand for domestic consumption (Nor et al., 2019).

One of the strategies to increase the production of field corn is by utilizing a proper information, management, technology and mechanization system (Aribe et al., 2019). Mechanization technology in agriculture which is economical and suitable to local physical and climatic conditions could potentially offer a long-term sustainable production for field corn, especially in dealing with labour shortage (Hamid et al., 2018). A proper mechanization package is also important to ensure grain corn can be harvested in a timely manner, avoiding grain losses due to unfavourable weather condition at a farm (Busato et al., 2007). Hence, the usage of a mid-size combine harvester by farmers could be considered as a good step to overcome labour shortage and improve operational time of machine in the corn fields (Hamid et al., 2018).

The analysis and prediction of agricultural machinery performance are essential in machinery management. The field performance and energy consumption are two important parameters in determining the efficiency and field capacity of a combine harvester (Busato et al., 2007). The working time required to perform any agricultural operation mostly depends on the design and technological parameters of the machinery used. For example, an increase in machine's working width and speed could increase the field efficiency and reduce the working time (Sarauskis et al., 2014). The working pattern during harvesting also has a great influence on the total time of some additional activities that are computed in the field (Busato et al., 2007). Although some of the non-productive activities are unavoidable,

optimising field performance is intended to minimise the sum of these non-productive activities (Henrichsmeyer & Ohls, 1995). Minimisation of the non-productive time, fuel consumption, in-field travelled distance, or the excessive wheeling of the field, may result in significant economic and environmental benefit (Bochtis et al., 2007).

Energy efficiency in machinery management is also one of the significant factors in agricultural production (Canakci et al., 2005). Efficient use of energy may increase productivity, profitability, sustainability and competitiveness of agricultural sector (Singh et al., 2002). Total physical energy input consists of human power, animal power, machinery power, electricity and fuel consumptions. In general, energy requirements in agriculture can be divided into two groups: direct and indirect energy (Ozkan et al., 2004a). Direct energy is consumed in the farm, in the form of energy products, such as fuel, lubricants and labour. Whereas, indirect energy is consumed outside the farm boundaries to produce any input used in the farm such as machinery and chemicals (Gemtos et al., 2013). The ability to identify and quantify different form of energies involved in field corn harvesting may help farmers to increase energy efficiency, thus minimise the production cost.

However, there are insufficient scientific researches that have been accomplished to quantify field performances and energy consumption for a mid-size combine harvester for harvesting grain corn in Malaysia. Therefore, the goal of this research is to evaluate the field performances and energy consumption for a typical mid-size combine harvester utilized for grain corn harvesting under Malaysian farm condition. The specific objective is to measure effective field capacity, field efficiency, human, fuel and machinery energy.

MATERIALS AND METHODS

Study Area

This research was conducted at a private field corn located in Labis, Johor, Malaysia ($2^{o}21'39"$ N – 102^{o} 56' 13" E). From the entire farm area (approximately 10 ha), 3 ha of the farm area which consisted of four plots were randomly selected for data collection during two consecutive dry harvesting seasons. The area of each plot was 0.75 ha consisting of three subplots (0.25 ha), which divided according to CRD (completely randomized design) method. All harvested plots were adjacent to each other in the same location and subjected to the same weather (sunny day) and terrain slopes ranging from 0^{o} to 30^{o} during harvest. The principal soil type in terms of soil texture found in this area was sandy loam. The area for each subplot was measured using measuring tape (Smith et al., 1994).

Combine Harvester

A mid-size combine harvester selected in this study (2.76 m reaping width and 103.53 hp engine output) was manufactured by Kubota (DC 105X) (Figure 1 & 2). This model was predominantly used by Malaysian Department of Agriculture (DOA) and National Farmers'

Association (NAFAS) for grain corn harvesting in their trial plots. The specification of the mid-size combine harvester is shown in Table 1.

Figure 1. Mid-size combine harvester in operation (model Kubota DC 105X)

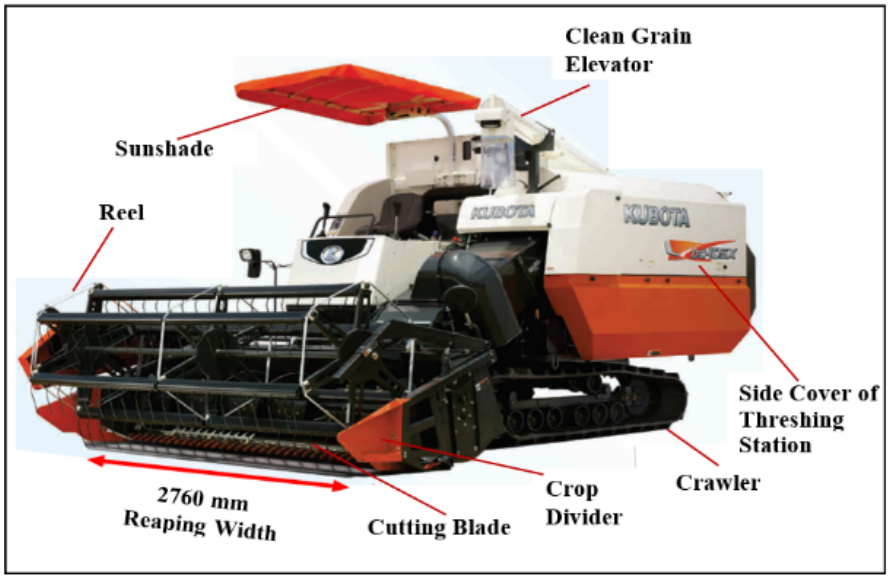


Figure 2. Typical parts of mid-size combine harvester (model Kubota DC 105X) (Sime Kubota Sdn Bhd, 2019)

Table 1
Specification of mid-size combine harvester for grain corn harvesting

Specification	Value
Brand	Kubota
Model	DC 105X-V3800DI-TIE2-CT
Overall length, mm	5460
Overall width, mm	3045
Overall height, mm	3040
Minimum ground clearance, mm	325 to 425
Weight, kg	4630
Total displacement, cc	3769
Engine output, kW	77.2
Rotation speed, rpm	2600
Fuel	Diesel
Fuel tank capacity, lit	105
Reaping width, mm	2760
Blade width, mm	2667
Grain tank capacity, lit	2350
Operator	1 person

Measurement of Operational Times

The time taken for each task involved during harvest operation were recorded manually using a stopwatch. These tasks included the harvesting time (time spent in performing the actual harvesting operation by cutting the plant), cornering and reversing time (time spent in turning and by the combine harvester without cutting the plant), unloading time (time spent in unloading the grain tank of the combine harvester and going to or from corn conveying trucks) and others (time taken in refuelling, adjusting or setting the machine). The total field time represents the time spent when a combine harvester engine was turned on, run and turned off once the job was completed (Olt et al., 2019). The data were used to compute the field performances which included forward speed (FS), effective field capacity (EFC), field efficiency (FE) and field machine index (FMI). FS of the combine harvester was obtained by recording the time taken to travel at a measured distance. The distance was measured with a measuring tape and the time was counted using a stopwatch. Such procedure was carried out in 12 repetitions to have the average theoretical speed by dividing the summation of total value of speed for all rows by the number of rows (Smith et al., 1994).

Determination of Fuel Consumption

The fuel consumption for the mid-size combine harvester was determined by measuring the difference of the fuel inside the fuel tank before and after the operation. The measurement of fuel consumption was conducted by refilling the fuel tank back to its full capacity after the harvesting operation using a measuring cylinder for each harvesting replications (Amponsah et al., 2017). The fuel consumption for each subplot was measured and the data characterizing fuel consumption were compared in lit/ha. The comparison of fuel consumption in lit/ha was a simple and cost-effective way to measure the fuel consumption of agricultural machinery (Jokiniemi et al., 2012).

Determination of Field Efficiency

Field efficiency (FE) is defined as the percentage of the time when the machine is operated at its full rated speed and width in the field under actual working condition (Nasri et al., 2016). FE describes how effective the time is spent to do the work and it becomes the ratio between the harvester's productivity under actual working conditions and the theoretical maximum possible productivity (Grisso et al., 2004). The FE is calculated based on the effective field capacity (EFC) and theoretical field capacity (TFC). EFC means the ability of a combine harvester to harvest crop under the actual field condition while TFC is a theoretical field capacity which was obtained from the width of machine and multiplied with the average forward speed (FS) during the field work (Zhou, 2012). The FE, TFC and EFC are calculated using Equation 1-3 (ASABE, 2005).

$$\mathrm{FE} = \frac{\mathrm{EFC}}{\mathrm{TFC}} \quad [1]$$

Where FE is field efficiency (%), EFC is an effective field capacity (ha/h) and TFC is a theoretical field capacity (ha/h).

$$\mathrm{TFC} = \frac{\mathrm{W} * \mathrm{S}}{10} \quad [2]$$

Where W is width of machine (m) and S is speed of machine (km/h).

$$\mathrm{EFC} = \frac{\mathrm{A}}{\mathrm{T}} \quad [3]$$

Where A is the harvested area (ha) and T is working time (h) which represents time spent in performing the operation from the beginning to the end which includes harvesting, turning, unloading and others for instance like machine setting, adjustment and refuelling.

Determination of Field Machine Index

Field machine index (FMI) is the index indicating the turning effectiveness of a combine harvester (Equation 4) (Wagiman et al., 2019).

$$FMI = \frac{EOT}{(EOT + \text{Turning time})} \quad [4]$$

Where EOT is effective operating time which means harvesting (h). Turning time includes cornering and reversing time is non-productive time (h).

Energy Sources of Field Corn Harvesting

Energy Conversion Coefficient. The recorded farm inputs namely machinery, fuel and human utilized during the harvesting operation were converted into equivalent energy values in MJ/ha by using specific conversion coefficients. Energy conversion coefficient is a value which expresses the energy input expended in the production and distribution of a unit physical material (Elsoragaby et al., 2019a). The equivalent energy sources of the harvesting operation in grain corn production were estimated by classical mathematical Equation 5-8.

Machinery Energy. Machinery energy is an indirect energy assumed to be embodied in a piece of equipment during manufacturing (Elsoragaby et al., 2019a). To compute the machinery energy, the total useful life and EFC of the machine were taken into consideration (Muazu et al., 2014a). The weight of the machine was included by equally distributing its weight over the total economic life of it. The general expression used to compute machinery energy is given in Equation 5 (Gezer et al., 2003).

$$ME = \frac{Cf * W}{EFC * L} \quad [5]$$

Where ME is machinery energy (MJ/ha), Cf is the energy conversion coefficient for the combine harvester, W is the weight of the combine harvester (kg), EFC is the effective field capacity (ha/h) and L is the economic life of the combine harvester (h). The 3000 hours of useful life for self-propelled combine harvester is adopted from ASABE Standard, ASAE D497.7 (ASABE, 2011). The machinery energy conversion factor used for the combine harvester was 87.63 MJ/kg (Muazu et al., 2014a).

Fuel Energy. Fuel energy per unit area is a function of the type and quantity of fuel consumed by the machinery used by the farmers. The quantity of fuel consumed by the machinery used to power engines in performing various operations in crop cultivation system was multiplied by an energy conversion factor for the fuel. Thus, the general

expression used to compute fuel energy is given as follows (Equation 6) (Muazu et al., 2014a):

$$FE = \frac{Fcon * Fc}{A} \quad [6]$$

Where FE is fuel energy (MJ/ha), Fcon is the quantity of fuel consumed (lit), Fc is the fuel energy conversion coefficient (MJ/lit) and A is the farm area covered (ha). The fuel energy conversion factor used for the combine harvester is 47.80 MJ/kg (Canakci et al., 2005).

Human Energy. Human energy expenditure in harvesting operation was evaluated based on the number of farm workers engaged in harvesting operation per unit area and the time spent in performing the operations multiplied by an energy conversion coefficient. The general expression used to compute human energy is given in Equation 7 (Muazu et al., 2014a).

$$HE = \frac{n * H * lc}{A} \quad [7]$$

Where HE is human energy (MJ/ha), n is the number of workers engaged in an operation, H is the duration of operation (h), lc is the energy conversion coefficient for human labour and A is the farm area covered (ha). The human energy conversion factor used for the combine harvester is 1.96 MJ/kg (Canakci et al., 2005).

Total Energy Input. Total energy input in harvesting operation per hectare was determined as the summation of energy from all the sources which is as follows (Equation 8) (Elsoragaby et al., 2019a):

$$TEI = ME + FE + HE \quad [8]$$

Where TEI is total energy input of harvesting operation (MJ/ha) and ME, FE, HE, are as previously defined.

Energy Forms

Direct and Indirect Energy. There were two types of energy inputs in agricultural productions namely the direct and indirect energy. Direct energy refers to the energy which was directly used in the field while indirect energy refers to the energy which was not directly used in the field (Ozkan et al., 2004b). In harvesting operation, direct energy was obtained through the summation of fuel and human energy while indirect energy was obtained from the machinery energy (Gemtos et al., 2013).

Renewable and Non-renewable Energy. Energy inputs in crop production can also be grouped into renewable and non-renewable energy. The renewable energy sources can

be replenished over time while non-renewable energy sources are depleted with time. In harvesting operation, the human energy is considered as renewable energy while the fuel and machinery energies are considered as non-renewable energy (Mohammadi et al., 2010).

Statistical Analysis

Statistical analysis was carried out using SAS v.9.1 software (SAS Institute Inc., Cary, NC). The variables were assessed for normal distribution and were log-transformed to fit a normal distribution. The one-way analysis of variance (ANOVA) was conducted using SAS GLM procedure and the mean results were compared for P-value with 95% confidence and 5% significance level ($\alpha = 0.05$).

RESULTS AND DISCUSSION

Field Performance Analysis

Field performance measurement is a very critical requirement in order to evaluate the effectiveness of the mid-size combine harvester employed during harvest in a field corn. Table 2 states that there are no significant differences for each parameter in the field performance of grain corn harvesting at 5% significance level ($P \leq 0.05$) between harvesting plots.

Effective Field Capacity and Field Efficiency

Table 2 illustrates EFC and FE of the mid-size combine harvester employed for grain corn harvesting in this study. There were no major mechanical or operational problems observed during harvesting which could affect either EFC or FE. Table 2 indicates that the mid-size combine harvester with 2.67 m working width had 0.23 ha/h and 34.97% of average EFC and FE, respectively. Previous research from Malaysian Agricultural Research and Development Institute (MARDI) showed that the mid-size combine harvester (Kubota DC 70G) with 1.98 m working width had EFC and FE of 0.32 ha/h and 62.79%, respectively (Hamid et al., 2018). Recently, the field performance of the mid-size combine harvester (WS 7.0Plus) with 2.2 m working width utilised in a paddy field was studied, and the EFC and FE for the machine were found to be 0.53 ha/h and 72% respectively (Elsoragaby et al., 2019a). In a similar study on wheat production, a mid-size combine harvester with 1.10 m working width had a 60% of mean FE and 0.86 ha/h of mean EFC (Chegini & Mirnezami, 2016).

Lower average values for FE and EFC were found in this study as compared to the previous studies due to lower FS (2.55 km/h) applied in this study to avoid over loading of the feeding rate, and thus reducing grain losses. Unsuitable FS poses high possibility in reducing the grain yield due to uncontrollable losses. Plot 4 with mean FS of 2.56 km/h had 9.09% greater mean EFC than plot 1 with 2.37 km/h of mean FS (Table 2). Plot 4 also

showed that the harvesting operation resulted in a higher number of acres covered per hour than plot 1. Hamid et al. (2018) found that the highest harvesting speed (4.73 km/h) of the mid-size combine harvester (Kubota DC 70G Plus) contributed to a lower yield (5814.22 kg/ha). The highest yield (7002.36 kg/ha) was obtained at a harvesting speed of 2.25 km/h for the same variety of grain corn. The study revealed that the difference in harvesting speed of 2.48 km/h between these 2 plots resulted in 16.97% (1188.14 kg/ha) of yield loss (Hamid et al., 2018). Zubko et al. (2018) reported that the productivity of winter wheat grain yield increased by 49% when FS increased from 3 to 6 km/h. Thus, it is concluded that the productivity would increase when FS increases without exceeding its optimum speed to minimize the grain losses.

Table 2

Field performance of the mid-size combine harvester for different plots

Field Performance	Plot 1 (Mean ± SD)	Plot 2 (Mean ± SD)	Plot 3 (Mean ± SD)	Plot 4 (Mean ± SD)	Average (Mean ± SD)	P- Value
FS, km/h	2.37 ± 0.19	2.72 ± 0.37	2.54 ± 0.15	2.56 ± 0.03	2.55 ± 0.23	0.36
EFC, ha/h	0.22 ± 0.02	0.23 ± 0.01	0.23 ± 0.02	0.24 ± 0.01	0.23 ± 0.01	0.70
FE, %	33.84 ± 3.15	35.35 ± 0.87	34.85 ± 2.63	35.86 ± 0.87	34.97 ± 1.99	0.70
FC, lit/ha	40.20 ± 2.80	34.97 ± 3.88	37.58 ± 2.54	36.27 ± 0.34	37.25 ± 3.07	0.19
FMI	0.91 ± 0.02	0.91 ± 0.01	0.90 ± 0.02	0.93 ± 0.01	0.91 ± 0.01	0.19

The P-values of experimental parameters at 95% confidence level (n=12). There were no significant differences of experimental parameters at $\alpha \leq 0.05$.

Fuel Consumption

Table 2 shows that the average value of FC for the mid-size combine harvester is 37.25 lit/ha. This amount is higher when compared to previous research on wheat and barley harvesting, using the combine harvester (Caterpillar C6.6), which consumed 29.03 and 17.28 lit/ha of diesel, respectively (Spokas et al., 2016). Table 2 which represents the comparison of the fuel consumption for all plots shows that the greater FE was associated with lesser FC and *vice versa*. For instance, in plot 1, FE at 33.84% consumed about 40.20 lit/ha of fuel which was 15.03% greater than FC for plot 2 with FE of 35.35%. Similar study was carried out using two combine harvesters at different capacities in paddy field

(Elsoragaby et al., 2019a). The conventional combine harvester (Clayson 8080) with a lower FE (64%) consumed 21.13 lit/ha of fuel, which was 14.46% greater than the fuel consumed by the mid-size combine harvester (WS 7.0Plus) with a FE of 72%.

Field Machine Index

The turning effectiveness of the mid-size combine harvester in terms of FMI is also shown in Table 2. The average value of FMI for the mid-size combine harvester is 0.91. This value is almost similar to the FMI value of 0.87 reported by previous study which employed the mid-size combine harvester (DC 95M) in a paddy field (Wagiman et al., 2019). According to Shamsiri et al. (2013), the value of FMI would be greater when the turning time of the combine harvester is lesser. A higher FMI value from both respective machineries in grain corn and paddy field as recently mentioned indicated the need for exclusion of unproductive movement such as cornering and reversing in carrying out the harvesting operation.

Field Time Distribution in Harvesting Operation

The field time distribution of the mid-size combine harvester for harvesting different plots considered several tasks which included productive (harvesting) and non-productive (cornering and reversing, machine setting and adjustment and unloading). Field time distribution is very important to be analysed for effective time management in farm machinery operation. Table 3 shows that there were no significant differences at 5% significant level ($P \leq 0.05$) of each task between harvesting for each plot.

Table 3

Field time distribution of the mid-size combine harvester for different plots

Task	Plot 1 (Mean ± SD)	Plot 2 (Mean ± SD)	Plot 3 (Mean ± SD)	Plot 4 (Mean ± SD)	Average (Mean ± SD)	P-Value
Harvesting, h/ha	2.87 ± 0.24	2.53 ± 0.32	2.67 ± 0.14	2.65 ± 0.03	2.68 ± 0.22	0.33
Turning, h/ha	0.30 ± 0.06	0.23 ± 0.02	0.29 ± 0.07	0.21 ± 0.01	0.26 ± 0.05	0.11
Machine Setting and Adjustment, h/ha	0.18 ± 0.06	0.22 ± 0.04	0.17 ± 0.03	0.14 ± 0.03	0.18 ± 0.05	0.22
Unloading, h/ha	1.16 ± 0.23	1.29 ± 0.13	1.30 ± 0.17	1.19 ± 0.02	1.24 ± 0.15	0.60
Total, h/ha	4.51 ± 0.49	4.27 ± 0.16	4.43 ± 0.34	4.20 ± 0.04	4.35 ± 0.29	0.61

The P-values of experimental parameters at 95% confidence level (n=12). There were no significant differences of experimental parameters at $\alpha \leq 0.05$.

Harvesting Time

The average harvesting time was computed for all plots as illustrated in Table 3. The mid-size combine harvester spent 61.58% of the total field time for harvesting task which is equivalent to 2.68 h/ha. Elsoragaby et al. (2019a) highlighted that the mid-size combine harvester (WS 7.0 Plus) spent 71.69% (1.36 h/ha) from the total field time in paddy field. In another study, Busato et al. (2007) reported that the wheat harvesting with 9 m working width combine harvester spent 61.76% of the total field time which was equivalent to 0.12 h/ha.

The differences of harvesting time between the combine harvesters as previously stated are dependent on machinery specification, terrain condition, types and varieties of grain crops. For instance, the mid-size combine harvester had recommended FS by the manufacturer which would be really suitable for specific grain corn harvesting to avoid losses of unrecovered corn from the cobs which are left behind due to excessive FS.

Turning Time

Generally, turning time which consists of cornering and reversing in harvesting consumed about 5.91% (0.26 h/ha) of the total field time (Table 3). For paddy and wheat harvesting, the turning time consumed 8.84% (0.17 h/ha) and 23.61% (0.05 h/ha), respectively of the total field time (Elsoragaby et al., 2019a; Busato et al., 2007). These studies showed that the turning time of the combine harvester for wheat harvesting was 80.77% lower than that of the rice harvesting. Lower mean value of turning time in wheat harvesting might due to less frequent turning of larger size combine harvester with the larger working width (5 m) as compared to the mid-size combine harvester employed in rice harvesting which had a smaller working width (2.67 m). It was mentioned by Elsoragaby et al. (2019a) that the advantage of having large working width made the conventional combine harvester to have greater EFC thus reducing the time loss in unproductive movement during turning.

Unloading time

The mid-size combine harvester spent about 28.41% (1.24 h/ha) of the total field time to unload the grain corn from the grain tank into the truck (Table 3). In comparison to paddy harvesting, earlier research finding showed that 18.95% (0.36 h/ha) of the total field time was spent for grain unloading (Elsoragaby et al., 2019a). Busato et al. (2007) revealed that for wheat harvesting, 12.22% (0.02 h/ha) of the total field time was spent to unload the grain. A relatively higher unloading time for the mid-size combine harvester as shown in Table 3 was justified by frequent grain unloading due to the small size of grain tank as compared to conventional combine harvester which generally has a larger size of grain tank.

Machine Setting and Adjustment

Table 3 presents the remaining time spent for other tasks such as machine setting and adjustment which consumed about 4.10% (0.18 h/ha) of the total field time in grain corn harvesting. A series of recent studies indicated that the remaining time for other tasks consumed 0.52% (0.01 h/ha) and 2.41% (0.01 h/ha) of the total field time in paddy and wheat harvesting, respectively (Elsoragaby et al., 2019a; Busato et al., 2007). The remaining time consumed for other tasks as previously discussed, is dependent on soil condition, effectiveness of the machine performance and the skills of operators. The extra unproductive time consumed in the corn field is due to frequent machine setting and adjustment in order to optimize the field performance and increase the harvesting efficiency.

Total Field Time

Table 3 indicates that the total harvesting field time in corn field is 4.35 h/ha. Similar research on the mid-size combine harvester (Kubota DC 70G Plus) conducted by MARDI revealed that the total field time are 3.13, 2.56 and 2.44 h/ha for 3 different varieties of grain corn at the same harvesting plot (Hamid et al., 2018). Elsoragaby et al. (2019a) stated that the total field time was 1.9 h/ha for paddy harvesting by the mid-size combine harvester (WS 7.0Plus). Chegini and Mirnezami (2016) presented the average total field time was 1.90 h/ha for wheat harvesting by John Deere 955 combine harvester. The higher total harvesting time in the current study as compared to the previous studies is due to a higher time consumed on unproductive task especially during grain corn unloading which accounts for 28.41% from the total average field time (Table 3).

Energy Consumption

Energy consumption is one of the most important factors required to be monitored and observed during field operation. Energy losses should be minimized in order to save cost of operation especially during harvesting. Table 4 shows that there are no significant differences of energy input between harvesting sub plots at 5% significance level ($P \leq 0.05$).

Fuel Energy

The fuel energy contributed 74.93% (1780.70 MJ/ha) of the total energy input in grain corn harvesting as shown in Table 4. Muazu et al. (2014a) reported that for wet land paddy harvesting, the self-propelled combine harvester consumed 853.54 MJ/ha of fuel energy which was equivalent to 73.59% of the total energy input. For other crops such as rapeseed, sunflower and sweet sorghum, the combine harvester with 7000 kg weight and 3.8 m working width consumed 1116 MJ/ha of fuel energy (Gemtos et al., 2013). The difference in fuel energy consumption between combine harvesters as previously mentioned is due to different in machinery specification, work load and duration of working time.

Table 3 and 4 show that a higher fuel energy corresponds to a higher total field time and *vice versa*. For instance, the mean fuel energy in plot 1 (1921.40 MJ/ha) with mean total field time of 4.51 h/ha was 14.94% greater in fuel energy than in plot 2 (1671.70 MJ/ha) which has a mean total field time of 4.27 h/ha. Similar findings could be seen between plot 3 and 4 where the mean fuel energy in plot 3 (1796.13 MJ/ha) with a mean total field time of 4.43 h/ha was 3.61% greater in fuel energy than plot 4 (1733.55 MJ/ha) which had a mean total field time of 4.20 h/ha. Chegini and Mirnezami (2016) found that 18 minutes and 33 minutes of total harvesting time resulted in 11.36 and 17.06 lit/ha of diesel fuel consumption, respectively in wheat harvesting.

Table 4

Energy input of grain corn harvesting for different plots

Energy Input	Plot 1 (Mean ± SD)	Plot 2 (Mean ± SD)	Plot 3 (Mean ± SD)	Plot 4 (Mean ± SD)	Average (Mean ± SD)	P-Value
Fuel Energy, MJ/ha	1921.40 ± 134.33	1671.70 ± 185.85	1796.13 ± 121.44	1733.55 ± 16.17	1780.70 ± 146.97	0.19
Machinery Energy, MJ/ha	609.24 ± 59.27	579.84 ± 14.15	590.15 ± 42.59	571.68 ± 14.15	587.73 ± 35.45	0.66
Human Energy, MJ/ha	8.85 ± 0.95	8.36 ± 0.33	8.70 ± 0.67	8.22 ± 0.08	8.53 ± 0.58	0.58
Total Energy, MJ/ha	2539.50 ± 193.56	2259.91 ± 200.24	2394.97 ± 164.58	2313.45 ± 17.76	2376.96 ± 176.68	0.24

The P-values of experimental parameters at 95% confidence level (n=12). There were no significant differences of experimental parameters at $\alpha \leq 0.05$.

Machinery Energy

Machinery energy was determined by considering weight, EFC and estimated working life. Table 4 represents the average value of machinery energy of 587.73 MJ/ha which is equivalent to 24.77% of energy distribution in the total energy input. Muazu et al. (2014b) pointed that the self-propelled combine harvester with 3000 hrs of estimated working life contributed 303.53 MJ/ha of machinery energy which was equivalent to 26.17% of energy distribution in the total energy input in paddy harvesting. Gemtos et al. (2013) stated that the harvesting of rapeseed, sunflower and sweet sorghum in 1 ha/h required 417.8 MJ/ha of machinery energy, where the combine harvester used had 7000 kg of weight and 2000 hrs estimated working life. In fact, the same model of the combine harvesters with the same specification used on the same crop can have different mean value of machinery energy

due to difference in EFC. For instance, plot 1 with 0.22 ha/h of EFC had 609.24 MJ/ha of machinery energy while plot 4 with 0.24 ha/h of EFC had 571.68 MJ/ha of machinery energy. The harvesting operation with lower EFC values has greater machinery energy than the harvesting operation with higher EFC values.

Human Energy

Table 4 also presents the average value of human energy consumed in corn field which is about 8.53 MJ/ha, equivalent to 0.36% of energy distribution in the total energy input. Muazu et al. (2014b) reported that the human energy in wet land paddy harvesting was 2.72 MJ/ha which was equivalent to 0.23% of energy distribution in the total energy input. Table 2 and 4 show that the lesser mean of EFC in plot 3 (0.23 ha/h) had 5.84% greater mean human energy (8.7 MJ/ha) than plot 4 which had 0.24 ha/h of mean EFC and 8.22 MJ/ha of the mean human energy. This study explained that greater human energy was consumed per hectare due to lesser harvested area per working hour. An additional working time per harvested area contributes to a lesser EFC.

Total Energy Input

Table 4 reveals that the average value of the total energy input of grain corn harvesting is 2376.96 MJ/ha. Muazu et al. (2014b) highlighted that the total energy input for wet land paddy harvesting by self-propelled combine harvester was 1159.79 MJ/ha. The average value of the total energy input or total operating energy for sunflower and sweet sorghum harvesting with a combine harvester of 7000 kg mass are 1585MJ/ha and 406 MJ/ha, respectively (Gemtos et al., 2013).

Table 2 and 4 represent plot 4 with mean FE (35.86%) and a mean total energy input (2313.45 MJ/ha) has 8.90% less value of mean total energy input than plot 1 which has a mean FE of 33.84%. Elsoragaby et al. (2019a) mentioned that the mid-size combine harvester (WS 7.0Plus) with mean FE of 72% had 19.77% lesser value of total energy input than the conventional combine harvester (Clayson 8080) which had a mean FE of 64%. Overall, the total energy input decreases when FE increases.

Energy Classification

From the result presented in Figure 3, the average value of direct and indirect energy distribution from the total energy input in grain corn harvesting were 75.23 and 24.77%, respectively. Elsoragaby et al. (2019b) reported that the average values of direct and indirect energies from the total energy input in paddy harvesting were 76.25 and 23.75%, respectively. For sweet sorghum, rapeseed and sunflower harvesting, the direct and indirect energy were 72.76 and 27.33%, respectively (Gemtos et al., 2013). Thus, it can be concluded that all studies reported that direct energy is greater than indirect energy for all the crops under investigation.

Figure 4 shows that the average value of renewable and non-renewable energy in grain corn harvesting are 0.36% and 99.64%, respectively. Recent research in paddy harvesting showed that the renewable and non-renewable energy are 0.17 and 99.83%, respectively (Elsoragaby et al., 2019b). Similar research conducted by Muazu et al. (2014b) in wet land paddy harvesting revealed that the renewable and non-renewable energy were 0.23% and 99.76%, respectively. These reports illustrate that the consumption of non-renewable energy is greater than the renewable energy in harvesting operation especially for grain crop like paddy and corn.

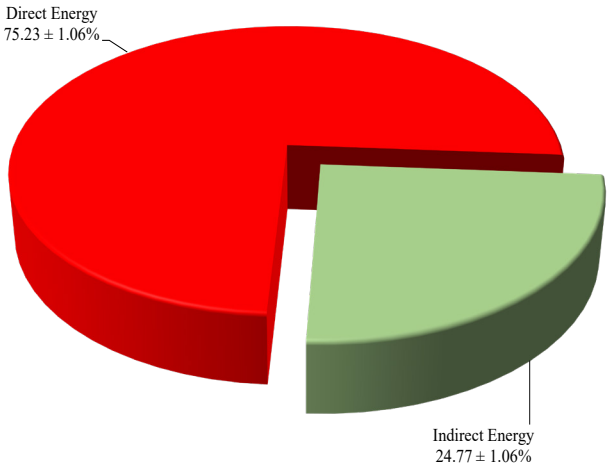


Figure 3. Direct and indirect energy for harvesting operation in grain corn production

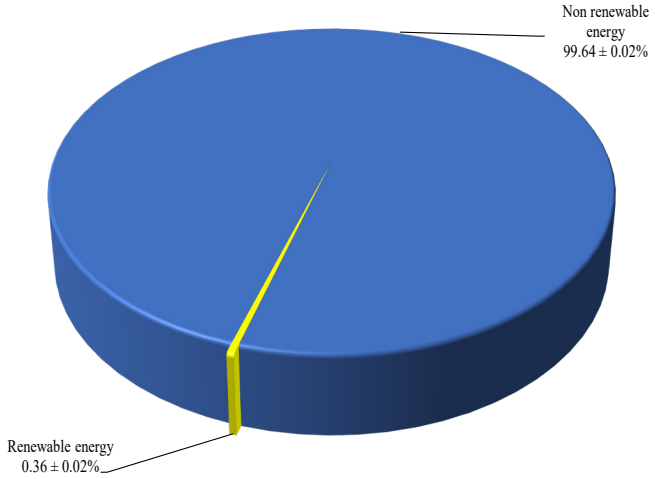


Figure 4. Renewable and non-renewable energy for harvesting operation in grain corn production

CONCLUSION

This study examined various parameters in determining the field performances and energy consumption involved in grain corn production. The average value of FE, EFC and FC were 34.97%, 0.23ha/h and 40.20 lit/ha, respectively. The average value of FS and FMI were 2.55 km/h and 0.91 respectively. Based on the findings, the average value of energy consumption for fuel, machinery and human energy were recorded at 1780.70 MJ/ha, 587.73 MJ/ha and 8.53 MJ/ha, respectively. The fuel, machinery and human energy contributed about 74.93, 24.77 and 0.36% to the total energy consumption, respectively. The share of the direct and indirect energy in the fields by average value from the total energy consumption were 75.23 and 24.77% respectively. While the share of the renewable and non-renewable energy were 0.36 and 99.64%, respectively. This research indicated that the field performance and energy consumption for grain corn harvesting in Labis, Malaysia was consistent with some other grain harvesting operation that were studied in previous scientific researches and are discussed as literature studies in this article. Based on these findings of this study, it is recommended that, when operating a mid-size corn harvester with 2.76 m reaping width and 2.67 m cutting blade, the following operating parameters would yield the most optimal corn productions and minimum energy consumption with forward speeds were between 2.37 to 2.72 km/h, total harvest time of 4.20 to 4.51 h/ha, and sandy loam soils. These parameters achieved an average value of the effective field capacity 0.23 ha/h with maximum value of 0.24 ha/h, the average value of fuel consumption 37.25 lit/ha with minimum value of 34.97 lit/ha, average percentage of field efficiency 34.97% with maximum value of 35.86%, and an average of the field machine index 0.91 with maximum value 0.93, as well as the total energy consumption between 2259.91 and 2539.50 MJ/ha with an average of consumption 2376.96 MJ/ha.

ACKNOWLEDGEMENTS

The authors acknowledge the research grant awarded by Universiti Putra Malaysia, classified under Putra Initiative Grant (Vot: 9635700). The authors also acknowledge technical supports from the staff of DOA and NAFAS throughout field measurement at corn field, Felda Chempelak, Johor, Malaysia.

REFERENCE

Amponsah, S. K., Addo, A., Dzisi, K., Moreira, J., & Ndindeng, S. A. (2017). Comparative evaluation of mechanised and manual threshing options for amankwatia and AGRA rice varieties in Ghana. *Journal of Agricultural Engineering: CIGR Journal, 48*(4), 181-189.

Aribe, J. S. G., Turtosa, J. M. H., Yamba, J. M. B., & Jamisola A. B. (2019). Ma-Ease: An android-based technology for corn production and management. *Pertanika Journal of Science and Technology*, *27*(1), 49-68.

ASABE. (2005). *ASAE S495.1 NOV2005, R2015. Uniform terminology for agricultural machinery management.* American Society of Agricultural and Biological Engineers. Retrieved March 23, 2020, from https://www.asabe.org/Publications-Standards/Standards-Development/National-Standards/Published-Standards

ASABE. (2011). *ASAE D497.7 MAR2011. Agricultural machinery management data.* American Society of Agricultural and Biological Engineers. Retrieved March 18, 2020, from http://www.abe-research.illinois.edu/Faculty/grift/TSM262_2017/Standards/ASAE_D497_7.pdf

Bochtis, D., Vougioukas, S., Tsatsarelis, C., & Ampatzidis, Y. (2007). Optimal dynamic motion sequence generation for multiple harvesters. *Agricultural Engineering International: CIGR Journal, 9*, 1-9.

Busato, P., Berruto, R., & Saunders, C. (2007). Optimal field-bin locations and harvest patterns to improve the combine field capacity: Study with a dynamic simulation model. *Agricultural Engineering International: CIGR Journal, 9*, 1-12.

Canakci, M., Topakci, M., Akinci, I., & Ozmerzi, A. (2005). Energy use pattern of some field crops and vegetable production: Case study for Antalya region, Turkey. *Energy Conversion Management, 46*, 655-666.

Chegini, G., & Mirnezami, S. V. (2016). Experimental comparison of combine performance with two harvesting methods: Stripper header and conventional header. *Agricultural Engineering International: CIGR Journal, 18*(1), 192-200.

Elsoragaby, S., Yahya, A., Mahadi, M. R., Nawi, N. M., & Mairghany, M. (2019a). Comparative field performances between conventional combine and mid-size combine in wetland rice cultivation. *Heliyon, 5*(4), 1-25.

Elsoragaby, S., Yahya, A., Mahadi, M. R., Nawi, N. M., & Mairghany, M. (2019b). Analysis of energy use and greenhouse gas emission (GHG) of 15transplanting and broadcast seeding wetland rice cultivation. *Energy, 189*, 116-160.

Gemtos, T. A., Cavalaris, C., Karamoutis, C., Tagarakis, A., & Fountas, S. (2013). Energy analysis of three energy crops in Greece. *Agricultural Engineering International: CIGR Journal, 15*(4), 52-66.

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

Grisso, R. D., Kocher, M. F., Adamchuk, V. I., Jasa, P. J., & Schroeder, M. A. (2004). Field efficiency determination using traffic pattern indices. *Applied Engineering in Agriculture*, *20*(5), 563-572.

Hamid, M. A., Rohazrin, A. R., Chan, C. S., Yahya, S., Shanmugavelu, S., Aris, A., & Saleh, B. (2018, April 17-19). Preliminary study of the performance of imported grain corn combine harvester machines. In *National Conference on Agricultural and Food Mechanization, NCAFM, 2018*. (pp. 35-39). Sarawak, Malaysia.

Henrichsmeyer, F., & Ohls, J. (1995). Leistung und kosten von arbeitsverfahren in grossbetrieben [Work requirements and costs on large farms]. *Landtechnik, 50*, 296-297.

Jokiniemi, T., Rossner, H., & Ahokas, J. (2012). Simple and cost-effective method for fuel consumption measurements of agricultural machinery. *Agronomy Research Biosystem Engineering Special Issue, 1*, 97-107.

Mohammadi, A., Rafiee, S., Mohtasebi, S. S., & Rafiee, H. (2010). Energy inputs–yield relationship and cost analysis of kiwifruit production in Iran. *Renewable Energy, 35*(5), 1071-1075.

Muazu, A., Yahya, A., Ismail, W. I. W., & Bejo, S. K. (2014a). Machinery utilization and production cost of wetland, direct seeding paddy cultivation in Malaysia. *Agriculture and Agricultural Science Procedia, 2*, 361-369.

Muazu, A., Yahya, A., Ismail, W. I. W., & Bejo, S. K. (2014b). Yield prediction modelling using data envelopment analysis methodology for direct seeding, wetland paddy cultivation. *Agriculture and Agricultural Science Procedia, 2*(0), 181-190.

Nasri, N. N., Nawi, N. M., Abdullah, N., Saripa, S., & Nasir, R. M. (2016, May 24-26). Measurement of field efficiency for different field operations in sweet corn production. In *Proceedings of the Technology and Innovation National Conference, TECHON 2016* (pp. 233-241). Sarawak, Malaysia.

Nor, N. A. A. M., Rabu, M. R., Adnan, M. A., & Rosali, M. H. (2019). *An overview of the grain corn industry in Malaysia*. FFTC Agricultural Policy Platform (FFTC-AP). Retrieved March 10, 2020, from http://ap.fftc.agnet.org/index.php.

Olt, J., Kuut, K., Ilves, R., & Kuut, A. (2019). Assessment of the harvesting costs of different combine harvester fleets. *Research in Agricultural Engineering, 65*(1), 25-32.

Ozkan, B., Akcaoz, H., & Fert, C. (2004a). Energy input–output analysis in Turkish agriculture. *Renewable Energy, 29*, 39-51.

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

Sarauskis, E., Buragien, S., Masilionyt, L., Romaneckas, K. S., Zienyt, D. A., & Sakalauskas, A. (2014). Energy balance, costs, and CO_2 analysis of tillage technologies in maize cultivation. *Energy, 69*, 227-235.

Shamsiri, R., Ehsani, R., Maja, J. M., & Roka, F. M. (2013). Determining machine efficiency parameters for a citrus canopy shaker using yield monitor data. *Applied Engineering in Agriculture, 29*(1), 33-41.

Sime Kubota Sdn Bhd. (2019). *Kubota combine harvester DC 105X*. Retrieved March 10, 2020, from https://www.kubotamalaysia.com/wp-content/uploads/2019/brochure/DC-105X_Brochure.pdf

Singh, H., Mishra, D., & Nahar, N. M. (2002). Energy use pattern in production agriculture of typical village in arid zone, India: Part I. *Energy Conversion Management, 43*, 75-86.

Smith, D. W., Sims, B. G., & O'Neill, D. H. (1994). *Testing and evaluation of agricultural machinery and equipment.* Rome, Italy: Food and Agriculture Org.

Spokas, L., Adamcuk, V., Bulgakov, V., & Nozdrovicky, L. (2016). The experimental research of combine harvesters. *Research in Agricultural Engineering, 62*(3), 106-112.

UNC. (2019). *Commodity trade statistics database.* United Nations Comtrade. Retrieved April 11, 2019, from http://data.un.org/Data.aspx?d=ComTrade&f=_l1Code%3A

USDA. (2018). *World agricultural production: Circular series WAP 12-18 December 2018.* United States Department of Agriculture. Retrieved January 8, 2019, from https://downloads.usda.library.cornell.edu/usda-esmis/files/5q47rn72z/qj72pc12b/sj1395841/production.pdf.

Wagiman, N. A., Nawi, N. M., Yahya, A., Su, A. S. M., & Nasir, R. M. (2019). Field performance comparison of the combine harvesters utilized for rice harvesting in Malaysia. *Food Research, 3*(2), 177-181.

Zhou, K. (2012). *Simulation modelling for in-field planning of sequential machinery operations in cropping systems* (PhD Thesis). Aarhus University, Denmark.

Zubko, V., Roubik, H., Zamora, O., & Khvorost, T. (2018). Analysis and forecast of performance characteristics of combine harvesters. *Agronomy Research, 16*(5), 2282-2302.

Pertanika J. Sci. & Technol. 28 (4): 1297 - 1313 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Development of a Self-regulated Bubble Irrigation System to Control the Size and Shape of Wetting Fronts

Yasir Layth Alrubaye[1]*, Badronnisa Yusuf[1] and Safaa Noori Hamad[2]

[1]Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Malaysia
[2]Engineering College, Uruk University, Baghdad, 10067 Iraq

ABSTRACT

The main objectives of this study were to introduce a bubble irrigation system, compare the wetting fronts created by the bubble and free-flow systems, and test the viability of a bubble irrigation system. Two laboratory experiments were conducted using 2D flow to measure the wetting fronts. The first experiment measured the free-flow infiltration using an inverted, open plastic bottle. The second experiment tested the bubble-flow technique using an inverted, closed plastic bottle (ICPB). The results showed that the bubble-flow system created a larger width of wetting fronts at the beginning of the infiltration and then expanded less than that of the free-flow system. In contrast, the infiltration depth of the wetting fronts created by the bubble-flow system was much lower than that of the free-flow system. In conclusion, the wetting front width and depth in the bubble-flow system were slightly smaller than those in the free-flow system. In addition, the wetting fronts created by the ICPB were not moved upwards significantly, which proves the ability of specific distribution of the bubble-flow system on the wetting fronts. Therefore, the bubble irrigation system can be used as an alternative for distributing the moisture content in soil profiles.

Keywords: Air-water exchange, bubble irrigation, plastic bottle, soil-water, uniformity and efficiency

ARTICLE INFO

Article history:
Received: 20 April 2020
Accepted: 18 August 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.09

E-mail addresses:
Yasir_alrubaye@hotmail.com (Yasir Layth Alrubaye)
nisa@upm.edu.my (Badronnisa Yusuf)
safaanori12@yahoo.com (Safaa Noori Hamad)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Recently, competition for water resources has increased for urban, industrial, and agricultural users (Kandelous & Šimůnek, 2010). Agricultural water use will increase by approximately 19% by 2050 (Kisekka et al., 2017). Therefore, more efficient and uniform irrigation methods, such as subsurface irrigation systems, must be developed for the distribution of irrigation water.

Recent studies have discussed subsurface irrigation systems to improve the design of irrigation systems (Gu et al., 2017; Gunarathna et al., 2017; Gunarathna et al., 2018), regulate the wetting fronts created by new products (Lima et al., 2019; Saefuddin et al., 2019), evaluate the hydraulic performance (Ren et al., 2017; Ren et al., 2018; Ghazouani et al., 2019), and measure the moisture distribution in the soil profile (Cai et al., 2017; Fan & Li, 2018; Cai et al., 2019a; Elnesr & Alazba, 2019).

Designing efficient subsurface irrigation systems is the most challenging because the irrigation water is applied directly to the soil profile. The design of subsurface irrigation systems is influenced by several factors including soil variety, environment, time, duration of the irrigation process, and plant type (Sakaguchi et al., 2019). Designers consider all these factors; however, these factors still reduce the efficiency and uniformity at the operation stage. To overcome the effects of these factors, new system components have been invented. Cai et al. (2019b) tested a ceramic patch in a subsurface irrigation line to control the saturation zone in the soil profile, which was created by the pressure head applied by the emitters. They found that there was a relationship among the pressure head, ceramic properties, and soil properties. Lima et al. (2019) showed that using a new irrigation product, named a permeable membrane, could be the solution for increasing water use efficiency and maximizing irrigation management.

Continuous evaluation of subsurface irrigation systems is essential for obtaining a more uniform irrigation process (Gunarathna et al., 2018). Evaluation of irrigation systems indicates what is the best method for irrigation or which irrigation system is more suitable under certain conditions. Nabayi et al. (2018) evaluated the performance of three different irrigation types (sprinkler, drip, and capillary wick irrigation systems) for raising rubber seedling crops. Their results indicated that the efficiency and water productivity was influenced by the type of irrigation system used. They also showed that the capillary wick and the drip irrigation systems had the highest water productivity, whereas the sprinkler irrigation system had the lowest water productivity due to the canopy intercept. Al-Ghobari and Dewidar (2018) evaluated deficit irrigation strategies for surface and subsurface drip irrigation systems and concluded that these improved water management by minimizing the effects on production.

The measurement of the size of the wetting front in the soil profile is critical for designing cost-effective and highly efficient subsurface irrigation systems (Fan et al.,

2018b). Measurements of wetting fronts indicate the size and shape of the moisture distribution in the soil. This measurement is used to reduce deep percolation to increase the efficiency of irrigation performance. Researchers have conducted experiments and simulations to measure the wetting patterns in different soil types, applied discharge, irrigation systems, and boundary conditions. Fan et al. (2018a) simulated a vertical line source using HYDRUS and produced empirical forms to measure the wetting patterns that were created. These authors concluded that the created empirical forms were required for future field studies to measure the wetting patterns and evaluate the created empirical forms. Elnesr and Alazba (2019) simulated the wetting fronts created by subsurface drip irrigation using 3D and 2D HYDRUS. They concluded that 2D simulation could be used confidentially by simulators.

The size and shape of the wetting patterns in the soil profile are influenced by the applied discharge, time of application, and the hydraulic properties of the soil (Amin & Ekhmaj, 2006; Dawood & Hamad, 2016; Moncef et al., 2002). Because the soil properties are the natural conditions, the applied discharge and time of application are specified to control the wetting patterns. In general, there are two traditional techniques for applying discharge into the soil profile: a free flow and pressurized discharge. However, these techniques are fully controlled, i.e., they are not influenced by the soil conditions.

To control the wetting patterns, several studies have used the basic principle of the buoyancy of bubbles for irrigation improvements. A study by Liu et al. (2019) aimed to improve the yield and irrigation water efficiency by evaluating a technique involving micro-nano bubble—water oxygenation using a subsurface drip irrigation system. They noticed that the speed and size of the bubbles in the saturated soil body increased the nutritional quality of the crops. To use the basic principle of buoyancy in subsurface irrigation systems, it is necessary to measure the movement and size of the bubbles in the hydraulic field. Mohagheghian and Elbing (2018) measured the size distribution of the bubbles within the water column using 2D imaging with a high-resolution camera. Another study by Barkai et al. (2019) investigated the hydrodynamics of bubble movement in a vacuum airlift column to optimize its design and operation.

There is a global need to save water, especially during the irrigation process. Therefore, the performance of subsurface irrigation systems must be made more efficient and uniform. Wetting fronts created in soil profiles by subsurface irrigation systems must be controlled. Novel design criteria of subsurface irrigation systems can be outlined to overcome the emissive losses owing to the uncontrolled size of wetting patterns. The key solution is to use the buoyancy principle of bubbles. This principle can be used in subsurface irrigation systems to improve irrigation performance and control the wetting fronts in the soil profile. The main objectives of the present study were to introduce a bubble irrigation system,

compare the wetting fronts created by the bubble and free-flow systems, and test the viability of using a bubble irrigation system.

MATERIALS AND METHODS

Description of Self-Regulated Irrigation System

In the present study, a novel self-regulated irrigation system was introduced, called the bubble irrigation system. The design of this system was dependent on the movement of water through soil caused by the air–water exchange technique, which was produced in the soil profile using inverted, closed plastic bottles (ICPBs; Figure 1). The system used in the present study consisted of an elevated closed tank, a main valve, an air valve, a pipe network, and an ICPB. The water movement in the bubble irrigation system started in the elevated tank and moved through the pipe network until it reached the ICPB by air–water exchange. The water flowed into the ICPB whereas the air travelled to the elevated tank through the pipe network. Water infiltrated into the soil profile from the ICPB and generated an inverted wetted bulb. The infiltration process depended on the air–water exchange in the soil voids. This exchange mechanism controlled the amount of water that infiltrated into the soil and thus the size of the wetting front was held under the control of the plants.

The bubble irrigation system is an alternative irrigation system that will improve the efficiency and uniformity of irrigation performance compared to traditional irrigation methods. This system controls the water that is applied and is a self-regulated system. The

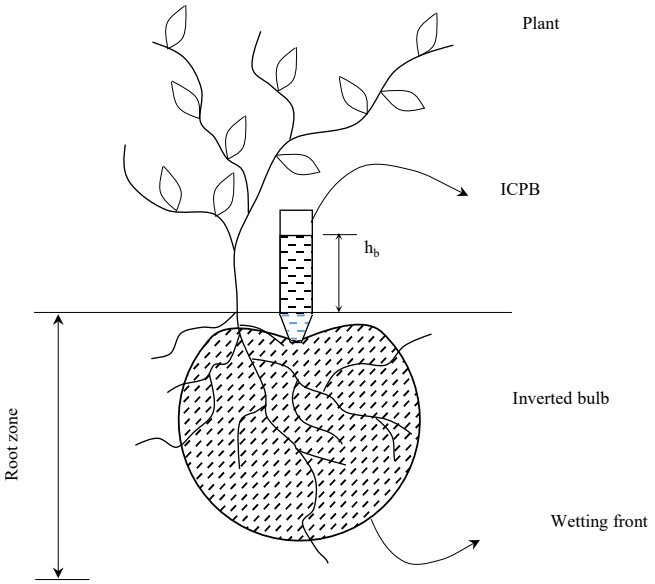


Figure 1. Schematic representation of the locations of the ICPB and root zone and the shape of the inverted bulb

water holding capacity of the soil and the water delivery capacity of the irrigation system are constraints on the ability to supply water to a crop (Klocke et al., 2009). Traditional irrigation systems adjust the water applied into or onto the soil to fit the required quantities. However, the bubble irrigation system adopts an air–water exchange technique to control the amount of applied water based on the soil moisture deficit, i.e., based on the consumptive use by the plants.

The major advantages of the bubble irrigation system are the avoidance of water losses from spray drift losses when sprinkling (Hobson et al., 1993; Holterman, 2003; Miller, 2003; Hilz & Vermeer, 2013), deep drainage losses (Paydar et al., 2009), surface runoff when surface irrigating (Hatiye et al., 2018; Masih & Giordano, 2014), and emitter clogging problems when drip irrigating (Cararo et al., 2006; de Oliveira et al., 2020). Adding water directly to the soil also decreases evaporation losses from the soil surface (Alrubaye et al., 2018). The bubble system continuously applies irrigated water directly to the soil depending on water infiltration. A suggested layout of a bubble irrigation system is shown in Figure 2. The operational cost of a bubble irrigation system depends on the method used for refilling the main tank.

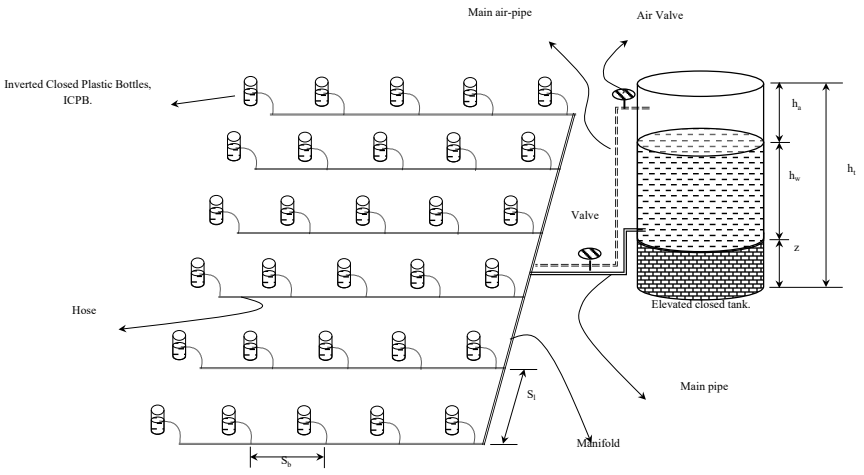


Figure 2. Suggested layout of a bubble irrigation system

Setup of Laboratory Experiments

Laboratory experiments were performed to compare the wetting fronts in the soil profile created by free flow from inverted, open plastic bottles (IOPBs) and the bubble flow from ICPBs. Bubble and free-flow experiments were designed to measure the flow rate and 2D wetting fronts in the soil profile during the experiments. The experiments were performed using a glass soil container that was 80 cm long, 80 cm deep, and 20 cm wide. An elevated

closed cubic tank edge with a length of 50 cm was used with an air valve at the top and a control valve at the bottom to assure the flow conditions of the bubbles (Figure 3).

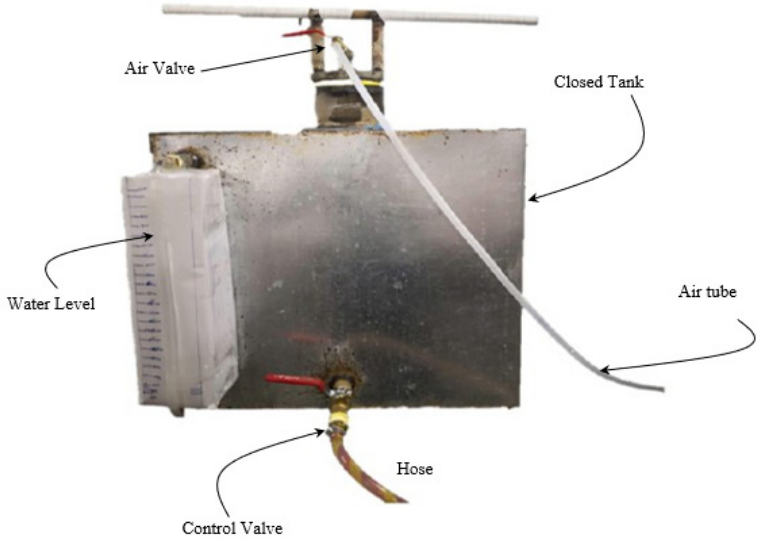


Figure 3. Elevated closed tank

The experimental setup of the free-flow experiments is shown in Figure 4. The equipment used included a soil container filled with sandy soil, a 2000 mL IOPB that was implanted 5 cm below the soil surface, and transparent paper. The constant head of water used in the experiment was 10 cm above the soil surface. Throughout the experiment, a transparent sheet was pasted on the front of the soil container. Wetting fronts were drawn manually on the transparent sheet at 5-min intervals and the flow rate was measured volumetrically every 5 min.

The bubble-flow experiments were conducted using the same soil container, a 2000 mL ICPB, a 1 m long hose, an air tube, and a closed elevated tank. The elevated tank was fully closed with two valves. The first valve was for water flow control and the second was for airflow control (Figure 3). The ICPB was connected to the elevated tank using the water and air tubes, in which these tubes were fully sealed to the ICPB by silicon and Teflon tape (Figure 5). The experimental setup is shown in Figure 6. Therefore, the condition inside the ICPB was fully controlled by the condition in the closed tank. When the water was raised in the ICPB until it reached the air tube, the water tube automatically stopped providing water into the ICPB. This is because the water level in the ICPB plugged the airflow to the closed tank. Water would raise in the air tube when there is a high vacuum pressure in the elevated tank. Therefore, the closed tank was determined as being in the right condition that allowed the mechanism of air–water exchange to control the supply of

water into the ICPB. This can be solved using a check valve at the head of ICPB connected to the air tube. In this paper, a slight vacuum was produced in the elevated tank to generate the air-water exchange condition. Thus, a slight water raised in the air tube did not affect the mechanism during the experiment.

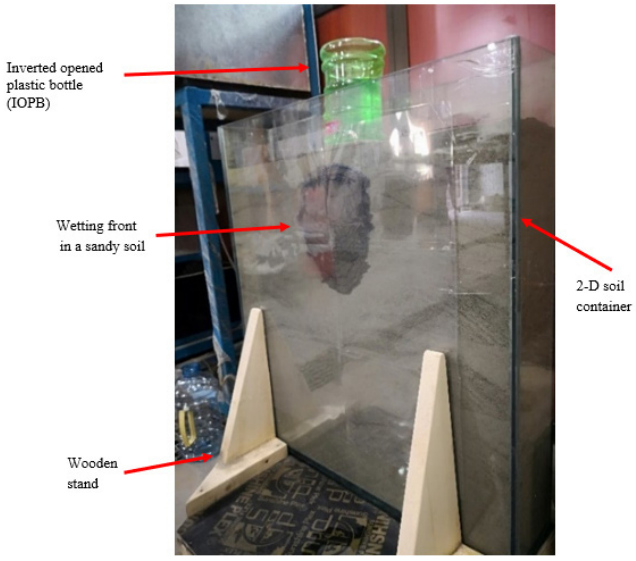


Figure 4. Experimental setup of the free-flow infiltration from the inverted open plastic bottle (IOPB)

Figure 5. Bubble-flow infiltration from the inverted closed plastic bottle (ICPB)

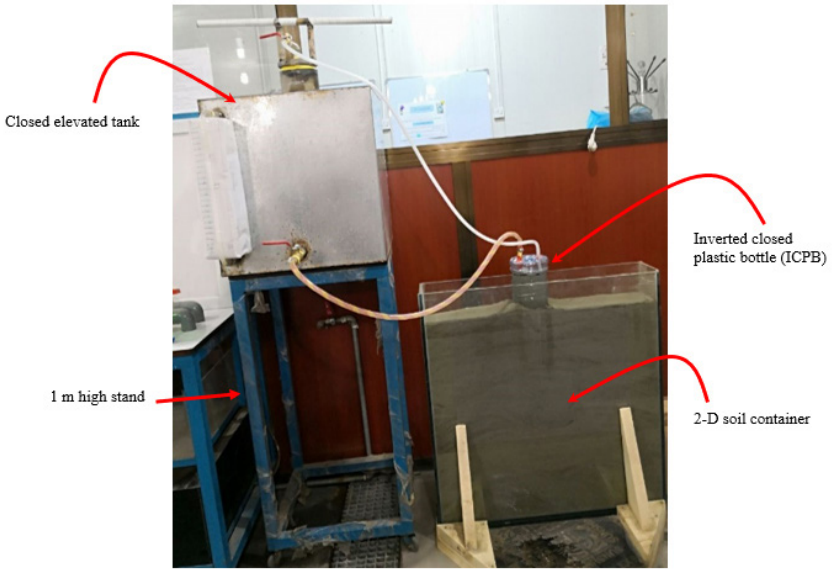


Figure 6. Experimental setup of bubble-flow infiltration from the inverted closed plastic bottle (ICPB)

Measurement of the Wetting Fronts

Two laboratory experiments were performed in the fluid mechanics laboratory at Al-Mansour University College. These experiments involved measuring the flow rate and drawing the movement of the wetting front throughout the bubble and free-flow experiments.

For the free-flow experiment, the container was filled with soil that was placed in layers to maintain a uniform and homogeneous soil density. Then, the IOPB was implanted in the soil and the transparent paper was pasted on the front face of the soil container. To allow water to flow through the soil, a constant head of water in the IOPB was maintained at the same level by a continual supply of water. The wetting front was traced on the transparent paper at 5-min intervals. The flow rate was measured volumetrically by dividing the volume of the infiltrated water by the time at 5-min intervals.

For the bubble-flow experiments, the container was filled with soil that had been placed in layers to maintain a uniform and homogeneous soil density. Then, the plastic bottle was connected to the hose by a plastic connection and the hose, in turn, was connected to the elevated tank. After that, transparent paper was pasted on the front face of the soil container. The control valve was opened to allow water to flow through the soil. At 5-min intervals, the wetting front was traced on the transparent paper. The flow rate was measured volumetrically by dividing the volume of the infiltrated water by the time in 5-min intervals.

For both the bubble and free-flow experiments, width measurements were taken at 5 cm below the plane at the point of infiltration throughout experimental period. Depth measurements were also taken along the central line of the wetting fronts over time.

Statistical Analysis

Three cases were statistically tested to determine the significant differences for the variables of the infiltration rate, the width of the wetting patterns, and the depth of the wetting patterns. These cases were tested using a one-sided paired t-test with a 95% confidence level to check the significance of the differences between free flow and bubble flow. The null hypothesis, Ho, assumed that the mean variable of free flow was equal to or less than that of the bubble flow, and the alternative hypotheses, H1, assumed that the mean variable of free flow was greater than that of bubble flow. Statistical results were performed using R Studio software.

RESULTS AND DISCUSSION

The bulk density of the soil used in the laboratory experiments was 1.3 gm/cm^3 and was classified as sandy soil based on the Unified Soil Classification System according to the ASTM Standards. The grain size distribution of the sandy soil is shown in Figure 7.

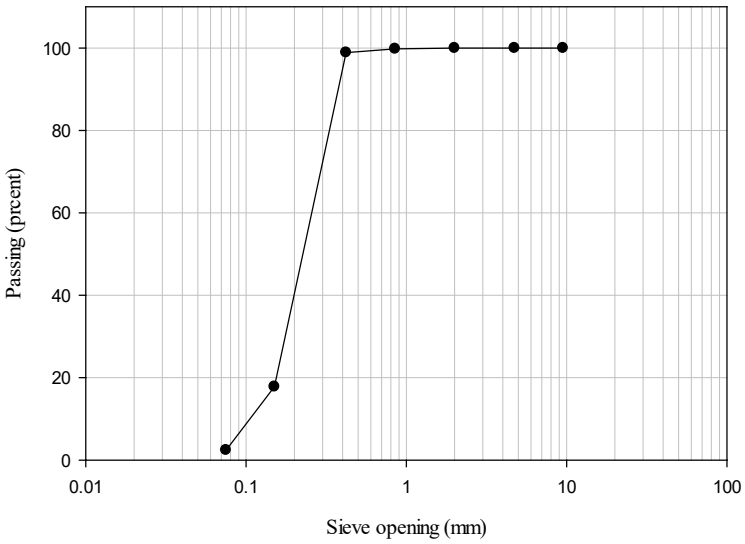


Figure 7. Grain size distribution of the sandy soil used in the laboratory experiments

The experiments showed that the wetting front area attained its maximum width approximately 5 cm below the plane at the point of infiltration, as can be seen in Figure 8A and 8B.

The inverted bulbs obtained from the bubble and free-flow experiments showed that the width and depth variation decreased over time. The wetting fronts for both the bubble and free-flow systems appeared 10 min after the start of the experiment. The wetting fronts for both bubble and free-flow experiments were almost similar in their general shape.

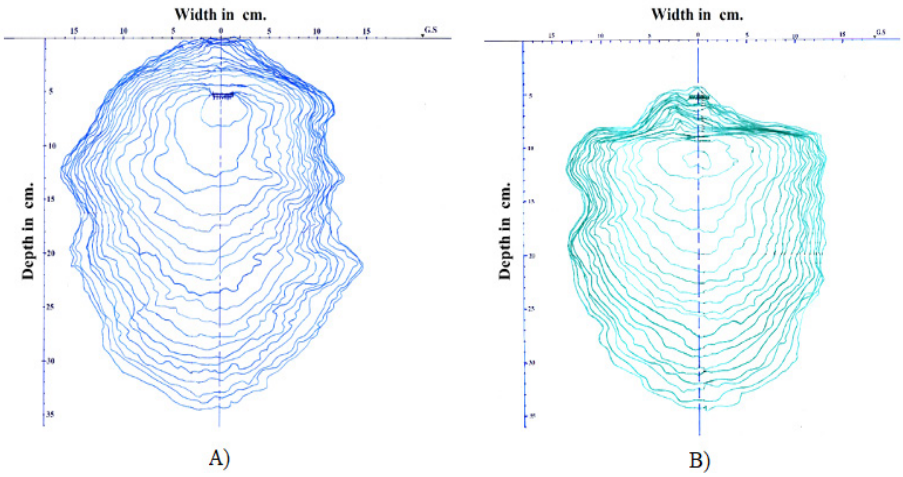


Figure 8. Wetting fronts of A) free flow and B) bubble-flow infiltration from an inverted bottle throughout the experimental period

The wetting fronts that were obtained from the IOPB showed that the free-flow infiltration extended in all directions and reached the soil surface. This was mainly because free flow occurs due to gravity and the pressure head in the IOPB. However, the wetting fronts obtained from the ICPB did not significantly move upwards and did not reach the soil surface; such distribution would decrease the evaporation from the soil surface and is one of the advantages of the bubble irrigation system.

The variation in the infiltration rate throughout the experimental period for both the bubble and free-flow systems is shown in Figure 9. The infiltration rate decreased with time for both flow types. The free-flow infiltration rate was, in general, larger than that for the bubble flow because bubble infiltration depends mainly on air–water exchange. This reduction in infiltration rate may indicate the effectiveness of the bubble irrigation system. Additionally, this comparison showed that the free-flow infiltration depended on pushing water through the soil profile. This phenomenon indicated the ability of the bubble irrigation system to control the irrigation water in the soil profile itself.

The results of the t-test for the first case involving the infiltration rate between the free-flow and bubble-flow systems found a t-value of 10.24, degrees of freedom of 7, and a p-value of 9.1 e^{-06}. Therefore, the null hypothesis was rejected and the alternative hypothesis was accepted, i.e., the mean infiltration rate of the free-flow system was greater than that of the bubble-flow system. The mean of the differences was 0.34 Lph. The statistical tests showed that the infiltration rate of the free-flow system was significantly greater than that of the bubble-flow system. Because the bubble-flow system is regulated by the air–water exchange technique, the infiltration rate of the bubble-flow system was lower than that of the free-flow system.

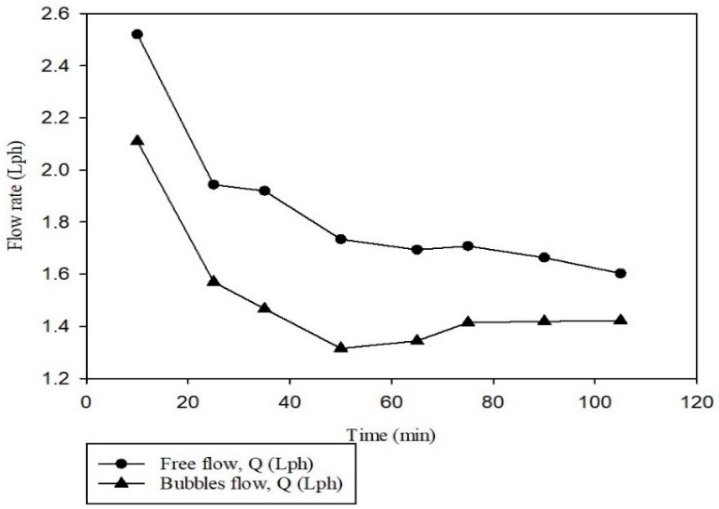


Figure 9. Variation in the infiltration rate throughout the experimental period for the bubble and free-flow irrigation systems

The growth of the wetting front width over time for both the bubble and free-flow systems is shown in Figure 10. At the beginning of the irrigation process, the wetting front width in both the bubble and free-flow systems was sharply expanded and then the rate of increase decreased over time. The two irrigation systems showed a reasonable phenomenon of reduction in the infiltration rate; however, there was a slight reduction in the wetting width for the bubble-flow irrigation system.

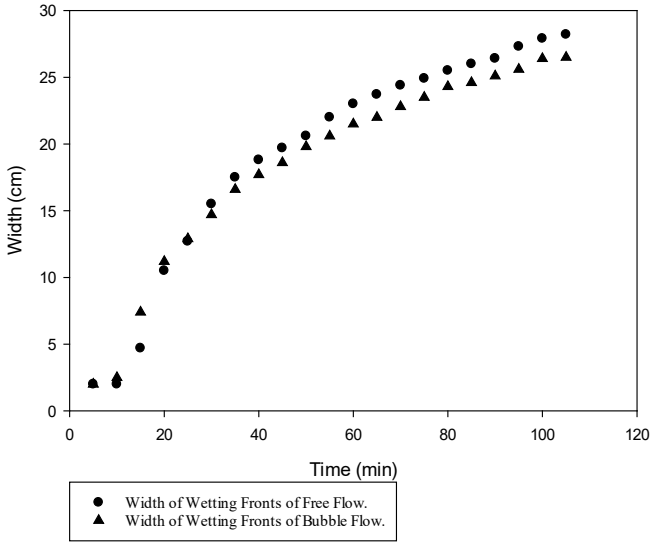


Figure 10. Variation in the wetting front width with time in the bubble and free-flow irrigation systems

The wetting front width in the bubble-flow system was generally smaller than that in the free-flow system (Figure 11). The bubble and free-flow systems distributed water close to each other in the width direction.

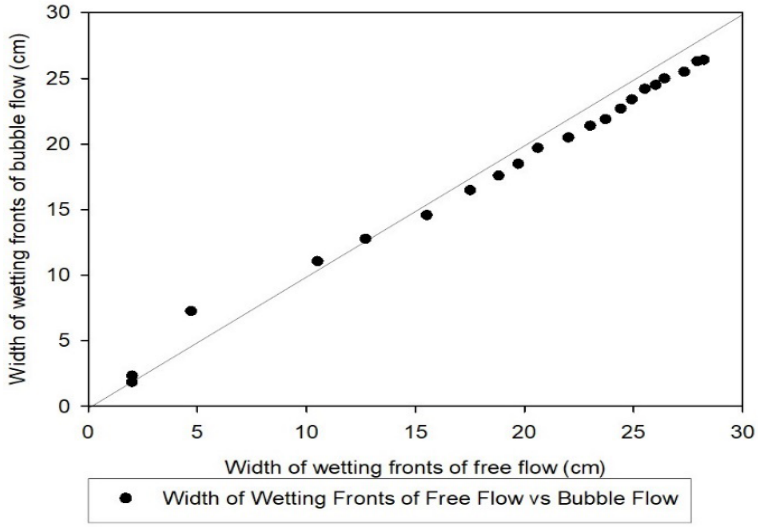


Figure 11. Comparison between the wetting front width of the bubble and free-flow irrigation systems

The results of the t-test for the second case of the wetting front width between the free-flow and bubble-flow systems showed that a t-value of 3.377, degrees of freedom of 21, and a p-value of 0.00142. Therefore, the null hypothesis was rejected and the alternative hypothesis was accepted, i.e., the mean wetting front width of the free-flow irrigation system was greater than that of the bubble-flow system. The mean of the differences was 0.772 cm. The statistical tests showed that the wetting front width of the free-flow irrigation system was significantly greater than that of the bubble-flow irrigation system. The infiltration rate of the free-flow system was also larger than that of the bubble-flow system; therefore, the response of the wetting front width was consistent.

The variation of the wetting front depth with time in both the bubble- and free-flow irrigation systems is shown in Figure 12. The wetting front depth increased sharply at the beginning of the irrigation process and then slowed down as time increased. In contrast, the wetting front depth in the bubble-flow irrigation system was, in general, lower than that of the free-flow irrigation system (Figure 13). This is one of the advantages of the bubble-flow system because it will reduce deep percolation losses and retain the nutrients within the root zone of the irrigated crop.

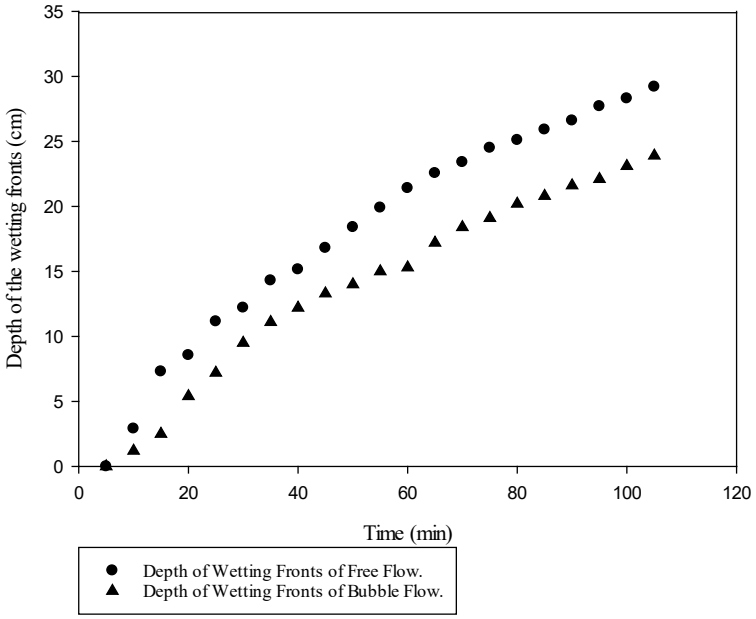


Figure 12. Variation in the wetting front depths in the bubble and free-flow irrigation systems

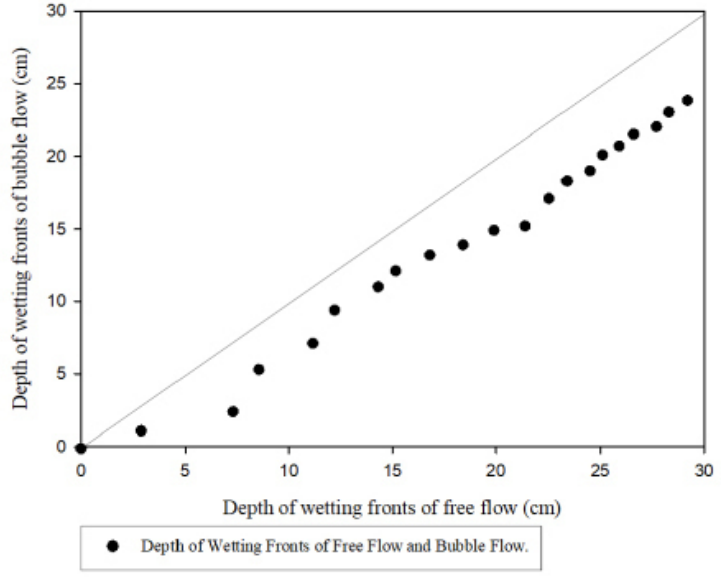


Figure 13. Comparison of the wetting front depths for the bubble and free-flow irrigation systems

The results of the t-test for the third case of the wetting front depth between the free- and bubble-flow systems showed a t-value of 11.014, degrees of freedom of 21, and a p-value of $1.738e^{-10}$. Therefore, the null hypothesis was rejected and the alternative hypothesis was accepted, i.e., the mean wetting front depth of the free-flow irrigation system was greater than that of the bubble-flow system. The mean of the differences was 4 cm. The response of the wetting front depth was consistent with the results of the infiltration rate, whereby the free-flow irrigation system was larger than that of the bubble-flow system. However, there was a great difference in depth and a slight difference in the width of wetting patterns between the free-flow and bubble-flow systems.

The statistical results indicated the differences between the behavior of these systems. The free-flow system pushes the water into the sandy soil using the free pressure head in the IOPB, which produces a higher depth than width. Yet, the bubble-flow system applies the air–water exchange and capillary pressure to reduce the pressure in the ICPB, and thus produce a higher width compared to depth. Therefore, the bubble-flow system can save water by self-regulating the wetting patterns. The experimental results from the present study showed that the bubble irrigation system controlled and almost held the wetting front compared to the free-flow irrigation system. Further studies are required to determine the bubble flow in different soil types, the contact area of infiltration, and other design parameters.

SUMMARY AND CONCLUSIONS

The main aim of the present study was to investigate the viability of using what was named 'bubble irrigation,' where water flow through the soil depends on capillary pressure and air–water interchange using an ICPB. Water flows into the soil and bubbles of airflow back into the bottle to replace the space of the infiltrated water.

The wetting front width and depth in the bubble and free-flow irrigation systems increased rapidly at the beginning of the irrigation process and then the rate of increase of the wetting front width and depth in both systems decreased with time. The wetting front width and depth in the bubble-flow system were smaller than those in the free-flow irrigation system.

Preliminary laboratory experiments indicated that the bubble irrigation system can be used as an efficient method for distributing irrigation water in the soil profile, eliminating surface runoff and deep percolation losses, and increasing water application efficiency. Field applications should be investigated in future studies to test the feasibility of such an irrigation technique and develop systematic procedures to design bubble irrigation systems.

ACKNOWLEDGEMENT

The authors are grateful to Al-Mansour University College for providing a suitable place for experiments.

REFERENCES

Al-Ghobari, H. M., & Dewidar, A. Z. (2018). Integrating deficit irrigation into surface and subsurface drip irrigation as a strategy to save water in arid regions. *Agricultural Water Management, 209*, 55-61.

Alrubaye, Y. L., Al-Tai, Z. H. A., & Al-Madhhachi, A. S. T. (2018). Laboratory and on-field experiments of measuring evaporation quantities from bare saturated soil surface. *International Journal of Civil Engineering and Technology (IJCIET), 9*(8), 1793-1801.

Amin, M. S., & Ekhmaj, A. I. (2006, September 10-16). DIPAC-drip irrigation water distribution pattern calculator. In *7th International Micro Irrigation Congress* (Vol. 1016, pp. 503-513). Kuala Lumpur, Malaysia.

Barkai, A. H., El Hajem, M., Lacassagne, T., & Champagne, J. Y. (2019). Experimental study of a gas–liquid flow in vacuum air-lift column using an optical bi-probe. *Fluids, 4*(2), 1-19.

Cai, Y., Wu, P., Zhang, L., Zhu, D., Chen, J., Wu, S., & Zhao, X. (2017). Simulation of soil water movement under subsurface irrigation with porous ceramic emitter. *Agricultural Water Management, 192*, 244-256.

Cai, Y., Zhao, X., Wu, P., Zhang, L., Zhu, D., & Chen, J. (2019a). Effect of soil texture on water movement of porous ceramic emitters: A simulation study. *Water, 11*(1), 1-13.

Cai, Y., Zhao, X., Wu, P., Zhang, L., Zhu, D., Chen, J., & Lin, L. (2019b). Ceramic patch type subsurface drip irrigation line: Construction and hydraulic properties. *Biosystems Engineering, 182*, 29-37.

Cararo, D., Botrel, T., Hills, D., & Leverenz, H. (2006). Analysis of clogging in drip emitters during wastewater irrigation. *Applied Engineering in Agriculture, 22*(2), 251-257.

Dawood, I. A., & Hamad, S. N. (2016). Movement of irrigation water in soil from a surface emitter. *Journal of Engineering, 22*(9), 103-114.

de Oliveira, F. C., Lavanholi, R., de Camargo, A. P., Ait-Mouheb, N., Frizzone, J. A., Tomas, S., & Molle, B. (2020). Clogging of drippers caused by suspensions of kaolinite and montmorillonite clays. *Irrigation Science, 38*(1), 65-75.

Elnesr, M., & Alazba, A. (2019). Computational evaluations of HYDRUS simulations of drip irrigation in 2D and 3D domains (ii-subsurface emitters). *Computers and Electronics in Agriculture, 163*, 1-12.

Fan, W., & Li, G. (2018). Effect of soil properties on Hydraulic characteristics under subsurface drip irrigation. *Earth and Environmental Science, 121*(5), 1-12.

Fan, Y. W., Huang, N., Zhang, J., & Zhao, T. (2018a). Simulation of soil wetting pattern of vertical moistube-irrigation. *Water, 10*(5), 1-19.

Fan, Y., Huang, N., Gong, J., Shao, X., Zhang, J., & Zhao, T. (2018b). A simplified infiltration model for predicting cumulative infiltration during vertical line source irrigation. *Water, 10*(1), 1-12.

Ghazouani, H., Rallo, G., Mguidiche, A., Latrech, B., Douh, B., Boujelben, A., & Provenzano, G. (2019). Assessing Hydrus-2D model to investigate the effects of different on-farm irrigation strategies on potato crop under subsurface drip irrigation. *Water, 11*(3), 1-18.

Gu, J., Yin, G., Huang, P., Guo, J., & Chen, L. (2017). An improved back propagation neural network prediction model for subsurface drip irrigation system. *Computers and Electrical Engineering, 60*, 58-65.

Gunarathna, M., Sakai, K., Nakandakari, T., Kazuro, M., Onodera, T., Kaneshiro, H., … & Wakasugi, K. (2017). Optimized subsurface irrigation system (OPSIS): Beyond traditional subsurface irrigation. *Water, 9*(8), 1-11.

Gunarathna, M., Sakai, K., Nakandakari, T., Momii, K., Onodera, T., Kaneshiro, H., … & Wakasugi, K. (2018). Optimized subsurface irrigation system: The future of sugarcane irrigation. *Water, 10*(3), 1-14.

Hatiye, S. D., Prasad, K. S. H., & Ojha, C. S. P. (2018). Deep percolation under irrigated water-intensive crops. *Journal of Irrigation and Drainage Engineering, 144*(8), 1-13.

Hilz, E., & Vermeer, A. W. (2013). Spray drift review: The extent to which a formulation can contribute to spray drift reduction. *Crop Protection, 44*, 75-83.

Hobson, P., Miller, P., Walklate, P., Tuck, C., & Western, N. (1993). Spray drift from hydraulic spray nozzles: the use of a computer simulation model to examine factors influencing drift. *Journal of Agricultural Engineering Research, 54*(4), 293-305.

Holterman, H. (2003). *Kinetics and evaporation of water drops in air* (Vol. 2012). Wageninger, The Netherlands: IMAG.

Kandelous, M. M., & Šimůnek, J. (2010). Numerical simulations of water movement in a subsurface drip irrigation system under field and laboratory conditions using HYDRUS-2D. *Agricultural Water Management, 97*(7), 1070-1076.

Kisekka, I., Oker, T., Nguyen, G., Aguilar, J., & Rogers, D. (2017). Revisiting precision mobile drip irrigation under limited water. *Irrigation Science, 35*(6), 483-500.

Klocke, N. L., Stone, L. R., & Bolton, D. A. (2009, May 17-21). *Irrigation scheduling for deficit irrigation.* In *World Environmental and Water Resources Congress 2009: Great Rivers* (pp. 1-9). Kansas City, Missouri, USA.

Lima, V., Keitel, C., Sutton, B., & Leslie, G. (2019). Improved water management using subsurface membrane irrigation during cultivation of Phaseolus vulgaris. *Agricultural Water Management, 223*, 1-11.

Liu, Y., Zhou, Y., Wang, T., Pan, J., Zhou, B., Muhammad, T., … & Li, Y. (2019). Micro-nano bubble water oxygation: Synergistically improving irrigation water use efficiency, crop yield and quality. *Journal of Cleaner Production, 222*, 835-843.

Masih, I., & Giordano, M. (2014). Constraints and opportunities for water savings and increasing productivity through resource conservation technologies in Pakistan. *Agriculture, Ecosystems and Environment, 187*, 106-115.

Miller, P. (2003). The measurement of spray drift. *Pesticide Outlook, 14*(5), 205-209.

Mohagheghian, S., & Elbing, B. (2018). Characterization of bubble size distributions within a bubble column. *Fluids, 3*(1), 1-17.

Moncef, H., Hedi, D., Jelloul, B., & Mohamed, M. (2002). Approach for predicting the wetting front depth beneath a surface point source: theory and numerical aspect. *Irrigation and Drainage: The Journal of the International Commission on Irrigation and Drainage, 51*(4), 347-360.

Nabayi, A., Teh, C., Husni, M., & Sulaiman, Z. (2018). Plant growth, nutrient content and water use of rubber (*Hevea brasiliensis*) seedlings grown using root trainers and different irrigation systems. *Pertanika Journal of Tropical Agricultural Science, 41*(1), 251-270.

Paydar, Z., Gaydon, D., & Chen, Y. (2009). A methodology for up-scaling irrigation losses. *Irrigation Science, 27*(5), 347-356.

Ren, C., Zhao, Y., Dan, B., Wang, J., Gong, J., & He, G. (2018). Lateral hydraulic performance of subsurface drip irrigation based on spatial variability of soil: Experiment. *Agricultural Water Management, 204*, 118-125.

Ren, C., Zhao, Y., Wang, J., Bai, D., Zhao, X., & Tian, J. (2017). Lateral hydraulic performance of subsurface drip irrigation based on spatial variability of soil: Simulation. *Agricultural Water Management, 193*, 232-239.

Saefuddin, R., Saito, H., & Šimůnek, J. (2019). Experimental and numerical evaluation of a ring-shaped emitter for subsurface irrigation. *Agricultural Water Management, 211*, 111-122.

Sakaguchi, A., Yanai, Y., & Sasaki, H. (2019). Subsurface irrigation system design for vegetable production using HYDRUS-2D. *Agricultural Water Management, 219*, 12-18.

Pertanika J. Sci. & Technol. 28 (4): 1315 - 1327 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Effect of Drying Temperature of Small Scale Spray Dryer on Physical Properties of Coffee Powder

Joko Nugroho Wahyu Karyadi*, Dwi Ayuni, Tsania Ayu Rohani and Devi Yuni Susanti

Departement of Agricultural and Bio-systems Engineering, Faculty of Agriculture Technology, Universitas Gadjah Mada, Jl. Flora Bulaksumur No.1, Kocoran, Caturtunggal, Depok, Sleman, Yogyakarta 5528, Indonesia

ABSTRACT

Nowadays, coffee has become one of the most favorable commodities for beverages, flavoring as well as for cosmetic industries. In Indonesia, coffee becomes more popular, especially among youngsters. In this study, the spray drying process was evaluated, giving the potential of how simple spray dryer can help local farmers of Indonesia to produce their coffee powders. One small scale of spray dryer was constructed with the total dimensions of 2.85 x 0.64 for length x width, with a height of 2.32 m. The spray dryer was equipped with a digital thermo regulator, the pneumatic nozzle system, and 4 finned heaters with the power of 2700 watt for each. The spray drying constructed was revealed to be able to produce coffee powders with fine quality. The inlet temperature of the drying chamber, as well as the initial Brix content of feed solution, were proven to affect the physical properties of powder produced such as moisture content, product yield, and solubility. The final moisture content of powders was ranged from 3 – 7% db, with the high product yield of the drying process, which could reach up to 70%.

ARTICLE INFO

Article history:
Received: 7 May 2020
Accepted: 3 August 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.10

Keywords: Coffee, product quality, product yield, spray-drying

E-mail addresses:
jknugroho@ugm.ac.id (Joko Nugroho Wahyu Karyadi)
dwi.ayuni@mail.ugm.ac.id (Dwi Ayuni)
tsania.ayu.r@gmail.com (Tsania Ayu Rohani)
deviyunisusanti@ugm.ac.id (Devi Yuni Susanti)
*Corresponding author

INTRODUCTION

Coffee is one of the most promising commodities in the world with over US$16 billion of the gross value of production, especially for green coffee. Concerning this

ISSN: 0128-7680
e-ISSN: 2231-8526

in world total, coffee production in Indonesia has a total production of more than 700,000 tonnes in 2018, putting Indonesia as the fourth highest country for coffee production (FAO, 2018). From their total production, at approximately 62% coffee in Indonesia are exported while 38% of the rest are for domestic consumption (Nugroho, 2014). However, according to Hirschmann (2019), Indonesia's domestic coffee consumption has increased almost four times now since 1990. Local coffee products are also gaining in popularity, which can be seen from the instant coffee market.

A growing youthful population and the latest trends of coffee-drinking lifestyles are some contributing factors to the increases coffee demand. Besides, coffee shops with its emphasis on buying local, single-sourced beans that continue to expand in urban areas also gave the local coffee great potential to be recognized in a wider market. As the example, one study revealed that Robusta coffee farming in East Java Province, Indonesia was proven to be profitable and feasible, and was expected to be able to give a real contribution in the national economy, such as creating added value through postharvest, processing, and distribution activities (Winarno et al., 2018).

According to the research done by Wirunthanakrit and Vittayaporn (2018), development of Instant coffee can be done to have value added coffee products. Their research revealed that producing the coffee into instant coffee increased the chance of consumers purchase up to 88%. Instant coffee, or soluble coffee, is a form of coffee available to consumers for quick hot coffee preparation. Instant coffee is made by taking extracts from coffee that has undergone a roasting process. This method was first introduced by Morgenthaler in Switzerland in 1938. The ground coffee has been extracted using certain pressures and extractors. The water steam used when taking the extract is 200°C. The dry component contained in the extracted coffee is 15%. Then the extraction results are dried using spray-dryer and freeze-dryer to form powder (Belitz & Grosch, 1987).

In the food industry, the most common process of powder formation is done by spray drying, since it has main advantages such as cost-effective, high yield, and a reduction of the product of exposure time to high temperature (Samantha et al., 2015). Spray drying is the process of converting a feed solution from a fluid into a dried particulate form by spraying the feed solution into a hot drying medium (Kuriakose & Anandharamakrishnan, 2010), through an atomization process. As the feed emulsion contact with the hot air in the chamber, the moisture is removed and each droplet dried to form the powder. There are several types of atomizers available for spray drying. Among them are the rotary, nozzle, two-fluid, pressure and ultrasonic nozzle (Walzel, 2011). Atomization is a crucial part of the spray dryer because it influences the drying rate, particle diameter, particle distribution, and dispersibility of the powder for rehydration (Chegini & Taheri, 2013). Drying airflow is another crucial factor in spray drying. The higher drying air flow rate will decrease the droplet drying time and lower the degree of water evaporation. This will lead to the

insufficient drying process, and consequently producing higher moisture content in the final product (Shishir & Chen, 2017).

Several studies have been done to evaluate the quality of coffee powder dried using mini lab spray dryer. The inlet temperature set as the drying condition were usually ranged from 150-220°C (Ishwarya & Anandharamakrishnan, 2014; Villegas-Santiago et al., 2020; Wongsa et al., 2019). One study reported that spray-dried coffee exhibited higher antioxidant activity that that in freeze-dried coffee, due to the shorter drying time (Ghirisan & Miclaus, 2017). Other studies also revealed the potential of spray-drying coffee oil to later be used as an enhancement of aroma to the coffee beverages, flavoring, as well as cosmetics (Carvalho et al., 2014; Frascareli et al., 2012; Rodrigues & Grosso, 2008).

However, in Indonesia the spray drying technology is still unaffordable for all coffee farmers, whereas only big scale factories are capable to produce instant coffee. This fact has encouraged the author to develop a simple and small scale of spray-drying that is relatively inexpensive to be made but of superior quality. Inlet temperatures of spray dryer as well as initial Brix content of feed solution are two important factors to determine the quality of the product. Therefore the objective of this research was to test the performance of spray drying constructed for coffee powder, based on the quality of the powder obtained including the moisture content, solubility, powder morphology, and color.

MATERIALS AND METHODS

Sample Preparations

Pure grinded coffee powder with specific renowned brand was purchased from a local market (Yogyakarta, Indonesia). Coffee powder was brewed in hot water (90°C) at mixing ratio between coffee and boiling water of 1:3 (w/w) before filtrating process. Diluted coffee extract obtained from the concentrated extract was used as feed solution, with three different variation of Brix content (Bx_i): 10, 20, and 30%.

Spray Drying Process

The spray dryer was designed in Laboratory of Food and Postharvest Engineering, and was manufactured in Laboratory of Energy and Agricultural Machinery of Universitas Gadjah Mada. The spray dryer had total dimensions of 2.85 x 0.64 for length x width, with height of 2.32 m. The drying chamber had diameter of 0.60 m, and the capacity spray dryer was 1.04 kg/h. Stainless steel 201 was chosen as construct material (Figure 1). The spray dryer was equipped with digital thermo regulator to set up the inlet temperature (Omron E5CSL-RTC, Omron Corporation, Kyoto, Japan), the blower to transfer hot air to the drying chamber (Sirocco DE150-6, Showa Denki Group, Osaka, Japan), the pneumatic nozzle system to form powder, 4 finned heaters with the power of 2700 watt for each, as well as cyclone to

separate the hot air from the final product. Feed solution was transferred from feed stock to nozzle using a compressor of 2 Bar pressure. In this study, the spray dryer was operated with three variations of inlet temperature (T_i): 180, 200, and 220°C.

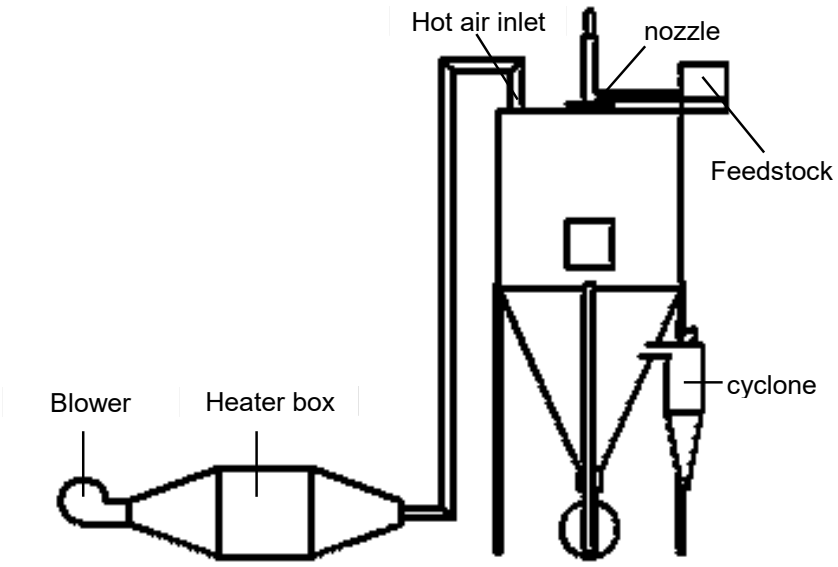


Figure 1. Spray dryer

Moisture Content

Particle moisture content was determined gravimetrically by drying in an oven (Memmert UM-400, Memmert GmbH + Co.KG, Schwabach, Germany) at 105°C for 24 hours (Park, 2008). The calculation of moisture content is presented by Equation 1.

$$M = \frac{W - W_1}{W_1} \times 100\% \quad [1]$$

where M is the moisture content (% db), W is the mass sample after spray drying process (g), and $W1$ is a dry matter of sample after 24 hours oven drying (g).

Product Yield

Product yield is the ratio of mass of powder dry matter collected after drying to the weight of total solids in feed (Tontul & Topuz, 2017), presented by Equation 2.

$$\eta_{po} = \frac{m_{po}(1 - M_{po})}{m_{feed}(1 - M_{feed})}\ x\ 100\ \% \quad [2]$$

where η_{po} is the product yield (%), m_{po} is the mass of powder, M_{po} is moisture content of powder, m_{feed} is the mass of feed solution, and M_{feed} is moisture content of feed solution.

Color Analysis

Color analysis was evaluated based on the components such as lightness (L^*), redness/greenness (a^*) and yellowness/blueness (b^*). The analysis was performed using Color Meter (Color Meter TES-135A, TES Electrical Electronic Corp., Taipei, Taiwan). Lightness, hue angle, and chroma values for each powder sample were reported. Hue angle (°) and chroma (-) were calculated using formula presented by Equation 3 and 4, respectively (Rahman et al., 2019; Da Silva & Moreira, 2008).

$$h_{ab} = \left(\frac{b^*}{a^*}\right) \quad [3]$$

$$C^* = \sqrt{\left(a^{*2} + b^{*2}\right)} \quad [4]$$

Water Solubility Index

Water solubility index (WSI) was determined using shake-flask method. 3 (three) g sample was suspended in 60 ml distilled water, using magnetic stirrer (Advantec SRS710HA, Seisakusho Kaisha Ltd., Japan) set at 50°C for 10 minutes. The suspension then was centrifuged at 1000 x for 10 minutes using table-top centrifuge (Kokusan H-27F, Kokusan Co.Ltd., Saitama, Japan). The insoluble residue was recovered and dried at 105°C for 24 hours. WSI (%) was calculated using formula presented by Equation 5 (Kang et al., 2014).

$$WSI = \frac{insoluble\ residue\ weight}{original\ sample\ weight} x100\% \quad [5]$$

Wettability

Wettability refers to the contact time of powder with water surface before it completely dissolved. The wettability was evaluated according to the method used by Edris et al. (2016). 0.5 g of the coffee powder was hydrated with 10 ml distilled water. The time needed for full dissolution after mild stirring was recorded as wetting time.

Powder Morphology

Scanning electron microscopy (SEM; JSM 6510LA JEOL Ltd., Tokyo, Japan) was used to observe the enlarged image of the powder morphology of spray-dried powder. The sample was coated with Pt-Pd using an Auto Fine Coater before placed over the SEM.

Statistical Analysis

Statistical analysis was performed using IBM SPSS Statistics 23 software (SPSS Inc., Chicago, IL). The properties of spray-dried powders were introduced to variance analysis (ANOVA, 95 % significance level). Duncan test was used to test the homogeneity of variances among samples after the conditions for homogeneity of variances were achieved.

RESULTS AND DISCUSSIONS

Physical Properties

Table 1 shows the physical properties of coffee powder resulted from the spray drying process at different inlet temperatures and initial Brix content. The physical properties observed included moisture content, product yield, solubility, and wettability. Moisture content plays an important role in coffee handling since the low moisture content will prolong the shelf-life and preserve the quality of the products (Nakilcioğlu-Taş & Ötleş, 2019). In general, the moisture content of coffee powders obtained ranged from 3 – 7 % db, with the average value different among each sample. This result is similar to that found in previous studies (Bhadra et al., 2013; Ramalakshmi et al., 2007), that revealed the moisture content of coffee powder was maximally 7%. In addition, according to the standard set by Indonesian National Quality Standard (SNI 2983:2014), the moisture content of instant coffee powder should not be more than 4 – 5%. From Table 1, it can be seen that sample with higher initial Brix content resulted in powder with the lowest moisture content, ranging from 3-5 % db ($p < 0.05$). The higher the initial Brix content, the lower the moisture content remaining in the powder.

The effect of initial Brix content had significantly affected the product moisture content rather than the variation inlet temperature. It might be due to the fact that solids content determined the total amount of moisture available for evaporation (Ferrari et al., 2012). Therefore initial Brix content is a a strong contributing factor to the product moisture content. This result is different from the study done by Tan et al. (2011), who revealed that an increase in temperature would increase the drying rate; enhancing the efficiency of heat and mass transfer.

The product yield refers to how much the powder recovery can be achieved after the spray-drying process. From Table 1, it can be seen that the product yield was mainly affected by the initial Brix content in the feed solution. The higher the Brix content, the higher the product yield obtained ($p < 0.05$). The highest yields achieved by 30% Brix were ranged from 70-75%. Increasing the Brix content will reduce the moisture content available in powder, therefore improving the drying efficiency and powder recovery. This result is in agreement with Maury et al. (2005), who stated that less water was needed to be removed from each droplet with increasing solid's content. Meanwhile from the results obtained we could not see the distinctive tendencies of product yield caused by different

Table 1

Physical properties of spray-dried powder of coffee

T_i (°C)/Bx_i (%)	Physical property			
	Moisture content (% db)	Product yield (%)	Solubility (%)	Wettability (s)
180/10	6.46 ± 0.22^{e}	31.98 ± 2.23^{a}	51.56 ± 7.44^{d}	2975.10 ± 150.30^{a}
200/10	7.21 ± 0.01^{f}	32.11 ± 2.14^{a}	42.69 ± 0.28^{abc}	2937.90 ± 96.30^{a}
220/10	5.58 ± 1.05^{d}	36.62 ± 6.47^{ab}	47.38 ± 5.95^{bcd}	3404.40 ± 378.83^{b}
180/20	5.48 ± 0.19^{cd}	41.40 ± 5.63^{b}	40.13 ± 4.90^{ab}	3685.50 ± 361.50^{b}
200/20	4.80 ± 0.32^{bc}	41.13 ± 0.14^{b}	36.48 ± 3.02^{a}	3671.10 ± 54.30^{b}
220/20	4.56 ± 0.14^{b}	51.54 ± 0.92^{c}	47.22 ± 1.73^{bcd}	3768.30 ± 143.70^{b}
180/30	4.26 ± 0.38^{b}	74.06 ± 2.58^{d}	45.96 ± 0.78^{bcd}	3462.00 ± 138.00^{b}
200/30	3.53 ± 0.19^{a}	71.79 ± 0.97^{d}	49.28 ± 1.09^{cd}	4779.60 ± 25.80^{c}
220/30	4.71 ± 0.21^{b}	71.81 ± 5.41^{d}	47.68 ± 3.54^{bcd}	4502.70 ± 301.50^{c}

Different superscript letters in the same column showed significant different at ($p < 0.05$).

T_i: inlet temperature; Bx_i: Brix content

inlet temperatures. Suzihaque et al. (2015) explained that increasing the temperature on spray drying could enhance the heat and mass transfer, but the extreme high temperature could cause the powder to melt and stick to the wall of the dryer, giving a negative effect to the product yield.

Solubility is known as the ability of the powder to fully dissolve in water to form a suspension. The wettability is the ability of the powder to overcome the surface tension between themselves and water, or a time required to give the powder wet completely. These parameters are very important as the indicator of coffee quality as drink powder. From Table 1, it can be seen that variances of inlet temperatures did not affect the solubility and wettability of powders obtained. Similar result was also reported by Al Mubarak et al. (2019), that the inlet spray drying temperature did not have a major effect on the water solubility index, bulk density, and hygroscopicity of Tamarillo powder. Meanwhile, the sample with 30% Brix were had the longest wetting time, ranged from 3500 – 4500 s. This result is similar to that in coconut powder reported by A-Sun et al. (2016), who revealed that the higher solid content would reduce the porous structure of powder. This makes the water difficult to be absorbed, therefore prolonging the time needed for powder to be wet.

Color

Figure 2 shows the results of lightness, hue angle, and chroma from the measuring of color using colorimeter. The lightness *(L*)* value of coffee powders obtained were ranged from 16 – 19, which was associated with a very dark grey color. The hue angle value (h) ranged from 0.3 – 0.5º. These values indicate that the red hue, similar to that reported by Villegas-Santiago et al. (2020). The chroma value ranged from 6 – 9. These small value of chroma indicates the color is less pure, with low saturation. In conclusion, the color of coffee powders is referred as very dark red with low saturation. The inlet temperatures and Brix contents do not give any significant difference to the color component value. This could occurr because feed solution was entirely comprised coffee and water. Otherwise, the lightness could increase due to the dilution effect caused by the addition of wall material, such as maltodextrin. Another coffee spray drying process done by Wirunthanakrit and Vittayaporn (2018) reported that the color of instant coffee with 10% maltodextrin presented a lower L (brightness) value than 15%.

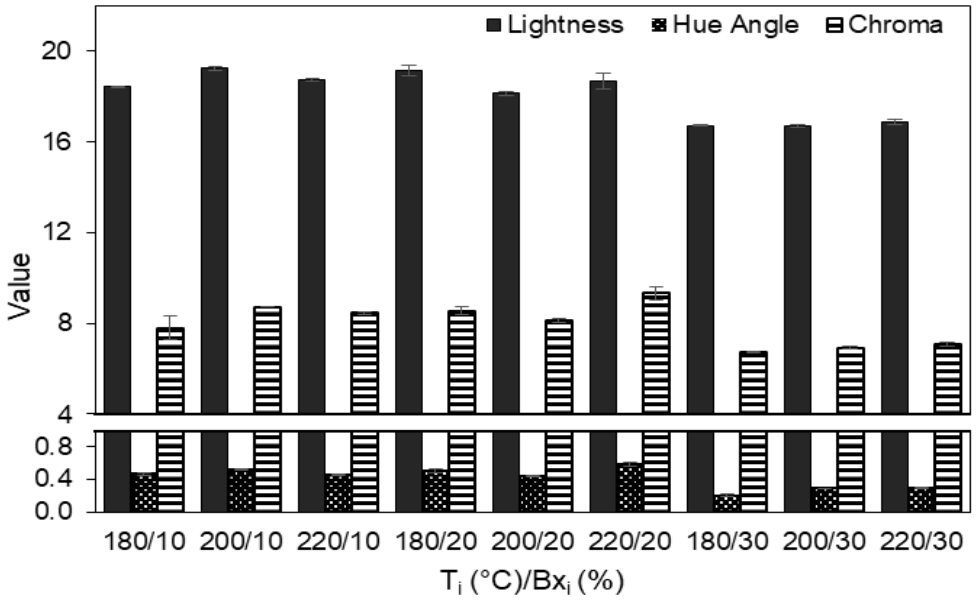


Figure 2. Color of spray-dried coffee at different inlet temperature (T_i) and Brix content (Bx_i)

Powder Morphology

Figure 3 shows the morphology of the spray-dried powders of coffee studied by scanning electron micrographs. Analysis of the shape revealed that all the powders had had almost

identical morphology, regardless of the operating conditions. The spray-dried powders were roughly spherical in shape, as described by other authors (Gallardo et al., 2013; Shamaei et al., 2017). The powders had a wide range of particle sizes, with the diameters of powders ranged from 11 – 31µm. The surface of powders covered with big "pockmarks" and wrinkles. Previous studies confirmed this, stated that the presence of wrinkles may be happened due to the mechanical stress produced by uneven drying at different parts of liquid droplets during the first step of drying (Klinkesorn et al., 2006; Pourashouri et al., 2014; Teixeira et al., 2004). The pockmarks formed on the surface were varied in diameter, ranged from 3 - 12µm. Besides, it can be presumed that the drying process has resulted in fine powders, proven by no agglomeration presented on the powder. The agglomeration could be formed due to the existence of surface oil, which caused the permanent link to form (Frascareli et al., 2012). The powders can be agglomerated and adhered one with another, formed a larger mass of powder compared to the rest samples.

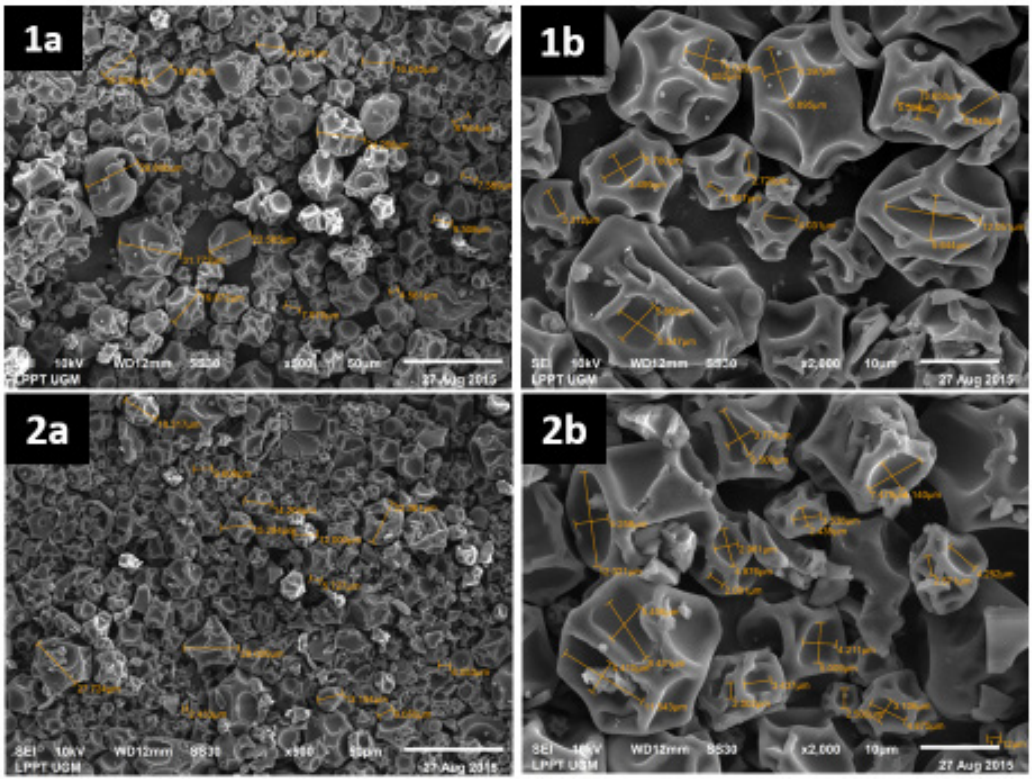


Figure 3. Micrographs of the outer morphology of coffee powders with initial Brix content of 20%, inlet temperature of 180 (1), 200 (2), and 220ºC (3) and magnification of 500 (a) and 2,000 (b)

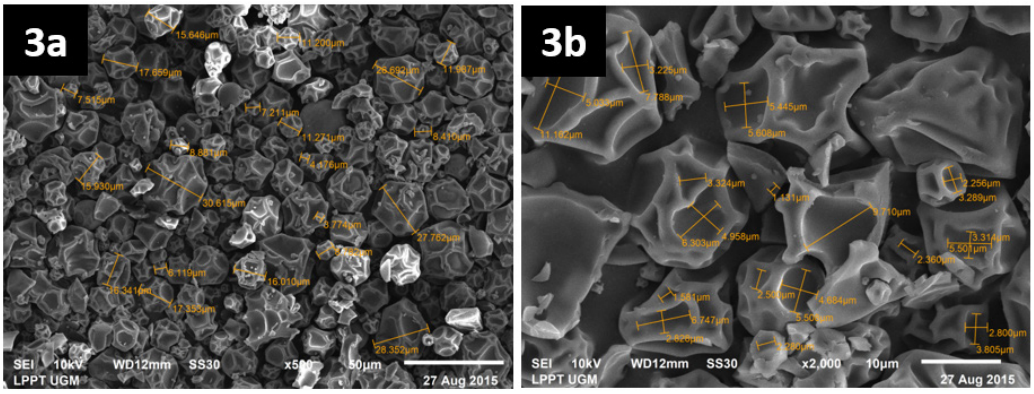


Figure 3. (Continued)

CONCLUSION

The performance evaluation of spray drying to form coffee powders was presented. The moisture content obtained ranged from 3 – 7%, with their colors, be associated as dark red with low saturation. The SEM results indicate the powders to be finely formed with no agglomeration. The variation of Brix content gave more significant effect to the physical properties of powder compared to the variation of inlet temperature. Feed solution with 30% Brix resulted in lowest moisture content and highest product yield. This study has suggested that the simple spray drying can be constructed and work effectively to form coffee powders, to help farmers to gain the value-added for their coffee products.

ACKNOWLEDGMENT

We thank Prof. Bambang Purwantana and technicians of Laboratory of Energy and Agricultural Machinery of Universitas Gadjah Mada who have supported in construction process of spray dryer. This research was funded by the Faculty of Agricultural Technology of Universitas Gadjah Mada.

REFERENCES

A-Sun, K., Thumthanaruk, B., Lekhavat, S., & Jumnongpon, R. (2016). Effect of spray drying conditions on physical characteristics of coconut sugar powder. *International Food Research Journal*, *23*(3), 1315-1319.

Al Mubarak, A., Hamid, N., Kam, R., & Chan, H. (2019). The effects of spray drying conditions on the physical and bioactive properties of New Zealand Tamarillo (*Solanum betaceum*) powder. *Acta Scientific Nutritional Health*, *3*(12), 121-131.

Belitz, H. D., & Grosch, W. (1987). *Food chemistry*. Berlin, Germany: Springer Verlag.

Bhadra, R., Rosentrater, K. A., & Muthukumarappan, K. (2013). Measurement of sticky point temperature of coffee powder with a rheometer. *International Journal of Food Properties*, *16*(5), 10711079.

Carvalho, A. G. S., Silva, V. M., & Hubinger, M. D. (2014). Microencapsulation by spray drying of emulsified green coffee oil with two-layered membranes. *Food Research International*, *61*, 236-245.

Chegini, G., & Taheri, M. (2013). Whey powder: Process technology and physical properties: A review. *Middle East Journal of Scientific Research*, *13*(10), 1377-1387.

Da Silva, P. F., & Moreira, R. G. (2008). Vacuum frying of high-quality fruit and vegetable-based snacks. *LWT - Food Science and Technology*, *41*(10), 1758-1767.

Edris, A. E., Kalemba, D., Adamiec, J., & Piaotkowski, M. (2016). Microencapsulation of Nigella sativa oleoresin by spray drying for food and nutraceutical applications. *Food Chemistry*, *204*, 326-333.

FAO. (2018). *Production of Coffee, green: Top 10 producers*. Retrieved May 5, 2020, from www.fao.org/faostat/en/#data/QC

Ferrari, C. C., Germer, S. P. M., & de Aguirre, J. M. (2012). Effects of spray-drying conditions on the physicochemical properties of blackberry powder. *Drying Technology*, *30*(2), 154-163.

Frascareli, E. C., Silva, V. M., Tonon, R. V., & Hubinger, M. D. (2012). Effect of process conditions on the microencapsulation of coffee oil by spray drying. *Food and Bioproducts Processing*, *90*(3), 413-424.

Gallardo, G., Guida, L., Martinez, V., López, M. C., Bernhardt, D., Blasco, R., … & Hermida, L. G. (2013). Microencapsulation of linseed oil by spray drying for functional food application. *Food Research International*, *52*(2), 473-482.

Ghirisan, A., & Miclaus, V. (2017). Comparative study of spray-drying and freeze-drying on the soluble coffee properties. *Studia UBB Chemia*, *4*, 309-316.

Hirschmann, R. (2019). *Total coffee consumption in Indonesia 1990 - 2018*. Retrieved May 5, 2020, from https://www.statista.com/statistics/314982/indonesia-total-coffee-consumption/

Ishwarya, S. P., & Anandharamakrishnan, C. (2014). Spray-freeze-drying approach for soluble coffee processing. *Journal of Food Engineering, 149*, 171-180.

Kang, O. L., Yong, P. F., Ma'aruf, A. G., Osman, H., & Nazaruddin, R. (2014). Physicochemical and antioxidant studies on oven-dried, freeze-dried and spray-dried agaro-oligosaccharide powders. *International Food Research Journal*, *21*(6), 2363-2367.

Klinkesorn, U., Sophanodora, P., Chinachoti, P., Decker, E. A., & McClements, D. J. (2006). Characterization of spray-dried tuna oil emulsified in two-layered interfacial membranes prepared using electrostatic layer-by-layer deposition. *Food Research International*, *39*(4), 449-457.

Kuriakose, R., & Anandharamakrishnan, C. (2010). Computational fluid dynamics (CFD) applications in spray drying of food products. *Trends in Food Science and Technology*, *21*(8), 383-398.

Maury, M., Murphy, K., Kumar, S., Shi, L., & Lee, G. (2005). Effects of process variables on the powder yield of spray-dried trehalose on a laboratory spray-dryer. *European Journal of Pharmaceutics and Biopharmaceutics*, *59*(3), 565-573.

Nakilcioğlu-Taş, E., & Ötleş, S. (2019). Physical characterization of Arabica ground coffee with different roasting degrees. *Anais Da Academia Brasileira de Ciencias*, *91*(2), 1-15.

Nugroho, A. (2014). The impact of food safety standard on Indonesia's coffee exports. *Procedia Environmental Sciences*, *20*, 425-433.

Park, Y. (2008). Moisture and water activity. In L. M. L. Nollet & F. Toldra (Eds.) *Handbook of processed meats and poultry analysis* (pp. 35-67). Boca Raton, Florida: CRC Press.

Pourashouri, P., Shabanpour, B., Razavi, S. H., Jafari, S. M., Shabani, A., & Aubourg, S. P. (2014). Impact of wall materials on physicochemical properties of microencapsulated fish oil by spray drying. *Food and Bioprocess Technology*, *7*(8), 2354-2365.

Rahman, N. F. A., Ismail, A., Shah, N. N. A. K., Varith, J., & Shamsudin, R. (2019). Effect of drying temperature on Malaysia pomelo (Citrus grandis (L.) osbeck) pomace residue under vacuum condition. *Pertanika Journal of Science and Technology*, *27*(S1), 57-66.

Ramalakshmi, K., Kubra, I. R., & Rao, L. J. M. (2007). Physicochemical characteristics of green coffee: Comparison of graded and defective beans. *Journal of Food Science*, *72*(5), S333-S337.

Rodrigues, R. A. F., & Grosso, C. R. F. (2008). Cashew gum microencapsulation protects the aroma of coffee extracts. *Journal of Microencapsulation*, *25*(1), 13-20.

Samantha, S. C., Bruna, A. S. M., Adriana, R. M., Fabio, B., Sandro, A. R., & Aline, R. C. A. (2015). Drying by spray drying in the food industry: Micro-encapsulation, process parameters and main carriers used. *African Journal of Food Science*, *9*(9), 462-470.

Shamaei, S., Seiiedlou, S. S., Aghbashlo, M., Tsotsas, E., & Kharaghani, A. (2017). Microencapsulation of walnut oil by spray drying: Effects of wall material and drying conditions on physicochemical properties of microcapsules. *Innovative Food Science and Emerging Technologies*, *39*, 101-112.

Shishir, M. R. I., & Chen, W. (2017). Trends of spray drying: A critical review on drying of fruit and vegetable juices. *Trends in Food Science and Technology*, *65*, 49-67.

Suzihaque, M. U. H., Hashib, S. A., & Ibrahim, U. K. (2015). Effect of inlet temperature on pineapple powder and banana milk powder. *Procedia - Social and Behavioral Sciences*, *195*, 2829-2838.

Tan, L. W., Ibrahim, M. N., Kamil, R., & Taip, F. S. (2011). Empirical modeling for spray drying process of sticky and non-sticky products. *Procedia Food Science*, *1*(May 2015), 690-697.

Teixeira, M. I., Andrade, L. R., Farina, M., & Rocha-Leao, M. H. M. (2004). Characterization of short chain fatty acid microcapsules produced by spray drying. *Materials Science and Engineering C*, *24*(5), 653-658.

Tontul, I., & Topuz, A. (2017). Spray-drying of fruit and vegetable juices: Effect of drying conditions on the product yield and physical properties. *Trends in Food Science and Technology*, *63*, 91-102.

Villegas-Santiago, J., Gómez-Navarro, F., Domínguez-Niño, A., García-Alvarado, M. A., Salgado-Cervantes, M. A., & Luna-Solano, G. (2020). Effect of spray-drying conditions on moisture content and particle size of coffee extract in a prototype dryer. *Revista Mexicana de Ingeniería Química*, *19*(2), 767-781.

Walzel, P. (2011). Influence of the spray method on product quality and morphology in spray drying. *Chemical Engineering and Technology, 34*(7), 1039-1048.

Winarno, S. T., Darsono, Harisudin, M., & Sudiyarto. (2018). Competitiveness analysis of Robusta coffee in East Java, Indonesia. *Academy of Strategic Management Journal*, *17*(6), 1-10.

Wirunthanakrit, W., & Vittayaporn, V. (2018). Development of instant coffee for value adding of coffee products from Ban Borsiliam, Ngao District, Lampang Province. *Journal of Food Health and Bioenvironmental Science*, *11*(3), 32-40.

Wongsa, P., Khampa, N., Horadee, S., & Chaiwarith, J. (2019). Quality and bioactive compounds of blends of Arabica and Robusta spray- dried coffee. *Food Chemistry*, *283*(May 2018), 579-587.

Pertanika J. Sci. & Technol. 28 (4): 1329 - 1343 (2020)

SCIENCE & TECHNOLOGY

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

Food Variety of Lesser Whistling Duck in Malaysian Lakes

Martins Chukwuemeka Onwuka[1], Muhammad Nawaz Rajpar[2] and Mohamed Zakaria[1]*

[1]Faculty of Forestry and Environment, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Department of Forestry, Faculty of Life Sciences, Shaheed Banazir Bhutto University Sheringal, Upper Dir, Pakistan

ABSTRACT

Food distribution and diversity is a significant factor that determines the habitat and site selection of avian species. Its effects on the health, reproduction, survival rate, diversity indices, population structure and home range of avian species. In wetland ecosystem, bird home range and population structure is influenced by richness and diversity of food resources and availability of suitable foraging sites. The aim of the study was to ascertain the food variety of Lesser Whistling Duck – *Dendrocygna javanica* in Paya Indah Wetland Reserve (PIWR). A scan method was employed from strategic places or blinds using a spotting scope and binocular to determine the food selection and density by Distance sampling point count method. Food items were categorised into the aquatic plants (AP), aquatic invertebrates (AIV), aquatic vertebrates (AV), terrestrial plants (TP), terrestrial vertebrates (TV) and terrestrial invertebrates (TIV). Distance analysis indicated that PIWR harboured 3.88 ± 0.00 birds per ha (n = 188 individuals). In addition, the results revealed that higher bird relative abundance of Lesser Whistling Duck concentrated in the shallow of Belibis lake that was rich in submerged and emergent vegetation (*Eleocharis dulcis*, *Philydrum lanuginosum, Utricularia vulgaris*, and *Potamogeton perfoliatus*). Kruskal–Wallis *H* test showed that food items were significantly different. Lesser Whistling Duck showed strong correlationship with AIV (r^2 = 1.00, P < 0.05) and negative relationship with water TP (r^2 = –0.061, P< 0.05). The regression model highlighted that Lesser Whistling

ARTICLE INFO

Article history:
Received: 5 May 2020
Accepted: 27 July 2020
Published: 11 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.11

E-mail addresses:
martinszy@outlook.com (Martins Chukwuemeka Onwuka)
rajparnawaz@gmail.com (Muhammad Nawaz Rajpar)
mzakaria@upm.edu.my (Mohamed Zakaria)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

Duck significantly preferred aquatic invertebrates, $r^2 = –0.686±0.68$; $P<0.001$ and aquatic vertebrates, –0.459±0.26) than other food items. The results revealed that shallow marshy-based lakes rich in aquatic vertebrates and invertebrates harboured higher population of Lesser Whistling Ducks to utilize it and performed multiple activities than other habitats.

Keyword: Density, duck, food, lakes, relative abundance, wetland habitat

INTRODUCTION

Lesser Whistling Duck (*Dendrocygna javanica*) is a waterbird with square head, pale buff face with a restricted yellow eye-rings, round, and more extensive wings, dim dark bill, leg, and feet (Zakaria & Rajpar, 2015). The colour of the body is commonly brownish, especially around the chest, while the other parts of the body may be light cinnamon with unnoticeable whitish flank quills. Lesser Whistling Duck are mostly found foraging around freshwater wetlands, for example, lakes, repositories, and swamps filled with aquatic vegetation, i.e., *Eleocharis dulcis, Stenochlaena palustris, Scleria purpurascens* and *Philydrum lanuginosum*. In addition, they feed on different pieces of aquatic plants (e.g., seed, shoots, tuber, and leaves). They also forage on different sources of aquatic vertebrates (fishes and creatures of land and water) and invertebrates, which constitute a significant portion of their diet. Their population is large in dense vegetation or grasslands, as dense vegetation keeps them safe from predators and provide a safe environment for nesting (Tellkamp, 2004; Zakaria & Rajpar, 2014). Although the current conservation status of this species is "least- concern", this species is currently declining in number, and if studies are not carried out to understand daily foraging requirement, they may become threatened (BirdLife International, 2016).

Family Anatidae (geese, ducks, and swans) comprises 49 genera and 148 species. Out of 148 species, 48 species have become threatened and endangered species due to human footprints, i.e. 27 vulnerable, 12 endangered, 9 critically endangered and 7 extinct (Andy, 1996; Howard, 2003; Martins et al., 2019). Currently, the members of anatidae are facing severe threats due to human interference, i.e. habitat loss due to urbanization and agriculture expansion (73.0%), illegal hunting (48.0%), and introduction of exotic species (31.0%) (Thiel et al., 2007; Asmawi, 2007; Johnsgard, 2010). Habitat loss and degradation may cause shift in home range, i.e. movements to unsuitable and less productive areas that ultimately affect reproduction success, increase chances of hunting and trapping (Shuford & Gardali, 2008; Gillespie, 2007). It has been stated that lesser whistling ducks spend most of their time to refuel the energy to maintain vigorous health and perform multiple activities for their survival and existence (Strasser & health, 2013; Rehnus et al., 2014).

The main threats to Lesser Whistling Duck are illegal hunting, trapping, habitat loss, habitat degradation and human intervention (Rajpar et al., 2017). There is a high incidence

of sedimentation, as well as alien and invasive species invasion on the wetland, creating less space or suitable habitat for waterbirds. The duck may be affected by excessive use of agro-chemicals in the farmlands. According to Rajpar and Zakaria, (2014), reduced quality of foraging habitat in terms of food resources, protection from prey while foraging is another major factor affecting the population of Lesser Whistling Duck in most Malaysian wetland (i.e. the ducks may not have adequate source of food for their daily requirement). Hence, understand diet composition is of great importance, as it will help ecologist to know the exact foraging types and manipulate their wetlands to enhance suitable diet for this species. In addition, there is potential for Paya Indah Wetland Reserve (PIWR) Reserve as staging ground for other ducks and other waterbird species that are migratory or native threatened. The conservation and protection of waterbirds especially game birds are crucial challenge for wildlife managers because they need immense work, mass awareness and efforts to protect these species. Hence, the aim of this study was to ascertain the food variety consumed by Lesser Whistling Duck in (PIWR).

METHODOLOGY

Study Site

This study was carried out at 14 lakes at the Paya Indah Wetland Reserve that covers an area of 3,050 ha. Out of the 3,050 ha, 450 ha area is under the administration of Wildlife Department while the rest is private owned and state land. The study area is located within 101°10′ to 101°50′ longitude and 2°50′ and 3°00′ latitude (Figure 1). The wetland is characterized by different types of habitats (Table 1). All plants were identified using "Flora of Peninsular Malaysia, Series II: Seed Plants, Volume 7" (Kiew et al., 2018). These lakes vary in size, vegetation structure, and composition, water depth, water quality, inflow, and outflow of water (Table 2). The dominant vegetation comprises *Eleocharis dulcis, Stenochlaena palustris, Philydrum lanuginosum, Nelumbo nucifera, Nymphaea pubescens* and *Scleria purpurascens*), lowered vegetation such as *Potamogeton perfoliatus, Myriophyllum spicatum, Utricularia vulgaris* and *Salvinia molesta.* The wetland edges encompassing of *Scirpus olneyi, Stenochlaena palustris, Phragmites karaka* and *Typha angustifolia*. PIWR serve as the green lung or super corridor for a wide array of avian species due to its strategic location, (i.e. 12 km to the west of Putrajaya, 30 km to the south of Kuala Lumpur, 15 km to the north of the International Airport in Kuala Lumpur, 4 km to Dengkil), Kuala Langat Peat Swamp Forest Reserve (in the east), Ayer Hitam Isolated Tropical Lowland Forest Reserve (Selangor), and Sungai Lalang Forest Reserve (in the west) (Rajpar & Zakaria, 2012; Martins et al., 2017).

Table 1

Different habitat types of Paya Indah Wetland Reserve

Habitat Type	Plants found in the Habitat
Marshy Swamp	*Lepironia articulata, Eleocharis dulcis,Stenochlaena palustris, Scirpus spp., Philydrum lanuginosum, Hydrilla* spp*., Carex* spp*., Sagittaria latifolia, Panicum repends, Nymphaea pubescens, Scleria purpurascens, Phragmites karka, Nymphaea rubra, Nelumbo nucifera, Gleichenia linearis, Lycopodium cernuum and scattered trees such as Acacia auriculiformis, A. mangium, Macaranga tanarius, Peltophorum pterocarpum, Cinnamonum iners, Melicope glabra and Melastoma malabathricum*.
Lotus swamp	*Nelumbo nucifera, Nelumbo nouchali, Nelumbo pubescens, Eleocharis dulcis, Elodea* spp*., Lepironia articulata*, *Phragmites karka* reeds and *Typha angustifolia*.
Open water body	*Nymphaea odorata, Potamogeton* spp*., Eleocharis dulcis, Myriophyllum spicatum, Salvinia molesta, Utricularia aurea, Scirpus holoschoenus, Scirpus sylvaticus, Scirpus californicus, Scirpus mucronatus* and *Scirpus maritimus*. The edges are predominated by *Eleocharis dulcis, Lepironia articulata*, *Philydrum lanuginosum, Scleria purpurascens, Scirpus* spp*., Carex* spp*., Sagittaria latifolia* and *Hydrilla* spp.
Dry land with scattered Vegetation	*Mimusops elengi, Fragraea fragrans, Cassia fistula, Tectona* spp*., Albizia julibrissin*, *Syzygium* spp., *Delonix regia, Samanea saman, Acacia auriculiformis, Acacia mangium, Melicope glabra, Melastoma malabathricum* and *Ficus* spp., *Imperata cylindrica, Cynodon dactylon, Wedelia trilobata, Nephrolepis acutifolia, Artocarpus altilis, Asystasia gangetica, Peltophorum pterocarpum, Plumeria obtuse* and *Passiflora caerulea.*
Shrub Patches	*Acacia auriculiformis, Acacia mangium, Fragraea fragrans, Delonix regia, Alstonia scholaris, Samanea saman, Macaranga lanrius, Ficus rubiginosa, Ficus benjamina, Ficus fistulosa, Lagerstroemia speciosa*, *Melastoma malabathricum, Wedelia trilobta, Nephrolepis acutifolia* and *Asystasia gangetica*.

Table 2

Habitat structure of 14 lakes in Paya Indah wetland Malaysia

Names of lakes	Vegetation type	Water Level (ft)		Size (ha)	WQI	WLF	NDVI
		Minimum	Maximum				
Belibis	Marsh swamp vegetation	0.65	3.38	5.09	88.53	7.81	0.51
Seroja	Open water vegetation	5.03	18.95	138.93	82.0	8.91	0.24
Telipok	Open water vegetation	3.84	20.74	85.09	83.6	7.14	0.33
Drift Wood	Lotus swamp vegetation	2.39	9.85	10.84	73.06	7.41	0.32
Tunira	Lotus swamp vegetation	1.47	5.46	7.50	80.22	7.55	0.54
Senduduk	Open water vegetation	0.73	3.10	7.52	50.26	6.56	0.53
Sendayan	Open water vegetation	2.98	10.89	28.17	82.24	6.5	0.21
Grebe	Lotus swamp vegetation	5.94	18.97	2.40	52.19	7.37	0.31
Resam	Lotus swamp vegetation	3.17	13.84	4.31	84.55	7.06	0.24
Teratai	Open water vegetation	1.54	20.60	149.04	82	5.86	0.33
Kemoning	Lotus swamp vegetation	2.92	8.23	1.91	85.49	6.65	0.27
Rusiga	Lotus swamp vegetation	3.38	16.79	37.45	72.3	7.07	0.36
Typha1	Lotus swamp vegetation	3.21	10.92	5.07	81.48	5.99	0.23
Typha2	Lotus swamp vegetation	1.19	5.90	6.81	72.3	7.07	0.36

Source: Martins et al.(2017) and Martins et al.(2019)

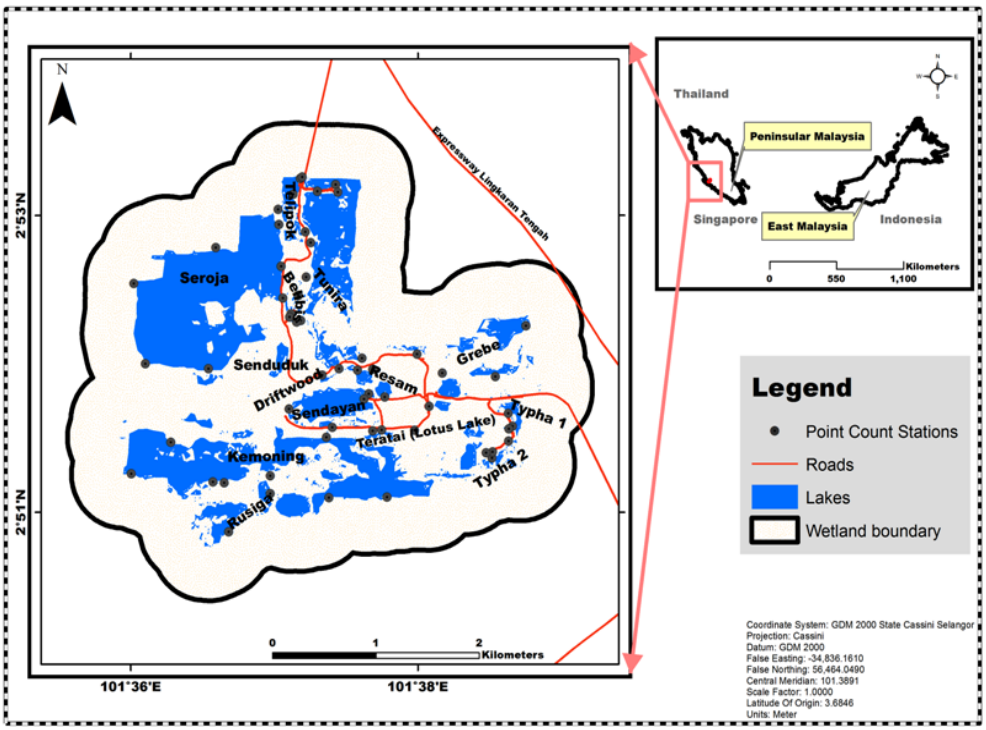


Figure 1. Map of 14 lakes in Paya Indah wetland

Foraging Ecology and Diet Composition of Lesser Whistling Duck

There are several sampling techniques to examine the foraging ecology of Anatidae. For foraging ecology of Lesser Whistling Duck, the scan through method was employed from strategic places (point count stations) using the spotting scope and binocular. All food types observed by Lesser Whistling Duck during the study were classified into the following diet composition classes, namely aquatic plants (AP), aquatic invertebrates (AIV), aquatic vertebrates (AV), terrestrial plants (TP), terrestrial vertebrates (TV), and terrestrial invertebrates (TIV). Each lake was observed daily from April to September 2019. Data collected were population of Lesser Whistling Duck in each lake and types of food seen eaten.

The distance sampling point count technique was employed. About 57 count stations in PIWR were systematically placed based on their visibility using binoculars and at least 100 m interval apart, to avoid the double count of the same avian species at more than one station. Bird count surveys in each count station with the maximum variable radius of 100m. The lakes were observed daily for 10 minutes each between 6 and 11am and between 5 and 7pm. Observation was carried out from a blind location (i.e., from a hideout to avoid disturbance). Thus, we avoided recording the foraging behaviour of Lesser Whistling

Duck individuals that occured outside of scan intervals. The methodology followed was as described by Boyce, (2010) and Rajpar and Zakaria, (2014).

Data Analysis

The distance software Version 7.2 was used to determine the population densities and detection functions (Equations 1, 2 & 3) (Thomas et al., 2010; Sebastián-González et al., 2018). According to Buckland et al. (2008), the distribution of the observed distances was used to estimate the "detection function," g(y) - the probability of detecting a bird at distance y. This function can be used to estimate the average probability of detecting a bird (denoted Pa), given that it is within mean radial distance to the point.

$$D = \frac{1}{a}\sum_{i=1}^{n}\frac{1}{p_a(z_i)} \quad (1)$$

Where, a = size of the covered region, n = number of birds seen, $P_{a(zi)}$ = the estimated probability of detecting the bird.

$$ER = \frac{\mathrm{n}}{\mathrm{K}} \quad (2)$$

Where, ER = Encounter rate (per point), n = total number of observed individuals in each point station, K = Point count effort (number of samples)

$$EDR = w(\sqrt{p}) \quad (3)$$

Where, EDR = Effective detection radius in meters, p = Detection probability, w = Radius of point transect in meters. However, six lakes were not analysed because there was no presence of Lesser Whistling Duck in the vicinity.

A Kruskal–Wallis *H* test was applied to examine if there was a significant difference between the diet composition classes of Lesser Whistling Duck among the lakes. After data transformation, Predictive Regression Model (PRM) was used to examine the correlation of Lesser Whistling Duck population with various food type classes, namely AIV, AV, AP, TIV, TV and TP in order to understand the food requirements of Lesser Whistling Duck. In addition, Pearson's Correlation Coefficient was also employed to examine the correlations between Lesser Whistling Duck density with the different food class types (Equation 4).

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \quad (4)$$

Where; *Xi*=standard score, *X*=sample mean, and *SX*=standard deviation

RESULTS

Bird Density

A total 188 bird individuals of Lesser Whistling Duck (density 3.88±0.00 birds per ha) were detected during the sampling period in PIWR (Table 3). Table 4 shows the monthly-observed density of Lesser Whistling Duck in each lake. The result reveals that Belibis harboured the highest density of Lesser Whistling Duck as compared to the other lakes. However, Lesser Whistling Duck avoided utilizing the six lakes and were not analysed.

Table 3

Density estimate of Lesser Whistling Duck in PIWR

Total Observation of Lesser Whistling Duck	188 bird individuals
Density	3.88 ± 0.00
Encounter rate (per meter)	0.01 ± 0.00
Detection probability	0.29 ± 0.00
Effective detection radius	2.18 ± 0.00

Table 4

The density of Lesser Whistling Duck in PIWR

Name of Lake	Monthly density of Lesser Whistling Duck individuals/months					
	April	May	June	July	August	September
Belibis	3.26±0.12	3.13±0.05	2.84±0.62	3.44±0.23	2.98±0.56	3.53±0.15
Tunira	0.93 ± 0.20	0.33±0.01	0.71±0.23	0.11±0.12	0.25±0.63	0.65±0.34
Senduduk	0.89 ±0.11	0.21±0.36	0.25±0.37	0.45±0.55	0.69±0.65	0.59±0.57
Grebe	0.42± 0.09	0.23±0.48	0.55±0.66	0.63±0.56	0.96±0.68	0.88±0.87
Resam	0.55± 0.09	0.66±0.44	0.98±0.46	0.88±0.55	0.22±0.23	0.87±0.65
Teratai	0.32± 0.31	0.52±0.25	0.23±0.33	0.43±0.11	0.26±0.04	0.22±0.05
Rusiga	0.25± 0.13	0.66±0.32	0.14±0.66	0.21±0.05	0.22±0.01	0.41±0.23
Typha1	0.88± 0.31	0.22±0.11	0.85±0.26	0.24±0.65	0.23±0.55	0.21±0.54

Foraging Ecology of Lesser Whistling Duck

Visual estimation and scan method revealed that Lesser Whistling Duck was a gregarious omnivorous species, mostly preferred to forage in open water column through dabbling, dipping, skimming, and even some time half diving in water column. It was observed that Lesser Whistling Duck preferred areas with submerged and emerged aquatic plants, i.e.,

Eleocharis dulcis, Philydrum lanuginosum, Utricularia vulgaris, *Potamogeton perfoliatus* etc. Furthermore, it was observed that Lesser Whistling Duck were also concentrated in shallow waters along the edges of the lakes where aquatic invertebrates (molluscs and worms) and aquatic vertebrates (tadpoles and small fishes) were present (Figure 2).

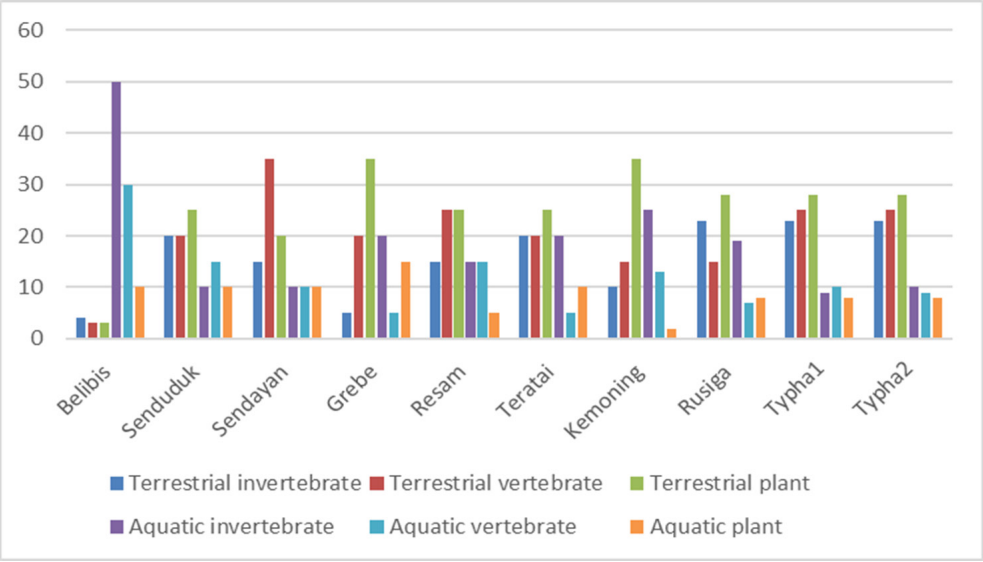


Figure 2. Food type foraged by Lesser Whistling Duck

Table 5

Kruskal–Wallis H test showing the temporal differences in diet composition of the ducks

Source of Variation	SS	df	MS	F	P-value	F crit
Between groups	2332.3	5	466.46	6.346395	0.00069	2.620654
Within groups	1764	24	73.5			
Total	4096.3	29				

Table 5 shows the Kruskal–Wallis *H* test result among the diet composition classes; the result revealed that there was a significant difference among the different diet composition classes. Pearson Correlation Coefficient was applied to determine the relationship density with the food types (Table 6). Lesser Whistling Duck showed strong relationship with aquatic invertebrates ($r2 = 1.00$, $P > 0.05$). Table 7 shows the predictive regression model correlation matrix. The result showed that there was a strong correlation of Lesser Whistling Duck with aquatic invertebrates and vertebrates.

Table 6

Pearson Correlation Coefficient test results indicating correlation of Lesser Whistling Duck with the diet composition classes

Name of food type	Association with Lesser Whistling Duck
Aquatic invertebrates (AIV)	Strong correlation with AIV ($r2 = 1.00$, $P > 0.05$)
Aquatic vertebrates (AV)	A positive relationship with water AV ($r2 = 0.595$, $P > 0.05$)
Aquatic plants (AP)	Moderate correlation with AP ($r2 = 0.640$, $P > 0.05$)
Terrestrial invertebrates (TIV)	A moderate negative relationship with TIV ($r2 = -0.651$, $P > 0.05$
Terrestrial vertebrates (TV)	A weak positive relationship with water TV ($r2 = 0.343$, $P > 0.05$)
Terrestrial plants (TP)	Weak negative relationship with water TP ($r2 = -0.061$, $P > 0.05$)

Table 7

Results of predictive regression model correlation matrix results

Food Type	Terrestrial invertebrate	Terrestrial vertebrate	Terrestrial plants	Aquatic invertebrate	Aquatic vertebrate	Aquatic plants
Terrestrial invertebrate	1					
Terrestrial vertebrate	0.888058	1				
Terrestrial plant	0.895427	0.981274	1			
Aquatic invertebrate	0.025659**	0.193363	0.155506	1		
Aquatic vertebrate	0.040306**	-0.36174	-0.35284	0.780217	1	
Aquatic plant	0.577926	0.166542	0.174283	0.906403	0.642829	1

Table 8

Slope and standard errors of ecological factors in parsimonious model

Food Type variable	Slope	Standard Error
Terrestrial invertebrate	0.587	0.524
Terrestrial vertebrate	0.356	0.236

Table 8 *(Continued)*

Food Type variable	Slope	Standard Error
Terrestrial plant	0.236	0.014
Aquatic invertebrate	-0.686	0.68
Aquatic vertebrate	-0.459	0.26
Aquatic plant	0.164	0.52

In addition, predictive regression model analysis test was carried out to ascertain the diet composition class variables that was most important and play a major role in foraging of Lesser Whistling Duck within the wetlands (Table 7). The results highlighted that the ducks significantly preferred more of aquatic-based food (i.e. aquatic invertebrates -0.686±0.68; $P<0.001$) (Table 8).

DISCUSSION

Lesser Whistling Duck are omnivorous in nature and the have a varied diet. The different diet composition sources provide good nutrition for the health of Lesser Whistling Duck and positively affect the species' growth, feather strength, muscle development, and breeding success. Understanding the different sources of food eaten by Lesser Whistling Duck will help conservationist to maintain the habitat that supports food resources. The diet composition classes were classified into six classes and this were based on feeding observations of this species. The wetland (PIWR) has 14 lakes, and Lesser Whistling Duck was found foraging in 8 of them. These lakes provide different sources of food, especially Lake Belibis. A large population of Lesser Whistling Duck was present in lake Belibis. This may be due to its richness in aquatic vertebrates and invertebrates (Figure 2), which agrees with Rajpar and Zakaria (2014), who stated that the presence of aquatic food resources was a major player for the distribution of water birds in various wetlands. Although some of the food classes and shelter vegetation might have been present in the avoided lakes, Lesser Whistling Duck still completely avoided six lakes which were characterized as having higher water level (deeper waters). Hence, this may have affected the population of Lesser Whistling Duck in these lakes. Hansson et al., (2010) stated that water quality parameter such as dissolved oxygen, water depth, salinity would affect the presence of macrophytes and aquatic invertebrates in wetlands, hence this might be the major reason why Lesser Whistling Duck avoided lakes that were not rich with this food source. It was also reported that water depth was an important variable affecting the habitat selection in water birds (Koli, 2014; Sulaiman et al., 2018) because it directly determined the accessibility of prey while foraging (McMahon & Moreira, 2014; Jayathilake & Chandrasekara, 2015). Wading

birds prefer shallow water because foraging efficiency decreases with increasing water depth. This indicated that water birds obtain higher net energy intake in shallower water than deeper water, as reported by Ma et al., (2009) and Sulaiman et al., (2018). It was observed that water level fluctuation created different foraging habitats from time to time, depending on inflow of water from catchment areas and rainfall pattern. When water level is reduced, the food resources such as fishes and tadpoles are concentrated in low–lying sites attracting a higher number of wading birds due to visibility of prey resources and increased foraging success.

Our study showed that both aquatic vertebrates and aquatic invertebrates were significantly associated (Table 6 and 7). Some aquatic plants provide sheltering, protection, nutrition and reproduction habitat for Lesser Whistling Duck These plants produce oxygen, acts as bio filters of the water, whereby they reduce the effect of pollutants. However, there also have negative impacts. Aquatic plants grow in lakes and fresh water and have positive or negative implication to the water bodies and to the waterfowls (Hansson et al., 2010). These plants are rooted in shallow waters with the ascend part emerging above the water surface. Water lily, for example, is found in fresh water ecosystem and is submerged in water (Sun et al., 2014). The impact of some macrophytes has been highlighted by many authors (Widyastuti & Haryono, 2016; Zhang et al., 2014). However, water lily can be an alternative protein source for Lesser Whistling Duck (Welsh et al., 2013; Russell et al., 2014).

Habitat characteristics play a vital role in determining habitat utilization for waterfowl, which include the availability of adequate food and shelter, and water body characteristics (Rajpar et al., 2017). Furthermore, it is known that human interference could lead to momentary modification in characteristics and sectionally change temporal distribution of waterfowl (Sun et al., 2014). The results of this study show that Lesser Whistling Duck prefered to utilize the lake area dominated with aquatic vegetation, such as; lotus (*Nelumbo nucifera*), water chestnut (*Eleocharis dulcis*), wooly waterlily (*Philydrum lanuginosum*), marsh sedge (*Sphenarium purpurascens*), common duckweed (Salvinia minima), spike water milfoil (*Myriophyllum spicatum*), and smooth cordgrass (*Spartina alterniflora*), for foraging and loafing. This highlights that aquatic vegetation composition is the key factor that plays a vital role in the habitat selection of the Lesser Whistling Duck, i.e., aquatic vegetation spread assumed a significant role in the circulation of prey resources, which is the major diet of Lesser Whistling Duck in the wetland ecosystem (Zhang et al., 2014; Zhang et al., 2018).

CONCLUSION

Based on the results, it is concluded that diversity and richness of food sources are driven factors that affect on habitat selection and home range of Lesser Whistling Ducks in PIWR.

In addition, it was observed that Lesser Whistling Duck often-select marshes densely occupied with emergent and submerged aquatic vegetation and having shallow ~~in~~ water depth. It may be that, these marshy areas rich in in aquatic vertebrates and invertebrates that are staple diet of Lesser Whistling Ducks.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Wildlife and National Parks (DWNP), Peninsular Malaysia for granting the permission to conduct this research at the Paya Indah Wetland Reserve. This study was funded by Putra Grant IPS of Universiti Putra Malaysia (Project No: GPIPS/2018/9638000).

REFERENCE

Andy, G. (1996). Analysis of globally threatened Anatidae in relation to threats, distribution, migration pattern and habitat use. *Conservation Biology, 10*(5), 1435-1445.

Asmawi, M. Z. (2007, September 4-6). The effects of town planning system in wetland management Peninsular Malaysia. In *International Seminar on Wetlands and Sustainability* (pp. 267-279). Johor Bharu, Malaysia.

BirdLife International. (2016). *Dendrocygna javanica*. The IUCN Red List of Threatened Species 2016. Retrieved September 3, 2020, from http://datazone.birdlife.org/species/factsheet/lesser-whistling-duck-dendrocygna-javanica/details

Buckland, S. T., Marsden, S. J., & Green, R. E. (2008). Estimating bird abundance: making methods work. *Bird Conservation International, 18*(S1), S91-S108.

Gillespie, A. (2007). *Protected areas and international environmental law*. Leiden, Netherland: Koninkilijke, Brill NV.

Hansson, L. A., Nicolle, A., Bronmark, C., Hargeby, A., Lindstrom, A., & Andersson G. (2010). Waterfowl, macrophytes and the clear water state of shallow lakes. *Hydrobiologyia, 646*(1), 101-109.

Howard, L. (2003). *Anatidae, animal diversity web*. Retrieved March 13, 2017, from http://animaldiversity.org/accounts/Anatidae/

Jayathilake, M. B., & Chandrasekara, W. U. (2015). Variation of avifaunal diversity in relation to land-use modifications around a tropical estuary, the Negombo estuary in Sri Lanka. *Journal of Asia-Pacific Biodiversity, 8*(1), 72-82.

Johnsgard, P. A. (2010) *Ducks, geese and swans of the world*. Lincoln, USA: University of Nebraska Press.

Kiew, R., Chung, R. C. K., Saw, L. G., & Soepadmo, E. (2018). *Flora of Peninsular Malaysia series II: Seed plants* (Volume 7). Kuala Lumpur, Malaysia: Forest Research Institute Malaysia (FRIM).

Koli, V. K. (2014). Diversity and status of avifauna in Todgarh-Raoli Wildlife Sanctuary, Rajasthan, India. *Journal of Asia-Pacific Biodiversity, 7*(4), 401-407.

Ma, Z., Wang, Y., Gan, X., Li, B., Cai, Y., & Chen, J. (2009). Waterbird population changes in the wetlands at Chongming Dongtan in the Yangtze River estuary, China. *Environmental Management, 43*(6), 1187-1200.

Martins, C. O., Rajpar, M. N., Nurhidayu, S., & Zakaria, M. (2017). Habitat selection of *Dendrocygna javanica* in heterogeneous lakes of Malaysia. *Journal of Biodiversity Management and Forestry, 6*(3), 1-6.

Martins, C. O., Zakaria, M., Mohd, H., Nurhidayu, S., & Olaniyi, O. (2019). Population and occupancy estimates of avian species in Payaindah and Putrajaya Wetlands, Peninsular Malaysia. *Journal of Biodiversity Management and Forestry, 8*(2), 1-9.

McMahon, G., & Moreira, S. (2014). *The contribution of the mining sector to socioeeconomics and human development*. Washington, DC: World Bank. Retrieved May 28, 2019, from http://documents.world-bank.org/curated/en/713161468184136844/The-contribution-of-the-mining-sector-to-socioeconomic-and-human-development

Rajpar, M. N., & Zakaria, M. (2012). Avian community parameters of freshwater wetland ecosystem in Peninsular Malaysia. *Asia Life Sciences, 21*(2), 1-19.

Rajpar, M. N., & Zakaria, M. (2014). Effects of habitat characteristics on waterbird distribution and richness in wetland ecosystem of Malaysia. *Journal of Wildlife and Parks, 28*, 105-120.

Rajpar, M. N., Zakaria, M., Ozdemir, I., Ozturk, M., & Gucel, S. (2017). Avian assemblages at Paya Indah natural wetland reserve, Malaysia. *Expert Opinion on Environmental Biology, 6*(3), 1-10.

Rehnus, M., Wehrle, M., & Palme, R. (2014). Mountain hares Lepus timidus and tourism: stress events and reactions. *Journal of Applied Ecology, 51*(1), 6-12.

Russell, I. A., Randall, R. M., & Hanekom, N. (2014). Spatial and temporal patterns of waterbird assemblages in the Wilderness Lakes Complex, South Africa*. WATERBIRDS: The International Journal of Waterbird Biology, 37*(1), 1-18.

Sebastián-González, E., Camp, R. J., Tanimoto, A. M., de Oliveira, P. M., Lima, B. B., Marques, T. A., & Hart, P. J. (2018). Density estimation of sound-producing terrestrial animals using single automatic acoustic recorders and distance sampling. *Avian Conservation and Ecology, 13*(2), 1-16.

Shuford, W. D., & Gardali, T. (2008). *California bird species of special concern: A ranked assessment of species, subspecies, and distinct populations of birds of immediate conservation concern in California*. Camarillo, California: Western Field Ornithologists.

Strasser, E. H., & Heath, J. A. (2013). Reproductive failure of a human-tolerant species, the American kestrel, is associated with stress and human disturbance. *Journal of Applied Ecology, 50*(4), 912-919.

Sulaiman, N. H., Khalit, S. I., Sharip, Z., Samsudin, M. S., & Azid, A. (2018). Seasonal variations of water quality and heavy metals in two ex-mining lake using chemometric assessment approach. *Malaysian Journal of Fundamental and Applied Sciences, 14*(1), 67-72.

Sun, B., Liu, H., Zhou, S., & Li, W. (2014). Evaluating the performance of polynomial regression method with different parameters during color characterizations. *Mathematical Problems in Engineering*, *2014*, 1-7.

Tellkamp, M. P. (2004). Ducks, geese, and swans (Anatidae). In B. Grzimek, D. A. Thoney, N. Schlager, J. E. Trumpey & M. Hutchins (Eds.), *Grzimek's animal life encyclopedia.* Detroit, USA: Thomson-Gale.

Thiel, D., Ménoni, E., BRENOT, J. F., & Jenni, L. (2007). Effects of recreation and hunting on flushing distance of capercaillie. *The Journal of Wildlife Management, 71*(6), 1784-1792.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., ... & Burnham, K. P. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology, 47*(1), 5-14.

Welsh, A. H., Lindenmayer, D. B., & Donnelly, C. F. (2013). Correction: Fitting and interpreting occupancy models. *Plos One, 8*(1), 1-21.

Widyastuti, M., & Haryono, E. (2016). Water quality characteristics of Jonge Telaga (Doline Pond) as water resources for the people of Semanu district Gunung Kidul regency. *Indonesian Journal of Geography, 48*(2), 157-167.

Zakaria, M., & Rajpar, M. N. (2015). Assessing the fauna diversity of Marudu Bay mangrove forest, Sabah, Malaysia, for future conservation. *Diversity, 7*(2), 137-148.

Zakaria, M., & Rajpar, M. N. (2014). Assessing the habitat suitability of two different artificial wetland habitats using avian community structures. *American Journal of Applied Sciences, 11*(8), 1321-1331.

Zhang, Y., Fox, A. D., Cao, L., Jia, Q., Lu, C., Prins, H. H., & de Boer, W. F. (2019). Effects of ecological and anthropogenic factors on waterbird abundance at a Ramsar Site in the Yangtze River Floodplain. *Ambio, 48*(3), 293-303.

Zhang, Y., Qi, W., Zhou, C., Ding, M., Liu, L., Gao, J., ... & Zheng, D. (2014). Spatial and temporal variability in the net primary production of alpine grassland on the Tibetan Plateau since 1982. *Journal of Geographical Sciences, 24*(2), 269-287.

Pertanika J. Sci. & Technol. 28 (4): 1345 - 1365 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Relating the Land-Use Changes to the Invasion of *Pneumatopteris afra* in Nigeria Using Remote Sensing

Gbenga Festus Akomolafe[1,2]* and Zakaria Rahmad[1,3]

[1]School of Biological Sciences, University Sains Malaysia, 11800 USM, Gelugor, Pulau Pinang, Malaysia
[2]Department of Botany, Federal University of Lafia, PMB 146, Lafia, Nigeria
[3]Center for Global Sustainability Studies (CGSS), level 5, Hamzah Sendut Library 1, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

ABSTRACT

The study aimed at using satellite remote sensing in identifying the land-use changes that have occurred in Lafia, Nigeria within the past 35 years, especially in relation to the current and the predicted future invasion of a fern (*Pneumatopteris afra*). Landsat satellite images OLI/TIRS, ETM+ and TM within the interval of 15 years from 1985 to 2020 were used for the extraction of land-use. Six broad classification systems were used to classify the land-use changes by employing a supervised classification technique. In 1985, the bare land dominated the land-use having an area of 69156 ha while the wetland was the least having an area of 3412 ha. However, in 2020, the built-up area has dominated the land-use of Lafia with an area of 144645 ha (52.21%) while the wetland still remained the least with area of 1477 ha. This is obviously due to the geometric increase in the urbanization of this city. There was a consistent loss of the forests from 1985 to 2020 with an annual rate of loss of 0.46%. This resulted in a loss of 44329 ha of forests in 2020 out of the 47643 ha in 1985. This approximately leads to a total loss of 172,732,045 USD of forest products. The current invasion of *Pneumatopteris afra* in Lafia was found to fall within the shrub and grasses land-use class. This indicates that the landsat satellite could not detect the wetlands where the plant dominated due to its massive covering. This study calls for immediate conservation of the remaining forests and wetlands in Lafia to prevent further encroachments and invasion by plants.

Keywords: Forests, GIS, invasion, Lafia, *Pneumatopteris afra,* remote sensing

ARTICLE INFO

Article history:
Received: 6 May 2020
Accepted: 27 July 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.12

E-mail addresses:
gfakomolafe@yahoo.com (Gbenga Festus Akomolafe)
rahmadz@gmail.com (Zakaria Rahmad)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Land-use change may be regarded as the changes that have occurred in the environment over a long period of time which can be measured by comparing the past and current land-use or vegetation data (Kapfer et al., 2017; Vellend et al., 2013). One of the main consequences of land-use change is the habitat fragmentation and loss, reduction in species diversity, changes in species composition and vegetation structure (Rodríguez-Echeverry et al., 2018; Savilaakso et al., 2014). These are mostly caused by the alteration in processes of community assemblages, for instance habitat filtering and limitation of dispersal (Bergeron et al., 2019). Land-use change may also result in changes in ecosystem functions (Laliberté & Tylianakis, 2012) and making the area susceptible to invasion by either plants or animals (Rembold et al., 2017).

Land-use types coupled with other factors such as soil functioning, ecosystem disturbances, community structure, species composition and diversity are known to determine the vulnerability of an ecosystem to biological invasions (Dimitrakopoulos et al., 2017; Menzel et al., 2017; Rembold et al., 2017; Schrama & Bardgett, 2016). This is regarded as the invasibility of such ecosystems (Catford et al., 2012; Milbau et al., 2009). One other thing that contributes to the high diversity of invasive species in an area is the broad environmental conditions of such area (Wilson et al., 2020). The climatic conditions most importantly influence the establishments of invasive plant species at larger scales while the local land-use activities mostly enhance the spread of invasive plants at habitat scales (Terzano et al., 2018).

In addition, several researches have revealed that the past land-use systems do influence the present distribution of species due to the alteration in the soil physic-chemical properties and disturbance regimes of the ecosystems (Flinn & Vellend, 2005; Johnson et al., 2015). Knowledge of land-use pattern of an area is important for the understanding of the vegetation cover changes of such area over a long period of time (Brůna, 2018). In actual form, land-use type of an area is a function of the degree of intensity and types of human disturbances in the ecosystems (Bart et al., 2015; Chytrý et al., 2012; Clotet et al., 2016; Csecserits et al., 2016). Therefore, there is a direct relationship between the land-use type and the environmental factors (mostly anthropogenic) of the ecosystem (Zhou et al., 2019). In addition, areas with high rate of disturbances in the land-use, for example urban areas are found to be more susceptible to the survival and spread of invasive plants (Portgieter et al., 2020). The construction of roads and railways in some cities in South Africa, which is a major disturbance, has been reported to have served as gateway for the rapid spread of many invasive plants (Faulkner et al., 2020).

Before the advent of remotely sensed data, researchers have utilized old maps such as topographic maps, military survey maps and economic maps for their studies on land-use changes (Fuchs et al., 2015; Godet & Thomas, 2013). However, the recent advances

in remote sensing and aerial photography have made it easier to map and understand the patterns of land cover changes of an area (Wachiye et al., 2013). The remote sensing approach has been widely used in urban planning, hydrological studies, prediction of drought and erosion and mapping of forest cover (Adamu, 2019; Sajjad et al., 2015). The advantages of using remote sensing techniques include data consistency, wide coverage, maximum data precision and accuracy (Adamu, 2019). Remote sensing can classify land-use features of an area based on their distinguish characteristics which can then be used for making specific land-use and land cover maps of the areas (Homer et al., 2004).

The most widely used satellite data for monitoring and classifying land-use in many countries is the Landsat TM image which is a medium-resolution data (Potapov et al. 2012; Zhuravleva et al. 2013). Several methods were already developed by researchers in classifying land-use of areas using the specific satellite image data (Saadat et al., 2011; Sivanpillai et al., 2007; Wardlow et al., 2007). Methods such as the supervised classification, unsupervised classification, image segmentation and the normalized differential vegetation index (NDVI) have been widely used for classifying land-use and land cover in remote sensing (Saadat et al., 2011). Other remote sensing methods of classifying images such as the non-parametric or knowledge-based and sub-pixel classification methods have also been used in many studies mostly in arid and semi-arid regions (Dawelbait & Morari, 2012; Salih et al., 2017). The only limitation to these methods is their unsuitability for limited resources and the specificity for spectrally distinctive components (Salih, 2018).

Therefore, combining GIS and remote sensing approaches can provide unswerving information on the land-use change of an area (Akingbogun et al., 2012). This study will reveal long-term changes in the vegetation and land-use pattern of Lafia, Nigeria which could not have been adequately covered by field work. It also aimed at using remote sensing and GIS approach to relate the current land-use and land cover type with the colonization of *Pneumatopteris afra* (Christ.) Holttum on several wetlands in Lafia, Nigeria. This plant has been reported as a colonizer of wetlands in many parts of Nigeria (Akomolafe & Rahmad, 2018).

MATERIALS AND METHOD

Study Area

Lafia is the capital city of Nasarawa State of Nigeria. It has a geographical extent of latitude 08^{o} 33$^{'}$ N and longitude 08^{o} 32$^{'}$ E (Figure 1). This is categorized to be within the north-central geopolitical zone of the country. Ecologically, it is also known to have the southern guinea savanna vegetation having an annual precipitation range of 1000 to 1500 mm and mean annual temperature range of 24^{o}C to 33^{o}C. This type of vegetation comprises mainly few trees, abundant woody shrubs and grasses. The soil of Lafia is predominantly sandy loam. Lafia, Nigeria is known to have two main seasons which are the wet and dry

seasons. Wet season occurs between May to September while the dry season falls between October and April. The major occupations of the indigenes of this area include fishing, mining and farming. The most widely cultivated crops include the maize, rice, cowpea, guinea corn, sesame and sugar cane. The use of wetlands for irrigation farming during the dry season is also very paramount in this study area. The methodological flowchart of the study is shown in Figure 2.

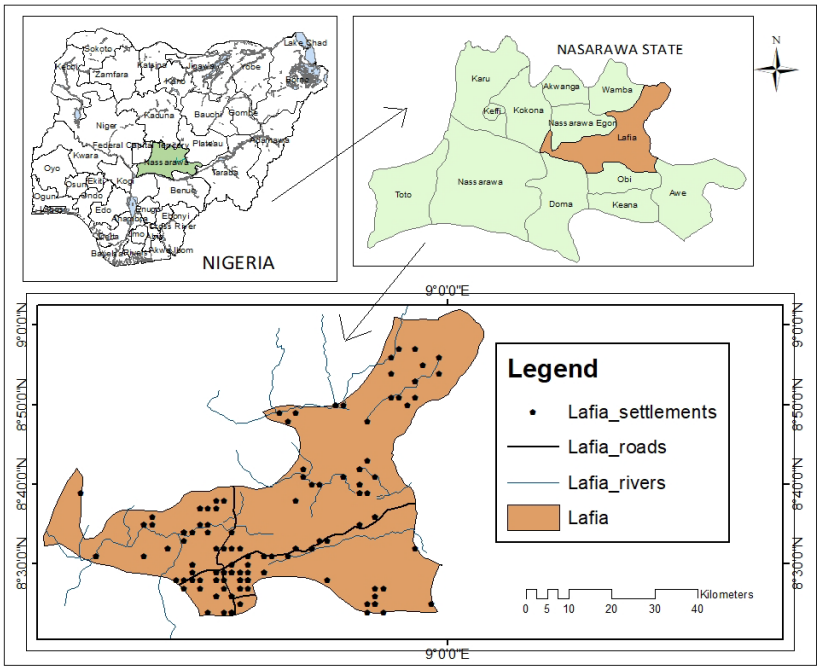


Figure 1. Study area map of Lafia, Nasarawa State, Nigeria (Source: author).

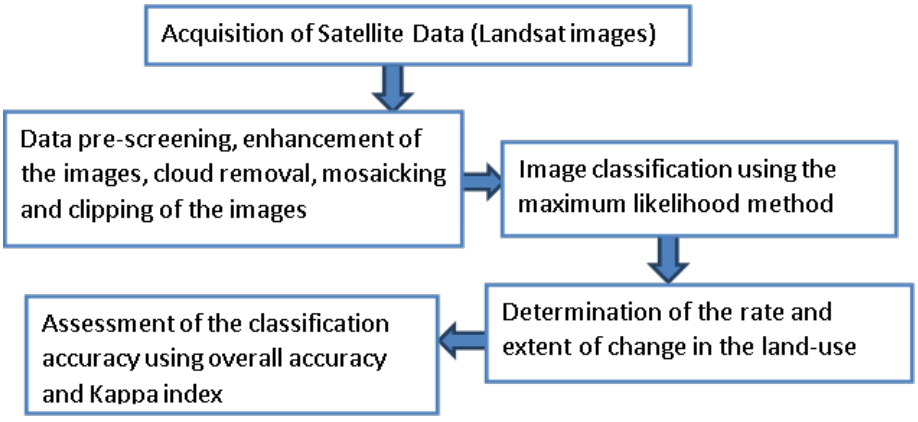


Figure 2. Methodological flowchart of the study

Spatial Data Acquisition

In order for the assessment of the changes in the land-use over the period of 1985 to 2020, Landsat satellite images were downloaded from the website of United State Geological Survey (https://earthexplorer.usgs.gov). The Landsat images include the Landsat 8 OLI / TIRS (operational land imager / thermal infrared sensor), Landsat 7 ETM+ (enhanced thematic mapper) and Landsat 4-5 TM (thematic mapper). These Landsat images were downloaded at Landsat level 1 dataset (Table 1). In order to reduce the cloud cover which is a major challenge of remote sensing of tropical countries (Hansen et al., 2008; Kim, 2016; Margono et al., 2012), additional criteria of land and scene cloud covers less than 10 were selected before downloading the Landsat images.

Table 1

The specifications of the satellite images used

Satellite	Path/Row	Sensor	Number of bands	Period covered	Date captured by sensor	Spatial resolution
Landsat 8	188/54	OLI/TIRS	11	2013 till date	31st January, 2020	30 m
Landsat 7	188/54	ETM+	8	2000 – 2012	2nd March, 2005	30 m
Landsat 5	188/54	TM	7	1984-1999	3rd March 1985	30 m

Georeferencing Techniques

These acquired Landsat raster files were loaded into the ArcMap 10.2.1 software for further georeferencing and analysis. The Landsat images were subjected to several techniques including the data pre-screening, enhancement of the images by cloud removal, correction of radiometric errors, mosaic and clipping of the images and image classification. All the Landsat 7 images acquired since 30th May 2003 till present were reported to have data gaps due to the scan line corrector (SLC) failure. In this study, the scan line stripping errors in the Landsat 7 bands were corrected using the QGIS 2.14.7 software package. The bands were loaded into the software and the "fill nodata" option was chosen from the "raster" tab. After this, the respective bands were chosen as input bands. The gap masks of the respective band which were initially downloaded with the Landsat images were then chosen as the "validity mask". By doing all these, the stripping effects were corrected. The image enhancement was done in order to ensure easier detection and classification of the land-use types (Jande et al., 2019). The boundary shape file of Lafia was used to clip out the area of study from the larger mosaic area.

Image Classification

The satellite image classification in this study involved both unsupervised random selection of sample training points and supervised classifications of the land-use types. The combination of these methods of image classifications has been advocated as more advantageous in ensuring a high level of accuracy (Saadat et al., 2011). For the Landsat 8 satellite image, bands 5, 4 and 3 were used for the analysis while for Landsat 5 and Landsat 7 images, bands 4, 3 and 2 were used (Table 2). The extraction by mask method in the spatial analyst tool was used to extract the Lafia from the selected bands. Each band served as the input while Lafia boundary map served as the feature mask. This extraction process was repeated for the remaining Landsat satellite images. The supervised image classification which had been widely used in remote sensing was employed in this study to classify the Landsat bands (Adamu, 2019; ILWIS, 2001; Jande et al., 2019). This method of image classification works on the principle of identifying training points of known targets and then using these to categorize other unknown sites with similar spectral signatures (Mather & Koch, 2011). In the data management tool, the raster processing was selected, and the three monochromatic bands were joined together using the composite band tool to form the false colour composite (FCC). The FCC was classified through selection of training sites by drawing polygons on the respective land cover type (built up, shrubs and grasses, rocks, forests, wetlands and bare surfaces). These land-use and land cover categories were modified from IPOC Good Practice Guidance (Change, 2003). Not less than 40 sample points were randomly selected for each land cover type. The prior knowledge of the study area was helpful in the selection of training sites as symbolized by different colours (Sinha et al., 2015). The signature file was hereafter created.

For the image transformation, we used the multivariate maximum likelihood classification (MLC) method. Besides MLC, there are other methods of image classification used by researchers. They include the extraction and classification of homogenous objects

Table 2

Characteristics of the satellite bands used for the classification

Spectral characteristics	Landsat 5 and Landsat 7			Landsat 8		
Types of bands used	4	3	2	5	4	3
Colour of bands	Near Infrared (NIR)	Red	Green	Near Infrared (NIR)	Red	Green
Wavelength range (μm)	0.772 – 0.898	0.631 – 0.692	0.519 – 0.601	0.851 – 0.879	0.636 – 0.673	0.533 – 0.590

(ECHO) classifier, fuzzy set classifier, neural networks (NN) classifier, sub-pixel classifier, per-field classifier, minimum distance classifier (MDC) (Lu et al., 2004), decision trees (DTs), support vector machines (SVMs) (Otukei & Blaschke, 2010) and so-on. The analyst choice and the efficiency of any of these classifiers are dependent on several factors such as the band selection, knowledge of the study area, complexity of landscape, accessibility of remote sensing data, the proficiency of the analyst on the classifier used and the classification algorithm (Otukei & Blaschke, 2010). In this study, MLC was preferred to other methods because it has been reported suitable for classification of land-use and land cover with high accuracy in northern Nigeria by recent studies (Adamu, 2019; Jande et al., 2019). MLC enabled the utilization of the prior knowledge of the area where we had already collected the ground-truth data of the respective land-use types (Kim, 2016). In addition, MLC was chosen over others because it was mostly available in many popular image processing and GIS software packages. The maximum likelihood algorithm has the principle of allocating pixels to the class of highest probability and then using it to ascertain the class ownership of that pixel. It is a parametric classifier which operates on the basis that the data follows a normal or near-normal distribution and that the featured classes have equal probability (Otukei & Blaschke, 2010). It has also been reported to perform better than the other parametric classifiers (Richards & Richards, 1999). Accuracy of the performance of MLC was ensured by selecting large number of training samples based on the knowledge of the area of interest.

Description of the Land-use Categories

The built-up is the portion of the land that has been fully developed into roads, railways, houses, industries, and other developed areas. It may also be referred to as settlement in other literature (Kim, 2016). Shrubs and Grasses is the portion of the land dominated by short woody plants called shrubs and other herbaceous plants including grasses (more than 90% cover). The shrubs are short trees usually less than 5m in height. They are generally products of forests degradation. Lafia is expected to have larger portion of land with shrubs and grasses because it falls within the southern guinea savanna zone of Nigeria. Rocks is the part of the land dominated by solid mineral material projecting out of the earth surface (rocks). Bare land is also regarded as bare soil and it falls within the areas of the land that are either made open by natural or anthropogenic activities. Some of these bare lands are lands cleared for agricultural activities particularly farming of rice, sugar cane, sorghum, millet and sesame in the study area. Forests comprises both primary and secondary forests. They are areas of the land occupied by trees. Forest has been defined as an area of land with more than 0.5 ha vegetation comprising trees of 5 m above in height with canopy greater than 10% (FAO, 2010). The forests in Lafia are generally secondary forests which have already experienced disturbances in the past. Some of them are also products of agroforestry.

Wetlands are land areas covered by water seasonally or permanently. They include rivers, streams, lakes or reservoirs which may either be man-made or natural (Kim, 2016).

Determination of Change in the Land-use

The determination of the rate and extent of change in the land-use and land cover of Lafia, Nigeria within the studied periods was done using the following Equation 1, 2 and 3 (Yesserie, 2009):

$$\text{Changed area } (C_a) = T_a(2^{nd} \text{ year}) - T_a(1^{st} \text{ year}) \quad [1]$$

$$\text{Changed extent } (C_e) = C_a / T_a(1^{st} \text{ year}) \quad [2]$$

$$\text{Percentage of change} = C_e \times 100 \quad [3]$$

Where T_a means total area

Classification Accuracy Assessment

The accuracy of the classification was assessed following the ideal method of taking ground truth data of the land-use and land cover of Lafia with the aid of a Garmin Etrex 10 device. These ground truth data (GPS coordinates) were then compared with the already classified land-use and land cover map (Jande et al., 2019). The ground truth assessment was done to also be familiar with the land-use features on the satellite. Areas that could not be accessible by field work were confirmed with the use of google earth images. An error matrix which generally explains the accuracy of the classified remotely sensed data by linking it with the ground truth data was produced. This error matrix utilizes the producer's accuracy, user's accuracy, overall accuracy and Kappa index (Jande et al., 2019). The producer's accuracy also known as omission error is the probability of the reference pixel correctly classified. The user's accuracy also known as commission error is the probability that the classified pixel on the map is exactly that on the ground. The Kappa coefficient was then calculated from this error matrix to give the entire statistical accuracy of the error matrix (Foody, 2004). The Kappa coefficient determines the agreement between the classified images and the ground truth data. The Kappa values ranges from -1 to +1 in increasing order of agreement (Borana & Yadav, 2017). The overall accuracy of the classification of the land-use was calculated by dividing the total number of pixels correctly classified by the total number of sampled ground data (Kim, 2016).

Relating the Invasion of *Pneumatopteris afra* to the Land-use of Lafia

This was done in order to relate the land-use classification map of Lafia, Nigeria to the incidence of the invasion of *Pneumatopteris afra* earlier reported from our previous study (Akomolafe et al., 2019). Ninety-five occurrence points of invasion of *P. afra* were identified through ground validation in the previous study. These points were obtained at 20m intervals using 200 m transect laid at the three sites already invaded by *P. afra* in Lafia,

Nigeria. This was done by superimposing the georeferenced points of present occurrence of *P. afra* on the land-use classification map of Lafia, Nigeria. By so doing, the portion of the land cover occupied by the invasion of this plant was identified. In addition, to identify the land-use classes that are affected by the predicted species distribution model of *P. afra,* the species distribution map that was generated from our previous study was laid side by side with the land-use map of Lafia, Nigeria. This species distribution model of *P. afra* was done using the Maxent algorithm to predict areas of future invasion of *P. afra* in Lafia, Nigeria (Akomolafe et al., 2019).

RESULTS AND DISCUSSION

The land-use and land cover of Lafia, Nigeria from 1985 to 2020 are presented in Table 3. These results were products of the land-use and land cover classification of Landsat images used. In 1985, bare lands (comprising unused land and farmlands) and rocks dominated the land cover of Lafia with areas of 69,156 ha (24.96% of the whole land cover area) and 58,573 ha (21.14%) respectively. The wetland was the least of all having an area of 3,412 ha (1.23% of the whole land cover area). The classification map of the land-use of Lafia in 1985 is shown in Figure 3. In 2005, the built up dominated the land cover with an area of 122,196 ha (44.11% of the total area) while the wetland still remained the least having an area of 2293 ha (0.83% of the total area). Also in 2020, the built up areas (comprising the roads and settlements) became the dominant land cover with an area of 144,645 ha (52.21% of the total land area) while the wetland still remained the least land cover with an area of 1477 ha (0.53% of the total land area). In 2020, the built-up areas (urbanized) have spread almost uniformly across the entire land in Lafia. This is a consistent increase in built up land-use class from 1985 to 2020. The land-use classification maps of Lafia in 2005 and 2020 are presented in Figures 4 and 5 respectively.

Table 3

Area of land use and land cover of Lafia, Nigeria (1985-2020)

Land-use class	1985		2005		2020	
	Area (ha)	Area (%)	Area (ha)	Area (%)	Area (ha)	Area (%)
Built up	48764	17.60	122196	44.11	144645	52.21
Shrubs and Grasses	49477	17.86	38629	13.94	54997	19.85
Rocks	58573	21.14	95666	34.53	28091	10.14
Bare land	69156	24.96	13417	4.84	44501	16.06
Forests	47643	17.19	4824	1.74	3314	1.19
Wetlands	3412	1.23	2293	0.83	1477	0.53
Total Area	277025	100	277025	100	277025	100

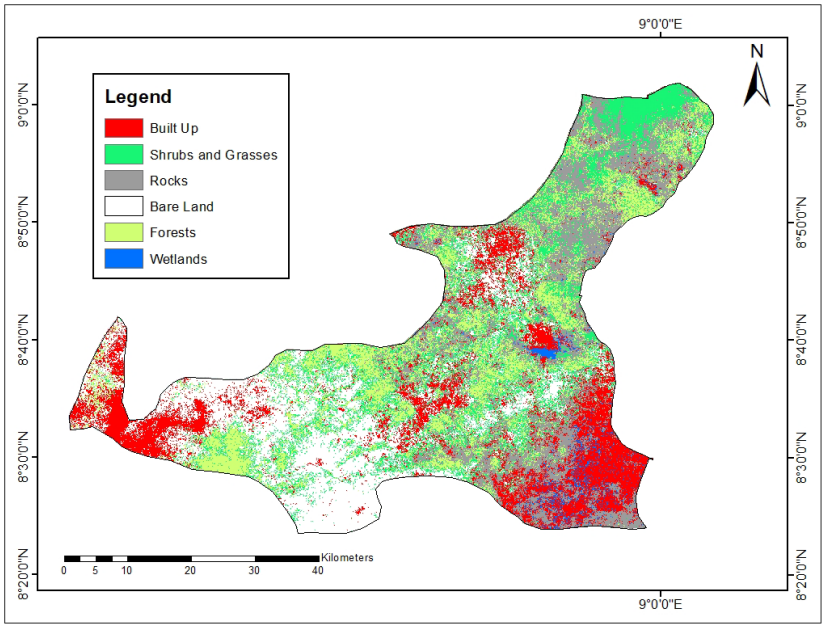


Figure 3. Land use and land cover map of Lafia (1985)

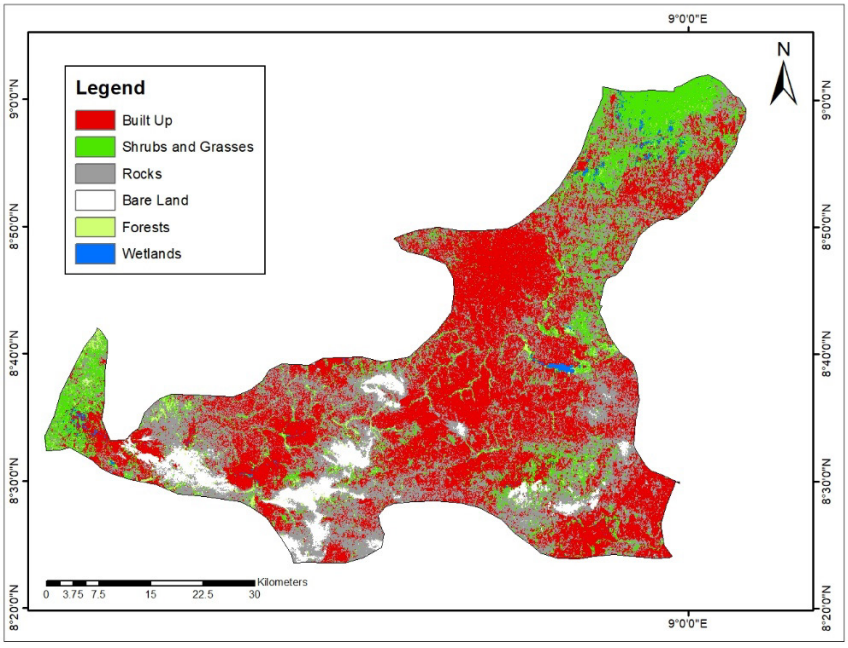


Figure 4. Land use and land cover map of Lafia (2005)

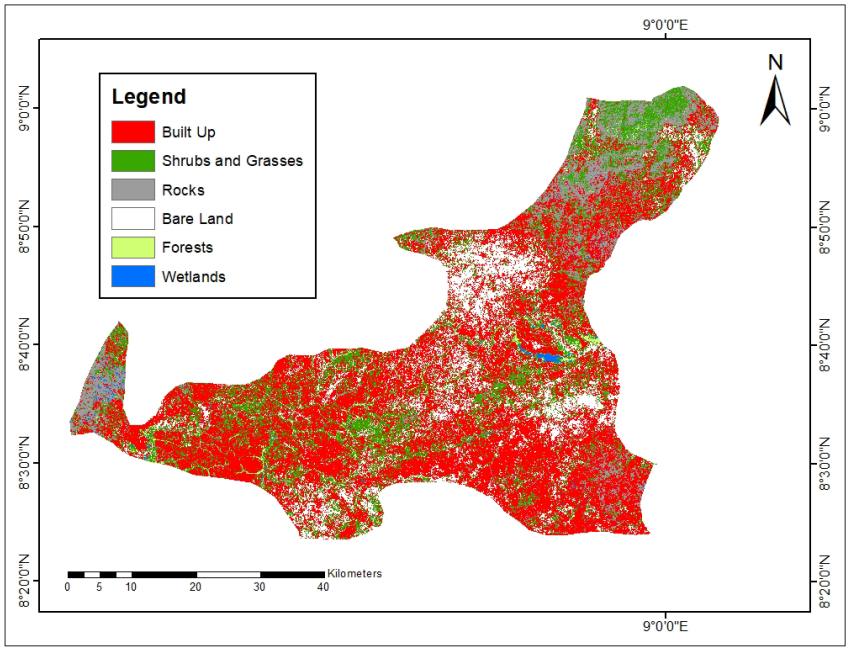


Figure 5. Land use and land cover map of Lafia (2020)

The analysis of the percentage change at interval of not less than 15 years and the annual rate of change in the land-use and land cover of Lafia, Nigeria from 1985 to 2020 are presented in Table 4 and Figure 6. In this result, there is a persistent decrease of forest land from 1985 to 2020 with an annual decrease rate of -4.49 (1985 – 2005), -2.09% (2005 – 2020) and -0.46% (1985 – 2020). This means that 42,819 ha of forests were lost to other land-use types between 1985 and 2005 while 1510 ha were lost between 2005 and 2020. The annual rate of loss of forest is lesser between 2005 and 2020 (2.09%) as compared to between 1985 and 2005 (4.49%). The loss of forest land in any land-use classifications has been attributed to several anthropogenic activities such as logging, clearing of land for farming, urbanization, and inefficiency of government agencies involved in the protection of forests (Curran et al., 2004; Kim, 2016). This could be the case in Lafia, Nigeria whereby the built-up and shrubs and grasses land-use gained from the other land-use classes and increased from 1985 to 2020. This loss in forest land from 1985 to 2020 can be equated to a total loss of 3.38 metric tons of CO_2 (Saka-rasaq, 2019). This resulted in a total loss of 172,732,045 USD of forest products in Lafia, Nigeria within 35 years. Lafia which is found within the guinea savanna vegetation zone of Nigeria was expected to have been dominated more by woody shrubs and grasses. The reverse was the case in this study whereby the shrubs and grasses were almost approximately having the same percentage

area with the forests in 1985 (17.86% and 17.19% respectively). However, in 2020, the shrubs and grasses were observed to have expanded in area of coverage than the forests, thereby justifying its classification as a guinea savanna zone of Nigeria. This result agrees with a similar analysis of land-use of Gboko town (a neighbouring State within the same guinea savanna) whereby the grassland was the dominant land cover increasing from 35.97% in 1987 to 67.54% in 2017 (Jande et al., 2019). The consistent loss of forests and increase in bare lands in the land-use and land cover maps of some parts of Pakistan was also reported (Qamer et al., 2012; Sajjad et al., 2015). It has been established that increase in human population is a major driver of forest loss in an area (Adamu, 2019). This same trend of persistent increased in urbanization and reduction of forest lands was also reported in Accra, Ghana (Addae & Oppelt, 2019) and Adamawa, Nigeria (Adamu, 2019).

Table 4

The annual rate of change in the land use and land cover of Lafia, Nigeria (1985-2020)

Land-use class	1985 - 2005		2005 - 2020	
	Change in area (ha)	% change	Change in area (ha)	% change
Built up	73432	150.59	22449	18.37
Shrubs and Grasses	-10848	-21.93	16368	42.37
Rocks	37093	63.32	-67575	-70.63
Bare land	-55739	-80.59	31084	231.68
Forests	-42819	-89.88	-1510	-31.30
Wetlands	-1119	-32.79	-816	-35.59

Land-use class	1985 - 2020		Annual rate of change (%)		
	Change in area (ha)	% change	1985-2005	2005-2020	1985-2020
Built up	95881	34.61	7.53	1.22	0.99
Shrubs and Grasses	5520	1.99	-1.09	2.83	0.06
Rocks	-30482	-11	3.16	-4.71	-0.31
Bare land	-24655	-8.89	-4.03	15.45	-0.25
Forests	-44329	-16	-4.49	-2.09	-0.46
Wetlands	-1935	-0.69	-1.64	-2.37	-0.02

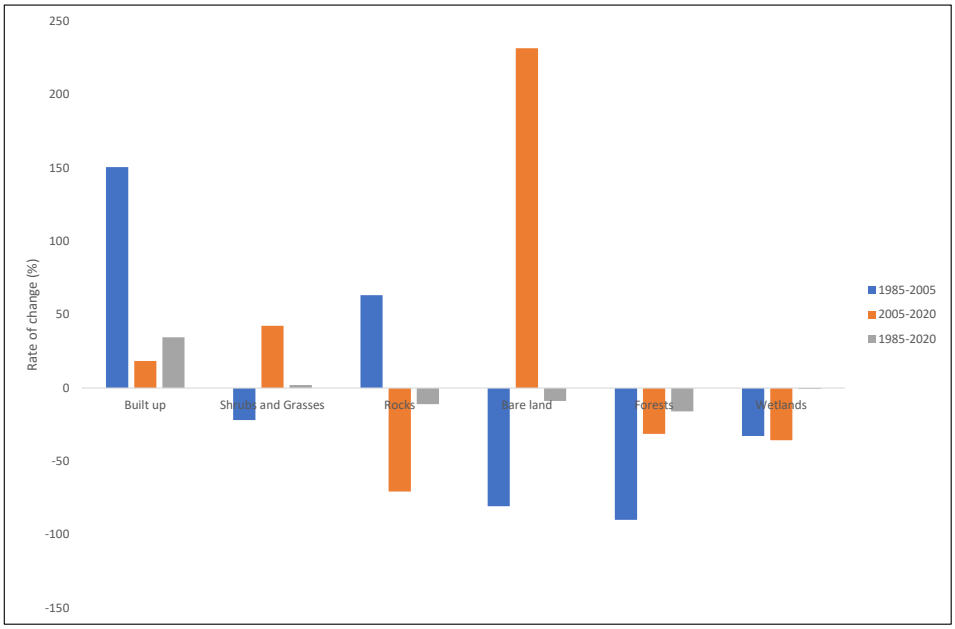


Figure 6. Rate of change in the land-use classes from 1985-2020

The wetlands and bare land also experienced decrease annual rate of change with -0.02% and - 0.25% (1985 – 2020) respectively. It was only the built-up areas that experienced an increase in the annual rate of change from 1985 to 2020. The built-up areas have 7.53% annual increase from 1985 – 2005, 1.22% from 2005 – 2020 and 0.99% from 1985 – 2020. The increased annual rate of change in the built-up areas of Lafia from 1985 to 2020 shows that there has been a steady rate of urbanization and human population of the city. The bulk of this urbanization might have occurred between 1985 and 2005 due to the higher annual rate of change. The shrub and grasses also had increase in the annual rate of change except from 1985 – 2005 where there was a slight loss of 10,843 ha of the area thereby giving rise to an annual rate of change of -1.09% during that period. Also, the decrease in the annual rate of change of the bare lands which are mostly farmlands is an indication that there is a reduction in the farming activities due to the non-involvement of youths in agriculture (Jande et al., 2019). It is very common in recent times that youths are mostly interested in ready-made jobs rather than engaging in farming activities. The elderly ones who have been practicing agriculture are reduced in number due to retirement or death.

The overall accuracy for the land-use classifications of 1985, 2005 and 2020 are 71.36%, 78.93% and 87.47% respectively (Table 5). These indicate that the analysis for the land-use and change detections within the period under consideration were fair enough.

Table 5

Accuracy assessment of land-use classification of Lafia, Nigeria

Land-use class	1985		2005		2020	
	Producer's accuracy (%)	User's accuracy (%)	Producer's accuracy (%)	User's accuracy (%)	Producer's accuracy (%)	User's accuracy (%)
Built up	74.61	78.95	80.31	82.33	89.92	94.53
Shrubs and Grasses	81.23	80.34	81.34	83.45	86.54	87.89
Rocks	60.56	76.81	70.87	80.21	70.54	76.12
Bare land	69.76	80.21	79.01	84.32	86.75	89.90
Forests	78.32	80.43	81.09	83.23	98.54	99.01
Wetlands	60.54	77.33	75.34	81.32	80.21	84.32
Overall Accuracy	**71.36%**		**78.93%**		**87.47%**	
Kappa	**0.71**		**0.83**		**0.89**	

The Kappa coefficients for the period of 1985 to 2020 also ranged between 0.71 – 0.89. This means that the agreement between the classified land-use and the observed ground data in Lafia, Nigeria ranged from substantial agreement to almost perfect agreement (Borana & Yadav, 2017). This is also an indication that the selected land-use classification method in this study is feasible and appropriate for the area under consideration. Other researchers have also used similar land-use classification methods in Northern Nigeria, and they reported higher degrees of classification accuracy (Adamu, 2019; Jande et al., 2019).

With respect to the invasion of the fern, *Pneumatopteris afra* in Lafia, Nigeria, the superimposition of its current invaded georeferenced points on the land-use classification map of Lafia in 2020 revealed that the plant was found within the shrub and grasses land-use class (Figure 7A). The area covered by *P. afra* within the land-use of Lafia is 75 ha. However, in the land-use and land cover maps of 1985 and 2005, the same areas invaded by *P. afra* fall within the bare lands and seasonally flooded wetlands. The landsat image was able to identify the current land-use category where *P. afra* invaded as shrubs and grasses. This could be probably due to the herbaceous growth form of this plant. The massive growth of this plant could have covered those wetlands so that the landsat satellite images could not detect the wetlands but see it rather as shrubs and grasslands.

However, comparing the predicted Maxent species distribution model of future invasion of *P. afra* in Lafia with the year 2020 land-use map revealed that the areas of high probability (0.8 – 1) of invasion of *P. afra* fall within shrubs and grasses, built-up, bare land, and wetlands (Figure 7B). This shows that the areas predicted to be susceptible to the future invasion of this plant in Lafia, Nigeria affect more land-use classes than the current areas of invasion. This also indicates that *P. afra* is a plant that has the potential to invade diverse land-use and land cover types in the future. This is buttressed by Oloyede et al., (2011) who reported that this plant could adapt to different types of habitats in Nigeria. The predicted areas of land affected was approximately 178407 ha (Akomolafe et al., 2019). Several researches have reported that the change in the land-use type of an area has direct influence on the success of invasive plants across the world (Bart et al., 2015; Chytrý et al., 2012; Clotet et al., 2016; Csecserits et al., 2016). This might be the case in this study whereby the land-use type of the areas currently invaded by *P. afra* changed from bare land and seasonally flooded wetlands in 1985 and 2005 respectively to shrub and grasslands in 2020. The changes in the land-use and land cover of Lafia from 1985 to 2020 as a result of environmental disturbances could have served as the gateway for the successful invasion of *P. afra* as confirmed by other studies whereby land-use change due to disturbances were promoters of invasion (Faulkner et al., 2020; Portgieter et al., 2020). If the spread of this plant is not controlled early, it poses serious threats on the large areas already predicted by the Maxent model to be affected which fall within the productive parts of the land-use in Lafia, Nigeria.

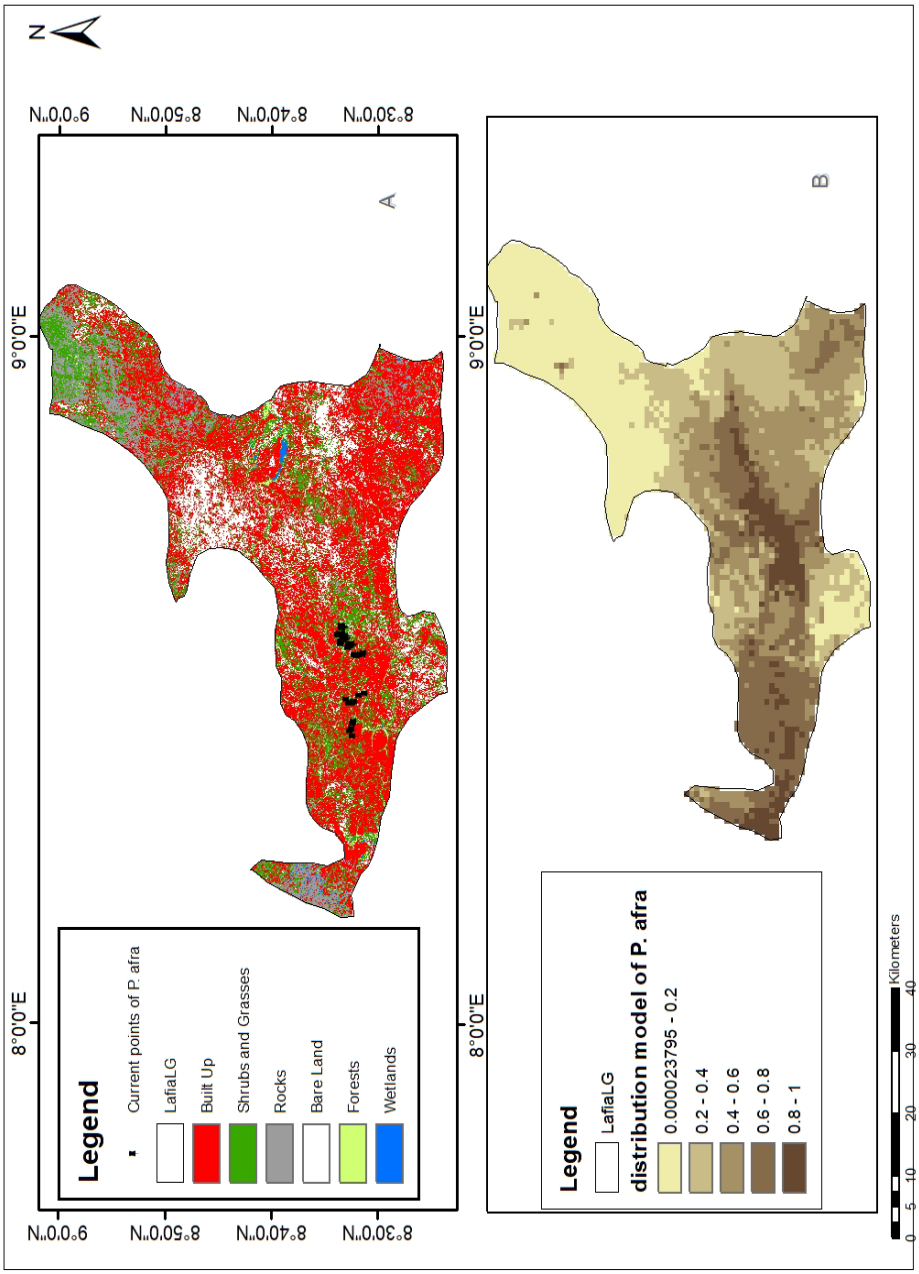


Figure 7A and 7B. Land use and land cover map of Lafia (2020) in relation with the invasion of *P. afra*

CONCLUSION

This study has been able to use the remote sensed data and GIS techniques to simulate the land-use and land cover changes in Lafia from 1985 to 2020 and relates it with the recent invasion of *Pneumatopteris afra* there. This is very critical in decision making by policy makers particularly as it reveals the rapid loss of forest lands and wetlands whereas built up areas have increased geometrically. The rapid increase in the built-up areas could be directly linked with increase in human population and its associated demand for several social amenities. It is so disturbing that the forest lands and wetlands are being depleted continuously for urbanization and invasion of plants. This has serious implication for increase in greenhouse gases in the environment and loss of native biodiversity. It is therefore highly imperative for relevant government agencies to prioritize the conservation of the remaining forests and wetlands in Lafia, Nigeria. The current and the predicted future invasion of *P. afra* which was observed to occupy large areas of the land-use classes in Lafia, Nigeria should be regarded as a major environmental threat which needs immediate attention by the policy makers.

ACKNOWLEDGEMENTS

We acknowledge the Nigerian Government Tertiary Education Trust Fund (TETFund) ASTD PhD Grant (FUL/REG/TETfund/002/VOL.II/182) for financially supporting this research.

REFERENCES

Adamu, S. (2019). Remote sensing and GIS application for forest reserve monitoring and prediction: A case of Girei Forest Reserve, Adamawa State, Nigeria. *Fudma Journal of Sciences, 3*(3), 83-94.

Addae, B., & Oppelt, N. (2019). Land-use/land-cover change analysis and urban growth modelling in the greater accra metropolitan area (GAMA), Ghana. *Urban Science, 3*(1), 1-20.

Akingbogun, A. A., Kosoko, O., & Aborisade, D. (2012, May 4-5). Remote sensing and GIS application for forest reserve degradation prediction and monitoring. In *First FIG Young Surveyors Conference Knowing to Create the Future* (pp. 1-27). Rome, Italy.

Akomolafe, G. F., & Rahmad, Z. B., & Karau, S. D. (2019). Modelling the distribution of a potential invasive tropical fern, Cyclosorus afer in Nigeria. *African Journal of Ecology, 57*(3), 304-313.

Akomolafe, G., & Rahmad, Z. (2018). A review on global ferns invasions: Mechanisms, management and control. *Journal of Research in Forestry, Wildlife and Environment, 10*(3), 42-54.

Bart, D., Davenport, T., & Carpenter, Q. (2015). Stress and land-use legacies alter the relationship between invasive-and native-plant richness. *Journal of Vegetation Science, 26*(1), 80-88.

Bergeron, A., Lavoie, C., Domon, G., & Pellerin, S. (2019). Changes in spatial structures of plant communities lead to functional homogenization in an urban forest park. *Applied Vegetation Science, 22*(2), 256-268.

Borana, S. L., & Yadav, S. K. (2017). Prediction of land cover changes of Jodhpur City using cellular automata markov modelling techniques. *International Journal of Engineering Science and Computing, 7*(11), 15402-15406.

Brůna, J. (2018). *Use of remote sensing data for monitoring of long-term vegetation dynamics on the landscape scale*. Prague, Czechia: Univerzita Karlova.

Catford, J. A., Vesk, P. A., Richardson, D. M., & Pyšek, P. (2012). Quantifying levels of biological invasion: Towards the objective classification of invaded and invasible ecosystems. *Global Change Biology, 18*(1), 44-62.

Change, I. (2003). *Good practice guidance for land use, land-use change and forestry*. Kanagawa, Japan: Institute for Global Environment Strategies.

Chytrý, M., Wild, J., Pyšek, P., Jarošík, V., Dendoncker, N., Reginster, I., … & Kühn, I. (2012). Projecting trends in plant invasions in Europe under different scenarios of future land-use change. *Global Ecology and Biogeography, 21*(1), 75-87.

Clotet, M., Basnou, C., Bagaria, G., & Pino, J. (2016). Contrasting historical and current land-use correlation with diverse components of current alien plant invasions in Mediterranean habitats. *Biological Invasions, 18*(10), 2897-2909.

Csecserits, A., Botta-Dukát, Z., Kröel-Dulay, G., Lhotsky, B., Ónodi, G., Rédei, T., Szitár, K., … & Halassy, M. (2016). Tree plantations are hot-spots of plant invasion in a landscape with heterogeneous land-use. *Agriculture, Ecosystems and Environment, 226*, 88-98.

Curran, L. M., Trigg, S. N., McDonald, A. K., Astiani, D., Hardiono, Y. M., Siregar, P., … & Kasischke, E. (2004). Lowland forest loss in protected areas of Indonesian Borneo. *Science, 303*(5660), 1000-1003.

Dawelbait, M., & Morari, F. (2012). Monitoring desertification in a Savannah region in Sudan using Landsat images and spectral mixture analysis. *Journal of Arid Environments*, *80*, 45-55.

Dimitrakopoulos, P. G., Koukoulas, S., Galanidis, A., Delipetrou, P., Gounaridis, D., Touloumi, K., & Arianoutsou, M. (2017). Factors shaping alien plant species richness spatial patterns across Natura 2000 Special Areas of Conservation of Greece. *Science of the Total Environment, 601*, 461-468.

Faulkner, K. T., Burness, A., Byrne, M. J., Kumschick, S., Peters, K., Robertson, M. P., … & Williams, V. L. (2020). South Africa's pathways of introduction and dispersal and how they have changed over time. In *Biological Invasions in South Africa* (pp. 313-354). Cham, Switzerland: Springer.

Flinn, K. M., & Vellend, M. (2005). Recovery of forest plant communities in post-agricultural landscapes. *Frontiers in Ecology and the Environment, 3*(5), 243-250.

FAO. (2010). *Global forest resources assessment 2010* (Report No. 163). Rome, Italy: Food and Agriculture Organization of United Nations.

Foody, G. M. (2004). Thematic map comparison: Evaluating the statistical significance of difference in classification accuracy. *Photogrammetric Engineering and Remote Sensing, 70*(5), 627-633.

Fuchs, R., Verburg, P. H., Clevers, J. G., & Herold, M. (2015). The potential of old maps and encyclopaedias for reconstructing historic European land cover/use change. *Applied Geography, 59*, 43-55.

Godet, L., & Thomas, A. (2013). Three centuries of land cover changes in the largest French Atlantic wetland provide new insights for wetland conservation. *Applied Geography, 42*, 133-139.

Hansen, M. C., Roy, D. P., Lindquist, E., Adusei, B., Justice, C. O., & Altstatt, A. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment, 112*(5), 2495-2513.

Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 national land-cover database for the United States. *Photogrammetric Engineering and Remote Sensing, 70*(7), 829-840.

ILWIS, I. (2001). *3.0 academic user's guide*. Enschede, The Netherlands: International Institute for Aerospace Survey and Earth Sciences.

Jande, J., Nsofor, G., & Abdulkadir, A. (2019). Assessment of land use and land cover changes and urban expansion using remote sensing and GIS in Gboko, Benue State, Nigeria. *Journal of Research in Forestry, Wildlife and Environment, 11*(3), 201-214.

Johnson, A. L., Tauzer, E. C., & Swan, C. M. (2015). Human legacies differentially organize functional and phylogenetic diversity of urban herbaceous plant communities at multiple spatial scales. *Applied Vegetation Science, 18*(3), 513-527.

Kapfer, J., Hédl, R., Jurasinski, G., Kopecký, M., Schei, F. H., & Grytnes, J. A. (2017). Resurveying historical vegetation data–opportunities and challenges. *Applied Vegetation Science, 20*(2), 164-171.

Kim, C. (2016). Land use classification and land use change analysis using satellite images in Lombok Island, Indonesia. *Forest Science and Technology, 12*(4), 183-191.

Laliberté, E., & Tylianakis, J. M. (2012). Cascading effects of long-term land-use changes on plant traits and ecosystem functioning. *Ecology, 93*(1), 145-155.

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing, 25*(12), 2365-2401.

Margono, B. A., Turubanova, S., Zhuravleva, I., Potapov, P., Tyukavina, A., Baccini, A., … & Hansen, M. C. (2012). Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010. *Environmental Research Letters, 7*(3), 1-16.

Mather, P. M., & Koch, M. (2011). *Computer processing of remotely-sensed images: An introduction*. Chichester, UK: John Wiley & Sons.

Menzel, A., Hempel, S., Klotz, S., Moora, M., Pyšek, P., Rillig, M. C., … & Kühn, I. (2017). Mycorrhizal status helps explain invasion success of alien plant species. *Ecology, 98*(1), 92-102.

Milbau, A., Stout, J. C., Graae, B. J., & Nijs, I. (2009). A hierarchical framework for integrating invasibility experiments incorporating different factors and spatial scales. *Biological Invasions, 11*(4), 941-950.

Oloyede, F., Aponjolosun, B., & Ogunwole, A. (2011). Reproductive potentials of a tropical fern *Cyclosorus afer* (Christ.) Ching (Thelypteridaceae: Pteridopyhte) at Obafemi Awolowo University, Ile-Ife, Nigeria. *Ife Journal of Science*, *13*(1), 143-148.

Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation, 12*, S27-S31.

Potapov, P. V., Turubanova, S. A., Hansen, M. C., Adusei, B., Broich, M., Altstatt, A., … & Justice, C. O. (2012). Quantifying forest cover loss in Democratic Republic of the Congo, 2000-2010, with Landsat ETM+ data. *Remote Sensing of Environment*, *122*, 106-116.

Potgieter, L. J., Douwes, E., Gaertner, M., Measey, J., Paap, T., & Richardson, D. M. (2020). Biological invasions in South Africa's urban ecosystems: Patterns, processes, impacts, and management. In *Biological Invasions in South Africa* (pp. 275-309). Cham, Switzerland: Springer.

Qamer, F. M., Abbas, S., Saleem, R., Shehzad, K., Ali, H., & Gilani, H. (2012). Forest cover change assessment in conflict-affected areas of northwest Pakistan: The case of Swat and Shangla districts. *Journal of Mountain Science, 9*(3), 297-306.

Rembold, K., Mangopo, H., Tjitrosoedirdjo, S. S., & Kreft, H. (2017). Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. *Biological Conservation, 213*, 234-242.

Richards, J. A., & Richards, J. A. (1999). *Remote sensing digital image analysis* (Vol. 3, pp. 10-38). Berlin, Germany: Springer.

Rodríguez-Echeverry, J., Echeverría, C., Oyarzún, C., & Morales, L. (2018). Impact of land-use change on biodiversity and ecosystem services in the Chilean temperate forests. *Landscape Ecology, 33*(3), 439-453.

Saadat, H., Adamowski, J., Bonnell, R., Sharifi, F., Namdar, M., & Ale-Ebrahim, S. (2011). Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(5), 608-619.

Sajjad, A., Hussain, A., Wahab, U., Adnan, S., Ali, S., Ahmad, Z., & Ali, A. (2015). Application of remote sensing and GIS in forest cover change in Tehsil Barawal, District Dir, Pakistan. *American Journal of Plant Sciences, 6*(09), 1-7.

Saka-rasaq, O. (2019). *Forest loss in Nigeria, the impact on climate and people from the perspectives of illegal forest activities and government negligence* (Degree Thesis). University of Applied Sciences, Yrkeshogskolan, Novia.

Salih, A. (2018). Classification and mapping of land cover types and attributes in Al-Ahsaa Oasis, Eastern Region, Saudi Arabia Using Landsat-7 Data. *Journal of Remote Sensing and GIS*, *7*(1), 228-234.

Salih, A. A., Ganawa, E. T., & Elmahl, A. A. (2017). Spectral mixture analysis (SMA) and change vector analysis (CVA) methods for monitoring and mapping land degradation/desertification in arid and semiarid areas (Sudan), using Landsat imagery. *The Egyptian Journal of Remote Sensing and Space Science*, *20*, S21-S29.

Savilaakso, S., Garcia, C., Garcia-Ulloa, J., Ghazoul, J., Groom, M., Guariguata, M.R., … & Zrust, M. (2014). Systematic review of effects on biodiversity from oil palm production. *Environmental Evidence, 3*(1), 1-21.

Schrama, M., & Bardgett, R. D. (2016). Grassland invasibility varies with drought effects on soil functioning. *Journal of Ecology, 104*(5), 1250-1258.

Sinha, S., Sharma, L. K., & Nathawat, M. S. (2015). Improved land-use/land-cover classification of semi-arid deciduous forest landscape using thermal remote sensing. *The Egyptian Journal of Remote Sensing and Space Science, 18*(2), 217-233.

Sivanpillai, R., Srinivasan, R., Smith, C. T., Messina, M. G., & Wu, X. B. (2007). Estimating regional forest cover in east Texas using advanced very high-resolution radiometer (AVHRR) data. *International Journal of Applied Earth Observation and Geoinformation*, *9*(1), 41-49.

Terzano, D., Kotzé, I., Marais, C., Cianciullo, S., Farcomeni, A., Caroli, P., … & Attorre, F. (2018). Environmental and anthropogenic determinants of the spread of alien plant species: insights from South Africa's quaternary catchments. *Plant Ecology*, *219*(3), 277-297.

Vellend, M., Brown, C. D., Kharouba, H. M., McCune, J. L., & Myers-Smith, I. H. (2013). Historical ecology: Using unconventional data sources to test for effects of global environmental change. *American Journal of Botany, 100*(7), 1294-1305.

Wachiye, S. A., Kuria, D. N., & Musiega, D. (2013). GIS based forest cover change and vulnerability analysis: A case study of the Nandi North forest zone. *Journal of Geography and Regional Planning, 6*(5), 159-171.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment*, *108*(3), 290-310.

Wilson, J. R., Foxcroft, L. C., Geerts, S., Hoffman, M. T., MacFadyen, S., Measey, J., … & van Wilgen, B. W. (2020). The role of environmental factors in promoting and limiting biological invasions in South Africa. In *Biological invasions in South Africa* (pp. 355-385). Cham, Switzerland: Springer.

Yesserie, A. G. (2009). *Spatio-temporal land use/land cover changes analysis and monitoring in The Valencia Municipality, Spain* (Master Thesis). Universitat Jaume I, Spain.

Zhou, Y., Su, Y., Zhong, Y., Xie, P., Xu, M., & Su, Z. (2019). Community attributes predict the relationship between habitat invasibility and land use types in an agricultural and forest landscape. *Forests, 10*(10), 1-17.

Zhuravleva, I., Turubanova, S., Potapov, P., Hansen, M., Tyukavina, A., Minnemeyer, S., … & Thies, C. (2013). Satellite-based primary forest degradation assessment in the Democratic Republic of the Congo, 2000-2010. *Environmental Research Letters*, *8*(2), 1-14.

Pertanika J. Sci. & Technol. 28 (4): 1367 - 1387 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Optimization and Modelling of Turbidity Removal of Sewage using High-Gradient Magnetic Separation (HGMS) by Response Surface Methodology (RSM)

Nur Sumaiyyah Supian*, Johan Sohaili and Nur Farhan Zon

Department of Civil Engineering, School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Johor, Malaysia

ABSTRACT

Endless industrial development and growing society occasionally create an enormous volume of wastewater, which leads to some issues on wastewater treatment. Existing conventional screening processes have various limitations and drawbacks. Therefore, this study investigated the use of a combination of non-corrosive stainless steel wool and a permanent magnet to increase magnetic gradient, hence reducing suspended matter in sewage through turbidity test. An approach for optimizing the reduction of suspended matter through turbidity analysis was conducted using central composite design (CCD) under response surface methodology (RSM). Three critical independent variables, such as magnet strength, circulation time, and steel wool, and turbidity removal as the response, were further studied to analyze their interaction effects. As a result, an optimal value of turbidity removal was found at 90.3% under the specified optimum conditions of magnet strength of 245 mT, 116 g of non-corrosive stainless steel wool, and 16 h of circulation time. Statistical analysis had shown that the magnet strength, circulation time, and steel wool significantly affected the turbidity removal performance. Furthermore, design of experiment was significantly verified by a small range of error between predicted and actual data. Consequently, a higher gradient of magnetic separation was proven to effectively remove suspended matter using inexpensive non-corrosive stainless steel wool without using magnetic adsorbent.

ARTICLE INFO

Article history:
Received: 27 May 2020
Accepted: 10 August 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.13

E-mail addresses:
maya_viaxzey89@yahoo.com (Nur Sumaiyyah Supian)
johansohaili@utm.my (Johan Sohaili)
nur_farhan02@yahoo.com (Nur Farhan Zon)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

Thus, the suggested approach was found to be cost-effective and environmentally friendly for sewage treatment.

Keywords: Magnetism, optimization, sewage, steel wool, turbidity

INTRODUCTION

Various types of effluent can be found from residence facilities, manufacturing, or any administration, as well as environmental surroundings, including from vehicle exhaust, oil emission, and road construction. The accumulated waste enters wastewater systems uncontrollably (Oliveira et al., 2007). In fact, the increase in water problems promotes economic constriction, decreases environmental quality, and interferes with human sustainability (Rajasulochana & Preethy, 2016). Therefore, there is an urgent need for a sustainable method that produces an efficient outcome and requires lower operating cost in order to reduce the rapid formation of polluted water. Commonly, the first step in wastewater treatment is the screening of coarse particles from raw sewage through settling (Kwon et al., 2014).

According to Zaidi et al. (2014), numerous conventional methods regarding wastewater treatment have been developed; however, most of them have limitations and drawbacks. For example, filtration requires higher energy consumption and expensive membrane filters, which have a short lifespan and inconvenient to use in a longer period of time (Gupta & Suhas, 2009). Meanwhile, chemical treatment usually uses a combination of flocculation agents in wastewater treatment. However, this treatment needs higher cost of chemicals, which leads to excess sludge formation as a by-product (Lee et al., 2006). In addition, this treatment highly depends on the pH level, which is quite challenging to maintain for treating raw wastewater (Kace & Linford, 1975). Recently, the separation by magnetic force has gained attention due to its capability to remove particles efficiently, cost-efficient technology, and adequate to treat a high volume of wastewater within a short time (Zhou et al., 2009; Wang et al., 2012).

The introduction of high-gradient magnetic separation (HGMS) for wastewater treatment becomes a popular method as the process involves a simple yet effective operation to eliminate unwanted particles from wastewater. During the separation that occurs in a field or a medium, the phenomenon that occurs due to different magnitude of diversification from magnetic flux density is called high gradient (Baik et al., 2010; Baik et al., 2013). In general, HGMS method is an efficient wastewater treatment process due to its capability to reduce various colloidal and suspended matter from the samples by handling weak magnetic particles involved during separation (Hournkumnuard & Chantrapornchai, 2011; Zheng et al., 2015). The method has been used in various treatment processes, such as

pharmaceutical research (Ueda et al., 2009; He et al., 2014), biotechnology application (Setchell, 1985; Hoffmann et al., 2002), water and wastewater (Karapinar, 2003; Tomska & Wolny, 2008; Hwang & Han, 2015), dye removal (Chen et al., 2011; Narian et al., 2015; Muntean et al., 2018), biological enhancement process (Bitton et al., 1975; Sirisha et al., 2017), and reduction of extracellular protein (Franzreb et al., 2006; Käppler et al., 2009).

One of the unique characteristics of HGMS compared to other separation methods is the selectivity mechanism that uses magnetic force to execute particle's capture movement. This is due to the different level of susceptibility between particles and magnetic field in the sample (Takayasu et al., 1984; Blamire et al., 2009; Ambashta & Sillanpää, 2010). Therefore, an additional method to increase the magnetic gradient other than using higher magnet strength is by adding steel wool. As claimed by Dhoble et al. (2011), greater magnetization generated by magnet strength and higher magnetic gradient through the presence of steel wool can increase the maximum removal performance of the sample. In achieving productive separation, the presence of steel wool combined with high magnetic strength is significantly important to achieve a successful HGMS application. Fibrous steel wool has a large surface area, which homogenizes with a uniform magnetic field that promotes a higher magnetic gradient (Zheng et al., 2019). Hence, weak magnetic particles are easily attracted to higher magnetic force and simultaneously trapped in the steel wool (Ebner et al., 1997; Mariani et al., 2010). It is proven by Podoynitsyn et al. (2016) who stated that the efficiency of HGMS approach depended on the level of magnetic susceptibility value on particles.

Another factor that promotes higher removal efficiency is the duration of particles exposed to a magnetic field. According to Busch and Busch (1997), the repetitive action of exposing particles to a magnetic field indicates a visible change in particle size. This phenomenon was observed by Rao and Sahu (2013) who insisted that a longer time span of particles exposed to a magnetic field could lead to the accumulation of ionic charges approaching pollutant charges in a sample. It was also convinced by Oshitani et al. (1999), that the presence of ionic charges during conformational changes of water molecules was significantly essential for removal efficiency. In addition, the magnetic effect is likely influenced by the area that consists of higher charge density (e.g., steel wool) in order to produce a competent result or output. Thus, a combination with the frequency of exposure to a magnetic field is necessary to increase the efficiency of wastewater treatment (Baker et al., 1997).

During previous study, application of response surface methodology (RSM) was widely used in optimizing the unwanted materials removal performance on water and wastewater treatment by adding magnetic particles or magnetic seeding. For example, zeolite magnetic nanocomposite was acknowledged as excellent adsorbent reacted to excessive amount of arsenic that majorly affected to the environment (Gugushe et al.,

2019). RSM analysis was conducted to determine the most optimum efficient variables to achieve higher arsenic removal performance. Captivating by the capability of magnetic adsorbent, Liu et al. (2016) modelled a novel magnetic ion imprinted polymer to optimize an optimum treatment of strontium (II) ions (Sr(II)) on wastewater via RSM. Similarly, a magnetic adsorbent known as clinoptilolite (CPL) used for lead removal on wastewater and RSM analysis was utilized to optimize pH, temperature, lead concentration as well as dosage adsorbent in order to excel the removal performance (Javanbakht & Ghoreishi, 2017). Additionally, Hwang and Han (2015) revealed that by applying 14.1 min of mixing time with 22.1 mgL^{-1} amount of magnetic iron oxide nanoparticles, produced 90.8% of reduction in sludge water content. It is much better than conventional method that takes longer time and depends only by gravity alone.

However, most of previous researches only limited to application of magnetic seeding in order to treat unwanted particles on water and wastewater treatment. Additionally, some of few drawbacks were encountered using magnetic seeding as such complex substances, highly in cost production as well as unable to recover completely if it flowed away by river. Hence, in this study, it would be an interest to expand the significant of high gradient performance only by combination between magnet strength and non-corrosive stainless steel wool. In the present study, the removal of suspended matter was investigated through turbidity test. Therefore, the main focus is given to increase magnetic gradient which is expected to boost turbidity removal by including non-corrosive stainless steel wool as a contribution factor together with a permanent magnet without using any magnetic adsorbent.

The magnet strength, circulation time, and steel wool were selected as the critical independent values and turbidity removal as the response, which were optimized using central composite design (CCD) under response surface methodology (RSM). CCD method was selected due to efficiency on experimental data information based on variable conditions effects. It was expected that from the sufficient number of runs in experimental design, hence, it provides a precise prediction as well as better performance for quadratic equation model (Asghar et al., 2014). A predicted model was generated using design of experiment to determine the optimum variable conditions. A statistical mathematical equation was developed in order to enhance turbidity removal for further study in the future.

MATERIALS AND METHODS

Material for Preparing of Operating Design

Initially, fresh samples were collected at Pulai Emas Indah Water Konsortium (IWK) every week. A magnetic reactor was built and fabricated at Universiti Teknologi Malaysia (UTM) with five slots for allocating permanent magnets. Polyvinyl chloride (PVC) pipe was used because of the pipe material is not influenced by a permanent magnet; thus, the particles

are exposed directly into the magnetic field without external disturbance. A permanent magnet was made from neodymium-iron-boron (NdFeB) composition and manufactured by Ningbo Newland International Trade Co. Ltd. China. Each permanent magnet has a square shape with two faces of 100 mm × 50 mm, including 5 mm of thickness.

All five sets of permanent magnets were allocated at an inverted position (i.e., different poles of the magnet faced each other) operating magnetic design. Three variables were used in the experiment, namely the amount of steel wool, magnet strength, and circulation time at a constant flow rate of 1 ml/s. The steel wool used in the experiment was classified as non-corrosive stainless steel with the diameter of 60 μ–80 μ. The steel wool was manufactured by Lih Feng Jiing Enterprise Co. Ltd. Taiwan. Installing non-corrosive stainless steel wool into the PVC pipe produces a mechanism to increase the gradient of magnetic separation in operating design. The operating magnetic design is shown in Figure 1.

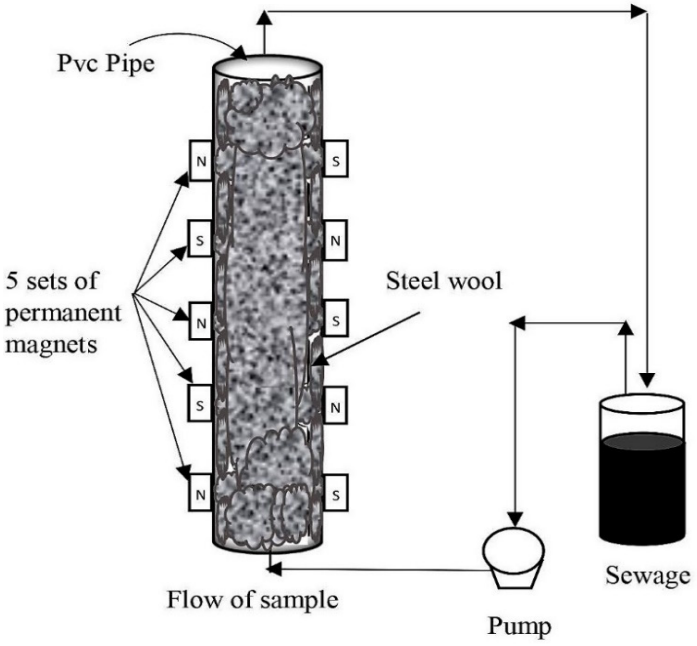


Figure 1. Design of experiment for operating high-gradient magnetic separation

Experimental Setup

The experiment was conducted with three different proportions of steel wool (40 g, 80 g, and 120 g) and magnet strength (140 mT, 195 mT, and 250 mT) under several time intervals of 24 h circulation time. Non-corrosive stainless steel wool was replaced each time after running a test. Therefore, no potential corrosion could occur during the circulation of the sample. During a certain interval time, a sample of 20 ml was taken from a plastic bottle to

conduct turbidity test. Turbidity test was measured by using Mi415 Turbidimeter following the Standard Method No. 2130 (APHA, 2005). The pH was measured using ORION 2 pH meter. pH reading and turbidity values throughout the experiment were observed at 7.0–7.2 and 19–125 NTU, respectively. The experiments were repeated twice to ensure a better estimation of the results and avoid bias. Initially, fresh sewage with no treatment was utilized as a control variable and compared with the treated samples.

Design of Experiment

The statistical prediction of design of experiment for turbidity test was carried out by using Design Expert 8.0.7 software. In this study, central composite design (CCD) under response surface methodology (RSM) was executed in order to optimize the optimum operating magnetic design. Three critical independent variables were selected, namely magnet strength (A), circulation time (B), and steel wool (C) for observing the turbidity removal performance as the response (Y_1). The range of each independent variable is presented in Table 1.

Table 1
The range of values of independent variables

Variables	Coding	Units	Level of range		
			-1	0	+1
Magnet strength	A	mT	140	195	250
Circulation time	B	h	1	12.5	24
Steel wool	C	g	40	80	120

By applying CCD, the number of run (N) in Equation 1 obtained by 2^n, which represents the factorial runs, $2n$ as the axial runs with the complementary of n_c as the center of runs, and n as the factors of experimental design (Behera et al., 2018). Three-level three-factors CCD was utilized to produce a second-order polynomial model that consists of 20 runs of experiments, including eight factorial points, six axial points, and six center points in this study.

$$N = 2^n + 2n + n_c = 2^3 + (2 \times 3) + 6 = 20 \quad [1]$$

Statistical Analysis

A predicted design matrix was produced by CCD for conducting experiments in a sequence.

Then, complete experimental data were further analyzed using RSM, which generated the optimum operating magnetic design equation for turbidity removal, as well as surface and contour plots. A general mathematical equation can be obtained by determining the correlation between the optimal critical independent values and the response as expressed in Equation 2:

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \beta_{ii} X_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{ij} X_i X_j + \varepsilon \quad [2]$$

where Y is defined as the output of experiment; β_0 is a constant regression coefficient; β_i represents a linear coefficient; β_{ii} portrays quadratic effect; β_{ii} represents the relation of coefficient; n is the number of factors involved in the study; X_iX_j is the coded value of critical independent values, and ε is the random error detected as the nuisance between the predicted and actual values in the design. There are two ways of predicting the well-fitted quality design in order to evaluate the capability of experimental design optimization: the coefficient of determination (R^2) and the analysis of variance (ANOVA) (Lee et al., 2011; Liu et al., 2016; Su et al., 2016).

The accuracy of the model based on R^2 coefficient analysis was validated and determined whether the model was accepted or rejected based on the probability value known as the *p*-value at a confidence level of 95%. Meanwhile, the ANOVA table was evaluated by Fisher's test, or also known as the *F*-test. Additionally, the quadratic effect from regression analysis in the ANOVA table displayed response and contour plot data by alternately using two critical independent values while keeping another critical independent value at constant in order to yield an optimum response.

RESULT AND DISCUSSION

Development of Model

By using Design Expert 8.0.7 software, the recommended output model was a quadratic model. A mathematical equation for optimizing turbidity removal through operational design was developed both in actual and coded factors, as shown in Table 2. Every positive sign in the mathematical equation is known as the synergistic effect where a negative sign indicates the antagonistic effect (Shuit et al., 2010; Lee et al., 2011).

The turbidity removal performance is presented in Table 3. In total, 20 experimental runs were performed according to CCD in actual values. The observed turbidity performance varied between 27% and 93% removal.

The highest removal of 93% was achieved when all the critical independent variables were at the highest values, including magnet strength of 250 mT, the proportion of non-corrosive stainless steel wool of 120 g, and running the experiment for 24 h of circulation

Table 2

An empirical model for turbidity removal both in coded and actual factors

Response/ Output	Coded factors	Actual factors
Turbidity (Y_1)	$+73.56 + 7.60A + 11.60B + 15.70C + 2.13AB - 2.87AC + 4.88BC - 1.91A^2 - 15.91B^2 - 0.41C^2$	$-43.4201 + 0.4469A + 2.5131B + 0.5558C + 0.0336 \times e^{-3}AB - 0.0131 \times e^{-3}AC + 0.0106\,BC - 0.0631 \times e^{-4}A^2 - 0.1203B^2 - 0.0256 \times e^{-4}C^2$

Notes. A: magnet strength; B: circulation time; C: steel wool; Y_1: turbidity removal

Table 3

A complete matrix design of experiment in actual values

Run no	A (mT)	B (hour)	C (gram)	Y_1 (%)
1	140	24	120	83
2	195	12.5	80	73
3	195	12.5	80	72
4	140	1	120	50
5	195	1	80	44
6	250	24	40	62
7	140	1	40	27
8	195	12.5	80	77
9	195	24	80	67
10	250	12.5	80	76
11	195	12.5	120	87
12	140	24	40	32
13	195	12.5	80	79
14	195	12.5	40	55
15	250	1	40	40
16	250	1	120	60
17	250	24	120	93
18	195	12.5	80	71
19	195	12.5	80	78
20	140	12.5	80	63

time. Meanwhile, the lowest removal of 27% was recorded for the magnet strength of 140 mT, 40 g of non-corrosive stainless steel wool, and the turbidity test was conducted during the first hour of circulation time. The ANOVA table is shown in Table 4. From the Table 4, the experimental data model was declared as significant according to the *F*-value of 49.15 and *p*-value < 0.0001, which simultaneously stated that the model is statistically significant as the *p*-value is less than 0.05 or significant at 95% of confidence level.

Table 4

Adequacy of turbidity removal through ANOVA table

Response	Source	Sum of square	Degree of freedom	Mean square	F value	Prob > F	Remarks
Y_1	Model	6191.00	9	687.89	49.15	< 0.0001	significant
	A	577.60	1	577.60	41.27	< 0.0001	
	B	1345.60	1	1345.60	96.15	< 0.0001	
	C	2464.90	1	2464.90	176.12	< 0.0001	
	AC	66.12	1	66.12	4.72	0.0548	
	BC	190.13	1	190.13	13.58	0.0042	
	B^2	696.02	1	696.02	49.73	< 0.0001	
	Residual	139.95	10	14.00			
	Lack of Fit	81.95	5	16.39	1.41	0.3568	not significant
	Pure Error	58.00	5	11.60			
	Cor Total	6330.95	19				
	Std. Dev.	3.74	R-Squared	0.9779			
	Mean	64.45	Adj R-Squared	0.9580			
	C.V. %	5.80	Pred R-Squared	0.8227			
	Press	1122.22	Adeq Precision	26.386			

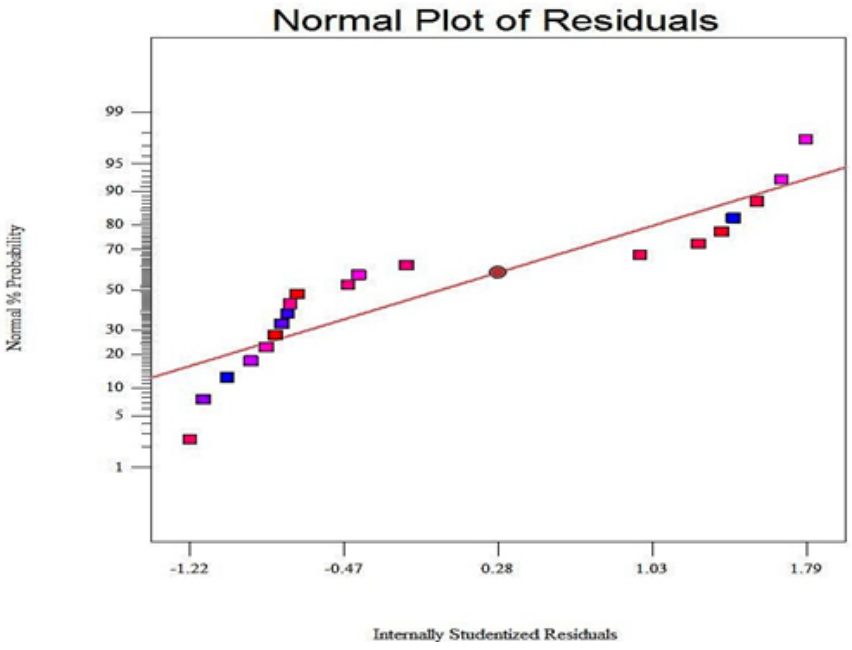
Notes. A: magnet strength; B: circulation time; C: steel wool; Y_1: turbidity removal; Cor Total: Correlation total; Std. Dev: Standard deviation; C.V.: Coefficient of variation; Adj R-Squared: Adjusted R^2; Pred R-Squared: Predicted R^2; Adeq Precision: Adequate precision

Furthermore, all three critical independent variables were pronounced as statistically significant as the *p*-value was less than 0.05 in the experimental model. However, it was found that non-corrosive stainless steel wool had a major effect on the turbidity removal, with the *F*-value of 176.12 compared to others. Nevertheless, all the independent variables were considered as potentially influential towards the turbidity removal performance in sewage treatment. In addition, at the first-order coefficient effect or linear effect, two combinations of independent variables were statistically significant. The combination between magnet strength and steel wool, as well as the combined effects between circulation time and steel wool were statistically significant with *p*-values of 0.0548 and 0.0042, respectively. The insignificant variable can be eliminated as it has limited influence, which degrades the optimum performance of a model (Amr et al., 2014).

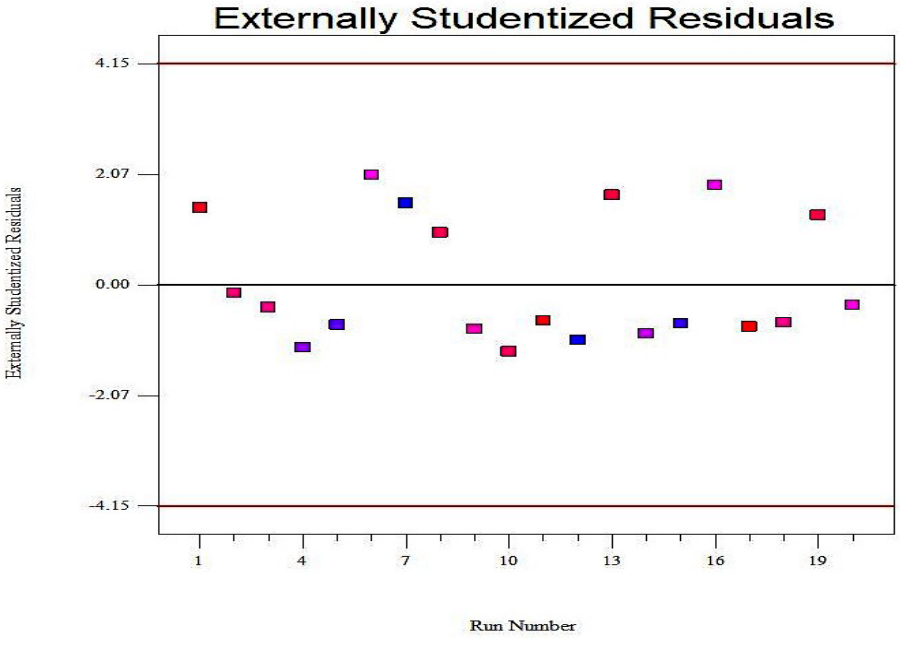
Furthermore, the evaluation of the well-fitted model could also be observed through several critical values from the ANOVA table. For example, the expected remark of lack-of-fit (LOF) should be not significant with *p*-value > 0.05, which is fundamental for determining that the model is adequately fitted to the experimental data (Bezerra et al., 2008). It is verified that the possibility of the input between the regressor-response relationships is unnecessary to be included in the model (Noordin et al., 2004). From this study, the lack-of-fit (LOF) was observed as insignificant with the *p*-value of 0.3568, which was higher than 0.05; thus, the obtained model was statistically significant. Moreover, the R^2 value closer to 1 can be suggested as reliable as predicted by the model Su et al. (2016), where R^2 was determined as 97.8%. This could explain that the data fitted to the model and promoted an optimum estimation of the output in the range of study.

In statistics, the signal of noise ratio is recognized as adequate if the value of adequate precision coefficient (AP) is higher than four (Markandeya et al., 2017). The AP in this study was 26.386 (i.e., higher than four), and the model could be claimed as desirable. The coefficient of variation (CV) should be less than 10% in order to achieve an optimum model; in this study, the CV obtained was significantly higher at 5.8%, but the value was still acceptable for a reliable model (Javanbakht & Ghoreishi, 2017). Apparently, the competency of a model can be evaluated by analyzing the pattern of a normal probability graph of residuals. Residuals are known as the distinct pattern between the predicted and observed response data (Lee et al., 2011). In general, a normal distribution is defined when all the residuals data would align and remain closer to the straight line; however, it is expected that some of the residual's plots would spread during the experiment, as shown in Figure 2(a) (Amr et al., 2014). Nevertheless, as long as the predicted experiment values were included between the tolerance limits represent by the red lines in Figure 2(b), the model could still be indicated as statistically significant (Zaidi et al., 2014).

In addition, all randomness and structureless plots signified that the model is satisfactorily adequate for the experimental data, and there is no need for the assumption



(a)



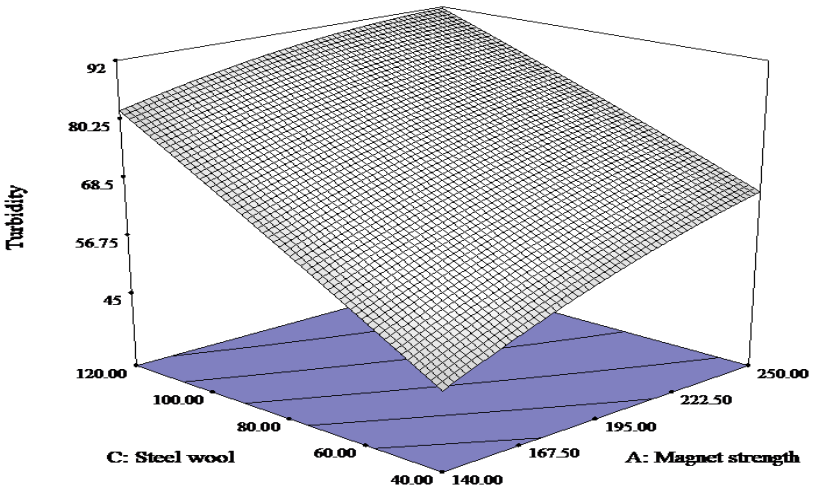
(b)

Figure 2. Turbidity removal for (a) normal probability plot with residuals and (b) predicted experiment data with residual plot

of variance that could promote the transformation of output variables (Noordin et al., 2004; Behera et al., 2018). Consequently, all the essential characteristics for determining the significance of the model are well accomplished by the experimental data; thus, the design model can be used for further studies by evaluating surface and contour plots.

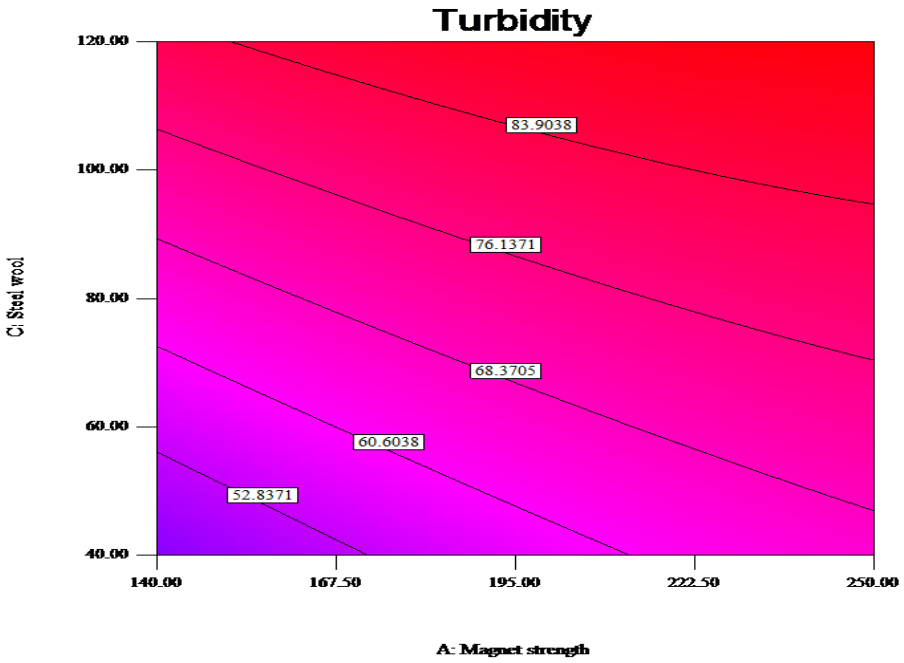
The Combined Effect of Magnet Strength and Steel Wool towards Turbidity

In this section, RSM was used to evaluate the relationship for the combined correlation effects of three-level factors on turbidity removal. Figure 3 represents the relationship between the significant interaction of magnet strength (A) and steel wool (C) through (a) a surface plot and (b) a contour plot. The values on both graphs are actual values. Turbidity removal displayed combined effects of *A* and *C* at the circulation time of 12.5 h, as shown in Figure 3 below. From the 3D surface plot in Figure 3(a), turbidity removal increased significantly due to the influence of increasing magnet strength (140 mT to 250 mT) and by adding 40 g to 120 g of non-corrosive stainless steel wool at constant circulation time (12.5 h). The maximum turbidity removal observed was 91.6%. On the contrary, both of the lowest values of steel wool and magnet strength did not accelerate turbidity removal. This could be explained by Ebner et al. (1997) and Li et al. (2012), who stated that when the high-gradient magnetic field engaged with non-corrosive stainless steel wool, the gradient of the magnetic field increased appropriately as steel wool started to magnetize.



(a)

Figure 3. Relationship between magnet strength (A) and steel wool (C) on turbidity removal at circulation time of 12.5 h through (a) surface plot and (b) contour plot



(b)

Figure 3.(Continued)

It was also claimed by Rao and Sahu (2013), that when magnet strength was increased to increase magnetic gradient, charged ions were produced, which attracted to different charges from water molecules in a sample. Thus, weak magnetized particles and tiny pollutant in sewage could be removed easily. The presence of non-corrosive stainless steel wool itself provides higher density or compression in a magnetic field medium (Hatch & Stelter, 2001; Ge et al., 2017; Ren et al., 2017). Whenever a particle circulates through a pipe, there is a higher chance of the particle being potentially exposed and magnetized, which is a significantly important process in HGMS. When a particle is magnetized, it is simultaneously attracted to magnetized steel wool; thus, higher turbidity removal can be achieved.

As proven from the contour plot in Figure 3(b), more than 50% of turbidity removal was observed in the early experiment, followed by a sharp increase of turbidity removal as the magnet strength and steel wool increased proportionately. Hence, the correlation between magnet strength and non-corrosive stainless steel wool is the most important relationship for turbidity removal in sewage treatment.

The Combined Effect of Circulation Time and Steel Wool towards Turbidity

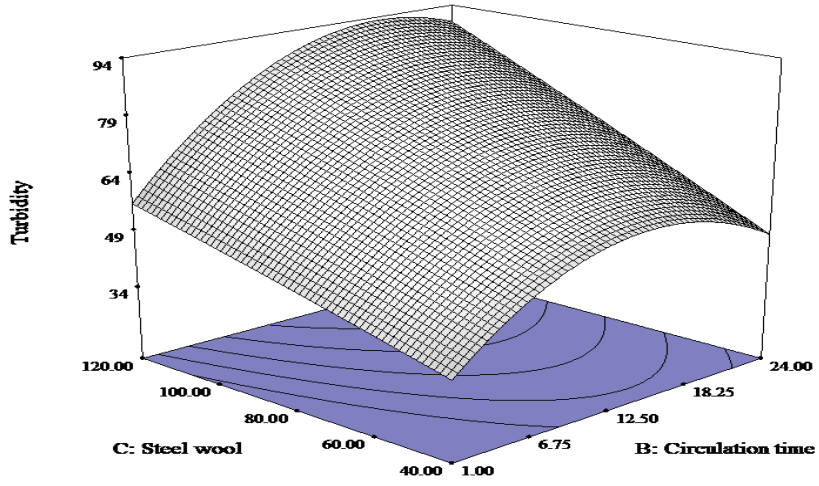
Figure 4 illustrates the relationship of circulation time (B) and steel wool (C) through (a) a surface plot and (b) a contour plot. Turbidity removal displayed the combined effects of B and C at magnet strength of 195 mT in Figure 4 (a). At a constant value of magnet strength, the 3D surface plot showed that the increase of circulation time explained the significant increase on turbidity removal (45% to 74%) at low and intermediate levels of steel wool. According to Figure 4 (b), the shape of the contour revealed that more than 75% of turbidity removal was achieved for steel wool and circulation time at intermediate and high levels. From this circumstance, it could be analyzed that the longer circulation time involved, the higher turbidity removal could be obtained.

This correlation is obtained when magnetized particles circulate repeatedly and continuously through a magnetic field, and the particles are able to retain magnetic memory. Magnetic memory is essential for particles as it affects magnetization behavior; thus, particles start to accumulate and attract particles with different charges (Baker & Judd, 1996; Colic & Morse, 1999). According to Higashitani et al. (1992), magnetic memory can be preserved by particles until 143 hours after being exposed to a magnetic field for 30 min. Meanwhile, Sohaili (2003) stated that particles could retain the magnetic memory almost permanently if they were continuously exposed in a longer period of time. Therefore, bigger particles are formed and could be easily trapped in fibrous steel wool. The increment of turbidity removal still occurred when non-corrosive stainless steel wool was added until 120 g and circulation time about to reach 24 h.

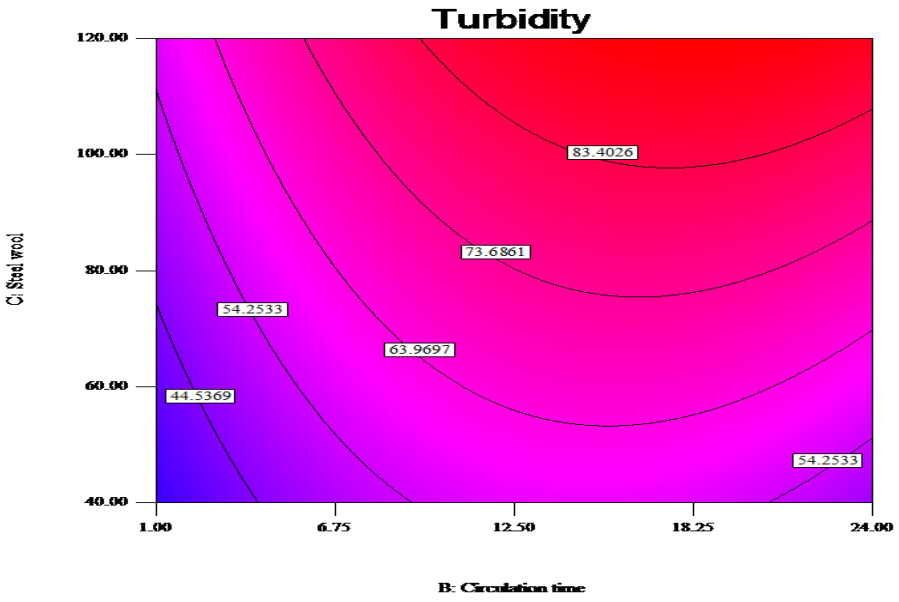
However, the turbidity removal reached a plateau when both independent variables reached the maximum values. This could be further explained that particles from the sample are almost fully acquired in fibrous non-corrosive stainless steel wool; thus, a treated sample with less colloidal and residual suspended matter can be achieved. Similarly, a previous study on chemical oxygen demand (COD) reduction using electromagnetic by Rao and Sahu (2013) found that a longer duration of exposure time to a magnetic field promoted a higher chance for each particle to achieve ionic charge. Therefore, charged particles potentially have a greater tendency to attract either positively or negatively charged particles and simultaneously form bigger particles that would be either trapped in the steel wool or settled down by gravity.

Optimization of Model

The optimization of the model was conducted using Design Expert 8.0.7 software. This software can statistically correlate each variable within the values in order to optimize the most optimum operating magnet design for this study, thus producing a desirable response. Each of the critical operational variable (i.e., A, B, and C) was selected within the range



(a)



(b)

Figure 4. Relationship between circulation time (B) and steel wool (C) on turbidity removal at magnet strength of 195 mT through (a) surface plot and (b) contour plot

of values. The output for turbidity removal (Y_1) was determined as the maximum removal performance according to the selected critical operational variables.

From the optimization, the optimum conditions of A of 245 mT, B of 16 h, and C of 116 g are expected to approximately achieve 94% of turbidity removal. In order to ensure that the validation of optimization process correlates significantly with experimental design, an additional experiment was conducted to analyze the optimum operational conditions for avoiding any bias result (Lee et al., 2011; Hwang & Han, 2015; Su et al., 2016). By using the same optimum operational conditions, turbidity removal of 90.3% was achieved with difference only 3.7% of error compared to predicted response value. Based on the small difference between the predicted response value and the experiment value, it can be concluded that the difference is still within the maximum range of response (Lee et al., 2011; Behera et al., 2018). Therefore, the predicted optimization model was valid for further study references. Moreover, the normal probability analysis graph acknowledged that the error distribution from experimental data was adequately distributed along the line which interprets the design model adjusted well with experimental data (Peng et al., 2020). Additionally, according to the ANOVA table analysis, the error between the R^2 value (i.e. 0.9779) and the adjusted R^2 value (i.e. 0.9580) was 0.02 which less than 0.2, verified that accuracy of experimental design (Su et al., 2016). Most prominently, R^2 value of actual data was higher than R^2 value of predicted data (i.e. 0.8227) which indicated that the experimental data suited well with design of experiment. Hence, it certified that the magnetic strength, circulation time and non-corrosive stainless steel wool had effects of 97.8% on the suspended matter removal by turbidity test. Therefore, all of optimum operating magnetic design values were showed that an acceptable predictability and precisely accuracy for determining suspended matter removal performance.

CONCLUSIONS

Combination of three different amounts of non-corrosive stainless steel wool and magnetic strengths were employed to maximize the removal of suspended matter under several time intervals of 24 h circulation time. As a conclusion, the maximal of turbidity removal was indeed increased by magnetic gradient as it occurred when higher amount of non-corrosive stainless steel wool combined with higher magnetic strength. Therefore, without using any additional magnetic adsorbent, suspended matter removal could be enhanced by using optimum magnetic gradient. According to the ANOVA analysis, all of three operating magnetic design conditions affected the suspended matter removal statistically significant as the *p*-value is less than 0.05. The extend of suspended matter removal was escalated at optimized operating magnetic design to obtain highest removal percentage by turbidity test. The experimental output was determined as 90.3% using the optimized conditions of magnet strength (A) of 245 mT, circulation time (B) of 16 h, and steel wool (C) of 116

grams. The empirical model stipulated that the suspended matter removal by turbidity, which subsequently improved water quality.

ACKNOWLEDGEMENT

A very special thank to both Universiti Teknologi Malaysia (UTM) and Ministry of Science, Technology and Innovation (MOSTI) for supporting this research titled " Optimization and Modelling of Turbidity Removal of Sewage using High Gradient Magnetic Separation (HGMS) by Response Surface Methodology (RSM)" financially (Project No. 4S032).

REFERENCES

Ambashta, R. D., & Sillanpää, M. (2010). Water purification using magnetic assistance: A review. *Journal of Hazardous Materials*, *180*, 38-49.

APHA. (2005). *Standard methods for the examination of water and wastewater* (21st Ed.). Washington DC, USA: American Public Health Association.

Amr, S. S. A., Aziz, H. A., & Bashir, M. J. K. (2014). Application of response surface methodology (RSM) for optimization of semi-aerobic landfill leachate treatment using ozone. *Applied Water Science*, *4*, 1-10.

Asghar, A., Raman, A. A. A., & Daud, W. M. A. W. (2014). A comparison of central composite design and taguchi method for optimizing fenton process. *The Scientific World Journal*, *2014*, 1-14.

Baik, S. K., Ha, D. W., Ko, R. K., & Kwon, J. M. (2010). Magnetic field and gradient analysis around matrix for HGMS. *Physica C*, *470*, 1831-1836.

Baik, S. K., Ha, D. W., Kwon, J. M., Lee, Y. J., & Ko, R. K. (2013). Magnetic force on a magnetic particle within a high gradient magnetic separator. *Physica C: Superconductivity*, *484*, 333-337.

Baker, J. S., & Judd, S. J. (1996). Magnetic amelioration of scale formation. *Water Research*, *30*(2), 247-260.

Baker, J. S., Judd, S. J., & Parsons, S. A. (1997). Antiscale magnetic pretreatment of reverse osmosis feedwater. *Desalination*, *110*, 151-165.

Behera, K. S., Meena, H., Chakraborty, S., & Meikap, B. C. (2018). Application of response surface methodology (RSM) for optimization of leaching parameters for ash reduction from low-grade coal. *International Journal of Mining Science and Technology*, *28*(4), 621-629.

Bezerra, M. A., Santelli, R. E., Oliveiraa, E. P., Villara, L. S., & Escaleira, L. A. (2008). Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta*, *76*, 965-977.

Bitton, G., Fox, J. L., & Strickland, H. G. (1975). Removal of algae from Florida lakes by magnetic filtration. *Journal of Applied Microbiology*, *30*(6), 905-908.

Blamire, M. G., Dinner, R. B., Wimbush, S. C., & MacManus-Driscoll, J. L. (2009). Critical current enhancement by Lorentz force reduction in superconductor - ferromagnet nanocomposites. *Superconductor Science and Technology*, *22*, 1-6.

Busch, K. W., & Busch, M. A. (1997). Laboratory studies on magnetic water treatment and their relationship to a possible mechanism for scale reduction. *Desalination*, *109*, 131-148.

Chen, W., Lin, H., & Zhang, S. (2011). Study on treatment of printing and dyeing wastewater by magnetic seed coagulation - high gradient magnetic separation. *Applied Mechanics and Material*, *71-78*, 2689-2694.

Colic, M., & Morse, D. (1999). The elusive mechanism of the magnetic memory of water. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, *154*, 167-174.

Dhoble, R. M., Lunge, S., Bhole, A. G., & Rayalu, S. (2011). Magnetic binary oxide particles (MBOP): A promising adsorbent for removal of As(III) in water. *Water Research, 45*(16), 4769-4781.

Ebner, A., D., Ritter, J., A., & Ploehn, H., J. (1997). Feasibility and limitations of nano-level high gradient magnetic separation. *Separation and Purification Technology, 11,* 199-210.

Franzreb, M., Siemann-Herzberg, M., Hobley, T. J., & Thomas, O. R. T. (2006). Protein purification using magnetic adsorbent particles. *Applied Microbiology and Biotechnology, 70,* 505-516.

Ge, W., Encinas, A., Araujo, E., & Song, S. (2017). Magnetic matrices used in high gradient magnetic separation (HGMS): A review. *Results in Physics*, *7*, 4278-4286.

Gugushe, A. S., Nqombolo, A., & Nomngongo, P. N. (2019). Application of response surface methodology and desirability function in the optimization of adsorptive remediation of arsenic from acid mine drainage using magnetic nanocomposite : Equilibrium studies and application to real samples. *Molecules*, *24*, 1-15.

Gupta, V. K., & Suhas. (2009). Application of low-cost adsorbents for dye removal - A review. *Journal of Environmental Management, 90*(8), 2313-2342.

Hatch, G. P., & Stelter, R. E. (2001). Magnetic design considerations for devices and particles used for biological high-gradient magnetic separation (HGMS) systems. *Journal of Magnetism and Magnetic Materials*, *225*(1-2), 262-276.

He, J., Huang, M., Wang, D., Zhang, Z., & Li, G. (2014). Magnetic separation techniques in sample preparation for biological analysis : A review. *Journal of Pharmaceutical and Biomedical Analysis, 101*, 84-101.

Higashitani, K. O., Okuhara, K., & Hatade, S. (1992). Effects of magnetic fields on stability of nonmagnetic ultrafine colloidal particles. *Journal of Colloid And Interface Science*, *152*(I), 125-131.

Hoffmann, C., Franzreb, M., & Höll, W. H. (2002). A novel high-gradient magnetic separator (HGMS) design for biotech applications. *IEEE Transactions on Applied Superconductivity*, *12*(1), 963-966.

Hournkumnuard, K., & Chantrapornchai, C. (2011). Parallel simulation of concentration dynamics of nano-particles in High Gradient Magnetic Separation. *Simulation Modelling Practice and Theory, 19*(2), 847-871.

Hwang, J., & Han, D. (2015). Optimization and modeling of reduction of wastewater sludge water content and turbidity removal using magnetic iron oxide nanoparticles (MION). *Journal of Environmental Science and Health*, *50*, 1307-1315.

Javanbakht, V., & Ghoreishi, S. M. (2017). Application of response surface methodology for optimization of lead removal from an aqueous solution by a novel superparamagnetic nanocomposite. *Adsorption Science and Technology*, *35*(1-2), 241-260.

Kace, J. S., & Linford, H. B. (1975). Reduced cost flocculation of a textile dyeing wastewater. *Journal of Water Pollution Control Federation, 47*, 1971-1977.

Karapinar, N. (2003). Magnetic separation of ferrihydrite from wastewater by magnetic seeding and high-gradient magnetic separation. *International Journal of Mineral Processing*, *71*(1-4), 45-54.

Käppler, T., Cerff, M., Ottow, K., Hobley, T., & Posten, C. (2009). In situ magnetic separation for extracellular protein production. *Biotechnology and Bioengineering, 102*(2), 535-545.

Kwon, H., Ha, D., & Kim, Y. (2014). Superconducting magnetic separation of powdered activated carbon from aqueous phase. *IEEE Transactions on Applied Superconductivity*, *24*(3), 3-6.

Lee, H. V., Yunus, R., Juan, J. C., & Tau, Y. H. (2011). Process optimization design for jatropha-based biodiesel production using response surface methodology. *Fuel Processing Technology*, *92*, 2420-2428.

Lee, J. W., Choi, S. P., Thiruvenkatachari, R., Shim, W. G., & Moon, H. (2006). Evaluation of the performance of adsorption and coagulation processes for the maximum removal of reactive dyes. *Dyes Pigments*, *69*, 196-203.

Li, S. Q., Wang, M. F., Zhu, Z. A., Wang, Q., Zhang, X., Song, H. Q., & Cang, D. Q. (2012). Application of superconducting HGMS technology on turbid wastewater treatment from converter. *Separation and Purification Technology*, *84*, 56-62.

Liu, Y., Liu, F., Ni, L., Meng, M., & Meng, X. (2016). A modeling study by response surface methodology (RSM) on Sr (II) ion dynamic adsorption optimization using a novel magnetic ion. *The Royal Society of Chemistry, 6*, 54679-54692.

Mariani, G., Fabbri, M., Negrini, F., & Ribani, P. L. (2010). High-gradient magnetic separation of pollutant from wastewaters using permanent magnets. *Separation and Purification Technology, 72*(2), 147-155.

Markandeya, Dhiman, N., Shukla, S. P., & Kisku, G. C. (2017). Statistical optimization of process parameters for removal of dyes from wastewater on chitosan cenospheres nanocomposite using response surface methodology. *Journal of Cleaner Production, 149*, 597-606.

Muntean, S. G., Nistor, M. A., Muntean, E., Todea, A. G, R. I., & Cornelia, P. (2018). Removal of colored organic pollutants from wastewaters by magnetite/carbon nanocomposites: Single and binary systems. *Journal of Chemistry*, *2018*, 1-16.

Narian, E., Arami, M., Bahrami, H., & Pajootan, E. (2015). Modification of nickel ferrite with cationic surfactant: Dye removal from textile wastewater using magnetic separation. *Journal of Environmental Engineering*, *141*(2), 1-10.

Noordin, M. Y., Venkatesh, V. C., Sharif, S., Elting, S., & Abdullah, A. (2004). Application of response surface methodology in describing the performance of coated carbide tools when turning AISI 1045 steel. *Journal of Materials Processing Technology, 145,* 46-58.

Oliveira, F. H., Osugi, M. E., Paschoal, F. M. M., Profeti, D., Olivi, P., & Zanoni, M. V. B. (2007). Electrochemical oxidation of an acid dye by active chlorine generated using Ti/Sn $_{(1-x)}$ Ir_xO_2 electrodes. *Journal of Applied Electrochemical*, *37*, 583-592.

Oshitani, J., Yamada, D., Miyahara, M., & Higashitani, K. (1999). Magnetic effect on ion-exchange kinetics. *Journal of Colloid and Interface Science*, *210*, 1-7.

Peng, Y., Khaled, U., Al-Rashed, A. A. A. A., Meer, R., Goodarzi, M., & Sarafraz, M. M. (2020). Potential application of response surface methodology (RSM) for the prediction and optimization of thermal conductivity of aqueous CuO (II) nanofluid: A statistical approach and experimental validation. *Physica A, 554,* 1-8.

Podoynitsyn, S. N., Sorokina, O. N., & Kovarski, A. L. (2016). High-gradient magnetic separation using ferromagnetic membrane. *Journal of Magnetism and Magnetic Materials*, *397*, 51-56.

Rajasulochana, P., & Preethy, V. (2016). Comparison on efficiency of various techniques in treatment of waste and sewage water - A comprehensive review. *Resource-Efficient Technologies*, *2*(4), 175-184.

Rao, M. S., & Sahu, O. (2013). Study of electromagnetic waves on industrial waste water. *Physics and Materials Chemistry*, *1*(2), 34-40.

Ren, Z., Lin, Z., Ye, Z., Xiangyu, Z., & Dongxing, M. (2017). Effects and mechanization of a high gradient magnetic separation process for particulate and microbe removal from ballast water. *International Journal of Environmental and Ecological Engineering*, *11*(6), 553-558.

Setchell, C. H. (1985). Magnetic separations in biotechnology - A review. *Journal of Chemical Technology and Biotechnology*, *35*(3), 175-182.

Shuit, S. H., Lee, K. T., Kamaruddin, A. H., & Yusup, S. (2010). Reactive extraction of *Jatropha curcas* L. seed for production of biodiesel: Process optimization study. *Environmental Science and Technology, 44*, 4361-4367.

Sirisha, K., Suganya, B., Sivasubramanian, V., Bs, V., Swaminathan, D., Babu, A. C., & Meyyappan, N. (2017). Studies on the effect of pulsed magnetic field on the productivity of algae grown in dye industry effluent. *Journal of Applied Biotechnology and Bioengineering*, *3*(5), 409-413.

Sohaili, J. (2003). *Capability of the magnetic technology to increase the suspended particles sedimentation rate from wastewater* (Doctoral Thesis). Universiti Teknologi Malaysia, Malaysia.

Su, Z., Li, X., Yang, Y., Xu, M., Ding, Y., & Zhou, Z. (2016). Optimization of magnetic-seeding coagulation in artificially polluted surface water treatment by response surface methodology. *Desalination and Water Treatment*, *57*, 20671-20682.

Takayasu, M., Hwang, J. Y., Friedlaender, F. J., Petrakis, L., & Gerber, R. (1984). Magnetic separation utilizing a magnetic susceptibility gradient. *IEEE Transactions on Magnetics, 20*, 155-159.

Tomska, A., & Wolny, L. (2008). Enhancement of biological wastewater treatment by magnetic field exposure. *Desalination*, *222*, 368-373.

Ueda, H., Agatsuma, K., Kajikawa, K., Furuse, M., Fuchino, S., & Ishiyama, A. (2009). Design and test of filter of high gradient magnetic separation system for trapping immunoglobulin in serum. *IEEE Transactions on Applied Superconductivity, 19*, 2157-2161.

Wang, Q., Guan, Y., Ren, X., Yang, M., & Liu, X. (2012). Removal of low concentration Cr(VI) from aqueous solution by magnetic-fluids fixed bed using the high gradient magnetic separation. *Journal of Colloid and Interface Science*, *374*(1), 325-330.

Zaidi, N. S., Sohaili, J., Muda, K., & Sillanpaa, M. (2014). Magnetic field application and its potential in water and wastewater treatment systems. *Journal of Separation and Purification Reviews*, *43*, 206-240.

Zheng, X., Wang, Y., & Lu, D. (2015). A realistic description of influence of the magnetic field strength on high gradient magnetic separation. *Minerals Engineering*, *79*, 94-101.

Zheng, X., Xue, Z., Wang, Y., Zhu, G., Lu, D., & Li, X. (2019). Modeling of particle capture in high gradient magnetic separation : A review. *Powder Technology*, *352*, 159-169.

Zhou, Y., Nie, H., Branford-white, C., He, Z., & Zhu, L. (2009). Removal of Cu^{2+} from aqueous solution by chitosan-coated magnetic nanoparticles modified with α-ketoglutaric acid. *Journal of Colloid and Interface Science*, *330*(1), 29-37.

Pertanika J. Sci. & Technol. 28 (4): 1389 - 1411 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Intelligent Bio-Inspired Whale Optimization Algorithm for Color Image Based Segmentation

Athraa Jasim Mohammed* and Khalil Ibrahim Ghathwan

Department of Computer Science, University of Technology, Baghdad 00964, Iraq

ABSTRACT

Color image segmentation is widely used methods for searching of homogeneous regions to classify them into various groups. Clustering is one technique that is used for this purpose. Clustering algorithms have drawbacks such as the finding of optimum centers within a cluster and the trapping in local optima. Even though inspired meta-heuristic algorithms have been adopted to enhance the clustering performance, some algorithms still need improvements. Whale optimization algorithm (WOA) is recognized to be enough competition with common meta-heuristic algorithms, where it has an ability to obtain a global optimal solution and avoid local optima. In this paper, a new method for color image based segmentation is proposed based on using whale optimization algorithm in clustering. The proposed method is called the whale color image based segmentation (WhCIbS). It was used to divide the color image into a predefined number of clusters. The input image in RGB color space was converted into L*a*b color space. Comparison of the proposed WhCIbS method was performed with the wolf color image based segmentation, cuckoo color image based segmentation, bat color image based segmentation, and k-means color image based segmentation over four benchmark color images. Experimental results demonstrated that the proposed WhCIbS had higher value of PSNR and lower value of RMSR in most cases compared to other methods.

ARTICLE INFO

Article history:
Received: 22 February 2020
Accepted: 27 May 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.14

E-mail addresses:
10872@uotechnology.edu.iq (Athraa Jasim Mohammed)
110039@uotechnology.edu.iq (Khalil Ibrahim Ghathwan)
*Corresponding author

Keywords: Clustering based technique, color image based segmentation, whale optimization algorithm

INTRODUCTION

Image segmentation is the separation of an image into homogeneous regions. It is frequently a main step in image investigation, item representation, visual

ISSN: 0128-7680
e-ISSN: 2231-8526

image and numerous other image processing jobs. The aim of image segmentation is to represent image into different description to make it simpler and easier for analyzing (Zaitoun & Aqel, 2015). There are four types of image: intensity image, binary image, indexed image, and color image (Gonzalez et al., 2003). Color image is classified into different color spaces such as RGB, L*a*b, YcbCr and HSV.The L*a*b space is excessively used due to less execution time (Mathur & Purohit, 2014). In this research, a color image based on RGB and L*a*b color spaces was applied.

In previous works, there are various segmentation techniques of image such as an edge or boundary based, threshold, clustering based, and soft computing approaches like genetic algorithm and Neural Network. Clustering based technique is an efficient technique (Dhanachandra et al., 2015). This study focused on clustering based technique.

Clustering can be defined as a process of splitting homogeneous objects in one cluster and heterogeneous objects in another cluster (Kapoor et al., 2017). In literature, there are different methods of clustering based such as: k-means clustering (Dhanachandra et al., 2015; Mathur & Purohit, 2014), DBScan clustering method (Ye et al., 2003), subtractive clustering (Dhanachandra et al., 2015), and most recently used intelligent inspired optimization algorithms (i.e.meta-heuristic algorithms) in clustering like Wolf optimization algorithm (Kapoor et al., 2017), Cuckoo search algorithm (Nandy et al., 2015), Particle Swarm Optimization (Zhang et al., 2011), Bat optimization algorithm (Alagarsamy et al., 2019; Mishra & Panda, 2018; Yang, 2010) and most recently introduced Whale Optimization algorithm (Mirjalili & Lewis, 2016).

Whale Optimization algorithm (Mirjalili & Lewis, 2016) is a nature inspired algorithm that uses in many fields and is proven successful (Mafarja & Mirjalili, 2017; Thennarasu et al., 2020). In this paper, a new method for Color Image based Segmentation is proposed based on intelligent inspired Whale optimization algorithm, term as WhCIbS. The proposed method was compared with Wolf Color Image based Segmentation, termed as WCIbS, Cuckoo Color Image based Segmentation, termed as CCIbS, Bat Color Image based Segmentation, termed as BCIbS, and k-means Color Image based Segmentation, termed as KmCIbS.

The organization of the remaining paper is as follows. Literature review includes related works and a brief description about image segmentation using intelligent bio-inspired clustering methods and k-means clustering presents in the next section, followed by materials and methods section that includes the proposed Whale Color Image based Segmentation Method (WhCIbS), then, section results and discusions that contain experiments that are compared proposed inspired Whale method with three bio-inspired clustering methods and K-means clustering methods. Finally, conclusions is explained in the last section.

LITERATURE REVIEW

Image segmentation is defined as "a processing technique used to classify or cluster an image into several disjoint parts by grouping the pixels to form a region of homogeneity based on the pixel characteristics like gray level, color, texture, intensity and other features" (Khan & Ravi, 2013).

The method for segmentation techniques in image are divided into two parts: Layer Based (LB) and Block Based (BB) (Zaitoun & Aqel, 2015). The Layer Based method is not discussed in this paper, for more information about it refer to Juliet et al. (2014). The BB segmentation method analyses and understands the image depending on the diverse characteristics that appear in it. The BB segmentation method is classified into two parts: Edge Based and Region Based, where, there are very interesting research papers done in these two areas.

In Edge Based method, the image is transformed to edge image by changing tones in the image. On the other side, Rigon Based classified the image into sub regions (clusters) (Zaitoun & Aqel, 2015). There are five categories of Rigon Based segmentation method; split and merge (Ning et al., 2010), normalized cuts (Shi & Malik, 2000), region growing (Tang, 2010), threshold (Mishra & Panda, 2018; Sarkar et al., 2011), and finally clustering (Alagarsamy et al., 2019; Zhang et al., 2011). This study is focused on the region based color image segmentation specified in clustering method.

Clustering method groups an image into k clusters of similar objects, homogeneous objects in one group while the heterogeneous objects are isolated. In literature, there are various works on medical image segmentation, satellite image segmentation, and color image segmentation that use traditional clustering methods combined with meta-heuristic algorithms.

In medical image segmentation, Dhanachandra, Manglen and Chanu (2015) proposed a method based on subtractive and k-means clustering algorithm, where, subtractive clustering algorithm found the initial centers from the data then k-means clustering algorithm works depending on these centers. After that, any unwanted regions are removed from segmentation image using medial filter. The result demonstrates that their proposed better than classical k-means algorithm in terms of PSNR and RMSE (Dhanachandra et al., 2015).

In satellite image segmentation, Kapoor et al. (2017) suggested satellite image segmentation method using the Grey wolf optimizer algorithm as an automatic clustering algorithm. The result of their proposed method is efficient in terms of Davies-Bouldin-index and inter-intra cluster distance.

In color image segmentation, Ye et al. (2003) proposed image segmentation method based on density based clustering. Nandy et al. (2015) proposed color image segmentation method optimized by Cuckoo Search algorithm CS. The researchers used CS algorithm to

cluster the image after finding optimal centers (Nandy et al., 2015). Mathur and Purohit (2014) developed color image segmentation methods using k-means clustering algorithm, where, it classified the image into different clusters depending on Euclidean distance metric.

In the current algorithms, there has been a tendency to combining the inspired meta-heuristic algorithms with clustering methods for improving the clustering performance. Although, some inspired meta-heuristic algorithms such as Particle Swarm Optimization has some disadvantages (i.e. low convergence rates and generating low quality solutions). This has happened because of the working mechanism of Particle Swarm Optimization keeping a single best solution through search space that can perform to local optima (Ghany et al., 2020).

Whale Optimization Algorithm is a well-known optimization algorithm. It has advantages such as minimum number of parameters, lack of local optima and acquiring a global optimal solution. These advantages lead WOA to become appropriate algorithm for finding the solutions of constrained or unconstrained optimization problems (Nasiri & Khiyabani, 2018). Mirjalili and Lewi (2016) had examined WOA algorithm against other inspired metaheuristic algorithms like: PSO, Differential Evolutional, and other using twenty nine mathematical benchmark optimization problems. The results gave evidence that WOA a competitive with others (Mirjalili & Lewis, 2016; Nasiri & Khiyabani, 2018). Further, Jadhav and Gomathi (2018) proposed a WOA algorithm for data clustering based on three constraints termed inter-cluster distance, intra-cluster and the cluster density. Their proposed outperformed PSO, modified PSO, grey wolf optimization (Jadhav & Gomathi, 2018). Also, Ghany et al. (2020) introduced combined WOA with Tabu search for data clustering. The results of their work can converge to global optima in less iteration, which it is better than PSO and other competitors (Ghany et al., 2020). In addition, Nasiri and Khiyabani (2018) proposed WOA algorithm for data collected from the UCI machine learning repository. Their proposing had shown success of WOA compared to PSO, ABC, GA, DE, and K-means (Nasiri & Khiyabani, 2018).

Further, In literature, Whale Optimization Algorithm uses to determine the optimal multilevel threshold for image segmentation, wherethe researchers: El Aziz et al. (2017) used Whale Optimization Algorithm and moth flame Optimization with Otsu's objective function. While in another work, El Aziz et al. (2018) utilized Whale Optimization Algorithm with multi objective function for optimizing the threshold that used to split the image. Whale Optimization Algorithm is also used in the segmentation of medical image (i.e. liver image), where Mostafa et al. (2017) utilized manual segmented image dataset, then they converted them into binary images. The researchers used Whale Optimization Algorithm to cluster the intensity values in these images (Mostafa et al., 2017). Based on our knowledge, there is no work about utilizing Whale Optimization Algorithm for color image segmentation.

Whale Optimization Algorithm

Whale Optimization Algorithm terms as WOA is a new bio-inspired optimization algorithm introduced by the researchers Mirjalili and Lewis in the year 2016 (Mirjalili & Lewis, 2016). It is based on the natural hunting behavior of humpback whales. The humpback whale can dive around 12m down a prey, then, creating bubbles net in a spiral shape and swim up towards the surface to hunt the prey. The humpback whale encircles the prey to preventing it from escaping. The pseudo code of WOA algorithm is shown in Figure 1, where, the code is divided into three phases; encircle prey, search for prey, and spiral bubble net (El Aziz et al., 2017; Mirjalili & Lewis, 2016; Mostafa et al., 2017).

In Encircle prey phase of WOA, humpback whale updates its position based on current best search agent (i.e. position of hunting prey). The other agents also update their position. This mechanism is represented mathematically in Equation 1 and 2 (Kaur & Arora, 2018; Mirjalili & Lewis, 2016).

$$\vec{D} = \left|\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)\right| \quad (1)$$

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{A}.\vec{D} \quad (2)$$

Where, $\vec{C}$ and $\vec{A}$ represent a coefficient vectors and they calculate using Equation 3 and 4 (Kaur & Arora, 2018), t is the current iteration, X^* indicates the best solution.

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \quad (3)$$

$$\vec{C} = 2.\vec{r} \quad (4)$$

Where, $\vec{a}$ is a value minimizing from 2 to 0, $\vec{r}$ is a random value between [0, 1].

In search for prey phase of WOA, humpback whale updates its position based on random position vector $\overrightarrow{X_{rand}}$ (i.e. random whale) that is chosen from the population. This phase is represented in Equation 5 and 6 (Kaur & Arora, 2018).

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_{rand}} - \vec{X}(t)\right| \quad (5)$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A}.\vec{D} \quad (6)$$

In the spiral bubble net phase of WOA, humpback whale mimics helix shaped movement when it changes its position towards the prey. The Equations 7 and 8 (Kaur & Arora, 2018) represent this action.

$$\overrightarrow{D'} = \left|\overrightarrow{X^*}(t) - \vec{X}(t)\right| \quad (7)$$

$$\vec{X}(t+1) = \overrightarrow{D'}.e^{bl}.\cos(2\pi l) + \overrightarrow{X^*}(t) \quad (8)$$

Where, b is a constant value that represents the logarithmic spiral shape, l represents a random value between [-1,1]. The p value in the code represents a random value between [0,1].

Initialization

- Randomly initializing the whale population Xi, where i=1,2,...,n
- Evaluating each search agent and calculate the fitness of them.
- Find best agent fitness X*.

Process

- While iteration < Maximum iteration
- For each search agent
- Updating a, A, C, I, and P by using Equations 3 and 4
- If (p<0.5)
- If (abs(A)<1)

Encircle Prey

- Updating current agent position using Equations 1 and 2
- Else if (abs(A)>=1)

Search for Prey

- Selecting random agent X_{random}.
- Updating current agent position using Equations 5 and 6
- End if

Bubble net

- Else if(p>=0.5)
- Updating current agent position using Equation 7 and 8
- End if
- End For
- calculating the fitness of each search agent.
- Updating best agent fitness X*.
- Iteration=iteration+1
- End while
- Return best solution X*.

Figure 1. Pseudo code of Original Whale Optimization Algorithm (Kaur & Arora, 2018)

MATERIALS AND METHODS

The flowchart of proposed Whale Color Image based Segmentation method can be seen in Figure 2. In proposed WhCIbS method, some changes had conducted to the original WOA to make it suitable for working on clustering. The goal of proposed WhCIbS is to segment the color image into clusters using the whale optimization algorithm in clustering. The input to this method is color image in RGB color space and the output is the optimal centers that return optimal clusters.

The input image to the Whale clustering algorithm is converted into L*a*b color space. In the initialization of the Whale clustering algorithm, the number of K clusters is pre-defined, and the number of Whale population n is determined. Each whale holds a solution X, where it represents a set of centroids that generated randomly. Then the clusters are constructed based on these centers by assigning a data value into clusters. The creating clusters are based on calculating the mean values of the attributes of the objects within their given clusters.

For each whale, the fitness value is computed based on Equation 9 (Tang et al., 2012). After that, the best whale is selected that has minimum fitness value and its corresponding position X*. Where, the data point belong to i clusters has minimum distance from i center.

$$F(C) = \sum_{i=1}^{k} \sum_{j=1}^{S} w_{j,i} \sum_{o=1}^{k*D} \left(X_{j,o} - C_{i,o}\right)^2 \quad (9)$$

Where, C is array containing cluster centers, k is the number of clusters, s is all solutions of X, w is set to 1 if x belongs to the cluster, otherwise w is set to 0, D is the search space dimension.

The Whale attacks their prey in either encircling or bubble-net strategies. The attacking is based on P value that represents a random number using for determining the probability of selecting one of two strategies. If the P is less than 0.5 the encircling step starts otherwise the bubble-net step starts.

In encircling step, the position of the whale is updated based on parameter A that can be calculated using Equation 3. If the value of A less than 1 the Whale updates the position based on best position X*. Otherwise, the Whale updates their position based on choosing a random search agent. This process is depending on Equation 5 and 6.

If the value of P is greater or equal 0.5 then the bubble-net strategy using Equation 7 and 8 is conducted to update the positions. The process of calculating the fitness value of whales and updating the best agent fitness continues until reaching the maximum number of iterations that is determined initially in the starting of the algorithm.

The output of the Whale clustering algorithm is optimal clusters to segment the color image. Then, the measuring segmented images using performance metrics are conducted. The performance of metrics equations is shown in the following results and discussion section.

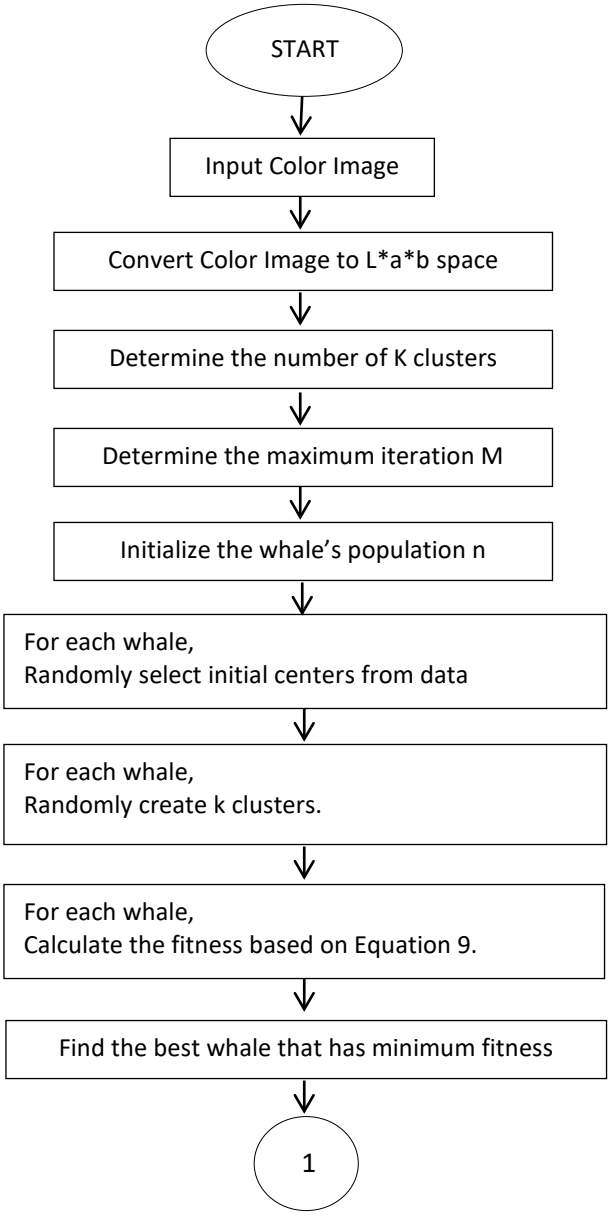


Figure 2. The flowchart of proposed Whale Color Image based Segmentation method

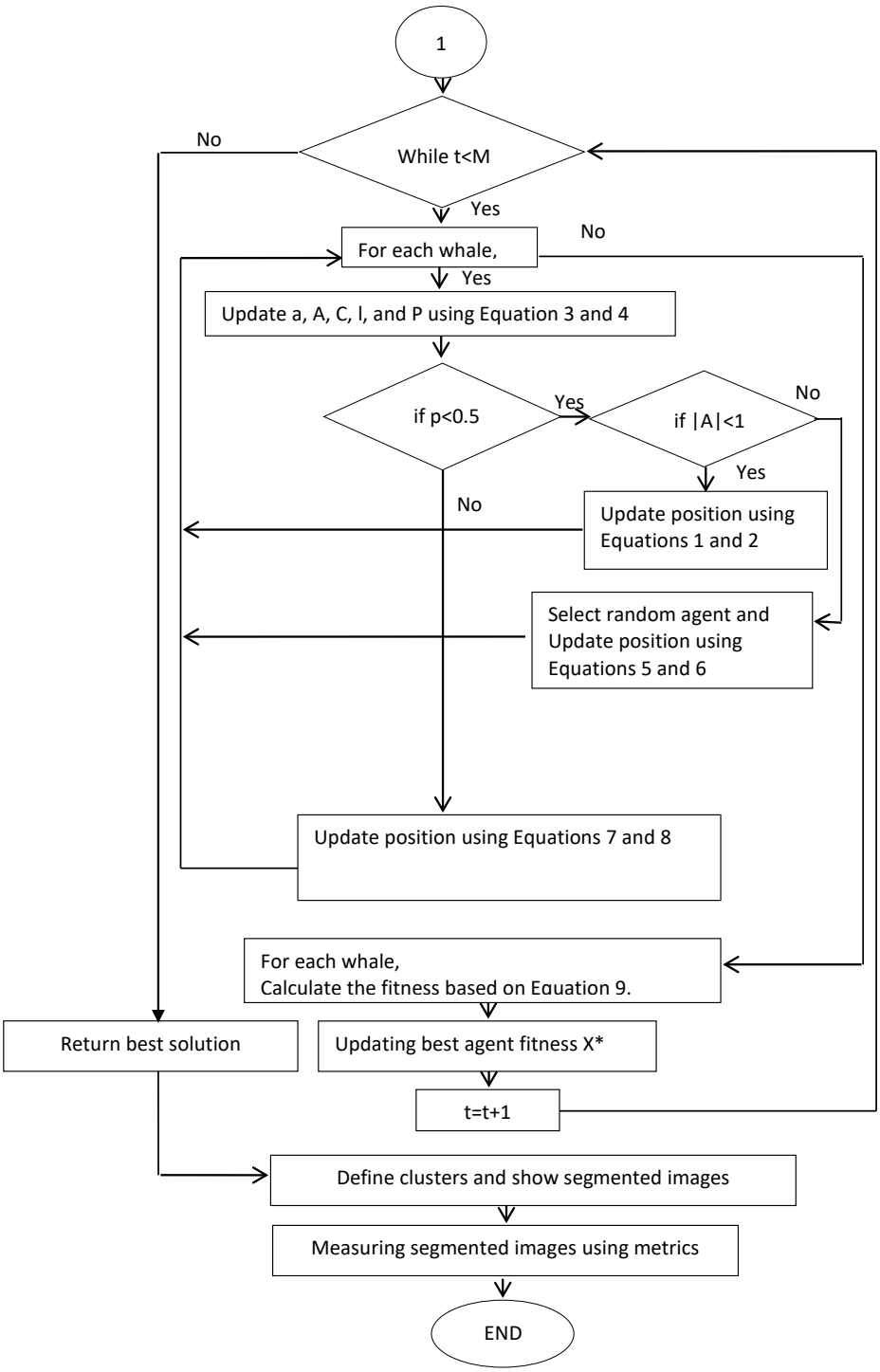


Figure 2. (Continued)

RESULTS AND DISCUSIONS

In this section, the environment of the experiments for the proposed Whale Color Image based Segmentation method is introduced. Firstly, the description of benchmark images is illustrated. Then, the parameters setup for proposed and comparative methods is briefly explained. Then the performance metrics for measuring the segmentation quality are described, followed by the experimental results.

Benchmark Images

In this paper, four test color images are utilized to test the performance of the proposed WhCIbS method. These images were collected from Berkeley Segmentation Dataset: Training Image (https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS300/html/dataset/images/color/train-076-100.html). The original test color images were named in the experiment; Deer, Dog, Lake, and Bear and they are shown in Figure 3.

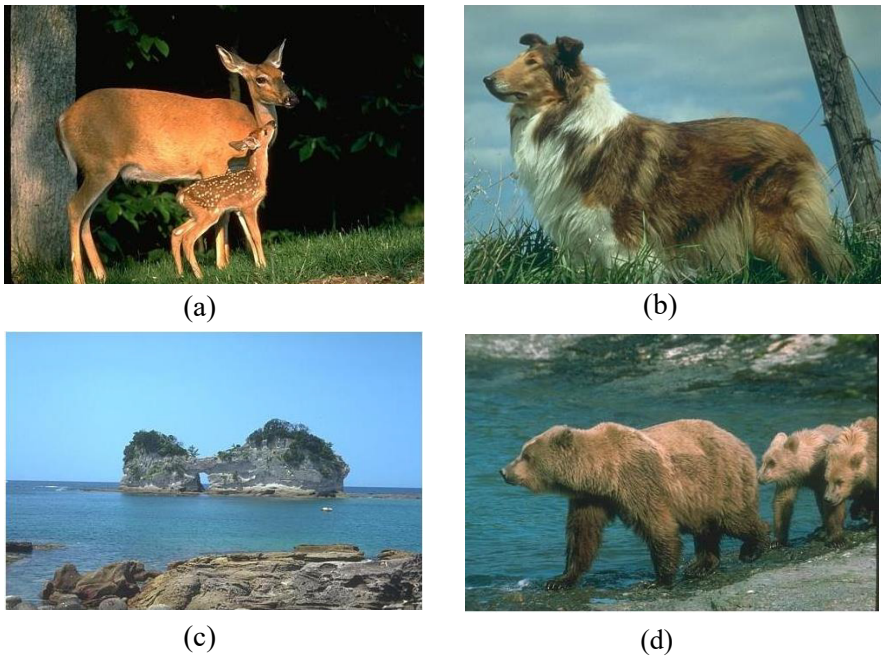


Figure 3. Original test image (a) Deer (b) Dog (c) Lake (d) Bear.

Parameters Setup

The proposed Whale Color Image based Segmentation method WhCIbS was preformed in MATLAB version 2018 in 64-bit windows 8 operating system with Intel Core i3 Duo 1.80 GHz processor. The results of proposed WhCIbS were compared with four methods; Wolf Color Image based Segmentation (WCIbS), Cuckoo Color Image based Segmentation

(CCIbS), BatColor Image based Segmentation (BCIbS) and K-means Color Image based Segmentation (KmCIbS) (Mathur & Purohit, 2014). For a fair comparison all methods had similar number of iterations, similar size of population and all methods testing on same test images. The parameters setup of the experiment for intelligent inspired optimization algorithms; Wolf, Cuckoo Search, Bat, and Whale algorithms are shown in Table 1.

Table 1

Parameters of the Wolf, Cuckoo, Bat, and Whale algorithms

Parameters	Wolf algorithm	Cuckoo algorithm	Bat algorithm	Whale algorithm
Number of population	20	20	20	20
Visual distance	1	-	-	-
Escape distance	5	-	-	-
Escape probability	0.25	-	-	-
Discovery rate	-	0.25	-	-
β	-	1.5	-	-
A: loudness	-	-	0.5	-
R: pulse rate	-	-	0.5	-
Qmin: minimum frequency	-	-	0	-
Qmax: maximum frequency	-	-	0.2	-
a	-	-	-	[0-2]
b	-	-	-	1
l	-	-	-	[-1,1]

Performance Metrics

Two performance metrics were used for evaluating the quality of the segmented images by the proposed WhCIbS and the comparative methods as follows; the Peak Signal to Noise Ratio (PSNR) which it was utilized for measuring the difference between segmented and reference images. The higher value of PSNR is the better performance of segmentation (Dhanachandra et al., 2015; El Aziz et al., 2017). Equation 10 shows the PSNR metric.

$$PSNR = 10\log_{10}\frac{\max[g(x,y)]^2}{RMSE} \tag{10}$$

The second metric was the Root Mean Squared Error (RMSE). The lower value of RMSE means better performance of segmentation (Dhanachandra et al., 2015; El Aziz et al., 2017). Equation 11 shows the RMSE metric.

$$RMSE = \sqrt{\frac{\sum_0^{mx-1}\sum_0^{my-1}[g(x,y)]^2}{mx\ my\ \sum_0^{mx-1}\sum_0^{my-1}[g(x,y)-s(x,y)]^2}} \quad (11)$$

Where, g(x,y) represents the test image and s(x,y) represents the segmented image of size (mxXmy).

Experimental Results

The results of the proposed Whale Color Image based Segmentation method WhCIbS and the comparative methods are clarified in Table 2.

As can see in Table 2 the PSNR of the proposed WhCIbS is highlighted in Table 2 (36.030, 38.801, 38.762, 38.868) for all test images: deer, dog, lake, and bear respectively, which is is a better result in most cases than the comparative methods WCIbS, CCIbS, BCIbS, and KmCIbS. Higher value of PSNR means better performance of segmentation (Dhanachandra et al., 2015; El Aziz et al., 2017). Figure 4 to 7 present the comparative results of PSNR for five methods WCIbS, CCIbS, BCIbS, WhCIbS and KmCIbS for four test images.

Table 2

Results of PSNR and RMSE for five methods: WCIbS, CCIbS,BCIbS, WhCIbS and KmCIbS.

Image name	Methods	PSNR	RMSE
Deer	WCIbS	36.023	10.572
	CCIbS	36.012	10.576
	BCIbS	36.025	10.571
	WhCIbS	36.030	**10.569**
	KmCIbS	36.025	10.571
Dog	WCIbS	38.740	12.999
	CCIbS	38.713	13.012
	BCIbS	38.741	12.999
	WhCIbS	**38.801**	**12.969**
	KmCIbS	38.741	12.999
Lake	WCIbS	38.745	13.036
	CCIbS	38.739	13.038
	BCIbS	38.745	13.035
	WhCIbS	**38.762**	**13.027**
	KmCIbS	38.745	13.035

Table 2 *(continued)*

Image name	Methods	PSNR	RMSE
Bear	BCIbS	38.822	12.993
	WhCIbS	**38.868**	**12.969**
	KmCIbS	38.819	12.995

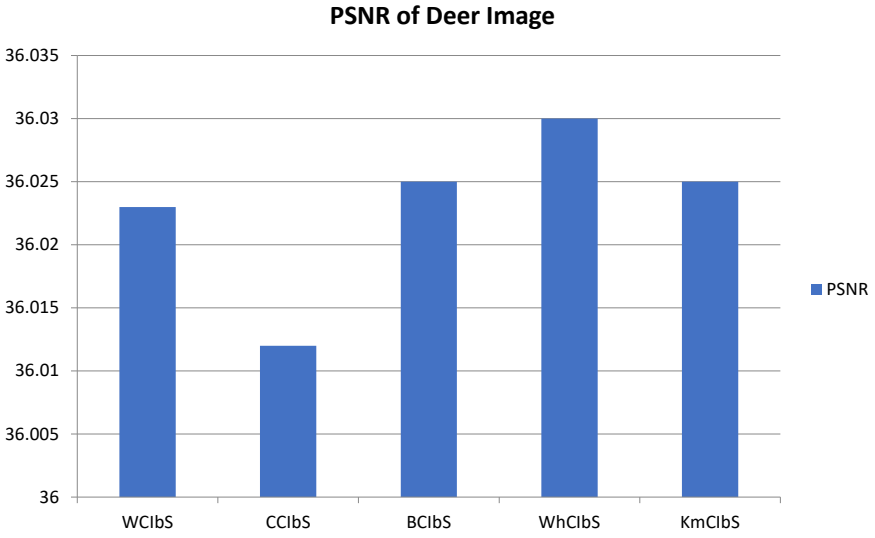


Figure 4. PSNR of five methods WCIbS, CCIbS,BCIbS, WhCIbS and KmCIbS for Deer image.

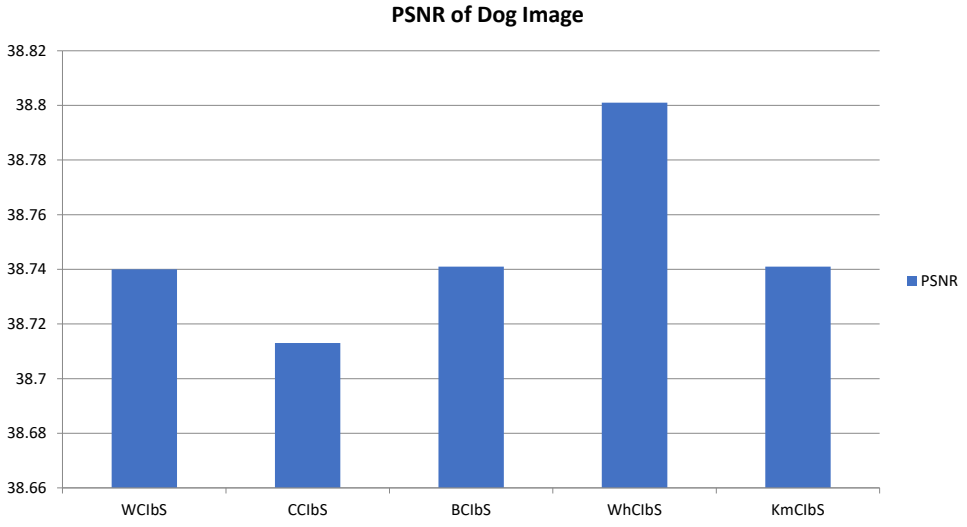


Figure 5. PSNR of five methods WCIbS, CCIbS,BCIbS, WhCIbS and KmCIbS for Dog image.

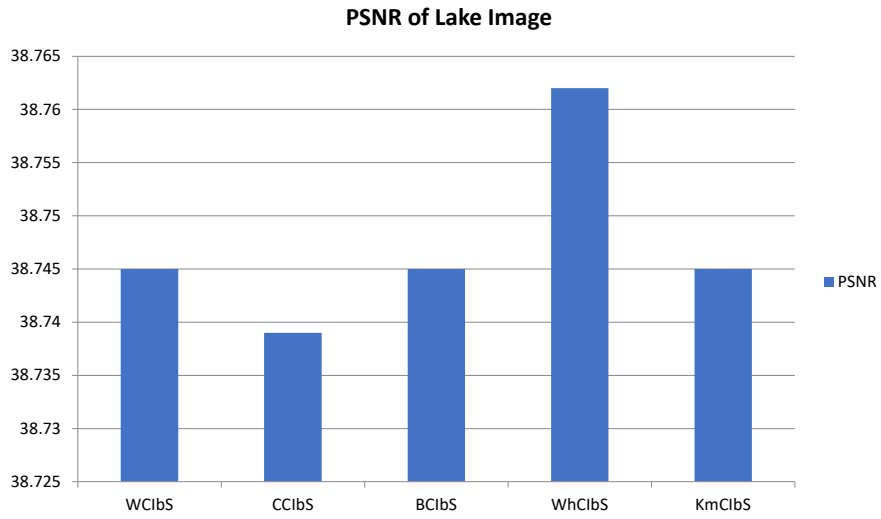


Figure 6. PSNR of five methods WCIbS, CCIbS,BCIbS, WhCIbS and KmCIbS for Lake image.

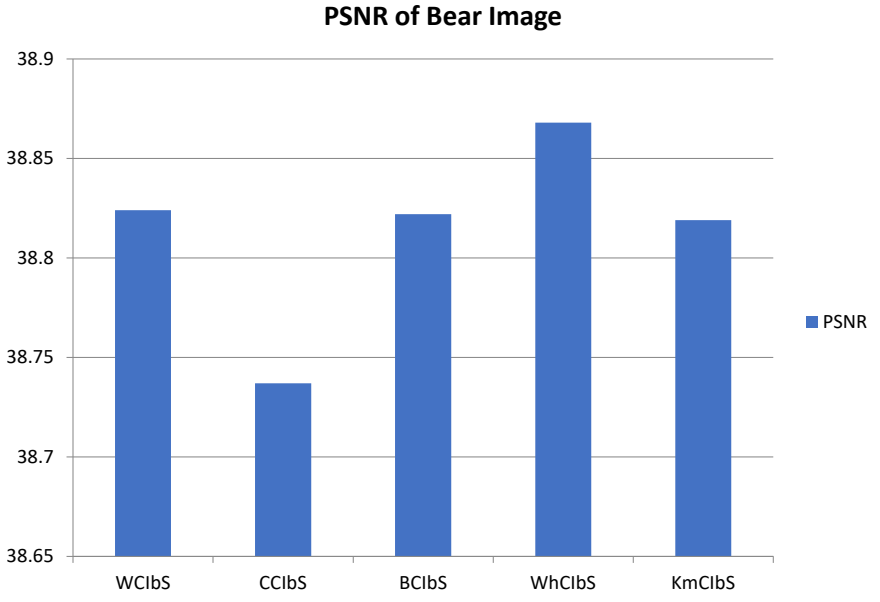


Figure 7. PSNR of five methods WCIbS, CCIbS,BCIbS, WhCIbS and KmCIbS for Bear image.

The RMSE for proposed WhCIbS and all comparative methods can be observed in Table 2. The highlighted values in Table 2, 10.569, 12.969, 13.027, 12.969 for proposed WhCIbS are minimum values of RMSE compared than other methods; WCIbS, CCIbS, BCIbS and KmCIbS. Lower value of RMSE means better quality of segmentation

(Dhanachandra et al., 2015; El Aziz et al., 2017). Figure 8 to 11 exhibit RMSE of five methods WCIbS, CCIbS, BCIbS, WhCIbS and KmCIbS for four test images.

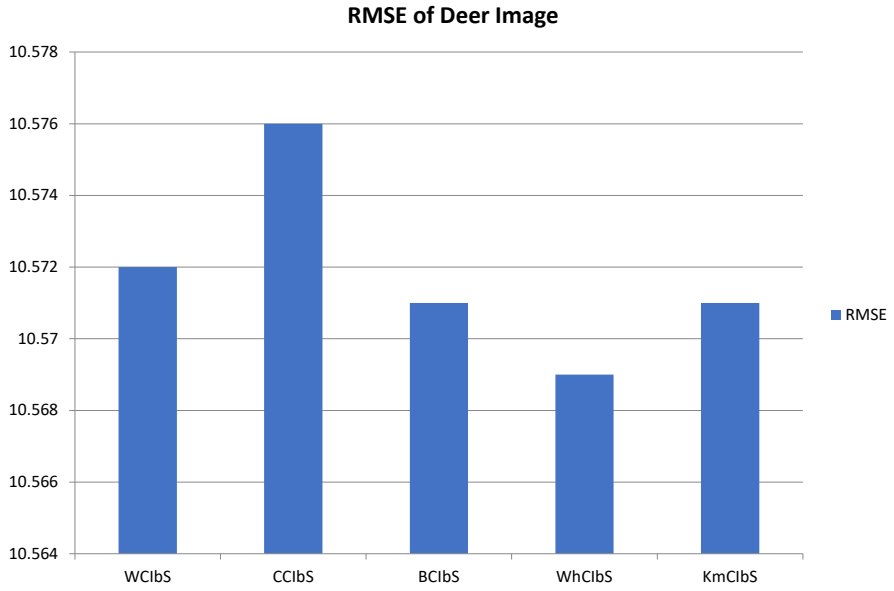


Figure 8. RMSE of five methods WCIbS, CCIbS, BCIbS, WhCIbS and KmCIbS for Deer image.

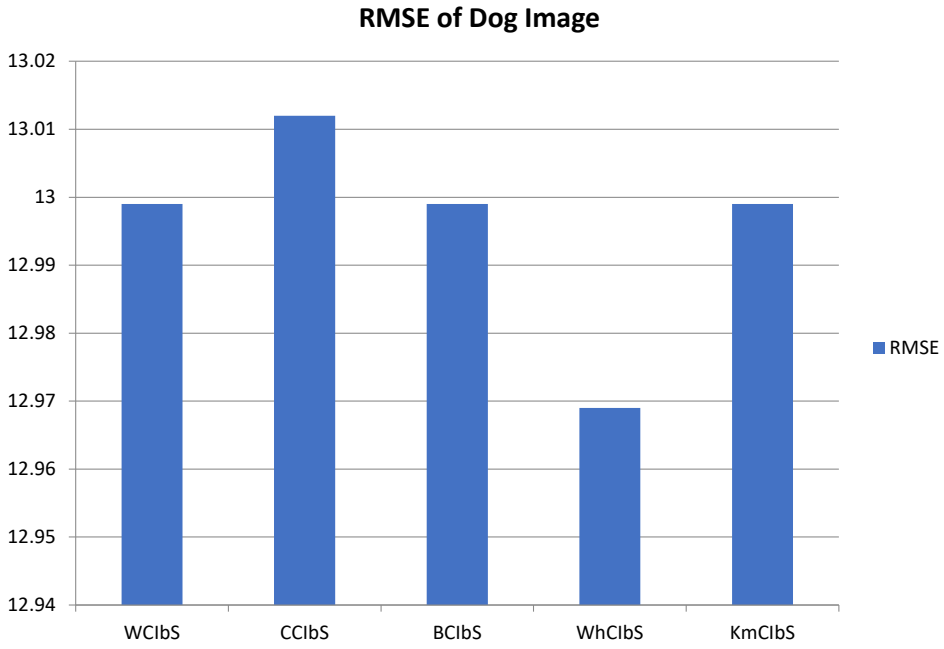


Figure 9. RMSE of five methods WCIbS, CCIbS, BCIbS, WhCIbS and KmCIbS for Dog image.

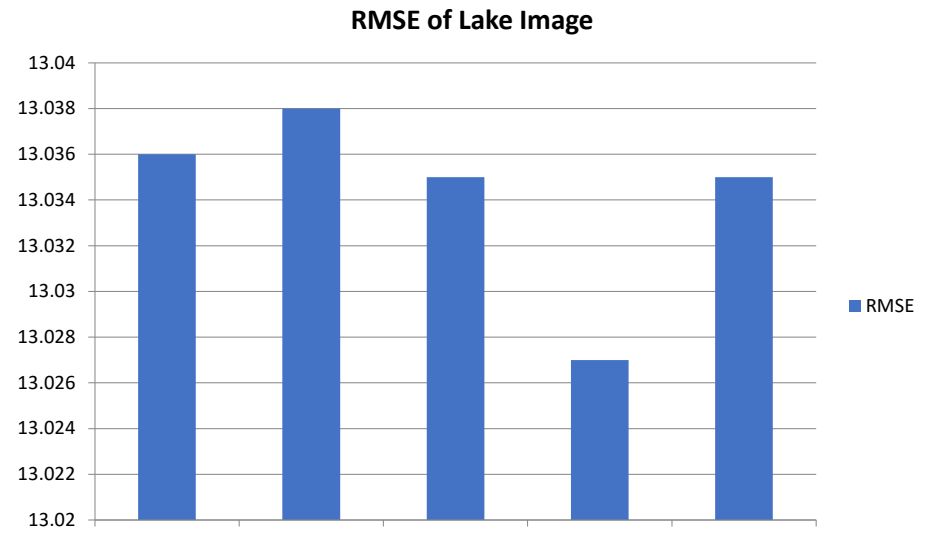


Figure 10. RMSE of five methods WCIbS, CCIbS, BCIbS, WhCIbS and KmCIbS for Lake image.

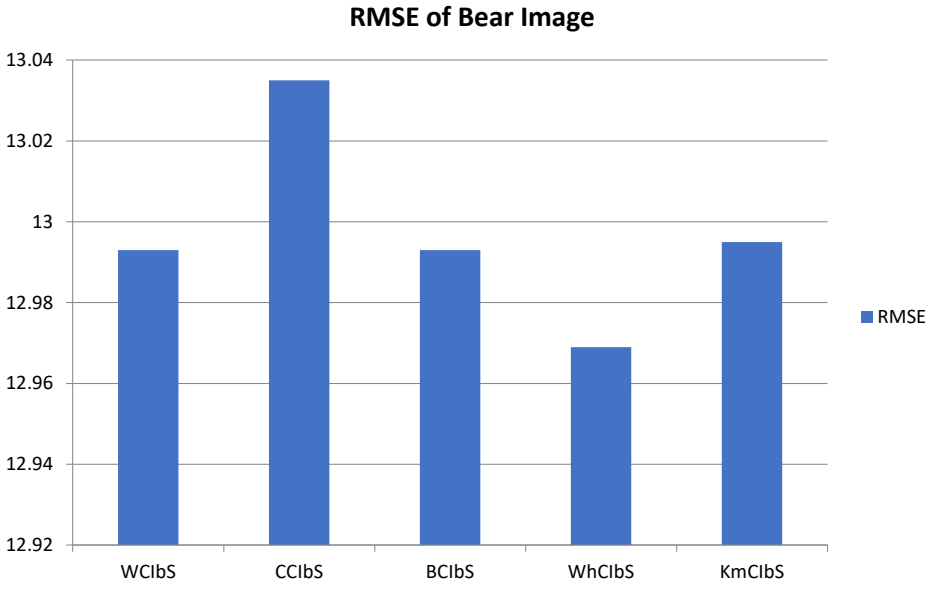


Figure 11. RMSE of five methods WCIbS, CCIbS, BCIbS, WhCIbS and KmCIbS for Bear image.

Figure 12 to 15 show color image segmentation for all methods. They indicate that WhCIbS is performed better whenever k=3.

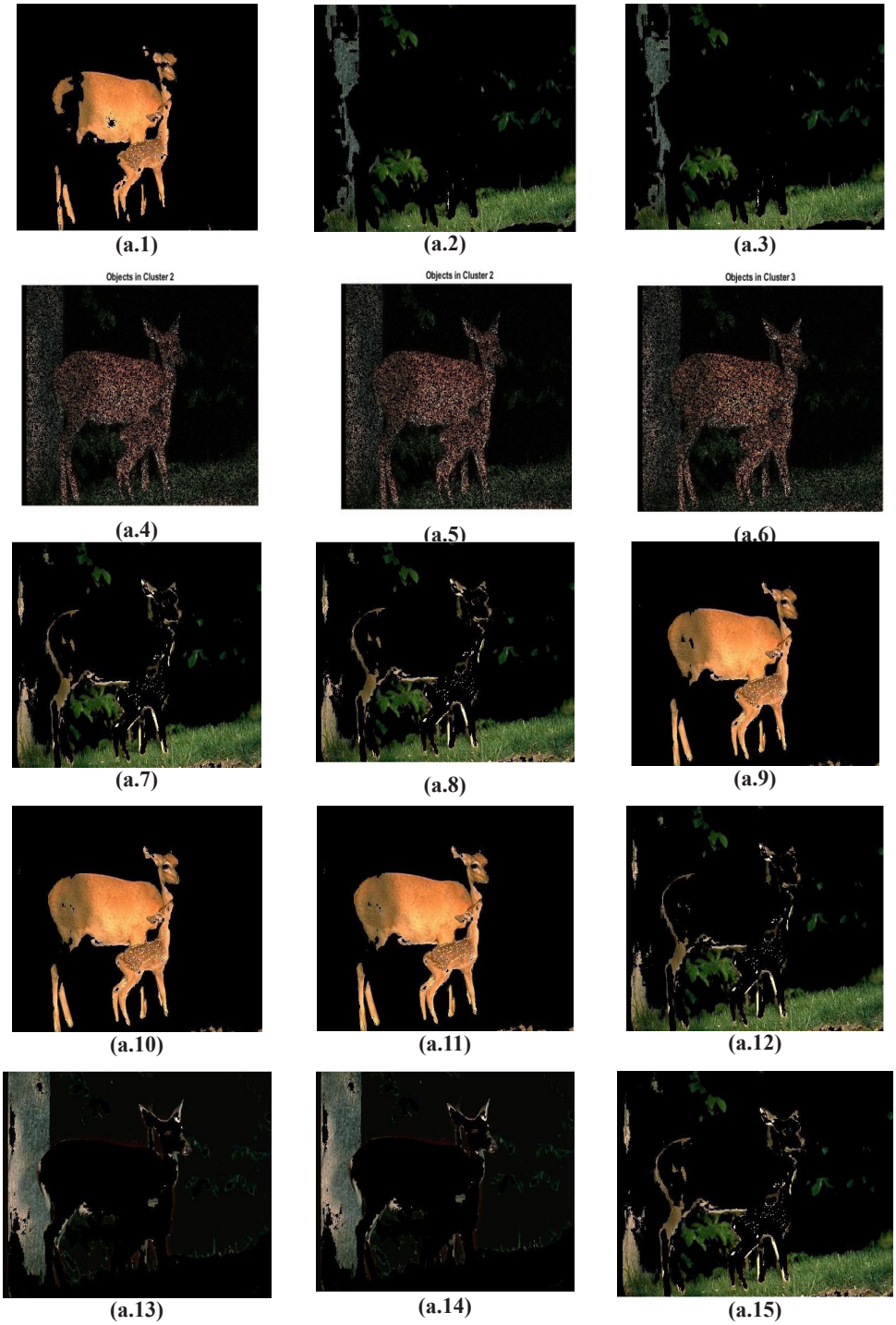


Figure 12. Color image segmentation[(a.1) (a.2) (a.3)] obtain by WCIbS method, [(a.4) (a.5) (a.6)] obtained by CCIbS method,[(a.7) (a.8) (a.9)] obtained by BCIbS method,[(a.10) (a.11) (a.12)] obtained by WhCIbS method, and [(a.13) (a.14) (a.15)] obtained by KmCIbS method.

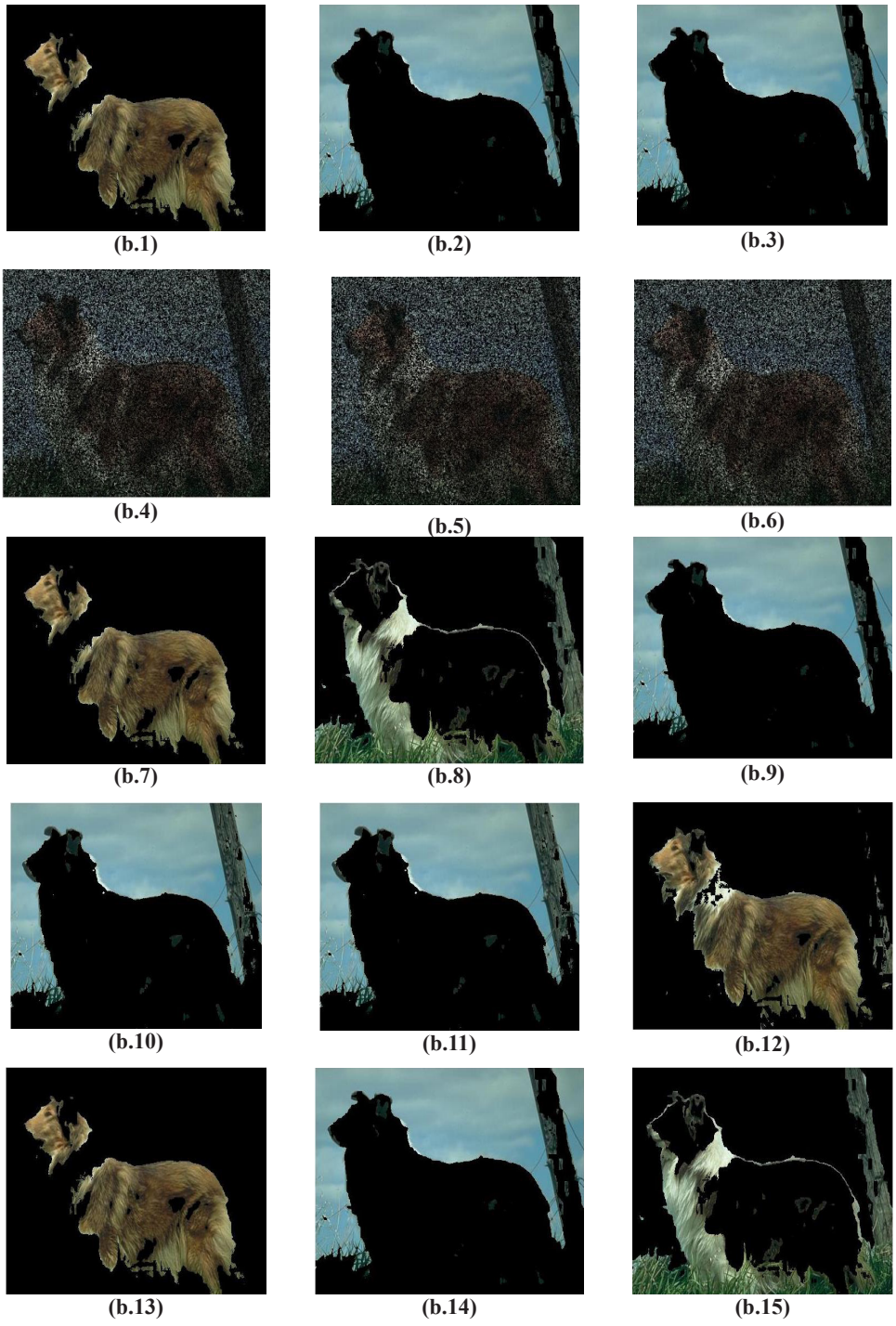


Figure 13. Color image segmentation[(b.1) (b.2) (b.3)] obtain by WCIbS method, [(b.4) (b.5) (b.6)] obtained by CCIbS method,[(b.7) (b.8) (b.9)] obtained by BCIbS method, [(b.10) (b.11) (b.12)] obtained by WhCIbS method, and [(b.13) (b.14) (b.15)] obtained by KmCIbS method.

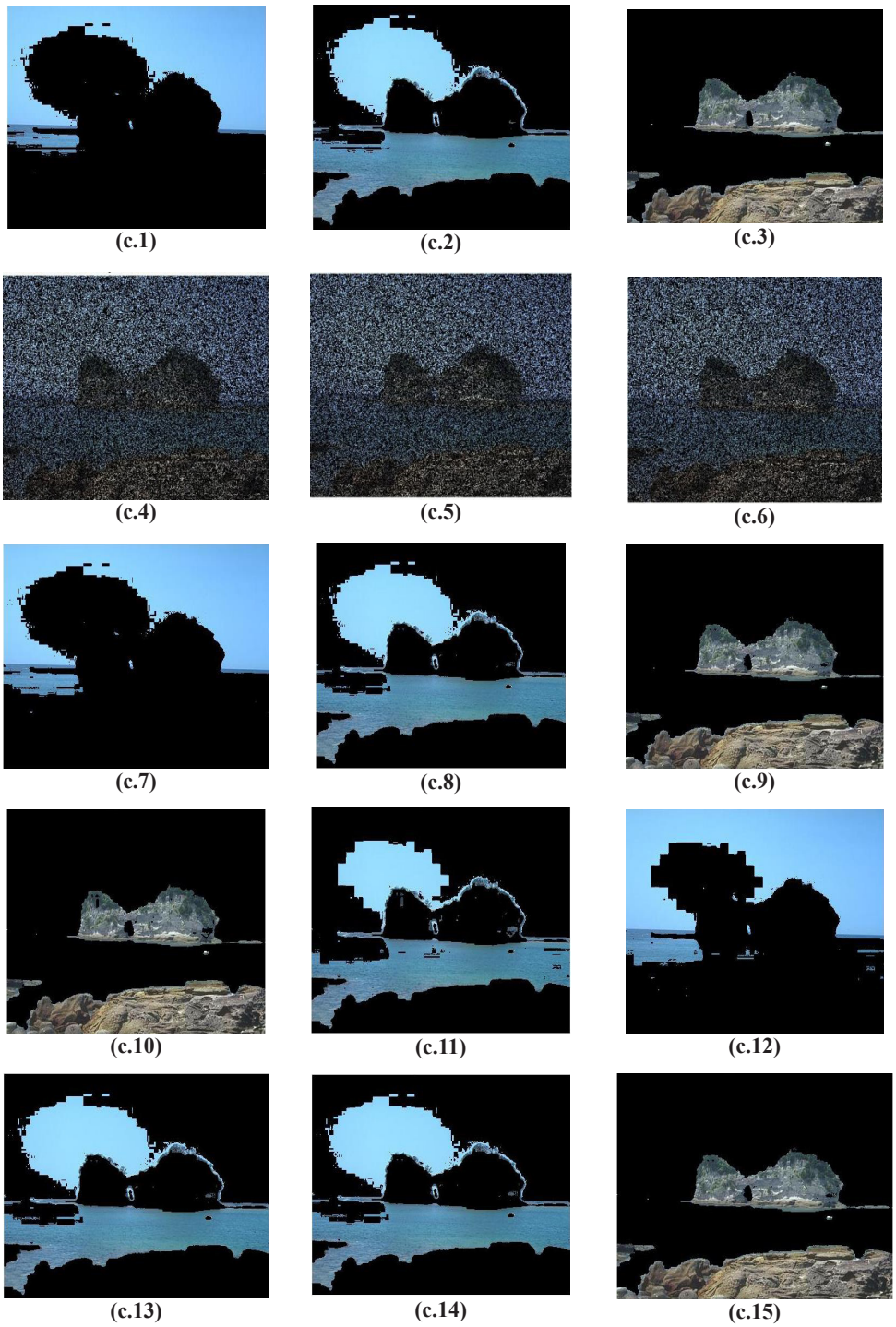


Figure 14. Color image segmentation[(c.1) (c.2) (c.3)] obtain by WCIbS method, [(c.4) (c.5) (c.6)] obtained by CCIbS method,[(c.7) (c.8) (c.9)] obtained by BCIbS method,[(c.10) (c.11) (c.12)] obtained by WhCIbS method, and [(c.13) (c.14) (c.15)] obtained by KmCIbS method.

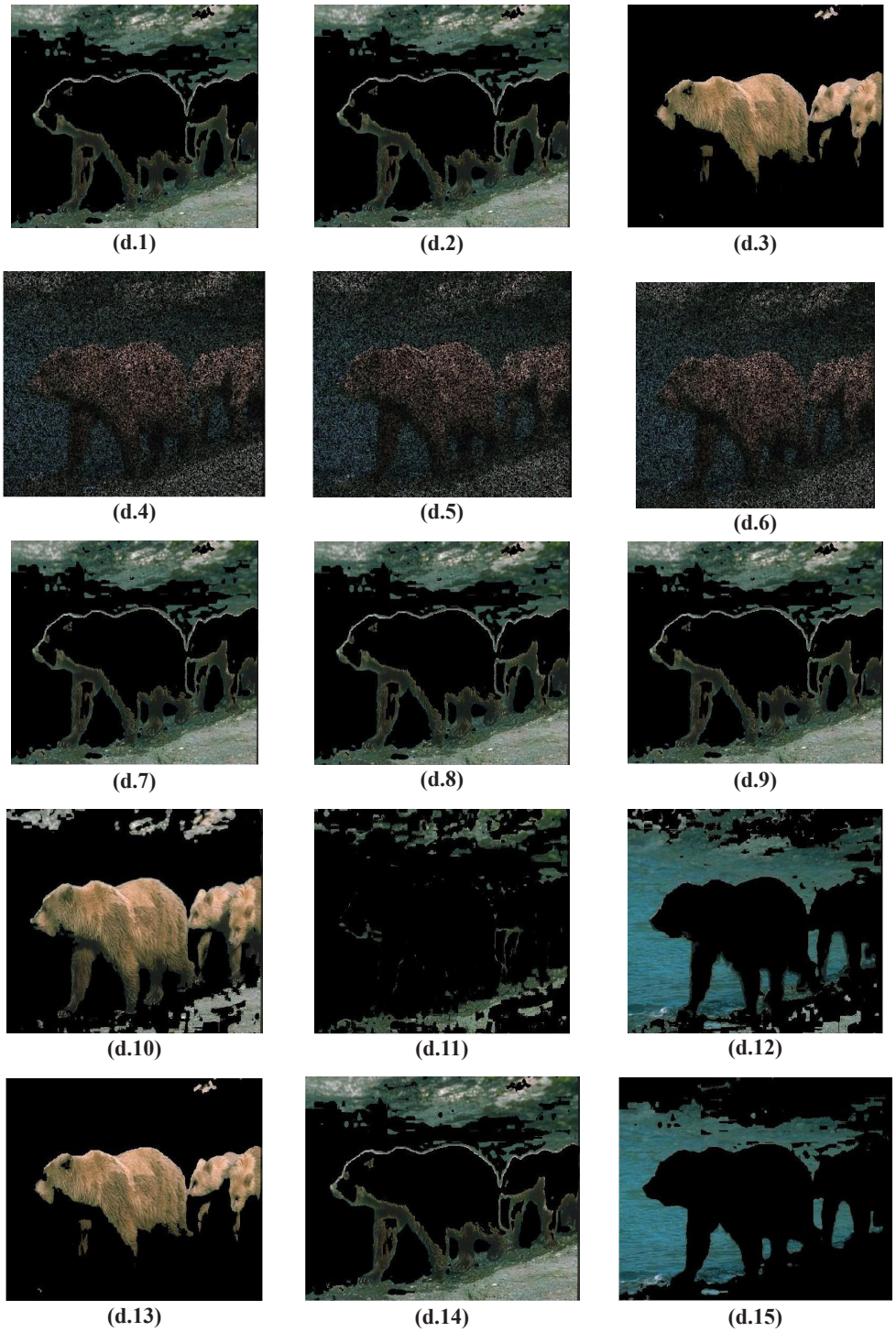


Figure 15. Color image segmentation[(d.1) (d.2) (d.3)] obtain by WCIbS method, [(d.4) (d.5) (d.6)] obtained by CCIbS method,[(d.7) (d.8) (d.9)] obtained by BCIbS method, [(d.10) (d.11) (d.12)] obtained by WhCIbS method, and [(d.13) (d.14) (d.15)] obtained by KmCIbS method.

CONCLUSIONS

Color image segmentation using a clustering technique is still an open problem. The determining of the optimal clusters has been considered as an optimization problem. Hence, a new swarm algorithm called the Whale optimization algorithm has been proposed to find a sloution for this problem using Euclidean distance as objective function. The goal of this algorithm is to obtain the best clusters for input color image. The proposed Whale Color Image based Segmentation (WhCIbS) method has been compared with three swarm algorithms; wolf color image based segmentation, cuckoo color image based segmentation, and bat color image based segmentation. Further, it is compared with k-means color image based segmentation. The performance of the proposed method and the comparativemethods has been evaluated based on two quality metrics; PSNR and RMSE. Experiments have been used four benchmark images that collected from the Berkeley segmentation dataset. The results demonstrate that the proposed method achieved better performance than the comparative methods in terms of PSNR and RMSE. In the future, we will study the performance of the proposed method with multi objective functions in order to obtain better quality outcomes for image segmentation.

ACKNOWLEDGEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Alagarsamy, S., Kamatchi, K., Govindaraj, V., Zhang, Y. D., & Thiyagarajan, A. (2019). Multi-channeled MR brain image segmentation: A new automated approach combining BAT and clustering technique for better identification of heterogeneous tumors. *Biocybernetics and Biomedical Engineering*, *39*(4), 1005-1035.

Dhanachandra, N., Manglem, K., & Chanu, Y. J. (2015). Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, *54*, 764-771.

El Aziz, M. A., Ewees, A. A., & Hassanien, A. E. (2017). Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation. *Expert Systems with Applications*, *83*, 242-256.

El Aziz, M. A., Ewees, A. A., Hassanien, A. E., Mudhsh, M., & Xiong, S. (2018). Multi-objective whale optimization algorithm for multilevel thresholding segmentation. In *Advances in soft computing and machine learning in image processing* (pp. 23-39). Cham, Switzerland: Springer.

Ghany, K. K. A., AbdelAziz, A. M., Soliman, T. H. A., & Sewisy, A. A. E. M. (2020, In Press). A hybrid modified step whale optimization algorithm with tabu search for data clustering. *Journal of King Saud University - Computer and Information Sciences*. DOI: https://doi.org/10.1016/j.jksuci.2020.01.015

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2003). *Digital image processing using MATLAB*. Upper Saddle River, USA: Prentice-Hall, Inc.

Jadhav, A. N., & Gomathi, N. (2018). WGC: Hybridization of exponential grey wolf optimizer with whale optimization for data clustering. *Alexandria Engineering Journal*, *57*(3), 1569-1584.

Juliet, S. E., Sadasivam, V., & Florinabel, D. J. (2014). Effective layer-based segmentation of compound images using morphology. *Journal of Real-Time Image Processing*, *9*(2), 299-314.

Kapoor, S., Zeya, I., Singhal, C., & Nanda, S. J. (2017). A grey wolf optimizer based automatic clustering algorithm for satellite image segmentation. *Procedia Computer Science*, *115*, 415-422.

Kaur, G., & Arora, S. (2018). Chaotic whale optimization algorithm. *Journal of Computational Design and Engineering*, *5*(3), 275-284.

Khan, A. M., & Ravi, S. (2013). Image segmentation methods: A comparative study. *International Journal of Soft Computing and Engineering, 3*(4), 84-92.

Mafarja, M. M., & Mirjalili, S. (2017). Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, *260*, 302-312.

Mathur, G., & Purohit, H. (2014). Performance analysis of color image segmentation using k-means clustering algorithm in different color spaces. *IOSR Journal of VLSI and Signal Processing*, *4*, 1-4.

Mirjalili, S., & Lewis, A. (2016). The Whale Optimization Algorithm. *Advances in Engineering Software*, *95*, 51-67.

Mishra, S., & Panda, M. (2018). Bat algorithm for multilevel colour image segmentation using entropy-based thresholding. *Arabian Journal for Science and Engineering*, *43*(12), 7285-7314.

Mostafa, A., Hassanien, A. E., Houseni, M., & Hefny, H. (2017). Liver segmentation in MRI images based on whale optimization algorithm. *Multimedia Tools and Applications*, *76*(23), 24931-24954.

Nandy, S., Yang, X., Sarkar, P. P., & Das, A. (2015). Color image segmentation by cuckoo search. *Intelligent Automation and Soft Computing*, *21*(4), 673-685.

Nasiri, J., & Khiyabani, F. M. (2018). A whale optimization algorithm (WOA) approach for clustering. *Cogent Mathematics and Statistics*, *5*(1), 1-13.

Ning, J., Zhang, L., Zhang, D., & Wu, C. (2010). Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition*, *43*(2), 445-456.

Sarkar, S., Patra, G. R., & Das, S. (2011). A differential evolution based approach for multilevel image segmentation using minimum cross entropy thresholding. In *International Conference on Swarm, Evolutionary, and Memetic Computing* (pp. 51-58). Heidelberg, Germany: Springer.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888-905.

Tang, J. (2010, April 16-18). A color image segmentation algorithm based on region growing. In *2010 2nd International Conference on Computer Engineering and Technology* (Vol. 6, pp. V6-634). Chengdu, China.

Tang, R., Fong, S., Yang, X. S., & Deb, S. (2012, August 22-24). Integrating nature-inspired optimization algorithms to K-means clustering. In *Seventh International Conference on Digital Information Management (ICDIM 2012)* (pp. 116-123). Macau, China.

Thennarasu, S. R., Selvam, M., & Srihari, K. (2020). A new whale optimizer for workflow scheduling in cloud computing environment. *Journal of Ambient Intelligence and Humanized Computing, 11*(5), 1-8.

Yang, X. S. (2010). A new metaheuristic bat-inspired algorithm. In *Nature inspired cooperative strategies for optimization (NICSO 2010)* (pp. 65-74). Heidelberg, Germany: Springer.

Ye, Q., Gao, W., & Zeng, W. (2003, July 6-9). Color image segmentation using density-based clustering. In *2003 International Conference on Multimedia and Expo. ICME'03. Proceedings (Cat. No. 03TH8698)* (Vol. 2, pp. II-401). Baltimore, MD, USA.

Zaitoun, N. M., & Aqel, M. J. (2015). Survey on image segmentation techniques. *Procedia Computer Science*, *65*, 797-806.

Zhang, Y., Huang, D., Ji, M., & Xie, F. (2011). Image segmentation using PSO and PCM with Mahalanobis distance. *Expert Systems with Applications*, *38*(7), 9036-9040.

Pertanika J. Sci. & Technol. 28 (4): 1413 - 1425 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Applying Genetic Algorithm to Optimization Second-order Bandpass MGMFB Filter

Maad Mohsin Mijwil* and Rana Ali Abttan

Computer Engineering Techniques Department, Baghdad College of Economics Sciences University, Baghdad Province, 11001 Iraq

ABSTRACT

In this paper, we have applied the genetic algorithm to the selection of the true values for RC (resistors/capacitors) as an essential role in the development of analogue active filters. The classic method of incorporating passive elements is a complex situation and can attend to errors. In order to reduce the frequency of errors and the human effort, evolutionary optimization methods are employed to select the RC values. In this study, Genetic algorithm (GA) is proposed to optimize the second-order active filter. It must find the values of the passive elements RC to get a filter configuration that reduces the sensitivities to variations as well as reduces design errors less than a defined height value, concerning certain specifications. The optimization problem which is one of the problems that must be solved by GA is a multi-objective optimization problem (MOOP). GA was carried out taking into account two possible situations about the values that resistors and capacitors could adopt. The obtained experimental results show that GA can be used to obtain filter configurations that meet the specified standard.

Keywords: Genetic algorithm, optimization, passive elements, second-order bandpass filter, sensitivities

ARTICLE INFO

Article history:
Received: 24 March 2020
Accepted: 10 July 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.15

E-mail addresses:
mr.maad.alnaimiy@baghdadcollege.edu.iq (Maad Mohsin Mijwil)
rana.ali.abttan@baghdadcollege.edu.iq (Rana Ali Abttan)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Active filters (AFs) are applied in high-frequency apps particularly in digital signal processing (DSP) circuits, communications, as well as health electronics (Soeiro et al., 2010; Shakoor et al., 2019). They can be executed in several ways, for example, using IC techniques, and requiring the presence of materials such as RC (Qazi & Mustafa, 2016). The characteristics of these filters depend on the passive elements because the selection of these values is considered very significant due to their influence on these filters (Kaya & Guler, 2018). There are various types of algorithms that are applied with AFs such as genetic algorithms (GAs), to find passive elements of RC (Benhala & Bouattane, 2014). Theses algorithms are an artificial intelligence techniques inspired by the idea that the individual that survives is the one that is best adapted to the environment (Bhasin & Bhatia, 2011). They are computational methods based on the theory of natural selection for the adaptive evolution of living things and have attracted attention as solutions to various optimization problems (Binitha & Sathya, 2012). Genetic algorithms are used successfully for a variety of problems that do not allow an efficient solution through the application of conventional techniques (Cassar et al., 2017). Besides, these algorithms permit a population made up of a specific set of individuals to develop according to designated rules, to minimize/maximize fitness (f) (Yan & Wang, 2010; Yang et al., 2014).

Recently, the modern electronic design includes AFs in many applications, such as conditioning and signal handling at audio and intermediate frequencies (IF) as well as DSP tasks (Li et al., 2018). In contrast to digital filters, assets can obtain excellent performance with significantly lower power demands (Vural et al., 2012). The active filter implementation alternatives present many options (He & Yin, 2018). The benefit of active filters is due essentially to integration ability and the extensive body of technical knowledge and they are very cheap, flexibility in gain and frequency adjustment, and no loading effect (Samadaei et al., 2011; Ibarra et al., 2011). Notwithstanding, the obtainable of active-filter ICs, most technicians and engineers are still resorting to RC active implementations consisting of operational amplifiers, RC (Corral, 2000; Zhang et al., 2020). Certainly, the prevalence of RC active filters has not declined since their bloom in the 1970s (Corral, 2000).

There are very few studies that apply the genetic algorithm to the second-order active filter. Starting with a study by Lovay et al. (2015), this study describes the application of the genetic algorithm to reduce sensitivities of second-order active filter called the SK (Sallen and Key) filter. The results show that configurations that meet established standards can be obtained. This study is considered closer to our study and its results have been compared to our results. In another study by Mostafa et al. (2018), eight components are applied as variables for improvement, as these components are compatible with the E96, E24, and E12 series by the genetic algorithm II (NSGA-II) with two analog filters which are a second order-filter and a fourth-order Butterworth with the operational amplifiers for testing

purposes. The results of this study have been compared with seeker optimization algorithm, evolutionary algorithms, differential evolution, artificial bee colony, and harmony search.

Moreover, the collection of discrete elements in RC active filter is an aspect of great quality since compliance with the specifications will depend largely on the true choice from users (El Beqal et al., 2019; Alkhazraji & Abttan, 2018). For reliable design, the values of the passive elements are selected from the E-series such as E12, E24, E48, & E96. Each of these series limits the values that passive elements can assume. Also, the approximation of the values of the elements achieved by those available in the E-series impacts the quality of the filters but also limits the scope of optimization (Makama et al., 2018).

The objective of this paper is to apply GA to optimize active filters taking into account their sensitivities (*Sens*) and to verify the feasibility of the proposed method, a second-order bandpass MGMFB (Multiple-Gain Multiple Feedback) filters is used. There are two possible situations about the values that resistors and capacitors can be used. For each situation, the algorithm must find the values of the elements that provides the established design requirements; and that reduces the sensitivities of the features of the filter concerning the values of the elements.

MATERIALS AND METHODS

MGMFB filter

MGMFB filter is widely applied as a bandpass filter because it allows an easy and reliable band-pass implementation, particularly below a quality factor (Q_p) of 20 or so (Zumbahlen, 2012). Figure 1 shows the second-order bandpass MGMFB filter. The major benefit of this filter is that without any modifying in the maximum gain at the centre frequency we can modify the value of the cut-off frequency. This modifies in the cut-off frequency can be done by the R2.

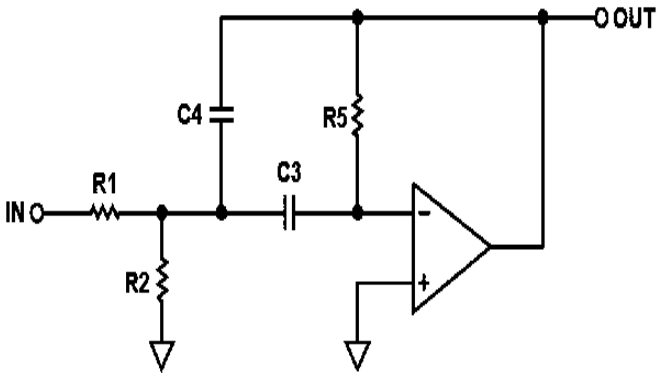


Figure 1. The MGMFB filter

The circuit in Figure 1 produces an infinity MGMFB filter. Due to this circuit the Q_p value increases maximum up to 20.

The transfer function (TF) of a second-order bandpass filter, presented in the Laplace transform (LT) as in Equation 1.

$$F(s) = \frac{-\frac{1}{R1C3}s}{s^2 + s\left(\frac{1}{R5C3} + \frac{1}{R5C4}\right) + \frac{1}{R5C4} + \left(\frac{1}{R1} + \frac{1}{R2}\right)} \quad (1)$$

This expression can be formulated based on the gain in the bandpass (H), the pole frequency $(a_p = 4\pi f_q)$ and the Q_p, ordinarily referred to as filter specifications. The TF is presented in Equation 2.

$$F(s) = \frac{1}{2\pi}\sqrt{\frac{Ha_p^2}{s^2 + \left(\frac{a_p}{Q_p}\right)s + a_p^2}} \quad (2)$$

For MGMFB filters, the values of , a_p, and Q_p can be calculated from the values of the passive elements, according to Equation 3-5.

$$H = \frac{R5 - C3}{R5C4 + C3} \quad (3)$$

$$a_p = \sqrt{\frac{2\pi}{R2 * R1 * C4 * C3}} \quad (4)$$

$$Q_p = \sqrt{\frac{C4}{C3}} + \frac{\sqrt{R1R5}}{0.5(R1)} + \sqrt{\frac{R5}{R2}} + \frac{1}{2}\sqrt{\frac{R2}{R5}} \quad (5)$$

Besides, MGMFB filters are less sensitive than other biquadratic filters. The less sensitive a filter to changes in its elements, the more stable the rest of its characteristics. Therefore, they are more likely to rest within its specifications, despite the presence of such variations.

Furthermore, if f is a function of many variables, $f = (x_1, x_2, \ldots, x_n)$, then the sensitivity of f concerning x_i is defined by Equation 6:

$$S_{x_i}^f = \frac{\%\ change\ in\ f}{\%\ change\ in\ x_i} = \frac{\partial f/f}{\partial x_i/x_i} \quad (6)$$

A filter is considered to have low sensitivity when all its *Sens* use values lower than unity. For MGMFB bandpass filters, the sensitivities of Q_p and a_p concerning each of the passive elements are (Equation 7-12):

$$S_{R1}^{a_p} = 0 \quad (7)$$

$$S_{R2}^{a_p} = S_{R5}^{a_p} = S_{C3}^{a_p} = S_{C4}^{a_p} = \left|\left(\frac{1}{2}\right)\right| \quad (8)$$

$$S_{C3}^{Q_p} = S_{C4}^{Q_p} = \left|\left(\frac{1}{2}\right)\right| \quad (9)$$

$$S_{R1}^{Q_p} = Q_p\left(\frac{1}{0.5(R1)}\sqrt{\frac{R2R5C4}{C3}}\right) \quad (10)$$

$$S_{R5}^{Q_p} = \frac{Q_p}{4}\left(\frac{1}{0.5(R5)}\sqrt{\frac{R2R3C4}{C3}} - \sqrt{\frac{R5C4}{R2C3}} + \sqrt{\frac{R2C4}{R5C3}}\right) \quad (11)$$

$$S_{R2}^{Q_p} = \frac{Q_p}{4}\left(\frac{2}{R_1}\sqrt{\frac{R2R5C4}{C4}} - \sqrt{\frac{R5C4}{R2C4}} + \sqrt{\frac{R5C3}{R2C3}}\right) \quad (12)$$

It is possible to estimate that in Equation 7-9 the sensitivities assume fixed values. In other words, the sensitivities specified in the remaining expressions indicate dependence on the values that the elements possess. For this purpose, the sensitivities expressed in Equation 10-12 must be recognized when choosing the values of both resistance and capacitor. In this paper, the specification chosen for the filter to be optimizing is (Equation 13-15):

$$H_f = 2 \quad (13)$$

$$a_{p_f} = \frac{797 * 2\pi * f}{2.5} rad/s \quad (14)$$

$$Q_{p_f} = 1.08. \quad (15)$$

Optimization by the Genetic Algorithm

This section describes the mechanism for using the GA to reduce the sensitivity of the filter, Figure. 2 presents a flow diagram of the GA.

Figure 2 presents the steps of the genetic algorithm, as follows:

1st step-Input samples: Insert all samples and parameters.

2nd step-Initialize population: This step includes the random generation of an initial population of individuals that are feasible solutions to the problem. In this study, every individual represents a filter configuration characterized by the values of RC. The GA describes every individual using a chromosome made up of 5 genes. Each gene represents the value of a resistor or a capacitor and is encoded using the entire encoding technique.

3rd step- Evaluation: In this step, every individual in the current population is estimated according to a predefined criterion, termed the fitness (f). In this study, the MOOP in Equation (19) is converted into an optimization problem with a single objective. For this, the weighted sum method is applied, which uses an aggregation function to carry out the

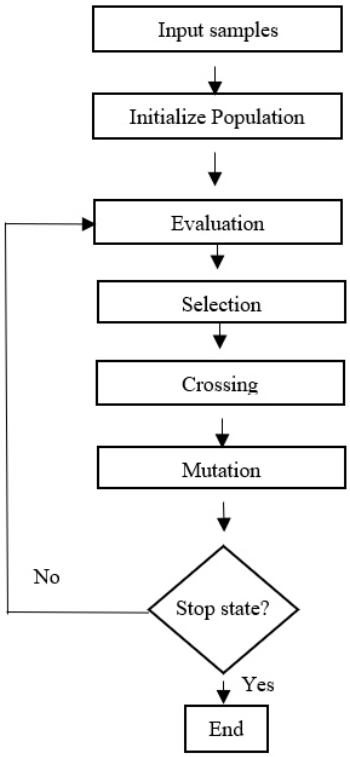


Figure 2. Flow diagram of GA steps

transformation. This is an easy way to get the best results. In this way, the f that assigns a fitness value to every individual (n), by Equation 16:

$$f(n) = (w_1 \left|S_{R1}^{Q_p}\right| + w_2 \left|S_{R2}^{Q_p}\right| + w_3 \left|S_{R5}^{Q_p}\right|) \quad (16)$$

In Equation 16, w_i denotes the weight assigned to each of the sensitivities total *(Senstotal)* that must be minimized. Because it is estimated that there is no preference between them, w_i adopts the same value for the 3 *Sens* (Equation 17-18).

$$f(n) = \frac{1}{3} Senstotal(n) \quad (17)$$

$$Senstotal(n) = (\left|S_{R1}^{Q_p}\right| + \left|S_{R2}^{Q_p}\right| + \left|S_{R5}^{Q_p}\right|) \quad (18)$$

In other words, the GA punishes those individuals that do not comply with them in every iteration. So, they adopt a higher fitness value according to the degree of violation of each restriction. This procedure is applied to Equation 16.

4th step –Selection: To produce a new population, the value of f is essential in selected individuals. For this, the GA employs the spinning roulette method as a selection method, which supposes that the individual's choice to crossing depends on the value of their f.

5th step- Crossing: The elected individuals are subjected to the action of the crossing operator, which enables the replacement of genetic elements between them, making it reasonable to create new individuals to replace their parents, forming the population of the next generation.

6th step- Mutation: The population of individuals resulting from the 4th step process is subjected to the mutation operator. It randomly affects an alteration in some of the genes on the chromosomes that make up the population, to avoid lack of diversity. For this purpose, the GA applies the consistent random mutation operator, which considers that each gene has the same probability of being mutated.

6th step- Stop state: The new generation passes through the process defined from the 2nd step to the 5th step process. The cycle is renewed till the stop state is fulfilled, including reaching a maximum number of generations.

Now, the GA find the passive elements values of the MGMFB (3 resistors & 2 capacitors) for which the sensitivities that rely on the values of the elements are minimum, and the errors in H, a_p and Q_p are less than a maximum error specified. The MOOP may be described by the Equation 19 and it is a possible filter configuration determined by the resistor and capacitor values.

$$min\, f(n) = S_{R1}^{Q_p}(n), S_{R2}^{Q_p}(n), S_{R5}^{Q_p}(n) \tag{19}$$

$$Subject\ to: Error_H(n) \leq E_{max}$$

$$Error_{a_p}(n) \leq E_{max}$$

$$Error_{Q_p}(n) \leq E_{max}$$

Where:

$S_{R1}^{Q_p}, S_{R2}^{Q_p}, S_{R5}^{Q_p}$ represents sensitivities that must be minimized, $Error_H$, $Error_{a_p}$ and $Error_{Q_p}$ represent the errors in H, a_p and Q_p, respectively and are calculated according to the Equation 20-22. Finally, E_{max} is described as the maximum common design error in the filter elements.

$$Error_H(n) = \left| \frac{H(n) - H_f}{H_f} \right| \tag{20}$$

$$Error_{a_p}\left(n = \left|\frac{a_p(n) - a_{p_f}}{a_{p_f}}\right|\right) \quad (21)$$

$$Error_{Q_p}(n) = \left|\frac{Q_p(n) - Q_{p_f}}{a_{p_f}}\right| \quad (22)$$

RESULTS AND DISCUSSION

The GA is aimed at optimization. The active filters are evaluated in this paper by taking into account the sensitivities and design errors obtained. For this, two possible situations are proposed to the values that the elements can adopt from the optimization process by GA. In situation 1, the capacitors and resistors can assume values according to the E24 and E96, respectively, considering that *Emax* = 5.00E-03 (0.5%). Moreover, situation 2, uses the E12 series for capacitor values and E24 series for resistance values, using the value 2.50E-02 (2.5%) for *Emax* . In both situations the range of defined values is 10^3- 10^6 Ω & 10^{-9} -10^{-6} F, respectively. The values of these ranges are assumed to have adverse impacts due to parasitic capabilities or very large modern signals. Therefore, the total search space amounts to 1.10E09 in situation 1 as well as 2.22E06 alternatives in situation 2.

Table 1 shows the *Senstotal* values optimized by the GA. In situation 1, the GA obtains an appropriate solution for population sizes from 20 to 70 individuals and 3 crossover probabilities estimated. *Senstotal* reaches the lowest value with a population of 60 individuals and crossover probability 0.9. In situation 2, the population sizes from 10 to 70, and 3 crossover probabilities are chosen. *Senstotal* reaches the lowest value with a population of 50 individuals and crossover probability 0.9 in situation 2.

Table 1

Senstotal values obtained for different GA parameter values

Situation 1				Situation 2			
Population	Crossover Probability			Population	Crossover Probability		
	0.7	0.8	0.9		0.7	0.8	0.9
20	0.8544	0.6790	0.7237	10	0.8076	0.6703	0.8329
30	0.6511	0.6725	0.9015	20	0.6852	1.1777	0.6541
40	0.8647	0.7327	0.6516	30	0.8740	0.6440	0.6792
50	1.2766	1.0110	1.0743	40	0.6690	0.8340	0.6542
60	0.6725	0.6725	0.6430	50	0.6540	0.6540	0.6203
70	0.6632	0.8329	1.0215	60	0.6540	0.6540	1.2046
				70	0.6540	0.6540	0.6518

Table 2 presents the results of the elements, their errors and the sensitivities achieved by the GA. In each situation, the described values correspond to the solution obtained by the genetic algorithm that has the lowest *Senstotal* value (Table 1). Besides, sensitivities reach similar values in both situations, with the exception of $S_{R_S}^{Q_p}$, where the result obtained from situation 1 is less than from the result obtained from situation 2. Nevertheless, *Senstotal* uses very similar values for the situations proposed. On the other hand, the specifications, (1) and (2) are applied to achieve the TF of a filter termed a NF (nominal filter). This filter does not consider into account the sensitivities, because it is not obtained using the values of its elements. But its frequency response (FR) is applied to graphically observe compliance with limitations by the solutions obtained by the GA. The FR of the NF and the response of the filter configuration are obtained by the algorithm in all situation.

Table 2
The result achieved by GA for each proposed situation

Parameter	Situation 1	Situation 2
*R*1	16600Ω	10900Ω
*R*2	36400Ω	21000Ω
*R*5	11000 Ω	6100 Ω
*C*3	1.21E-08 F	0.40E-08 F
*C*4	1.01E-09 F	1.21E-09 F
H	2.0042	1
*Q*p	0.7066	0.7151
a_p	5202.2015	4241.1778
$Error_H$	0.1622%	0
$ErrorQ_p$	0.0472%	1.0083%
$Error_{ap}$	0.1916%	1.3820%
$\left\|S_{R1}^{Q_p}\right\|$	0.2641	0.2529
$\left\|S_{R2}^{Q_p}\right\|$	0.2652	0.2564
$\left\|S_{R5}^{Q_p}\right\|$	0.0002	0.0011
Senstotal	0.6516	0.6521

The GA implements a stochastic process, the results of this process may be changed according to the statistical distribution of the initial population. To observe how they can be influenced by the establishment of the initial population, 70 runs are made for each situation, changing the speed in the random generation of the initial population of each run. In both situations, it is possible to note that the three sensitivities always use values lower than 1. Also, S_{R1}^{Qp} reaches the highest values and S_{R2}^{Qp} has the lowest values, while S_{R5}^{Qp} is the sensitivity with the lowest variations.

On the other hand, to complement the analysis, characterization of *Senstotal* and the errors received by the algorithm in the 70 runs performed are presented in Table 3. In it, the minimum, maximum and median values are displayed for each situation. This measure of central tendency is adopted because the data presents a non-normal distribution. In the same table, it is possible to appreciate that *Senstotal* varies in a slightly greater range in situation 2, concerning situation 1, presenting similar minimum values and a lower median in situation 2. As for errors, in both situations, the maximum error is less than E_{max} and minimum errors, in situation 1 the lowest value corresponds to $Error_{ap}$ and the highest one to $Error_H$, while in situation 2, $Error_H$ reaches zero and the remaining two errors adopt similar values. For the median, the highest value corresponds to $Error_G$ in both situation, while the lowest value is presented for $Error_{ap}$ in situation 1 and for $Error_{Qp}$ in situation 2.

Table 3
Senstotal characterization and the errors obtained by the GA

	Situation 1			Situation 2		
	Minimum	Maximum	Median	Minimum	Maximum	Median
Senstotal	0.6513	0.9322	0.7520	0.6521	2.4117	0.8710
ErrorH(%)	0.1682	0.3742	0.2409	0	1.2332	0.9223
$Error_{ap}$(%)	0.1269	0.4626	0.3721	0	2.4760	1.3539
$Error_{Qp}$ (%)	0.0219	0.4126	0.2579	0.2232	3.3228	2.1731
Execution times	43,8345 seconds			51,7782 second		
Estimated total time	472,100 seconds			2,330 seconds		

The GA gets an appropriate solution using considerably lower execution times than the exhaustive search method. Table 4 presents a comparison between the current results of this study and the previous study regarding the sensitivities (*Sens*) value of RC in both situations.

Table 4

Comparison between current study and previous study

	Current study			Previous study (Lovay et al., 2015)		
Situation 1	Minimum	Maximum	Median	Minimum	Maximum	Median
Senstotal	0.6513	0.9322	0.7520	1.994	2.222	1.964
Situation 2	Minimum	Maximum	Median	Minimum	Maximum	Median
Senstotal	0.6521	2.4117	0.8710	1.964	2.499	2.059

The above table shows the difference between these two studies despite the difference in the type of filter used, but the genetic algorithm gave better results in the current study in terms of reducing sensitivities of RC values.

CONCLUSIONS

Within this work, a genetic algorithm to optimize the second-order bandpass MGMFB filter has been applied. For the values that RC can use, two possible situations are suggested. In each situation, the algorithm must find the values of the passive elements of the filter in order to have the configuration as robust as possible to the variations in them, by reducing the sensitivity. In addition, it should also be borne in mind that design errors relating to certain specifications must be less than the maximum value. Finally, the simulation outcomes indicate that the genetic algorithm can obtain filter configurations matching the defined requirements. Moreover, the outcomes can provide several design alternatives because of their stochastic nature.

ACKNOWLEDGEMENT

We thank the anonymous editors and reviewers for their useful suggestions to develop this manuscript.

REFERENCES

Alkhazraji, A. A., & Abttan, R. A. (2018). Multiple fault detection and classification in power system. *Advances in Natural and Applied Sciences*, *12*(5), 1-13.

Benhala, B., & Bouattane, O. (2014). GA and ACO techniques for the analog circuits design optimization. *Journal of Theoretical and Applied Information Technology, 64*(2), 413-419.

Bhasin, H., & Bhatia, S. (2011). Application of genetic algorithms in machine learning. *International Journal of Computer Science and Information Technologies, 2*(5), 2412-2415.

Binitha, S., & Sathya, S. S. (2012). A survey of bio inspired optimization algorithms. *International Journal of Soft Computing and Engineering (IJSCE)*, *2*(2),137-151.

Cassar, I. R., Titus, N. D., & Grill, W. M. (2017). An improved genetic algorithm for designing optimal temporal patterns of neural stimulation. *Journal of Neural Engineering*, *14*(6), 1-32.

Corral, C. (2000). Designing RC active filters with standard-component values. *EDN, 45*(24), 141-149.

El Beqal, A., Benhala, B., Garbay, A., Kotti, M., Fakhfakh, M., & Zorkani, I. (2019, April 18-19). Synthesis of a current mode second order band pass filter using the genetic algorithm. In *Proceedings of 5th International Conference on Optimization and Applications (ICOA)* (pp. 1-9). Kenitra, Morocco.

He, J., & Yin, J. (2018, July 28-30). A practical evolution model for filter automatic design. In *Proceedings of International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 1-8). Huangshan, China.

Ibarra, F. S., Claros, M. C., Espinosa, R. M., Levy, E. O., & Betancourt, L. P. (2011). Design of 2nd order low-pass active filters by preserving the physical meaning of design variables. *Revista Mexicana de Física E, 57(1),1-10.*

Kaya, T., & Guler, H. (2018). A hybrid genetic algorithm for analog active filter component selection. *AEU - International Journal of Electronics and Communications, 86*, 1-7.

Li, S., Zou, W., & Hu, J. A. (2018). Novel evolutionary algorithm for designing robust analog filters. *Algorithms Journal*, *11*(3),1-22.

Lovay, M., Peretti, G., & Romero, E. (2015, July 30-31). Application of genetic algorithms in the design of robust active filters. In *Proceedings of International Conference for Argentine School of Micro-Nanoelectronics, Technology and Applications (EAMTA)* (pp. 01-06). Villa Maria, Argentina.

Makama, A., Binns, R., & Basha, M. (2018, July 14-16). Application of an infinite gain multiple feedback filter in VLC link range extension. In *Proceedings of IEEE 2nd International Conference on Circuits, System and Simulation (ICCSS)* (pp. 1-7). Guangzhou, China.

Mostafa S. S., Horta N., Garcíac A. G. R., & Dias F. M. (2018). Analog active filter design using a multi objective genetic algorithm. *AEÜ - International Journal of Electronics and Communications*, *93*(1), 93-94.

Qazi, S. H., & Mustafa, M. W. (2016). Review on active filters and its performance with grid connected fixed and variable speed wind turbine generator. *Renewable and Sustainable Energy Reviews, 57*(12), 420-438.

Samadaei, E., Lesan, S., & Cherati, S. M. (2011, April 18-19). A new schematic for hybrid active power filter controller. In *Proceedings of the 2011 IEEE Applied Power Electronics Colloquium (IAPEC) Malaysia* (pp. 143-148). Johor Bahru, Malaysia.

Shakoor, A., Abbas, S., & Abbas, Z. (2019). Optimization of design parameters of low pass filter using genetic algorithm. *UW Journal of Science and Technology*, *3*(1), 55-60.

Soeiro, T., Biela, J., Linnér, J., Ranstad, P., & Kolar, J. W. (2010, November 7-10). Line power quality improvement for ESP systems using multi-pulse and active filter concepts. In *Proceedings of International Conference on IEEE Industrial Electronics Society (IECON 2010)* (pp. 1-6). Glendale, USA.

Vural, R. A., Yıldırım, T., Kadioglu, T., & Basargan, A. (2012). Performance evaluation of evolutionary algorithms for optimal filter design. *IEEE Transactions on Evolutionary Computation*, *16*(1), 135-147.

Yan, X., & Wang, X. (2010). Fitness function of genetic algorithm in structural constraint optimization. In *Proceedings of International Conference on Advances in Swarm Intelligence (ICSI 2010)* (pp. 432-483). Berlin, Germany: Springer.

Yang, M. D., Yang, Y. F., Su, T.C., & Huang, K. S. (2014). An efficient fitness function in genetic algorithm classifier for land use recognition on satellite images. *The Scientific World Journal, 2014*(264512), 1-12.

Zhang, C., Shang, L., Wang, Y., & Tang, L. (2020). A CMOS programmable fourth-order butterworth active-RC low-pass filter. *Electronics Journal-MDPI*, *9*(2), 1-9.

Zumbahlen, H. (2012). *Multiple feedback band-pass design example*. Massachusetts, USA: Analog Devices, Inc.

Pertanika J. Sci. & Technol. 28 (4): 1427 - 1444 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

A Compression System for Unicode Files Using an Enhanced Lzw Method

Rincy Thayyalakkal Anto[1]* and Rajesh Ramachandran[2]

[1]Department of Computer Science, Prajyoti Niketan College, Pudukad, Thrissur, Kerala, 680301 India
[2] Department of Computer Science, CHRIST (Deemed to be University), Hosur Road, Banglore, 560029 India

ABSTRACT

Data compression plays a vital and pivotal role in the process of computing as it helps in space reduction occupied by a file as well as to reduce the time taken to access the file. This work relates to a method for compressing and decompressing a UTF-8 encoded stream of data pertaining to Lempel-Ziv-welch (LZW) method. It is worth to use an exclusive-purpose LZW compression scheme as many applications are utilizing Unicode text. The system of the present work comprises a compression module, configured to compress the Unicode data by creating the dictionary entries in Unicode format. This is accomplished with adaptive characteristic data compression tables built upon the data to be compressed reflecting the characteristics of the most recent input data. The decompression module is configured to decompress the compressed file with the help of unique Unicode character table obtained from the compression module and the encoded output. We can have remarkable gain in compression, wherein the knowledge that we gather from the source is used to explore the decompression process.

Keywords: Compression algorithms, dictionary-based data compression, LZW, unicode encoding, UTF-8

ARTICLE INFO

Article history:
Received: 17 April 2020
Accepted: 27 July 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.16

E-mail addresses:
rincyanto@gmail.com (Rincy Thayyalakkal Anto)
ryanrajesh@hotmail.com (Rajesh Ramachandran)
*Corresponding author

INTRODUCTION

Since from the distant past, every human society has undoubtedly tried to compress data which is very evident as attempts are made to make devices for briefing as well as for abridgment of relayed messages. The limited transmission bandwidth and necessity of quick and prompt transmission

ISSN: 0128-7680
e-ISSN: 2231-8526

in the digital world makes data compression an imperative and indispensable method. The methodology of minimizing the potential size of a data block with the exception of diminution the quality of information that it is carrying, can be termed as Data Compression. Compression is helpful because it helps in reducing the data storage such as hard disk space and increases its feasibility to be transmitted in the allowed transmission bandwidth. We can tenably state that devoid of fruitful compressions many of the sumptuous and contemporary contrivances, formulations as well as progressions, will be a mirage for all the populations.

Shannon (1948), actualized a formal intellectual discipline for compression, even though there existed compression of data in a very informal manner. Numerous methods are available for data compression. According to the flexibility of methods, compression methods are broadly classified into different types as follows: By pruning and sacrificing some information, lossy compression makes it feasible to attain a better compression rate. Certain files which will become defective and futile if even a bit gets altered, must be compressed merely by lossless compression method. The dissimilarity among lossless and lossy codecs can be evidently featured by taking a cascade of compressions. Perceptive compression methods take advantage of the data being compressed by perceiving our psycho-acoustic and psycho-visual perception. In Symmetric compression, essentially the same algorithm is made to work in "opposite" directions both in the cases of compressor and decompressor. A non-adaptive compression technique which is very inflexible and does not recast its parameters, process, or table in reciprocation to the distinct data selected for compression.

Compression ratio, compression factor and compression gain. are the criteria's commonly used to express the efficiency and performance of a compression method. For byte-based compressors, the common test data used for compression algorithms and implementations are Calgary Corpus and Canterbury Corpus.

Welch (1984) proposed Lempel-Ziv-Welch (LZW) algorithm that belonged to the category of dictionary compressors. If the size of the file is immense, the dictionary based LZW compressor works as an entropy encoder. Always it is desirable to select an adaptive method which is dictionary-based. While static dictionary is a good choice for a special-purpose compressor, an adaptive dictionary-based method is always favoured for a general-purpose compressor. The ideas of Ziv and Lempel (1977; 1978), led to the development of many common LZ compression methods. In LZW, which itself is a leading derivative of LZ78, the token comprises only a pointer to the dictionary. Before inputting any data, the LZW compression commences by initializing the dictionary with all the symbols of the alphabet in the first 256 entries. Hereafter, the succeeding character can be visualized in the dictionary, as it is initialized. Many personages could come up with divergent application and derivatives of this method, which is an evidence of intimidation created by publication of LZW algorithm in 1984.

The most important derivatives of LZW are LZMW (LZ Miller and Wegman), LZAP (LZ All Prefixes), LZY (LZ Yabba) and LZP (LZ Predictor). As the string in the dictionary expand only one character lengthier at each instance, LZW, which is an adaptive method, lags to adjust to its inputs. The LZMW method survives the problems as the existing string is summated to the whole adjacent phrase to the dictionary, rather than adding character one by one. In LZAP method, rather than directly concatenating the last two phrases and positioning the result in the dictionary, it keeps all of the prefixes of the concatenation in the dictionary itself. LZP is a LZ77 variant and is based on the principle of context prediction.

Now-a-days, it is feasible to display and print characters of any shape and size. As a result, 128 characters ASCII (American Standard Code for Information Interchange) code now stands superannuated for modern computing. The needs of present hardware and software are very well met with Unicode encodings. The world's most supreme Unicode character encoding UTF-8, presently hosts the greater part of online contents penned in Non-English dialects. The prevailing compression algorithms routinely act on discrete bytes. Even though, this mode of approach finishes poorly in case of UTF-8, as characters generally span multiple bytes, it operates well for the single-byte ASCII encodings. Repertoire of Unicode 12.1, which spans 150 novel and monumental scripts, have its justified ground with a solid base consisting 137,994 characters in total, as well as multiple symbol sets and emoji. This firm base owes for the dominance of Unicode at unifying character sets, have become the reason of its extensive and tremendous use in the global platforms as well as for the computer software localization.

Even though, various compression techniques are in service for English as well as for other languages, the grammatical structure and syntactic structure differs from one dialect to another. This fact necessitates the need of a peculiar compression technique for natural languages. A small corpus of Unicode files has been compressed on several widely available text compressors of the various types by Fenwick and Brierley (1998), confirming that Unicode files have different compression characteristics from those known for 8-bit data. The Malayalam text compression by variable length encoding was explained by Divakaran et al. (2013), after an informational analysis of Malayalam Language. Barua et al. (2017) projected an enhanced LZW compression technique for Bangla dialect considering the unique features of that language. Gleave et al. (2017) represented modified techniques with escaping on LZW and PPM (Prediction with Partial Matching). An abridged bit representation in the dictionary is an indicative for each Unicode character. A novel approach that suggests modifications to cater the characteristic of Gujarati text to achieve better compression was discussed by Maniya et al. (2012). Vijayalakshmi and Sasirekha (2018) used a static dictionary compression technique in which Unicode characters were replaced with ASCII characters for compression and the original file was retained in the decompression process. Marjan et al. (2014) had proposed a novel approach in which Bengali text was represented efficiently with a better compression ratio.

Yamagiwa et al. (2019) proposed a work known as LCA-DLT (Lowest Common Ancestor- Dynamic invalidation and Lazy compression Technique) and it focused on a technique to reduce the number of searches in the dictionary using a bank separation technique. The aptitude of wavelet transform to be multilingual lossy text compression was discussed and a new strategy that is based on the application of wavelet transform to text compression of files was proposed by Al-Dubaee et al. (2010). A novel technique of compressing a more symbolic dialect like Bengali through a less symbolic language like English was proposed by Hossain et al. (2014) in which Huffman principle was used to its maximum.

The established algorithms such as bzip2, usually dominates over newly developed Unicode compressors from scratch, as per research findings. Here, we are trying to introduce a technique to modify byte-based compressors to percolate directly on Unicode characters and to execute a variant of LZW to accomplish better compression.

In this article, we apply an interactive approach to data compression and discuss its prominent benefits as well as significant advantages. To execute this, we can presume some degree of interaction between the compressor and the decompressor. It can allow a more competent usage of the information gained during compression of the original file, whilst the data compression is used for data transmission purposes.

The systematic organization of the paper is presented as follows: In section II, LZW Compression technique is explained for both ASCII text and Unicode Text. In section III, we propose our new method for Unicode text as a variation of existing LZW Compression method. We are discussing the results of our innovative method in section IV. The future research directions are concluded in Section V.

LZW COMPRESSION

LZW is characterized by the fact that its conditions are fulfilled only by one pass over the given data. A code table is created wherein each entry is made up of peculiar strings of characters. As detailed by the algorithm in Figure 1, the principle of LZW derived by Welch (1984) is that the encoder puts in symbols one after another, to get them accumulated in a string α. The dictionary is searched for string α, after receiving and concatenating each discrete symbol to α. The process continues as long as α is discovered in the dictionary. At a certain point, adding the immediate successive symbol x causes a failure of search characterized by the presence of string α and absence of string αx in the dictionary. At this juncture, the encoder

a. Output the dictionary pointer that points to string α
b. Saves string αx, in the next available dictionary entry, and
c. Initializes string α to symbol x.

Algorithm 1 The Existing LZW Encoder

```
1: for i←0 to 255 do
2:    append i as 1-symbol string to the dictionary;
3: end for
4: α← empty string
5: x ← first input character
6: while not end of stream do
7:     if α+x is in the dictionary then
8:         α← α+x
9:     else
10:        output code for α
11:    end if
12:    append α+x to the dictionary
13:    α ← x
14:    x ← next input character
15: end while
16: output code for α
```

Figure 1. Algorithm for the existing LZW Encoder

The decoder starts to make its operations with the first 256 entries of its dictionary initialized to the whole symbols of the alphabet. Later, it reads the input stream which includes pointers to the dictionary and utilizes each pointer to retrieve uncompressed symbols from its dictionary and writes them on its output stream. The algorithm for the existing LZW decoder is provided as Figure 2. In this process, it constructs its dictionary in lockstep with the encoder.

Algorithm 2 The Existing LZW Decoder

```
1:  for i← 0 to 255 do
2:     append i as 1-symbol string to the dictionary
3:  end for
4:  read first input code
5:  output translation of code, α
6:  w ← α
7:  while not end of input do
8:     x ←read next input code
9:     β← translation of x
10:    output β
11:    append w+β[0] to dictionary
12:    w←β
13:  end while
```

Figure 2. Algorithm for the existing LZW Decoder

When we encode the string “agh eats aghagha” with LZW, the dictionary entries obtained are shown in Figure 3.

0	NULL	256	ag	264	␣a
1	SOH	257	gh	265	agh
.....		258	h␣	266	ha
32	SP	259	␣e	267	agha
.....		260	ea		
97	a	261	at		
.....		262	ts		
255	255	263	s␣		

Figure 3. LZW dictionary entries for the string “agh eats aghagha”.

But, if we compress a string including Unicode characters, then it will also do byte-wise compression. Figure 4 shows the dictionary entries for the Malayalam (A Dravidian language spoken over the Indian territory of Kerala) string 'കേന്ദ്രതാഴില് മന്ത്രാലയം'.

0	NULL	260	´?	272	à´°	284	²à	296	°à
1	SOH	261	?à	273	°	285	àµ?	297	à´¾
.....		262	àµ	274	à	286	à´	298	¾à
32	SP	263	µ?	275	à´¤	287	´®	299	à´²
.....		264	?à	276	¤à	288	®à	300	²à´
97	a	265	à´¨	277	àµ?	289	à´¨à	301	´¯
.....		266	¨à	278	?à	290	àµ?à´	302	¯à
255	255	267	àµ?	279	à´´	291	´¤	303	à´?
256	ï»	268	?à	280	´à	292	¤àµ		
257	»¿	269	à´	281	à´¿	293	µ?		
258	¿à	270	¦à	282	¿à´	294	?à´		
259	à´	271	àµ?à	283	´²	295	´°		

Figure 4. LZW dictionary entries for the string 'കേന്ദ്രതാഴില് മന്ത്രാലയം'.

A code table (dictionary), with 4096 as a usual choice for the number of table entries, is utilized in LZW compression. At the commencement of encoding, the code table comprises of only the initial 256 entries of single byte characters, with the rest of the table being kept blank. Codes 256 through 4095 are used to indicate the sequential bytes for achieving the compression. While the encoding advances, LZW determines frequented sequences in the data and adds them to the code table. For the above text, it builds the dictionary with 303 entries, even though the text contains only 25 Unicode characters. This prompts to making an attempt to recommend a technique to alter byte-based compressors to permeate

precisely on Unicode characters and to actualize a versatile variant of LZW to attain better compression.

METHODOLOGY

An optimised method and system for compressing and decompressing a UTF-8 encoded stream of data using an enhanced LZW algorithm was made available. In this method of compression, the initial part is to read data from the input stream present in the form of UTF-8 characters. The only known work which reads data in the form of UTF-8 characters belongs to Gleave et al. (2017). In their work, they had investigated the effectiveness of different token distributions while being used as a base distribution for LZW. They had proposed the algorithms of LZW and PPM with three base models. The initial two models characterised by uniform base distribution over the byte alphabet as well that of token alphabet, and another one with Polya tree-based model over the alphabet. They could find that adaptive character model based on Polya trees was well suited for learning Unicode character distributions. In this method, UTF-8 text is decoded into a sequence of Unicode code points. Beginning with an empty initial dictionary, the method is implemented by escaping to a base model over tokens, immediately when a new symbol is seen. This implementation places no upper limit on the dictionary size N and hence, encodes each index in $\log_2 N$ bits using arithmetic coding. As the mapping used by them had the appealing property of allocating semantically related tokens to nearby integers, they claimed that their approach yielded a 12.2% average improvement in compression effectiveness for LZW over a corpus of UTF-8 files. We propose a work without converting UTF-8 text to Unicode code points and with no initial identification of token distributions.

In our work, we propose a different version of the traditional LZW encoding algorithm in which no characters are preloaded in the dictionary. The newly derived algorithm for the encoder is shown in Figure 5. As the dictionary is empty, the occurrence of each new character creates a new entry in the dictionary. As the process of encoding goes on, it further decides whether to output a code associated with the data, in case if the data is already in the dictionary or to add the character to the dictionary as a new character, in case if it is not found in the dictionary. The system identifies duplicated sequences in the data as well as the new characters which are unavailable in the dictionary and adds them to the dictionary. All the new characters along with its position in the dictionary are stored in a separate table and it is passed as part of the compressed data to the decompressor. Even though this may be considered as an overhead, the price we accept to pay is negligible compared to the benefits we achieve. In the existing LZW method, 4096 is used as the dictionary size and 12 bits are used to output the code. When the dictionary has filled up, existing LZW becomes static. Further it compresses with unchanging dictionary and may lead to degradation in compression performance. So in our proposed system, when the dictionary

had been filled up to a specified maximum, we cleared the dictionary after checking the boundary conditions and then reloaded the dictionary with all new characters found so far along with a certain number of most frequently used character sequences from the previous dictionary. Rather than using 12 bits for every outputted code, we used a variable-width coding system in which each code was represented exactly as their binary representation.

LZW methods convert a sequence of strings (<string (1)> . . . <string (n)>) to a sequence of codes (<code (1)> . . . <code (n)>). The proposed strategy and methodology lessen the number of entries in the dictionary and hence caters the need of better compression in a file. It can boost the activities of many LZW-based applications.

Algorithm 3 The Proposed LZW Encoder

```
1:  α ← empty string
2:  x ← first input character
3:  while not end of stream do
4:     if x is not in the dictionary then
5:        append x to dictionary
6:        add x to a string table 'dict_add' with its dictionary index position
7:     end if
8:     s← α+x
9:     if s is in the dictionary then
10:       α ← s
11:    else
12:       output code for α
13:       append s to the dictionary
14:       α←x
15:    end if
16:    x← next input character
17: end while
18: output code for α
19: pass the string table 'dict_add' to the decoder
```

Figure 5. Algorithm for the proposed LZW Encoder

The same string in Malayalam 'കേന്ദ്രത്തിന്റെ മന്ത്രാലയം'. was tested with the proposed compression and the number of dictionary entries was reduced to 37. Malayalam which is one among 22 scheduled Indian languages is a one with Dravidian origin spoken in the state of Kerala. With the largest number of alphabets accounting to 56, Malayalam tops the other Indian languages in this aspect. The dictionary entries created are shown in Figure 6.

Decompressor built the dictionary in the same pattern as that of compressor with the help of the table that it received as part of the compressed file. The decompressor reads the input stream containing the pointers to the dictionary and used each pointer to recover original symbols from its dictionary, if found present in the dictionary and wrote them on its output stream. On the other hand, if it was not present in the dictionary, the system extracted the next character from the table that it received from the compressor to create a

0	ക	10	ര	20	ല	30	ൽ
1	േ	11	്ര	21	ില	31	ി
2	കേ	12	ന	22	ല്	32	്രാ
3	ന	13	രന	23		33	ിലെ
4	നേ	14	ര	24	്	34	ള
5	്	15	രന്ത	25		35	ലള
6	ന്	16	ഴ	26	ള	36	ിം
7	ര	17	രിഴ	27	ള	37	ളം
8	്ര	18	ി	28	ളന		
9	ര്	19	ഴി	29	ന്ത		

Figure 6. The proposed LZW dictionary entries for the string ˙കകകരിരിരിനിലെ ളരിരിരിലെളം˙.

new dictionary entry, if the position extracted from the table matched with the dictionary index. The number of dictionary entries and the search time are considerably reduced due to the proposed system and it contributes directly to faster compression and decompression. Thus, we have made an attempt to recommend a technique to alter byte-based compressors to permeate precisely on Unicode characters and to actualize a versatile variant of LZW to attain better compression.

In the proposed LZW decoding method, based on algorithm detailed in Figure 7, the decoder reads the first character α from the string table 'dict_add' as the first decoding step and append it to dictionary. For every following steps of decoding after the initial one, the decoder inputs the next pointer, fetches the immediate string β from the dictionary if it is present, writes it on the output, followed by isolation of its initial symbol, and saves string w+ initial character of β in the next obtainable entry in the dictionary. If not present in the dictionary, it extracts the next character from the table 'dict-add' and adds it to the dictionary if the position matches with the dictionary index. String β is the extracted character in this case, and it is noted down on the output. The decoder then proceeds w to β and is all set for further step.

Some of the existing LZW compression or decompression methods use 8 to 16 bits to represent the dictionary entries. Accordingly, the number of entries in the LZW dictionary stands limited to 2^number of bits. The application program will now be in a position either to fortify the dictionary once it is full, or to cast off a few entries in the dictionary once it becomes filled. Henceforth, we can undoubtedly say that the capacity of LZW dictionary is obtained from the size of code (i.e., compression code) used in a particular implementation of the LZW variant (algorithm). The system and method of the present disclosure provided

Algorithm 4 The Proposed LZW Decoder

```
1: read first character α from the string table 'dict_add'
2: append it to the dictionary
3: remove the first input code
4: output α
5: w ← α
6: while not end of input do
7:     x ← read next input code
8:     if dictionary contains translation of x then
9:         β ← translation of x
10:    else if the position of next character in 'dict_add' = current dictionary
                index then
11:        extract the next character y from the string table
12:        append y to the dictionary
13:        β ← y
14:    end if
15: output β
16: append w+β[0] to dictionary
17: w ← β
18: end while
```

Figure 7. Algorithm for the proposed LZW Decoder

for enhanced LZW compression or decompression utilizes the method of clearing the dictionary once it is full and reloading the dictionary with all the new characters found so far. Implementation of this algorithm with dictionary update makes use of a certain number of most frequently used entries from the previous dictionary.

The most frequently used entries in the dictionary were found by using an auxiliary dictionary. Whenever the code corresponding to a character sequence in the dictionary was added to the output of an encoder, the count corresponding to the character sequence was incremented in the auxiliary dictionary. Once the dictionary reaches its specified maximum, the dictionary was cleared, and all the new characters found so far were added to the dictionary. Thereafter, the auxiliary dictionary was sorted and a specified number of most recently used entries are added to the dictionary. This helps the compression process very much as there is no need to start all from scratch.

We have tested this algorithm both with 32 KB and 64 KB dictionaries. We were using auxiliary dictionary of the same size, with dictionary update variant of this algorithm. The results obtained and its comparison with two popular methods Unix compress and gzip are detailed below.

RESULTS AND DISCUSSION

The modus operandi of the present invention in one actualization executes an enhanced LZW procedure to create dictionary entries in a dynamic manner and acquires codes from those entries at the stage of compression. It retrieves the code and reinstates the original stream.

After testing the proposed system with Malayalam String, we had also done it with many Unicode files in languages like Tamil, Hindi, Bengali and Arabic. Some of the sample files, number of dictionary entries (with less than 4096 entries) and execution time are given in Table 1. From Table 1, we could see that number of dictionary entries was considerably reduced in the proposed LZW compression scheme and the time spent for compression and decompression was much lesser as the search time was reduced.

Table 1

Comparison between The Existing LZW method and The Proposed LZW method

File Name	No. of dictionary entries		Time spent (in milliseconds)			
			ELZW[1]		PLZW[2]	
	ELZW[1]	PLZW[2]	Comp[3]	Decomp[4]	Comp.[1]	Decomp.[2]
malayalam.txt	3411	2433	112	21	23	7
tamil.txt	1334	823	71	17	17	4
hindi.txt	1896	1277	85	21	21	5
Bengali.txt	4096	3873	103	28	25	8
Arabic.txt	3267	2581	117	20	40	10

[1] Existing LZW [2]Proposed LZW [3]Compression [4]Decompression

According to Uthayakumar et al. (2018), the performance of data compression algorithms can be evaluated in multiple dimensions. The competence of compression can be estimated using parameters like algorithm complexity, computational memory, speed, amount of compression and quality of reconstructed data. Even though the performance of data compression can be evaluated using multiple parameters, most frequently it is done by Compression Ratio. It is defined as the ratio of total number of bits required to store uncompressed data and total number of bits required to store compressed data and is given in Equation 1.

$$CR = \frac{\text{No: of bits in uncompressed data}}{\text{No: of bits in compressed data}} \quad [1]$$

Another measure called Space savings is also used, which defines the reduction in file size relative to the uncompressed size and is given in Equation 2.

$$\text{Space Savings} = 1 - \frac{\text{No : of bits in compressed data}}{\text{No : of bits in uncompressed data}} \quad [2]$$

Table 2

Compression Ratio – The Existing LZW method and The Proposed LZW method

File Name	Original File Size (KB)	Compressed File Size		Additional overhead	Total File Size	Compression Ratio		Space Savings (%)	
		ELZW[1] (KB)	PLZW[2] (KB)			ELZW[1]	PLZW[2]	ELZW[1]	PLZW[2]
malayalam.txt	14.7	4.62	3.46	0.37	3.83	3.18	3.83	68.57	73.94
tamil.txt	3.89	1.58	1.13	0.26	1.39	2.46	2.8	59.38	64.26
hindi.txt	6.16	2.39	1.77	0.36	2.13	2.58	2.9	61.2	65.58
Bengali .txt	23.1	7.28	5.55	0.46	6.01	3.17	3.85	68.48	74.02
arabic.txt	10.2	4.41	3.68	0.28	3.96	2.31	2.57	56.76	61.17

[1] Existing LZW [2]Proposed LZW

Table 2 shows the improved compression performance for the same files in terms of Compression Ratio and Space Savings. It describes the total compressed file size after adding the additional overhead caused by the proposed algorithm. By additional overhead, we mean the space required to store the various new characters detected and its position during the compression process. As the dictionary is initialized as empty with this algorithm, it is passed as part of the compressed file. Compression performance is substantially improved as per the data in Table 2.

From the above tables, it is clear that the Unicode character-based compressor based on an enhanced LZW scheme outweighs the traditional byte based LZW compressors in performance. The number of dictionary entries is considerably reduced. Both the sender and the receiver are able to construct a code table either synchronously or in lockstep. Though the code table may occupy more space with the Unicode characters than that with ASCII characters, but it is likely to be compensated considerably as the number of entries is less in the code table.

After the analysis of the proposed algorithm with some small files, we made a thorough analysis of it with some large files from Canterbury Corpus and from the public domain. This analysis was done with 32 KB and 64 KB dictionaries. The first one is the proposed LZW method wherein the dictionary size used is 32 KB. When the dictionary size reaches its specified maximum, it clears the dictionary and only the new characters found so far is used to initialize the dictionary. In the second method, we were updating the dictionary as in the previous method and in addition, a certain number of most recently used dictionary entries from the previous dictionary were also appended. Auxiliary dictionaries of the same size were used with this algorithm. Unix Compress which is a popular method that use LZW algorithm was also tested with 32KB and 64KB dictionaries. Another LZ variant compression tool gzip was also used for the comparison.

The details of test data files are given in Table 3 and comparison between the compression ratios of different LZ implementations are detailed in Table 4.

Table 3
Test Corpus

File	Size in KB	Description
ben-kobita.txt	23.1	Bengali Novel
hin-baital.txt	420	Hindi
mal-travelblog.txt	209	Malayalam
Indulekha.txt	290	Malayalam Novel
Mal-Bible.txt	1453	Malayalam Bible
ben-blog.txt	1104	Bengali blog

Table 3 *(Continued)*

File	Size in KB	Description
Bhagavat Geetha.txt	1068	Malayalam
Kundalatha.txt	174	Malayalam Novel
Umakeralam.txt	347	Malayalam Poetry
alice29.txt	676	Alice in wonderland, English
Asyoulik.txt	148	As you like it, Early modern English
lcet10.txt	122	Technical Writing, English
plrabn12.txt	416	Paradise Lost, English Novel
ara-tabula.txt	470	Arabic
Beowulf.txt	2277	Old English Poetry
crime_and_punishment.txt	156	Russian Novel
genchi-all.txt	1889	Japanese Novel
Kokoro.txt	1421	Japanese Novel
ziemia_obiecana.txt	473	Polish Novel
cedict_small.txt	1202	Mixed language, Chinese & English

Table 4

Compression Ratio with proposed LZW, Unix Compress and gzip. Highest Compression Ratio is shaded in grey color

File name (.txt File)	**Original File Size (KB)**	**Compression Ratio**						
		Proposed LZW				**Unix Compress**		**Gzip**
		32 KB	**64 KB**	**32KB Dictionary update**	**64KB Dictionary update**	**32KB**	**64KB**	
ben-blog	23.1	4.17	4.17	4.17	4.17	3.30	3.30	4.15
ben-kobita	420	5.24	5.42	5.42	5.42	4.31	4.29	4.42
hin-baital	209	6.13	6.13	6.13	6.13	4.55	4.55	5.89
mal-travelblog	290	5.40	6.12	6.12	6.12	4.65	4.63	5.27
Indulekha	1453	7.30	7.30	7.38	7.41	4.96	5.36	6.08
Mal-Bible	1104	6.92	7.34	7.46	7.51	5.21	5.66	6.24
Khuran	1068	7.80	7.80	7.85	7.91	5.42	5.84	6.47

Table 4 *(Continued)*

File name (.txt File)	Original File Size (KB)	Compression Ratio						
		Proposed LZW				Unix Compress		Gzip
		32 KB	64 KB	32KB Dictionary update	64KB Dictionary update	32KB	64KB	
Bhagavat Geetha	174	5.34	5.34	5.34	5.34	4.24	4.24	4.73
Kundalatha	347	6.46	6.46	6.46	6.46	4.82	4.82	5.48
Umakeralam	676	5.24	5.24	5.28	5.28	4.12	4.17	4.23
alice29	148	2.39	2.43	2.44	2.44	2.45	2.44	2.79
asyoulik	122	2.27	2.27	2.27	2.27	2.27	2.27	2.55
lcet10	416	2.41	2.51	2.57	2.57	2.43	2.62	2.95
plrabn12	470	2.26	2.35	2.39	2.40	2.34	2.37	2.47
ara-tabula	2277	5.17	5.41	5.42	5.53	3.92	4.13	4.69
beowulf	156	2.54	2.58	2.59	2.59	2.52	2.50	2.69
crime_and_ punishment	1889	3.80	3.99	4.03	4.09	3.25	3.37	3.65
genchi-all	1421	3.90	4.10	4.14	4.24	2.99	3.17	3.30
Kokoro	473	3.45	3.62	3.58	3.64	2.83	2.99	3.20
ziemia_ obiecana	1202	2.28	2.40	2.42	2.47	2.25	2.36	2.54
cedict_small	745	1.65	1.72	1.75	1.77	1.71	1.80	1.97

Figure 8 depicts a comparison between the compression ratios obtained with the above said implementations with 32 KB dictionaries. Figure 9 depicts the same with 64 KB dictionaries. From the results obtained, it is evident that the proposed method compressed UTF-8 files by an average of 5% while using 32 KB dictionaries and 4% in case of 64 KB dictionaries than Unix Compress. The proposed method also showed a better compression ratio than gzip for languages with 3 to 4 bytes encodings.

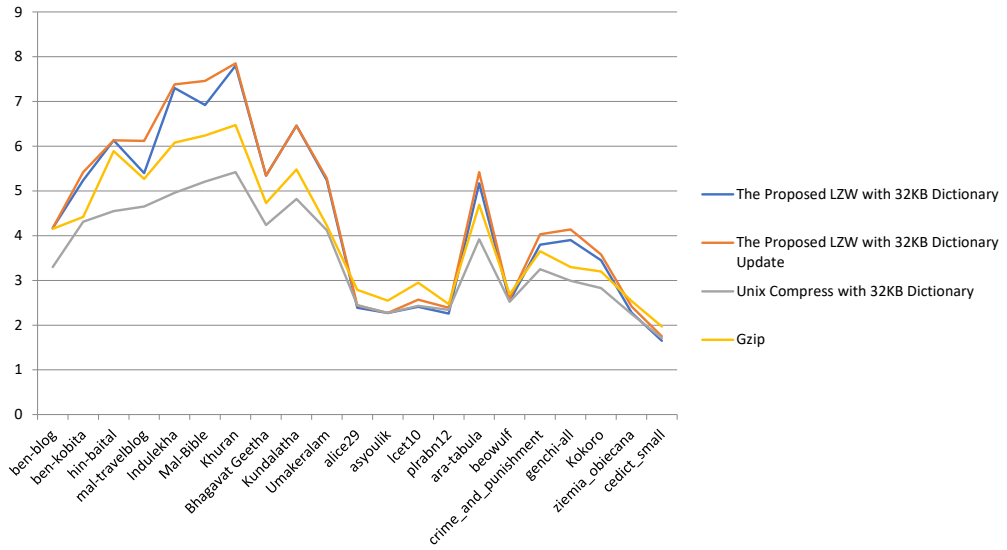


Figure 8. Compression Ratios with 32 KB dictionaries and Gzip

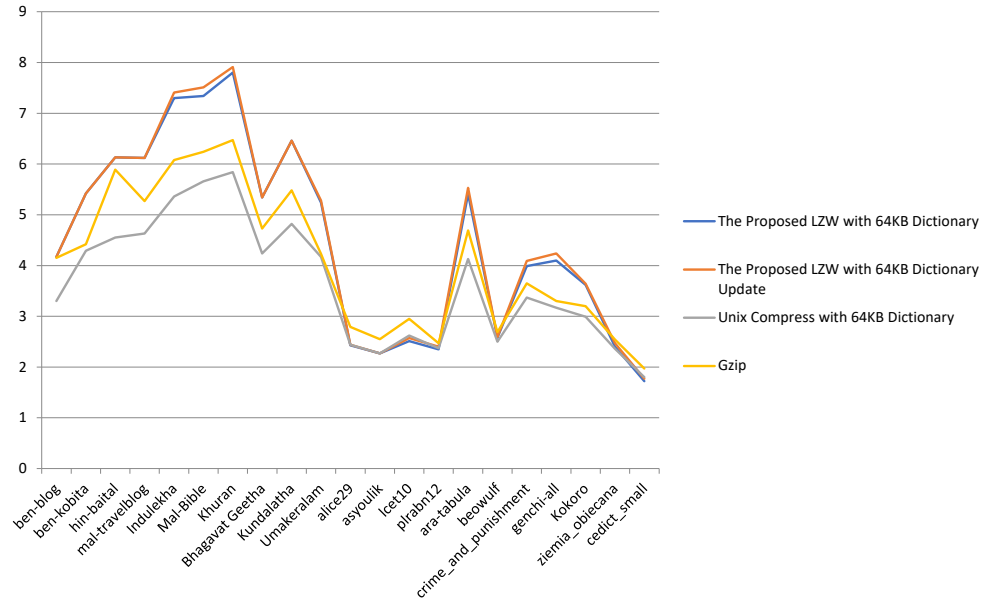


Figure 9. Comparison of Compression Ratios with 64 KB dictionaries and Gzip

CONCLUSION

This research demonstrates the use of enhanced LZW approach which permits empty dictionaries and it offers an efficient dictionary update mechanism whenever the dictionary reaches its specified maximum. This approach makes use of Unicode characters to lessen the number of dictionary entries and to achieve better compression. From the experiments, we can see that the proposed method is highly efficient in compressing text files with 3 or 4 byte encodings, even though there is performance degradation while compressing files of 1 or 2 byte encodings (Figure 8 & 9). This research will be the basis for future research to improve compression rate. It may be interpreted that the current invention is restricted to these specific patterns, but rather consider them as few explanatory examples alone. Hence, diverse alteration and improvisation shall be invoked by anyone who is talented in the art, without departing from the spirit or scope of the invention.

ACKNOWLEDGEMENT

We thank the Bharathiar University Research Centre, Coimbatore for all the facilities extended so far during the period of research.

REFERENCES

Al-Dubaee, S. A., & Ahmad, N. (2010, August 5-7). Multilingual lossy text compression using wavelet transform. In *2010 First International Conference on Integrated Intelligent Computing* (pp. 39-44). Bangalore, India.

Barua, L., Dhar, P. K., Alam, L., & Echizen, I. (2017, February 16-18). Bangla text compression based on modified Lempel-Ziv-Welch algorithm. In *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 855-859). Cox's Bazar, Bangladesh.

Divakaran, S., Anjali, C., Biji, C. L., & Nair, A. S. (2013). Malayalam Text Compression. *International Journal of Information Systems and Engineering, 1*(1), 7-11.

Fenwick, P., & Brierley, S. (1998, August 6). Compression of unicode files. In *Proceedings of DCC '98 Data Compression Conference* (pp. 547). Snowbird, Utah, USA.

Gleave, A., & Steinruecken, C. (2017, April 4-7). Making compression algorithms for Unicode text. In *Proceedings of the Data Compression Conference* (pp. 22-25). Snowbird, Utah, USA.

Hossain, M. M., Habib, A., & Rahman, M. S. (2014, May 23-24). Transliteration based bengali text compression using huffman principle. In *2014 International Conference on Informatics, Electronics & Vision* (ICIEV) (pp. 1-6). Dhaka, Bangladesh.

Maniya, S. V., Sheth, M. J., & Lad, K. (2012). Compression Technique based on Dictionary approach for Gujarati Text. *International Journal of Engineering Research and Development, 4*(8), 101-108.

Marjan, M. A., Uddin, M. P., Afjal, M. I., & Haque, M. D. (2014, October 21-23). Developing an efficient algorithm for representation and compression of large Bengali text. In *2014 9th International Forum on Strategic Technology (IFOST)* (pp. 22-25). Cox's Bazar, Bangladesh.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379-423.

Uthayakumar, J., Vengattaraman, T., & Dhavachelvan, P. (2018). A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications. *Journal of King Saud University-Computer and Information Sciences, 2018*, 1-22.

Vijayalakshmi, B., & Sasirekha, N. (2018). Lossless text compression for unicode Tamil documents. *ICTACT Journal on Soft Computing, 8*(2), 1635-1640.

Welch, T. A. (1984). A technique for high-performance data compression. *Computer, 17*(6), 8-19.

Yamagiwa, S., Morita, R., & Marumo, K. (2019, March 26-29). Bank select method for reducing symbol search operations on stream-based lossless data compression. In *2019 Data Compression Conference (DCC)* (p. 611). Snowbird, Utah, USA.

Ziv, J., & Lempel, A. (1978). Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory, 24*(5), 530-536.

Ziv, J., & Lempel, A. (1977). A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory, 23*(3), 337-343.

Pertanika J. Sci. & Technol. 28 (4): 1445 - 1468 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Electricity Generation Potential and Energy Cost of Wind Conversion Systems in Ikeja Southwest Nigeria

Adetona Tayo Fatigun*, Ebenezer Babatope Faweya, Funmilola Olusola Ogunlana and Taiwo Hassan Akande

Department of Physics, Ekiti State University, P.M.B. 5363, Ado-Ekiti, Ekiti State, Nigeria

ABSTRACT

In this study, the wind electricity generation potential and energy cost at Ikeja were investigated using 31 years wind speed data obtained from Nigeria Meteorological Agency. The study addresses the challenges of inadequate electricity supply and the development of alternative source of electricity. The measured data, captured at 10m height were subjected to 2-parameter Weibull and other statistical analysis. Weibull analysis of wind speed showed good fit between actual data and Weibull predicted data confirming the adequacy of the model. The value of wind speed at 10m height ranged between 3.47m/s and 5.33m/s with annual average of 4.5m/s. Also, the Wind Power Density (WPD) ranged between 116.3 W/m^2 and 423.3W/m^2 with annual average value of 257.85W/m^2. The mean electric power outputs from the model turbines varied between 11KW and 290KW while its Capacity Factor (CF) ranged between 13.8% and 0.36%. Also, the generation cost per kilowatt-hour varied between \$0.11 and \$2.39 annually. Therefore, the wind energy potential at Ikeja could be adjudged marginal and belonging to wind power class 2. The generation cost of wind electricity is cost-effective in the months of April and August while cost-deficit in the remaining months of the year. The location is considered suitable for small to medium

ARTICLE INFO

Article history:
Received: 31 August 2019
Accepted: 2 January 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.17

E-mail addresses:
fatigunadetona@gmail.com (Adetona Tayo Fatigun)
febdeprof@yahoo.co.uk (Ebenezer Babatope Faweya)
funmieola@yahoo.com (Funmilola Olusola)
thaiwohassan@gmail.com (Taiwo Hassan Akande)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

scale wind power generation, but economically infeasible for large scale grid connected wind electricity generation.

Keywords: Capacity factor, energy cost, mean power output

INTRODUCTION

Energy availability has been identified to be one of the factors that have major effect on the socio-economic life of any society (Herman, 2001). The poor state of electrical power supply in Nigeria has been widely viewed as one of the major constraint to the nation economic growth (Nkalo & Agwu, 2019). The electrical energy in particular is an important component for the development of any economy. Besides capital and labor, electricity is the third important factor of production in economic model (Ezema et al., 2016). Lack of electricity or its inadequacy had been said to be a source of social and economic poverty (Nkalo & Agwu, 2019). Nigeria, despite abundance natural resources, including fossil fuel and renewable energy resources, has one of the lowest net electricity generation (Adebayo, 2014). The report has it that about 75millions Nigerian lack access to adequate electricity (Adebayo, 2014). Lagos state which is Nigeria economic center is not an exception in the terms of inadequate electricity supply from the national grid.

The present electricity demand of Lagos state alone is in needs of additional 4000MW (Oladipo et al., 2018). While the entire country installed electricity generation capacity is 12,522MW and available capacity of 6,056MW (NERC, 2019). The actual electricity generated and distributed in Nigeria is often less than 4000MW and sometimes drops as low as below 1000MW (Oladipo et al., 2018). This shows that the Lagos state could not attain self-sufficiency in her energy need by depending solely on the supplies from the national grid. The inadequate electricity supply from national grid has made most of the residents and industries to resort to power generation from small, medium and heavy-duty generating sets (Odunola et al., 2018).These have become major source of environmental pollution and health hazard. The environmental pollution from fossil fuel combustion has been described as the world most significant threat to children's health and future and is significant contributors to global inequality and environmental injustice (Federica, 2018). Considering the high energy demand of Lagos, no single energy resources can sustainably meet its demand without energy mix; therefore, integrating all exploitable energy sources is a viable way of achieving stability in energy supply of Lagos state (Uzoma et al., 2014).. The economic growth of Lagos city has been influenced by the high population density, the commercial and industrial activities within the metropolis. Since energy growth, economic development and sustainable development are grossly inseparable (Uzoma et al, 2014).

Therefore, there is a need for the development of alternative and sustainable energy source that will be commensurate to the population growth and economic development of the state.

It has been widely accepted that wind energy is becoming the fastest growing renewable sources of energy in both developed and developing countries (Ajayi et al., 2014). In Sub-Sahara Africa, particularly the West Africa region, no country has yet generated grid electricity from wind despite the identified opportunities (Ajayi, 2013). The challenge of wind energy project development in Sub-Sahara Africa may however be linked to inadequate measurement, lack of assessment studies and/or improper classification of the location wind regime (Okeniyi et al., 2015).

Some of the previous research study that have assessed wind energy potential in different locations in Nigeria include the work of Ajayi, (2010), Agbetuyi et al. (2012), Adaramola et al. (2011) and Fagbenle et al. (2011). Majority of the studies have focused on the northern part of the country. This may probably be due to fact that wind speeds are generally believed to be weak in the southern part. But coastal regions and offshore areas of southwest and south-south have been reported to have potentialities for strong wind (Oyewole & Aro, 2018; Ajayi, 2010). Also, due to varying roughness of the country, large differences may exist within the same locality in term of wind energy potential (ECN-UNDP, 2015). Therefore, more needs to be done to expose the wind profile of the southern regions of Nigeria (Ajayi, et al., 2014). Some of the previous studies aimed at assessing the wind energy potential and energy cost of locations within the southwest geopolitical zone includes the work of Ajayi et al. (2014). They presented study on the wind energy potential and the cost-benefit analysis of wind power generation in ten selected sites within the southwest geopolitical zones of Nigeria. There result showed that locations in Lagos and Oyo state were suitable for large scale wind electricity generation while energy cost analysis showed that generation cost varied between €0.02 and €5.03 depending on the turbine model employed. Likewise, Nze-Esiaga and Okogbue (2014) studied the wind power generation potential as a power generation sources in five locations of south western Nigeria. There result showed that Ikeja had better wind speed profile in the wet season compared to the other locations considered. They concluded that the wind speed of most of the locations was viable for wind electricity.

In this study, Weibull model was employed in the wind speed analysis. The 2-parameter Weibull distribution function out of all probability density functions has enjoyed wide use (Fadare, 2008; Carta, et al., 2009). This is because the 2-parameter Weibull model produces better fit to the observed data and could be use to determine the two important wind speeds (maximum energy carrying wind speed and most probable wind speed) for wind farm evaluation and further technical analyses (Akpinar & Akpinar, 2005). Also, the quantitative assessment of wind electricity generated was carried out using power-output simulation of four different commercial wind turbines while energy cost was based on present value

cost (PVC) method. The result will lead to increase in the reliability and predictability of wind as alternative source of energy. It will also show how wind electricity potential and energy cost vary across months and seasons of the year. This will enable the electricity grid to make informed decision on the management and accommodation of variable nature of wind energy. It will also be useful to individuals, government at all levels, wind energy assessors and policy makers regarding wind electricity generation and investment.

MATERIAL AND METHOD

Thirty-one years (1980-2010) monthly mean wind speed data measured at Ikeja airport was obtained from the Nigeria Meteorological Agency (NIMET), Oshodi, Lagos State, Nigeria. The data was recorded continuously using cup-generator anemometer at a height of 10m. The high variability in wind speed data necessitated the use of probability distribution function in the data analysis. Though there are many probability distribution functions that describe wind speed distribution in a particular location. However, Weibull distribution has been found to be most accurate and adequate in analyzing and interpreting wind speed data (Carta et al., 2009). In this study, two-parameter Weibull distribution model was used in the analysis of wind speed data. The estimated Weibull scale (c) and shape parameters (k) were used in the calculation of other wind parameters while the cost of wind electricity was estimated using the present value cost method (PVC).

Weibull Parameters Estimation

The shape parameters (k) and scale parameter (c) were estimated based on the maximum likelihood method (MLM). The MLM technique estimates the Weibull shape parameter (k) and the scale parameter (c) using the following Equation 1 and 2

$$k = \left\{ \frac{\sum_{i=1}^{n} v_i^k \ln v_i}{\sum_{i}^{n} v_i^k} - \frac{\sum_{i=1}^{n} \ln v_i}{n} \right\}^{-1} \quad [1]$$

$$\mathbf{c} = \left\{ \frac{1}{\mathbf{n}} \sum_{\mathbf{i}}^{\mathbf{n}} = \mathbf{v_i^k} \right\}^{1/\mathbf{k}} \quad [2]$$

where vi is the wind speed in time step i and n is the number of non-zero wind speed data points.

The mean wind speed (Vd) can be evaluated from the actual data using the Equation 3

$$v_d = \frac{1}{n} \left(\sum_{i=1}^{n} v_i \right) \quad [3]$$

while the predicted weibull mean wind speed (Vw) is defined in terms of the Weibull parameters k and c is given according to Equation 4

$$v_{weibull}\left(= c\Gamma 1 + \frac{1}{k}\right) \quad [4]$$

Performance Evaluation of Weibull Distribution Model

The performance of Weibull model used in the prediction of monthly mean wind speeds was evaluated using the following statistical test; correlation coefficient (R2), Chi-square ($\chi 2$), root mean square error (RMSE) and coefficient of efficiency (COE). These test models are estimated using the following Equation 5, 6, 7 and 8

$$R^2 = \frac{\sum_{i=1}^{N}(y_i - z_i)^2 - \sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i - z_i)^2} \quad [5]$$

$$\chi^2 = \frac{\sum_{i=1}^{n}(y_i - z_i)^2}{N - n} \quad [6]$$

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2\right]^{\frac{1}{2}} \quad [7]$$

$$COE = \frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i - z_i)^2} \quad [8]$$

Wind Power Density

The wind power density (WPD) is the amount of wind energy transported across a unit area in unit time. This important parameter can be estimated from the values of Weibull scale and shape parameters using the expression given in Equation 9

$$WPD = \frac{1}{2}\rho c^3\left(1 + \frac{3}{k}\right) \quad [9]$$

where ρ is the location's air density(kg/m^3).

Wind resources are usually represented by wind power classes. Table 1 shows the international wind power classification at 50m hub height. Each class represents a range of wind power densities and equivalent mean wind speed.

Table 1

International wind power classification at 50m hub height (Ahmed, 2016)

Wind power class	WPD (W/m^2)	Wind speed (m/s)	Remark
1	≤200	≤5.6	Poor
2	≤300	≤6.4	Marginal
3	≤400	≤7.0	Fair
4	≤500	≤7.5	Good

Table 1 *(Continued)*

Wind power class	WPD (W/m^2)	Wind speed (m/s)	Remark
5	≤600	≤8.0	Excellent
6	≤800	≤8.8	Outstanding
7	≤2000	≤11.9	Superb

Wind Turbulence Intensity

Wind turbulence is the rapid disturbances or irregularities in the wind speed, wind direction and wind vertical component.. The most common indicator of turbulence for sitting purpose is the standard deviation from the mean wind speed. Turbulence intensity can be estimated by normalizing standard deviation with the mean wind speed using the expression given in Equation 10

$$TI = \frac{\sigma}{\bar{v}} \quad [10]$$

Turbulence intensity, a dimensionless quantity is a relative indicator of turbulence level. Low level is indicated by values less than or equal to 0.10 (≤ 0.10), moderate turbulence level is indicated by values greater than 0.10 and up to 0.25 ($0.10 \leq TI \leq 0.25$) while high turbulence level is indicated by values greater than 0.25 ($TI > 0.25$) (Ahmed, 2016).

Power Law Exponent

The mean horizontal wind speed is zero at the earth's surface and increases with altitude in the atmospheric boundary layer. Wind data sets are usually collected at 10m hub height, thereby necessitating extrapolation of wind speed data for higher hub heights using the power index law. The wind speed extrapolation according to power law is given in Equation 11

$$v = v_o \left(\frac{H}{H_o}\right)^{\alpha} \quad [11]$$

where v is the wind speed at the turbine hub height H, vo is the wind speed at original height Ho and α is the surface roughness coefficient or empirical wind shear exponent. The values of α is taken as 1/7 for most sites with uniform terrain (Ahmed, 2016). Since wind speed varies with height, Weibull parameters c and k must be corrected for different hub heights. This can be determined using Frost Equation 12 and 13

$$k = k_r \left[\frac{1 - 0.088\ln\left(\frac{H_o}{10}\right)}{1 - 0.088\ln\left(\frac{H}{10}\right)}\right] \quad [12]$$

$$c = c_r \left(\frac{H}{H_o}\right)^{\alpha} \quad [13]$$

where krand cr represent the shape and scale parameters at higher hub heights

Maximum Energy Carrying Speed and Most Probable Wind Speed

The other parameters of utmost interest to wind resource assessors are the maximum energy carrying speed (vEmax) which is the wind speed carrying maximum wind energy and the most probable wind speed (vmp) which is the modal wind speed for the given wind distribution. These parameters can be estimated from the shape parameter k and scale parameter c using the given Equation 14 and 15

$$v_{Emax} = c\left(\frac{k+2}{k}\right)^{\frac{1}{k}} \quad [14]$$

$$v_{mp} = c\left(\frac{k-1}{k}\right)^{\frac{1}{k}} \quad [15]$$

Mean Power Output

The performance of the wind turbine can be examined by the mean power output. The mean electrical power output of a model wind turbine requires simulation of the equation given according to the Equation 16

$$P_e = \begin{cases} P_R \dfrac{v^k - v_c^k}{v_R^k - v_c^k} & v_c \le v \le v_R \\ P_R & v_R \le v \le v_F \\ 0 & v > v_F and v < v_C \end{cases} \quad [16]$$

where PR is the rated electrical power, vC is the cut-in wind speed, vR is the rated wind speed, vF is the cut-out speed and k is the shape parameter of the model wind turbine. Table 2 shows the technical specification of the model wind turbines used for this study which was retrieved from (http://en.windturbinesmodel.com/).

Table 2

Technical specifications and wind speed parameter of model wind turbines (http://en.windturbinesmodel.com/)

TECHNICAL PARAMETERS	WIND TURBINE MODELS			
	VENSYS 87	VESTAS V90	GAMESA G97	SUZZLON S88
Rated speed	12m/s	14m/s	12m/s	11.2m/s
Cut-in speed	3m/s	4m/s	3.5m/s	3.5m/s
Cut-out speed	22m/s	25m/s	25m/s	25m/s
Rotor swept area	5890m^2	6362m^2	7390m^2	6082m^2
Rotor diameter	87m	90m	90m	88m
Speed range	9-17.3rpm	9.9-18.4rpm	10-18rpm	7.8-15rpm
Rated power	1500KW	3000KW	2000KW	2100KW
Hub height	85m	80m	78m	80m

Capacity Factor

The capacity factor represents the fraction of the mean power output to the rated electrical power of the turbine. It can be used to predict wind turbine energy production and its economic feasibility. The capacity factor can be estimated from the Equation 17

$$C_f = \frac{P_e}{P_R} \quad [17]$$

whereP_e and P_R are the mean power output and the rated electrical power of the wind turbine respectively.

Annual Energy Production

The annual energy production for a particular site is the total energy output of the wind turbine considering the power output at any wind speed and frequency of occurrence of that wind in a year. The accumulated annual energy output of a wind turbine can be estimated from the given Equation 18

$$E_c = P_e \times 8760(\mathrm{KWh}) \quad [18]$$

whereP_e is as defined above and the constant (8760) represents the total number of hours in the year.

Energy Cost

There are two notable methods used for estimating the generation cost of unit energy (1kilowatt-hour). They are the levelized cost of electricity (LCOE) method and the present value cost (PVC) method. Considering the instability of the Nigeria economy, the present value cost method was most suitable for this study. The cost of a kilowatt-hour of wind-generated electricity was estimated using the given Equation 19

$$C = \frac{C_I}{8760n}\left(\frac{1}{P_R C_F}\right)\left[1 + m\left\{\frac{(1+I)^n - 1}{I(1+I)^n}\right\}\right] \quad [19]$$

where CI= initial investment on the project, PR = rated power of the turbine, CF = capacity factor, m = annual operation and maintenance cost, I = interest rate and n = project life span.

The economic analysis of the model wind turbines was carried out based on the following assumptions;

(i) The useful life span of the wind turbine is assumed to be 20years

(ii) The interest rate and inflation rates were 20% and 16% respectively

(iii) Operating and maintenance cost was assumed to be 3.5% of the initial investment

(iv) Turbines cost is assumed to be $Million per Mega-Watt.

RESULT AND DISCUSSION

The thirty-one year's monthly mean wind speed data (1980-2010) obtained from NIMET was analyzed to obtain the mean wind speeds across months and seasons. Matlab computer programme was used for the estimation of the weibull shape (k) and scale (c) parameters based on the maximum likelihood technique (MLT). The result shown in Tables 3 was obtained from Weibull model while result shown in Table 4 was obtained from power simulation of the model wind turbine and cost analysis using the present value cost (PVC) method.

Wind Speed Characteristic at Ikeja

The variation of monthly mean wind speed for thirty-one years at different hub heights is shown in Figure 1. It could be observed from the plot that the value of mean wind speed varied with months and seasons of the year. It also showed that the magnitude of wind speed increased with elevation (hub heights). At the hub height of 10m, the monthly mean wind speed ranged between the minimum value of 3.47m/s and a maximum of 5.33m/s with annual average value of 4.5m/s.. The wind speed had a seasonal average value of 4.09m/s in dry season and 4.65m/s during the rainy season. It was observed that the monthly mean speed exceeded 5m/s in the months of March, April, July and August while it ranged between 4m/s and 4.9m/s in January, February, May, June and September.

Table 3

Results of Weibull based parameters estimated at 10m hub height

Period	V_{data}(m/)	$V_{weibull}$	C(m/s)	K	WPD(W/m^2)	TI	Vemax(m/s)	Vmp(m/s)
January	4.13	4.14	4.56	4.07	100.473	0.285714	5.030568	4.254783
February	4.53	4.52	5	3.8	136.4474	0.275938	5.588529	4.613904
March	5.07	5.06	5.61	3.65	196.2218	0.285996	6.323401	5.138872
April	5.08	5.07	5.57	4.39	177.4497	0.244094	6.06727	5.251488
May	4.32	4.32	4.75	4.26	111.4135	0.263889	5.199172	4.460859
June	4.41	4.42	4.89	3.76	128.2376	0.301587	5.477387	4.50398
July	5.07	5.07	5.59	4.06	185.2866	0.274162	6.169583	5.21392
August	5.33	5.31	5.89	3.67	226.5351	0.283302	6.631188	5.40096
September	4.51	4.51	5	3.67	138.5797	0.299335	5.629192	4.584855
October	3.83	3.82	4.21	4.08	78.98584	0.261097	4.642412	3.929645
November	3.47	3.47	3.84	3.75	62.17217	0.288184	4.303624	3.535181
December	3.73	3.73	4.14	3.66	78.76332	0.308311	4.663708	3.7943
Whole Year	4.46	4.5	4.95	3.45	138.3203	0.307175	5.651502	4.48247
Dry Season	4.09	4.18	4.66	3.32	117.5078	0.332518	5.371118	4.183149
Rainy Season	4.65	4.64	5.15	3.62	152.3705	0.286022	5.81537	4.709998
1980	5.5	5.53	5.95	6.18	190.8688	0.223636	6.226156	5.78246
1981	4.75	4.74	5.12	5.77	124.441	0.189474	5.391002	4.95387
1982	5.03	5.04	5.37	7.41	132.7044	0.163022	5.545983	5.265961
1983	3.85	3.85	4.17	5.44	68.62487	0.207792	4.417039	4.017167
1984	3	3	3.35	3.13	44.91386	0.346667	3.922821	2.962358

Table 3 *(Continued)*

Period	V_{data}(m/)	$V_{weibull}$	C(m/s)	K	WPD(W/m^2)	TI	Vemax(m/s)	Vmp(m/s)
1985	3.17	3.19	3.53	3.73	48.4128	0.29653	3.960593	3.246648
1986	2.75	2.78	2.92	10.1	19.69828	0.163636	2.972703	2.890012
1987	3.03	3.01	3.08	24.45	20.00993	0.036304	3.089921	3.074744
1988	4.21	4.22	4.66	3.85	109.8291	0.301663	5.194927	4.309826
1989	4.28	4.29	4.67	5.04	99.10738	0.238318	4.990165	4.469506
1990	3.42	3.42	3.73	4.9	51.03726	0.24269	3.999873	3.560229
1991	2.93	2.94	3.22	4.34	34.44322	0.276451	3.51384	3.031431
1992	3.47	3.48	3.77	5.68	49.94882	0.216138	3.97564	3.643633
1993	4.04	4.06	4.35	6.7	72.69333	0.178218	4.522945	4.24631
1994	4.83	4.84	5.21	6.17	128.2116	0.190476	5.452562	5.062806
1995	4.97	4.98	5.24	9.82	114.5777	0.132797	5.339855	5.183003
1996	4.71	4.72	4.95	10.95	94.25524	0.123142	5.026419	4.906897
1997	4.33	4.33	4.7	5.19	99.94014	0.2194	5.00465	4.510117
1998	4.58	4.58	4.88	7.33	99.90464	0.155022	5.043293	4.78332
1999	5.33	5.32	5.78	5.28	184.7182	0.213884	6.142537	5.554654
2000	6.29	6.29	6.71	7.42	258.7982	0.163752	6.929328	6.58036
2001	6.93	6.91	7.39	7	351.6941	0.15873	7.660136	7.22904
2002	6.69	6.69	7.35	4.34	409.6361	0.284006	8.020723	6.919572
2003	4.23	4.24	4.5	8.32	75.62937	0.165485	4.618036	4.431271
2004	3.67	3.66	3.96	5.6	58.17358	0.19891	4.181945	3.823312

Table 3 *(Continued)*

Period	V_{data}(m/)	$V_{weibull}$	C(m/s)	K	WPD(W/m^2)	TI	Vemax(m/s)	Vmp(m/s)
2005	3.87	3.85	4.11	7.52	59.2452	0.147287	4.240933	4.032748
2006	4.13	4.13	4.4	7.48	72.80271	0.150121	4.541614	4.316386
2007	4.17	4.16	4.53	5	90.72864	0.22542	4.845335	4.332277
2008	5.75	5.74	6.08	8.38	186.1825	0.130435	6.237289	5.988498
2009	5	5	5.31	7.98	125.6646	0.148	5.460921	5.221651
2010	5.25	5.23	5.47	11.45	125.9953	0.085714	5.547453	5.426515

Table 4

Result of power output simulation of model turbines and energy cost per kilo-watt-hour90m hub heights

Period	V	C	K	Mean Power Output (KW)				Capacity Factor			
				VESTAS V90	GAMESA G97	SUZZLON S88	VENSYS 87	VESTAS V90	GAMESA G97	SUZZLON S88	VENSYS 87
January	5.65	6.24	5.04	25.59	40.99	61.01	32.33	0.008	0.020	0.029	0.021
February	6.20	6.84	4.71	56.73	83.50	121.49	64.86	0.018	0.041	0.057	0.043
March	6.94	7.68	4.52	115.51	161.05	231.39	123.4	0.038	0.080	0.110	0.082
April	6.96	7.62	5.44	63.423	100.46	153.64	76.34	0.021	0.050	0.073	0.050

Table 4 *(Continued)*

Period	V	C	K	Mean Power Output (KW)				Capacity Factor			
				VESTAS V90	GAMESA G97	SUZZLON S88	VENSYS 87	VESTAS V90	GAMESA G97	SUZZLON S88	VENSYS 87
May	5.91	6.50	5.28	27.71	44.773	67.72	34.80	0.009	0.022	0.032	0.023
June	6.04	6.69	4.66	50.93	75.234	109.09	58.80	0.016	0.037	0.051	0.039
July	6.94	7.65	5.08	82.51	123.41	183.54	94.10	0.027	0.061	0.087	0.062
August	7.30	8.06	4.54	145.25	201.50	289.99	153.63	0.048	0.100	0.138	0.102
September	6.17	6.84	4.54	62.56	90.303	129.96	70.38	0.020	0.045	0.061	0.046
October	5.24	5.76	5.05	15.61	26.49	39.47	21.45	0.005	0.013	0.018	0.014
November	4.75	5.25	4.64	10.89	20.49	29.69	17.84	0.003	0.010	0.014	0.011
December	5.10	5.66	4.53	20.77	34.12	49.06	28.36	0.006	0.017	0.023	0.018
Whole Year	6.10	6.77	4.27	72.50	101.45	143.35	79.64	0.024	0.050	0.068	0.053
Dry Season	5.59	6.38	4.11	52.07	74.75	104.48	60.33	0.017	0.037	0.049	0.040
Rainy Season	6.36	7.05	4.48	76.79	108.45	155.	84.43	0.025	0.054	0.074	0.056

Table 4 *(Continued)*

Period	Energy Cost ($ per Kilo-Watt-hour)			
	VESTAS V90	GAMESA G97	SUZZLON S88	VENSYS 87
January	1.018	0.423	0.298	0.402
February	0.459	0.208	0.150	0.200
March	0.225	0.107	0.078	0.105
April	0.410	0.172	0.118	0.170
May	0.940	0.387	0.269	0.374
June	0.511	0.230	0.167	0.221
July	0.315	0.140	0.099	0.138
August	0.179	0.086	0.062	0.084
September	0.416	0.192	0.140	0.185
October	1.669	0.655	0.462	0.607
November	2.390	0.847	0.614	0.729
December	1.254	0.509	0.371	0.459
Whole Year	0.359	0.171	0.127	0.163
Dry Season	0.500	0.232	0.174	0.215
Rainy Season	0.339	0.159	0.116	0.154

The monthly mean wind speed was less than 4m/s in October, November and December. This suggested that the potential of wind energy is fair in March, April, July and August, marginal in the months of February, May, June and September and poor in the months of January, October, November and December.

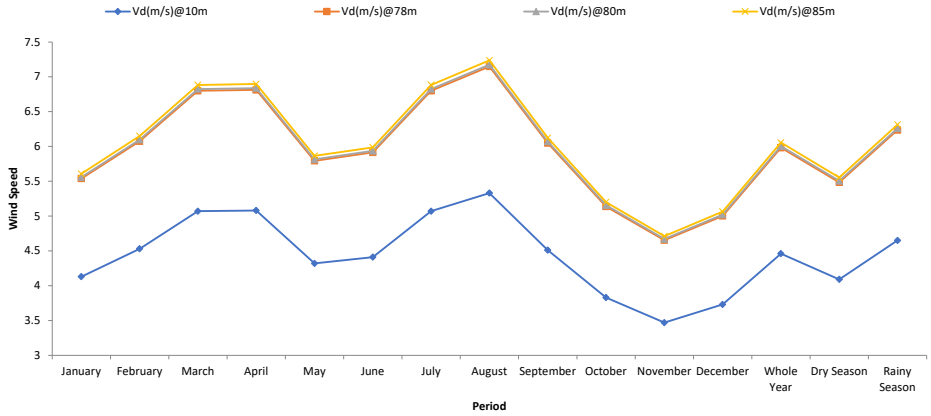


Figure1. Monthly and seasonal variation of average wind speed (1980–2010) at different heights

Wind Speed Frequency Distribution

Wind speed frequency distribution shows the cumulative time the wind blows at prescribed value or range of values. It is a relative indicator of energy input to a wind turbine. The frequency of occurrence of a particular wind speed range is critical information required in wind resource assessment. Figure 2 shows the wind speed frequency distribution of Ikeja based on the thirty-one year's monthly mean wind speed. The frequency distribution showed that there were three dominant wind speed ranges with a percentage frequency higher than 20%. These included ranges of 3-3.9m/s, 4-4.9m/s, and 5-5.9m/s. The wind speed range with the highest frequency of occurrence of 27% was between the range 3-3.9m/s. The plot also indicated that wind speed above 3m/s occurred with a cumulative frequency of over 70%. Since the cut-in speed of most modern wind turbine is between 3m/s to 3.5m/s, therefore, it can be inferred that most modern wind turbine will generate electricity regularly at Ikeja location.

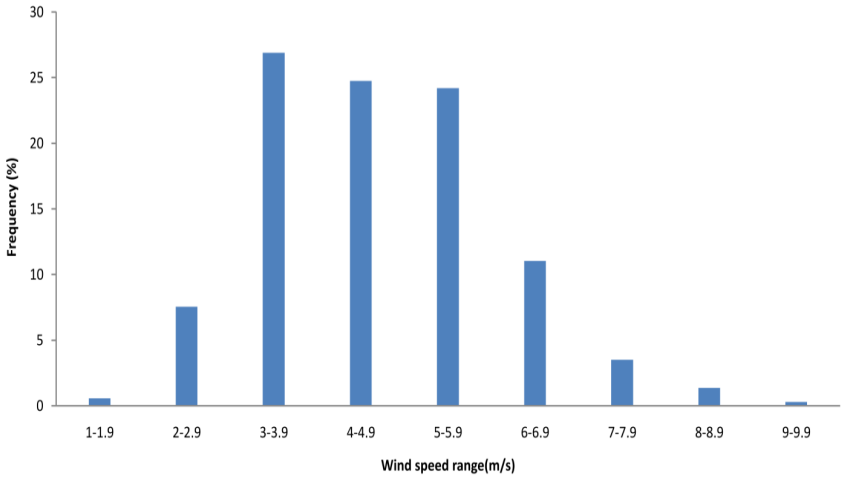


Figure 2. Wind speed frequency distribution for Ikeja (1980-2010)

Wind Turbulence Intensity

Figure 3 shows the plot of yearly variation of wind turbulence intensity. The plot reveals that the turbulence intensity varies with year. The turbulence intensity ranges between low values of 0.03 and a high value of 0.34 within the thirty-one years considered. High-level turbulence (TI > 0.25) was observed in the years 1985, 1986, 1989, 1992, 2002 and 2003 while low-level turbulence (<0.1) was observed in the year 1987. The turbulence intensity of the remaining years of 1980, 1981, 1982, 1983, 1984 , 1988, 1990, 1991, 1993, 1994,

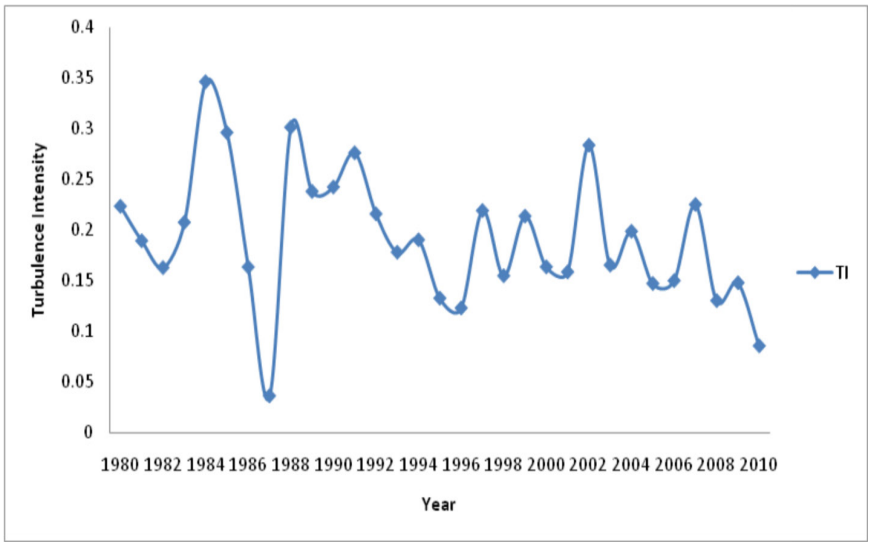


Figure 3. Yearly variation of wind turbulence intensity (1980 – 2010)

1995, 1996, 1997, 1998, 1999, 2000, 2001, 2004, 2005, 2006, 2007, 2008, 2009 and 2010 were moderate (<0.25). Based on the thirty-one years average of 0.193, the turbulence level of Ikeja location can be said to be moderate and suitable for wind energy generation without overloading the wind turbine components

Performance Evaluation of Weibull Model

The adequacy and reliability of the Weibull based model used in this study was assessed using statistical test tools. These include the correlation coefficient between the actual wind speed and Weibull predicted wind speed, the root mean square error, chi-square test and the coefficient of efficiency of the Weibull model. The plot of monthly and seasonal variation of actual wind speed and Weibull predicted wind speed is as shown in Figure 4. The correlation coefficient between the actual data and weibull predicted data was determined to be 0.99. This implies that the Weibull model produces a good fit to the actual data. The chi-square test gave 100% while the root means square error was 0.0268. Also, the coefficient of efficiency of the Weibull model was determined to be 0.0026. The results obtained shows that there is a high level of agreement between the actual data and the Weibull predicted data with minima error. Therefore, it can be concluded that the 2-parameter Weibull model can adequately and reliably predict the wind situation at Ikeja location.

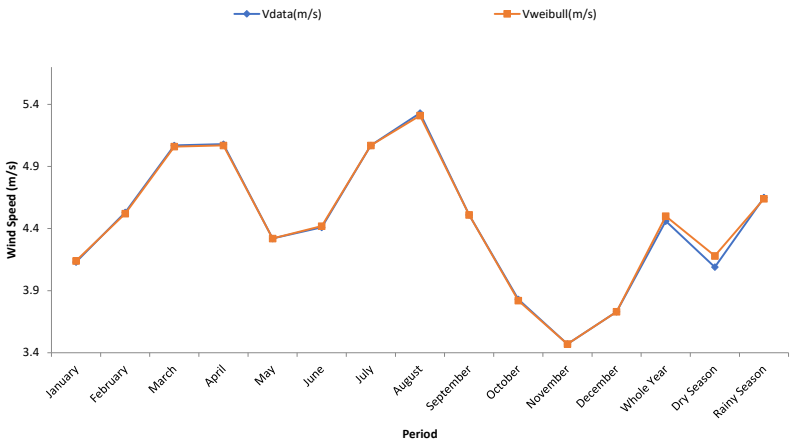


Figure 4. Monthly and seasonal variation of actual data and weibul predicted mean wind speed

Wind Power Density

Figure 5 shows the monthly and seasonal variation of wind power density estimated at hub heights of 10m, 50m. It could be observed from the plot that wind power density changed with months, season and hub heights. Monthly variation showed two peak points in March and August while minimum values of WPD were observed in October, November, December and January and May. At 50m height which was the reference height for wind power classification (Table 2), the WPD ranged between 116.3 W/m^2 and 423.3W/m^2 with annual average value of 257.85W/m^2. According to the international wind power classification shown in Table 2, the wind energy potential of the study area can therefore be remarkably described as fair in the months of March, April, and July belonging to power class 3. The wind potential is good in August belonging to power class 4 while it is marginal in February, May, June and September belonging to power class 2. The wind potential is remarkably poor in the months of January, October, November and December belonging to power class 1. Seasonally, the wind energy potential is marginal in rainy season belonging to power class 2 and poor during dry season belonging to power class 1. Based on the whole year assessment the wind potential at Ikeja is marginal belonging to power class 2. The result indicated that Ikeja location will be most suitable for standalone small to medium scale wind electricity generation.

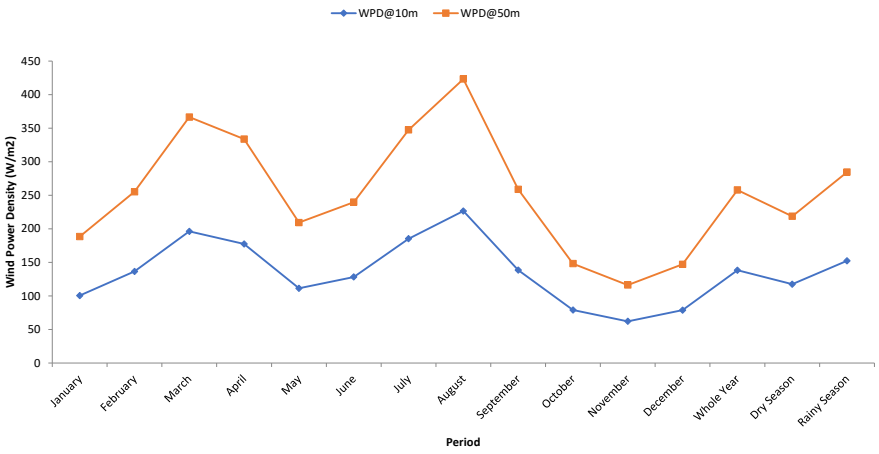


Figure 5. Monthly and seasonal variation of wind power density at different hub height

Maximum Energy Carrying Wind Speed and Most Probable Wind Speed

Figure 6 shows the monthly and seasonal variation of maximum energy carrying wind speed and the most probable wind speed estimated at different hub heights. The closer the value of these parameters to the rated speed (VR) of the wind turbine, the higher will be the conversion efficiency of the turbine. The technical specification presented in Table 1 showed that model turbine Suzzlon S88 had the lowest rated speed of 11.2 m/s which was relatively closer to the two speed parameters. This implies that the turbine will likely perform better than the other wind turbines considered. It could also be observed from the plot that the two speed parameters varied with months and seasons of the year. The values of maximum energy carrying wind speed and the most probable wind speed increased with increase in elevation or hub heights. At 90m hub height, the most probable wind speeds varied between 4.99 m/s and 7.64m/s with an annual average value of 6.37m/s. There were four months with values of most probable wind speed exceeding 7.0m/s. These are the months of March, April, July, and August representing months within the period of the rainy season. This implies that the wind electricity generation potentials will be maxima within those months. Also, there were four months with the value of most probable wind speed less than 6.0m/s at 90m hub height. These include the months of October, November, December and January, which are the first four months of the dry season. This observation likewise implies low wind electricity generation potential within those months. Seasonally, it could be observed that the value of most probable wind speed was higher during the rainy season compared to the dry season. This implies higher wind electricity generation potential during the rainy season compared to the dry season. The variation of maximum energy carrying wind speed follows similar trend as that of most probable wind speed. The

values of maximum energy carrying wind speed at 90m height varied between 5.68m/s and 8.74m/s with an annual average of 7.4m/s. Those months with values exceeding 8m/s were March, April, July, and August while values below 7m/s were obtained in the months of October, November, December and January and May. This observation likewise implies higher wind electricity potential during rainy season compared to the dry season.

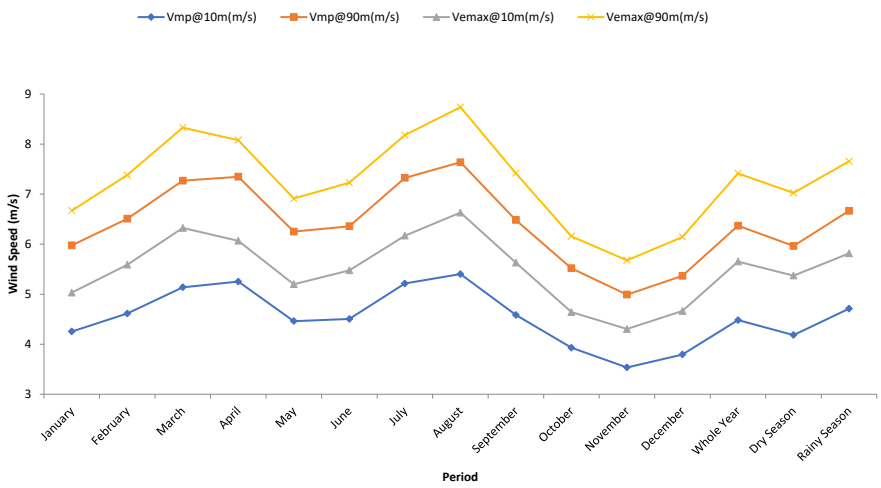


Figure 6. Monthly and seasonal variation of maximum energy carrying wind speed and most probable wind speed at different hub heights

Mean Power Output and Capacity Factor

The monthly and seasonal variation of mean power output was estimated based on the technical specifications and speed parameters of the model turbines highlighted in Table 2. The plot in Figure 7 shows that the mean power output and the capacity factor of the model turbines vary with months and season of the year. It could be observed from the plot that the variation of mean power output and capacity factor exhibited a similar trend. They both attained maximum values in March and August. They both dropoed to a minimum in October, November, December, January and May which are mostly months of the dry season except May which is month within the rainy season. The mean electrical power output from the model turbine varied between 289.99KW and 10.9KW while the capacity factors varied between 13.8% and 0.36%. Seasonally, the mean power output and the capacity factor of the model turbine were higher during the rainy season compared to the dry season. This implies that the wind turbines will generate more electrical power during the rainy season compare to the dry season. The turbine model Suzzlon S88 performed best under the wind regime having maximum mean power output and highest conversion efficiency.

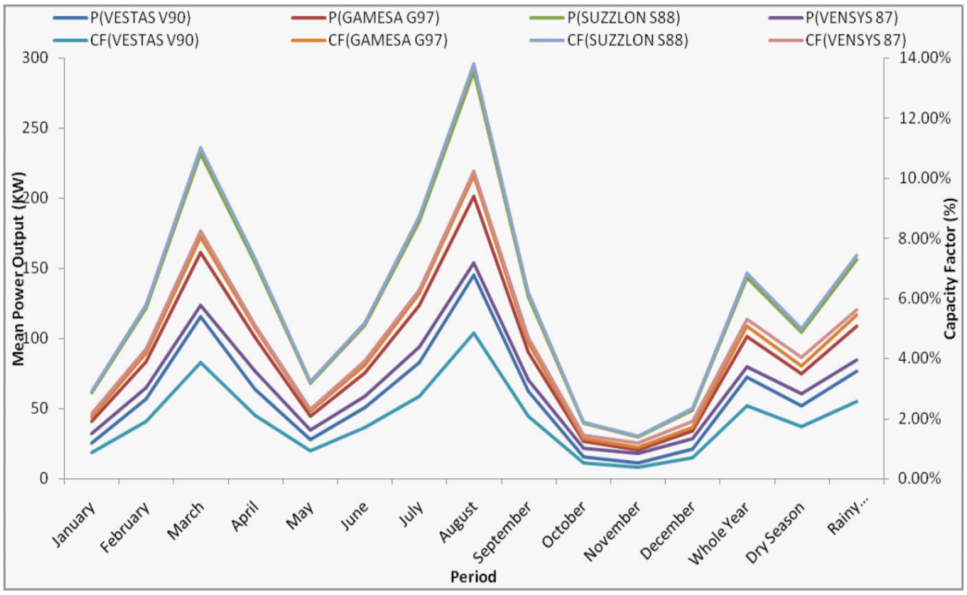


Figure 7. Monthly and seasonal variation of mean power output and capacity factor of the four model wind turbine

Total Annual Energy Production

The plot of annual energy production for the four different models of wind turbines is shown in Figure 8. It could be observed that the annual energy production differ for the four turbines models considered. The turbine model Suzzlon S88 was found to produce the highest total annual energy of 1,255,772KWh; this was followed by Gamesa G97 with annual production of 888,787KWh. Vensys 87 had total annual energy production of 697,652.7KWh while Vestas V90 turbine produced 635,126 KWh of electrical energy annually. Wind turbine model Suzzlon S88 performed best with highest annual energy production.

Energy Cost

The monthly and seasonal variation of energy cost per kilowatt-hour for the four turbine model considered is shown in Figure 9. It could be seen from the plot that the cost of generation of wind electricity per kilowatt-hour was inconsistent but varies with months and seasons of the year. The cost of wind electricity generation ranged between a minimum of $0.11/KWh and maximum of $2.39/KWh. The estimated cost of generation is higher than $0.5/KWh in the months of January, May, October, November and December which are mostly months within the period of the dry session with the exception of May. The electricity generation cost is less than $0.2/KWh in March, July and August which are

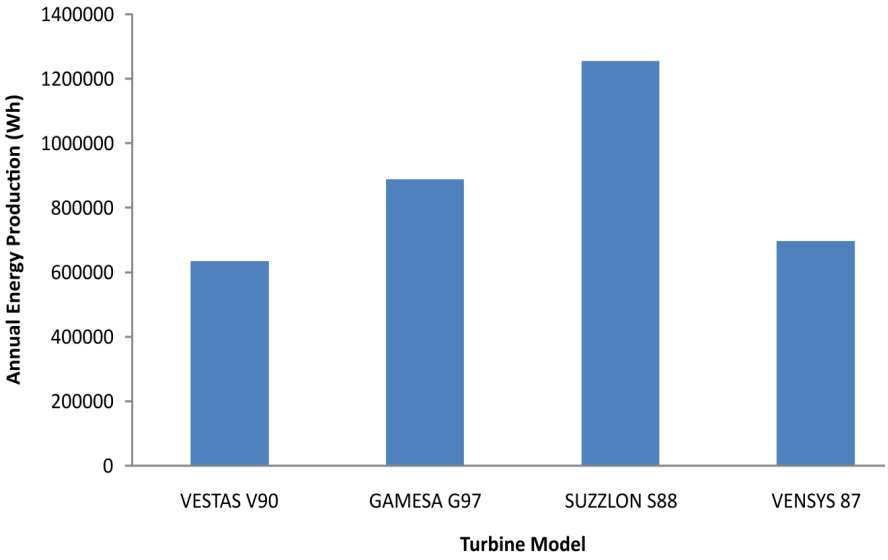


Figure 8. Annual energy production of the four model wind turbine

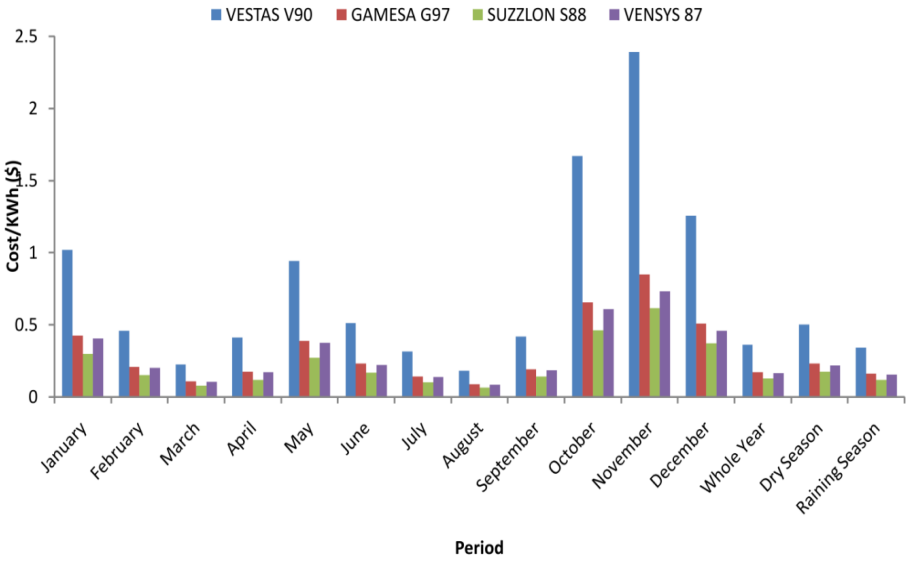


Figure 9. Monthly and seasonal estimate of energy cost ($/kWh) of four wind turbine models

months within the the rainy season. The Ikeja electricity tariff presently ranges between $0.07/KWh and $0.12KWh. Comparing the estimated energy cost with electricity tariff showed that wind electricity is yet to cost effective year round. Though the energy cost was lower than electricity tariff during some months of the year, the economic feasibility of wind energy development depends on its ability to generate electricity at a low operating cost throughout the year. The seasonal variation showed that the cost of wind electricity generation was higher during dry season compared to the rainy season. This indicates that the cost of wind electricity is relatively cheaper during the rainy season compared to the dry season. Improvement in the country economic situation and government intervention in the form of green subsidy, tax waiver and reduced interest on the loan, will go a long way in reducing the generation cost of wind electricity. Also, Nigeria electricity tariff is expected to increase biannually for years to come as contained in the country Multi-Year Tariff order 2.1 of 2015 (NERC, 2019). The expected increase in electricity tariff will further help in making the cost of wind electricity to be competitive with conventional electricity sources and subsequently, the wind energy might become economically feasible year round.

CONCLUSIONS

In this study, the wind electricity generation potential and its cost implication for Ikeja location were assessed using the two-parameter Weibull distribution function and PVC energy cost method. From the result obtained it can be concluded that;

i) The wind speed data shows good fit between actual data and Weibull predicted data confirming the adequacy and reliability of the Weibull model.

ii) The value of wind speed at 10m height ranges between 3.47m/s and 5.33m/s with annual average of 4.5m/s. While it has seasonal average value of 4.09m/s in dry season and 4.65m/s during the rainy season. While the Wind Power Density (WPD) at 50m hub height ranges between 116.3 W/m^2 and 423.3W/m^2 with annual average value of 257.85W/m^2.

iii) The location wind regime belongs to power class 2 with marginal potential for wind electricity generation.

iv) The location is suitable for small to medium scale stand-alone wind electricity generation.

v) The generation cost of wind electricity is not cost-effective year round when compared with the present Ikeja electricity tariff while the location is not economically feasible for large scale grid connected wind power generation.

RECOMMENDATION

Further study on the assessment of wind electricity generation potential of Lagos Island and the coast of Gulf of Guinea for comprehensive mapping is hereby recommended.

ACKNOWLEDGEMENTS

Our gratitude goes to the management of Tertiary Education Trust Fund (tetfund) for supporting this project. Also, special thanks to the management of Nigeria Meteorological Agency (NIMETs) for making their data available for this work.

REFERENCES

Adaramola, M. S., Paul, S. S., & Oyedepo, S. O. (2011). Assessment of electricity generation and energy cost of wind energy conversion systems in north-central Nigeria. *Energy Conversion and Management*, 52, 3363-3368.

Adebayo, C. (2014). *How is 100% renewable energy possible for Nigeria?* San Diego, California: Global Energy Network Institute.

Agbetuyi, A. F., Akinbulire, T. O., Abdulkareem, A., & Awosope, C. O. A. (2012). Wind energy potential in Nigeria. I*nternational Electrical Engineering Journal, 3*(1), 595-601.

Ahmed, S. (2016). *Wind energy: Theory and practice* (3rd Ed.). Delhi, India: PHI Learning Private Limited.

Ajayi, O. O. (2010). The potential for wind energy in Nigeria. *Wind Engineering, 34*(3), 303-312.

Ajayi, O. O. (2013). Sustainable energy development and environmental protection; The case of five West Africa Countries. *Renewable and Sustainable Energy, 26,* 532-539.

Ajayi, O. O., Fagbenle, R. O., Katende, J., Ndambuki, J. M., Omole, D. O., & Badejo, A. A. (2014). Wind energy study and energy cost of wind electricity generation in Nigeria: Past and recent results, a case study for South West Nigeria. *Energy,* 7, 8508-8534.

Akpinar, E. K., & Akpinar, S. (2005). A statistical analysis of wind speed data used in installation of wind energy conversion systems. *Energy Conservation and Management, 46,* 515-532.

Carta, J. A., Ramirez, P., & Velazquez, S. (2009). A review of wind speed probability distribution used in wind energy analysis: Case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews, 13*(5), 933-955.

ECN-UNDP. (2015). *Renewable energy master plan final draft report*. Energy Commission of Nigeria and United Nation Development Programme. Retrieved October 12, 2018, from http://www.icccdnigeria.org/REMP%20final%report.pdf/.

Ezema, I. C., Olotuah, A. O., & Fagbenle O. I. (2016). Evaluation of energy use in public housing in Lagos, Nigeria: Prospects for renewable energy sources. *International Journal of Renewable Energy Development, 5*(1), 15-24.

Fadare, D. A. (2008). Statistical analysis of wind energy potential in Ibadan, Nigeria, based on Weibull distribution function. *Pacific Journal of Science and Technology, 9*, 110-119.

Fagbenle, R. O., Katende, J., Ajayi, O. O., & Okeniyi, J. O. (2011). Assessment of wind energy potential of two sites in north-east Nigeria. *Renewable Energy, 36,* 1277-1283.

Federica, P. (2018). Pollution from fossil fuel combustion is the leading environmental threat to global pediatric health and equity; Solution exists. *International Journal of Environmental Research and Public Health (MDPI), 15*(1), 1-17.

Hermann, S. (2001). *A solar manifesto*. London, England: James and James LTD.

NERC. (2019). *Multi-Year Tariff Order (MYTO) for the determination of charges and Tariffs for electricity generation, transmission and retail tariffs.* Nigeria Electricity Regulatory Commission. Retrieved June 7, 2019, from http://www.nerc.org/index.php

Nkalo, U., & Agwu, O. (2019). Review of the impact of electricity supply on economic growth; A Nigeria case study. *IOSR Journal of Electrical and Electronics Engineering, 14*(1), 28-34.

Nze-Esiaga, N., & Okogbue, E. C. (2014). Assessment of wind energy potential as a power generation source in five locations of South Western Nigeria. *Journal of Power and Energy Engineering, 2,* 1-13.

Odunola, O. O., Odunsi, O. M., Kasim, O. F., & Alabi, A. T. (2018). Implication of fossil fuel generating set on resident wellbeing in Lagos, Nigeria. *African Journal of Psychological Study of Social Issues, 21*(2), 253-265.

Okeniyi, J. O., Ohunakin, O. S., & Okeniyi, E. T. (2015). Assessment of wind energy potential in selected sites from three geopolitical zones in Nigeria: Implication for renewable/ sustainable rural electrification. *The Scientific World Journal, 2015,* 1-13.

Oladipo, K., Agbetuyi, A. F., Owolabi, B., Obiakor, C., & Fagbuaro, O. (2018). Power sector reform in Nigeria; Challenges and solutions. *IOP Publishing, Material Science and Engineering, 4*(13), 12-37.

Oyewole, J. A., & Aro, T. O. (2018). Wind speed pattern in Nigeria: A case study of some coastal and inland areas. *Journal of Applied Science Environment Management, 22*(1), 119-123.

Uzoma, C. C., Nnaji, C. E., & Nnaji, M. (2014). The role of energy mix in sustainable development of Nigeria. *Continental Journal of Social Science, 5*(1), 21-29.

Pertanika J. Sci. & Technol. 28 (4): 1469 - 1486 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

A Multi-State Model for Reliability Analysis of Metal Sheet Manufacturing Process using Artificial Neural Network Technique

Anil Chandra[1], Surbhi Gupta[1]* and Chandra Kant Jaggi[2]

[1]*Amity Institute of Applied Sciences, Amity University Uttar Pradesh, 201313 Noida, India*
[2]*Department of Operational Research, University of Delhi, 110007 Delhi, India*

ABSTRACT

A manufacturing system is governed by its various processes upon which its efficiency is dependent. Since, failure results in considerable losses, many manufacturing systems have certain redundancies for some processes. These redundancies cause the system to work under different efficiency states called multi-state elements. In this paper various processes of metal sheet manufacturing unit have been categorized as subsystems to determine the multi-state probabilities of its different efficiency states. Artificial Neural Network Technique (ANN) has been used to estimate the change in these multi-state probabilities over time. The ANN has also been used to estimate variation in upstate and downstate probabilities of the system for a particular-time period. The results have been used to determine variation in profit over time for the system.

Keywords: Artificial neural network, downstate, metal sheet manufacturing, reliability, state transition, upstate

ARTICLE INFO

Article history:
Received: 5 April 2020
Accepted: 2 June 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.18

E-mail addresses:
achandra@amity.edu (Anil Chandra)
sgupta11@amity.edu (Surbhi Gupta)
ckjaggi@yahoo.com (Chandra Kant Jaggi)
*Corresponding author

INTRODUCTION

A well-established industrial process is designed to provide the best quality product within optimum cost. However, these processes are susceptible to failures due to various reasons. Most of the industrial processes are designed to accommodate parallel redundancy with reduced functionality of plant so as to minimize the losses during corrective maintenance of the process. This leads to the whole system

ISSN: 0128-7680
e-ISSN: 2231-8526

being working under different efficiency states. These states are referred to as multi-state elements (MSE) and such a system is known as multi-state system (MSS). The conventional methods for reliability analysis assume a system to be a binary-state system (BSS) i.e. it has only two states – perfectly working state and failed state (Rausand & Høyland, 2004). However, most of the real-world systems are complex and they fall under the classification of MSS as they undergo many levels of degradation states between perfectly working state and a failed state (Natvig, 2011). Thus studying industrial processes, like metal sheet manufacturing process, as an MSS rather than a BSS is useful for practical assessment of its profitability and reliability. Furthermore, predicting the probability of these multi-states is essential to make many important decisions from selecting appropriate maintenance strategy to other measures for reducing downtime probability. The study on reliability assessment of MSS has its history since mid 1970s wherein, the fundamental concepts were introduced (Murchland, 1975) and Boolean methods extension technique was used for reliability modeling of such systems (Barlow & Wu, 1978). Since then researchers have studied various methods for reliability modeling of MSS including Multiple Valued Logic (MLV) (Zaitseva & Levashenko, 2017) universal generating function (UGF), Markov and Semi-Markov Processes (Lisnianski & Levitin, 2003; Lisnianski et al., 2010). While MLV is an extension of Boolean method, in UGF the distribution of performance output of system is obtained on the basis of performance distribution of its elements. Markov and Semi-Markov process modeling analyze the reliability of MSS under assumption that failure and repair times are exponentially distributed (Lisnianski et al., 2012; Li et al., 2018b; Liu et al., 2014; Fang et al., 2016). Researchers have also used a combination of Markov Model with dynamic Bayesian Network for reliability assessment of MSS (Alyson & Aparna, 2007; Li et al., 2018a).

The modeling technique of MSS mentioned above has its own usages and limitations. These techniques have certain pre-assumptions regarding statistical distributions of various states, failure and repair rates and are generally used to ascertain the steady state behavior of the system. However, a metal sheet manufacturing plant is a high demand industrial process as it is required to be in continuous working state for most of its useful life to meet the requirements of industry. Due to which it has a tendency to reach its deteriorating phase quite rapidly and failures no longer follow any particular distribution. This makes the relationship between each of the parameters viz., failure rate, repair rate and states with respect to time more complex. Additionally, to keep the system in profitably working state, the repairs, replacements and maintenance have to be optimized quite frequently duly considering those complexities. Artificial Neural Network (ANN) Models are capable to learn and model such complex real-life systems. On the basis of observed industrial data, the ANN can not only model but also predict the future states by understanding the levels of adjustment in the weights assigned (in this case failure and repair rates) to its neurons.

Furthermore, no prior assumption, especially regarding the distributions of failure/repair rates is required to model a system using ANN. Now-a-days ANN modeling computations are convenient and yield more precise estimates due to availability of a variety of software and high processing speeds of modern computers. Due to these reasons ANN has wide range of applications in many fields including that of reliability (Karunanithi et al., 1992) and, in recent times, ANN models are gaining popularity for reliability estimation of systems (Hurtado & Alvarez, 2001; Sharma et al., 2016; Reshid et al., 2017; Bhargava & Handa, 2018; Chandra et al., 2019).

Evaluating Steady State Probabilities are useful for evaluating reliability of a system for a considerably long period of time. However, an industrial system, like metal-sheet manufacturing plant, which is under high demand and susceptible to deterioration in short span of time, short term state probability analysis is preferred to steady state probabilities.

Considering above-mentioned factors, the objective of this study is to understand short-term behavior of various state probabilities of metal sheet manufacturing plant, using ANN model, in its two phases, (i) "Useful life" during which failure rates of its sub-systems remain constant with time and (ii) "Wear-out phase" or "deterioration period" during which failure rates of its sub-systems start to increase with time. The effect of these variations on profitability of system, under two specific types of preventive maintenance has also been discussed.

Metal sheet production Industry has a global market size of 265 billion US dollars with expected increase of 5% per year (Grand View Research, 2020). It has application in important industrial sectors like automobile, railways, construction and machinery (Kozaki et al., 2017). The workers generally work in an eight to ten hour shift, with machines being run for 24 X 7, to meet this ever-increasing demand. As per report of World Steel Association, India is second highest producer of raw steel after China (Angel, 2019). In the financial year 2019, India produced 82.4 million tonnes of finished steel product (Bhati, 2019). Due to ever increasing demand of metal sheet the evaluation of reliability of metal sheet manufacturing process becomes equally important.

The metal sheet manufacturing process consists of different processes and these processes can be put under the following operations (Bhattacharyya, 1997; Kalpakjian & Schmid, 2001; de Sousa, 2016):

a. Cutting, in which various techniques like roll forming, shearing, blanking, fine blanking and punching are used to cut metal sheets as per requirement

b. Bending, in which various techniques like, V-shaped, edge bending, Shearing etc. are used to bend the metal as per requirement

c. Drawing, in which trimming and slitting of sheets are done to make convex or concave shapes

d. Dye and Coloring of final product

MATERIALS AND METHODS

System Configuration

Considering the processes as subsystems for reliability estimation, the series-parallel configuration for successful operation of metal sheet manufacturing is given in the form of a block diagram given in Figure 1 and its state transition diagram is given in Figure 2.

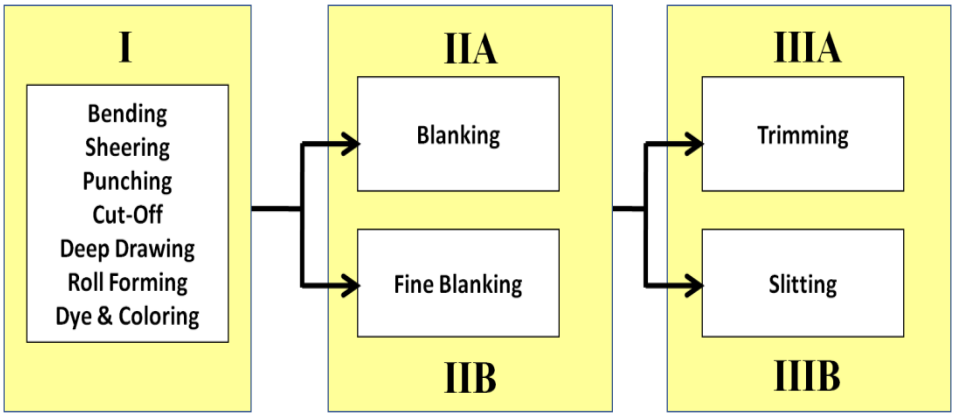


Figure 1. Block Diagram for Successful Operation of Metal Sheet Manufacturing Plant

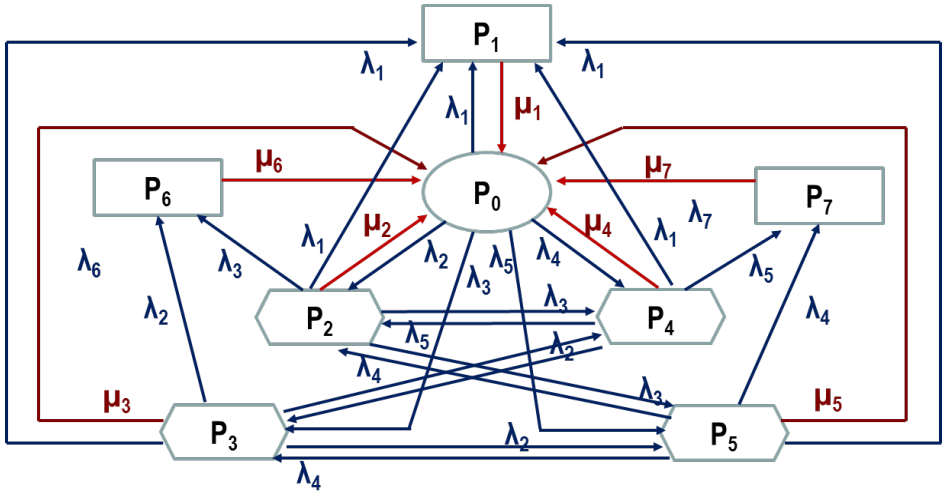


Figure 2. State Transition Diagram of the System

The subsystems are:
I = Bending, Sheering, Punching, Cut-Off, Deep Drawing, Roll Forming, Dye & Coloring
IIA= Blanking
IIB = Fine Blanking

IIIA = Trimming
IIIB = Slitting
The state probabilities are
P_0 = All subsystems are in working state
P_1 = Subsystem I fails and system fails
P_2 = Subsystem IIA fails and system works with degraded efficiency
P_3 = Subsystem IIB fails and system works with degraded efficiency
P_4 = Subsystem IIIA fails and system works with degraded efficiency
P_5 = Subsystem IIIB fails and system works with degraded efficiency
P_6 = Subsystem IIA &IIB fails and system fails
P_7 = Subsystem IIIA & IIIB fails and system fails

Notations

λ_1 = Failure Rate of subsystem I
λ_2 = Failure Rate of subsystem IIA
λ_3 = Failure Rate of subsystem IIB
λ_4 = Failure Rate of subsystem IIIA
λ_5 = Failure Rate of subsystem IIIB
λ_6 = Failure Rate of subsystem IIA & IIB
λ_7 = Failure Rate of subsystem IIIA & IIIB
μ_1 = Repair Rate of subsystem I
μ_2 = Repair Rate of subsystem IIB
μ_3 = Repair Rate of subsystem IIB
μ_4 = Repair Rate of subsystem IIIA
μ_5 = Repair Rate of subsystem IIIB
μ_6 = Repair Rate of subsystem IIA & IIB
μ_7 = Repair Rate of subsystem IIIA & IIIB

Assumptions

The following assumptions have been associated with this model

i. Initially, the state probabilities are known
ii. The states of all processes are statistically independent
iii. Failure of each process follows arbitrary failure time
iv. Repair facility is available and Repair Rates are constant
v. Maintenance Strategy adopted is Corrective Type, unless specified otherwise
vi. No defect is caused in the sheets during shifting from one process to the other

Formulation of Artificial Neural Network (ANN) Model

When the metal sheet manufacturing system is in downstate, due to failure in any one of the processes, the losses incurred are high (Tang et al., 2007). Due to this reason the availability of repair facility and formulating appropriate maintenance strategy becomes equally important.

ANN Model of the System

The ANN is designed to mimic functioning of human brain. Input data and target output data is fed into the ANN model. The model trains itself, by reducing error, after certain number of trials (called epochs) to give the best estimates for output data.

The proposed ANN model contains an input layer, a hidden layer and output layer.

Input Layer. Input are defined as Equation 1

$$Xi = Pi(t), i = 1, \ldots 7 \quad [1]$$

Neurons. The numbers of neurons are equal to number of states given in transition diagram given in Figure 4. The neural weights of each state are assigned according to the failure or repair rates of that particular state as per details given below:

The weights of neural network are (Equation 2-23):

$$W_{01} = W_{21} = W_{31} = W_{41} = W_{51} = \lambda_1 \Delta t \quad [2]$$

$$W_{02} = W_{34} = W_{35} = W_{36} = \lambda_2 \Delta t \quad [3]$$

$$W_{03} = W_{24} = W_{26} = \lambda_3 \Delta t \quad [4]$$

$$W_{04} = W_{52} = W_{53} = W_{57} = \lambda_4 \Delta t \quad [5]$$

$$W_{05} = W_{25} = W_{42} = W_{43} = W_{47} = \lambda_5 \Delta t \quad [6]$$

$$W_{06} = \lambda_6 \Delta t \quad [7]$$

$$W_{07} = \lambda_7 \Delta t \quad [8]$$

$$W_{10} = \mu_1 \Delta t \quad [9]$$

$$W_{20} = \mu_2 \Delta t \quad [10]$$

$$W_{30} = \mu_3 \Delta t \quad [11]$$

$$W_{40} = \mu_4 \Delta t \quad [12]$$

$$W_{50} = \mu_5 \Delta t \quad [13]$$

$$W_{60} = \mu_6 \Delta t \quad [14]$$

$$W_{70} = \mu_7 \Delta t \quad [15]$$

$$W_{00} = 1 - W_{01} - W_{02} - W_{03} - W_{04} - W_{05} - W_{06} - W_{07} \quad [16]$$

$$W_{11} = 1 - W_{10} \quad [17]$$

$$W_{22} = 1 - W_{21} - W_{24} - W_{25} - W_{26} - W_{20} \quad [18]$$

$$W_{33} = 1 - W_{31} - W_{34} - W_{35} - W_{36} - W_{30} \quad [19]$$

$$W_{44} = 1 - W_{41} - W_{42} - W_{43} - W_{47} - W_{40} \quad [20]$$

$$W_{55} = 1 - W_{51} - W_{52} - W_{53} - W_{57} - W_{50} \quad [21]$$

$$W_{66} = 1 - W_{60} \quad [22]$$

$$W_{77} = 1 - W_{70} \quad [23]$$

Hidden Layer. Contains activation function defined by Equation 24

$$f(Z = Z) \quad [24]$$

Output Layer. Outputs are represented by Equation 25

$$Yi = Pi(t + \Delta t) \quad \textit{where } i = 1\text{-}7 \quad [25]$$

ANN (Figure 3) is given by Equation 26 and 27

$$Y(j((= f\ \Sigma WijXi + bj \quad i, j = 0 \text{ to } 7, \quad [26]$$

$$bj \text{ is bias, } j = 0 \text{ to } 7 \quad [27]$$

with linear activation function as defined in Equation 24.

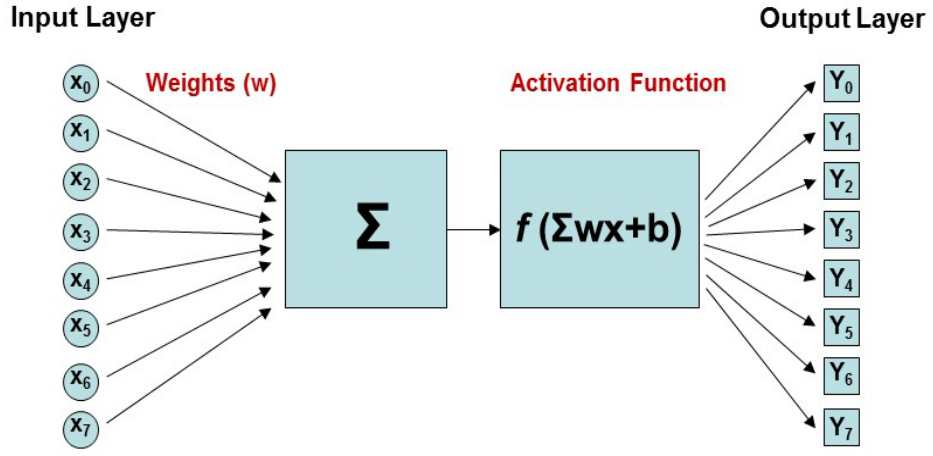


Figure 3. The Artificial Neural Network (ANN) Model of System

Solving ANN model, the outputs are given by substituting the values of Equation 2-23 in Equation 26 and using Equation 24 and 27 the outputs are given by Equation 28-35

$$Y_0 = W_{00}X_0 + W_{10}X_1 + W_{20}X_2 + W_{30}X_3 + W_{40}X_4 + W_{50}X_5 + W_{60}X_6 + W_{70}X_7 + b_0 \quad [28]$$

$$Y_1 = W_{01}X_0 + W_{11}X_1 + W_{21}X_2 + W_{31}X_3 + W_{41}X_4 + W_{51}X_5 + b_1 \quad [29]$$

$$Y_2 = W_{02}X_0 + W_{22}X_2 + W_{42}X_4 + W_{52}X_5 + b_2 \quad [30]$$

$$Y_3 = W_{03}X_0 + W_{33}X_3 + W_{43}X_4 + W_{53}X_5 + b_3 \quad [31]$$

$$Y_4 = W_{04}X_0 + W_{24}X_2 + W_{34}X_3 + W_{44}X_4 + b_4 \quad [32]$$

$$Y_5 = W_{05}X_0 + W_{25}X_2 + W_{35}X_3 + W_{55}X_5 + b_5 \quad [33]$$

$$Y_6 = W_{06}X_0 + W_{26}X_2 + W_{36}X_3 + W_{66}X_6 + b_6 \quad [34]$$

$$Y_7 = W_{07}X_0 + W_{47}X_4 + W_{57}X_5 + W_{77}X_7 + b_7 \quad [35]$$

The Upstate and Downstate probabilities are given by Equation 36 and 37

$$P_{upstate} = Y_0 + Y_2 + Y_3 + Y_4 + Y_5 \quad [36]$$

$$P_{downstate} = Y_1 + Y_6 + Y_7 \quad [37]$$

In the above-mentioned model, State Probabilities at time 't' (input of ANN) are defined by Equation 1 and variations in those State Probabilities at time increment $(t + \Delta t)$ (output of ANN) are defined by Equation 25 and 26.

RESULTS AND DISCUSSION

Estimated Variations in State Probabilities, with Time

A system, at a particular time t, time factor being in months, has been considered for numerical computations and comparison of state probabilities. It has been assumed that the system has been under continuous operation, the initial state probabilities of the system are taken as given in Table 1. The Repair Rates (constant over time) are as given in Table 2. While Table 3 represents the two cases of Failure Rates, (i) Useful life of system when failure rate is constant and (ii) Wear-out phase of system when failure rate becomes time dependent. In case (i) constant failure rate of each subsystem is defined by its respective exponentially distributed survival function and in case (ii) time-dependent failure rate of each subsystem is defined by its respective Weibull distributed survival function and failure rate is defined as Equation 38:

$$\lambda(t) = \beta t^{\beta-1}/\eta^{\beta} \quad [38]$$

The shape parameters (β) and scale parameters (η) of time dependent failure rates of each subsystems at time, t = 0 have been taken to match the initial failure rates of respective subsystems assuming constant failure rates.

Table 1

Initial State Probabilities of Subsystems (Input Values for ANN)

$P_0(t) = X_0$	$P_1(t) = X_1$	$P_2(t) = X_2$	$P_3(t) = X_3$	$P_4(t) = X_4$	$P_5(t) = X_5$	$P_6(t) = X_6$	$P_7(t) = X_7$
0.45	0.05	0.1	0.1	0.1	0.1	0.05	0.05

Table 2

Repair Rates of Subsystems

Repair Rates (per month)	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7
Values	0.02	0.01	0.01	0.01	0.01	0.02	0.02

Table 3

Failure Rates of Subsystems Assuming Time-Dependent Failure Rates v/s Constant Failure Rates

Failure Rates (per month)	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
Scale Parameter	0.75	1.65	2.36	3	5	3.5	1.36
Shape Parameter	2	2	2	2	2	2	2
Failure Rate at t=0 (Weibull)	0.0497	0.0102	0.005	0.0031	0.0011	0.0022	0.015
Failure Rate (Constant)	0.05	0.01	0.005	0.003	0.001	0.02	0.015

In the proposed ANN model, bias is a constant added to activation function to shift the estimated values state probabilities to closer to observed state probabilities. But since in this numerical computation observed state probabilities have not been taken hence assuming $b_j = 0$.

Applying ANN to the input state probabilities given in Table 1, taking time as a factor of months and $\Delta t = 10\ hours = 0.014\ month$ and taking weights as combination of repair and failure rates as mentioned in Equation 2-23, the estimated changes in State Probabilities of subsystems with time increment of 10 hours, computed using ANN model outputs given in Equation 28-35, for constant failure rate over time (useful life), are given in Table 4 and Figure 4. Similarly, the estimated changes in State Probabilities of subsystems with time increment of 10 hours, for time dependent failure rate (wear-out phase under assumption that survival function follows Weibull distribution) are given in Table 5 and Figure 5.

It can be observed from Table 4 and Table 5 that under both the circumstances i.e. "useful life" and "wear-out phase", the estimated decrease over time in state probabilities of degraded efficiencies P_3 versus P_5 are very close to each other. Similarly, it is also evident that the estimated decrease over time in state probabilities of P_2 versus P_4 are also observed to be close to each other under both the circumstances. In ANN model, the output (variation in state probabilities at time, $(t + \Delta t)$) are derived by summation of product of weights and input states. The closeness in values of states P_2 versus P_4 and P_3 versus P_5 are due to the two facts (i) this summation is dominated by value of their respective weights rendering cumulative effect of other weights and states as negligible (ii) the values of state probability at time t=0 (input value of states) and are assumed to be similar (0.1 each). Theoretically, this similarity can be explained by the fact that there is only one process working in each of these states and these processes are similar in nature.

As evaluated in Table 4 the estimated increase in failed State probability P_1 was from 5% to 5.5% during 80 hours when failure rates of subsystems are constant over time (useful life period). However, as evaluated in Table 5 it can be observed that estimated value of P_1

increased considerably from 5% to 7% for same time period when failure rates are time-dependent (wear-out phase). The failed state P_1 is reached due to failure of subsystem I which has seven processes. On top of that each process has heavy loaded equipment having tendency to deteriorate rapidly with time thereby affecting the shape parameter (β) of time-dependent failure rate, unless appropriate maintenance strategy is adopted. Both these factors add up to the considerable decrease in state probability P_1 with time when the system goes to deteriorating condition referred to as to wear-out phase in this study.

The changes in respective Upstate and Downstate probabilities, computed using Equation 36 and 37 respectively are given in Table 6 and Figure 6. It is evident that the estimated decrease in upstate probability is from 85% to 84.53% for useful life, and from 85% to 82.87% for wear-out phase. Again, this considerable decrease in downstate probability during wear-out phase is mostly attributed to its direct correlation with P_1.

It can be verified from given data that $P_{upstate} + P_{downstate} = 1$

Table 4

Change in State Probabilities with Time (Subsystems with Constant Failure Rates)

TIME (Hour)	P_0	P_1	P_2	P_3	P_4
0	0.45	0.05	0.1	0.1	0.1
10	0.44944	0.05058	0.09997	0.09992	0.09995
20	0.44889	0.05116	0.09994	0.09984	0.0999
30	0.44833	0.05174	0.09991	0.09975	0.09986
40	0.44778	0.05232	0.09988	0.09967	0.09981
50	0.44722	0.0529	0.09985	0.09959	0.09976
60	0.44667	0.05348	0.09981	0.09951	0.09971
70	0.44612	0.05405	0.09978	0.09943	0.09966
80	0.44557	0.05463	0.09975	0.09935	0.09961

TIME (Hour)	P_5	P_6	P_7
0	0.1	0.05	0.05
10	0.09992	0.05013	0.05009
20	0.09984	0.05027	0.05017
30	0.09975	0.0504	0.05026
40	0.09967	0.05053	0.05034
50	0.09959	0.05066	0.05043
60	0.09951	0.05079	0.05051
70	0.09943	0.05093	0.0506
80	0.09935	0.05106	0.05068

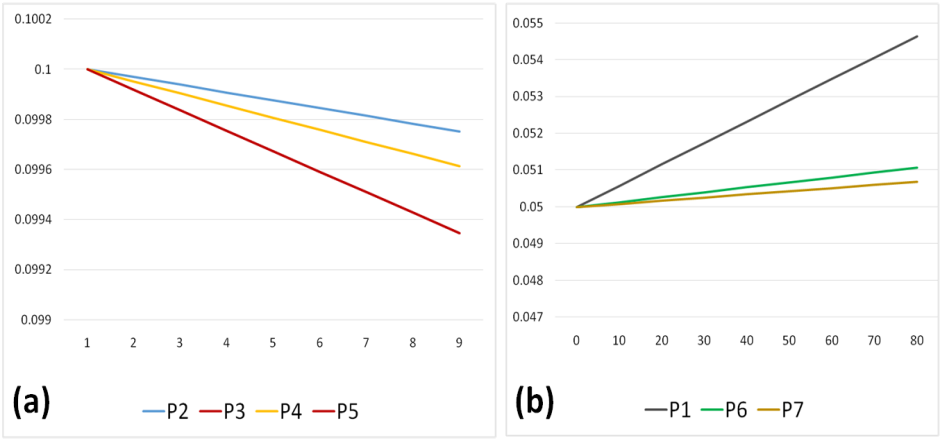


Figure 4. Variation in State Probabilities (a) P_2, P_3, P_4, P_5 (b) P_1, P_6, P_7 with Time (Subsystems with Constant Failure Rates)

Table 5

Change in State Probabilities with Time (Subsystems with Time-Dependent Failure Rates)

TIME (Hour)	P_0	P_1	P_2	P_3	P_4
0	0.45	0.05	0.1	0.1	0.1
10	0.44962	0.05058	0.09997	0.09992	0.09995
20	0.44893	0.05175	0.09987	0.09977	0.09984
30	0.44793	0.05351	0.09971	0.09955	0.09965
40	0.44662	0.05585	0.09947	0.09926	0.09939
50	0.445	0.05878	0.09916	0.0989	0.09907
60	0.44308	0.06228	0.09879	0.09847	0.09868
70	0.44086	0.06634	0.09835	0.09798	0.09822
80	0.43834	0.07097	0.09784	0.09742	0.09769

TIME (Hour)	P_5	P_6	P_7
0	0.1	0.05	0.05
10	0.09992	0.05002	0.050001
20	0.09977	0.05004	0.050003
30	0.09955	0.05007	0.050004
40	0.09926	0.05009	0.050006
50	0.0989	0.05011	0.050007
60	0.09848	0.05013	0.050008
70	0.09799	0.05015	0.05001
80	0.09743	0.05017	0.050011

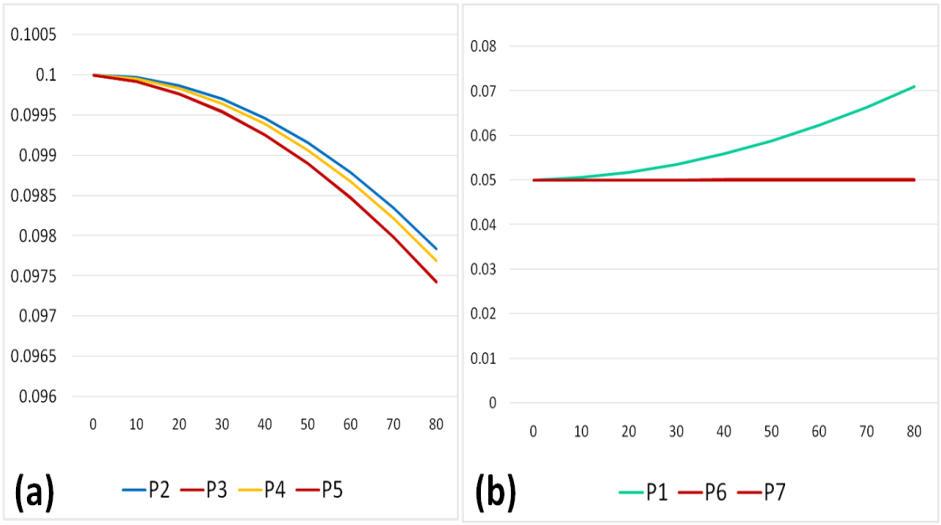


Figure 5. Variation in State Probabilities (a) P_2, P_3, P_4, P_5 (b) P_1, P_6, P_7 with Time (Subsystems with Time Dependent Failure Rates)

Table 6

Variation of Upstate and Downstate Probabilities with Time (Subsystems with Constant Failure Rates v/s Time- Dependent Failure Rates)

	Constant Failure Rate		Time Dependent Failure Rate	
TIME (Hour)	$P_{upstate}$	$P_{downstate}$	$P_{upstate}$	$P_{downstate}$
0	0.85	0.15	0.85	0.15
10	0.8492	0.1508	0.849384	0.1506
20	0.8484	0.1516	0.848177	0.15179
30	0.8476	0.1524	0.846381	0.15358
40	0.84681	0.15319	0.843998	0.15594
50	0.84601	0.15399	0.841035	0.15889
60	0.84522	0.15478	0.837495	0.16242
70	0.84442	0.15558	0.833387	0.16651
80	0.84363	0.15637	0.828719	0.17116

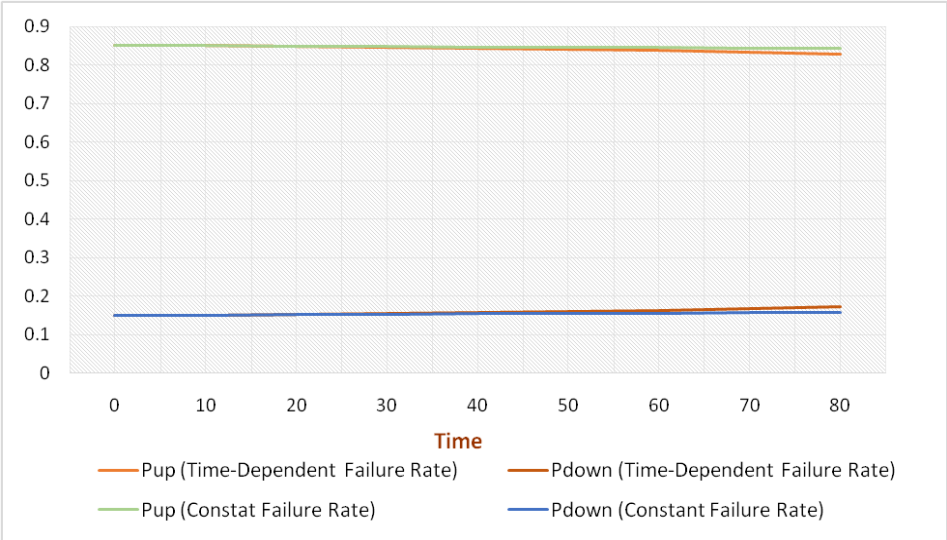


Figure 6. Variation of Upstate and Downstate Probabilities with Time (Subsystems with Constant Failure Rates v/s Time-Dependent Failure Rates)

Profit Analysis Considering Preventive Maintenance (PM)

One of the prominent objectives to estimate reliability of an industrial process is to optimize the costs to maximize the profits. Although, maximizing the uptime probability to more than 90% seems to be an obvious choice. However, its implementation leads to maximizing maintenance costs. There are various maintenance performance measures suited to meet the industrial requirements with objective to minimize costs (Samat et al., 2011), still their implementation affects profit. Many researchers have used reliability parameters to estimate the profit function of various industrial processes (Taneja et al., 2007; Yaqoob et al., 2017; Nasir et al., 2019).

In metal sheet manufacturing plant, total cost consists of many factors including material cost, labour cost, engineering cost including machine running cost, maintenance costs, and overhead cost (Dallan Newsletter, 2017). Profit function, or Expected Total Profit per unit time, is computed by Equation 39 and 40:

$$Profit\ per\ unit\ time\ = K_1 A_0(t) - C_T - C_{PM} \quad [39]$$

K_1= Revenue per unit uptime; $A_0(t)$ =Steady state availability of the system;
C_T = Total Cost per unit time; C_{PM} = Total Cost of PM per unit time

$$C_{PM} = C_{PRS} + C_{DT} \quad [40]$$

Where,
C_{PRS} = *Preventive Replacement and Service Cost;* C_{DT} = *Downtime Cost*

Numerical Computation

Taking K_1 = 14000 US$, C_T = 8400 US$ each for time increment Δt = 10 *hour*
Suppose the system enters wear-out phase (deteriorating condition) at $t = 0$, wherein failure rate of each sub-system is time dependent as defined by Equation 38, with shape parameter $\beta = 2.5$ and the respective scale parameters taken in such a manner so as to match the initial failure rates of sub-systems given in Table 3.

To compare profitability of system under different Preventive Maintenance (PM) costs, computed using Equation 40 proposed further actions are enumerated in Table 7.

Table 7

Proposed further actions (i) PM with maximum repairs/replacements (ii) PM with minimum repairs/ replacements (iii) Without PM

Proposed Further Actions →	(i) System under PM with maximum repairs/replacement	(ii) System under PM with minimum repairs/ replacements	(iii) System without PM
Scheduled PM Cycle	After 2160 hours	After 720 hours	-
Time duration of PM	5 hours	1 hour	0
C_{DT} (@1400$ per hour)	US $ 7000	US $ 1400	0
C_{PRS}	US $ 3000	US $ 600	0
C_{PM}	US $ 10000	US $ 2000	0
C_{PM} per 10 hours	US $ 46.29	US $ 27.77	0

The effects of proposed further actions (Table 7) on system are given below:

(i) System under PM with maximum repairs/replacements

After this PM action, failure rate of each sub-system is defined by Equation 38, but its shape parameter reduces from $\beta = 2.5$ to $\beta = 2$ and initial failure rates are as given in Table 3.

(ii) System under PM with minimum repairs/replacements

After this PM action, failure rate of each sub-system is defined by Equation 38, but its shape parameter reduces from $\beta = 2.5$ to $\beta = 1.5$ and initial failure rates are as given in Table 3.

(iii) No action is taken and condition of the system continues to deteriorate ($C_{PM} = 0$)

Using numerical values of C_{PM} per unit time (10 hours) as given in Table 7 in Equation 39 and using estimated values of State Probabilities modeled by ANN, the variation in profit with respect to time for all three cases mentioned above are given in Table 8 and Figure 7. It is evident that if the system continues to be in working state without undergoing any PM, the profit with respect to time starts declining rapidly after certain period. As discussed earlier, the equipment also used in various processes of metal sheet manufacturing plant

are susceptible to quite rapid deterioration and downtime losses are high, therefore PM becomes inevitable. In this particular case, although, the costs per unit time associated with PM strategy of "maximum repairs and replacements" is twice that of PM strategy of "minimum repairs and replacements", but the running costs under it become more profitable over time.

Table 8

Variation in Profit with Time for system under (i) PM with maximum repairs/replacement, (ii) PM with minimum repairs, (iii) Without PM

TIME (Hour)	Profit per unit time (US $)		
	System under PM with max. repairs/replacement	System under PM with min. repairs/replacements	System without PM
0	3453.71	3472.23	3500
10	6931.19	6949.85	6983.12
20	10371.51	10380.29	10399.91
30	13764.65	13738.89	13681.83
40	17102.46	17001.22	16746.67
50	20377.98	20143.17	19501.88
60	23585.04	23141.06	21847.57
70	26718.10	25971.67	23679.27
80	29772.11	28612.35	24890.87

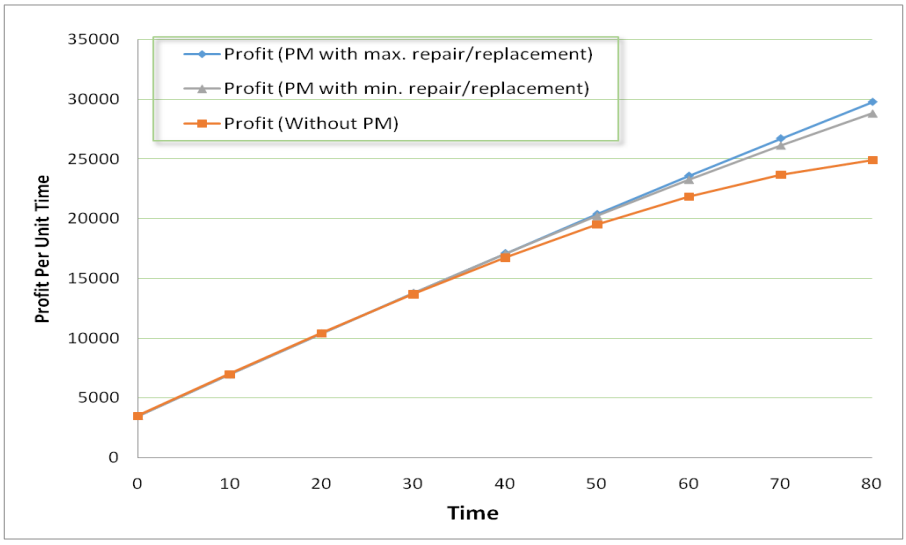


Figure 7. Variation in Profit with Time for system under (i) PM with maximum repair/replacement, (ii) PM with minimum repairs and (iii) Without PM

CONCLUSIONS

In this study, metal sheet manufacturing plant has been considered for examining its reliability in terms of multi-state probabilities of its different efficiency states. The system is susceptible to reach "wear-out phase" rapidly due to its high industrial demand. Therefore, the variations in multi-state probabilities with respect to time have been compared for system's (i) useful life period and (ii) wear-out phase (deterioration) using ANN model. During "useful life period" of system, failure rates of sub-systems are constant over time. While, during "wear out phase" of system, failure rates of sub-systems tends to increase with time due to deterioration of associated equipment. This factor, in turn, affects the state probabilities over time. These variations in state probabilities were found to decline/ increase linearly under system's "useful life". While, for "wear-out phase" these variations were mostly found to decline/increase sharply. `

It can also be concluded that the variation in failed state probability, P_1 with time, shows considerable increment during wear-out phase i.e. when failure rates are time-dependent. This effect is cumulatively attributed to more number of processes associated with state P_1 and deterioration of heavy-loaded equipment required in these processes. Similarly, decline in downstate probability was also higher over time when failure rates of sub-systems were assumed to be time-dependent but this influence is mainly due to correlation of downstate probability with P_1.

Considering a particular case of metal sheet manufacturing system the variations in upstate probability were used to estimate and compare the variation in profit per unit time, for system without PM, and under two PM strategies, (i) minimum repairs and replacements, (ii) maximum repairs and replacements. It has been concluded that although, PM strategy with maximum repairs and replacements is costlier, but it is more profitable in the long run.

ACKNOWLEDGMENTS

The authors are thankful to Dr. Prakriti Rai, Head of Mathematics Department and Dr. Sunita Rattan, Head of Institution of Amity Institute of Applied Science for their support. The authors are also thankful to reviewers for their critical comments which have helped in improving the representation of this study.

REFERENCES

Alyson, G. W., & Aparna, V. H. (2007). Bayesian networks for multilevel system reliability. *Reliability Engineering and System Safety, 92*, 1413-1420.

Angel, M. (2019). *Global crude steel output increases by 4.6% in 2018.* Retrieved March 30, 2020, from https://www.reuters.com/article/us-steel-output-global/global-crude-steel-output-jumps-46-percent-in-2018-worldsteel-idUSKCN1PJ1MF

Barlow, R., & Wu, A. (1978). Coherent systems with multi-state components. *Mathematics of Operations Research, 3*(4), 275-281.

Bhargava, C., & Handa, M. (2018). An intelligent reliability assessment technique for bipolar junction transistor using artificial intelligence techniques. *Pertanika Journal of Science and Technology, 26*(4), 1765-1776.

Bhati, B. C. (2019). *Steel: Long and flat productions.* Retrieved March 31, 2020, from http://www.careratings.com/upload/NewsFiles/Studies/Long%20and%20Flat%20Steel%20Products%20Feb%202019.pdf

Bhattacharyya, D. (1997). *Composite sheet forming* (vol. 11). Amsterdam, The Netherlands: Elsevier.

Chandra, A., Gupta, S., & Naithani, A. (2019). Assessment of reliability factors in chocolate manufacturing plant using boolean function technique and neural networking. *International Journal of Innovative Technology and Exploring Engineering , 8*(12), 2562-2567.

Dallan Newsletter. (2017). *Four steps to calculate the manufacturing cost of sheet metal products.* Retrieved March 30, 2020, from https://www.dallan.com/en/news/four-steps-to-calculate-the-manufacturing-cost-of-sheet-metal-products/

de Sousa, A. R. (2016). Incremental sheet forming technologies. *Reference Module in Material Sciences and Materials Engineering, 2016*, 1-10.

Fang, Y., Tao, W., & Tee, K. F. (2016). Reliability analysis of multi-state engine units utilizing time-domain response data. *Journal of Systems Science and Information , 4*(4), 354-364.

Grand View Research. (2020). *Steel market size, share and trend analysis report 2020-2027.* Retrieved March 30, 2020, from https://www.grandviewresearch.com/industry-analysis/structural-steel-market

Hurtado, J. E., & Alvarez, D. A. (2001). Neural-network based reliability analysis: A comparative study. *Computer Methods in Applied Mechanics and Engineering , 191*, 113-132.

Kalpakjian, S., & Schmid, S. (2001). *Manufacturing engineering and technology.* New Jersey, USA: Prentice Hall.

Karunanithi, N., Whitley, D., & Malaiya, Y. K. (1992). Using neural network in reliability prediction. *IEEE Software , 9*(4), 53-59.

Kozaki, M., Kozaki, S., Tohnai, S., Wakayama, Y., Fujisawa, H., Onari, K., & Ikeda, S. (2017). *Design and construction of steel-structured houses and future efforts.* Technical Report No. 115. Chiyoda-ku, Tokyo: Nippon Steel & Sumitomo Metal.

Li, Z., Xu, T., Gu, J., Dong, Q., & Fu, L. (2018a). Reliability modelling and analysis of a multi-state element based on a dynamic bayesian network. *Royal Society Open Science , 5*(4), 1-18.

Li, Z. Q., Xu, T. X., Gu, J. Y., Fu, L. Y., & Zhao, J. Z. (2018b). Performance degradation modeling for multi-state element considering aging factor. *Mechanika , 24*(2), 260-267.

Lisnianski, A., & Levitin, G. (2003). *Multi-state system reliability* (vol. 6). New York, NY: World Scientific.

Lisnianski, A., Elmakias, D., Laredo, D., & Haim, H. B. (2012). A multi-state Markov model for a short-term reliability analysis of a power generating unit. *Reliability Engineering and System Safety , 98*(1), 1-6.

Lisnianski, A., Frenkel, I., & Ding, Y. (2010). *Multi-state system reliability analysis and optimization for engineers and industrial managers.* London, UK: Springer.

Liu, Z., Liu, Y., & Cai, B. (2014). Reliability analysis of the electrical control system of subsea blowout preventers using markov models. *PLOS One, 9*(11), 1-9.

Murchland, J. (1975). Fundamental concepts and relations for reliability analysis of multi-state systems. In R. E. Barlow (Ed.), *Reliability and fault tree analysis, theoretical and applied aspects of system reliability* (pp. 581-618). Berkeley, United States: Society for Industrial and Applied Mathematics.

Nasir, M., Muhamad, W. M. W., & Maarof, R. A. R. (2019). Reliability based redundancy assessment of a cogeneration plant. *Pertanika Journal of Science and Technology , 27*(1), 225-246.

Natvig, B. (2011). *Multi-state system reliability theory with applications.* Chichester, United Kingdom: John Wiley and Sons, Inc.

Rausand, M., & Høyland, A. (2004). *System reliability theory.* Haboken, USA: John Wiley and Sons, Inc.

Reshid, M. N., Muhamad, W. M. W., & Maarof, R. A. R. (2017). Empirical analysis of chilled water generation for off peak period of cogeneration plant using neural network. *International Journal of Applied Engineering Research , 12*(24), 14669-14676.

Samat, H. A., Kamaruddin, S., & Azid, I. A. (2011). Maintenance performance measurement: A review. *Pertanika Journal of Science and Technology, 19*(2), 199- 211.

Sharma, P. S., Thakur, G. K., & Priya, B. (2016). Reliability measures for tele-communication system with redundant transferring machine by using algebraic method. *American Journal of Operational Research, 6*, 371-377.

Taneja, G., Singh, D. V., & Minocha, A. (2007). Profit evaluation of 2-out-of-3 unit system for an ash handling plant wherein situation of system failure did not arise. *Journal of Information and Optimization Sciences, 4*, 195-204.

Tang, D., Eversheim, W., & Schuh, G. (2007). Qualitative and quantitative cost analysis for sheet metal stamping. *International Journal of Computer Integrated Manufacturing , 17*(5), 394-412.

Yaqoob, A. R., Rizwan, S. M., Alkali, B. M., Andrew, C., & Taneja, G. (2017). Reliability analysis of rodding anode plant in aluminium industry. *International Journal of Applied Engineering Research, 12*(16), 5616-5623.

Zaitseva, E., & Levashenko, V. (2017). Reliability analysis of multi-state system with application of multiple-valued logic. *International Journal of Quality and Reliability Management, 34*(6), 862-878.

Pertanika J. Sci. & Technol. 28 (4): 1487 - 1515 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Seasonal Upwelling in the Northern Arafura Sea from Multi-datasets in 2017

Agus Saleh Atmadipoera[1]*, Agits Agnia Almatin[1], Rina Zuraida[2] and Yani Permanawati[3]

[1]Department of Marine Science and Technology, Faculty of Fisheries and Marine Sciences, IPB University, 16680 Bogor, West Java, Indonesia
[2]Center for Geological Survey, Geological Agency, 40122 Bandung, Indonesia
[3]Marine Geological Institute, Balitbang ESDM, 40174 Bandung, Indonesia

ABSTRACT

Seasonal upwelling phenomenon in the Arafura Sea plays an important role on supplying upwelled nutrient-rich water to sustain biogeochemistry processes and thus contributes to high marine primary productivity and fisheries resources in this region. The objective of this research was to investigate physical process and dynamics of upwelling by analyzing stratification of seawater properties, evolution of surface ocean-atmosphere parameters, and current structure and transport volume in the northern Arafura Sea. The multi-datasets in 2017 were used in this study, acquired from field CTD measurement, satellite-derived sea surface parameters, and the ocean general circulation model outputs, which were processed and analyzed using the available standard procedure. It was found that upwelling event was associated with a sharp subsurface thin layer that upsloping isotherms (23.5 - 25.5 °C), isohalines (33.50 - 34.25 psu), and isopycnals (21.8 - 23.2 kg/m^3) from the shelf-break region to the inner shelf region at a distance of approximately 167 km. This barrier layer separated the first surface mixed layer from the second mixed layer beneath the subsurface layer. The model suggests that the current in these two layers is in the opposite direction, to the west in the first layer as a response to the Ekman drift and to the

ARTICLE INFO

Article history:
Received: 5 April 2020
Accepted: 27 July 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.19

E-mail addresses:
atmadipoera_itk@apps.ipb.ac.id (Agus Saleh Atmadipoera)
agits.agnia@gmail.com (Agits Agnia Almatin)
rina.zuraida@grdc.esdm.go.id (Rina Zuraida)
yani.permanawati@esdm.go.id (Yani Permanawati)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

east in the second layer as a current extension from deep Aru basin. Therefore, upwelling dynamics here is not only generated by the southeasterly monsoon winds from May (onset) to November (termination) that transport warm and fresh surface water away from the shelf, but also modulated by the presence of strong inflow currents beneath subsurface that supply colder saltier nutrient-rich water into the shelf. During the upwelling period, mean transport volume in the upper 25 m depth between Aru and Papua at 134.25°E was -0.28 (±0.34) Sv (westward), but the transport volume between 25m and 110m depth was +1.06 (±0.29) Sv (eastward), suggesting this inflow may regulate the upwelling and supply Arafura shelf water.

Keywords: Arafura sea, chlorophyll-a, ocean current, salinity, temperature, upwelling

INTRODUCTION

The shallow Arafura shelf sea is located in the eastern Indonesian Archipelago and northern Australia waters, and is a semi-enclosed sea with complex bathymetry ranging from 30 to 90 m depth. The northern part of Arafura Sea faces on eastern deep Banda Sea and Aru basin with high pelagic and demersal fisheries productivity. Arafura and eastern Timor Seas show a seasonal high marine productivity and abundance of marine living resources with primary productivity greater than 300 gC/cm2/year (Alongi et al., 2011). This high marine productivity is associated significantly with upwelling phenomenon (Alongi et al., 2011). Hence, Arafura Sea is one of important fishing grounds in Indonesia, contributing 21% of marine fisheries at about 2.64 million ton/year (Sari et al., 2018).

Upwelling is defined simply as a physical oceanographic process in which displaced surface warmer waters are replaced by colder and nutrients-rich water that wells up from deeper layer, and its major driving factor is surface wind fields (Sarhan et al., 2000). According to Wyrtki (1961) upwelling event in Arafura Sea is categorized as a seasonal coastal upwelling, which is driven by the southeasterly monsoon winds (Figure 1). Modeling study of Kämpf (2015) assumed that upwelling in Arafura Sea was forced by eastward flow of under-currents derived from eastern Banda slope and westward flow in the surface layer. A schematic surface circulation during the southeast monsoon (SEM) period from May to September (Figure 1) is indicated by westward and southwestward flows as a response to the southeasterly monsoon winds (Wyrtki, 1961; Condie, 2011; Kämpf, 2015).

In some interior Indonesian seas such as Banda Sea, southern Makassar Strait, and Banggai Maluku Sea, the major forcing of coastal upwelling is generated by the persistent southeasterly monsoonal winds which is fully developed during the SEM period, and peak of upwelling period usually appears between August and September. Gordon and Susanto (2001) reported that Banda upwelling was caused by a surface layer divergence in eastern

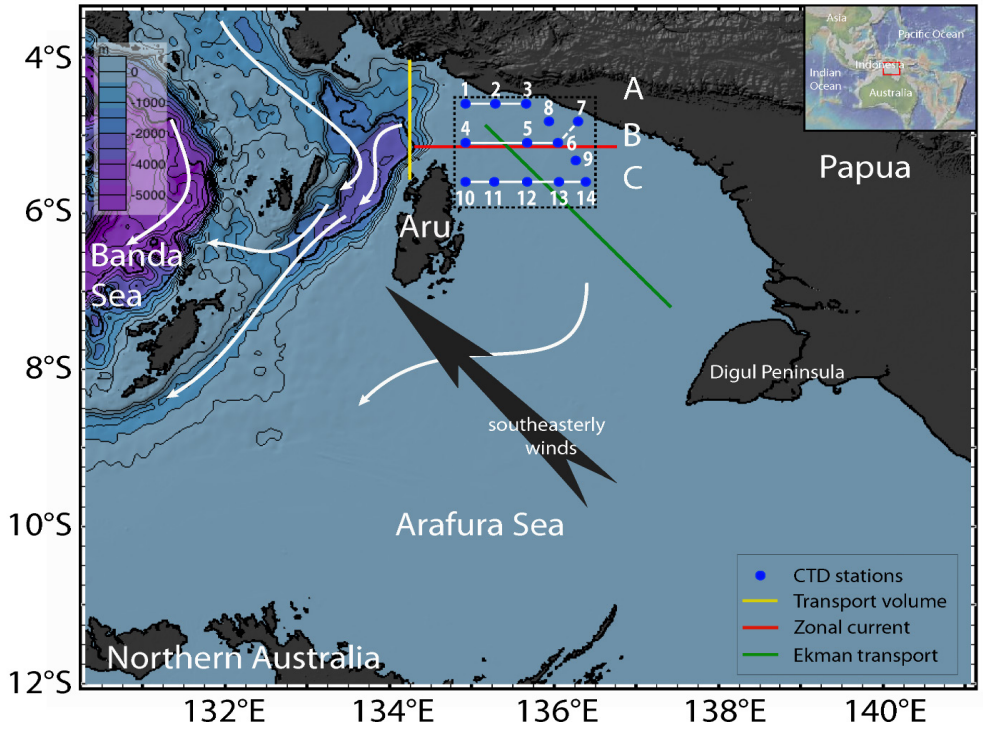


Figure 1. A schematic surface current circulation (white arrows) and southeasterly monsoonal winds (black large arrow) during the Southeast Monsoon (SEM) period over the study area. Bathymetric contours in the Arafura and Banda Sea are drawn for every 500 m. Yellow, red, and green section lines denote for upper 25 m transport volume in east-west direction, depth-distance zonal current section, and Ekman transport calculation, respectively. CTD station measurement is blue dots. Bathymetry map is made with GeoMapApp (www.geomapapp.org).

part of Banda during the SEM period. In southern tips of Makassar Strait, modelling study showed that southeasterly wind-driven upwelling was modulated by southward surface Indonesian Throughflow from and its intensity is also reduced by leakage of inflow from Selayar Strait (Atmadipoera & Widyastuti, 2015). In western Maluku Sea near Banggai waters close to equatorial region, coastal upwelling is forced by fully developed southerly monsoon winds that blow over small islands of Taliabu and Banggai islands creating upwelled water near coastal waters and flowing northward (Atmadipoera et al., 2018).

Coastal upwelling brings colder, higher saline, and nutrients-rich from deeper layer to sea surface. Surfacing of relatively colder water with isotherms of 24.5°C and 26°C from about 60 m depth to the sea surface have been found during peak of upwelling period in southern Makassar Strait and in Banggai waters in Maluku Sea (Utama et al., 2017; Atmadipoera et al., 2018). Regeneration of the upwelled nutrients-rich water are then

utilized by phytoplankton in the photosynthetic processes with chlorophyll-a pigment receives light energy to be chemical energy. Rosdiana et al. (2017) found that at the center area of upwelling in southern Makassar Strait nutrients tended to be low but the chlorophyll-a was high, suggesting that regeneration of nutrients from deeper layer had been used for photosynthesis processes.

Previous studies on upwelling dynamics from satellite data and ocean model in Arafura Sea and its adjacent waters had been done by Condie (2011), Schiller (2011), Kida and Wijffels (2012) and Kämpf (2015). They showed model seasonal circulation driven by the monsoonal winds and barotropic tides around northern coastal of Australia and part of southern Arafura Sea (Condie, 2011). Model circulation in northern Australian Shelf, including Arafura Sea was also done by Schiller (2011), where his main focus was in Australian Shelf and did not discuss Arafura upwelling. Kida and Wijffels (2012) found negative anomaly of sea surface temperature in northern Arafura Sea during the SEM monsoon, associated with seasonal upwelling event. Using 3-dimension hydrodynamics model Kämpf (2015) suggested classical lee effect that created undercurrent which drove nutrients-rich deep water from eastern Banda Sea slope to northern Arafura Sea by about 300 km, and then this deep nutrients-rich water was lifted up to surface layer, as upwelling process. These modeling studies have revealed a better general understanding on upwelling dynamics and circulation in Arafura Sea and its adjacent waters. However, evolution of surface ocean-atmosphere from the onset to the termination of upwelling, observed stratification of seawater properties, and upper-layer transport volume estimate have still less been well understood.

Here, a field hydrological measurement has been carried out onboard research vessel Geomarin III of Marine Geological Institute Bandung Indonesia during the peak on upwelling period, from 2 to 12 September 2017. The objective of present study was to investigate physical processes and dynamics of upwelling in northern Arafura Sea by analyzing stratification of seawater properties, evolution of upwelling, current structure and transport volume estimates from multi-datasets (field observation, satellite-derived parameters, and general ocean circulation model outputs).

MATERIALS AND METHODS

Study Area

The study area was located in the northern Arafura Sea with geographical boundary between longitude of 130°E–141°E and latitude of 4°S–12°S. Figure 1 shows the location of the 12 CTD casts, the dashed black rectangle box and three cross-sections applied in this study, namely section A, B, and C. Black dashed rectangle is a sampling box for data and model validation and extraction of time-series analysis of sea surface temperature (SST), sea surface height anomaly (SSHA), air temperature, and surface chlorophyll-a (chl-a). Red-

line is a cross-section for analysis of depth-longitude zonal current component; yellow-line is for calculation of transport volume in the upper 25 m depth, 25-55 m depth, 55-110 m depth; and northwest-southeast transect (green-line) is for calculating Ekman transport derived from the wind fields datasets (Figure 1).

Data

Field measurement of CTD data casts from sea surface down to near seabed that vary between 43 m and 80 m depth, had been carried out at 12 stations (Table 1). The Sea-Bird Electronics SBE-CTD 19plus with 4 Hz sampling rate (Sea-Bird Scientific, 2017) had been used during the field measurement. The profiling of conductivity-temperature-depth with SBE-CTD can be done when the CTD sensors are lowered (downcast) and lifted up (upcast). However, in this study, only downcast CTD profiling data were used for analysis, since the profiling was not interrupted by bottle rosette stops.

The daily averaged model output datasets (3-dimension zonal and meridional current components, seawater temperature, and salinity) were obtained from simulation results of 1/12° ocean general circulation model of INDESO configuration, performed by CLS

Table 1

Hydrographic CTD measurement during the cruise in the northern Arafura Sea onboard the R.V. Geomarine 3 of Marine Geological Institute Bandung

Station	Geographical coordinate		Date	Local Time (UTC+9)	Bottom depth (m)	CTD depth (m)
	longitude (°E)	latitude (°S)				
1	135.098	4.6123	02/Sep/2017	23.42	80	72
2	135.466	4.61884	03/Sep/2017	03.19	63	57
3	135.816	4.61371	03/Sep/2017	05.25	43	30
4	135.096	5.06561	10/Sep/2017	15.05	75	74
5	135.82	5.0681	03/Sep/2017	15.48	69	65
6	136.132	5.06545	03/Sep/2017	13.59	56	48
7	136.432	4.82533	03/Sep/2017	10.59	43	40
8	136.129	4.8149	03/Sep/2017	08.41	54	53
9	136.37	5.27615	12/Sep/2017	03.54	44	38
10	135.101	5.57434	11/Sep/2017	11.40	60	55
11	135.47	5.58089	11/Sep/2017	22.08	45	42
12	135.819	5.57964	11/Sep/2017	20.10	45	42
13	136.133	5.5745	12/Sep/2017	05.10	52	49

France. Model configuration and validation has been described in detail in Tranchant et al. (2015). INDESO is a program of developing and utilizing oceanographic data for Indonesian fisheries and marine needs, which was developed by the Ministry of Maritime Affairs and Fisheries of the Republic of Indonesia through the Agency of Marine Fisheries Research and Development.

In brief, the configuration of the INDESO model can be described as follows, horizontal resolution of 1/12° or approximately 9.25 km and vertical resolution consisting of 50 depth levels, where at the depth of the first 10 meters, the thickness of the layer is less than 2 meters and then increasing by 10 meters at 50 meters depth; bathymetry of this model from ETOPO2 v2 with a 2-minute grid resolution and from GEBCO with a 1-minute grid resolution; atmospheric forcing used data from operational analysis of the European Center for Medium-Range Weather Forecasts (ECMWF); the ocean-atmosphere interaction model uses the "bulk" formula from CORE; tidal forcing has tidal forces geopotential components M2, S2, N2, and K2 (the main component of semidiurnal tides) and components K1, O1, P1, and Q1 (the main components of diurnal tides); and monthly climatology of freshwater runoff is obtained from freshwater runoff data on the coast and 99 major rivers in Indonesia (Tranchant et al., 2015).

The global ocean OSTIA satellite-derived sea surface temperature datasets were obtained from Marine Copernicus (https://resources.marine.copernicus.eu/). The OSTIA global sea surface temperature reprocessed product provides daily gap-free maps of foundation sea surface temperature and ice concentration (referred to as an L4 product) at 0.05° x 0.05° (5.6 km) horizontal grid resolution and daily-mean temporal resolution, using in-situ and satellite data (Worsfold et al., 2020). The daily-mean of SST datasets in the study area (130-141°E; 4-12°S) were used for the upwelling analysis.

Satellite derived surface chlorophyll-a (chl-a) datasets were also obtained from European Centre for Operational Oceanography of Copernicus in Toulouse France. The daily-mean chlorophyll-a and bio-products refer to chlorophyll-a and primary production (PP) are based on a multi sensors/algorithms approach to provide to end-users the best estimate. Two dailies Chlorophyll-a products are distributed: one limited to the daily observations (called L3), and the other based on a space-time interpolation: the "Cloud Free" (called L4). Products of daily surface chl-a datasets are based on the merging of the sensors SeaWiFS, MODIS, MERIS, VIIRS-SNPP&JPSS1, and OLCI-S3A&S3B with the spatial resolution is 4 km (Copernicus Marine Service, 2019; Garnesson et al., 2019).

Wind-stress datasets in 2017 were derived from scatterometer satellite datasets with 25 km spatial and 6-hour temporal resolutions were obtained from Copernicus Marine Service. In addition, surface atmospheric time-series datasets (zonal and meridional wind speed components, and air temperature) with 6-hour temporal and 0.125° spatial resolutions were obtained from ECMWF (https://www.ecmwf.int/en/forecasts/datasets).

Data Analysis and Validation

The general terminology used for winds fields in the north-south orientation of coastline is cross-shore wind component for zonal (u) and along-shore for meridional (v) components (Emery & Thomson, 2004). However, in northern Arafura Sea where coastline of Papua mainland is oriented northwest-southeast, the wind field components need to be rotated (Figure 1). These schematic rotated wind components represented for the peaks of monsoon period (August and February). Angle of rotation (θ) is estimated by about 58° counter-clockwise from the north (0°) and parallel to the mainland coastline. Rotated zonal (u') and meridional (v') wind-stress components can be calculated as described by Emery and Thomson (2004), and Kutsuwada (1998) (Equation 1):

$$u' = u\cos\theta + v\sin\theta; \;\; and \;\; v' = -u\sin\theta + v\cos\theta \qquad [1]$$

According to Pond and Pickard (1983) and Stewart (2006), offshore Ekman upwelling transport caused by southeasterly monsoon winds can be calculated with the following Equation 2,

$$M_xE = \frac{\tau_y}{f} \qquad \text{and} \qquad M_yE = -\frac{\tau_x}{f} \qquad [2]$$

where, MxE and MyE are along-shore and cross-shore Ekman transport components (kg/m s), respectively; τx and τy are cross-shore and along-shore wind stress component (Pa), respectively; f is Coriolis parameter (rad/s), calculated from f = 2 Ω sin θ; Ω is Earth' rotation speed (7.29x10-5 rad/s); θ is geographical latitude (°).

Estimate of transport total in northern Arafura Sea is calculated by using Ekman transport's approach which is integrated along a transect line, as described by Sprintall and Liu (2005) (Equation 3),

$$Q_E = \Delta x \sum_{i=1}^{n} \frac{\tau_y}{\rho_s f} \qquad [3]$$

where, QE is total Ekman transport (m^3/s); Δx is grid length (m); ρs is density of seawater (1024 kg/m3); n is number of grid (55).

Estimate of transport volume. Transport volume is calculated by integrating zonal current component for each grid (uz,y) to transect length (y) and depth (z), following Emery and Thomson (2004) (Equation 4),

$$Q = \int_{z1}^{z0} \int_{y1}^{y2} u_{z,y}\, dy\, dz \qquad [4]$$

where, Q is transport volume in Sv unit (1 Sv = 106 m^3/s); z is depth integration from sub-surface (z1) of 25 m depth to the sea surface (z0); y is length of section from y1 (4 °S) to y2 (5°24' S) or about 156 km; uz,y is zonal current component (m/s) at depth z and cell y.

Model and data comparison. Deviation of model output from the data measurement is estimated from the root mean square error (RMSE) (Wilks, 2006) (Equation 5),

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X\,model_{(i)}-X\,data_{(i)})^2} \quad [5]$$

where, Xmodel(i) is variable of model output (i = 1, 2, 3 …, N); Xdata(i) is variable of data (i = 1, 2, 3 …, N); and N is data number.

Seawater temperature is a key parameter for analysis of upwelling event in the study area. Here, model temperature is compared to observed 12 CTD temperature profiles at the same date and geographical coordinates (colocation). The average of RMSE of temperature from 12 CTD profiles was 0.60 (±0.21)°C. An example of comparison of model and CTD profile is shown in Figure 2. Here, both model and CTD temperature was in good agreement with RMSE of 0.22°C (Figure 2). Difference of model temperature and CTD data at 0.5 m depth was about 0.20°C. Below 35 m depth, both model and data temperature were just overlaid.

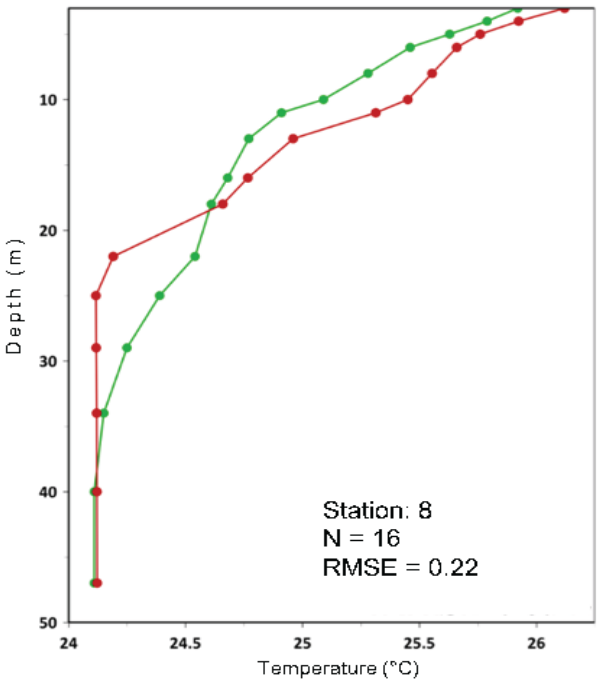


Figure 2. An example for comparison of temperature profiles between model (green) and CTD data (red) at CTD station 8

These small differences may be related to different spatial-temporal resolution between model and CTD data, e.g., model spatial resolution is 9.25 km with daily average, while CTD data is at one-point measurement and very short time snapshot. Since CTD data was only a snapshot field measurement between 2 and 12 September 2017, model temperature datasets were used to evaluate evolution of temperature prior to – during – and after the upwelling event in year 2017.

RESULTS AND DISCUSSION

Evolution of Surface Ocean Parameters: Onset and Termination of Upwelling

Onset and termination of upwelling in Arafura Sea were analyzed from several key ocean-atmosphere parameters, such as surface wind field, SSH, SST, air temperature (Ta), and surface chl-a (Figure 3). The southeasterly monsoon winds predominate from May to October (the SEM period), which are much stronger and much persistent compared to that during the northwesterly monsoon winds – the NWM period (November-April) (Figure 3a). Here, if we apply the Ekman upwelling theory, the southeast monsoon period from May to October is proposed as the onset and termination of upwelling event (in 2017). This is also consistent with other key parameters that fluctuate consistently with the winds data.

Negative anomaly of SSH commenced in May and its minimum was found by about -0.18 cm in August/September, then recovered to positive anomaly of SSH from October to April as under the period of the NMW. Gordon and Susanto (2001) found minimum of anomaly of SSH of about 20 cm (Figure 3b). This minimum of negative SSH anomaly is caused by upwelling process (Kämpf, 2016). The SEM period is also indicated by negative anomaly of SST from June to October with its minimum of -3.0°C in August/September (Figure 3c). Positive anomaly of SST is found during the rest of months (the NWM period). Fluctuation of negative air temperature anomaly (Ta) is similar with anomaly of SST (Figure 3d). However, during the NWM period fluctuation of Ta anomaly is completely different to SST anomaly (Figure 3d). Cooling air temperature may be caused by change of air-sea heat flux over the upwelling area (Sproson & Sahlée, 2014). Fluctuation of chl-a anomaly is out-of-phase to SSH and SST and Ta parameters (Figure 3e). Anomaly positive of chl-a during the SEM is coincident with negative anomaly of temperature and SSH anomaly. This may be related to upwelled nutrients- rich water from deeper layer to surface layer stimulate phytoplankton growth that increase chl-a concentration.

Seasonal Variation of Wind-stress, Seas Surface Temperature and Chlorophyll-a

Monthly averaged surface wind-stress, SST and Chl-a in February and September 2017 are shown in Figure 4, 5 and 6. In September (as a representative of the peak of SEM period) the southeasterly wind field vectors were much stronger than that the northwesterly

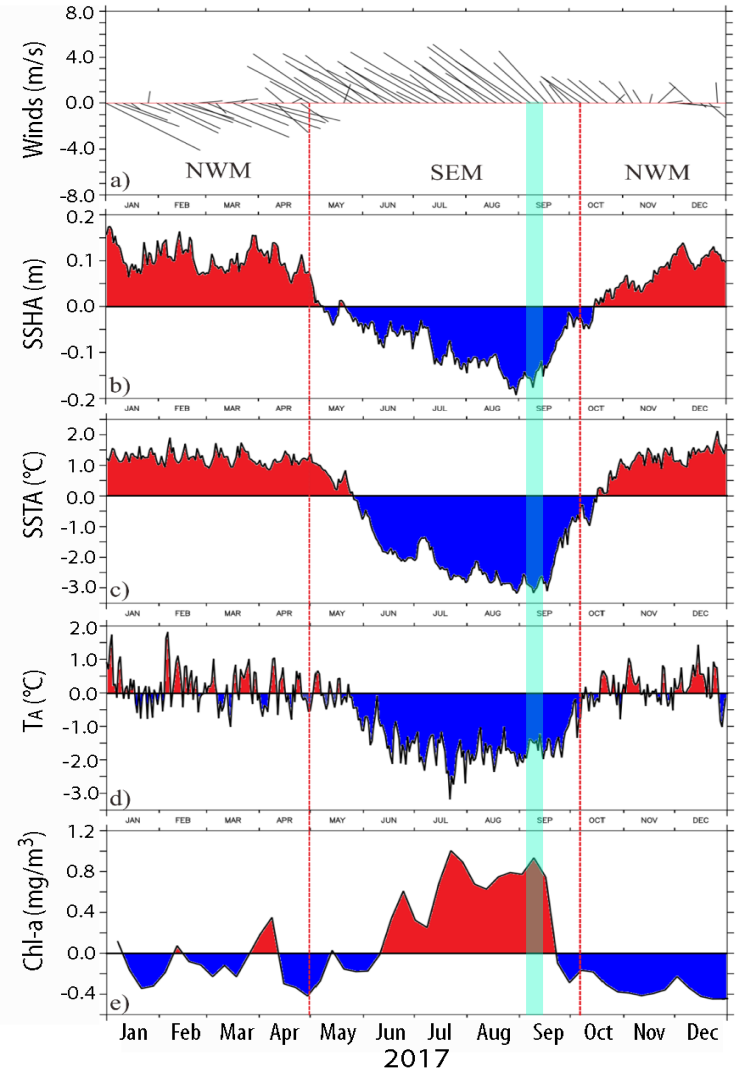


Figure 3. Time series of surface ocean-atmosphere parameters, averaged from a dashed black rectangle in Figure 1; (a) surface winds vectors; (b) anomaly of SSH; (c) anomaly of SST; (d) anomaly of air temperature; and (e) anomaly of surface chl-a in 2017. Vertical cyan-line denotes for field measurement time from 2 to 12 September 2017.

monsoon winds vectors in February (during the NWM period) (Figure 4a). Magnitude of wind-stress in 2017 in the northern Arafura Sea varied between 0.01 - 0.18 Pa with average of 0.06(±0.013) Pa. Maximum wind-stress vectors during the NWM appeared in February with average winds speed is 6.11 m/s. During the SEM period, maximum wind stress was found in June with wind speed of 8.23 m/s. However, on the climatology mean

the maximum of wind-stress occurs in August (Sprintall & Liu, 2005). The southeasterly wind fields are fully developed in central Arafura Sea with magnitude of wind-stress were much stronger during the SEM than that during the NWM. In particular, the maximum of wind-stress field in September 2017 exhibited around southern peninsula of Digul Papua (Figure 4b).

Seasonal upwelling event was examined from evolution of satellite-derived sea surface temperature (SST) and surface chl-a in 2017, since we had field CTD measurement in September 2017. Here, we contrasted spatial distribution of monthly mean SST and chl-a during the NWM period (February 2017) and during the SEM period (September 2017) (Figure 5).

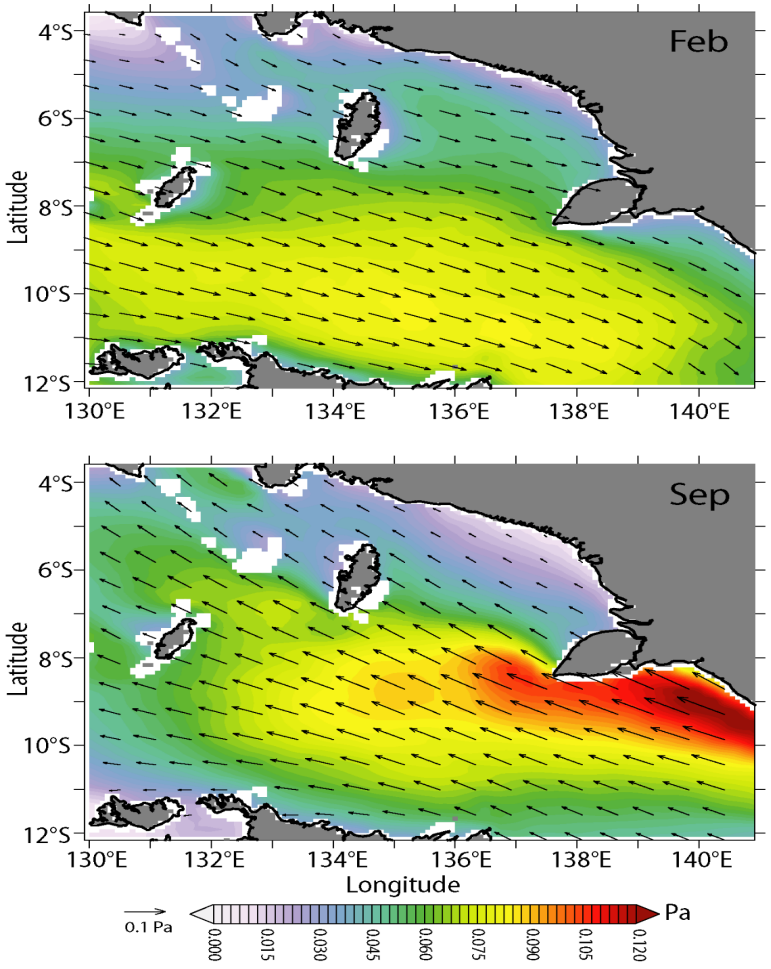


Figure 4. Monthly average of wind-stress vectors. Color denotes wind-stress magnitude in the study area for February (upper) and September 2017 (lower)

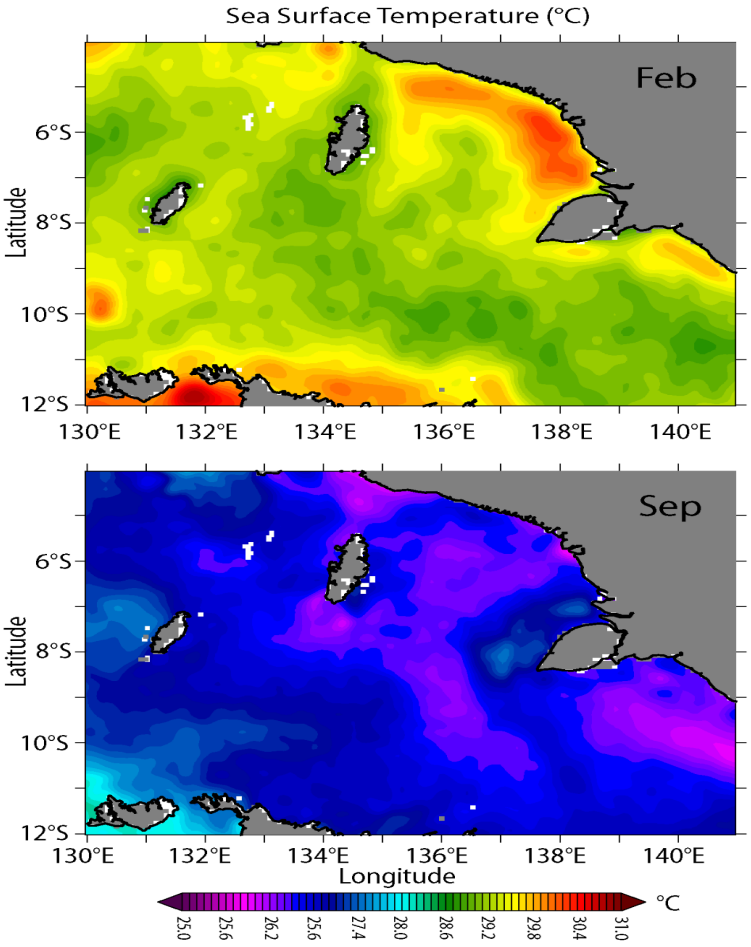


Figure 5. Monthly average of satellite-derived sea surface temperature (SST) in February (upper panel) and September 2017 (lower panel)

Distribution of monthly average SST in the study area in 2017 varied between 24.1 - 33.0°C with mean SST of 28.7(±1.5)°C. Seasonal changes of solar radiations and upwelling processes contribute to this large SST variation in this region (Gordon & Susanto, 2001; Condie, 2011; Kida & Richards, 2009;). During the SEM period, SST was much colder since heat budget from solar radiation was surplus in northern hemisphere. In contrast, during the NWM period, when southern hemisphere was surplus of heat budget, warmer SST was found (Figure 5). Furthermore, during the SEM period the southeasterly winds were much stronger than that during the NWM period. The SEM winds have also a significant spatial variation with strongest winds over the southeastern Banda that create upwelling (Condie, 2011; Kämpf, 2015).

During the NWM (February) the SST varied between 28.4 and 33.1°C with mean SST of 30.1(±0.6)°C. Much warmer SST was found around Aru Islands and northern Arafura, southwest Arafura, and around Digul Peninsula (Figure 3). In February 2017, SST ranged between 28.4 and 31.6°C. An increased temperature was found in March 2017 (28.7-31.7°C) and commenced to decrease in April and May 2017 with SST varied 27.5-31.1°C and 27.0-31.1°C, respectively. This SST variation is consistent with past studies (Dewi et al., 2018).

During the 2017 SEM period (September), SST varied between 24.3-29.8 °C with mean value of 27.0 °C. It shows that development of colder water is derived from southwestern coastal area of Papua. This colder water is commenced from May to September (Figure 3). In northern Arafura (around Aru island), evolution of colder SST appeared from June to October indicated by a patch of colder temperature of about 27.0°C which gradually grew much cooler and larger area and reached its minimum of 24.0°C in September. Extension of colder water distribution decreased in September and near the coastal Papua region colder temperature of 24.1°C was present. After that, SST commenced to increase in October 2017 with an increase temperature of 1.8°C. Here, temperature varied between 26.2 and 31.9°C. In November 2017 SST varied between 27.9 and 32.4°C (Figure 3).

Figure 6 shows a contrast of surface chl-a during the NWM (February) and SEM (September) in Arafura Sea. It is seen clearly a high chl-a in coastal area of the northern Arafura Sea both during different monsoon period. However, in the SEM period, high chl-a extended to the west between Aru and peninsula of Digul Papua.

In 2017 surface chl-a concentration in the study area varied between 0.02 - 12.85 mg/m^3 with average of 0.53(±0.89) mg/m^3. In the coastal waters of the northern Arafura Sea surface chl-a is always high around the year. This may be related to high contribution of rivers discharge from many rivers in Papua (Kämpf & Chapman, 2016). Mean surface chl-a (in 2017) during the NWM period was lower than that during the SEM period (Figure 6).

Distribution of surface chl-a during the NWM period (February) ranged between 0.02 and 6.53 mg.m-3 with mean value of 0.31(±0.61) mg/m^3 (Figure 6). In Arafura Sea during this period precipitation rate exceeds evaporation rate (Wyrtki, 1961). Thus, high nutrient inflow from river discharge into the coastal waters influence significantly on high chl-a concentration near the coastal area. Monthly mean of surface chl-a increased gradually from May with chl-a concentration of 0.57 mg/m^3. A drastic increase of surface chl-a was revealed during the SEM period that varied between 0.02 and 12.85 mg/m^3 with mean value of 0.98(±1.11) mg/m^3.

Surface chl-a reaches its maximum in the peak of SEM period (July-August-September). According to Kämpf (2016), spatial distribution of chl-a between June and October increases significantly, where horizontal distribution of chl-a extends to area of

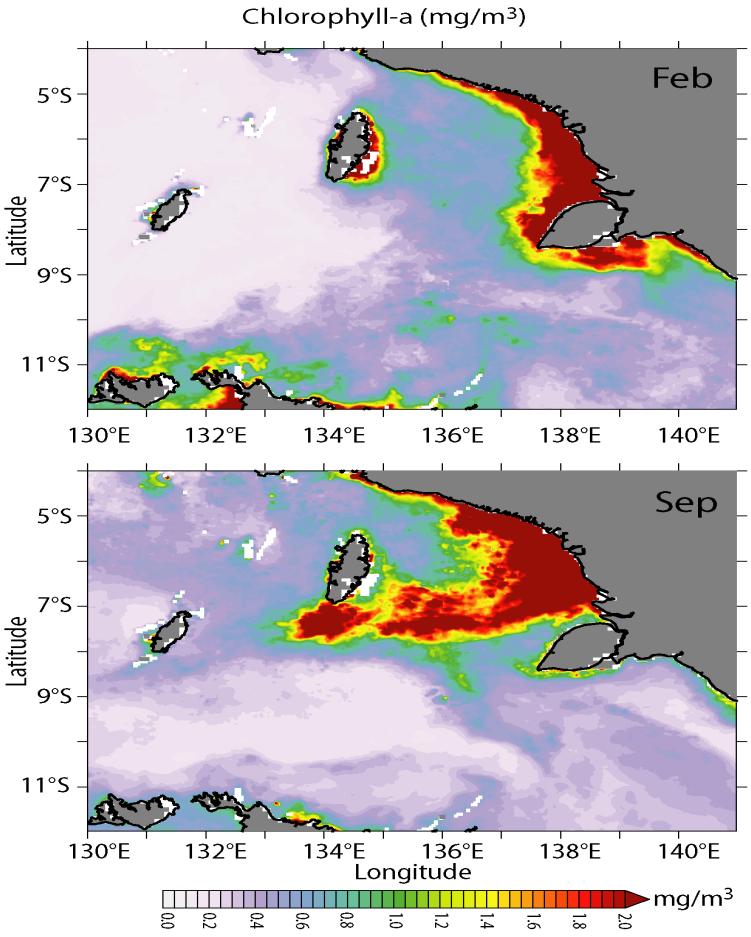


Figure 6. Monthly average of satellite derived surface chl-a in February (upper panel) and September 2017 (lower panel)

about 90,000 km^2. Starting from October, concentration of chl-a decreased again to about 0.33 mg/m^3, and this trend continued decreasing in December to about 0.15 mg/m^3.

This maximum surface chl-a concentration which is related to minimum SST during the SEM period may corroborate upwelling event. In contrast, during the NWM period maximum SST is followed by minimum of chl-a concentration. Upwelling process may contribute to that out-of-phase fluctuation between surface chl-a and SST, in which during the upwelling event colder and nutrients-rich deeper water mass is lifted up to near-surface layer. The availability of this unique water mass is utilized by phytoplankton in photosynthetic process, so that a drastic growth of phytoplankton is evident (Taufikurahman & Hidayat, 2017).

Stratification of Water Mass During Upwelling Event

Profiles of temperature, salinity and potential density anomaly (hereafter referred to as the parameters) in 3 CTD stations in section A (Figure 7a) revealed a step-like pattern with two strong vertical gradients of parameters in the upper 7 m (0.08-0.22°C m^{-1}), and between 31-35 m depth (0.10-0.18°C/m). In-between, a homogeneous layer was found between 7-30 m and 35-70 m depth. Step-like profiles may be formed due to different surface winds. The homogeneous layers at the beginning may be created by strong surface winds that mix surface layer. These conditions take place continuously, but the strength of the wind may vary and weaken causing the formation of this stratification (Tomczak, 2014). Other factor may be related to different flows direction in the sub-surface and near-bottom flow, as suggested by Kämpf (2015) where near-bottom flow directed eastward and westward flow in near-surface layer.

Spatial variation of temperature in the upper 30 m depth from station 1 to 3 was also seen, where colder water was found in station 3 (green) close to coastal area, and temperature at station 1 was much colder than station 2 at depth between 35 and 55 m depth. Salinity profiles showed a small variation and very fresh water in the upper 5 m depth. Density anomaly profiles are controlled strongly from temperature profiles.

Transect A was closed to coastal area (Figure 7b), so the upwelling process is expected much stronger compared to other two sections B and C. Here, slope from offshore to coastal area of isotherm of 25°C, isohaline of 34.25 psu, and isopycnal of 22.75 75 kg/m^3 from about 30 m depth to the sub-surface (about 5-10 m depth) were seen clearly. Strong vertical stratification in the upper 10 m was also observed as shown with plenty contours of parameters. Near surface layer temperature was about 26.5°C.

Profiles of the parameters in section B formed a step-like structure with different mixed homogeneous layers depth for each station (Figure 8a). For example, at station 4 offshore area (grey profiles), a homogeneous temperature was found in the upper 28 m depth, then from 29 m to 40 m depth a high vertical gradient of temperature of 0.19°C/m was seen before the second homogeneous layer.

Consistently, a cross-section of temperature-salinity and potential density anomaly in transect B (this is the longest section, about 150 km from offshore to the coastal area) revealed a clear signature of shallowing contours of these parameters from offshore to near coastal waters. Here, isotherm contours of 25.0°C and 25.5°C at depth near 40 m in the offshore area (as well as isohaline contours of 33.75 psu, 34 psu, and 34.25 psu; and isopycnal contours of 22.75 kg/m^3, 22.5 kg/m^3, and 22.25 kg/m^3) were uplifted to shallower depth of 5-10 m depth (Figure 8b). A very freshwater (<32.5 psu) and less density water surface layer water were found only in station 7, which might be related to runoff from rivers in the region. The shallowing isotherm, isohaline, and isopycnal contours from deeper layer in offshore toward the coastal waters indicated an upwelled water as a response on

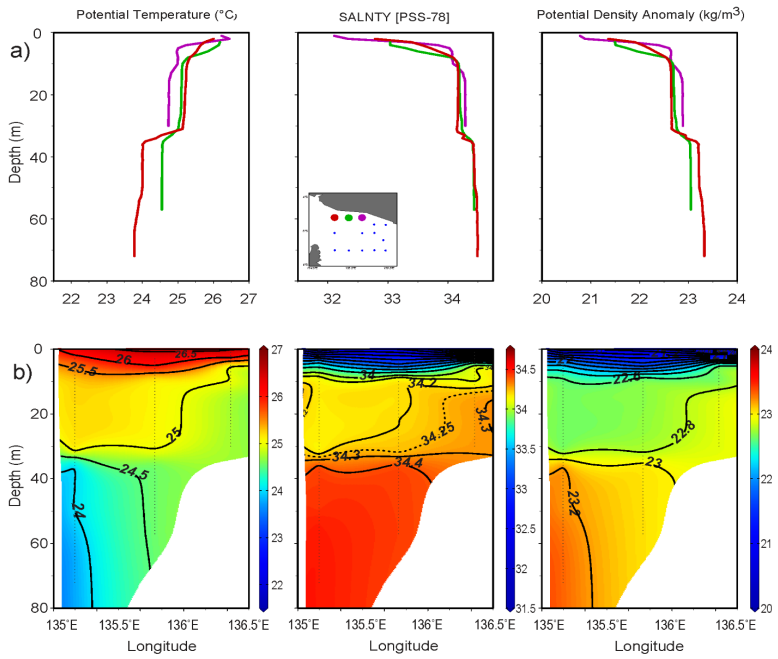


Figure 7. Profile of parameters (a) and its depth-longitude distribution (b) in section A

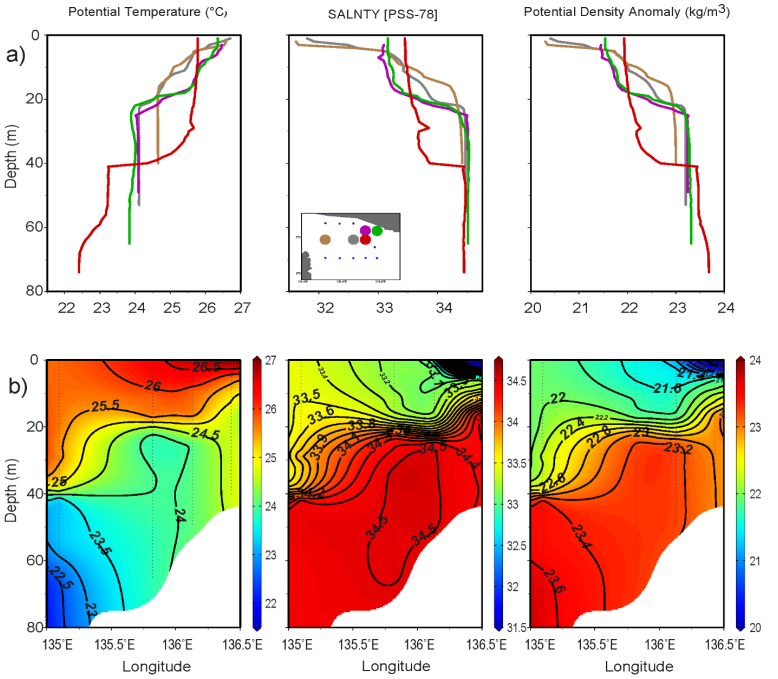


Figure 8. Profile of parameters (a) and its depth-longitude distribution (b) in section B

surface Ekman transport that brought near surface water away from coastal area, and then deeper water filled these surface water. This implies that there should be present near-bottom flow eastward that contrast to those in near-surface layer. Indeed, this reversal flows are found from the model current which will be discussed later in next section (Figure 13b).

Vertical profiles of the parameters at stations in section C showed a similar pattern (except for profile at station 7) with first homogeneous layer in the upper 20-25 m depth, then strong vertical gradient of temperature (0.36 – 0.45°C/m between 23-29 m depth, and then second homogeneous layer below 23 – 29 m depth to the bottom (Figure 9a). Vertical profile at station 7 revealed high vertical gradient of 0.15 in the upper 16 m depth, and a homogeneous layer below that.

In Transect C, shallowing of contours of isotherm, isohaline and isopycnal was also seen, but it lessened from 30 m to 20 m depth (Figure 9b). However, contour of isotherm 26°C extended from surface to about 20 m depth, as well as isohaline contour of 33 psu, and isopycnal of 22.0 kg/m^3. It is also shown an outcropping of isohaline 33.25 psu and isopycnal 21.75 kg/m^3 from 25 m depth in the offshore region to surface near the coast.

Vertical profile and stratification of the parameters in three CTD sections showed an evidence of upwelling processes in the study area, where colder, saltier and heavier water from deeper layer (30-40 m depth) in the offshore area were consistently uplifted toward the shallow coastal area. Here, upwelling indicator from CTD datasets could be defined for isotherm between 25.5 and 26.0°C, isohaline of 33.5-33.75 psu, and 22.5-22.75 kg/m^3. Surfacing of these parameters varied spatially, for example, these contours appeared in the upper 5 m, 9 m, and 20 m in the section A, B, and C, respectively.

Temperature-salinity relationship exhibits a structure of upwelling coastal water mass with the deepest CTD measurement of approximately 80 m depth. This may represent upper-thermocline water derived from adjacent deep Aru basin (Figure 10). Salinity varied from 31.62 psu at surface layer to 34.63 psu at deeper layer with mean salinity of 33.955 psu (±0.605). Potential temperature ranged from 21.995 °C to 26.707 °C with mean temperature of 24.559°C (±1.131), and potential density anomaly from 20 to 24 kg/m^3. Very fresh surface water was found at the stations closed to the coastal waters, but salty water was found at stations in the offshore and central area. Water mass with salinity maximum (34.6 psu) was found at surface density between 23.0 and 23.5 kg/m^3 which represented upper thermocline water. Relatively salty water (>34.0 psu) was found at approximately 40 m depth at the offshore but was uplifted to about 20 m depth in the onshore area. This spatial variation of seawater properties between shelf-break and inner-shelf regions may be associated with upwelling process. Mechanism of this coastal upwelling is discussed in the following section.

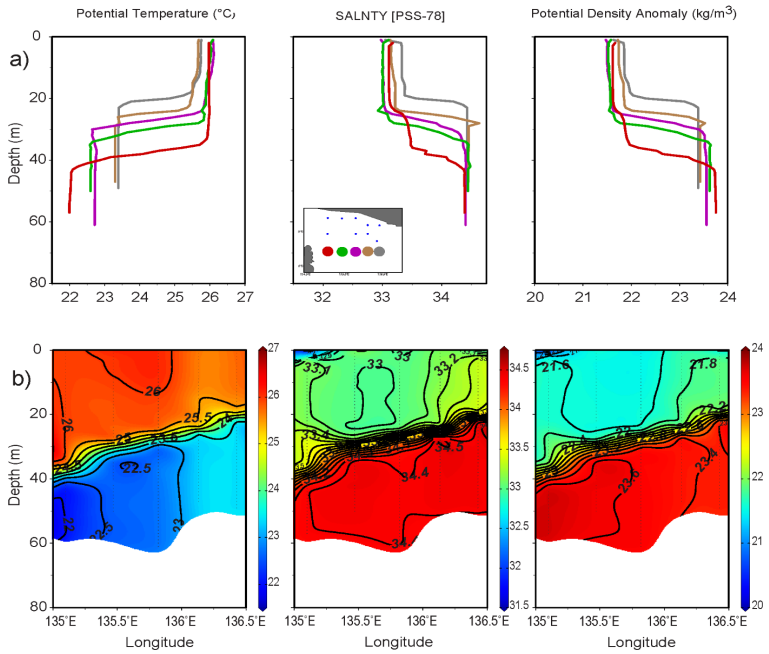


Figure 9. Profile of parameters (a) and its depth-longitude distribution (b) in section C

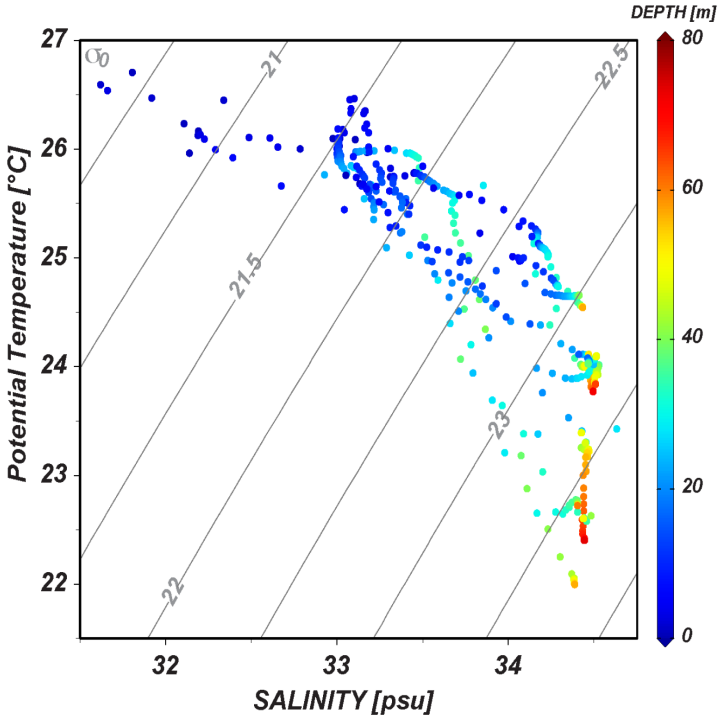


Figure 10. Temperature-salinity relationship (T-S diagram) plotted from 13 CTD casts during the Geomarine 3 cruise on 2-12 September 2017 in the northern Arafura Sea

Upwelling Mechanism in the Northern Arafura Sea

Previous analysis from derived-satellite parameters and field measurement datasets have demonstrated a strong indication of upwelling event. Here we proposed, a classical theory of coastal Ekman upwelling, as a physical mechanism of generating upwelling forced by the SEM winds field. Then, we evaluate regional circulation which may modulate the Ekman upwelling dynamics.

The strong persistently southeasterly monsoon winds during the SEM period which are fully developed from May to November are considered to be a main generating forcing of coastal upwelling in Arafura Sea (Figure 4). By applying Ekman upwelling theory, during the SEM period strong persistent southeasterly winds drive surface current westward-northwestward (or 45° from wind direction to the left) and integrated Ekman transport from surface to Ekman depth is expected drifting southwestward from wind direction which is away from the coastline of Papua. The continuity implies that transported surface water offshore must be replaced by water from deeper layer, and the upwelled water must be created. On the other hand, wind-stress fields in February 2017 may lead to generate downwelling since the Ekman drift directed to the coastal waters. The southeasterly (northwesterly) wind-driven surface current in the study area can be inferred from the surface and subsurface current field during different monsoon peaks, as discussed below.

Model surface current vectors at two depth levels (5 m and 25 m) overlaid with model temperature in February (the peak of NWM period) and September 2017 (the peak of SEM period) revealed a contrast of surface and subsurface flows pattern (Figure 11a-b). In February, circulation in the northern Arafura Sea, adjacent Aru island (north of 8°S), exhibited warm water and revealed two anti-clockwise eddies in northern and western Aru (Figure 11a). Relatively strong current vectors from Banda Sea and northern Aru basin flowed eastward into northern and southern Aru, which was consistent at 5 m and 25 m depths. In the Arafura shelf, surface circulation at 5 m depth during this period tends to flow eastward, but it is slightly different at 25 m depth, particularly along northern coastal water where weak northwestward flows are found.

During the SEM period (in September), a contrast of surface and subsurface circulation was seen (Figure 11b). At 5 m depth level, strong current vectors from Aru basin flow southeastward and deflected westward into Banda Sea in west of Aru. Furthermore, a relatively strong current vector that brought colder water from northern Aru flow westward which then join with current from Aru basin into Banda Sea. The westward flowed at this 5 m depth level appeared mostly over the region, such as at around latitude band of 7°S – 10°S (Figure 11b). As expected from Ekman upwelling theory, during the SEM period, surface current flows generally westward away from the coastal waters, as a response to the persistent southeasterly monsoon winds (Figure 4). Here, the minimum SST of 26°C appeared in September with much colder water occurring near northeastern Aru.

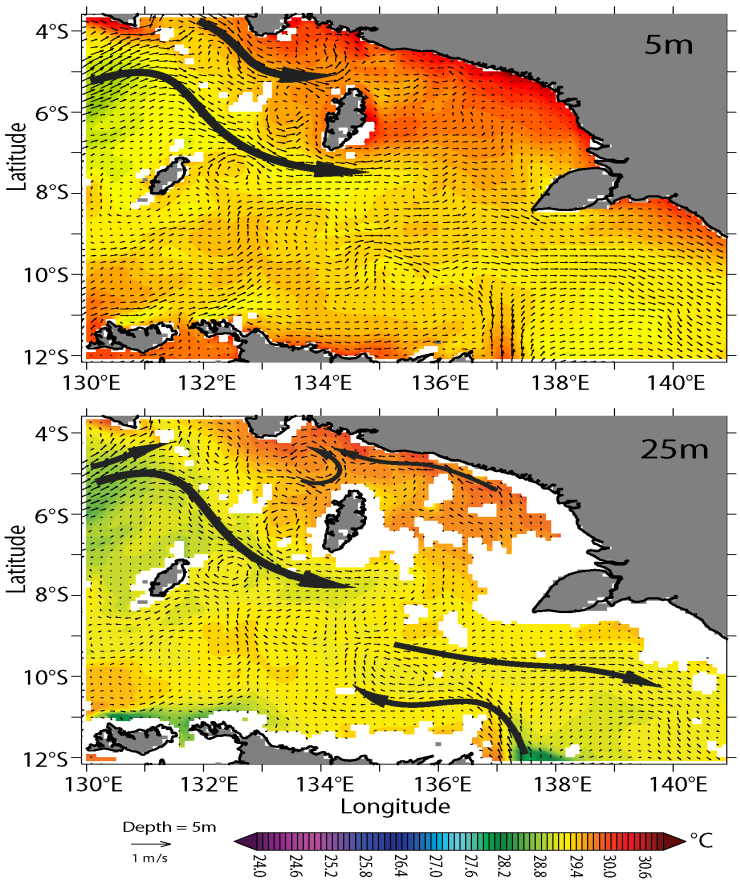


Figure 11a. Model surface current vectors, overlaid with model sea surface temperature at 5 m and 25 m depth in February 2017

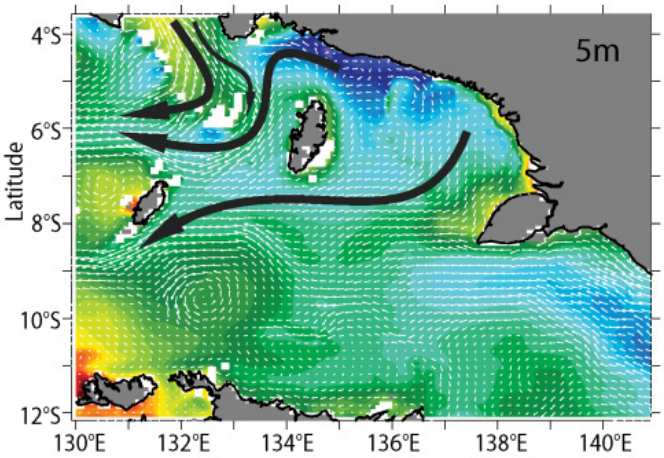


Figure 11b. Model surface current vectors, overlaid with model sea surface temperature at 5 m and 25 m depth in September 2017

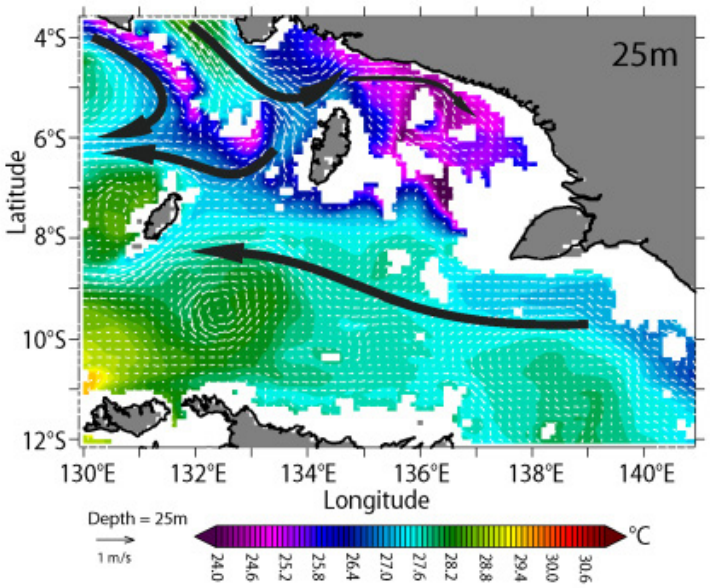


Figure 11b. (Continued)

Nevertheless, at 25 m depth to the deeper layer, reversed current vectors were found in the northern Aru island, where much stronger southeastward currents from Aru basin entered the Arafura shelf-slope (between Aru and Papua), and flowed eastward into the shelf, and the remaining flow southward and deflected westward into Banda Sea (Figure 11b). Here, a contrast of circulation was found: westward flow in surface layer that bring surface water away from coastal waters, and eastward inflow below subsurface that bring Aru basin water into the shelf waters. This strong southward Aru basin inflow may depict regional circulation of the Indonesian Throughflow (ITF) from Halmahera Sea that flows along outer Banda Arcs and enters into the study area (Masoleh &Atmadipoera, 2018). Trajectory analysis of water particles from Halmahera Sea confirmed the pathways of strong seasonal flow from Halmahera along outer Banda arcs during the SEM period (Wattimena, 2020). Thus, the presence of strong subsurface Banda arcs - ITF inflow into the Arafura shelf-break between Aru and Papua at 134°E can supply colder, saltier and nutrient-rich water and modulate the intensity of Ekman upwelling there. Kämpf (2015; 2016) suggested that this inflow was derived from Banda sea slope water since the model domain covered only eastern part of Banda Sea constrained to 132°E in the west and to 4°S in the north.

Schematic upwelling processes during the SEM period in the study area is presented in Figure 12. Much stronger southeasterly winds during the SEM period is associated with much stronger magnitude of surface current speeds. Susanto et al. (2001) suggested

that the monsoonal winds were the main forcing for generating coastal upwelling. Here, along-shore southeasterly winds generate surface current deflected about 45° to the left due to Coriolis effect from the wind direction (about westward). Below surface layer current directions are also deflected consecutively until the Ekman depth layer, forming the Ekman spiral. If we integrated this flow from surface to Ekman depth, then we have Ekman transport (Et) flowing southwestward or away from the coastal water (Figure12a). This creates horizontal pressure gradient between offshore and coastal waters, indicating by low sea surface height and reducing coastal water mass which is restored by upwelling water from deeper layer (Figure12b). Thus, surface Ekman transport flowing offshore generates upwelled deeper water in the coastal waters (Stewart, 2006).

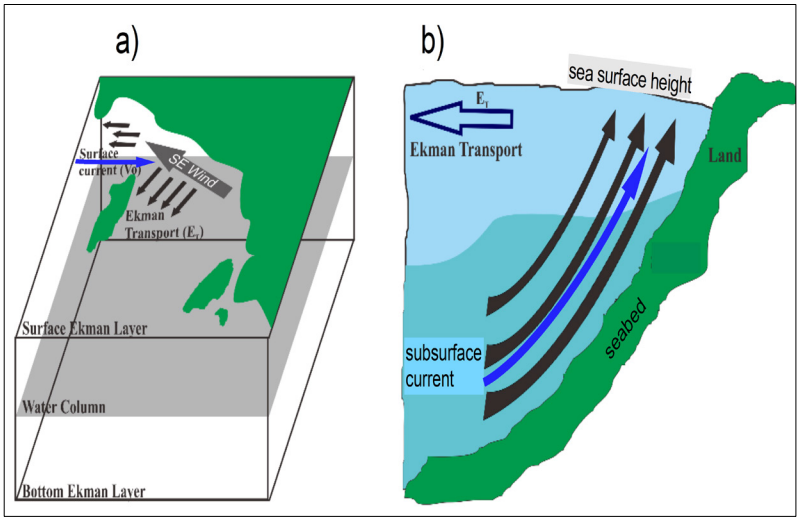


Figure 12. Schematic upwelling processes in the study area, (a) 3-dimension landscape figure; and (b) schematic cross-section in water column

In particular, in northern Arafura shelf sea, upwelling dynamics is not only controlled by the southeasterly winds that generate Ekman upwelling as described above, but also may be modulated by the subsurface current that inflows from the shelf-break into the inner shelf waters. Regional current system along the outer Banda arcs as a branch current of Indonesian throughflow (ITF) may contribute significantly on the upwelling process which then modulate intensity of upwelling. The presence of outer Banda arcs current – ITF may favor the upwelling intensity during the SEM period.

Further analysis on this subsurface deeper layer from a depth-latitude between Aru and Papua at 134.25 °E (please refer to yellow-line in Figure 1), revealed a reversed current

between surface and subsurface layer (Figure 13). During different monsoon period the core of strong eastward inflow into Arafura shelf slope appeared much stronger and larger from subsurface (20 m depth) to approximately 100 m depth during the SEM period than that during the NWM period from surface to about 50 m depth (Figure 13a). The westward flow in the upper 25 m was seen clearly during the SEM period. This current was a direct response to the southeasterly winds. At least, in the mid-section to the south the eastward flow was present during both monsoon periods, but the eastward flow during the 'SEM' upwelling period played a significant role to modulate upwelling dynamics by charging drifted surface water of Ekman transport.

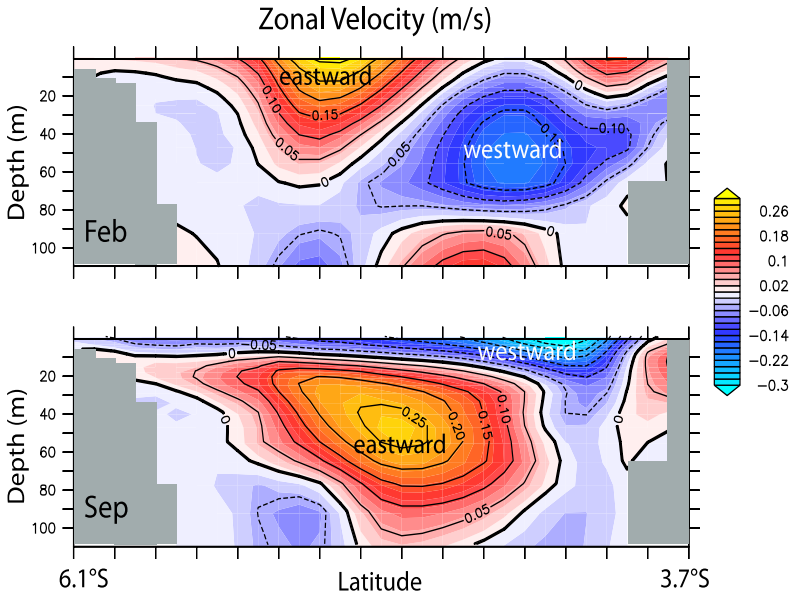


Figure 13a. Cross-section of model zonal current component in (a) February, and (b) September 2017. (Please refer to yellow line in Figure 1)

Furthermore, a depth-longitude plot of model zonal current component in the study area (please refer to red-line in Figure 1) shown in Figure 13b, exhibits the upwelling processes during the SEM period (September) as described above. Surface current flowed westward or away from coastline, while subsurface shelf-break current below 20 m depth to the bottom flows eastward that supplied colder, saltier and nutrient-rich water from deeper layer to the upwelling area (Figure 13b). This is contrast to the NWM period (February) when the surface current flowed eastward, accumulating surface water in coastal area, and created subsurface current flowing westward (Figure 13a).

This zonal current pattern in September 2017 is in good agreement with cross-section of temperature, salinity and potential density anomaly (Figure 7-9), in which subsurface eastward shelf-break flow (Figure 13b, in Sep) is associated with a homogeneous water mass in the second mixed layer or below the barrier layer with sharp stratification of uplifting isotherm of 25°C, isohaline of 34 psu and isopycnal of 22.5 kg/m^3 from deeper layer (30-40 m depth) at shelf-break area to near-surface layer (10 m) in the inner shelf area. This subsurface eastward flow has been also discussed by Kämpf (2015; 2016), suggesting that this existence of undercurrent-driven upwelling below 20-30 m depth was derived from Banda slope water. We found that this eastward inflow was as a branch current originated from the outer Banda arcs current flowing from Halmahera Sea into eastern Seram Sea and flowed along outer Banda arcs and entered northern Arafura shelf sea, as part of the ITF.

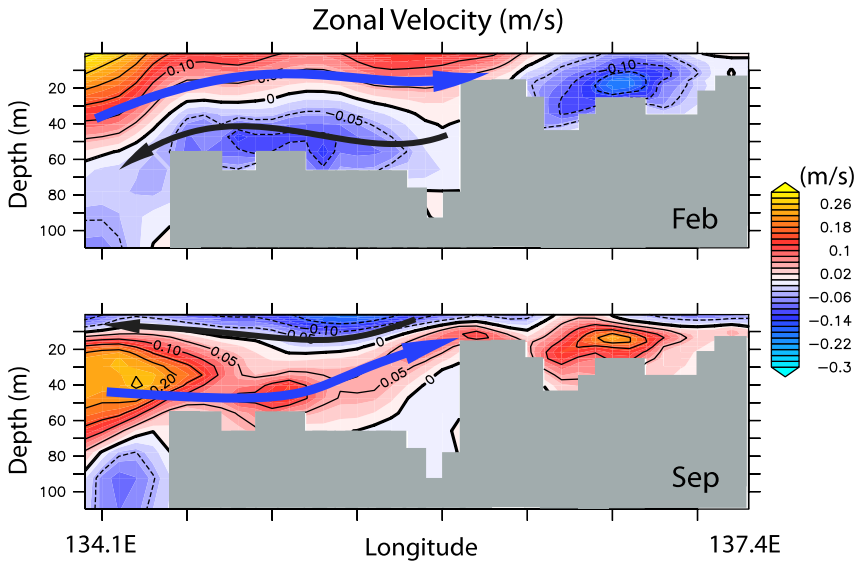


Figure 13b. Cross-section of model zonal current component in (a) February, and (b) September 2017. (Please refer to red line in Figure 1)

Estimates of Transport Volume in the Upwelling Region

In this study, two methods had been applied to calculate transport volume for quantifying Ekman transport, near-surface, and subsurface transports due to upwelling: by using classical "atmospheric" methods which is based on wind-stress calculation at a particular section, and by using "ocean" method, based on integrating of model zonal current component at a meridional section.

Monthly average of Ekman transport (TE) was calculated from wind stress data along a section (please refer to green-line in Figure1) to quantity of transported water mass away from coastline during upwelling event or toward coastline (downwelling period). It is found that from December to April (during the NWM period) the positive Ekman transport was toward the coastal area, which is associated with downwelling process, with maximum transport in February (0.46 Sv) (Figure 14). On contrary, negative Ekman transport (toward offshore) was revealed from May to November during the SEM period (upwelling period) with maximum transport to offshore in June (-0.71 Sv) (Figure 14). Negative Ekman transport was associated with coastal upwelling event from May to November, which is consistent with previous discussion (Wyrtki, 1961). As described in the method, Ekman transport estimate is sensitive to the wind-stress datasets. For example, wind-stress series is maximum in June 2017 which is related to the maximum of Ekman transport in June 2017 (Figure 14)

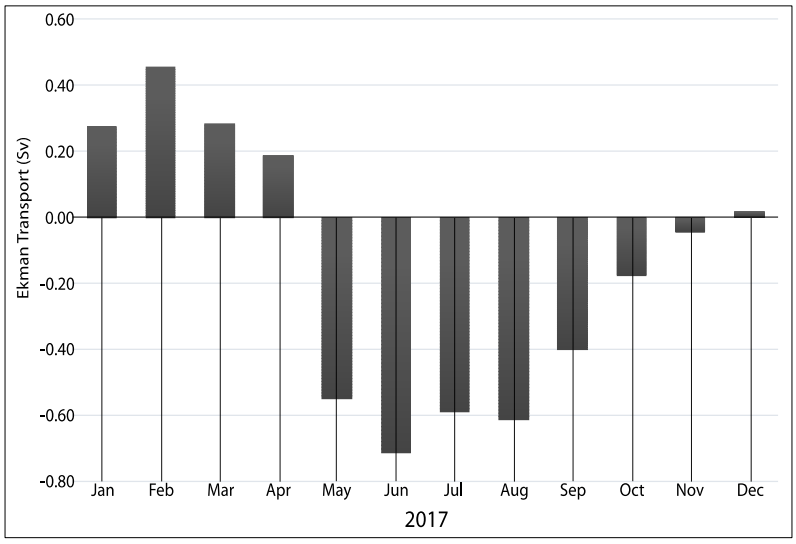


Figure 14. Monthly average of Ekman transport in coastal upwelling region in 2017 (please refer to green line in Figure 1)

Transport volume is also estimated from integration of zonal current component from the ocean circulation model which has been configured with both atmospheric and oceanic forcing in the model. Here, based on monthly average of zonal current in September 2017 (Figure 13a), we calculated transport volume at three layers (0-25m, 25-55m, and 55-110m). As expected, in the first layer (0-25 m depth) during the upwelling period a westward transport was seen clearly from May to end of September 2017 with its mean transport of -0.28 (±0.34) Sv (westward) and its transport maximum of -0.899 Sv at the end of August

- early of September (Figure 15). At second layer (25-55 m depth) the mean transport volume during the SEM period was +0.65 (±0.19) Sv (eastward), and at third layer (55-110 m depth) the mean transport was +0.41 (±0.213) Sv (eastward). The mean transport from 25 m to 110 m depth was +1.06 (±0.29) Sv. This implies that an inflow from Aru basin into the shelf waters during the SEM period may regulate the Ekman upwelling and also supply Arafura shelf sea with much colder, saltier and nutrient-rich water originated from deeper outer Banda arcs throughflow.

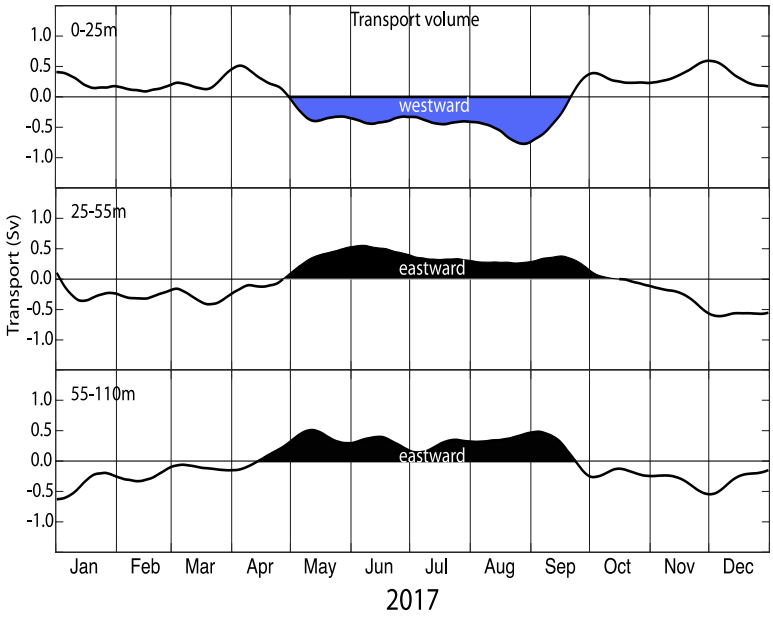


Figure 15. Daily transport volume series, integrated from 25m to sea surface (upper panel), from 55 m to 25 m depth (middle panel), and from 110 m to 55 m depth (lower panel) (please refer to yellow line in Figure 1). The transport volume time-series are smoothed with 15-day cut-off.

CONCLUSION

Physical processes and dynamics of upwelling in the northern Arafura Sea in 2017 had been investigated, by using multi-datasets from CTD casts, satellite-derived parameters, and validated ocean circulation model outputs. The evidence of upwelling clearly observed from a sharp barrier layer separating first mixed layer from second homogeneous layer beneath subsurface layer. This barrier layer formed large upsloping isotherms of 23.5 - 25.5°C, isohalines of 33.5 - 34.25 psu, and 21.8 - 23.2 kg/m^3 from deeper layer (30-40 m depth) in the shelf-break area to subsurface layer in the inner shelf waters, where much

colder and saltier water occured in shallower depth. A homogeneous water in the upper-layer is related to Ekman upwelling drift that is generated by the southeasterly monsoon winds, but the second homogeneous water beneath the barrier layer is belong to a distinct water mass which may be originated from the deeper shelf-break area. The existence of two mixed layer separated by the barrier layer corroborates a reversed current direction, where upper-layer current flowing westward – away from the shelf and strong subsurface current flowing eastward into the shelf. Therefore, the upwelling dynamics here is generated mainly by the persistent southeasterly winds leading to Ekman drift westward in the upper-layer, and is modulated by the subsurface eastward current from deeper shelf-break into the inner-shelf. The latter is not local current but a regional current which is considered derived from the outer Banda arcs current system as a current branch of the Indonesian Throughflow. Mean transport volume in the upper layer (0-25 m depth) between Aru and Papua during upwelling period (May – September) was -0.28 (±0.34) Sv (westward), and transport between 25m - 110m was +1.06 (±0.29) Sv (eastward), suggesting this inflow may regulate the upwelling and supply colder, saltier, and nutrient-rich water into Arafura shelf water. Transport volume derived from wind-stress calculation revealed its maximum in June, which is different from the ocean current calculation from May to September and its maximum in August-September. Surface ocean parameters reveal that onset (termination) of upwelling in May (October) is indicated by the reversed southeasterly winds, negative anomaly of SST, SSH, air temperature, but positive anomaly of chlorophyll-a.

ACKNOWLEDGEMENTS

The authors express their gratitude to the captain and the crews of Research Vessel GEOMARIN 3 for professional works and helps during field experiment. We also thank H. Catur Widiatmoko, MSc., team leader, for an excellent onboard coordination. Satellite derived sea surface temperature and surface chlorophyll-a dataset were obtained from Mercator Ocean/ Copernicus, available at (https://www.marine.copernicus.eu). Ocean general circulation model output datasets performed by CLS Toulouse France and were obtained from INDESO, available at (https://www.indeso.go.id). Surface meteorological datasets were obtained from the ECMWF, available at (https://www.apps.ecmwf.int). The Author also thank to the reviewers who gave a constructive criticism for the paper improvement.

REFERENCES

Alongi, D., Edyvane, K., do Ceu Guterres, M., Pranowo, W., Wirasantosa, S., & Wasson, R. (2011). *Biophysical profile of the Arafura and Timor Seas.* Jakarta, Indonesia: Arafura and Timor Seas Ecosystem Action (ATSEA) Program.

Atmadipoera, A.S., Khairunnisa, Z., & Kusuma, D. W. (2018). Upwelling characteristics during El Nino 2015 in Maluku Sea. *IOP Conference Series: Earth and Environmental Science, 176*(1), 1-18.

Atmadipoera, A. S., & Widyastuti, P. (2015). A numerical modeling study on upwelling mechanism in Southern Makassar Strait. *Jurnal Ilmu dan Teknologi Kelautan Tropis, 6*(2), 355-371.

Condie, S. A. (2011). Modeling seasonal circulation, upwelling and tidal mixing in the Arafura and Timor Seas. *Continental Shelf Research, 31*(14), 1427-1436.

Copernicus Marine Service. (2019). *Product user manual for all ocean colour products.* Retrieved June 18, 2020, from https://resources.marine.copernicus.eu/documents/PUM/CMEMS-OC-PUM-009-CHL-A.pdf

Dewi, D. M. P. R., Fatmasari, D., Kurniawan, A., & Munandar, M. A. (2018). The impact of ENSO on regional chlorophyll-a anomaly in the Arafura Sea. *IOP Conference Series: Earth and Environmental Science, 139*(1), 1-18.

Emery, W. J., & Thomson, R. E. (2004). *Data analysis methods in physical oceanography.* Massachusetts, USA: Elsevier.

Garnesson, P., Mangin, A., d'Andon, O. F., Demaria, J., & Bretagnon, M. (2019). The CMEMS GlobColour chlorophyll a product based on satellite observation: multi-sensor merging and flagging strategies. *Ocean Science, 15*(3), 819-830.

Gordon, A. L., & Susanto, R. D. (2001). Banda sea surface-layer divergence. *Ocean Dynamics, 52*(3), 0002-0010.

Kämpf, J. (2015). Undercurrent-driven upwelling in the northwestern Arafura Sea. *Geophysical Research Letters, 42,* 9362-9368.

Kämpf, J. (2016). On the majestic seasonal upwelling system of the Arafura Sea. *Journal of Geophysical Research Oceans, 121,* 1218-1228.

Kämpf, J., & Chapman, P. (2016). *Upwelling systems of the world: A Scientific journey to the most productive marine ecosystems.* Berlin, Germany: Springer International Publishing.

Kida, S., & Wijffels, S. (2012). The impact of the Indonesian throughflow and tidal mixing on the summertime sea surface temperature in the western Indonesian Seas. *Journal of Geophysical Research: Oceans, 117*(9), 1-14.

Kida, S., & Richards, K. J. (2009). Seasonal sea surface temperature variability in the Indonesian seas. *Journal of Geophysical Research Oceans, 114*(C6), 1-17.

Kutsuwada, K. (1998). Impact of wind/wind-stress field in the North Pacific constructed by ADEOS/NSCAT data. *Journal of Oceanography, 54*(5), 443-456.

Masoleh, V.C., & Atmadipoera, A.S. (2018). Coherence of transport variability along the outer Banda arcs. *IOP Conference Series: Earth and Environmental Science, 176,* 1-13.

Pond, S., & Pickard, G.L. (1983). *Introductory dynamical oceanography.* Massachusetts, USA: Elsevier Butterworth-Heinemann.

Rosdiana, A., Prartono, T., Atmadipoera, A. S., & Zuraida, R. (2017). Nutrient and chlorophyll - A distribution in Makassar upwelling region: From MAJAFLOX cruise 2015. *IOP Conference Series: Earth and Environmental Science, 54*(1), 1-10.

Sarhan, T., García-Lafuente, J., Vargas, M., Vargas, J. M., & Plaza, F. (2000). Upwelling mechanisms in the northwestern Alboran Sea. *Journal of Marine Systems, 23*(4), 317-331.

Sari, Y. D., Syaukat, K., Kusumastanto, T., & Hartoyo, S. (2018). Management of demersal fishery in the Arafura Sea: A bio-economic approach. *Jurnal Sosial Ekonomi Kelautan dan Perikanan, 13*(1), 43-57.

Schiller, A. (2011). Ocean circulation on the North Australian Shelf. *Continental Shelf Research, 31*(10), 1087-1095.

Sea-Bird Scientific. (2017). *Product manual SBE 19plus V2 SeaCAT profiler CTD.* Retrieved October 25, 2018, from https://www.seabird.com/asset-get.download.jsa?id=54627862329

Sprintall, J., & Liu, W. (2005). Ekman mass and heat transport. *Oceanography, 18*(4), 88-97.

Sproson, D., & Sahlée, E. (2014). Modelling the impact of Baltic Sea upwelling on the atmospheric boundary layer. *Tellus A: Dynamic Meteorology and Oceanography, 66*(1), 1-16.

Stewart, R. H. (2006). *Introduction to physical oceanography.* Texas, USA: Department of Oceanography, Texas A & M University.

Susanto, R. D., Gordon, A. L., & Zheng, Q. (2001). Upwelling along the coasts of Java and Sumatra and its relation to ENSO. *Geophysical Research Letters, 28*(8), 1599-1602.

Taufikurahman, Q., & Hidayat, R. (2017). Coastal upwelling in Southern Coast of Sumbawa Island, Indonesia. *IOP Conference Series: Earth and Environmental Science, 54*(1), 1-7.

Tomczak, M. (2014). *Lecture notes in oceanography.* California, USA: CreateSpace Independent Publishing Platform.

Tranchant, B., Reffray, G., Greiner, E., Nugroho, D., Koch-Larrouy, A., & Gaspar, P. (2015). Evaluation of an operational ocean model configuration at 1/12° spatial resolution for the Indonesian seas – Part 1: Ocean physics. *Geoscientific Model Development Discussions, 8*(8), 6611-6668.

Utama, F. G., Atmadipoera, A. S., Purba, M., Sudjono, E. H., & Zuraida, R. (2017). Analysis of upwelling event in Southern Makassar Strait. *IOP Conference Series: Earth and Environmental Science, 54*(1), 1-9.

Wattimena, M. C. (2020). *Structure and variability of the Indonesian throughflow in Halmahera Sea* (Doctorate Dissertation). IPB University, Indonesia.

Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (2nd Ed.). New York, NY: Academic Press.

Worsfold, M., Good, S., Martin, M., McLaren, A., Robert-Jones, J., & Fiedler, E. (2020). *Global ocean OSTIA sea surface temperature reprocessing SST-GLO-SST-L4-REP-OBSERVATIONS-010-011.* Product user manual, Copernicus Marine Environment Monitoring Service. Retrieved June 20, 2020, from https://resources.marine.copernicus.eu/documents/PUM/CMEMS-SST-PUM-010-011.pdf

Wyrtki, K. (1961). *Physical Oceanography of the Southeast Asian Waters, Naga Report Volume 2.* San Diego, California: Scripps Institution of Oceanography, the University of California.

Pertanika J. Sci. & Technol. 28 (4): 1517 - 1529 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Pattern and Trend of Land Surface Temperature Change on New Guinea Island

Munawar[1,2], Tofan Agung Eka Prasetya[1,3], Rhysa McNeil[1,4]* and Rohana Jani[5]

[1]Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Muang Pattani, 94000 Thailand
[2]Faculty of Mathematics and Science, Syiah Kuala University, Jl. Syech Abd.Rauf, Kopelma Darussalam, Banda Aceh, Aceh 23111, Indonesia
[3]Vocational Faculty, Universitas Airlangga, Jl. Dharmawangsa Dalam Selatan No 68, Airlangga, Gubeng, Surabaya, East Java, Indonesia
[4]Centre of Excellence in Mathematics, Commission on Higher Education (CHE), Ministry of Education, Ratchathewi, Bangkok, 10400 Thailand
[5]Faculty of Economics and Administration, University of Malaya, 50603 UM, Kuala Lumpur, Malaysia

ABSTRACT

Global warming will have an impact on nature in many ways, including rising sea levels and an increasing spread of infectious diseases. Land surface temperature is one of the many indicators that can be used to measure climate change on both a local and global scale. This study aims to analyze the change in land surface temperatures on New Guinea Island using a cubic spline method, autoregressive model, and multivariate regression. New Guinea Island was divided into 5 regions each consisting of 9 subregions. The data of each subregion was obtained from the National Aeronautics and Space Administration moderate resolution imaging spectroradiometer database from 2000 to 2019. The average change in temperature was +0.012°C per decade. However, the changes differed by region; significantly decreasing in the northwest at -0.107°C per decade (95% CI: -0.207, -0.007), significantly increasing in the south at 0.201°C per decade (95% CI: 0.069, 0.333), and remaining stable in the centralnorth, southeast and northeast.

Keywords: Cubic spline, global warming, land surface temperature, New Guinea Island

ARTICLE INFO

Article history:
Received: 12 April 2020
Accepted: 15 June 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.20

E-mail addresses:
munawar@unsyiah.ac.id (Munawar)
tofank3@gmail.com (Tofan Agung Eka Prasetya)
rhysa.m@psu.ac.th (Rhysa McNeil)
rohanaj@um.edu.my (Rohana Jani)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Climate change, particularly rising temperatures, is one of the important environmental problems facing the world today. Land surface temperature (LST) can provide insights into climatological processes, land surface energy relations and water stability at regional and global scales (Li et al., 2013; Wongsai et al., 2017), including climate change effects. These changes in climate can severely affect human health, the environment, and economic and social development (Marjuki et al., 2016; Mboera et al., 2011; Mishra et al., 2010). Climate change has a significant association with human disease vulnerability (Wu et al., 2016). This is manifested in the slowing down of the long-period decrease in the incidence of undernutrition, which is somewhat linked to extreme climatic events (Wheeler & Braun, 2013). LST is commonly used to assess rising temperatures.

The LST average around the world will continue to increase (Mildrexler et al., 2018). In tropical areas there has been wide variations in the level of increase in average surface temperatures. The variation depends on many factors such as elevation, normalized difference vegetation index (NDVI), and land cover (Alavipanah et al., 2015; Sun et al., 2012). The great variation on land elevation has a significant impact on the LST (Gao et al., 2008).

In South East Asia, a relationship exists between landscape composition and average LST in Jakarta, Bangkok, and Manila, in which green space was found to be cooler by 3^{o}C compared to those of impervious surfaces (Estoque et al., 2017). Some parts of South East Asia such as Indonesia (including Papua province), Malaysia and Papua New Guinea have experienced land use change, mainly for palm oil cultivation (Agus et al., 2013). Indonesia is the region with the highest level of land use change as a result of the development of palm oil farms and agriculture, and this has had a direct impact on LST (Ramdani et al., 2014; Sabajo et al., 2017). Papua New Guinea has also suffered from forest degradation and half of its forests will be damaged by 2021 (Filer et al., 2009).

New Guinea Island is the second biggest island after Greenland (Permana, 2011). The western half of the island forms a part of Indonesia and the eastern half contains the sovereign state of Papua New Guinea. Natural vegetation in this region consists of tropical rain forests in the lowlands and mountains, although there is a savannah area on the southern coast which has a different seasonal climate (Bowler et al., 1976). Annual maximum and minimum temperatures in New Guinea have been increasing in accordance with the global pattern (International Climate Change Adaptation Initiative, 2007). The New Guinea LST variation will affect LST in the surrounding islands, especially those in Indonesian archipelago, Australia, and even the larger Asia continent (Mildrexler et al., 2018). The extreme temperature was one of the factors of significant loss of suitable habitats for plant species in Papua and Papua New Guinea (Robiansyah, 2018).

Several statistical techniques were used to investigate the temperature changes. Cubic splines are widely used for smoothing data, especially data obtained from satellites. Data is fitted using least squares linear regression (Mao et al., 2017; Smith et al., 1974; Wüst et al., 2017). The technique has been used to model the vegetation index in Nepal (Wongsai et al., 2017; Sharma et al., 2018). There are approaches for the best possible selection of those parameters based on procedures to add knots in intervals where the residuals show trends as signaled by autocorrelation or in intervals where the residuals are inadmissibly significant (Wold, 1974). A first-order autoregressive model was used to fit fluxes in humid subtropical monsoon areas (Kumar et al., 2009). Therefore, the objective of this study was to investigate the change in day land surface temperatures on New Guinea Island during 2000 to 2019 using appropriate statistical methods.

MATERIALS AND METHODS

Study Area

The area of this study was New Guinea Island located at 130° to 152° east longitude and -11° to 0° south latitude (Figure 1). The New Guinea super region contains 5 regions with each region consisting of 9 subregions. The first region is located in the northwest and includes subregions 1 to 9 while the last region is located in the southeast and includes subregions 37 to 45. The sample points are located around parallels of latitude 210 pixels widths (190 km) apart. The subregion as the sample point spread with the equal distance, each comprising 49 pixels in a 7×7 array, covering the New Guinea mainland.

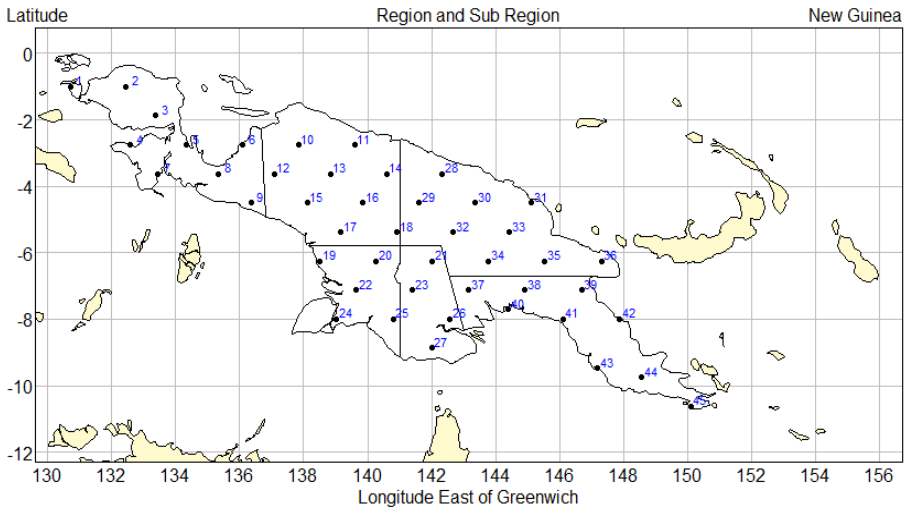


Figure 1. New Guinea Island: area of study

Data

The LST data, which is freely available to download from the MODIS LST (ORNL DAAC, 2018) database, contains average temperatures every 8 days of clear sky for pixels each of area 0.859 km^2. To ensure area equality of all pixels, a sinusoidal projection with tiles of size 10 ×10 latitude degrees was used, with each tile in turn divided into 1200 ×1200 pixels. To avoid missing data, the download was based on the center of the subregion on the island. If there were any missing values, then they were deleted. An unexpected natural disaster (e.g. forest fire, landslide, or tsunami) that may possibly cause a sudden change of data behavior, was excluded. Outliers were kept in the data set to have a comprehensive view of the LST data. The original LST temperature measurements were stored in degrees Kelvin and then converted into degrees Celsius.

The MODIS LST time series data has coverage of global and regular monitoring, include the topography of an area. The Terra (land) satellite will provide the LST data during the daytime and nighttime. The LST data and climate component inland is also influenced by atmospheric and land processes (Luintel et al., 2019; Wan et al., 2015).

MODIS LST data were collected over time with fluctuations due to the season (Wongsai et al., 2017). The seasonal pattern was assumed to be the same for every year and the change in other parameters such as the land cover change that has a direct or indirect effect on the LST data is consistently increasing or decreasing.

Methods

Cubic spline functions are defined as piecewise polynomials of degree *r*. The joined pieces are called knots. A spline function of degree *r* is a continuous function with *r* – 1 continuous derivatives (Wahba, 1990; Wold, 1974). The formula of the cubic spline function is (Equation 1):

$$S\left(t = a + bt + \sum_{k=1}^{p} c_k (t - t_k)_+^3\right) \quad [1]$$

where *S* is the spline function, *t* denotes the time in Julian calendar, specified knots are $t_1 < t_2 < ... < t_p$ and $(t - t_k)_+$ is $(t - t_k) > 0$ for $t > x$ and 0 otherwise. The boundary conditions require that $S(t)$ for $t < t_1$ equals $S(t)$ for $t > t_p$. a, b, c_k are the coefficients of the combination between a linear and cubic spline model.

Selecting the position and number of knots for smoothing the spline curve is an important issue. The placement of the knots relies on the LST data in a tropical area with a rainy and dry season (Kohavi, 1995; Lukas et al., 2010; Wongsai et al., 2017). In the different regions of the biosphere, inter-seasonal variation can affect the land surface temperature variation (Singh et al., 2014). Areas that have both dry and rainy seasons will have lower LST during the rainy season (de Jesus & Santana, 2017). In a tropical area, changes in LST may be linked with heatwaves (during April and May) and rainfall (during

June-September) (Gogoi et al., 2019). We used 8 knots and placed 4 knots at the beginning of the year and the remaining 4 knots at the end of the year based on the seasons that are characteristic of tropical regions.

The LST was seasonally adjusted using the Equation 2:

$$Y_a = Y - S_f + \bar{x} \quad [2]$$

where Y_a is a seasonally adjusted time series for LST, Y is an observed data (LST per day) for 18 years, S_f is a vector of spline fitted values that we estimated from the cubic splines and $\bar{x}$ is the average LST per year.

A second order autoregressive model AR(2) was used to fit the LST seasonally adjusted. The model is given by Equation 3:

$$Y_{at} = \alpha_1 Y_{at-1} + \alpha_2 Y_{at-2} + \varepsilon_t \quad [3]$$

where Y_{at} is the seasonally adjusted LST at time t, and Y_{at-1} is the LST at time t-1, t = 1, …, 365 days, α_1 and α_2 are unknown parameters to be estimated and ε_t is the random error with zero mean and finite variance (Chan & Wei, 1987).

A multivariate regression model (Mardia et al., 1979) was then used to analyse the seasonally adjusted LST data. The model is given by Equation 4:

$$Y = XB + U \quad [4]$$

where Y is the outcome matrix of variables with dimension $n \times m$, n is the number of observations, m is the number of subregions, X is a matrix of independent variables $n \times q$, q is the number of independent variables, B is a regression parameter matrix with dimension $q \times m$ and U is an unobserved random disturbance matrix.

All analyses and graphical displays were carried out using R (R Core Team, 2018).

RESULTS AND DISCUSSION

Region 1 of New Guinea Island was used to represent the results of this study. Figure 2 shows the LST for each day in region 1 of New Guinea Island where Figures 1 and 2 depict the distribution of LST records.

The vertical axis denotes average temperatures on the same day for each of the 18 years. The solid red curves are the fitted natural spline functions with 8 knots denoted by blue crosses. Land surface temperatures showed a moderate seasonal pattern with two summer peaks commonly found in tropical zones.

During the years from 2000-2019 the lowest LST corresponded to day 210 which was during the rainy season and the highest LST corresponded to day 324. The highest average LST was 27.7°C which occurred in subregion 9 of Region 1. The R-squared was 0.31. The lowest R-squared was 0.034 which occurred in the model for subregion 4. The cubic spline with 8 knots was fitted to the LST data and the annual LST data showed a seasonal pattern.

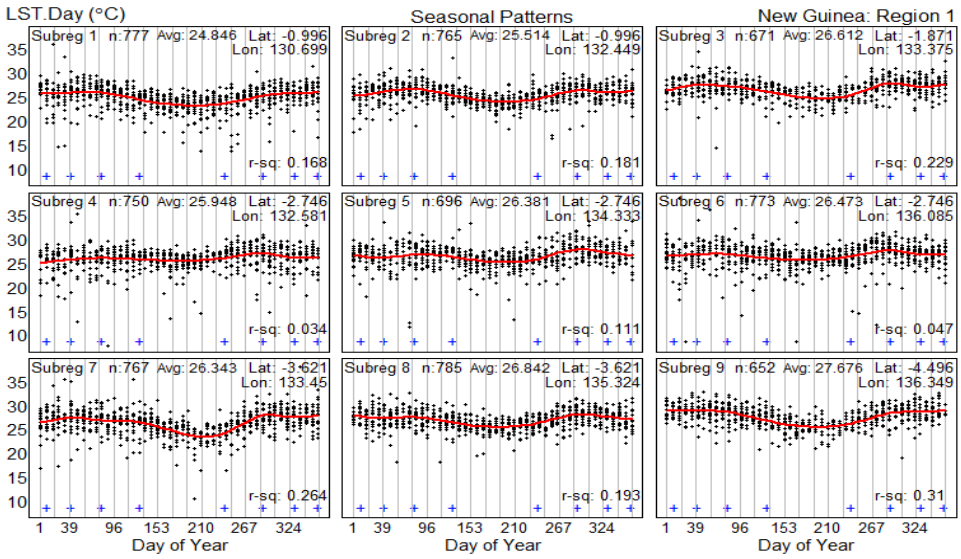


Figure 2. Land surface temperatures in New Guinea Region 1 showing a seasonal pattern

Figure 3 displays the seasonally adjusted LST. The estimated coefficients of a_1 and a_2 were very low, indicating that the time series of daily temperatures are independent.

In Figure 3, the dotted lines in the right panel indicate that LST decreased in subregions 1, 2, 7, and 9, increased in subregions 3, 4, and 8 and remained stable in subregions 5 and 6 over the 18 year period, but p-values for the linear models (zero knots) with two parameters indicate that none of these changes were statistically significant. The thick curves in the right panel show fitted cubic splines with seven knots (with significant p-values for subregions 4, 7 and 9) whereas the dashed lines (cubic splines with zero knots) show the same trends in 18 years for Region 1. The number of knots depends on the LST variation between years. We used 7 knots which divided the data equally into 4-year intervals (assuming that the variation of LST happens every 3 years). In the bottom-right panel of Figure 3, multivariate regression was used to reduce spatial correlation and estimate the mean LST for this region. We found a statistically significant decrease, with a z-value -2.098 and 95% confidence interval (-0.21 - 0.01) ^{o}C per decade.

The increases in LST (^{o}C per decade) for each region is shown in Figure 4. The overall mean increase was 0.024^{o}C per decade. There was a wide variation for each region. The mean change in daily land surface temperatures for the northwest, central-north, south, northeast and southeast regions were -0.11^{o}C, 0.002^{o}C, 0.20^{o}C, -0.07^{o}C and 0.03^{o}C, respectively. Only the south and northwest regions had significant changes in day land surface temperatures.

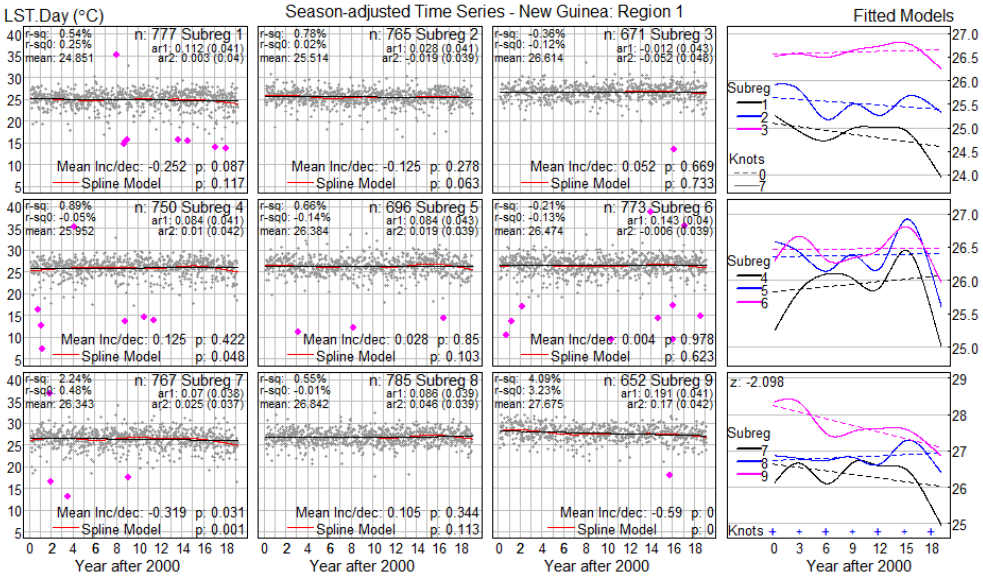


Figure 3. Seasonally adjusted land surface temperatures for New Guinea, Region 1

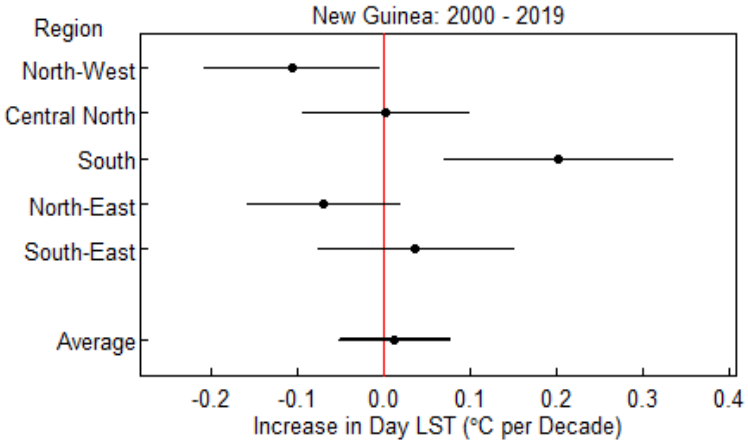


Figure 4. Increase in mean day land surface temperatures with 95% confidence intervals for New Guinea Island

This finding illustrates that the land surface temperatures in New Guinea Island would increase by 0.12°C per decade which was lower than that predicted for Papua New Guinea by the Papua New Guinea National Weather Service (predicted increase in temperatures ranging from 0.4–1.0°C, International Climate Change Adaptation Initiative 2007). The change in land surface temperatures in New Guinea Island was not significant. The islands laying on the equator line tend to have warm temperatures (stable) and the island close to

the pole consistently cold and the variation of LST depend on the position of a place on the earth and its elevation (Gillespie, 2014).

Figure 5 shows the results of LST change for the 45 subregions in New Guinea using the multivariate regression model. The probability that temperatures in each of the 5 regions increased, decreased, or remained stable were determined by averaging the LST trends for each subregion. The day LST decreased in the northwest, increase in the south, and is likely to be stable in the central-north and southeast, while it is likely to decrease in the north-east.

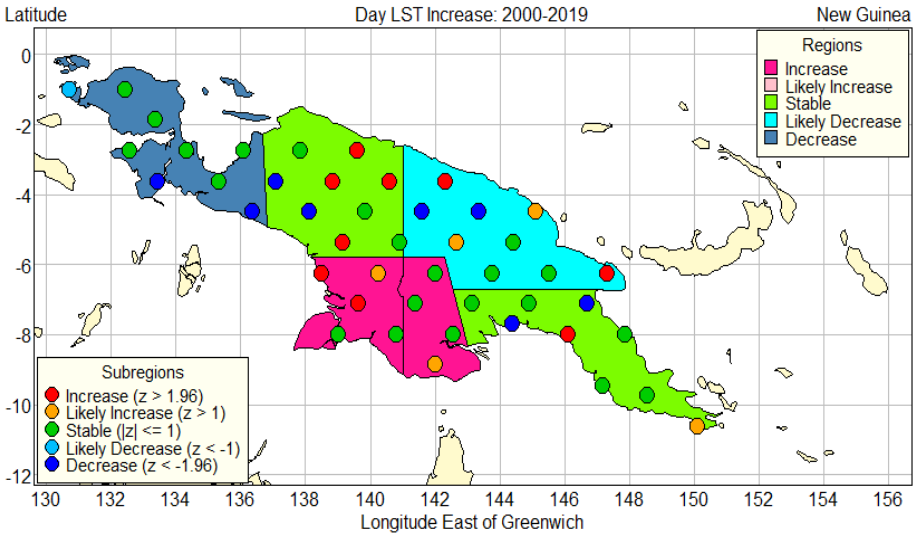


Figure 5. Trends in land surface temperatures (°C/decade) in New Guinea Island, 2000-2019

The seasonal pattern of the data showed that the highest LST for region 1 occurred between day 267 and day 324. However, the findings indicate that there was a seasonal pattern. Other studies conducted in areas with four seasons showed that the highest LST appeared during summer (Singh et al., 2014). LST has also been shown to be influenced by vegetation, land use/land cover (LULC) and surface solidity (Khandelwal et al., 2018).

Several studies have shown that LULC affects LST temperatures (Odindi et al., 2015; Rasul et al., 2017). The area will experience an increase in temperatures if the land cover is greatly reduced (Parmesan & Hanley, 2015). Green trees or other plants that cover the ground surface can absorb heat as a result of the reflection of sunlight through the evaporation process. The heat on the surface of the trees will metabolize the heat and convert it into other forms of energy therefore reducing the temperature on the land surface. This means that LULC in the form of healthy vegetation tends to not produce an increase in LST (Babalola & Akinsanola, 2016).

Vegetation or healthy plants in an area were affected by LST variation (Rahmad et al., 2019). Two studies reported on the increasing rate of forest conversion in the areas of Papua New Guinea and Papua (Alamgir et al., 2019; Austin et al., 2019). This deforestation is suspected to be the cause of an increase in temperatures in the area. A study found that areas with a low vegetation index will have high LST (Buyadi et al., 2014).

Temperature changes on the island of New Guinea occurred at a minor rate. Research shows that an increase in the temperature of an area can be reduced if the vegetation can be recovered again (Cooper et al., 2017). In other words, if the NDVI is increased in an area that experiences an increase in temperature, it is expected that this will reduce or prevent further increases in LST.

Use of a cubic spline and multivariate regression analysis was suitable for examining the seasonal pattern and variation in the LST for the 5 study regions. Only 2 of the 5 regions showed a significant change in LST. This might due to the number of subregions when compared to the size of New Guinea Island. We included 210 pixels in each grid, which equated to a longitudinal distance of 190 km between each subregion. It has been shown that increasing the sample size can improve the estimation of the true population mean (Storch & Zwiers, 1999; Mehta & Pocock, 2011).

CONCLUSIONS

With the appropriate number and placement of knots, the cubic spline model provided a satisfactory fit to the LST data on New Guinea Island. This study demonstrated a decreased LST in the northwest and south regions. There were variations in the increase of LST, although the increases were not significant. An increase in LST on the Island of New Guinea is an indication of global warming at the regional level. However, further investigations are needed to confirm these findings on a wider scale. Another approach is needed to improve the accuracy of estimation, especially to validate our findings for the same area such as Sumatra and Borneo islands, which are on the equator. The sample size can be increased by increasing the number of regions or subregions. Finally, since New Guinea Island contains many rainforests and mountainous areas, including other variables such as NDVI or land elevation in the analysis may help to improve the model fit.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Professor Don McNeil for his invaluable assistance. This research was supported by the Thailand's Education Hub for the Southern Region of ASEAN Countries (TEH-AC), Prince of Songkla University graduate school research grant and Centre of Excellence in Mathematics, commission on higher Education, Thailand.

REFERENCES

Agus, F., Gunarso, P., Sahardjo, B. H., Harris, N., Noordwijk, M. Van, & Killeen, T. J. (2013). *Historical CO_2 emissions from land use and land use change from the oil palm industry in Indonesia, Malaysia, and Papua New Guinea*. Kuala Lumpur, Malaysia: Roundtable on Sustainable Palm Oil.

Alamgir, M., Sloan, S., Campbell, M. J., Engert, J., Kiele, R., Porolak, G., … & Laurance, W. F. (2019). Infrastructure expansion challenges sustainable development in Papua New Guinea. *PLoS ONE*, *14*(7), 1-20.

Alavipanah, S., Wegmann, M., Qureshi, S., Weng, Q., & Koellner, T. (2015). The role of vegetation in mitigating urban land surface temperatures: A case study of Munich, Germany during the warm season. *Sustainability*, *7*(4), 4689-4706.

Austin, K. G., Schwantes, A., Gu, Y., & Kasibhatla, P. S. (2019). What causes deforestation in Indonesia? *Environmental Research Letters*, *14*(2), 1-10.

Babalola, O., & Akinsanola, A. (2016). Change detection in land surface temperature and land use land cover over Lagos Metropolis, Nigeria. *Journal of Remote Sensing and GIS*, *5*(3), 2-7.

Bowler, J. M., Hope, G. S., Jennings, J. N., Singh, G., & Walker, D. (1976). Late quaternary climates of Australia and New Guinea. *Quaternary Research*, *6*(3), 359-394.

Buyadi, S. N. A., Mohd, W. M. N. W., & Misni, A. (2014). Impact of vegetation growth on urban surface temperature distribution. *IOP Conference Series: Earth and Environmental Science, 18*(1), 1-7.

Chan, N. H., & Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *The Annals of Statistics, 15*(3), 1050-1063.

Cooper, L. A., Ballantyne, A. P., Holden, Z. A., & Landguth, E. L. (2017). Disturbance impacts on land surface temperature and gross primary productivity in the western United States. *Journal of Geophysical Research: Biogeosciences, 122*(4), 930-946.

de Jesus, J. B., & Santana, I. D. M. (2017). Estimation of land surface temperature in Caatinga area using Landsat 8 data. *Journal of Hyperspectral Remote Sensing, 7*(3), 150-157.

Estoque, R. C., Murayama, Y., & Myint, S. W. (2017). Effects of landscape composition and pattern on land surface temperature: An urban heat island study in the megacities of Southeast Asia. *Science of the Total Environment*, *577*, 349-359.

Filer, C., Keenan, R. J., Allen, B. J., & McAlpine, J. R. (2009). Deforestation and forest degradation in Papua New Guinea. *Annals of Forest Science*, *66*(8), 813-825.

Gao, M., Qin, Z., Qiu, J., Liu, S., Xu, B., Li, W., & Yang, X. (2008, September 15-18). Retrieving spatial-temporal variation of land surface temperature in Tibetan Plateau for the years 2005-2006 from MODIS satellite data. In *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VIII* (Vol. 7110, p. 71101A). Cardiff, Wales, United Kingdom.

Gillespie, A. (2014). Land surface temperature. In *Encyclopedia of Earth Sciences Series* (November 2019) (pp. 314-320). Retrieved December 1, 2019, from https://doi.org/10.1007/978-0-387-36699-9_79

Gogoi, P. P., Vinoj, V., Swain, D., Roberts, G., Dash, J., & Tripathy, S. (2019). Land use and land cover change effect on surface temperature over Eastern India. *Scientific Reports*, *9*(1), 1-10.

International Climate Change Adaptation Initiative. (2007). Current and future climate of Papua New Guinea. In *Pacific climate change science program*. Retrieved January 1, 2020, from http://www.pacificclimatechangescience.org/wp-content/uploads/2013/06/14_PCCSP_PNG_8pp.pdf

Khandelwal, S., Goyal, R., Kaul, N., & Mathew, A. (2018). Assessment of land surface temperature variation due to change in elevation of area surrounding Jaipur, India. *Egyptian Journal of Remote Sensing and Space Science*, *21*(1), 87-94.

Kohavi, R. (1995, August 20-25). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *14th International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 1137-1143). Montreal, Quebec, Canada.

Kumar, A., Kumar, M., Mahanti, N. C., & Mallik, C. (2009). Surface flux modelling using ARIMA technique in humid subtropical monsoon area. *Journal of Atmospheric and Solar-Terrestrial Physics*, *71*(12), 1293-1298.

Li, Z. L., Tang, B. H., Wu, H., Ren, H., Yan, G., Wan, Z., … & Sobrino, J. A. (2013). Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, *131*, 14-37.

Luintel, N., Ma, W., Ma, Y., Wang, B., & Subba, S. (2019). Spatial and temporal variation of daytime and nighttime MODIS land surface temperature across Nepal. *Atmospheric and Oceanic Science Letters*, *12*(5), 305-312.

Lukas, M. A., De Hoog, F. R., & Anderssen, R. S. (2010). Efficient algorithms for robust generalized cross-validation spline smoothing. *Journal of Computational and Applied Mathematics*, *235*(1), 102-107.

Mao, F., Li, X., Du, H., Zhou, G., Han, N., Xu, X., … & Cui, L. (2017). Comparison of two data assimilation methods for improving MODIS LAI time series for bamboo forests. *Remote Sensing*, *9*(5), 1-17.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. New York, NY: Academic Press, Inc.

Mboera, L. E. G., Mayala, B. K., Kweka, E. J., & Mazigo, H. D. (2011). Impact of climate change on human health and health systems in Tanzania: A review. *Tanzania Journal of Health Research*, *13*(5 SUPPL. ISS), 1-23.

Mehta, C. R., & Pocock, S. J. (2011). Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Statistics in Medicine*, *30*(28), 3267-3284.

Mildrexler, D. J., Zhao, M., Cohen, W. B., Running, S. W., Song, X. P., & Jones, M. O. (2018). Thermal anomalies detect critical global land surface changes. *Journal of Applied Meteorology and Climatology*, *57*(2), 391-411.

Mishra, A. K., Singh, V. P., & Jain, S. K. (2010). Impact of global warming and climate change on social development. *Journal of Comparative Social Welfare*, *26*(2-3), 239-260.

Odindi, J. O., Bangamwabo, V., & Mutanga, O. (2015). Assessing the value of urban green spaces in mitigating multi-seasonal urban heat using MODIS land surface temperature (LST) and landsat 8 data. *International Journal of Environmental Research, 9*(1), 9-18.

ORNL DAAC. (2018). *MODIS and VIIRS land products global subsetting and visualization tool*. Retrieved February 2, 2019, from https://doi.org/10.3334/ornldaac/1379

Parmesan, C., & Hanley, M. E. (2015). Plants and climate change: Complexities and surprises. *Annals of Botany*, *116*(6), 849-864.

Permana, D. S. (2011). Meteorological data analysis based on automatic weather stations at different elevation and radiosondes data in Papua. *Jurnal Meteorologi Dan Geofisika*, *12*(2), 151-162.

R Core Team. (2018). *R: A Language and environment for statistical computing*. Retrieved February 2, 2019, from https://www.r-project.org/

Rahmad, R., Nurman, A., & Pinem, K. (2019). Impact of NDVI change to spatial distribution of land surface temperature (A study in Medan city, Indonesia). In *1st International Conference on Social Sciences and Interdisciplinary Studies (ICSSIS 2018)* (pp. 167-171). Amsterdam, The Netherlands: Atlantis Press.

Ramdani, F., Moffiet, T., & Hino, M. (2014). Local surface temperature change due to expansion of oil palm plantation in Indonesia. *Climatic change, 123*(2), 189-200.

Rasul, A., Balzter, H., Smith, C., Remedios, J., Adamu, B., Sobrino, J., … & Weng, Q. (2017). A review on remote sensing of urban heat and cool islands. *Land*, *6*(2), 1-10.

Robiansyah, I. (2018). Assessing the impact of climate change on the distribution of endemic subalpine and alpine plants of new Guinea. *Songklanakarin Journal of Science and Technology*, *40*(3), 701-709.

Sabajo, C. R., Maire, G., June, T., Meijide, A., Roupsard, O., & Knohl, A. (2017). Expansion of oil palm and other cash crops causes an increase of the land surface temperature in the Jambi province in Indonesia. *Biogeosciences*, *14*, 4619-4635.

Sharma, I., Ueranantasun, A., & Tongkumchum, P. (2018). Modeling of satellite data to identify the seasonal patterns and trends of vegetation index in Kathmandu Valley, Nepal from 2000 to 2015. *Jurnal Teknologi*, *80*(4), 125-133.

Singh, R. B., Grover, A., & Zhan, J. (2014). Inter-seasonal variations of surface temperature in the urbanized environment of Delhi using landsat thermal data. *Energies*, *7*(3), 1811-1828.

Smith, J. R. E., Price, J. M., & Howser, L. M. (1974). *A smoothing algorithm using cubic spline functions*. Hampton, USA: Nasa Langley Research Center.

Storch, H. Von, & Zwiers, F. W. (1999). *Statistical analysis in climate research*. Cambridge, UK: Cambridge University press.

Sun, Q., Wu, Z., & Tan, J. (2012). The relationship between land surface temperature and land use/land cover in Guangzhou, China. *Environmental Earth Sciences*, *65*(6), 1687-1694.

Marjuki, van der Schrier, G., Tank, A. M. G. K., van den Besselaar, E. J. M., Nurhayati, & Swarinoto, Y. S. (2016). Observed trends and variability in climate indices relevant for crop yields in Southeast Asia. *Journal of Climate, 29*(7), 2651-2669.

Wahba, G. (1990). *Spline models for observational data (CBMS-NSF Regional Conference Series in Applied Mathematics)*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.

Wan, Z., Hook, S., & Hulley, G. (2015). *MOD11A2 MODIS/Terra land surface temperature/emissivity 8-day L3 global 1km SIN grid V006*. NASA EOSDIS Land Processes DAAC. Retrieved December 1, 2019, from https://doi.org/10.5067/MODIS/MOD11A2.006

Wheeler, T., & Braun, J. V. (2013). Climate change impacts on global food security. *Science, 341*(6145), 508-513.

Wold, S. (1974). Spline functions in data analysis. *Technometrics*, *16*(1), 1-11.

Wongsai, N., Wongsai, S., & Huete, A. R. (2017). Annual seasonality extraction using the cubic spline function and decadal trend in temporal daytime MODIS LST data. *Remote Sensing*, *9*(12), 1-17.

Wu, X., Lu, Y., Zhou, S., Chen, L., & Xu, B. (2016). Impact of climate change on human infectious diseases: Empirical evidence and human adaptation. *Environment International*, *86*, 14-23.

Wüst, S., Wendt, V., Linz, R., & Bittner, M. (2017). Smoothing data series by means of cubic splines: Quality of approximation and introduction of a repeating spline approach. *Atmospheric Measurement Techniques*, *10*(9), 3453-3462.

Pertanika J. Sci. & Technol. 28 (4): 1531 - 1547 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

The Effect of Alkaline Treatment onto Physical, Thermal, Mechanical and Chemical Properties of Lemba Leaves Fibres as New Resources of Biomass

Nur Aina Farhana Mat Nasir[1], Jamarosliza Jamaluddin[1]*, Zuraidah Zainudin[1], Mahirah Muhammad Busheri[1], Nadia Adrus[1], Fatria Syaimima Syaiful Azim[1] and Rosnani Hasham[2]

[1]Department of Chemical Engineering, School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[2]Institute of Bioproduct Development, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

ABSTRACT

The main purpose of this paper is to investigate the effect of alkaline treatment on the physical, thermal, mechanical and chemical properties of pristine lemba leaves fibres (LeLeFs). LeLeFs were treated with 6, 8, and 10 wt% sodium hydroxide (NaOH) solution at room temperature for 24 h. In order to determine the functional group presence after the alkaline treatment, LeLeFs were analyzed using Fourier Transform Infrared (FTIR) Spectroscopy. The density of LeLeFs treated with 10 wt% NaOH solution recorded the highest density with 1.168 g/cm^3. Morphology study showed that the diameter of fibre reduced with the increment of NaOH concentration. The removal of lignin and hemicellulose could be observed in the thermogravimetric analysis (TGA). Alkaline treatment enhanced the tensile properties of fibre and 10 wt% alkaline treated fibre resulted in the highest tensile strength, modulus and elongation of the fibre at 511.10 MPa, 11.76 GPa and 3.69% respectively. Chemical resistance analysis found that the treated fibre had better chemical resistance compared to untreated fibre. Therefore, it is substantiated that alkaline treatment affects the properties of LeLeF.

Keywords: Alkaline treatment, chemical resistance, density, FTIR, morphology, tensile, TGA

ARTICLE INFO

Article history:
Received: 11 February 2020
Accepted: 18 June 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.21

E-mail addresses:
nurainafarhana94@gmail.com (Nur Aina Farhana Mat Nasir)
jamarosliza@cheme.utm.my (Jamarosliza Jamaluddin)
zuraidah39@graduate.utm.my (Zuraidah Zainudin)
myrabusheri@gmail.com (Mahirah Muhammad Busheri)
nadia@utm.my (Nadia Adrus)
fatriaazim@gmail.com (Fatria Syaimima Syaiful Azim)
r-rosnani@utm.my (Rosnani Hasham)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

In recent years, the natural fibres have been regarded as great resources in replacing the synthetic polymers that widely used in various applications such as modern apparel, home furnishings, textile, medicine, aeronautics, building, construction and others. This is mainly due to raising awareness of the environment and the ever-depleting trend of petroleum supplies. Natural fibres are renewable, biodegradable and environmentally friendly. Moreover, they provide a lot of advantages over man-made fibres, including low-cost and abundantly available, relatively high specific strength and modulus, light weight, low density, less abrasiveness, and minimal health hazards (Hashim et al., 2017). Due to their abundant availability, natural fibres can be obtained or processed from a wide range of natural resources around the globe.

Malaysia, as a tropical country, has been blessed with an abundance amount of fibrous plants and agricultural resources. One of them is lemba (Figure 1) or scientifically known as *Curculigo latifolia*. Lemba is a member of the Hypoxidaceae family or known as flowering plants (Shaari, 2005). There are about 20 species of genus Curculigo that distributed in the tropical regions of Africa and Asia (Ranjbarfard et al., 2014). *Curculigo capitulata* and *Curculigo latifolia* are the familiar species that can be found in Malaysia and Borneo Island. Lemba grows about one meter tall and the blade elliptical leaf of 30-100 cm × 5-10 cm. It grows well in the hilly area and requires less exposure to sunlight, with abundant water supply (Shaari, 2005).

Figure 1. The lemba plant

Every part of this plant can be utilized to obtain its several advantages. The leaves have been used as wrapping, like banana leaves, by the people in Lahu, Thailand (Brink & Escobin, 2003). In Malaysia and Borneo, lemba leaves fibres (LeLeFs) have been used for making the rope, fishing net and twines (Farzinebrahimi et al., 2016). Meanwhile, in Japan, this plant also known to be a natural sweetener that gives a very sweet taste when its seeds or parts of the plant are chewed (Shaari, 2005). Not only that, high fever also can be treated by using the combination of leaves and flowers, while the concoction of flowers and roots has been used to treat stomachache and frequent urination (Shaari, 2005). The fruit has been recorded to be used to increase the appetite (Farzinebrahimi et al., 2016). It has been reported that lemba rhizome extract could be used to inhibit hepatitis B virus (Wiart & Wong, 2002).

Despite its useful applications, the knowledge of lemba is still restricted and limited to rural people only. There is an investigation on the potential use of LeLeF as a new material for textiles, where it found that LeLeF was much stronger compared to the cotton (Shaari, 2005). However, there is no scientific study in depth that explaining on the chemical and physical properties of LeLeF. Thus, it is worth further investigate the lemba leaves as a new resource of natural fibres in order to fully develop their potential application. Furthermore, newly identified fibres must be analyzed to identify their physical and chemical properties as this knowledge is important to evaluate the properties of the fibres efficiently.

Alkaline treatment is a method that commonly used by scientists and researchers on cellulosic fibres in order to produce high-quality fibres. This method will disperse bulk lignocellulosic materials into lignocellulosic fibres and remove lignin together with hemicellulose. Previously, the alkaline treated napier grass fibres have been studied for their chemical and physical properties. It shows that the alkali treatment enhanced the tensile properties of the fibres compared to those of the untreated fibres (Reddy et al., 2012). Besides, a few researches have been reported on the improvement of mechanical properties of pineapple leaf fibres after alkali treatment (Asim et al., 2016; Motaleb, 2018; Zin et al., 2018).

In this study, experiments were conducted to characterize the properties of pristine and treated LeLeF at different concentrations of alkaline. The characterization conducted includes determining the physical, thermal, mechanical and chemical properties. This investigation is very crucial to gain an established platform on the properties of LeLeF in determining the suitable potential application to be utilized. The removal of amorphous structure which are lignin and hemicellulose is very important as it can enhance the properties of the cellulose thus making it suitable to be used as a filler in polymer matrix and also can be used as a raw material for hydrogel preparation

MATERIALS AND METHODS

Materials

Lemba leaves used in this research were collected in the rural area at Jerantut, Pahang. Sodium hydroxide (NaOH, 98%), sodium carbonate (Na_2CO_3, 99.5%), ammonium hydroxide (NH_4OH,25% NH_3), toluene (99.5%), nitric acid (HNO_3, 69-72%), hydrochloric acid (HCl, 35-37%) and acetic acid (CH_3COOH, 99.8%) were purchased from Sigma Aldrich. All chemicals were analytical reagent grade and were used as received without further purifications.

Sample Preparation

LeLeFs were extracted from the leaves by using the hand-scrapping method. Firstly, lemba leaves were scrapped with a knife. The outer layers of the leaves were removed gently to ensure the fibres did not break and finally, LeLeFs were obtained.

LeLeFs were immersed in 6, 8 and 10 wt% NaOH solution at room temperature for 24 h in order to remove lignin and hemicellulose. After that, LeLeFs were washed several times with distilled water to remove any traces of alkali on the surface of the fibres. Then, they were neutralized with a diluted acetic acid solution and again were washed thoroughly with distilled water. Finally, the treated LeLeFs were air dried for 24 h.

Physical Properties

Determination of Fibre Yield. Ten samples of lemba leaves were taken randomly. The weight of the raw leaves (W_L) and the extracted fibres (W_F) were measured. The yield of fibre extracted from the leaves was calculated by using the following Equation 1:

$$\text{Yield} = \frac{W_F}{W_L} \times 100\% \quad (1)$$

Determination of Fibre Density. The density determination of LeLeF was carried out according to the Archimedes principle. The fibre was placed in a beaker containing 50 ml of water. The fibre was weighed before putting into the beaker. The total mass of water and fibre was recorded then the mass of water displaced by fibre was recorded. The weight to volume ratio yielded the density of fibre. The density of LeLeF before and after alkaline treatment was calculated by dividing the mass of LeLeF with the volume of water displaced.

Morphological Observation. The morphology of pristine and alkaline-treated LeLeF was examined using a scanning electron microscope (SEM) model JEOL JSM-6390LV.

All the surfaces of the solid fibres were sputtered with gold before the testing to avoid any electron charging on the image and poor image resolution.

Thermal Properties

Thermogravimetric Analysis. Thermogravimetric Analysis (TGA) of LeLeFs was conducted using a thermal gravimetric analyzer (TGA-Model: TGA7 Perkin Elmer Pyris). It was used to measure changes in the weight loss (mass) of the sample and determine the degradation of LeLeF. The fibres were chopped into micro size (about 1 mm) for the analysis. The samples were heated up from 25 to 500°C at a rate of 10°C/min in a nitrogen gas (60 ml/min). Each fibre was analyzed separately and overlapped for comparison.

Mechanical Properties

Evaluation of Tensile Properties. The tensile properties of LeLeF were performed using a Texture Analyzer according to ASTM D3822. The gauge length was kept constant at 50 mm and test speed 0.50 mm/s. Fibres length was maintained 50 mm for all samples. In each case, five specimens were tested and the average value was tabulated. Tensile strength (TS), Young's modulus (YM) and elongation at break (EB) were determined from the graphs plotted. The fundamental relationships applied to determine these properties are as shown in Equation 2, 3 and 4:

$$TS = \frac{F}{A_F} \tag{2}$$

$$EB = \frac{L_f - L_i}{L_i} \times 100\% \tag{3}$$

$$YM = \frac{S_S}{S_T} \tag{4}$$

Where F is force, A_F is average fibre area, L_i is initial length, L_F is final length, S_s is stress and S_T is strain

Chemical Properties

Evaluation of Chemical Resistance. Chemical resistance test was performed according to the method used by Gupta and Kumar (2012). The effect of some solvents such as toluene, the effect of some acids i.e.nitric acid, hydrochloric acid and the effect of some alkalies such as NaOH, Na_2CO_3, and NH_4OH were studied. For each case, five pre-weighed samples were dipped for 24 h in the respective chemical reagents. Then, the samples were

removed and washed with distilled water. They were dried at room temperature by using filter paper. The samples were then weighed and the percentage weight loss/gain (WL) was determined using the following Equation 5:

$$WL = \frac{W_f - W_i}{W_i} \times 100\% \quad (5)$$

where W_i is the initial weight of the fibre and W_f is the weight of the fibre after dipped in chemical reagents.

Functional Group Analysis. Fourier Transform Infrared (FTIR) spectra were acquired using IR Tracer-100 FTIR Spectrophotometer SHIMADZU. The LeLeFs before and after treatment were chopped into micro size (about 1 mm) and placed on the ATR crystal surface. The ATR-FTIR spectra were obtained with an accumulation of 40 scans and with a wavelength from 4000 to 800 cm^{-1}.

RESULTS AND DISCUSSION

Physical Properties

Determination of Fibre Yield. Initially, the fibre content in LeLeF based on fibre yield was analysed by using the hand-scraping method. Based on the data collection, it was found that the fibre yield was in a range of 6.73 to 13.21% with an average of 9.59%. The previous study on pineapple leaves stated that conventional methods like retting and scraping provide fibre yield of 1.8 % and 1.4% respectively (Kengkhetkit & Amornsakchai, 2012). As compared to the conventional method, the fibre content in the lemba leaves was higher than the pineapple leaves.

The appearance of the obtained LeLeF before and after alkaline treatment was then observed as shown in Figure 2. As seen in Figure 2, the pristine LeLeFs were naturally dark brownish colour. After alkaline treatment, the colour of LeLeFs appeared to be relatively lighter compared to the pristine one. This difference can be attributed to an increase in removal of lignin and hemicellulose content with the increasing of NaOH concentration. The result is in agreement with the FTIR result where the peaks at which associated to the lignin and hemicellulose decreased as the NaOH concentration increased. Hence, this shows that the changes in the colour of the fibres in Figure 2 were due to lignin and hemicellulose removal. Besides, previous study by Wunna et al. (2017) on the pre-treatment of sugarcane bagasse also proved that the amount of lignin removal increased as the concentration of NaOH increased. Fareez et al. (2018) reported that the colour of pineapple leaf fibre changed from brown to white after undergoing alkali treatment and bleaching.

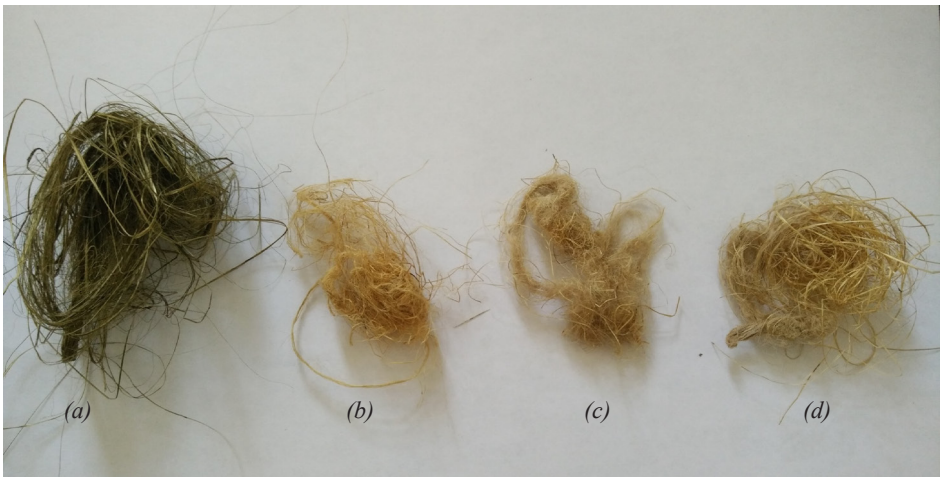


Figure 2. Pictures of a) The pristine LeLeF and LeLeF treated with b) 6 wt% NaOH, c) 8 wt% NaOH and d) 10 wt% NaOH

Determination of Fibre Density. Table 1 shows the densities of pristine LeLef and alkaline treated LeLef. From Table 1, it is clearly evident that the density values of LeLeF were increased gradually with an increase of the NaOH concentration. This is attributed to the densification of the cell wall as a result of removing non-cellulosic parts of fibre which were lignin, hemicellulose, pectin and others after increasing the concentration of sodium hydroxide (Sawpan et al., 2011). The density values mainly depended on the environment in which the plant was grown, extraction procedure of fibres, soil condition, and age of the plant (Hulle et al., 2015). The extraction process with higher NaOH concentration increased the density of the fibre as more of the amorphous part of the plant which contributed to low density was removed. Non-treated fibre consists of cellulosic and non-cellulosic part and the density of the fibre was the average density of all the components. Meanwhile, fibre treated with higher NaOH concentration resulted in higher purity of cellulose. High purity of cellulose indicates high crystallinity where the crystalline structure was packed in an orderly manner thus increase the density of the fibre. Our finding shows that the density of LeLeF was in the range of other natural fibres obtained from leaves as shown in Table 2.

Table 1

Density Properties of LeLeF

Condition of LeLeF	Pristine	6 wt% NaOH	8 wt% NaOH	10 wt% NaOH
Density (g/cm^3)	0.809±0.006	0.899±0.008	1.021±0.011	1.168±0.020

Table 2

Density of natural fibres

Natural Fibre	Density (g/cm^3)	Reference
Sisal	1.3	(Li et al., 2007)
Curaua	1.1	(Spinacé et al., 2009)
Pineapple Leaf	1.07	(Zin et al., 2018)

Morphological Observation. The morphologies of pristine, 6, 8, and 10 wt% treated LeLeFs were studied by using SEM (Figure 3). Analogous to other lignocellulosic reinforcing plants, LeLeF is a multicellular composite fibre. Each unit cell of fibre composed of cellulose microfibrils with different fibrillary orientations surrounded and cemented together with lignin and hemicellulose. Due to alkali treatment, NaOH disrupted the hydrogen bonding in the network structure of the fibre thus removing a certain amount of lignin, wax and oils, covering the external surface of the fibre cell wall and also depolymerized the cellulose (Li et al., 2007).

The effect of different concentrations of alkaline treatment on the LeLeFs was further investigated via SEM analysis at 300x magnification, as seen in Figure 3 (a-d). As expected, the morphological images of surfaces and cross-sectional of pristine LeLeF and treated

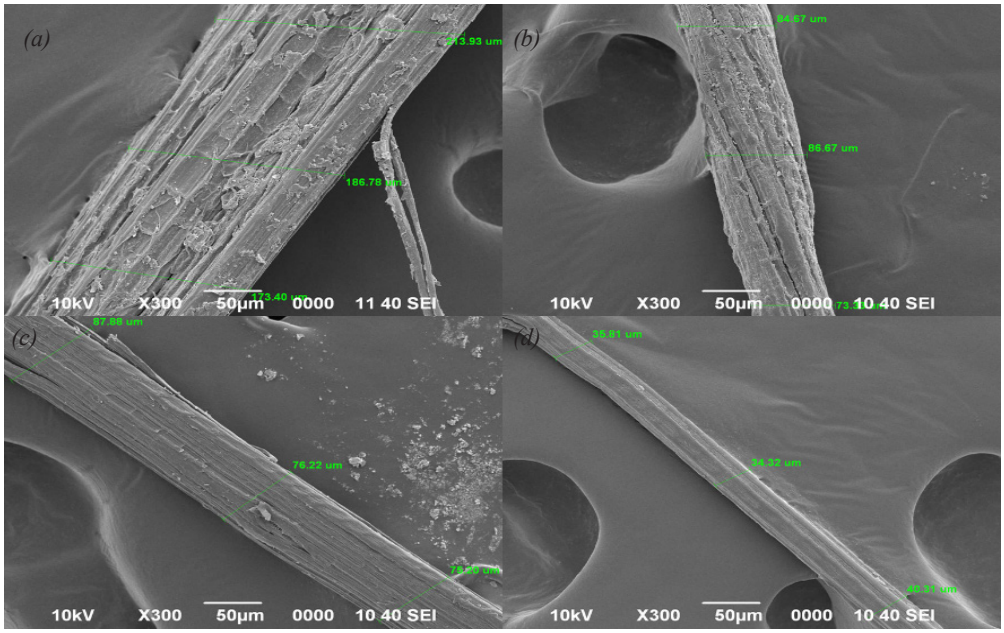


Figure 3. Scanning electron micrographs surface of a) pristine LeLeF and LeLeF treated with b) 6 wt% NaOH, c) 8 wt% NaOH and d) 10 wt% NaOH at magnification x300

LeLeF demonstrated significant different images in the terms of their level of smoothness, fibre binding and diameter (Li et al., 2007). As a benchmark, the pristine LeLeF fibres appeared to have many uplifted filaments that were randomly embedded and stacked together on the surface, as indicated in Figure 3a. Owing to the weak binding of lignin and hemicellulose (Ramadevi et al., 2012), they can be easily separated from the fibre surface by an external force (Mwaikambo et al., 2007).

After treatment, these filaments disappeared and the fibre surface became cleaner as well as a reduction in the fibre diameter from 191.37 μm to 36.81 μm. This suggests that alkaline treatment caused the removal of the surface impurities such as wax, hemicellulose, and other inorganic impurities, as seen from Figure 4c-d (George et al., 1997; Zannen et al., 2014).

Thermal Properties

Thermogravimetric Analysis. TGA was carried out to study the weight loss occurring on fibres at different temperatures (Figure 4). In general, the weight loss steps of natural fibres consists of three phases which are; i) moisture evaporation, ii) decomposition of components of hemicelluloses with lignin and iii) decomposition of cellulose. From the curve in Figure 4, the first stage of weight loss started from 37-100°C which is 12% for pristine and 5%, 4% and 3% for 6, 8 and 10 wt% NaOH respectively. This low-temperature weight loss occured due to the loss of moisture or water vaporization. The weight loss decreased for higher NaOH concentration treated LeLeF. This is in agreement with the FTIR result where the NaOH concentration increased, the broadness band of the hydroxyl group decreased which showed decreasing of OH group that can absorb water. This resulted in the decrease of the water absorption of LeLeF which means that the moisture was lost through alkali treatment thus decreasing the weight. Some other lignocellulosic fibres showed similar results for this low-temperature loss of weight like piassava fibre (5.18%) (D'Almeida et al., 2006), jute fibre (10.52%) (Das et al., 2000), flax (6.3%) and wheat straw (7.3%) (Hornsby et al., 1997).

Weight loss temperature ranged at about 250-420°C, indicating the degradation of cellulose. The cellulose degradation temperature of other lignocellulosic fibres like sugarcane was 350°C (Hoi & Martincigh, 2013), amazon piassava (361°C) (Rebelo et al., 2019), pineapple's crown, rice husks and cotton (257-390°C) (Prado & Spinacé, 2015). From this figure, it can be observed that the pristine LeLeF started to decompose at 286 °C while for 6, 8 and 10 wt% NaOH treated LeLeF, they started to decompose at 287°C, 291°C, and 293°C respectively. The treated fibres showed higher initial temperature of thermal decomposition which related to the partial removal of lignin and hemicellulose (Santos et al., 2013). The decomposition temperature increased as the concentration for the NaOH treatment increased. The thermal stability of fibres increased with increasing of

NaOH concentration thus requires higher temperature to decompose. Due to the complex structure of lignin, the fibre decomposition occured slowly within the whole temperature range (Rosa et al., 2010).

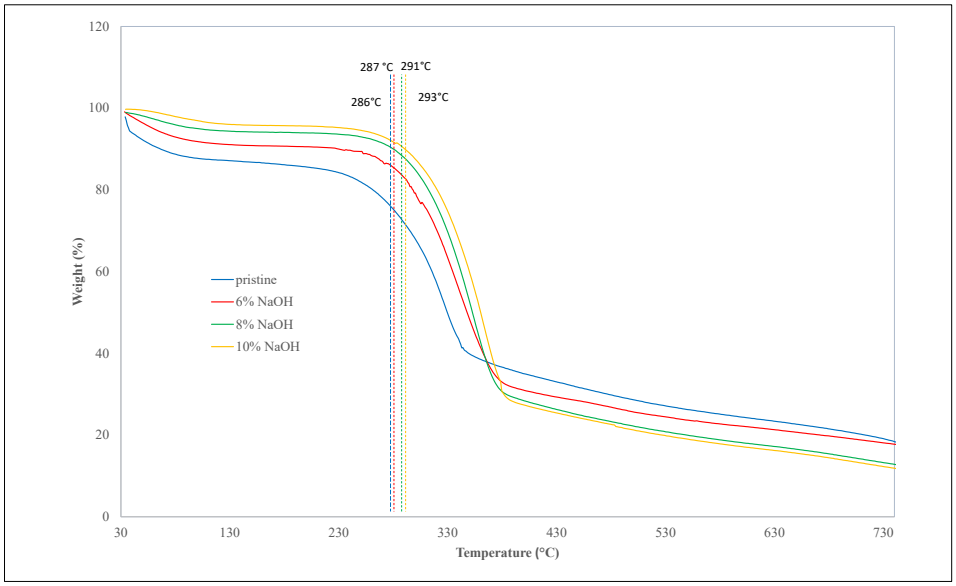


Figure 4. TGA analysis for pristine and treated LeLef at different concentrations

Mechanical Properties

Evaluation of Tensile Properties. The result in Table 3 shows that the maximum tensile strength of the LeLeF was achieved after 10 wt% NaOH treatment at 429.5 MPa. The tensile strength of LeLeF was significantly increased from pristine LeLeF, 6 wt% NaOH treated and 8 wt% NaOH treated which were recorded at about 14.5 MPa, 80.8 MPa and 121.7 MPa, respectively. This result proves that the treated fibres show higher tensile strength compared to pristine fibre. Changes of cellulose crystallinity during alkaline treatment causes increased in the tensile strength of LeLeF from pristine to 10 wt% NaOH treated. This result is in agreement with the density data where the remaining structure left was high crystallinity cellulose. The crystalline structure of cellulose was packed in an orderly manner and more compact thus increasing the density resulting in an increment of tensile strength as the brittle amorphous structure had been removed during the alkaline treatment. Ridzuan et al. (2015) also reported that the removal of the weak amorphous component and retaining only the crystalline component enhanced the strength of the fibre.

Table 3

Tensile properties of pristine LeLeF and different concentration of NaOH treated LeLeF

Fibre	Tensile Properties		
	Maximum stress (MPa)	*Young's modulus (GPa)*	*Elongation at break (%)*
Pristine	14.45	0.83	2.06
6 wt% NaOH	80.79	4.16	2.44
8 wt% NaOH	130.12	4.19	2.78
10 wt% NaOH	511.10	11.76	3.69

The result for the elongation at break in Table 3 shows a similar increasing trend as the tensile strength. However, the increment was insignificant as it was only about 2.4 to 3.7% for pristine LeLeF to 10 wt% NaOH treated LeLeF. As for the modulus of elasticity, the result in Table 3 shows that the highest modulus of elasticity was achieved with 10 wt% NaOH treatment at 11.8 GPa. Modulus of elasticity increases from pristine LeLeF and 6 wt% NaOH treated LeLeF at 0.83 GPa and 4.16 GPa, respectively. The modulus of elasticity for 8 wt% NaOH treated LeLeF was found to be the same as the modulus of elasticity for 6 wt% NaOH treated LeLeF. The results obtained for the tensile strength, elongation at break and modulus of elasticity exhibited a similar trend with research done on napier grass fibres which showed an increment from pristine napier grass fibre to 5% NaOH treated napier grass fibre for tensile strength, elongation at break and modulus of elasticity (Reddy et al., 2012). The dispersion of hemicellulose and lignin in the interfibrillar region of pristine fibres separated the cellulose chain from one another thus making it always in a state of constraint. After alkali treatment, the removal of hemicellulose eased the fibrils to rearrange themselves in a compact manner, resulting a closer packing of the cellulose chains and thus contributing to improve the strength and tensile properties of the fibres (Reddy et al., 2012).

Chemical Properties

Evaluation of Chemical Resistance. The resistance of the fibres to water and certain chemicals was studied (Figure 5). The acids (HCl and CH_3COOH), alkaline (NaOH) were used as a polar inorganic solvent while toluene was used as a non-polar organic solvent. The weight gain for all pristine fibres, 6, 8 and 10 wt% NaOH treated fibres with different chemicals are shown in Figure 5. From the figure, it is seen that the weight gained was observed in all cases. Therefore, it did not seem as if any erosion of fibres occurred as the fibres did not lose weight (Jayaramudu et al., 2015). The percentage of weight gain for the pristine fibres in all chemicals was higher compared to treated fibres. According to Reddy

et al. (2018), the weight gain of the fibre corresponded to a better interaction with the fluid which indicated poor chemical resistance while less weight gain corresponded to the less absorption of solvent thus indicated better chemical resistance. This shows that fibres treated at highest NaOH concentration had the highest chemical resistance. Higher NaOH removed more of the non-cellulosic part of fibres thus increased the crystallinity structure of the cellulose. The crystalline structure of cellulose was built up of denser packing structure thus possesing higher resistance to chemical as indicated by the percentage of weight gained.

The percentage weight gain of all the polar solvents were almost the same while for the non-polar solvent the percentage weight gain was low compare to polar solvent. This indicates that cellulose was more susceptible to polar solvents compared to non-polar solvents because of OH interaction.

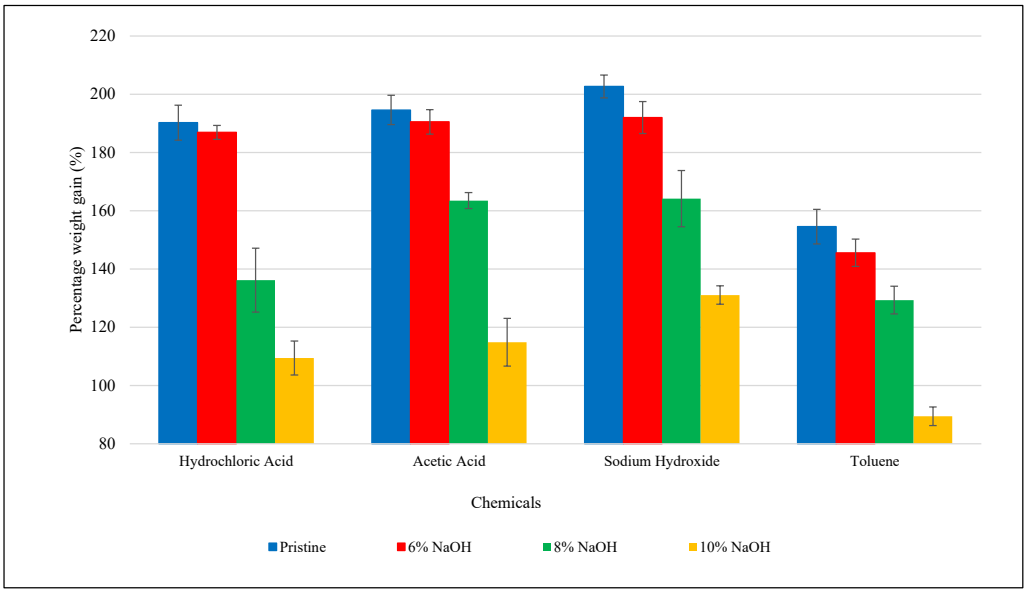


Figure 5. Chemical resistance properties of pristine, 6, 8, and 10 wt% alkaline treated LeLeF

Functional Group Analysis. The FTIR analysis was employed to investigate remarkable differences in the functional groups of pristine LeLeF and LeLeF after the NaOH treatments (Figure 6). Figure 6 shows that the spectra obtained have a similar shape to those of other types of plant fibres (Guimarães et al., 2009; Hoi & Martincigh, 2013; Santos et al., 2013). The appearance of a broad absorption band was observed in the region of 3117–3480 cm^{-1}, with a peak at 3334 cm^{-1} which corresponded to the hydroxyl groups of cellulose (Garside & Wyeth, 2006; Neto et al., 2013). The peaksat 2918 cm^{-1} and 2848 cm^{-1} were attributed

to the asymmetric stretching of CH and CH_2 in cellulose and hemicellulose (Alvarez & Vázquez, 2006; Fan et al., 2012; Rosa et al., 2010). Meanwhile, the peak at 1713 cm^{-1} was associated with the C=O stretching vibration linkage of the ester group in hemicellulose (Guimarães et al., 2009). The peak at 1578 cm^{-1} was corresponded to the C=C stretching of the aromatic ring, characteristic of the lignin (Guimarães et al., 2009) while the peak at 1240 cm^{-1} which only present in the spectra of pristine LeLeF was represented C-O stretch of acetyl group of lignin (Sgriccia et al., 2008).

In comparison to the pristine LeLeF, a reduction in hydroxyl (-OH) stretching intensity could be observed as LeLeF was treated with a higher concentration of NaOH. This is probably due to the free hydroxyl group participated in the chemical reaction (Samal & Ray, 1997). The intensity of peak at 2918 cm^{-1}, 2848 cm^{-1} and 1713 cm^{-1} decreased as LeLeF was treated with a higher concentration of alkali solution. This clearly indicates that the amount of hemicellulose on the LeLeF was successfully reduced by the treatment with NaOH. In placing more emphasis, Sgriccia et al. (2008) reported that the peak 1730 cm^{-1} of hemp fibre that attributed to the C=O stretching of the acetyl group of hemicellulose was not present in the alkali treated samples as the removal of the hemicellulose caused this peak to disappear. After the alkaline treatment, the reducing peak occurred at 1578 $cm^{-1}$and no distinct peak could be observed at the peak 1240 cm^{-1}, indicating that the removal of lignin was achieved.

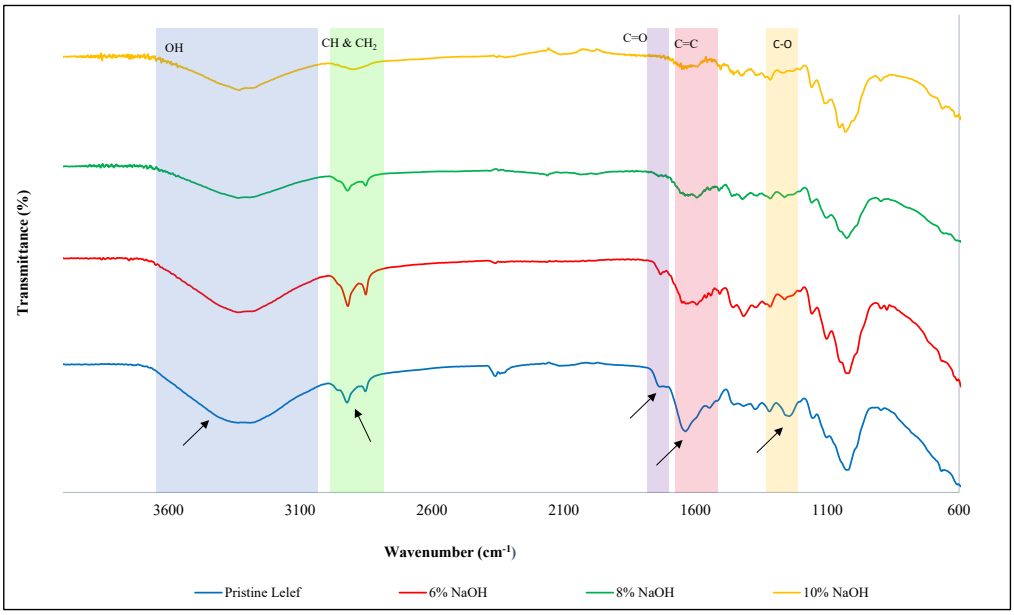


Figure 6. FTIR spectra of pristine and treated LeLeF with different concentration of NaOH

CONCLUSIONS

FTIR analysis showed that the higher the NaOH concentration, the more lignin and hemicellulose removed from the fibre as the intensity of peak associated with hemicellulose and lignin were reduced upon increase of NaOH. The density of LeLeF increased with the increment of the NaOH concentration. Higher NaOH concentration removes more surface impurities such as wax, hemicellulose and lignin resulted in the decreasing of fibre diameter from 191.37 μm to 36.81 μm. This study shows that as the NaOH concentration increased the tensile properties of LeLeF also increased up to 3400% (from 14.45 to 511.10 MPa). The highest tensile strength, elongation and modulus were achieved at 10 wt% NaOH treated LeLeF. TGA results showed that the decomposition temperature of fibre increased from 286 °C to 293 °C as the fibre was treated with higher NaOH concentration. Thus, alkaline treatment enhanced the thermal stability of the fibre. Chemical resistance analysis showed that alkaline treated LeLeF had better chemical resistance compared to pristine LeLeF where the percentage weight gain of 10 wt% NaOH treated for all polar solvent were less than 130% while the percentage weight gain of pristine LeLeF for all polar solvents were more than 180%.

ACKNOWLEDGEMENT

This study was supported by Collaborative Research Grant (CRG Vote No 08G10), Fundamental Research Grant (UTMFR Vote No 20H81) and (UTMFR Vote No 20H84) funded by the Universiti Teknologi Malaysia.

REFERENCES

Alvarez, V. A., & Vázquez, A. (2006). Influence of fiber chemical modification procedure on the mechanical properties and water absorption of MaterBi-Y/sisal fiber composites. *Composites Part A: Applied Science and Manufacturing*, *37*(10), 1672-1680.

Asim, M., Jawaid, M., Abdan, K., & Ishak, M. R. (2016). Effect of alkali and silane treatments on mechanical and fibre-matrix bond strength of kenaf and pineapple leaf fibres. *Journal of Bionic Engineering*, *13*(3), 426-435.

Brink, M., & Escobin, R. P. (2003). *Plant resources of south-east Asia No 17: Fibre plants*. Leiden, Netherlands: Backhuys.

D'Almeida, J., Aquino, R., & Monteiro, S. (2006). Tensile mechanical properties, morphological aspects and chemical characterization of piassava (*Attalea funifera*) fibers. *Composites Part A: Applied Science and Manufacturing*, *37*(9), 1473-1479.

Das, S., Saha, A. K., Choudhury, P. K., Basak, R. K., Mitra, B. C., Todd, T., … & Rowell, R. M. (2000). Effect of steam pretreatment of jute fiber on dimensional stability of jute composite. *Journal of Applied Polymer Science*, *76*(11), 1652-1661.

Fan, M., Dai, D., & Huang, B. (2012). Fourier transform infrared spectroscopy for natural fibres. *Fourier Transform - Materials Analysis, 3*, 45-68.

Fareez, I. M., Ibrahim, N. A., Yaacob, W. M. H. W., Razali, N. A. M., Jasni, A. H., & Aziz, F. A. (2018). Characteristics of cellulose extracted from Josapine pineapple leaf fibre after alkali treatment followed by extensive bleaching. *Cellulose, 25*(8), 4407-4421.

Farzinebrahimi, R., Taha, R. M., Rashid, K. A., Ahmed, B. A., Danaee, M., & Rozali, S. E. (2016). Preliminary screening of antioxidant and antibacterial activities and establishment of an efficient callus induction in *Curculigo latifolia Dryand* (Lemba). *Evidence-Based Complementary and Alternative Medicine*, *2016*, 1-9.

Garside, P., & Wyeth, P. (2006). Identification of cellulosic fibres by FTIR spectroscopy differentiation of flax and hemp by polarized ATR FTIR. *Studies in Conservation*, *51*(3), 205-211.

George, J., Bhagawan, S., & Thomas, S. (1997). Improved interactions in chemically modified pineapple leaf fiber reinforced polyethylene composites. *Composite Interfaces*, *5*(3), 201-223.

Guimarães, J., Frollini, E., Silva, C. D., Wypych, F., & Satyanarayana, K. (2009). Characterization of banana, sugarcane bagasse and sponge gourd fibers of Brazil. *Industrial Crops and Products*, *30*(3), 407-415.

Gupta, A., & Kumar, A. (2012). Chemical Properties of Natural Fiber Composites and Mechanisms of Chemical Modifications. *Asian Journal of Chemistry, 24*(4), 1831-1836.

Hashim, M. Y., Amin, A. M., Marwah, O. M. F., Othman, M. H., Yunus, M. R. M., & Huat, N. C. (2017). The effect of alkali treatment under various conditions on physical properties of kenaf fiber. *Journal of Physics: Conference Series*, *914*, 1-15.

Hoi, L. W. S., & Martincigh, B. S. (2013). Sugar cane plant fibres: Separation and characterisation. *Industrial Crops and Products*, *47*, 1-12.

Hornsby, P. R., Hinrichsen, E., & Tarverdi, K. (1997). Preparation and properties of polypropylene composites reinforced with wheat and flax straw fibres: Part II Analysis of composite microstructure and mechanical properties. *Journal of Materials Science, 32*(2), 443-449.

Hulle, A., Kadole, P., & Katkar, P. (2015). Agave Americana leaf fibers. *Fibers*, *3*(4), 64-75.

Jayaramudu, J., Agwuncha, S. C., Ray, S. S., Sadiku, E. R., & Rajulu, A. V. (2015). Studies on the chemical resistance and mechanical properties of natural polyalthia cerasoides woven fabric/glass hybridized epoxy composites. *Advanced Materials Letters*, *6*(2), 114-119.

Kengkhetkit, N., & Amornsakchai, T. (2012). Utilisation of pineapple leaf waste for plastic reinforcement: 1. A novel extraction method for short pineapple leaf fiber. *Industrial Crops and Products*, *40*, 55-61.

Li, X., Tabil, L. G., & Panigrahi, S. (2007). chemical treatments of natural fiber for use in natural fiber-reinforced composites: A review. *Journal of Polymers and the Environment*, *15*(1), 25-33.

Motaleb, K. Z. M. A. (2018). Improvement of mechanical properties by alkali treatment on pineapple and jute fabric reinforced polyester resin composites. *International Journal of Composite Materials*, *8*(2), 32-37.

Mwaikambo, L. Y., Tucker, N., & Clark, A. J. (2007). Mechanical properties of hemp-fibre-reinforced euphorbia composites. *Macromolecular Materials and Engineering*, *292*(9), 993-1000.

Neto, A. R. S., Araujo, M. A., Souza, F. V., Mattoso, L. H., & Marconcini, J. M. (2013). Characterization and comparative evaluation of thermal, structural, chemical, mechanical and morphological properties of six pineapple leaf fiber varieties for use in composites. *Industrial Crops and Products*, *43*, 529-537.

Prado, K. D. S. D., & Spinacé, M. A. D. S. (2015). Characterization of fibers from pineapples crown, rice husks and cotton textile residues. *Materials Research*, *18*(3), 530-537.

Ramadevi, P., Sampathkumar, D., Srinivasa, C. V., & Bennehalli, B. (2012). Effect of alkali treatment on water absorption of single cellulosic abaca fiber. *BioResources*, *7*(3), 3515-3524.

Ranjbarfard, A., Saleh, G., Abdullah, N. A. P., & Kashiani, P. (2014). Genetic diversity of lemba (*Curculigo latifolia*) populations in peninsular malaysia using ISSR molecular markers. *Australian Journal of Crop Science, 8*(1), 9-17.

Rebelo, V., Silva, Y. D., Ferreira, S., Filho, R. T., & Giacon, V. (2019). Effects of mercerization in the chemical and morphological properties of amazon piassava. *Polímeros*, *29*(1), 1-6.

Reddy, B. M., Reddy, Y. V. M., & Reddy, B. C. M. (2018). Effect of alkali treatment on mechanical, water absorption and chemical resistance properties of cordia-dichotoma fiber reinforced epoxy composites. *International Journal of Applied Engineering Research*, *13*(6), 3709-3715.

Reddy, K. O., Maheswari, C. U., Shukla, M., & Rajulu, A. (2012). Chemical composition and structural characterization of Napier grass fibers. *Materials Letters*, *67*(1), 35-38.

Ridzuan, M. J. M., Majid, M. A., Afendi, M., Kanafiah, S. A., & Nuriman, M. (2015). Effects of alkaline concentrations on the tensile properties of napier grass fibre. *Applied Mechanics and Materials*, *786*, 23-27.

Rosa, I. M. D., Kenny, J. M., Puglia, D., Santulli, C., & Sarasini, F. (2010). Morphological, thermal and mechanical characterization of okra (Abelmoschus esculentus) fibres as potential reinforcement in polymer composites. *Composites Science and Technology*, *70*(1), 116-122.

Samal, R. K., & Ray, M. C. (1997). Effect of chemical modifications on FTIR spectra. II. Physicochemical behavior of pineapple leaf fiber (PALF). *Journal of Applied Polymer Science*, *64*(11), 2119-2125.

Santos, R. M. D., Neto, W. P. F., Silvério, H. A., Martins, D. F., Dantas, N. O., & Pasquini, D. (2013). Cellulose nanocrystals from pineapple leaf, a new approach for the reuse of this agro-waste. *Industrial Crops and Products*, *50*, 707-714.

Sawpan, M. A., Pickering, K. L., & Fernyhough, A. (2011). Effect of various chemical treatments on the fibre structure and tensile properties of industrial hemp fibres. *Composites Part A: Applied Science and Manufacturing*, *42*(8), 888-895.

Sgriccia, N., Hawley, M., & Misra, M. (2008). Characterization of natural fiber surfaces and natural fiber composites. *Composites Part A: Applied Science and Manufacturing*, *39*(10), 1632-1637.

Shaari, N. (2005, December 12-14). Lemba (*Curculigo latifolia*) leaf as a new materials for textiles. *N 2005 4th International Symposium on Environmentally Conscious Design and Inverse Manufacturing* (pp. 109-111). Tokyo, Japan.

Spinacé, M. A., Lambert, C. S., Fermoselli, K. K., & Paoli, M. A. D. (2009). Characterization of lignocellulosic curaua fibres. *Carbohydrate Polymers*, *77*(1), 47-53.

Wiart, C., & Wong, F. K. (2002). *Medicinal plants of Southeast Asia*. Kuala Lumpur, Malaysia: Pelanduk.

Wunna, K., Nakasaki, K., Auresenia, J. L., Abella, L. C., & Gaspillo, P. A. D. (2017). Effect of alkali pretreatment on removal of lignin from sugarcane bagasse. *Chemical Engineering Transactions*, *56*, 1831-1836.

Zannen, S., Ghali, L., Halimi, M. T., & Hssen, M. B. (2014). Effect of chemical extraction on physicochemical and mechanical properties of doum palm fibres. *Advances in Materials Physics and Chemistry*, *04*(10), 203-216.

Zin, M. H., Abdan, K., Mazlan, N., Zainudin, E. S., & Liew, K. E. (2018). The effects of alkali treatment on the mechanical and chemical properties of pineapple leaf fibres (PALF) and adhesion to epoxy resin. *IOP Conference Series: Materials Science and Engineering*, *368*, 1-10.

Pertanika J. Sci. & Technol. 28 (4): 1549 - 1558 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

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

Pattern of Histopathological Alterations in N-methyl-N-Nitrosourea (MNU) Induced Breast Cancer in Female Sprague Dawley (SD) Rats Treated with Crude Honey

Urmila Banik[1,2]*, Sarfarz Ahamed[3], Swe Swe Latt[4], Nur Asyilla Che Jalil[1], Wan Faiziah Wan Abdul Rahman[1] and Nor Hayati Othman[1]

[1]Department of Pathology, School of Medical Sciences, Universiti Sains Malaysia, 16150 USM, Kubang Kerian, Kelantan, Malaysia
[2]Unit of Pathology, Faculty of Medicine, AIMST University, Semeling, 08100 Bedong, Kedah, Malaysia
[3]Department of Biochemistry, Bahauddin Zakariya University, Multan, Pakistan
[4]Unit of Community Medicine, AIMST University, Faculty of Medicine, Semeling, 08100 Bedong, Kedah, Malaysia

ABSTRACT

Natural products are becoming primary investigative molecules creating hope for finding new powerful classes of anticancer agents for breast cancer. One of the most remarkable of these is honey. To explore the mechanism of action of any anticancer agent the initial step is to analyse its effect in the histopathological tissue section. This study was designed to describe the histopathology of N-methyl-N-nitrosourea (MNU) induced breast cancer in Sprague Dawley rat (SD rats) treated with crude honey. Female rats were distributed into 4 groups: Group0 (normal), Group1 (MNU control), Group2 and 3: Tualang and Manuka honey-treated, respectively. Rats were sacrificed and histopathology of both non-treated and treated tumours was done. Lower histological grade, infrequent combination histologic pattern ($p<0.001$), prominent cytoplasmic vacuolization ($p<0.001$), aggregates of atypical macrophages ($p< 0.001$) and lesser necrosis ($p= 0.005$) were

ARTICLE INFO

Article history:
Received: 28 March 2020
Accepted: 15 June 2020
Published: 21 October 2020

DOI: https://doi.org/10.47836/pjst.28.4.22

E-mail addresses:
urmila_banik@yahoo.co.in; ub15_med074@student.usm.my (Urmila Banik)
sarfraz_114@yahoo.com (Sarfarz Ahamed)
sweswelattdr@gmail.com (Swe Swe Latt)
asyilla@usm.my (Nur Asyilla Che Jalil)
wfaiziahwar@gmail.com (Wan Faiziah Wan Abdul Rahman)
hayatikb@usm.my (Nor Hayati Othman)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

major outcomes in treated cancers. This data will aid in the histopathological interpretation of honey-treated breast cancer model and future anticancer study of honey.

Keywords: Breast cancer, histopathology, honey, MNU, Sprague Dawley rats

INTRODUCTION

Breast cancer is the most common cancer among women, affecting 2.1 million women every year and accounting for the highest number of cancer-related deaths among them (WHO, n.d.). According to traditional classification, primary breast adenocarcinoma is categorized as in situ and invasive. In situ cancers can be ductal carcinoma in situ (DCIS) or lobular (LCIS) and have a great prognosis. While 50–80% of the invasive cancers are invasive ductal carcinoma (IDC), only 25% of invasive breast cancers are 'special type' (invasive lobular carcinoma, invasive cribriform carcinoma and tubular carcinoma) (Masood, 2016). Histologic grading of invasive cancers is highly correlated with disease-free and overall survival. Posttreatment histological alterations that are evident in breast cancer tissues are quite diverse. Histo-morphologically the Russo and Russo classification scheme for the N-methyl-N-nitrosourea (MNU) induced breast cancers in Sprague Dawley (SD) rats are quite alike to the traditional classification of human breast cancer (Russo & Russo, 2000). This breast cancer model is simple to develop, cost-effective and mimics human breast cancer in terms of tumour histopathology, origination from mammary ductal epithelial cells and dependence on ovarian hormones for tumourigenesis. Hence it is a commonly used cancer model to study the anticancer mechanism of natural products and other potential anticancer agents (Han et al., 2019; Lu et al., 2013). However, cancers in rats may be composed of single histologic type or combinations of several patterns (Russo & Russo, 2000).

Nowadays natural products are becoming the primary investigative molecules creating hope for finding new powerful classes of anticancer agents for breast cancer (Banik et al., 2017). One of the most remarkable natural compounds with substantial anticancer properties is honey. Polyphenols are the primary anticancer components of crude honey (Spilioti et al., 2014). Anticancer effect of honey is attributable to its anti-proliferative and pro-apoptotic activities (Jaganathan et al., 2015). Depending upon concentration crude honey can demonstrate either oestrogenic or antioestrogenic mode of action (Jaganathan & Mandal, 2009). Although anticancer studies of honey on breast cancer primarily focuses on the molecular mechanistic effect of honey (Jaganathan et al., 2015), its detail mechanism of action is still indistinct. Generally, to explore the mechanism of action of any anticancer agent the initial step is to analyse its effect in the histopathological tissue section. Based on the histopathological observation subsequent suitable molecular methods are applied to understand the mechanism of action in detail. Precise correlation of post-treatment

histopathological changes with underlying pathological processes helps to interpret the results accurately. Study shows that crude form of Tualang Honey (TH) and Manuka honey (MH) have a significant anticancer effect on breast cancer SD rat model (Ahmed et al., 2017). Our present study has been designed for the first time to do a detail analysis and comparison of the pattern of histological alterations in honey-treated MNU induced breast cancer in female SD rats with that of the non-treated ones.

MATERIALS AND METHODS

Animals, MNU and Honey

SD female rats (28 and 33 days old) were obtained from the Animal Research and Service Centre (ARASC), Universiti Sains Malaysia (USM). TH was supplied by the Federal Agricultural Marketing Authority (FAMA), Ministry of Agriculture and Agro based Industry, Malaysia. MH was purchased from the market (packed under license number 1003 for Vitaco Health (NZ) Ltd., New Zealand and imported and distributed by Cambert (M) Sdn.Bhd, Malaysia). MNU (Catalog number N1517-1G, Sigma, USA) and honey samples were prepared as described previously (Ahmed et al., 2017).

Study Design

Forty female SD rats were distributed into 4 groups (10 rats/group): Group0 (healthy normal control), Group1 (positive control, non-treated rats), Groups2 and 3: 1g/kg body weight/ day of TH and MH was given orally, respectively. MNU was injected intraperitoneally (80 mg/kg body weight) into the rats around the age of 40 days. The rats in groups 1, 2 and 3 were induced with cancer using carcinogen MNU. Mammary areas of the rats were palpated twice weekly to detect the appearance of masses. Oral honey treatment (using 1 ml syringes without needles administered to mouth) was started for groups 2 and 3 when the first palpable mass reached 10–12mm size and continued till day 120th. The rats were maintained on a standard balanced rat feed diet with water ad libitum and a 12 h day/ night cycle. On the day 120th, the rats were sacrificed following intraperitoneal injection of pentobarbital 100mg/kg body weight. Tumour masses were examined in vivo, resected and fixed in neutral-buffered formalin. Subsequently, masses were paraffin-embedded, sectioned to 4μ thickness and stained with hematoxylin and eosin.

Total 39 non-treated and 36 honey-treated (20TH, 16MH) tissue sections were examined under a light microscope (Olympus Optical Co., Ltd., Tokyo, Japan). From the non-treated slide sections, 3 were excluded due to scanty tissue or poor slide quality. Russo and Russo guideline (Russo & Russo, 2000) was followed for histological analysis. Grading was done by modified Bloom-Richardson system (Mukhopadhyay et al., 2006). Post-treatment alterations were analysed mainly by focusing on vacuolar degeneration,

microcyst formation, tumour necrosis, tumour infiltrating lymphocytes and macrophages, hyalinization and fibrosis.

Statistical Analysis

Collected data was entered to an Excel sheet and transferred to and analysed with IBM SPSS version 20.0. The result of the descriptive statistics was shown as frequency and percentages. The chi-square test was used to analyse the relationship between outcome variables and non-treated & treated groups with the p-value less than 0.05 as significant level.

RESULTS AND DISCUSSION

Total 103 breast cancer masses (positive control: 47, TH: 23 & MH: 33) were collected from experimental animals. Grossly, tumour masses in non-treated control group were larger and harder as described previously (Ahmed et al., 2017). Both benign and malignant tumours were observed. Cancer masses in honey treated groups showed a lower median tumour size, weight, multiplicity compared with the non-treated control ($p < 0.05$). Honey treatment also demonstrated a dramatic slower growth rate (up to 70.82%) compared with the non-treated control (0%) ($p < 0.05$). Routine histopathology was done. Total 72 histopathology slides (36 non-treated and 36 treated) were microscopically examined. All were malignant tumours of epithelial origin. *In situ* breast cancers were found in both non-treated control group (n=3) and honey-treated group (n=2). All were of single histologic pattern.

Among the non-treated invasive cancers (n=33), a combination of ≥2 patterns were frequently detected (57.6%) (Table 1). Histologic patterns observed in these invasive cancers were: papillary, cribriform, tubular and occasionally comedocarcinoma. However, no IDCs were seen. Papillary combined with cribriform was the most frequently (30%) observed combination pattern (Table 2). The histologic grading ranged from grade 2 (51.5%) to 3 (48.5%).

Table 1

Histological types of breast cancers seen in non-treated control and honey treated female SD rats

Types of breast cancer	Histologic Findings	Non-treated control		Honey-treated		p-value
		Number	%	Number	%	
In situ	Single histologic pattern	3	60.0	2	40.0	p=0.643
Invasive	Single histologic pattern	14	32.6	29	67.4	p<0.001
	Combination of ≥ 2 histologic patterns	19	79.2	5	20.8	p<0.001

Table 2

Distribution of different combinations of histologic patterns of invasive breast cancers seen in the non-treated control and honey treated cancers of female SD rats

Breast tissue of invasive breast cancer			
Histologic patterns observed in Non-treated control	Number	Histologic patterns observed in Honey-treated	Number
Papillary+Cribriform	10	Papillary+Cribriform	4
Papillary+Cribriform+Comedocarcinoma	1	Papillary+Cribriform+Comedocarcinoma	1
Cribriform	4	Cribriform	6
Cribriform+Tubular	3	-	-
Papillary+ Tubular	2	-	-
Tubular	5	Tubular	4
Papillary	5	Papillary	10
Papillary+Cribriform+Tubular	1	-	-
Tubular+Comedocarcinoma	1	-	-
Tubular+Comedocarcinoma	1	Tubular+Comedocarcinoma	1
-	-	IDC	8
-	-	Comedocarcinoma	1
Total	33		34

The degenerative features seen in the non-treated cancers (n=36) were microcyst formation (39%), hyalinization (75%) and fibrosis (72%) (Figure 1B-1D & 1G-1I; Table 3). While hyalinization was more frequent in the nontreated group, fibrosis was slightly common in the treated ones. However, microcyst formation was equivalent in both the groups. Vacuolar degeneration with tumour cell swelling was quite significant in the treated cancers in comparison to the non-treated ones ($p<0.001$). Eighty-nine percent of the treated cancers showed vacuolar degeneration. Swollen cancer cells coalesced to form cystic spaces followed by disintegration. In the non-treated controls, this vacuolar degeneration was very mild or absent (Figure 1E, 1J; Table 3).

In the non-treated cancers mild to moderate tumour infiltrating lymphocytes (TIL) with or without aggregates of foamy large a typical macrophages and eosinophil were frequently seen. Likewise, in treated cancers TILs were associated with tumour associated aggregates of foamy large atypical macrophages. Interestingly although in the treated group these a typical bizarre-looking macrophages with large hyperchromatic nucleus were seen in many of the cancers (71.4%), compared with the non-treated cancers, they were somewhat inconspicuous (Figure 1F, 1K; Table 3). Mild to massive tumour necrosis (65.8%) were

noted in the non-treated control cancers (Table 3). However, in the treated ones, necrosis was present in 34.2% of cases with massive necrosis seen in only 5.55% cases.

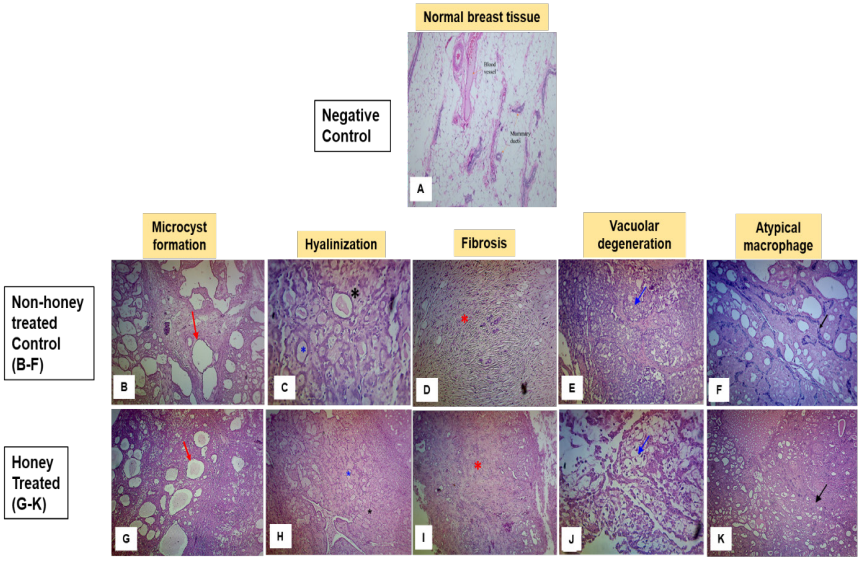


Figure 1. Histopathological findings of honey-treated and non-treated control breast cancer in female SD rats (200X magnification). (A): Histology of normal mammary gland of female Sprague-Dawley rat. Mammary ducts are surrounded by adipose and fibrous tissue with the varied distribution. H&E staining magnification x100. (B, G): Microcyst formation was equivalent in both nontreated control and honey treated group (Red arrow). (C, H): Hyaline material surrounding nests of malignant cells; extraluminal hyaline (Black asterisk): the intraluminal hyaline (Blue asterisk) is due to the secretory material. (D, I): Small nests of malignant cells surrounded by dense fibrous tissue (Red asterisk) deposition. (E, J): Cellular swelling (Blue arrow) was more prominent in the treated group. Absent or mild cytoplasmic vacuolization was seen in the nontreated control group. (F, K): Bizarre atypical macrophages in tumour parenchyma and tumour stroma of both honey-treated and nontreated control cancers (Black arrow). These cells are more prominent in the control group.

Table 3

Histomorphological alterations seen in non-treated control and honey treated breast cancers

Traits	Non-treated control		Honey treated		p-value
	Number	%	Number	%	
Microcyst	14	50.0	14	50.0	p>0.99
Hyalinization	27	60.0	18	40.0	p=0.028
Fibrosis	26	46.4	30	53.6	p=0.257
Vacuolar degeneration	2	5.9	32	94.1	p<0.001
Tumor necrosis	25	65.8	13	34.2	p=0.005
Foamy macrophages	12	28.6	30	71.4	p<.001

Post-treatment histopathological changes in tumour morphology play an important role in evaluating the therapeutic response. With increasingly accumulating data on distinct molecular-morphologic correlates, there is a resurgence of attention in the role of tissue evaluation in in vivo cancer study. In breast cancer, diverse histopathologic alterations are observed with the application of different therapeutic modalities. With radiation therapy, bizarre atypical nuclear changes and abnormal mitotic figures are usually seen. Yet again widespread tumour necrosis may develop and become bounded by fibrosis (Schnitt et al., 1984). Hormone therapy can cause connective tissue deposition with striking stromal fibrosis, hyalinization. Cytoplasmic vacuolization, cell membrane disruption, nuclear aberrations, and eventual necrosis are frequent changes seen in post endocrine therapy (Kennedy et al., 1990). Chemotherapy can induce vacuolar degeneration in the cancer cells. Neoadjuvant chemotherapy does not usually affect the histologic grading of the carcinoma (Kennedy et al., 1990). Our study showed that honey treatment led to various degenerative changes in breast cancer tissue. These are mainly due to cellular response to the injurious effect of crude honey on cancer cells. These histologic changes could imply good prognosis. In breast cancer tumour grading is more clinically significant than the number and type of patterns observed. Histologic grading involves identification of cellular pleomorphism, tubular formation and mitoses of the breast cancer cells. Nevertheless, the tumour histologic patterns may also have some impact on the prognosis.

Cytoplasmic vacuolization is a well-recognized morphological alteration seen in mammalian cells after exposure to many pathogens and different natural compounds (Shubin et al., 2016). It often accompanies cell death (Bouzas-Rodríguez et al., 2012; Li et al., 2013; Shubin et al., 2016; Singha et al., 2013). A portion of the inducers of irreversible vacuolization causes known types of caspase-independent cell death (Christofferson & Yuan, 2010; Overmeyer et al., 2008; Weerasinghe & Buja, 2012). Our present study indicates that cellular vacuolization is a characteristic outcome observed in the histopathology of crude honey-treated MNU induced breast cancer in female SD rats. This indicates that crude honey has a cytotoxic effect on the breast cancer cells and vacuoles can be a side effect of the action of cytotoxic factors in crude honey. It is important to correlate this alteration with the effect of honey on the cancer cell cycle and cell death.

Presence of necrosis indicates aggressiveness of a malignant tumour and implies a high proliferative rate of the cancer cells. Usually the higher the rate of tumour cells proliferation, the higher the occurrence of ischaemic necrosis in tumour tissue. Yet again post-treatment necrosis in a treated malignant tumour may indicate the positive response towards therapy. In our study, we found that the frequency of necrosis was less in the honey-treated group ($p=0.005$) compared with that of the control ones. It can be presumed that due to the growth inhibitory activity of honey, cancer cell proliferation drops and there is a lesser amount of ischemic necrosis.

In early tumours, tumour associated macrophages appear to have an inflammatory, tumouricidal M1 phenotype. However, with tumour progression, these macrophages polarise towards M2-like phenotype that is protumour, angiogenic and immune-inhibitory (Fridlender & Albelda, 2013). In the case of TILs, density and the phenotypic profile are important prognostic factors. Our study revealed that aggregates of macrophages were present in the stroma and parenchyma of both the non-treated and treated MNU induced breast cancers. However, these were less obvious in the treated ones. It would be interesting to analyse the phenotype of the lymphocytes and macrophages, especially in the treated cancers.

Among the broad range of breast cancer animal models, orthotopic xenograft models are cutting-edge cancer model for the evaluation of novel anticancer agents (Whittle et al., 2015). Patient-derived xenografts (PDXs) are not only expensive but also need many elaborate settings and requirements like fresh patient material, immunocompromised host. Furthermore, in PDX, mouse stroma replaces human with increasing passages (Holen et al., 2017). For our current study, we have used MNU induced SD rat breast cancer model as it is not only simple to develop but also cost-effective and addresses the specific research question and aim of our study quite effectively.

CONCLUSION

The present study is focused mainly on the various histomorphological alterations observed in MNU induced experimental breast cancer model in female SD rats after crude honey treatment. Crude honey has anti-cancer effect as seen histologically in this in vivo study. In this study for the first time, we have described the morphological alterations in MNU induced breast cancer in SD rats after crude honey treatment. Presence of single morphological pattern, lower histopathological grade, prominent cytoplasmic vacuolization in the cancer cells, frequent aggregates of tumour-associated atypical large macrophages with scanter necrosis and slightly augmented fibrosis were the major post-treatment outcomes observed. Lower histological grade of cancer with cytoplasmic vacuolisation in cancer cells and less necrosis implicates the positive efficacy of crude honey as an anticancer agent on breast cancer tissue. The overall post-therapeutic histomorphological alterations showed similarity with that of the post-therapeutic alterations caused by endocrine therapy and to some extent chemotherapy. These histopathological findings will direct towards a future anticancer study on honey and its derivatives. It will also help in molecular correlation and histological interpretation of honey-treated breast cancer tissue in the animal model. The histopathological parameters used in this study to describe the anticancer effect of honey in the animal model can be applied to evaluate the effect of other potential anticancer agents.

ACKNOWLEDGEMENT

The project is supported by USM RUI Grant 1001 PPSP 8012299.

REFERENCES

Ahmed, S., Sulaiman, S. A., & Othman, N. H. (2017). Oral administration of Tualang and Manuka honeys modulates breast cancer progression in Sprague-Dawley rats model. *Evidence-Based Complementary and Alternative Medicine: ECAM, 2017,* 1-15.

Banik, U., Parasuraman, S., Adhikary, A. K., & Othman, N. H. (2017). Curcumin: The spicy modulator of breast carcinogenesis. *Journal of Experimental and Clinical Cancer Research: CR, 36*(1), 1-16.

Bouzas-Rodríguez, J., Zarraga-Granados, G., del Rayo Sanchez-Carbente, M., Rodríguez-Valentín, R., Gracida, X., Anell-Rendón, D., ... & Castro-Obregón, S. (2012). The nuclear receptor NR4A1 induces a form of cell death dependent on autophagy in mammalian cells. *PLoS One, 7*(10), 1-12.

Christofferson, D. E., & Yuan, J. (2010). Necroptosis as an alternative form of programmed cell death. *Current Opinion in Cell Biology, 22*(2), 263-268.

Fridlender, Z. G., & Albelda, S. M. (2013). Modifying tumor-associated macrophages. *Oncoimmunology, 2*(12), 1-3.

Han, B., Peng, X., Cheng, D., Zhu, Y., Du, J., Li, J., & Yu, X. (2019). Delphinidin suppresses breast carcinogenesis through the HOTAIR/microRNA-34a axis. *Cancer Science, 110*(10), 3089-3097.

Holen, I., Speirs, V., Morrissey, B., & Blyth, K. (2017). In vivo models in breast cancer research: Progress, challenges and future directions. *Disease Models and Mechanisms, 10*(4), 359-371.

Jaganathan, S. K., Balaji, A., Vellayappan, M. V., Asokan, M. K., Subramanian, A. P., John, A. A., … & Marvibaigi, M. (2015). A review on antiproliferative and apoptotic activities of natural honey. *Anti-Cancer Agents in Medicinal Chemistry, 15*(1), 48-56.

Jaganathan, S. K., & Mandal, M. (2009). Antiproliferative effects of honey and of its polyphenols: A review. *Journal of Biomedicine and Biotechnology, 2009,* 1-13.

Kennedy, S., Merino, M. J., Swain, S. M., & Lippman, M. E. (1990). The effects of hormonal and chemotherapy on tumoral and nonneoplastic breast tissue. *Human Pathology, 21*(2), 192-198.

Li, X., Sui, C., Chen, Q., Chen, X., Zhang, H., & Zhou, X. (2013). Promotion of autophagy at the maturation step by IL-6 is associated with the sustained mitogen-activated protein kinase/extracellular signal-regulated kinase activity. *Molecular and Cellular Biochemistry, 380*(1), 219-227.

Lu, Y., You, M., Ghazoui, Z., Liu, P., Vedell, P. T., Wen, W., … & Lubet, R. A. (2013). Concordant effects of aromatase inhibitors on gene expression in ER+ Rat and human mammary cancers and modulation of the proteins coded by these genes an uninvited scientific article. *Cancer Prevention Research, 6*(11), 1151-1161.

Masood, S. (2016). Breast cancer subtypes: Morphologic and biologic characterization. *Women's Health, 12*(1), 103-119.

Mukhopadhyay, S., Ballard, B. R., Mukherjee, S., Kabir, S. M., & Das, S. K. (2006). Beneficial effects of soy protein in the initiation and progression against dimethylbenz [a] anthracene-induced breast tumors in female rats. *Molecular and Cellular Biochemistry, 290*(1-2), 169-176.

Overmeyer, J. H., Kaul, A., Johnson, E. E., & Maltese, W. A. (2008). Active ras triggers death in glioblastoma cells through hyperstimulation of macropinocytosis. *Molecular Cancer Research: MCR, 6*(6), 965-977.

Russo, J., & Russo, I. H. (2000). Atlas and histologic classification of tumors of the rat mammary gland. *Journal of Mammary Gland Biology and Neoplasia, 5*(2), 187-200.

Schnitt, S. J., Connolly, J. L., Harris, J. R., & Cohen, R. B. (1984). Radiation-induced changes in the breast. *Human Pathology, 15*(6), 545-550.

Shubin, A. V., Demidyuk, I. V., Komissarov, A. A., Rafieva, L. M., & Kostrov, S. V. (2016). Cytoplasmic vacuolization in cell death and survival. *Oncotarget, 7*(34), 55863-55889.

Singha, P. K., Pandeswara, S., Venkatachalam, M. A., & Saikumar, P. (2013). Manumycin A inhibits triple-negative breast cancer growth through LC3-mediated cytoplasmic vacuolation death. *Cell Death and Disease, 4*(1), e457- e457.

Spilioti, E., Jaakkola, M., Tolonen, T., Lipponen, M., Virtanen, V., Chinou, I., … & Moutsatsou, P. (2014). Phenolic acid composition, antiatherogenic and anticancer potential of honeys derived from various regions in Greece. *PLoS ONE, 9*(4), 1-10.

Weerasinghe, P., & Buja, L. M. (2012). Oncosis: An important non-apoptotic mode of cell death. *Experimental and Molecular Pathology, 93*(3), 302-308.

Whittle, J. R., Lewis, M. T., Lindeman, G. J., & Visvader, J. E. (2015). Patient-derived xenograft models of breast cancer and their predictive power. *Breast Cancer Research: BCR, 17*(1), 1-13.

WHO. (n.d.).| *Breast cancer.* World Health Organization. Retrieved August 14, 2019, from http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/

REFEREES FOR THE PERTANIKA
JOURNAL OF SCIENCE AND TECHNOLOGY

VOL. 28 (4) OCT. 2020

The Editorial Board of the Journal of Science and Technology wishes to thank the following:

Abd Muhaimin Amiruddin
(UPM, Malaysia)

Abdullah Hisam Omar
(UTM, Malaysia)

Ahmad Fikri Abdullah
(UPM, Malaysia)

Ahmad Ismail
(UPM, Malaysia)

Ainun Zuriyati Mohamed Asaari
(UPM, Malaysia)

Aishah Hani Azil
(PPUKM, Malaysia)

Amir Izzwan Zamri
(UMT, Malaysia)

Ana Sakura Zainal Abidin
(UNIMAS, Malaysia)

Anasyida Abu Seman
(USM, Malaysia)

Andrew Alek Tuen
(UNIMAS, Malaysia)

Chuan Zun Liang
(UMP, Malaysia)

Dat Tran Van
(IWEM, Vietnam)

Dzun Noraini Jimat
(IIUM, Malaysia)

Fatimah Khalid
(UPM, Malaysia)

Habshah Midi
(UPM, Malaysia)

Hashibah Hamid
(UUM, Malaysia)

Helmi Zulhaidi Mohd Shafri
(UPM, Malaysia)

Idham Khalil
(UMT, Malaysia)

Jasronita Jasni
(UPM, Malaysia)

Jayaraj Vijaya Kumaran
(UMK, Malaysia)

Kek Sie Long
(UTHM, Malaysia)

Khuzaimah Zailani
(UPM, Malaysia)

Lai Kee Huong
(Sunway University, Malaysia)

Lee Sau Har
(Taylor's University, Malaysia)

Lee Sin Chang
(UPM, Malaysia)

Lili Nurliyana Abdullah
(UPM, Malaysia)

Lim Chee Kau
(UM, Malaysia)

Mohamad Syazarudin Md Said
(UPM, Malaysia)

Mohammad Abdul Mojid
(BAU, Bangladesh)

Mohd Amran Mohd Radzi
(UPM, Malaysia)

Mohd Azri Mohd Izhar
(UTM, Malaysia)

Mohd Fadzil Mohd Akhir
(UMT, Malaysia)

Mohd Shareduwan Mohd Kasihmuddin
(USM, Malaysia)

Muhammad Nomani Kabir
(UMP, Malaysia)

Mundzir Abdullah
(USM, Malaysia)

Mus`ab Abd. Razak
(UPM, Malaysia)

Muthu Krishnan
(KIT, India)

Naji Mordi Al-Dosary
(KSU, Saudi Arabia)

Ng Kok Haur
(UM, Malaysia)

Nik Nor Liyana Nik Ibrahim
(UPM, Malaysia)

Norharyati Harum
(UTeM, Malaysia)

Noridayu Manshor
(UPM, Malaysia)

Norizah Kamarudin
(UPM, Malaysia)

Nur Anisah Mohamed Rahman
(UM, Malaysia)

Nuramidah Hamidon
(UTHM, Malaysia)

Nurul Hashimah Ahamed Hassain Malim
(USM, Malaysia)

Nwankwo Wilson
(Edo University, Nigeria)

Peter David Kulyakwave
(CAAS, China)

Rabiah Abdul Kadir
(UKM, Malaysia)

Razali Yaakob
(UPM, Malaysia)

Razana Mohd Ali
(UPM, Malaysia)

Syazwani Idrus
(UPM, Malaysia)

Syed Abdul Rahman Al-Haddad Syed Mohamed
(UPM, Malaysia)

Tay Chia Chay
(UiTM, Malaysia)

Wang Seok Mui
(UiTM, Malaysia)

BAU	- Bangladesh Agricultural University
CAAS	- Chinese Academy of Agricultural Sciences
IIUM	- International Islamic University Malaysia
IWEM	- Institute for Water Resources Economics and Management
KIT	- Kalaignarkarunanidhi Institute of Technology
KSU	- King Saud University
PPUKM	- Pusat Perubatan Universiti Kebangsaan Malaysia
UiTM	- Universiti Teknologi MARA
UKM	- Universiti Kebangsaan Malaysia
UM	- Universiti Malaya
UMK	- Universiti Malaysia Kelantan
UMP	- Universiti Malaysia Pahang
UMT	- Universiti Malaysia Terengganu
UNIMAS	- Universiti Malaysia Sarawak
UPM	- Universiti Putra Malaysia
USM	- Universiti Sains Malaysia
UTeM	- Universiti Teknikal Malaysia Melaka
UTHM	- Universiti Tun Hussein Onn Malaysia
UTM	- Universiti Teknologi 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 executive_editor.pertanika@upm.edu.my

Any inclusion or exclusion of name(s) on this page does not commit the *Pertanika* Editorial Office, nor the UPM Press or the University to provide any liability for whatsoever reason.

Pertanika Journals

Our goal is to bring high-quality research to the widest possible audience

INSTRUCTIONS TO AUTHORS

(REGULAR ISSUE)

(Manuscript Preparation & Submission Guide)

Revised: July 2020

Please read the *Pertanika* guidelines and follow these instructions carefully. 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 **4** types of manuscripts
- that have not been published elsewhere (including proceedings)
- that are not currently being submitted to other journals

1. Regular article

Regular article is a full-length original empirical investigation, consisting of introduction, methods, results, and discussion. Original research work should present new and significant findings that contribute to the advancement of the research area. *Analysis and Discussion* must be supported with relevant references.

Size: Generally, each manuscript is **not to exceed 6000 words** (excluding the abstract, references, tables, and/or figures)**,** a maximum of **80 references**, and **an abstract of less than 250 words.**

2. Review article

A review article reports a critical evaluation of materials about current research that has already been published by organising, integrating, and evaluating previously published materials. It summarises the status of knowledge and outlines future directions of research within the journal scope. A review article 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.

Size: Generally, it is expected **not to exceed 6000 words** (excluding the abstract, references, tables, and/or figures), a maximum of **80 references,** and **an abstract of less than 250 words.**

3. Short communications

Each article should be timely and brief. It is suitable for the publication of significant technical advances and maybe used to:

(a) reports 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) reports/discuss on significant matters of policy and perspective related to the science of the journal, including 'personal' commentary;

(c) disseminates information and data on topical events of significant scientific and/or social interest within the scope of the journal.

Size: It is limited to **3000 words** and have a maximum of **3 figures and/or tables, from 8 to 20 references, and an abstract length not exceeding 100 words.** The information must be in short but complete form and it is not intended to publish preliminary results or to be a reduced version of a regular paper.

4. Others

Brief reports, case studies, comments, concept papers, letters to the editor, and replies on previously published articles may be considered.

Language Accuracy

Pertanika **emphasises** on the linguistic accuracy of every manuscript published. Articles must be in **English** and they must be competently written and presented 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) **may be required to provide a certificate** confirming that their manuscripts have been adequately edited. **All editing costs must be borne by the authors**.

Linguistically hopeless manuscripts will be rejected straightaway (e.g., when the language is so poor that one cannot be sure of what the authors are really trying to say). This process, taken by authors before submission, will greatly facilitate reviewing, and thus, publication.

MANUSCRIPT FORMAT

The paper should be submitted in one-column format with 1.5 line spacing throughout. Authors are advised to use Times New Roman 12-point font and *MS Word* format.

1. Manuscript Structure

The 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's information***

General information: This page should contain the **full title** of your paper **not exceeding 25 words**, with the name of all the authors, institutions and corresponding author's name, institution and full address (Street address, telephone number (including extension), handphone number, and e-mail address) for editorial correspondence. **The corresponding author must be clearly indicated with a superscripted asterisk symbol (*).**

Authors' name: The names of the authors should be named **in full without academic titles.** For Asian (Chinese, Korean, Japanese, Vietnamese), please write first name and middle name before surname (family name). The last name in the sequence is considered the surname.

Authors' addresses: Multiple authors with different addresses must indicate their respective addresses separately by superscript numbers.

Tables/figures list: A **list** of the number of **black and white/colour figures and tables** should also be indicated on this page. See ***"5. Figures & Photographs"*** for details.

Example (page 2):

***In vivo* Fecundity Evaluation of *Phaleria macrocarpa* Extract Supplementation in Male Adult Rats**

Sui Sien Leong[1]* and Mohamad Aziz Dollah[2]

[1]Department of Animal Sciences and Fishery, Universiti Putra Malaysia, 97008 Bintulu, Sarawak, Malaysia
[2]Department of Biomedical Sciences, Universiti Putra Malaysia, 43400 Serdang, Malaysia

leongsuisien@upm.edu.my (Sui Sien Leong), Contact number
azizdollah@gmail.com (Mohamad Aziz Dollah), Contact number
**Corresponding author*

List of Table/Figure:
Table 1.
Figure 1.

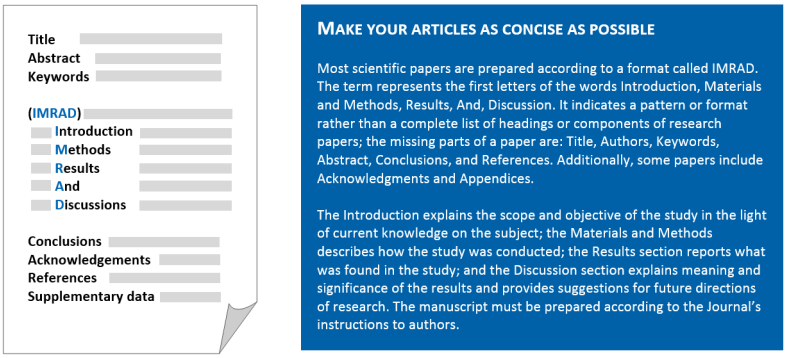
- **Page 3: *Abstract***

This page should **repeat** the **full title** of your paper with only the ***Abstract***, usually in one paragraph and ***Keywords*.**

Keywords: **Not more than 8 keywords in alphabetical order** must be provided to describe the content of the manuscript.

- **Page 4: *Text***

A regular paper should be prepared with the headings *Introduction, Materials and Methods, Results and Discussions, Conclusions, Acknowledgements, References,* and *Supplementary data* (if any) in this order. The literature review may be part of or separated from the *Introduction*.



2. *Levels of Heading*

Level of heading	Format
1st	**LEFT, BOLD, UPPERCASE**
2nd	**Flush left, Bold, Capitalise each word**
3rd	**Bold, Capitalise each word, ending with .**
4th	***Bold italic*, Capitalise each word, ending with .**

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 (Table 1, Table 2).
- A brief title should be provided, which should be shown at the top of each table **(APA format):**

Example: Table 1

PVY infected Nicotiana tabacum *plants optical density in ELISA*

- Explanatory material should be given in the table legends and footnotes.
- Each table should be prepared on a new page, embedded in the manuscript.
- Authors are advised to keep backup files of all tables.

***** Please submit all tables in Microsoft word format only, 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 a 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 (Figure 1, Figure 2).
- Provide a brief title, which should be shown at the bottom of each table **(APA format):**

  Example: Figure 1. PVY-infected *in vitro* callus of *Nicotiana tabacum*

- If a figure has been previously published, acknowledge the original source, and submit written permission from the copyright holder to reproduce the material.
- Authors are advised to keep backup files of all figures.

***** Figures or photographs must also be submitted separately as TIFF or JPEG, 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.***

6. Acknowledgement

Any individuals and entities who have contributed to the research should be acknowledged appropriately.

7. 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) (refer to the examples). 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* (https://apastyle.apa.org/) for further details.

Examples of reference style are given below:

Books

Books	**Insertion in Text**	**In Reference List**
Book with 1-2 authors	**Information prominent' (the author's name is within parentheses):** … (Cochrane, 2007) … Or **'Author prominent' (the author's name is outside the parentheses):** Cochrane (2007) …	Cochrane, A. (2007). *Understanding urban policy: A critical approach.* Malden, United States: Blackwell Publishing.
Book with 3 or more authors (*Pertanika*'s format)	*For all in-text references, list only the first author's family name and followed by* 'et al.' **Information prominent' (the author's name is within parentheses):** … (Seeley et al., 2011) … Or **'Author prominent' (the author's name is outside the parentheses):** Seeley et al. (2011) …	Seeley, R., VanPutte, C., Regan, J., & Russo, A. (2011). *Seeley's anatomy & physiology*. New York, United States: McGraw-Hill.

Books	Insertion in Text	In Reference List
Book with 6-7 authors	*For all in-text references, list only the first author's family name and followed by* 'et al.' **Information prominent' (the author's name is within parentheses):** … (Bulliet et al., 2011) … Or **'Author prominent' (the author's name is outside the parentheses):** Bulliet et al. (2011) …	Bulliet, R. W., Crossley, P. K., Headrick, D. R., Hirsch, S. W., Johnson, L. L., & Northrup, D. (2011). *The earth and its peoples: A global history* (5th ed.). Boston, United States: Wadsworth.
Book with more than 8 authors	*For all in-text references, list only the first author's family name and followed by* 'et al.' A recent study (Edge et al., 2011) concluded that… Or Edge et al. (2011) concluded that ….	For books with eight or more authors, please follow the guidelines for journal articles with eight or more authors.
Chapter in an edited book	**Information prominent' (the author's name is within parentheses):** … (Richards, 1997) … Or **'Author prominent' (the author's name is outside the parentheses):** Richards (1997) …	Richards, K. C. (1997). Views on globalization. In H. L. Vivaldi (Ed.), *Australia in a global world* (pp. 29-43). Sydney, Australia: Century.
e-book/online book	**Information prominent' (the author's name is within parentheses):** … (Niemann et al., 2004) … Or **'Author prominent' (the author's name is outside the parentheses):** Schiraldi (2001) …	Niemann, S., Greenstein, D., & David, D. (2004). *Helping children who are deaf: Family and community support for children who do not hear well.* Retrieved June 1, 2019, from http://www.hesperian.org/publications_download_deaf.php Schiraldi, G. R. (2001). *The post-traumatic stress disorder sourcebook: A guide to healing, recovery, and growth* [Adobe Digital Editions version]. doi:10.1036/0071393722
Editor	**Information prominent' (the author's name is within parentheses):** … (Zairi, 1999) … Or **'Author prominent' (the author's name is outside the parentheses):** Zairi (1999) …	Zairi, M. (Ed.). (1999). *Best practice: Process innovation management.* Oxford, United Kingdom: Butterworth-Heinemann.

Books	Insertion in Text	In Reference List
Several works by the same author in the same year	**Information prominent' (the author's name is within parentheses):** … (Fullan, 1996a, 1996b) … Or **'Author prominent' (the author's name is outside the parentheses):** Fullan (1996a, 1996b) …	Fullan, M. (1996a). Leadership for change. In *International handbook for educational leadership and administration.* New York, United States: Kluwer Academic. Fullan, M. (1996b). *The new meaning of educational change.* London, United Kingdom: Casell.
Several authors, different years referred to collectively in your work	**List sources alphabetically by the family name in the in-text reference in the order in which they appear in the Reference List.** The cyclical process (Carr & Kemmis, 1986; Dick, 2000) suggests…	Carr, W., & Kemmis, S. (1986). *Becoming critical: Education knowledge and action research.* London, United Kingdom: Falmer Press. Dick, B. (2000). *A beginner's guide to action research.* Retrieved June 1, 2019, from http://www.scu.edu.au/schools/gcm/ar/arp/guide.html

Journals

Journals	Insertion in Text	In Reference List
Journal article with 1-2 authors	**Information prominent' (the author's name is within parentheses):** … (Kramer & Bloggs, 2002) … Or **'Author prominent' (the author's name is outside the parentheses):** Kramer and Bloggs (2002) …	Kramer, E., & Bloggs, T. (2002). On quality in art and art therapy. *American Journal of Art Therapy*, *40*, 218-231.
Journal article with 3 or more authors (*Pertanika*'s format)	*For all in-text references, list only the first author's family name and followed by* 'et al.' **Information prominent' (the author's name is within parentheses):** … (Erlo et al., 2008) … Or **'Author prominent' (the author's name is outside the parentheses):** Erlo et al. (2008) …	Erlo, A., Ervasti, J., Kuosma, E., & Mattila, P. (2008). Evaluation of an organizational stress management program in a municipal public works organization. *Journal of Occupational Health Psychology*, *13*(1), 10-23. doi: 10.1037/1076-8998.13.1.10

Journal article with 6-7 authors	*For all in-text references, list only the first author's family name and followed by* 'et al.' **Information prominent' (the author's name is within parentheses):** … (Restouin et al., 2009) … Or **'Author prominent' (the author's name is outside the parentheses):** Restouin et al. (2009) …	Restouin, A., Aresta, S., Prébet, T., Borg, J., Badache, A., & Collette, Y. (2009). A simplified, 96-well-adapted, ATP luminescence-based motility assay. *BioTechniques, 47*, 871-875. doi:10.2144/000113250
Journal article with more than 8 or more authors	**Information prominent' (the author's name is within parentheses):** … (Steel et al., 2010) … Or **'Author prominent' (the author's name is outside the parentheses):** Steel et al. (2010) …	Steel, J., Youssef, M., Pfeifer, R., Ramirez, J. M., Probst, C., Sellei, R., ... & Pape, H. C. (2010). Health-related quality of life in patients with multiple injuries and traumatic brain injury 10+ years postinjury. *Journal of Trauma: Injury, Infection, and Critical Care, 69*(3), 523-531. doi: 10.1097/TA.0b013e3181e90c24
Journal article with DOI	**Information prominent' (the author's name is within parentheses):** … (Shaw et al., 2005) … Or **'Author prominent' (the author's name is outside the parentheses):** Shaw et al. (2005) …	Shaw, K., O'Rourke, P., Del Mar, C., & Kenardy, J. (2005). Psychological interventions for overweight or obesity. *The Cochrane Database of Systematic Reviews* (2). doi:10.1002/14651858.CD003818.pub2

Newspapers

Newspapers	Insertion in Text	In Reference List
Newspaper article – with an author	… (Waterford, 2007) ...	Waterford, J. (2007, May 30). Bill of rights gets it wrong. *The Canberra Times*, p. 11.
Newspaper article – without an author	… ("Internet pioneer", 2007) …	Internet pioneer to oversee network redesign. (2007, May 28). *The Canberra Times*, p. 15.
Article in a newsletter	... ("Australians and the Western Front", 2009) ...	Australians and the Western Front. (2009, November). *Ozculture newsletter*. Retrieved June 1, 2019, from http://www.cultureandrecreation.gov.au/newsletter/

Conference / Seminar Papers

Conference / Seminar Papers	Insertion in Text	In Reference List
Print – *If the paper is from a book, use the book chapter citation format. If it is from regularly published proceedings (e.g. annual), use the Journal article citation format*	… (Edge, 1996) … Or Edge (1996) …	Edge, M. (1996). Lifetime prediction: Fact or fancy? In M. S. Koch, T. Padfield, J. S. Johnsen, & U. B. Kejser (Eds.), *Proceedings of the Conference on Research Techniques in Photographic Conservation* (pp. 97-100). Copenhagen, Denmark: Royal Danish Academy of Fine Arts.
Online	… (Tester, 2008) … Or Tester (2008) …	Tester, J. W. (2008). The future of geothermal energy as a major global energy supplier. In H. Gurgenci & A. R. Budd (Eds.), *Proceedings of the Sir Mark Oliphant International Frontiers of Science and Technology Australian Geothermal Energy Conference*. Canberra, Australia: Geoscience Australia. Retrieved June 1, 2019, from http://www.ga.gov. au/image_cache/ GA11825.pdf

Government Publications

Government Publications	Insertion in Text	In Reference List
Government as author	**First in-text reference: Spell out the full name with the abbreviation of the body.** … (Department of Finance and Administration [DOFA], 2006) … **Subsequent in-text reference/s: Use the abbreviation of the body.** … (DOFA, 2006) …	Department of Finance and Administration. (2006). Delivering Australian Government services: Managing multiple channels. Canberra, Australia: Author.
Government report - online	**First in-text reference: Spell out the full name with the abbreviation of the body.** … (Department of the Prime Minister and Cabinet [PM&C], 2008) … **Subsequent in-text reference/s: Use the abbreviation of the body.** … (PM&C, 2008) …	Department of the Prime Minister and Cabinet. (2008). *Families in Australia: 2008.* Retrieved June 1, 2019, from http://www.....

8. 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.

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*.

Similarity Index: All articles received must undergo the initial screening for originality before being sent for peer review. *Pertanika* does not accept any article with a similarity index exceeding **20%**.

Copyright Permissions: Authors should seek necessary permissions for quotations, artwork, boxes or tables taken from other publications or other freely available sources on the Internet before submission to *Pertanika*. The *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, and tables 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. Each manuscript should follow one type of spelling only.

SUBMISSION OF MANUSCRIPTS

All submissions must be made electronically using the **ScholarOne™ online submission system**, a web-based portal by Clarivate Analytics. For more information, go to our web page and click "**Online Submission (ScholarOne™)**".

Submission Checklist

1. **MANUSCRIPT:**

   Ensure your manuscript has followed the *Pertanika* style particularly the first-4-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.

2. **DECLARATION FORM:**

- Author has to sign a declaration form. In signing the form, authors declare that the work submitted for publication is original, previously unpublished, and not under consideration for any publication elsewhere.
- Author has to agree to pay the publishing fee once the paper is accepted for publication in *Pertanika*.

3. **COVER LETTER:**

   In Step 6 of the ScholarOne system, author is asked to upload a cover letter in *Pertanika* format. Please ignore this instruction and replace the cover letter with the Declaration Form.

Note:

COPYRIGHT FORM: Author will be asked to sign a copyright form when the paper is accepted. 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.

Visit our Journal's website for more details at http://www.pertanika.upm.edu.my/

ACCESS TO PUBLISHED MATERIALS

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, "*Regular Issue*", "*Special Issue"* or "*Archives*". Here you will get access to all current and back-issues from 1978 onwards. No hard copy of journals or offprints are printed.

Visit our Journal's website at
http://www.pertanika.upm.edu.my/regular_issues.php for "Regular Issue"
http://www.pertanika.upm.edu.my/cspecial_issues.php for "Special Issue"
http://www.pertanika.upm.edu.my/journal_archives.php for "Archives"

PERTANIKA
JOURNALS

PUBLICATION CHARGE

Upon acceptance of a manuscript, a processing fee of RM 750 / USD 250 will be imposed on authors; RM 750 for any corresponding author affiliated to an institution in Malaysia; USD 250 for any corresponding author affiliated to an institution outside Malaysia. Payment must be made online at https://paygate.upm.edu.my/action.do?do=

Any queries may be directed to the **Chief Executive Editor's** office via email to
executive_editor.pertanika@upm.edu.my

Pertanika Journal of Science & Technology
Vol. 28 (4) Oct. 2020

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.edu.my
Tel: +603 9769 1622

PENERBIT UPM
UNIVERSITI PUTRA MALAYSIA
BERILMU BERBAKTI
PRESS

http://penerbit.upm.edu.my
E-mail : penerbit@upm.edu.my
Tel : +603 9769 8855

ISSN 0128-7680
10 >
9 770128 768083