About the Journal

Overview

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

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

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

The Journal is available world-wide.

Aims and scope

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

History

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

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

Goal of *Pertanika*

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

Quality

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

Abstracting and indexing of *Pertanika*

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

Future vision

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

Citing journal articles

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

Publication policy

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

Code of Ethics

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

International Standard Serial Number (ISSN)

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

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

Lag time

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

Authorship

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

Manuscript preparation

Refer to Pertanika's **Instructions to Authors** at the back of this journal.

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

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

Editorial process

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

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

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

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

The Journal's peer-review

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

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

Operating and review process

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

1. The Journal's chief executive editor and the editorial board examine the paper to determine whether it is appropriate for the journal and should be reviewed. If not appropriate, the manuscript is rejected outright and the author is informed.

2. The chief executive editor sends the article-identifying information having been removed, to three reviewers. Typically, one of these is from the Journal's editorial board. Others are specialists in the subject matter represented by the article. The chief executive editor asks them to complete the review in three weeks.

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

3. The chief executive editor, in consultation with the editor-in-chief, examines the reviews and decides whether to reject the manuscript, invite the author(s) to revise and resubmit the manuscript, or seek additional reviews. Final acceptance or rejection rests with the Edito-in-Chief, who reserves the right to refuse any material for publication. In rare instances, the manuscript is accepted with almost no revision. Almost without exception, reviewers' comments (to the author) are forwarded to the author. If a revision is indicated, the editor provides guidelines for attending to the reviewers' suggestions and perhaps additional advice about revising the manuscript.

4. The authors decide whether and how to address the reviewers' comments and criticisms and the editor's concerns. The authors return a revised version of the paper to the chief executive editor along with specific information describing how they have answered' the concerns of the reviewers and the editor, usually in a tabular form. The author(s) may also submit a rebuttal if there is a need especially when the author disagrees with certain comments provided by reviewer(s).

5. The chief executive editor sends the revised paper out for re-review. Typically, at least one of the original reviewers will be asked to examine the article.

6. When the reviewers have completed their work, the chief executive editor in consultation with the editorial board and the editor-in-chief examine their comments and decide whether the paper is ready to be published, needs another round of revisions, or should be rejected.

7. If the decision is to accept, an acceptance letter is sent to all the author(s), the paper is sent to the Press. The article should appear in print in approximately three months.

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

Pertanika Journal of Science & Technology
Vol. 26 (2) Apr. 2018

Contents

Case Studies

Foreword

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

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

This issue contains **30 articles,** of which **five** are review articles, **two** are case studies and **23** are regular research articles. The authors of these articles hail from several countries namely, **Malaysia**, **the United Kingdom**, **Nigeria**, **Australia**, **India**, **Iraq**, the **Republic of Ghana**, **Iran**, **the Kingdom of Morocco** and **Thailand**.

The first review article in this issue reports briefly on termite mounds as bio-indicators of groundwater, reporting on prospects and constraints (*Jamilu Ahmed* and *Biswajeet Pradhan*), while the second is on the use of waste materials in concrete (*Davoud Tavakoli, Masoumeh Hashempour* and *Ali Heidari*). The third review article is a comparative study that discusses modelling parallel programmes for static mapping (*Sina Zangbari Koohi, Nor Asilah Wati Abdul Hamid, Mohamed Othman* and *Gafurjan Ibragimov*). The next review article looks at design of a new safe operating product, and is a review of classification and opportunities related to the subject (*Fatimazahra Guini, Abdellah El Barkany, Abdelouahhab Jabri* and *El Hassan Irhirane*), while the final review article examines space medicine as the next frontier for the heart (*Balasingam, M.*).

One of the case studies discusses the rare case of subglottic ectopic thyroid tissue causing upper airway obstruction and its detection by CECT and MRI examination of the neck (*Ngah, N. A., SitiJusnaMuhammad, Suraini, M. S., Mohd Noh, M. S. F., AbHamid, S., Salleh, H.* and *Suppiah, S.*). The other case study looks at an uncommon vascular cause of spontaneous hydropneumothorax (*Balakrishnan, D., Suraini, M. S., Hazman, M. N., Hariati, J., R, Mahmud.* and *Ezamin, A. R.*).

The 23 regular articles cover a wide range of topics. The first article is on the evaluation of single reservoir performance for flood risk reduction using a developed simulation model, and is a case study of the Makhoul Reservoir (*Yousif H. Al-Aqeeli, S. Abd Aziz, Badronnisa Yusuf* and *Aimrun Wayayok*). The following articles look at: a comparison of goat skin and x-ray film membranes on the traditional musical instrument, the kompang (*W. A. Siswanto* and *M. Syiddiq*); an efficient method for secure 2D image visualisation and transmission through chaotic confusion and pixel diffusion (*Gunasekaran, G.* and *Venkatesan, M.*); socio-demographic factors associated with low birth weight in Nepal using imputation of missing determinants (*Singh, U., Ueranantasun, A.* and *Kuning, M.*); production of lentiviral vector with polycistronic transcripts for reprogramming of mouse fibroblast cells (*Akram Al Abbar, Norshariza Nordin, Siew Ching Ngai* and *Syahril*

Abdullah); increased endothelial progenitor cells with age and grade of malignancy in astrocytic glioma patients (*Priscilla Das, Nyi Nyi Naing, Nadiah Wan-Arfah, KON Noorjan, Yee Cheng Kueh* and *Kantha Rasalingam*); synthesis, characterisation and biological activities of Ru(iii), Mo(v), Cd(ii), Zn(ii) and Cu(ii) complexes containing a novel nitrogen-sulphur macrocyclic Schiff base derived from glyoxal (*Chah, C. K., Ravoof, T. B. S. A.* and *Veerakumarasivam, A.*); the effect of pre-processing and testing methods on online Kannada handwriting recognition using signal processing and statistical techniques (*S. Ramya* and *Kumara Shama*); statistical modelling of daily rainfall variability patterns in Australia (*Bright Emmanuel Owusu* and *Nittaya McNeil*); simulation and optimisation of bioethanol purification using extractive distillation with an additive solvent (*S. M. Anisuzzaman, D. Krishnaiah, A. Bono, F. A. Lahin, E. Suali* and *I. A. Z. Zuyyin*); the feasibility of 24-hour discharge for post-laparoscopic ovarian cystectomy patients (*Wan Ahmad Hazim, Nur Hidayah Aeshah Ng* and *Salleha Khalid*); the performance comparison of classification algorithms for medical diagnosis (*Anju Jain, Saroj Ratnoo* and *Dinesh Kumar*); estimating the economic impact of climate change on agricultural water management indicators (*Hayat Lionboui, Tarik Benabdelouahab, Fouad Elame, Aziz Hasib* and *Abdelali Boulli*); integrated rainwater drainage system for groundwater improvement and economic benefit (*Manisha D. Desai* and *Jayantilal N. Patel*); mathematical modelling and residual life prediction of an aluminium electrolytic capacitor (*Bhargava, C., Banga, V. K.* and *Singh, Y.*); comparison extraction of peanut skin between CO_2 supercritical fluid extraction and Soxhlet extraction in terms of oil yield and catechin (*Nicky Rahmana Putra, Mohd Azizi Che Yunus, Muhammad Syafiq Hazwan Ruslan, Zuhaili Idham* and *Fadillah Nur Idrus*); watermelon radioprotection against radiation stress in mice (*Md Saad, W. M., Mohd Nor, N. A., Abdul Razak, H. R., Mat Salleh, M. S., Abdul Rasid, A.* and *Hanifah, R.*); utilisation of PET-CT in oesophageal cancer management from a clinician's perspective (*Abdul Razak, H. R., Azmi, N. A.* and *Vinjamuri, S.*); a mono-window algorithm for land surface temperature estimation from Landsat 8 thermal infrared sensor data, reporting on a case study of the Beas River basin, India (*Gopinadh Rongali, Ashok Kumar Keshari, Ashvin Kumar Gosain* and *Rakesh Khosa*); application of almost an increasing sequence for the absolute Riesz $|\bar{N}, p_n^\alpha; \delta; \gamma|_k$ summable factor (*Sonker, Smita* and *Munjal, Alka*); the effect of tool-pin profile on weld zone and mechanical properties in friction stir welding of the aluminium alloy (*Prashant Prakash, Sanjay Kumar Jha* and *Shree Prakash Lal*); computation of antifractals i.e. tricorns and multicorns and their complex nature (*Narayan Partap, Sarika Jain* and *Renu Chugh*) and an explicit drain current model in a subthreshold regime for the graded channel Schottky barrier gate all around MOSFET to improve analog/RF performance (*Suman Sharma, Rajni Shukla* and *Malay Ranjan Tripathy*).

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

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 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.

I would also like to express my gratitude to all the contributors, namely, the authors, reviewers and editors for their professional contribution towards making this issue feasible.

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

Chief Executive Editor
Nayan Deep S. KANWAL, FRSA, ABIM, AMIS, Ph.D.
nayan@upm.my

Review Article

Termite Mounds as Bio-Indicators of Groundwater: Prospects and Constraints

Jamilu Bala Ahmed II[1,2] and Biswajeet Pradhan[1,3]*

[1]*Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

[2]*Department of Geology, Faculty of Science, Federal University Lokoja, Kogi State, 1154, Lokoja Nigeria*

[3]*School of Systems, Management and Leadership, Faculty of Engineering and Information Technology, University of Technology Sydney, New South Wales, Australia*

ABSTRACT

Reliance on modern sophisticated equipment for making 'discoveries' has limited the human power of observing subtle clues in the environment that are capable of saving cost and labour that come with researching new resources and methods to improve life for all. Due to the growing scarcity of potable water, especially in African and Asian countries, newer, cheaper and reliable methods of investigating groundwater resources are becoming critical. One such potentially promising method is mapping the distribution of termite mounds in the environment. Termite mounds are conspicuous landscape features in tropical and sub-tropical regions of the world. Built from surrounding soils by several species of termite, the properties of mound soil are relatively different from the surrounding soil in most cases, indicating improved hydraulic properties. In this paper, the aim is to review the possibility of employing termite mounds as prospecting tools for groundwater search from three spatial scales of observation. From assessing the smallest to the highest scale of observation, it can be concluded that termite mounds' prospect as surface indicators of groundwater is apparent. Review findings indicate increased surface water infiltration, presence of riparian tree vegetation and other trees with tap-root system around termite mounds, linear assemblage of termite mounds along aquiferous dykes and seep-lines as well as the dependence of termites on water but avoidance of places with risk of inundation. Whether they indicate permanent groundwater reserves in all cases or whether all species depend largely on water for their metabolism is a subject for further research.

INTRODUCTION

Access to water is recognised as a key limiting factor to socioeconomic development of any nation (United Nations, 1997). Management of water resources has been a major issue for stakeholders in recent years (Adelana et al., 2008; Tijani et al., 2016) as competition for economic development, associated with rapid growth in population, agricultural mechanisation and urbanisation has brought in significant land use changes and increased demand of water (Pradhan, 2009; Akankpo & Igboekwe, 2012; Fashae et al., 2013). This is especially true for developing countries like those in Africa where water scarcity has brought untold hardship, including poverty and civil unrest among the population (Ferriz & Bizuneh, 2002; El-baz, 2008). Water is plentiful on our mother Earth (Plummer et al., 2010), but it is usually not available where and when needed nor is the quality always suitable for all purposes, especially human use (Al-Abadi & Al-Shamma, 2014). Of all the accessible freshwater available on Earth, groundwater accounts for about 60% (Manap et al., 2011), which is about 35 times the water in all rivers and lakes (Plummer et al., 2010). In addition, groundwater has the advantages of low development cost, excellent natural quality, limited vulnerability, drought reliability and availability in vast geological formations. In addition, it is a dependable water supply source across all climatic zones (Jha et al., 2007). Notwithstanding the advantages, one constraint with groundwater exploration is actually locating the favourable spot or water-bearing fissure for its development as there is no one-hundred-percent chance of success in locating the resource anywhere. Thus, the search for groundwater, especially in difficult or complex geological terrains, is not only expensive but time consuming, labourious (Fenta et al., 2014) and often associated with a high rate of failure (Edet et al., 1998; Bala et al., 1999). As most communities in both developed and developing parts of the world are increasingly depending on groundwater for various uses (Oh et al., 2011; Kura et al., 2014; Park et al., 2014), it is necessary to identify natural features in the environment that can aid narrowing down our groundwater search to places that hold promise.

Places with good promise could be indicated by the prevalence of conspicuous soil mounds constructed by termites, also referred to as "ecosystem engineers" or "soil engineers" (Dangerfield et al., 1998; Bottinelli et al., 2014; Jouquet et al., 2016a). Termites are one of the most successful groups of social insects whose engineering resourcefulness include selection of soil particles (sorting) from depths for the construction of their below and/or above ground nest (Arhin et al., 2015), construction of galleries, tunnels and foraging holes below ground level (Mando et al., 1996) and burrowing of large quantities of soil underground to bring to the surface for litter harvesting (Bottinelli et al., 2014; Jouquet et al., 2015a) among others. The result is that they successfully modify the physical (change soil porosity and bulk density), hydraulic (increase soil water retention, induce higher infiltration rate), chemical (change soil pH, C, N and CEC) and biological (change soil organic matter, nutrient cycling) characteristics of their nests soil compared to the adjacent soil (e.g. Choosai et al., 2009; Jouquet et al., 2015a; Moura et al., 2014). The concentration of termite activities in a location increases ecological heterogeneity, and the nest environments serve as fertility islands or hotspots for plant and tree growth (Jouquet et al., 2005; Bonachela et al., 2015), reflecting improvement in soil characteristics including water content. Consequently, the engineering activities of termites have been an interesting subject area for research by soil scientists, plant scientists,

entomologists, agronomists, veterinarians, ecosystem managers, civil engineers, architects and mineral explorationists. The hydrogeologist whose concern is to locate groundwater in appreciable quantity and quality should also be featured in this interest list.

It has been observed in various parts of Africa that termite mound locations are locally exploited for locating groundwater sources. For example, in the North-central part of Nigeria, largely covered by basement complex rocks of Eburnean-Pan African age, some villagers rely on epigeal termite mounds for locating groundwater sources (personal observation). A similar observation has been made in Cameroon (Van Ranst, personal communication) and Ghana (Dowuona et al., 2012). This paper, therefore, aimed to review the possibility of utilising termite mounds as prospecting tools for locating suitable groundwater sources from three different spatial scales of observation viz. mound soil aggregate scale, mound profile scale and landscape scale. Jouquet et al. (2016a) reviewed the impact of termites on soil structure and water dynamics on four scales of observation, where they focussed on soil water content (shallow depth) as it affected plant species heterogeneity as opposed to groundwater storage for community supply.

BASIC FEATURES OF TERMITES

Termites are eusocial insects that are mainly divided into lower (paraphyletic) and higher (monophyletic) species (Jouquet et al., 2016a). The higher species (infraorder Isoptera) is most important in this regard, with over 3000 known species (Rajeev & Sanjeev, 2011; Sarcinelli et al., 2009). They are said to comprise about 75% of modern termite species (Krishna et al., 2013 from Jouquet et al., 2016a) that feed on various food types ranging from wood, leaf litter, crop residue and soils to animal dung (Bottinelli et al., 2014; Jouquet et al., 2016a). Although termites are found worldwide with the exception of Antarctica, they are particularly important in the tropical and subtropical savannahs of Africa, Australia, Asia, and South America (Bonachela et al., 2015) as they account for higher diversities in these continents and build high-rising and long-lasting above-ground nests (Jouquet et al., 2005; Sarcinelli et al., 2009; Mujinya et al., 2013; Nauer et al., 2015). Termite nests, preferably referred to as 'mounds' are built from soil material (mostly clay), saliva and excreta (Denovan et al., 2001; Rajeev & Sanjeev, 2011; Jouquet et al., 2016a). Mound shapes vary in terms of their architecture from conical, dome, cathedral and mushroom to lenticular in shape (Arhin & Nude, 2010; Arhin et al., 2015) which according to Korb and Linsenmair (1998a, 1998b) are dependent on environmental temperature conditions. The height of mounds could be a function of species type, clay availability and level of disturbance in the environment, but generally range from almost a few centimetres above ground level to as high as 2 m as recorded in Southern India (Jouquet et al., 2015a), 4.4 m in Namibia (Grohmann et al., 2010), and up to about 8 m in D. R. Congo (Mujinya et al., 2013, 2014).

Table 1
Termite diversity by bontinent (order isoptera)

Continent	Africa	Asia	S. America	Australia	N. America	Europe
No. of species	1,000	435	400	360	50	10

Source: UNEP (2000)

Termite mounds, especially those built by the fungus-growing species (Termitidae, Macrotermitinae subfamily) present only in Africa and Southeast Asia, are conspicuous features of many landscapes of tropical and sub-tropical regions of the continents (Levick et al., 2010; Jamali et al., 2013). Termites build these mounds to protect themselves against predators and sunlight (Korb & Linsenmair, 2000; Jouquet et al., 2016a) and for the maintenance of high humidity, temperature and food (Jouquet et al., 2016a). These mounds can be viewed from remotely sensed images when their density is high and the patterning is regularly spaced (Francis et al., 2012; Mujinya et al., 2014; Adhikary et al., 2016). They are, however, mostly mapped by field measurements (Roose-Amsaleg et al., 2005; Ackerman et al., 2007; Sako et al., 2009). This is because field mapping avails the opportunity to take measurements such as height and diameter of mounds as well as record activity status, all of which is not achievable with satellite images. However, the use of airborne surveys carrying a Light Detection and Ranging (LIDAR) sensor was successfully used to map termite mounds in Kruger National Park, South Africa (Levick et al., 2010; Davies et al., 2014a, 2014b, 2016). The surveys were successful in recording mound heights and diameters but for mostly heights >1 m. However, this survey type may not be applicable if activity status and species identification are required.

(a) (b)

(c) (d)
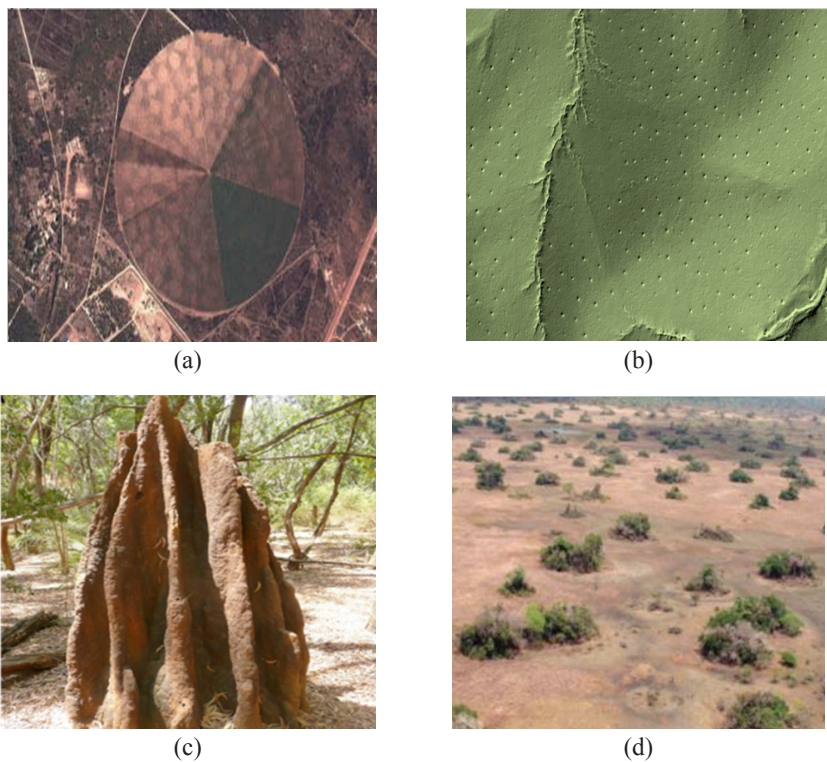
Figure 1. Termite mounds as viewed from remotely sensed images and field mapping. **(a)**: Patches left behind by termite mounds after two years of clearing as seen from Google Earth (Adhikary et al., 2016), **(b)**: Over-dispersed termite mounds at Kruger National Park, South Africa viewed from LIDAR image (Davies et al., 2014b), **(c)**: Single cathedral termite mound devoid of vegetation (source: Jouquet et al., 2016b), **(d)**: Termite mounds covered by trees in Sofala, Mozambique, image taken from helicopter (Source: Bonachela et al., 2015)

Various studies have underscored the benefits of termite mounds to the environment. These include growth of certain species of vegetation as a result of improved physical, chemical and biological conditions of mound soil (e.g. Mando et al., 1996; Choosai et al., 2009; Bonachela et al., 2015), accumulation of minerals such as Ag, U, Cu, Cd, Ni, Co, Mn, Pb and Zn in mounds as a result of termites' burrowing activities below ground (Kebede, 2004; Arhin et al., 2015), human exploitation of mound soil for erection of structural buildings because of their sesquioxide content (Yamashina, 2010) and consumption of mound soil for their geophagic effects (Kalumanga et al., 2016). Although unstudied, it is likely, therefore, that termite mounds can serve as pointers to local groundwater distribution in the environment because of the improved physical and hydraulic conditions of the mound soil (Ackerman et al., 2007; Moura et al., 2014) as well as the strategic positioning of the mounds themselves (Mege and Rango, 2009; Davies et al., 2014b; Kalumanga et al., 2016).

Improved hydraulic conditions of mound soil may significantly increase surface water infiltration and reduce runoff, thereby increasing the chances of groundwater recharge (Bargués-Tobella et al., 2014). This phenomenon is somewhat understood by some rural Africans who site their water wells around high-rise termite mounds. Also, trees springing out from termite mounds have their leaves ever fresh even during extended dry periods, an indication of a water source below. On the basis of positioning, places prone to surface water submergence are avoided by termites but because water is essential for livelihood, they may line their nests along seepage lines through which groundwater is recharged (Mege & Rango, 2009) and in other cases along groundwater outcrop areas (spring) (Davies et al., 2014b).

DISCUSSION OF THE THREE SCALES OF OBSERVATION

Analysing termite mounds' prospects as groundwater indicators from different spatial scales could serve the purpose of discerning subtle clues from the small scale (often considered trivial) to more significant evidence at intermediate and larger scales.
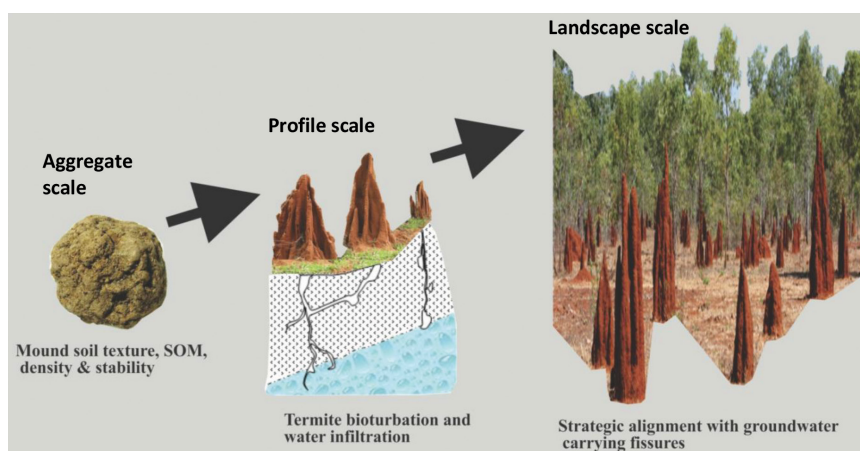
Figure 2. Termite mounds at three scales of observation. At aggregate scale, changes in soil texture, organic matter and bulk density have implication on water hydraulics. At profile scale, termite burrowing activities can influence surface water infiltration, while at landscape scale it can indicate water-bearing fractures and suitable places for groundwater exploration

Aggregate Scale

At the smallest scale of observation, the important factors to be considered are soil texture, organic matter and bulk density. Numerous studies have investigated mound soil texture in comparison with off-mound soil, and the results have shown some variability. For example, clay content has been reported to be higher (Roose-Amsaleg et al., 2005; Mujinya et al., 2010; Dowuona et al., 2012; Jouquet et al., 2015b; Adhikary et al., 2016), proportionately (Levick et al., 2010) and slightly lower (Ackerman et al., 2007) than surrounding off-mound soils. Clay enrichment on mound soil is mostly attributed to fungus-growing termite species that have the ability to select certain size fractions of soils in mound construction (Mujinya et al., 2010; Jouquet et al., 2016a). Soils generally are weathering products of rocks and their textures have varying implications to water drainage and storage. Depending on the mineral constituents present, some rocks weather to produce more clay than others. It will then be expected that termite mounds will thrive more in such clay abundant areas because of the availability of construction material, but instead, such areas tend to be avoided. For example, Levick et al. (2010) and Davies et al. (2014b) observed the abundance of termite mounds on sandy granitic soils than on clayey basaltic soils in Kruger National Park, South Africa. If termites depend on soil properties of the environment as pointed out by Jouquet et al. (2015b), the reason as to why they neglect clay-rich environments has largely not been given. Clay on the surface store large quantities of water and at the same time serve as an impediment to subsurface infiltration. Accordingly, it can be viewed that termites require clay fractions for stability of their mounds because clay has more binding and cementing capability than the remaining soil fractions but at the same time, requires good drainage conditions that might not be offered in clay-rich environments. As a result, an appreciable percentage of silt and sand is mixed with enriched clay content in moderate clay environments for mound construction (Seymour et al., 2016). The higher the percentage of clay used for mound construction, the more the increase in soil porosity and water-holding capacity at mound sites, but these translate to reduced soil drainage; hence, the soil exhibits low hydraulic conductivity. However, termites are reported to avoid soils with high water holding capacity (Schuurman & Dangerfield, 1997) and a decrease in clay content with depth beneath termite mounds has been reported in the Upper Katanga, D. R. Congo (Mujinya et al., 2010). Alternatively, the mounds are restricted to drier climatic zones or where prolonged episodes of rainfall are not experienced (Kandasami et al., 2016). Hence, it can be summarised that although termite mounds are constructed from clay, which may serve as impediment to vertical drainage, finer particles contribute to capillary action that allows water to be transported from the ground to the upper reaches of the mounds (Dangerfield et al., 1998; Turner, 2000).

Soil organic matter (SOM) is another factor that determines the stability of mound soil aggregates (Jouquet et al., 2016a) as well as the physical properties of mound soil. SOM is directly related to soil texture as it tends to increase as the clay content increases and results in aggregate formation by binding the soil particles together with their micropores and macropores (Food and Agriculture Organisation [FAO], 2005). This makes termite mounds relatively resistant to degradation by rain and other agents, decreasing runoff and erosion. The literature discloses higher SOM content in termite mounds compared with off-mound soil, especially for mounds constructed by soil-feeding termite species because they add their faeces

and other biogenic structures as binding materials to the construction medium (Moura et al., 2014; Jouquet et al., 2004; 2011; 2016a). The effect of higher SOM content in soil indirectly impacts positively on soil porosity as a consequence of increased micro-organism activity in the soil, promoting continuity of pore space; hence, soil water storage is increased and probably continues its infiltration into the saturation zone. However, SOM decreases with increase in depth (Brauman, 2000; Jouquet et al., 2015b) and because investigations at this scale have rarely exceeded the foot zone (ground level) of termite mounds (Muyinya et al., 2010; Dowuona et al., 2012), it is difficult to suggest if modification in physical and hydraulic conditions of soil below mound structures extend to greater depths in reflection of SOM content. If not, surface water infiltration will be limited to mound soil and might not exceed a few metres beneath the mound as demonstrated by Choosai et al. (2009), where a flooded paddy field was found to be unsaturated at depth 0-30 cm. Furthermore, an increase in porosity following enrichment in SOM content on termite mounds might be insignificant compared with adjacent soil that is porous and sandy.

Bulk density is dependent on soil organic matter, soil texture and their packing arrangement. Generally, loose, well aggregated, porous soil and soil rich in organic matter have lower bulk density (USDA, 1998). The engineering dexterity of termites separates them into two groups as regards density of their mounds. While some species such as Cubitermes, Trinervitermes and Macrotermes can compact their mound soils raising bulk density high to frustrate plant colonisation, others such as Amitermes do not compact their mound soils (Malaka, 1977). Ekundayo and Aghatise (1997) obtained slightly higher bulk densities on mound surfaces constructed by Macrotermes sp. compared to surrounding soils in fallow and cropped lands of Mid-Western Nigeria. Similarly, Dowuona et al. (2012) obtained higher bulk density on mounds sited on rhodic acrisol soil of coastal savannah of Ghana and also showed the bulk density increasing with depth. On the other hand, Ackermann et al. (2007) and Tilahun et al. (2012) found no significant difference between the mean bulk density of mounds and control soil in Amazonia, Brazil and Southern Ethiopia, respectively. Bulk density has direct impact on soil porosity (Haghnazari et al., 2015) and its severity is dependent on soil texture (Abdel-Magid et al., 1987).

Increased bulk density of mound soil can be attributed to its higher clay and silt fractions, which increase its water-holding capacity, making it much easier for termites to work on its compaction. Needless to say, the activities of predators, especially browsing animals like elephants, reported to feed on mound soil (Kalumanga et al., 2016), can contribute to this compaction as does the thumping effect of rain especially on mounds without canopy protection. Although compaction increases the aggregation and stability of mound soils, it leads to closing up of pore space and reduced hydraulic conductivity. This can be aggravated when mound soils are composed of carbonate minerals such as $CaCO_3$, which causes crust formation thereby, further reducing infiltration rate (Mujinya et al., 2010; Adhikary et al., 2016). Conversely, plants and trees are observed to flourish on other mounds, a manifestation of enhanced physical and hydraulic characteristics that provide a local source of water to sustain vegetation all through the year (Turner, 2006; Pardeshi & Prusty, 2010). Tree species such as Combretum imberbe and Phalaenopsis violacea described as lowland and riparian habitat species were found with increased density around termite mounds on hill crests in Kruger National Park, South Africa

(Davies et al., 2016). The soil conditions of these mounds, even though on hill crests several meters above the surrounding plains, mimic that of lowland areas, where the soil properties such as texture, bulk density and organic matter content favour nutrient and water availability. Although with reduced density, high water demanding tree species such as Combretum apiculatum and Colophospermum mopane were also found around termite mounds (Davies et al., 2016).

The pros and cons at this scale abound, making it difficult to deduce whether or not modification in mound soil properties will result to any significant increase in runoff collection and storage underground, especially when considering the limited area covered by individual mounds, density coverage of mounds per hectare and the volume of soil material used for mound construction. The almost generalised conclusion of clay enrichment on termite mound soils further heighten the constraints of large groundwater storage below the mounds as the clay soil will not only wick in water into the mound but will also hold it tightly (Turner, 2006), thus restricting storage below ground.
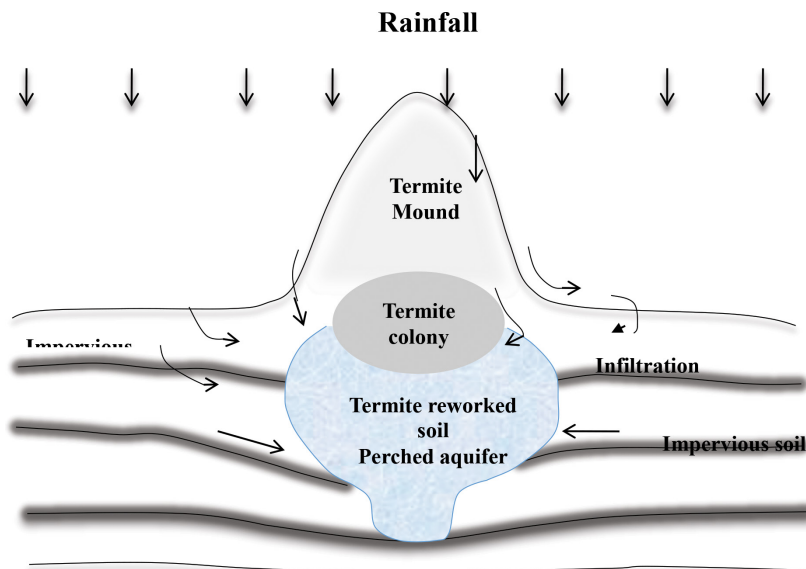
Figure 3. Schematic sketch of rainwater infiltration around termite mound due to improved soil physical properties and storage in perched aquifer below colony chamber. Modified from Turner (2006)

Mound Profile Scale

Perhaps a larger scale of observation might clear some ambiguity and present a better understanding of this topic. At the profile scale, the construction of the network of foraging galleries, tunnels and shafts of varying inclination from the mound surface to depths below the mound in a process known as bioturbation is among the reckoned engineering skills of termites (Arhin et al., 2015). Termites burrow large quantities of soil below ground for the construction of surface mounds and other features within the interior of mounds such as the

queen's chamber, fungus garden, vertical chimney and networked tunnels and galleries (Turner, 2000; Kandasami et al., 2016) for maintaining efficient exchange of respiratory gasses and regulation of humidity and temperature (Jouquet et al., 2006; Nauer et al., 2015). Termite excavating activities create macropores on mound soil as well as destruction of the crust below the mounds, increasing the rate of surface water infiltration, which can continue long after termite activity has ended (Léonard & Rajot, 2001) conducted infiltration tests using both simulated rainfall and ponding methods on termite mounds in Niger and found the infiltration capacity of mounds built by *Macroterme Subhyalinus* to be as high as 9 cm³/s. However, they stressed the possibility of this to the large number of foraging holes exceeding 30 per square metre and located on positions topographically too low to be able to intercept surface water. In Amazonia, Brazil, Ackerman et al., (2007) used the constant head single ring method to obtain infiltration rates six times greater on mounds than in the surrounding control soils. However, other instances have recorded lower infiltration rates on termite mounds than in the surrounding soils. This is mostly attributed to the low permeability of the clay that makes up the mound and the near absence of surface conduits or foraging holes on some mounds (Dowuona et al., 2012; Bargués Tobella et al., 2014). Again, studies at this scale have been limited by depth of investigation, thereby constraining our appreciation of how deeply the soil profile is affected by termite tunnelling and burrowing activities. At a soil depth of about 3 m below the termite mound foot, Mujinya et al. (2013) observed evidence of groundwater from the mottling of Mn-Fe oxides and attributed it to reducing and oxidising conditions in the mound as a result of water table fluctuation in the perched aquifer below. Other studies with clearly different objectives discovered by chance a high groundwater table during both wet and dry seasons in places with high densities of termite mounds (Mujinya et al., 2011; Dowuona et al., 2012; Kandasami et al., 2016). However, a high groundwater table in the dry season is not known with perched aquifers as they lack surface water replenishment during this season, thereby drying up. This can be catastrophic to the termite colonies that depend so much on water for their metabolism (Dangerfield et al., 1998). It can, therefore, be suggested that the aquifers are more inclined to being permanent than perched.

Investigations into deeper soil depths have greatly contributed to the understanding of why termites create burrows in soil profiles. It appears that their dependence on permanent water supply and avoidance of periodic inundation (Mege & Rango, 2009; Levick et al., 2010; Davies et al., 2014b) are the primary reasons. In Rhodesia (present-day Zimbabwe), underground mining of gold availed a number of researchers the opportunity to study the relationship between termites and gold anomalies recorded on termite mounds. At a depth of about 200 ft, termites were observed to have created burrows from soil surface through rock fissures below to access groundwater in permanent underground reservoirs (West, 1965). Watson (1972) identified Odontotermes laterius as tunnelling through Kalahari sand to locate fractured sections of the underlying basement complex rock and continue down through a gold mine shaft to access groundwater-bearing fissures at a depth of about 27 m. On their return to the surface, they picked gold samples that were deposited on mounds. For this reason, termite mounds have become valuable prospecting sites for mineral exploration (e.g. West, 1965; 1970; Arhin et al., 2015) but ironically, not for groundwater search. Fluctuations in water level in the fissure occur seasonally, rising during the wet season and lowering in dry season. To keep up

with the lowering water table in dry season, the termites excavate deeper into the fissures and deposit the material in the mound (West, 1970). This could be the explanation why Amitermes, Cubitermes and Trinervitermes termite species are acknowledged to enlarge or reconstruct their mounds after the rainy season (Skaife, 1955; Sands, 1961a as cited in Bouillon, 1970).
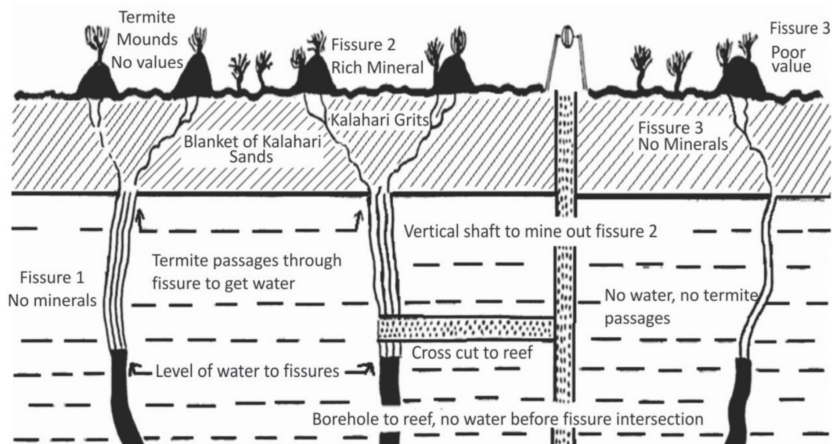
Figure 4. Schematic section through Leopard Mines in present-day Zimbabwe. Termite mounds built at surface from where the termite colony burrow through Kalahari sands to access basement fractures leading to water- bearing fissures below with water table at 27 m depth. On their way back, they pick gold grains from fissure 2 where a horizontal reef lies at 23 m depth to deposit in their mound. Figure based on West (1970)

Landscape Scale

Investigation at this scale of observation is sparse and is mainly limited to mound density estimation, evaluation of relationships between mounds and woody tree or plant vegetation distribution, mound distribution across mean annual precipitation changes and across broad geology. Termites strategically site their mounds where nutrients such as food and water are readily available. It has been observed of termites to avoid lowland environments and erect their structures on hill crests where building material is limited but where water and perhaps woody food can easily be sourced without risk of inundation. Levick et al. (2010) and Davies et al. (2014b) described mounds in Kruger National Park of South Africa as lining above seep-lines on hill crests, offering better drainage obviously because of the sandy nature of the soil. The seep-line could be the colony's source of permanent groundwater (groundwater outcrop) and because woody vegetation is higher on the hill crest, making it an idle habitat. The dependence of termites on groundwater was also stressed following their relationship with aquiferous dykes in North-western Ethiopian lowlands (Mege & Rango, 2009). Termite mounds were strategically constructed in linear pattern on outcropped, fracture dense aquiferous sections of the dykes with a remarkable regular spacing between mounds and were completely absent on surrounding basaltic flow country rock, which has similar mineralogy as the dyke. Although the termite species was not identified, it is clear evidence that mound locations serve as pointers

to identifying water-bearing sections of the dykes and that groundwater is chiefly among the factors governing termite mound distribution. A similar observation has been made in central Nigeria, where termite mounds are assembled in a linear fashion though without immediate evidence of a dyke or any linear structure (Ahmed II pers. obs.), but their strong relationship with Mangifera indica and Azadirachta indica tree species with large trunks known to have tap-root systems might just be an indication of what lies beneath. On the land surface, termite mounds associated with trees also influence the preferential flow of runoff towards their course (Bargues-Tobella et al., 2014), but what remains poorly understood is the preference of termites for regions of lower annual rainfall as demonstrated by Davies et al. (2014b). Higher annual rainfall would translate to more water available for underground storage but probably for security of their nests from degradation by rain, termites would prefer lower rainfall regions.

Table 2

Direct and indirect evidence of termite mound impact on groundwater

Author(s)	Country	Scale of observation	Method(s)	Comments
Ackerman et al., 2007	Brazil	Aggregate and profile	Field survey and experiment, laboratory tests	Improve infiltration, low soil water retention on termite mounds show promise for groundwater recharge
Bargués-Tobella et al., 2014	Burkina Faso	Landscape	Field experiment	Surface water flow preferentially around trees associated with termite mounds, increasing chances of infiltration
Davies et al., 2014b	South Africa	Landscape	Airborne LIDAR survey	Geology, drainage, wood cover accounted for termite mound patterning and distribution, which is a reflection of soil texture and permeability.
Davies et al., 2016	South Africa	Landscape	Airborne LIDAR survey	Tree species associated with lowlands had increased density near termite mounds at hill crest. Lowlands and valleys are suitable sites for groundwater exploration.
Dowuona et al., 2012	Ghana	Aggregate and profile	Field sampling and laboratory analysis	Physical and hydraulic properties to support runoff interception are poor but wells in the study area are known to be very productive.
Jouquet et al., 2015a	India	Aggregate	Field sampling and laboratory analysis	Cracks on termite mounds developed due to shrinking in the dry season can allow rainwater to directly penetrate into the mound.

Table 2 (*continue*)

Author(s)	Country	Scale of observation	Method(s)	Comments
Kalumanga et al., 2016	Tanzania	Aggregate and profile	Field survey and sampling	Termite mound distribution is controlled by surface and groundwater availability and not by clay or woody vegetation.
Leonard & Rajot, 2001	Niger	Profile	Field experiment and simulation	Macropores formed by termites increased infiltration rates in both ponding and rainfall simulation experiments.
Levick et al., 2010	South Africa	Landscape	Airborne LIDAR survey and clay content analysis	Termite mounds are built above seep-lines on well-drained sites where clay content is average.
Mege & Rango, 2009	Ethiopia	Landscape	Field survey	Groundwater storage in dolerite dyke controls the distribution of termite mounds.
Moura et al., 2014	Brazil	Aggregate and profile	Field and laboratory experiments	Enhanced mound soil porosity and infiltration as well as reduced bulk density hold promise for groundwater recharge.
Mujinya et al., 2013	D. R. Congo	Profile	Field investigation	There is evidence of a seasonally high groundwater table from the mottling of Mn-Fe oxides below the termite mounds.
Yamashina, 2010	Namibia	Landscape	Field survey	Termite mounds are distributed along seasonal streams, lowlands and areas with sharp changes in inclination (probably discontinuities) on mountainous regions.
Watson, 1972	Zimbabwe	Profile	Field and underground mines survey	Termite species *O. latericus* builds mounds directly above basement fissures where the termites burrow to access water at 27 m depth.
West, 1965	Zimbabwe	Profile	Field and underground mines survey	Reconstruction of mounds during dry season is evidence of termites' burrowing further into the ground, trying to reach the groundwater table that has dropped in response to the season.

CONCLUDING REMARKS AND SUGGESTIONS FOR FURTHER STUDY

As surface water is becoming scarce due to climate change causing unusual drying up of rivers and streams on which most of the rural populace rely for their supply, groundwater development becomes a favourable alternative. It is known that groundwater is drought reliable and available in almost all geologic formations. However, there is less than 1% chance of hitting water-bearing formations should wells be sited without adequate prospecting (West, 1965). Our quest for employing only sophisticated equipment such as geophysical tools in groundwater prospecting with enormous cost and labour attached has limited our power of observation. Identifying termite mound locations could mean that the arduous part of the prospecting job has been done at zero cost. Termites' reliance on water for their metabolism and avoidance of inundation from surface water somewhat constrain their habitation to areas with groundwater reserves that can last through the long dry season. Whether these reserves are permanent in all cases is a subject of further research. Furthermore, the current practices in groundwater prospecting entail the analysis of lineaments from remotely sensed images. Lineaments are linear to curvilinear lines of weaknesses on the earth surface through which runoff are thought to recharge aquifers. Many studies have employed this method to zone prospective groundwater-potential areas with impressive success rates (Solomon & Quiel, 2006; Jasrotia et al., 2013; Rahmati et al., 2014; Pinto et al., 2015). A major limitation with this method lies in its inability to distinguish between open lineaments (or water-bearing lineaments) from closed ones; hence, only zones with high lineament densities are assigned greater prospects. This limitation can be addressed when termite mound locations are integrated into such data to show only those fractures that harbour groundwater. Likewise, high topographic areas are usually characterised as poor in terms of groundwater prospect but the distribution of termite mounds on hill crest along seep-lines in Kruger National Park should counter this understanding. For future studies, the use of invasive electrical surveys on mounds to reveal the extent of regolith, underlying rock formation and its structural disposition beneath mounds can be explored. Across broad geology, distinction on termite mound distribution has only been made between granitic and basaltic rock covers; how this pattern varies across other rock types is not known. Finally, the use of the Geographic Information System (GIS) to integrate data from the three scales of observation discussed can prove very effective in further shaping our understanding of the relationship that exists between termite mounds and groundwater and other factors that control their distribution.

REFERENCES

Abdel-Magid, A. H., Schuman, G. E., & Hart, R. H. (1987). Soil bulk density and water infiltration as affected by grazing systems. *Journal of Range Management*, *40*(4), 307–309.

Ackerman, I., Lehmann, J., Fernandes, E. C. M., Ackerman, I. L., Teixeira, W. G., & Riha, S. J. (2007). The impact of mound-building termites on surface soil properties in a secondary forest of Central Amazonia. *Applied Soil Ecology*, *37*(2007), 267–276.

Adelana, S. M. A., Olasehinde, P. I., Bale, R. B., Vrbka, P., Edet, A. E., & Goni, I. B. (2008). An overview of the geology and hydrogeology of Nigeria. In S. M. A Adelana & A. M. MacDonald (Eds.), *Applied groundwater studies in Africa* (pp. 171–179). London: Taylor & Francis.

Adhikary, N., Erens, H., Weemaels, L., Deweer, E., Mees, F., Mujinya, B. B. … & Van Ranst, E. (2016). Effects of spreading out termite mound material on ferralsol fertility, Katanga, D. R. Congo. *Communications in Soil Science and Plant Analysis*, *47*(9), 1089–1100.

Akankpo, A., & Igboekwe, M. (2012). Geographic information system in mapping of groundwater quality for Michael Okpara University of Agriculture Umudike and its Environs, Southeastern Nigeria. *Archives of Applied Science Research*, *4*(3), 1483–1493.

Al-abadi, A., & Al-Shamma, A. (2014). Groundwater potential mapping of the major aquifer in northeastern Missan Governorate, south of Iraq by using analytical hierarchy process and GIS. *Journal of Environment and Earth Science*, *4*(10), 125–150.

Arhin, E., Boadi, S., & Esoah, M. C. (2015). Identifying pathfinder elements from termite mound samples for gold exploration in regolith complex terrain of the Lawra belt, NW Ghana. *Journal of African Earth Sciences*, *109*, 143–153.

Arhin, E., & Nude, P. M. (2010). Use of termitaria in surficial geochemical surveys: Evidence for >125μm size fractions as the appropriate media for gold exploration in northern Ghana. *Geochemistry, Exploration, Environment, Analysis, 10*(4), 401–406.

Bala, A. E., Batelaan, O., & Smedt, F. (1999). Using Landsat 5 imagery in the assessment of groundwater resources in crystalline rocks around Dutsin-Ma, Northwest Nigeria. *Journal of Mining Geology, 36*(1), 85–92.

Bargués Tobella, A., Reese, H., Almaw, A., Bayala, J., Malmer, A., Laudon, H., & Ilstedt, U. (2014). The effect of trees on preferential flow and soil infiltrability in an agroforestry parkland in semiarid Burkina Faso. *Water Resources Research*, *50*(4), 3342–3354.

Bonachela, J. A., Pringle, R. M., Sheffer, E., Coverdale, T. C., Guyton, J. A., Caylor, K. K. … & Tarnita, C. E. (2015). Termite mounds can increase the robustness of dryland ecosystems to climatic change. *Science, 347*(6222), 651–655.

Bouillon, A. (1970). Termites of the Ethiopian region. In K. Krishna, & F. M. Weesner (Eds.), *Biology of termites* (pp. 153–280). New York: Academic Press.

Bottinelli, N., Jouquet, P., Grimaldi, M., & Peng, X. (2014). Why is the influence of soil macrofauna on soil structure only considered by soil ecologists? *Soil and Tillage Research*, *146*, 118–124.

Brauman, A. (2000). Effect of gut transit and mound deposit on soil organic matter transformations in the soil feeding termite: A review. *European Journal of Soil Biology*, *36*(3–4), 117–125.

Choosai, C., Mathieu, J., Hanboonsong, Y., & Jouquet, P. (2009). Termite mounds and dykes are biodiversity refuges in paddy fields in north-eastern Thailand. *Environmental Conservation*, *36*(1), 71–79.

Corgne, S., Magagi, R., Yergeau, M., & Sylla, D. (2010). An integrated approach to hydro-geological lineament mapping of a semi-arid region of West Africa using Radarsat-1 and GIS. *Remote Sensing of Environment*, *114*(9), 1863–1875.

Dangerfield, J. M., Mccarthy, T. S., & Ellery, W. N. (1998). The mound-building termite Macrotermes michaelseni as an ecosystem engineer. *Journal of Tropical Ecology*, *14*(4), 507–520.

Davies, A. B., Robertson, M. P., Levick, S. R., Asner, G. P., Van Rensburg, B. J., & Parr, C. L. (2014a). Variable effects of termite mounds on African savanna grass communities across a rainfall gradient. *Journal of Vegetation Science*, *25*(6), 1405–1416.

Davies, A. B., Levick, S. R., Asner, G. P., Robertson, M. P., Van Rensburg, B. J. & Parr, C. L. (2014b). Spatial variability and abiotic determinants of termite mounds throughout a savanna catchment. *Ecography*, *37*(9), 001–011.

Davies, A. B., Baldeck, C. A., & Asner, G. P. (2016). Termite mounds alter the spatial distribution of African savanna tree species. *Journal of Biogeography*, *43*(2), 301–313.

Denovan, S. E., Eggleton, P., Dubbin, W. E., Batchelder, M. & Dibong, L. (2001). The effects of a soil-feeding termite, Cubitermes fungifaber (Isoptera: Termitidae) on soil properties: Termites may be an important source of soil microhabitat heterogeneity in tropical forests. *Pedobiology, 45*(1), 1–11.

Dowuona, G. N. N., Atwere, P., Dubbin, W., Nude, P. M., Mutala, B. E., Nartey, E. K., & Heck, R. J. (2012). Characteristics of termite mounds and associated Acrisols in the coastal savanna zone of Ghana and impact on hydraulic conductivity. *Natural Science*, *4*(7), 423–437.

Edet, A. E., Okereke, C. S., Teme, S. C., & Esu, E. O. (1998). Application of remote-sensing data to groundwater exploration: A case study of the Cross River State, southeastern Nigeria. *Hydrogeology Journal*, *6*(3), 394–404.

Ekundayo, E. O., & Aghatise, V. O. (1997). Soil properties of termite mounds under different land use types in a Typic Paleudult of Midwestern Nigeria. *Environmental Monitoring and Assessment*, *45*(1), 1–7.

El-Baz, F. (2008). Remote sensing of the earth: Implications for groundwater in Darfur. *The Bridge, 38*(3), 5–13.

Food and Agriculture Organisation. (2005). The importance of soil organic matter. *FAO Soils Bulletin No. 80*. Rome, Italy.

Fashae, O. A., Tijani, M. N., Talabi, A. O., & Adedeji, O. I. (2013). Delineation of groundwater potential zones in the crystalline basement terrain of SW-Nigeria: An integrated GIS and remote sensing approach. *Applied Water Science*, *4*(1), 19–38.

Fenta, A. A., Kifle, A., Gebreyohannes, T., & Hailu, G. (2014). Spatial analysis of groundwater potential using remote sensing and GIS-based multi-criteria evaluation in Raya Valley, northern Ethiopia. *Hydrogeology Journal*, *23*(1), 195–206.

Ferriz, H., & Bizuneh, G. (2002). Development and management of water resources. In *Proceedings of Ethioforum* (pp. 361–391). Addis Ababa.

Francis, M. L., Ellis, F., Lambrechts, J. J. N., & Poch, R. M. (2012). A micromorphological view through a Namaqualand termitaria (Heuweltjie, a Mima-like mound). *Catena, 100*(2013), 57–73.

Grohmann, C., Oldeland, J., Stoyan, D., & Linsenmair, K. E. (2010). Multi-scale pattern analysis of a mound-building termite species. *Insectes Sociaux*, *57*(4), 477–486.

Haghnazari, F., Shahgholi, H., & Feizi, M. (2015). Factors affecting the infiltration of agricultural soils: REVIEW. *International Journal of Agronomy and Agricultural Research, 6*(5), 21–35.

Jamali, H., Livesley, S. J., Hutley, L. B., Fest, B., & Arndt, S. K. (2013). The relationships between termite mound CH4/CO2 emissions and internal concentration ratios are species specific. *Biogeosciences*, *10*(4), 2229–2240.

Jasrotia, A. S., Bhagat, B. D., Kumar, A., & Kumar, R. (2013). Remote sensing and GIS Approach for delineation of groundwater potential and groundwater quality zones of Western Doon Valley, Uttarakhand, India. *Journal of the Indian Society of Remote Sensing*, *41*(2), 365–377.

Jha, M. K., Chowdhury, A., Chowdary, V. M., & Peiffer, S. (2007). Groundwater management and development by integrated remote sensing and geographic information systems: Prospects and constraints. *Water Resources Management, 21*(2), 427–467.

Jouquet, P., Tessier, D., & Lepage, M. (2004). The soil structural stability of termite nests: Role of clays in Macrotermes bellicosus (Isoptera, Macrotermitinae) mound soils. *European Journal of Soil Biology*, *40*(1), 23–29.

Jouquet, P., Barré, P., Lepage, M., & Velde, B. (2005). Impact of subterranean fungus-growing termites (Isoptera, Macrotermitiane) on chosen soil properties in a West African savanna. *Biology and Fertility of Soils*, *41*(5), 365–370.

Jouquet, P., Dauber, J., Lagerlöf, J., Lavelle, P., & Lepage, M. (2006). Soil invertebrates as ecosystem engineers: Intended and accidental effects on soil and feedback loops. *Applied Soil Ecology*, *32*(2), 153–164.

Jouquet, P., Traoré, S., Choosai, C., Hartmann, C., & Bignell, D. (2011). Influence of termites on ecosystem functioning. Ecosystem services provided by termites. *European Journal of Soil Biology, 47*(4), 215–222.

Jouquet, P., Guilleux, N., Caner, L., Chintakunta, S., Ameline, M., & Shanbhag, R. R. (2015a). Influence of soil pedological properties on termite mound stability. *Geoderma*, *262*(2015) 45–51.

Jouquet, P., Guilleux, N., & Shanbhag, R. R. (2015b). Influence of soil type on the properties of termite mound nests in Southern India. *Applied Soil Ecology, 96*(2015) 282–287.

Jouquet, P., Bottinelli, N., Bourguignon, T., Jouquet, P., Bottinelli, N., Shanbhag, R. R., & Bourguignon, T. (2016a). Termites: The neglected soil engineers of tropical soils. *Soil Science, 181*(3/4), 157–165.

Jouquet, P., Airola, E., Guilleux, N., Harit, A., Chaudhary, E., Grellier, S., & Riotte, J. (2016b). Abundance and impact on soil properties of cathedral and lenticular termite mounds in Southern Indian woodlands. *Ecosystems, 20*(4), 1–12.

Kalumanga, E., Mpanduji, D. G., & Cousins, S. A. O. (2016). Geophagic termite mounds as one of the resources for African elephants in Ugalla Game Reserve, Western Tanzania. *African Journal of Ecology, 55*(1), 1–10.

Kandasami, R. K., Murthy, T. G., & Borges, R. M. (2016). Effect of biocementation on the strength and stability of termite mounds. *Environmental Geotechnics*, *3*(2016), 99–113.

Kebede, F. (2004). Use of termite mounds in geochemical exploration in North Ethiopia. *Journal of African Earth Science, 40*(1-2), 101–103.

Korb, J., & Linsenmair, K. E. (1998a). The effects of temperature on the architecture and distribution of Macrotermes bellicosus (Isoptera, Macrotermitinae) mounds in different habitats of a West African Guinea savanna. *Insectes Sociaux, 45*(1), 51–65.

Korb, J., & Linsenmair, K. E. (1998b). Experimental heating of Macrotermes bellicosus (Isoptera, Macrotermitinae) mounds: What does microclimate play in influencing mound architecture? *Insectes Sociaux, 45*(3), 355–342.

Korb, J., & Linsenmair, K. E. (2000). Ventilation of termite mounds: New results require a new model. *Behavioral Ecology*, *11*(5), 486–494.

Krishna, K., Grimaldi, D. A., Krishna, V., & Engel, M. S. (2013). Treatise on the Isoptera of the world. *Bulletin of the American Museum of National History, 377*(4), 1–200.

Kura, N. U., Ramli, M. F., Ibrahim, S., Sulaiman, W. N. A., Aris, A. Z., Tanko, A. I., & Zaudi, M. A. (2014). Assessment of groundwater vulnerability to anthropogenic pollution and seawater intrusion in a small tropical island using index-based methods. *Environmental Science and Pollution Research*, *22*(2), 1512–1533.

Léonard, J., & Rajot, J. L. (2001). Influence of termites on runoff and infiltration: Quantification and analysis. *Geoderma*, *104*(1–2), 17–40.

Levick, S. R., Asner, G. P., Chadwick, O. A., Khomo, L. M., Rogers, K. H., Hartshorn, A. S. … & Knapp, D. E. (2010). Regional insight into savanna hydrogeomorphology from termite mounds. *Nature Communications*, *1*(6), 65-71.

Malaka, S. L. O. (1977). A note on the bulk density of termite mounds. *Austrialian Journal of Soil Resources. 15*(1), 93–94.

Manap, M. A., Sulaiman, W. N. A., Ramli, M. F., Pradhan, B., & Surip, N. (2013). A knowledge-driven GIS modeling technique for groundwater potential mapping at the Upper Langat Basin, Malaysia. *Arabian Journal of Geosciences*, *6*(5), 1621–1637.

Mando, A., Stroosnijder, L., & Brussaard, L. (1996). Effects of termites on infiltration into crusted soil. *Geoderma, 74*(1-2), 107–113.

Mège, D., & Rango, T. (2009). Permanent groundwater storage in basaltic dyke fractures and termite mound viability. *Journal of African Earth Sciences*, *57*(1–2), 127–142.

Moura, E. G., Aguiar, A., Das, C. F., Piedade, A. R., & Rousseau, G. X. (2014). Contribution of legume tree residues and macrofauna to the improvement of abiotic soil properties in the eastern Amazon. *Applied Soil Ecology*, *86*(2014), 91–99.

Mujinya, B. B., Van Ranst, E., Verdoodt, A., Baert, G., & Ngongo, L. M. (2010). Termite bioturbation effects on electro-chemical properties of ferralsols in the Upper Katanga (D. R. Congo). *Geoderma*, *158*(3–4), 233–241.

Mujinya, B. B., Mees, F., Boeckx, P., Bodé, S., Baert, G., Erens, H. … & Van Ranst, E. (2011). The origin of carbonates in termite mounds of the Lubumbashi area, D. R. Congo. *Geoderma*, *165*(1), 95–105.

Mujinya, B. B., Mees, F., Erens, H., Dumon, M., Baert, G., Boeckx, P. … & Van Ranst, E. (2013). Clay composition and properties in termite mounds of the Lubumbashi area, D. R. Congo. *Geoderma, 192*, 304–315.

Mujinya, B. B., Adam, M., Mees, F., Bogaert, J., Vranken, I., Erens, H. … & Van Ranst, E. (2014). Spatial patterns and morphology of termite (Macrotermes falciger) mounds in the Upper Katanga, D. R. Congo. *Catena*, *114*(2014), 97–106.

Nauer, P. A., Hutley, L. B., Bristow, M., & Arndt, S. K. (2015). Are termite mounds biofilters for methane? Challenges and new approaches to quantify methane oxidation in termite mounds. In *Proceedings of EGU General Assembly.* Vienna, Austria.

Ndiaye, D., Lapage, M., Sall, C. E., & Brauman, A. (2004). Nitrogen transformations associated with termite biogenic structures in a dry savanna ecosystem. *Plant Soil, 265*(1-2), 189–196.

Oh, H. J., Kim, Y. S., Choi, J. K., Park, E., & Lee, S. (2011). GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *Journal of Hydrology*, *399*(3-4), 158–172.

Pardeshi, M., & Prusty, B. A. K. (2010). Termites as ecosystem engineers and potentials for soil restoration. *Current Science*, *99*(1), 11.

Park, I., Kim, Y., & Lee, S. (2014). Groundwater productivity potential mapping using evidential belief function. *Ground Water*, *52*(S1), 201–207.

Pinto, D., Shrestha, S., Babel, M. S., & Ninsawat, S. (2015). Delineation of groundwater potential zones in the Comoro watershed, Timor Leste using GIS, remote sensing and analytic hierarchy process (AHP) technique. *Applied Water Science*, *7*(1), 503–519.

Plummer, C. C., Carlson, D. H., & Hammersley, L. (2010). *Physical Geology* (13th ed., pp. 284–301). New York: McGraw-Hill Publishers.

Pradhan, B., 2009. Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. *Central European Journal of Geosciences, 1*(1), 120–129.

Rahmati, O., Nazari Samani, A., Mahdavi, M., Pourghasemi, H. R., & Zeinivand, H. (2014). Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arabian Journal of Geosciences, 8*(9), 7059–7071.

Rajeev, V., & Sanjeev, A. (2011). Impact of termite activity and its effect on soil composition. *Tanzanian Journal of Natural and Applied Sciences 2*(2), 399–404.

Roose-Amsaleg, C., Mora, P., & Harry, M. (2005). Physical, chemical and phosphatase activities characteristics in soil-feeding termite nests and tropical rainforest soils. *Soil Biology and Biochemistry*, *37*(10), 1910–1917.

Sako, A., Mills, A. J., & Roychoudhury, A. N. (2009). Rare earth and trace element geochemistry of termite mounds in central and northeastern Namibia: Mechanisms for micro-nutrient accumulation. *Geoderma*, *153*(1–2), 217–230.

Sarcinelli, T. S., Schaefer, C. E. G. R., de Souza Lynch, L., Arato, H. D., Viana, J. H. M., Filho, M. R. de A., & Gonçalves, T. T. (2009). Chemical, physical and micromorphological properties of termite mounds and adjacent soils along a toposequence in Zona da Mata, Minas Gerais State, Brazil. *Catena*, *76*(2), 107–113.

Schuurman, G., & Dangerfield, J. M. (1997). Dispersion and abundance of Macrotermes michaelseni colonies: A limited role for intraspecific competition. *Journal of Tropical Ecology*, *13*(1), 39–49.

Seymour, C. L., Joseph, G. S., Makumbe, M., Cumming, G. S., Mahlangu, Z., & Cumming, D. H. M. (2016). Woody species composition in an African savanna: Determined by centuries of termite activity but modulated by 50 years of ungulate herbivory. *Journal of Vegetation Science*, *27*(4), 824–833.

Solomon, S., & Quiel, F. (2006). Groundwater study using remote sensing and geographic information systems (GIS) in the central highlands of Eritrea. *Hydrogeology Journal*, *14*(5), 729–741.

Tilahun, A., Kebede, F., Yamoah, C., Erens, H., Mujinya, B. B., Verdoodt, A., & Van Ranst, E. (2012). Quantifying the masses of Macrotermes subhyalinus mounds and evaluating their use as a soil amendment. *Agriculture, Ecosystems and Environment*, *157*, 54–59.

Tijani, M., Crane, E., Upton, K., & O'Dochartaigh, B. E. (2016). *Africa groundwater atlas: Hydrogeology of Nigeria*. British Geological Survey. Retrieved April 16, 2016, from http://earthwise.bgs.ac.uk/index.php/Hydrogeology_of_Nigeria

Turner, J. S. (2000). Architecture and morphogenesis in the mound of Macrotermes michaelseni (Sjostedt) (Isoptera: Termitidae, Macrotermitinae) in northern Namibia. *Cimbebasi*, *16*, 143–175.

Turner, J. S. (2006). Termites as mediators of the water eonomy of arid savanna ecosystems. *Dryland Ecohydrology* (pp. 303-313). Springer, Dordrecht.

United Nations. (1997). Comprehensive assessment of the freshwater resources of the World. *World Meteorology Organisation*. Geneva, Switzerland.

United Nations Environmental Program. (2000). Termite biology and ecology. *Division of Technology, Industry and Economics Chemicals Branch,* 1–5.

United States Department of Agriculture. (1998). Inherent factors affecting bulk density and available water capacity. *Soil quality kit-Guides for educators*. Iowa.

Watson, J. P. (1972). The distribution of gold in termite mounds and soils at a gold anomaly in Kalahari sand. *Soil Science*, *113*(5), 317–321.

West, W. F. (1965). Some unconventional ideas on prospecting. *Chamber of Mines Journal*, 7, 40–42.

West, W. F. (1970). The Bulawayo symposium papers: No. 2 termite prospecting. *Chamber of Mines Journal, 12*(10), 32–35.

Yamashina, C. (2010). Interactions between termite mounds, trees, and the Zemba people in the Mopane savanna in north-western Namibia. *African Study Monographs, 40*, 115–128.

SCIENCE & TECHNOLOGY

Review Article

Use of Waste Materials in Concrete: A review

Davoud Tavakoli[1]*[#], Masoumeh Hashempour[2] and Ali Heidari[2]

[1]Department of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran

[2]Department of Civil Engineering, Shahrekord University, Shahrekord, Iran

ABSTRACT

Millions of tons of waste is produced in the world each year and most of it is not recyclable. Furthermore, recycling waste consumes energy and produces pollution. In addition, accumulation of waste in the suburbs and the disposal of waste are very dangerous for the environment. Using waste material in concrete production is an appropriate method for achieving two goals: eliminating waste and adding positive properties in concrete. Since the green concrete industry is expanding, it is necessary to evaluate concrete that contains waste from all aspects in order to determine its capability. This literature study consists of two parts i.e. the use of waste as a substitute for cement and as a substitute for aggregates. Leading waste material that has been used as substitutes is highlighted and the characteristics of the resulting concrete is evaluated. Among other findings, rubber was found to have improved fire resistance and ductility in concrete and agricultural and PET wastes were successfully used in non-structural concrete, while glass helped to improve thermal stability.

Keywords: Concrete, environment, sustainable development, waste materials

INTRODUCTION

Concrete, one of the most important construction materials in the construction of infrastructure and development facilities, has the potential for significant and positive environmental participation (Tavakoli et al., 2012). Waste material in concrete can be used as cement or aggregate replacement, fillers or fibres. As cement is a dangerous pollutant of the environment, waste material can be used as a substitute for cement as well as for aggregates. The environmental advantages of using waste material as a replacement for cement can be investigated in two ways. One is the removal

of a part of the cement from concrete and the other is the use of waste material that is useless in concrete. Due to the volume of cement consumption around the world, a lot of waste can be used as a replacement for concrete. From the standpoint of reducing cement, there are many benefits attached to the use of pozzolans, including the reduction of greenhouse gasses, the most hazardous of which are carbon dioxide and nitrogen oxides.

The consumption of cement is 60 million tons per year in Iran (Tavakoli et al., 2012); if 5% of the concrete projects in our country used 10 to 15% of waste material to replace cement, in one year a large amount of the waste generated in Iran can be reused. Also, if this consumption of cement continued and extended to 10% of the country's projects, the total waste from past years can be completely consumed by reusing it in concrete in a few years. This will save on consumption of the country's other resources and at the same time, reduce environmental pollution in the country. Therefore, the use of waste material as a substitute for cement is beneficial for both reducing cement consumption and waste consumption. As it is impossible to use only waste material instead of cement in construction, waste material can be reused as an aggregate. Aggregates occupy about 70% of the volume of concrete, thus a large amount of it can be reused. To avoid consumption of raw materials that are already scarce, this option is crucial, especially for European countries that are facing mineral deficiency.

The survey investigates aggregate replacement and cement replacement. It also discusses the advantages and disadvantages of each.

WASTE MATERIAL AS AGGREGATE REPLACEMENT

Glass

Glass is a colourless, transparent, hard and fragile material, with a hardness of 6.5. According to ASTM, glass is an inorganic material. In its molten state, as glass gets colder, it becomes more rigid without crystallisation. Silica is a fundamental constituent of glass, but in most common glasses MgO, CaO, Al_2O_3 and Na_2O are also found. Studies have shown that it is possible to use glass in concrete in three forms: as Coarse Glass Aggregate (CGA), Fine Glass Aggregate (FGA) and Glass Powder (GP).

When mixed with cement, glass undergoes a chemical reaction that produces a secondary hydrated calcium silicate (C-S-H) and a pozzolanic reaction with cement hydrates (Islam et al., 2017). Meyer and Baxter (1997, 1998), pioneers in this field of research, tried to prove the practicality of concrete production containing 100% glass aggregate and 20% metakaolin. The studies showed that an increase in the amount of glass waste reduced not only the specific weight of concrete but also the compressive strength of the concrete due to reduced adhesion with it (Topcu & Canbaz, 2004). This type of concrete has high thermal stability due to high thermal conductivity in comparison with conventional aggregate concrete; thus, it can be used for buildings that require thermal stability. This type of concrete is ideal for buildings in cold, mountainous areas (Poutos et al., 2006). Cazacliu et al. (2010) and, Ling and Poon (2012), showed that the use of glass in concrete is practical, but its size can influence the effectiveness of concrete. Reduction of particle size improves workability but reduces the 28-day compressive strength of cement. The combination of both fine and coarse glass can improve water absorption and bring the shrinkage of concrete to its lowest value (de Castro & de Brito, 2013).

Due to the fact that glass aggregates are composed of a high percentage of silica, it is possible that the alkaline reaction of the cement causes a faster reaction between alkali and silica in the aggregate. However, it is possible to prevent this by controlling the exact percentage of the cement used. In 2015, Cota's survey focused on three parameters: impact of the particle size of the glass, the impact of the percentage of the glass and the amount of metakaolin added.

The results indicated that the use of metakaolin and glass particles can cause specific weight loss, especially when the glass is replaced with finer particles. As a result, an increase of 79% was observed in comparison with control samples when substituting with fine quartz particles; however, the dynamic modulus of concrete increases when replaced with larger particles and also, concrete containing coarser-grain particles is more prone to reaction between alkali and silica (Cota et al., 2015). The studies showed that glass particles can cause ASR expansion and can also reduce compressive strength. If there are pores that provide enough space for ASR production, it may be possible to mitigate this reduction of strength. Adjusting the water-to-cement ratio and the amount of fine particles can provide this space. Figure 1 shows that the highest expansion is achieved by adding coarse glass particles, the use of 15% glass and non-use of metakaolin. As can be seen in the charts, the use of metakaolin reduces alkali-silica expansion. In another study, liquid crystal display (LCD) was used as fine aggregates, with 0%, 10%, 20% and 30% replacement by ordinary sand. The results showed that compressive strength and ultrasonic pulse velocity increased with curing time but decreased with an increase of w/b ratio (Wang & Wang, 2017).

Eventually, glass waste can produce not only suitable concrete for harsh climatic conditions due to its high thermal stability, but also concrete of appropriate strength. In general, glass can be used as an aggregate with optimum percentage in many cases, but it is necessary to control the ASR expansion. This is the main problem with using silica.
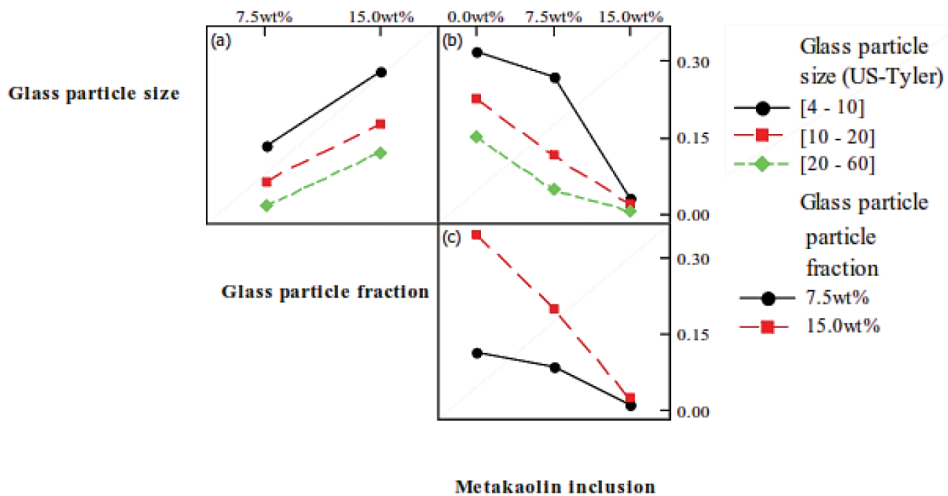
Figure 1. The third interaction effect related to the mean ASR expansion (Cota et al., 2015)

Poly(Ethylene Terephethalate) (PET)

PET, or Poly(Ethylene Terephethalate), belongs to the long-chain polymers of the polyester structure. The components of PET are pure terephthalic acid and ethylene glycol. Both are derived from petroleum products. The polyester processes involve other chemicals that can be made via polymerisation between an acid and an alkali. PET is an amorphous glass material. Using PET waste in concrete can save and preserve the environment. Yoon et al. (2005) analysed the microstructure of PET lightweight aggregates and then examined the effectiveness of the granular slag of molten metal (GBFS) on it. In these experiments, the density of the concrete containing PET aggregates increased from 1940 to 2260 kg/m^3, and the transition zone between PET particles and the cement paste in comparison with natural aggregates expanded more. Overall, it was expected that molten metal slag grain could boost the PET level and limit the transition zone leading to the reaction of calcium hydroxide. Other experiments showed that using PET in concrete can increase ductility and reduce shrinkage cracks (Sehaj et al., 2004; Won et al., 2009). Due to the lower specific density of PET compared with ordinary aggregates, light concrete with high quality can be produced using PET (Akcaozoglu, Atis, & Akcaozoglu, 2010). Fresh concrete containing PET has lower workability, density, modulus of elasticity and tensile strength than ordinary concrete. Increasing PET to 15% reduced the compressive strength of 15.9 % by up to 18% and reduced the modulus of elasticity of 20% by up to 23%. It also reduced the specific weight loss of 3.1% by up to 3.3%. Figure 2 shows the size and the type of PET particles in this study (Rahmani et al., 2013).
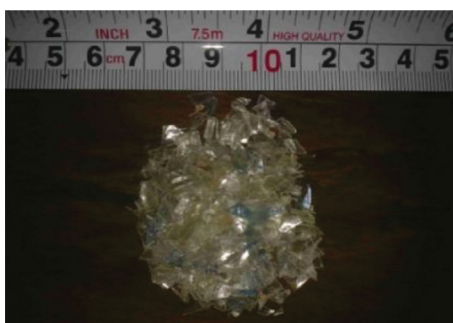
Figure 2. Size and type of PET particles (Rahmani et al., 2013)

Fresh concrete slump was also affected by the size and amount of PET particles, so the high amount of plastic-shaped aggregates reduces the slump of fresh concrete. Figure 3 shows that compressive strength was reduced by increasing the amount of PET aggregates, unlike the action of natural aggregates. PET aggregates cannot react with cement paste. In addition, the transfer zone in this type of concrete was weaker than in the control sample. PET aggregates reduced tensile strength, modulus of elasticity and bending strength. The physical structures of PET aggregates can also change the percentage of water to cement and the slump of concrete. The flaky aggregates of PET can connect two different parts, ensuring stability against the erosion of concrete containing PET aggregates (Saikia & De Brito, 2004). The ultrasonic wave speed of samples that contain PET with 0%, 5%, 10% and 15% reduced compressive strength

to 32.56%, 22.65%, 32.75% and 20.7%, respectively. Thus, samples containing 15% PET are integrated and dense. The reduction of compressive strength in these samples was 53.92%, 51.95%, 49.8% and 32.59, respectively. This was similar to the results obtained by Rahmani in 2013 (Araghi et al., 2015).
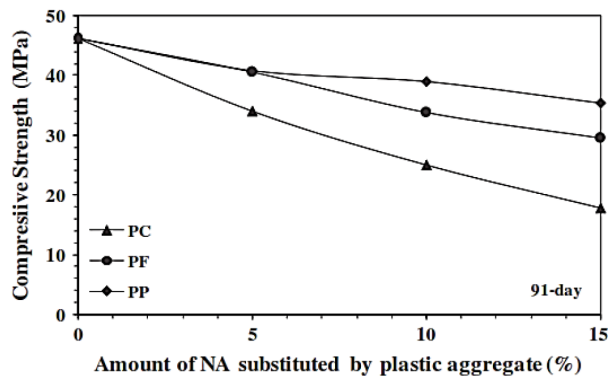
Figure 3. Compressive strength of concrete incorporation of PET aggregates (Saikia & De Brito, 2004)

As the results showed, using pet in concrete was practical and had no significant negative effects on the properties of concrete but for structural concretes that need rather high compressive strength this waste is not a good choice because it cannot react with cement and also, it has lower mechanical strength than ordinary aggregates. However, this kind of aggregate is a good for building lightweight concrete or corrosion-resistant concrete.

Tile and Sanitary Ceramics

A tile is a piece of artificial stone with thickness of a few millimetres and a glassy, soft and smooth surface on one side. Ceramic is a non-metallic and non-organic material. It is classified in two categories of crystalline and non-crystalline. Tile and ceramic waste is created during the transfer process, during or after burning, due to human error, manufacturing error or use of inappropriate material and much of it is due to the destruction of buildings (Tavakoli et al., 2012). Many studies have been done to dump this waste in concrete. The results of experiments showed that it would be feasible to use tile waste in concrete as pozzolan or aggregate (Ay & Unal, 2000; Portella et al., 2006). Using white ceramic aggregates as fine aggregate and substituted with ratios of 10% to 50%, the quality of concrete improved (Lopez et al., 2007). Moreover, if porcelain sanitary waste is used as coarse aggregate in concrete at a rate of 3% to 9%, its resistance is more than that of concrete without additives at a rate of 2% to 8% (Guerra et al., 2009). If the curing process takes a long time, about 28 days, and 15% to 20% of aggregates containing porcelain sanitary waste are used, then the concrete's resistance may be increased (Medina et al., 2012; Medina et al., 2012). Heidari et al. (2013) examined the effect of ceramic aggregates in concrete. For this purpose, ceramic was used as both coarse grain at a rate of 0% to 40% and sand at a rate of 0 to 100%. Figure 4 shows the size of the ceramic particles used in these studies.

Figure 4. Different sizes of ceramic tile waste (Tavakoli et al., 2013)

The results showed that the use of ceramics did not have a significant negative effect on the properties of concrete. The optimum sample of ceramic as an alternative to sand was about 25 to 50%. The best example of using ceramics as coarse grain used between 10% and 20% of it. Not only was there increase in compressive strength in this instance, there was a decrease in specific weight without a significant negative effect on water absorption. Table 1 and Table 2 show the summary of these results.

Table 1

Physical and mechanical properties of concrete mixes (Phase A) (Tavakoli et al., 2013)

Sample	Slump (mm)	Specific Weight (kg/m³)	Water Absorption (%)	Average Strength (MPa)	
				7 days	28 days
C	60	2441	5.05	26.9	33.1
CS25	55	2430	4.96	28.1	35.7
CS50	50	2382	4.79	27.2	35.1
CS75	40	2341	5.10	25.8	34.6
CS100	40	2294	5.30	24.1	33.7

Table 2

Physical and mechanical properties of concrete mixes (Phase B) (Tavakoli et al., 2013)

Sample	Slump (mm)	Specific Weight (kg/ m³)	Water Absorption (%)	Average Strength (MPa)	
				7 days	28 days
C	60	2441	5.05	26.9	33.1
CG10	50	2427	4.9	28.2	34.8
CG20	50	2407	5.2	27	34.3
CG30	45	2397	5.45	27.2	34.1
CG40	40	2385	5.7	25.7	33

Mechanical strength increased by addition of waste aggregates and it was higher than that of the control sample. In addition, maximum water penetration depth in the waste aggregates in the treated concrete was lower than in ordinary concrete; the amount used was less than 30 mm. Thus, the replacement of natural aggregates with waste aggregates can increase water penetration resistance (Medina et al., 2013). In order to create concrete with good performance, concrete with a mixture of 20% natural and ceramic aggregates can be used as it gives the same compressive strength as normal concrete with 100 MPa. The corrosion probability can be reduced by the 180th day of curing and by using 50% of the mixture aggregates (Gonzalez & Etxeberria, 2014). Therefore, it is possible to use tile waste in the form of fine and coarse aggregates. It should be noted that ceramic particles can also be porous or hard, in which case they are not only effective in the process of water absorption, but also in the elasticity of concrete (Anderson et al., 2016).

Generally, tile and ceramics with low specific weight and pozzolanic properties are a good choice for manufacturing concrete, but the results have shown that this material must be tested before concrete production and cannot be relied on to give the results of these preliminary studies because the type of burning of the tile and its constituents and even the type of mixing plan are factors that influence the behaviour of the concrete.

Clay Bricks

Fired bricks are burnt in a kiln and most of them contain silica, alumina, lime, iron oxide and magnesia. Because of this chemical structure, the use of bricks in concrete production seems to be practical. Researchers have suggested different mixtures to create this type of concrete. One study showed that the compressive strength of this concrete has a downward trend. This reduction was equal to 10% to 35% for coarse aggregates and 30% to 40% for fine aggregates (Akhtaruzzaman & Hasnat, 1983). Using clay bricks as sand in concrete increased water absorption and this may affect the durability of concrete parameters, so this subject needs more investigation (Tavakoli et al., 2014). In terms of durability, the concrete containing clay brick waste was no different from the control sample. Nevertheless, the brick aggregates had a negative effect on the durability of the reinforced concrete. A high amount of clay brick aggregate can reduce the corrosion time of bars, although this concrete has better performance in freezing and thawing. As the amount of bricks increase, the stability against the chloride ion penetration reduces. This reduction can be due to the higher absorption of the bricks because of their porosity. The 28-day compressive strength of concrete with brick coarse aggregates was slightly greater than that of the control sample and workability was also improved by increasing the amount of the brick coarse aggregates in the concrete (Adamson et al., 2015).
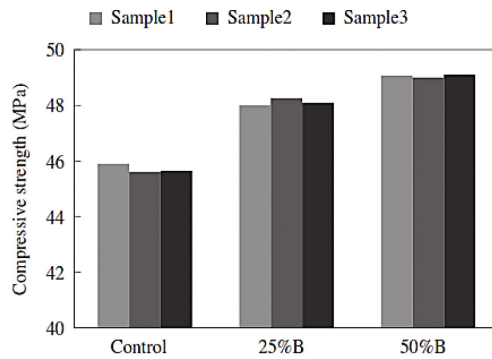
Figure 5. The 28-day compressive strength of the samples (Adamson et al., 2015)

In general, the results showed that using this type of waste was economical and practical. In addition there were no significant negative effects on the concrete. However, bricks are not suitable for use in reinforced concrete because they cause corrosion of bars.

Tyres and Rubber

Tyres are rubber pieces that are mounted on vehicle wheels. Tyres are made of natural rubber, styrene-butadiene, polybutadiene, carbon black and silica, which is used in high-performances tyres. The main idea of using this elastomeric material in cementitious matrix is to reduce the stiffness of concrete in order to make it more flexible and to improve its resistance to fire (Olivares & Barluenga, 2004). The use of tyre scrap as a substitute for aggregates and cement in concrete is new. Figure 6 shows grading tyre that can be used in concrete.
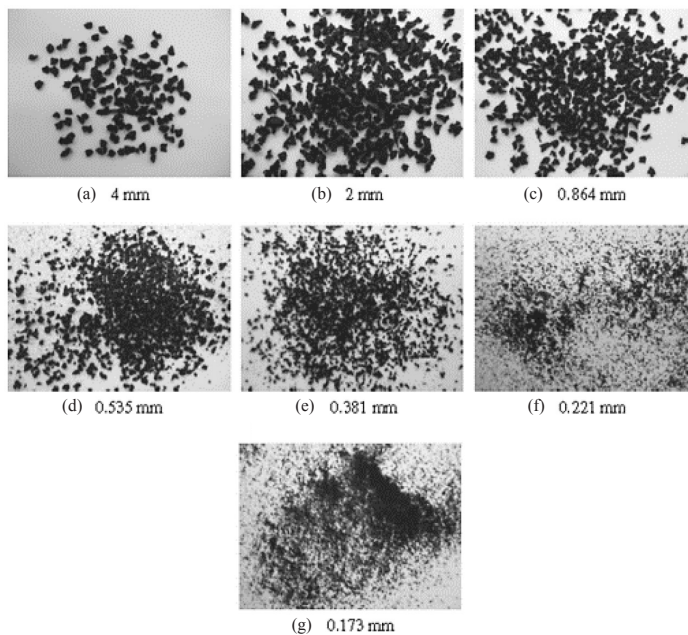
Figure 6. Various sizes of crumb rubber (Li et al., 2014)

Olivares and Barluenga (2004) used rubber waste as fibre. The results showed that by increasing the amount of rubber, the performance of concrete decreased. A flexural strength test showed that the concrete samples containing rubber tyre as fibre were stronger by up to 20% compared with the strength of the control samples. This could have been due to the conversion of the concrete to a more flexible material through the addition of rubber fibre. The control samples showed fractures caused by brittleness, and immediately split after cracking, while the samples containing plastic fibre became deformed but did not collapse (Yilmaz & Degirmenci, 2009). Sohrabi and Karbalai (2011) also showed that using silica fume with increased adhesion between the cement paste and rubber particles improved the filling of the pores and increased the compressive strength. The density of this type of concrete is also 13% less than that of the control samples (Pelisser et al., 2011). The increase of waste tyres can influence the carbonation depth, especially if the tyres are used as coarse-grained rubber (Bravo & Brito, 2012). Workability is also reduced by keeping the ratio of water to cement and increasing the amount of the ash of rubber. By increasing the percentage of rubber ash for water-to-cement ratios of 0.35 and 0.45, compressive strength was decreased. Figure 7 shows the 90-day compressive strength of cement with various ratios of water-to-cement and rubber ash (Gupta et al., 2014). In freezing-thawing resistance, the experiments showed that replacing or adding fine crumb rubber improved this property (Gesoğlu et al., 2014a, 2014b; Thomas & Gupta, 2016). It would seem that the use of tyre waste as an alternative in concrete still needs to be studied and further explored to determine the durability and strength of this material on concrete.
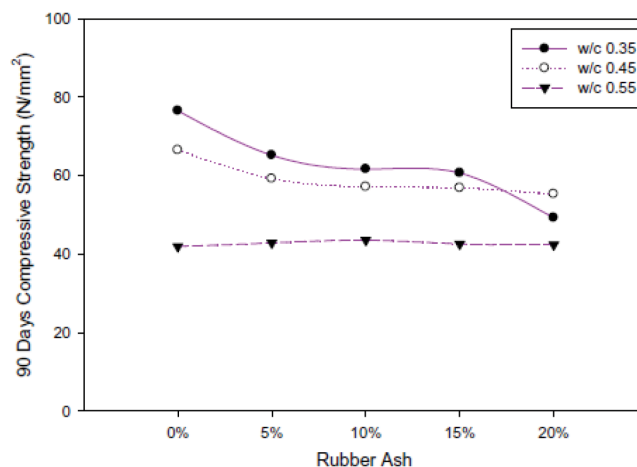
Figure 7. The 90-day compressive strength of rubber ash concrete (Gupta et al., 2014)

Metal

One ton of steel produces 17% of slag, which is used in depot sites. Due to the high production of metal products, a plan is required for its use. The major components of slag include the oxides of calcium, magnesium, silicon, iron and other metals, so it seems that the use of this waste is practicable in concrete production.

The first step in studying these materials was done by Akinmusuru in 1991 regarding the use of metal slag as an aggregate in the concrete (Akinmusuru, 1991). The initial studies proved that slag can be used in the manufacturing of non-structural concretes (Rai et al., 2002), but gradually it was shown that slag conversion to aggregate can be used as a substitute for coarse grain (Tay et al., 2003). In another study, industrial solid waste completely replaced the coarse aggregate and it was proved that this type of concrete had higher shear modulus and chemical stability in acidic and alkaline solutions than ordinary concrete (Ghailan, 2005). Later, the results of using slag in high performance concrete showed that compressive strength, water absorption and tensile strength of this type of concrete were higher than those of the control sample (Demirboğa & Gül, 2006). Slump decreases with the increasing metal waste and, as expected, density and bending strength increase compared to those of the control samples (Ismail & Al-Hashemi, 2008). Studies have shown the possibility of creating concrete with a compressive strength of higher than 150 MPa by copper slag when copper is used as a substitute for fine grain. With full replacement of standard sand and copper slag, it was concluded that the largest reduction in 28-day compressive strength was about 15 to 25%. Figure 8 shows the results on samples of 100 mm cubes. Mixture 1 is concrete without aggregate and steel fibre, mixture 2 is concrete without sand, mixture 3 is concrete without copper slag and steel fibre and mixture 4 is concrete without copper slag. It can be observed that the control sample immediately breaks at the end of the linear region. These studies also showed that the bending strength of fibre-reinforced concrete is approximately two times higher than that of concrete without fibre (Ambily et al., 2015). In a recent study about high-volume slag concrete the results showed that with an increase in slag, carbonation depth increases (Han-Seung & Wang, 2016).

In general, what follows from these results is that, due to the hardness and high density of the steel furnace slag compared to those of the natural aggregates, compressive strength and flexural strength of concrete increase. However, it should be noted that this waste can increase the weight of the concrete and may turn the concrete into a non-consumable material for the building industry.
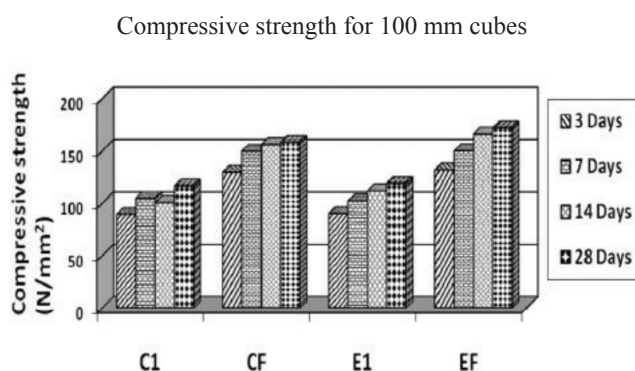
Figure 8. Compressive strength for 100 mm cubes (Ambily et al., 2015)

Concrete Waste

Research into the use of demolished concrete as aggregate for the new concrete production goes back to the end of World War II. Experiments on the use of concrete waste in the production of concrete began in 1993. Evaluation of different concrete mixtures containing fine-grain concrete waste concluded that the addition of a superplasticiser improved the concrete shrinkage (Merlet & Pimienta, 1993). Other experiments showed that drying shrinkage of concrete containing 30% of fine-grain concrete waste with a constant water-cement ratio and varying amounts of water reducer, is equal to the loss of natural aggregate concrete in 180 days (Zega & di Maio, 2011). It should be noted that, by adding fly ash as an alternative to cement, it is possible to reduce the shrinkage of concrete that is due to the addition of waste aggregate (Jeong, 2011). Some studies were done on the effects of increasing the coarse recycled concrete aggregates on mechanical characteristics and deformation caused by shrinkage (Cartuxo et al., 2015). In another study considering two samples with clay brick powder as cement and recycled concrete aggregates, mechanical properties were investigated. Results showed that clay brick powder compensated the decreasing of compressive strength due to the use of recycled aggregates because this powder can fill the porosity of concrete well (Letelier et al., 2017).

Research conducted from 1993 up to now indicates that concrete waste is reusable in concrete production. However, depending on the project, type and amounts should be specified precisely and it is noted that this waste is susceptible to carbonate reaction and may cause corrosion of the reinforcement, so the carbonation depth should be measured.

Agricultural Waste

Between 20 and 30% of agricultural production in the world becomes waste. The using of agricultural product waste has attracted researchers' attention to return the investment to the economic cycle. The most popular agricultural waste is almond and coconut shell.

Almond shell is not very common and it is more used in research to produce light weight concrete. A study by Siamardy and Vahedi (2008) was conducted by using almond shell as the coarse aggregate. Their research showed that this type of concrete had average performance slump, high air content and low density compared to those of ordinary concrete. Another study on coconut shell showed that this material had the capability to produce light concrete with good quality (Gunasekaran, 2008). In the long-term experiments, tests carried out in 365 days of concrete, this type of concrete also had good quality. Even the ultimate adhesion strength of this type of concrete is much higher than the theory of adhesion strength (Olanipekun et al., 2006).

Investigating the characteristics of concrete containing coarse grain coconut shell showed that this concrete had lower weight and its mechanical properties were equal to concrete with ordinary coarse aggregate. The long-term compressive strength of concrete containing coconut shell aggregate showed good quality and showed flexural behaviour comparable to that of the control sample (Gunasekaran et al., 2012, 2013; Gunasekaran et al., 2010; Gunasekaran et al., 2011a, 2011b). A recent study showed that 40% replacement of conventional aggregate with coconut shell could decrease the compressive strength of concrete by about 22% and to improve this property, reduction of the water-to-cement-ratio was necessary (Kanojia & Jain,

2017). Generally, it can be said that concrete made with coconut shell has higher compressive strength than concrete containing oil palm shell because of the roughness of coconut shell and its better adhesion with cement flakes (Shafigh et al., 2014). The microscopic surface of both materials can be seen in Figure 9 and Figure 10.
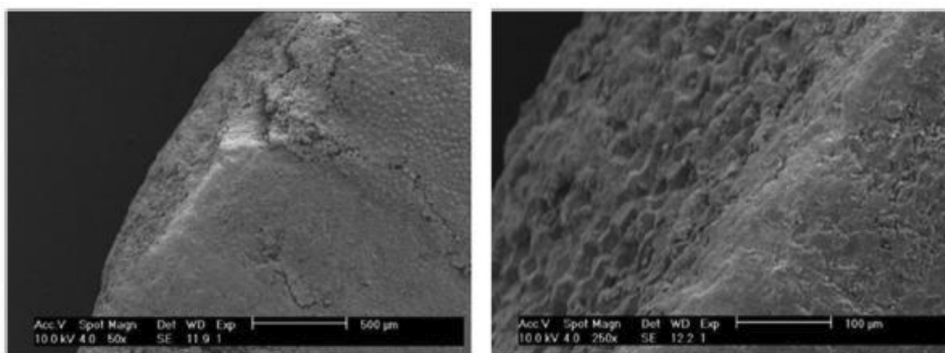
Figure 9. Microscopic images of the surface of an OPS grain in two scales (Shafigh et al., 2014)
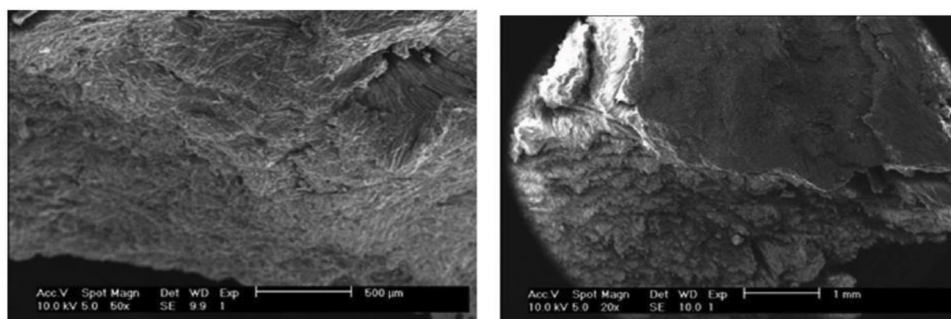
Figure 10. Microscopic images of the surface of coconut shell grain (Shafigh et al., 2014)

WASTE MATERIALS AS CEMENT REPLACEMENT

Silica Fume

Silica fume is a product of electrostatic capturing and tranquilising of silica dust with gasses discharged from electric arcs or alloys in the production process of silicon metal, particularly ferrosilicon alloys. This material has more than 80% non-crystalline silica with a diameter between 0.01 and 0.3 microns, which is about 50 to 100 times smaller than cement particles (Tavakoli et al., 2012). It is a 'super pozzolan' that can improve portland cement production properties. It modifies the physical characteristics of early cement paste and the micro-structural characteristics of cement paste after hardening. Research on the effects of silica fume on concrete started in 1987. The researchers of this study examined the effects of silica fume on high-strength concrete. The highest compressive strength achieved was for a sample

containing 15% cement replacement with silica fume due to the filler effect and pozzolanic reactions (Yogendran et al., 1987). Brooks et al. (1998) also examined the factors affecting the high-strength concrete shrinkage containing this material. Mechanical characteristics in the short and long term on the high strength of concrete has also shown that by replacing 10% and 15% silica fume with cement, its compressive strength did not increase after 90 days and enhancement of strength occurred at the early curing ages. In addition, initial and final creep decreased with increasing amounts of silica fume (Mazloom et al., 2004). Studies on high performance concrete have shown that increasing the superplasticiser from 5 to 20% and decreasing the water-cement ratio from 0.31 to 0.26 caused an increase in compressive strength from 86 to 97 MPa (Sobolev, 2004). Table 3 shows the compressive strength in this scheme. Ganjian and Pouya (2005) investigated the effects of silica fume on resistance to sulfate attack in sea waters subjected to tidal waves and simulated by dry-wet conditions. The results showed that silica fume had more harmful effects on the durability of concrete (Tanyildizi & Coskun, 2008) but if fly ash were added to silica in concrete, the resistance to sulfate attack would improve (Wang et al., 2017) and in the case of reinforcing concrete to steel fibre containing microsilica, tensile strength significantly increased (Köksal et al., 2008). The simultaneous effects of silica fume and nano silica in concrete were investigated. The results showed that using both of these materials in concrete could increase the compressive strength of the concrete. This was because of the filler ability of the nanoparticles and a decrease in porosity (Heidari & Tavakoli, 2013).

Table 3
Details of HPC mixtures (Heidari & Tavakoli, 2013)

Proportions (kg/m³)	SF (5%)	SF (10%)	SF (15%)	SF (20%)
Cement	426	449	468	478
Silica fume	22	50	83	120
Age	Compressive Strength (MPa)			
1 day	16.8	24.1	34.4	45.1
3 days	28.6	42.2	63	84.9
7 days	50.1	67.2	84.8	102.5
28 days	60	80	100	120

As the results showed, although silica fume could improve some of the mechanical features of concrete, it could reduce some durability characteristics. So it is recommended to use other admixtures beside silica fume in concrete to mitigate some of its negative features such as low durability.

Agricultural Waste

From agricultural waste, rice husk ash is the most applicable. The heating value of 1 ton of rice husk is equal to the thermal value of 0.48 tons of coal or 0.36 tons of fuel oil. If rice husk is used for fuel, it burns uncontrolled and many particles change to crystalline, which dramatically

reduces pozzolanic activity. Therefore, if rice husk is going to be used in concrete, it must be burnt under controlled conditions and milled in the long run so that its pozzolanic properties increase. In a study, by calcining rice husk in 500 and using microsilica in high-performance concrete, the researchers obtained porosity decreased by development of the hydration of cement. Using these two materials improve the compressive strength and water absorption of concrete (Huang et al., 2017) and it was proved that rice husk ash has high pozzolanic potential (Mehta, 1992). It can also improve resistance to chloride attack, compressive strength and other mechanical properties (Antiohos et al., 2014). Adding a superplasticiser can also increase slump and decrease viscosity. Using rice husk ash can reduce the filling ability of concrete; however, paste viscosity and segregation rose sharply. By combining rice husk ash and fly ash the self-compacting and compressive strength properties of concrete improved (Le & Ludwig, 2016). Finally, studies have showed that in countries with limited production, rice husk ash can be a valuable additive in concrete products such as high-strength concrete and reconstructive mortars. Corncob has also been used in some studies. Corncob ash consists of more than 65% silicon dioxide and more than 70% combination of aluminium oxide and silicon dioxide. This reflects that the material is cemented and may have a viscous role in concrete (Adesanya, 1996; Adesanya & Raheem, 2009, 2010). Therefore, the use of these materials in the construction of concrete is practical, but they should be used more carefully in order to preserve their pozzolanic property and to strengthen the microstructure of the concrete.

Fly Ash

In power plants that are fuelled by coal, there are spherical particles in the gas that come from burning coal with a diameter of 0.1 to 0.15 mm; the particles are made up of about 85% of silicon, aluminium, iron, magnesium and calcium. Investigating the effects of fly ash started in 1999, and it was observed that the creep and shrinkage level of fly ash samples can be less than those of materials that do not use fly ash (Day, 1990). Also, if the curing temperature increases, the fly ash reaction will increase in the cement paste (Hanehara et al., 2001). This is not limited to only the amount of cement hydration but has effects on the type, characteristics, stability and the production process of hydration (Rojas & Cabrera, 2002; Ma, 2013).

According to ASTM C618, fly ash has two classes, Class F and Class C. The main difference between the two is on the levels of calcium, aluminium, silicon and iron content in the ash. Haque et al. (1984) pioneered the investigation of high-volume fly ash concrete. In 1987 and 1989, studies explored the incorporation of large quantities of fly ash in concrete. The method was about aerated and non-aerated concrete with 55% weight of fly ash substituted for cement in three strength levels of 21, 28 and 35 MPa. The results showed that the initial and final setting were not significantly influenced by replacing fly ash with cement by up to 55%. In addition, concrete containing 40 to 60% of fly ash showed lower compressive strength at an early curing age. However, it showed higher compressive strength compared to similar concrete without fly ash in 28 days (Naik & Ramme, 1987, 1989). Generally, Class F fly ash with good pozzolanic activities cause good mechanical properties, durability and low chloride permeability (Malhotra, 1990). By using a superplasticiser, poor abrasion resistance was generated compared to concrete without fly ash (Bilodeau & Malhotra, 1992). By using

50% fly ash in Class F, an appropriate concrete was obtained for the construction of reinforced concrete structures (Siddique, 2004). Chung-Ho et al. (2013) showed that in fly ash concrete, setting time and air percentage increased with enhancement of fly ash dose. Due to the fact that a high amount of fly ash and good pozzolanic activity reacts with CH of cement, the porosity of concrete decreases. Moreover, compressive and flexural strength of concrete shows an ongoing trend in 91 days and 365 days. Concrete mixtures containing fly ash with low loss of ignition had higher mechanical properties compared to concrete mixtures containing fly ash with high loss of ignition; in addition, increasing fly ash in concrete caused higher shrinkage due to drying at different ages (Chung-Ho et al., 2013). Figure 11 shows the development of shrinkage at different curing ages.
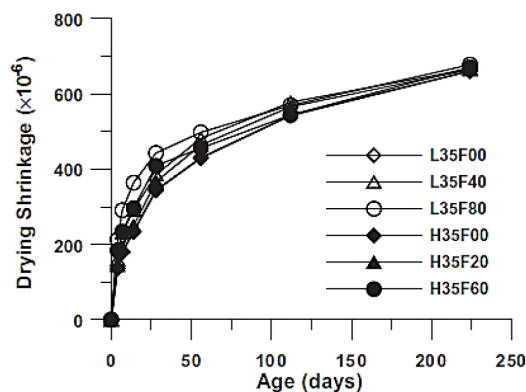
Figure 11. Development of shrinkage (Chung-Ho et al., 2013)

Investigations have also shown that fly ash can increase the compressive strength of concrete on long curing days and improve concrete durability by reducing permeability and increasing density. This type of material, in contrast to concrete waste and brick, is suitable for use in reinforced concrete.

Glass

Researchers have shown that glass particle size has an obvious effect on concrete performance .Smaller particles increase activity with lime, improve compressive strength and reduce shrinkage. The other results showed that the size of the glass does not lead to alkali-silica reaction but the high potential of high alkali glass powder particles leads to destructive expansion (Schwarz & Neithalath, 2007; Shao et al., 1999; Shayan & Xu, 2006; Shi et al., 2005). The experimental tests showed that the compressive strength of concrete samples with 10% glass powder was higher than the samples using fly ash. However, with 90 days' curing, the compressive strength and water absorption of fly ash concrete was higher than that of the glass powder samples. Also, increasing the curing time and adding fly ash to glass powder decreased the chloride diffusion and expansion of the alkali-silica reaction (Schwarz, 2007;

Schwarz et al., 2008; Schwarz & Neithalath, 2008). In high-performance concrete, if glass powder is shattered in micro scale, it will establish useful reactions with cement over time. The formation of calcium silicate hydrate (C-S-H) is very useful for the structure and characteristics of high-performance concrete (Vaitkevičius et al., 2014). Particle distribution and glass size curves can be seen in Figure 12.
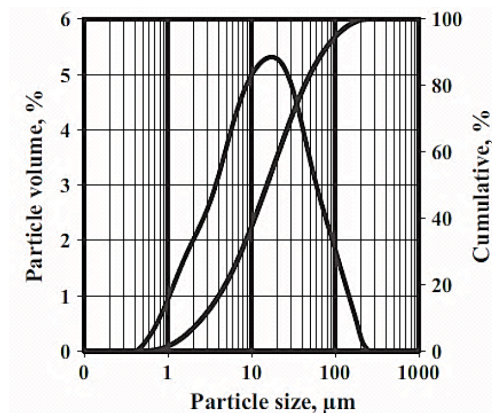
Figure 12. Particle distribution (Vaitkevičius et al., 2014)

In order to reduce ASR reaction, adding fly ash to the mixture of concrete and glass powder is recommended.

Ceramic Tile, Sanitary Ceramic and Clay Brick

The first studies on chemical properties of waste ceramic tiles were done in 2000. The results showed that waste tile has pozzolanic properties and the ability to be used in concrete construction (Shang, 2000; Toledo et al., 2007). Some studies investigated the use of clay brick waste from demolished buildings as a pozzolanic material. The results showed that this material could be replaced in cement (Lin et al., 2010). Tourgal and Jalali (2010) replaced 20% of the samples of ceramic waste as a pozzolan. The experimental results showed that a strength equal to 91% of the control sample can be reached by using these materials. It also reduced the permeability of concrete and increased its efficiency. Heidari and Tavakoli (2014) investigated waste ceramic tile concrete with silica fume to determine the effects of pozzolanic tile waste and confirmed pozzolan activity. Tile powder was used in different amounts in the concrete and its properties were measured. The results showed that increasing ceramic tile reduces compressive strength (Figure 13). However, if silica fume is added, the good effects will be doubled and the concrete defects will be covered. In these experiments the highest compressive strength was observed in the for 20% ceramic tile and 15% silica fume, while the lowest compressive strength was related to the 25% ceramic tile and 5% silica fume (Heidari & Tavakoli, 2010).
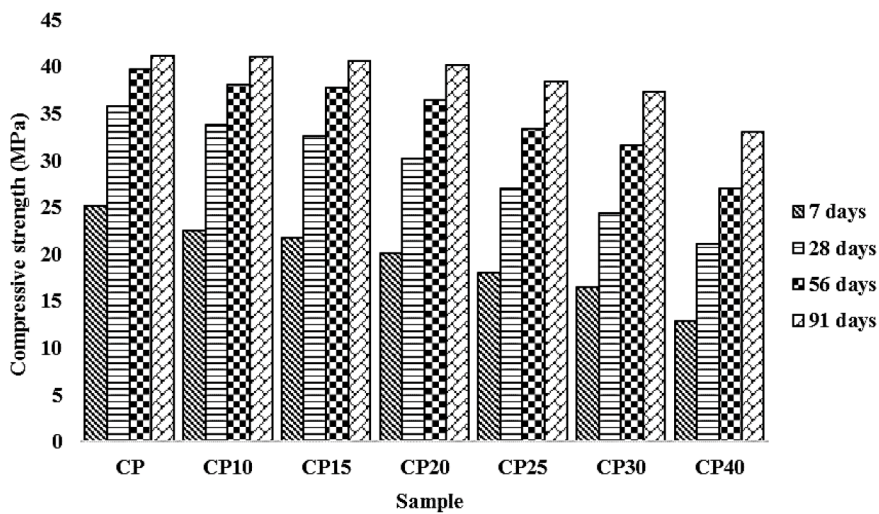
Figure 13. Compressive strength of samples (Phase A) (Heidari & Tavakoli, 2010)

CONCLUSION AND FUTURE WORKS

In the present study, the use of waste materials in concrete was reviewed. The intention of this study was to express the advantages and disadvantages of using waste in concrete, the effect of doing this and introducing materials that can be used in concrete. In order to clarify the details, the reviews were conducted in two parts, considering substitutes as aggregates and substitutes as part of cement. The most important studies in this field were reviewed and the properties of concrete that used waste were investigated. The results showed that waste material can be used in concrete. If waste is used as aggregates in concrete, it can lead to disposal of a large amount of waste. On the one hand, when a material is used as a substitution for cement in concrete, it has two advantages: less use of cement, an element that is very destructive upon the environment and the recycling of waste. It should be noted that further research is necessary to consider different conditions. It is not possible to make use of every kind of waste on every type of concrete as this may endanger the quality of the concrete, and in turn, is especially harmful to the environment. For functions such as filler, binder and separator, a portion of concrete can be replaced with waste as this causes little damage to the final mixture. On the other hand, usually in the process of production and transportation of waste material, harmful additives are added to the waste that can also be harmful to concrete. So further research into using modern equipment to recycle waste and the use of waste material in large projects in a way that causes no damage to concrete, leading to low environmental cleanup, which can be an action towards sustainable development, is needed. Finally, based on other research, the author's experience, lower negative effects and better performance, some types of waste were proposed for the production of concrete. These suggestions are shown in Figure 14 and Figure 15.
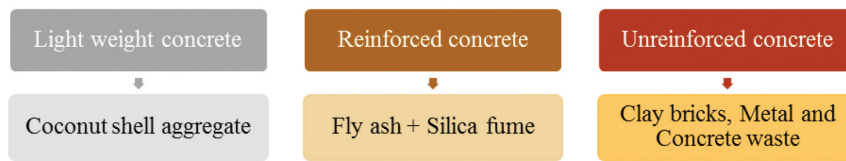
Figure 14. Suggestion of waste for special concrete production

Figure 15. Suggestion of waste based on environmental conditions

However, it should be noted that regional and climatic conditions, as well as access to waste materials, are also effective in this selection. For future study, the authors suggest that researchers could do the following:

- Investigate using waste material and the feasibility of using sand or aggregates to create concrete for the construction of dams, tunnels, roads etc.

- Make sure of the durability and health, through long-term tests, of the concrete that uses waste

- Use neural networks or fuzzy systems for the prediction of parameters of other samples and for developing the results

- Evaluate the properties of other waste and new types of waste such as LCD waste etc.

- Investigate the usage of waste in concrete for improving the environment

- Investigate the use of waste in improving the durability of concrete sewerage pipes

- Improve concrete containing waste by finding its weaknesses and improving its characteristics by using combinations

REFERENCES

Adamson, M., Razmjoo, A., & Poursaee, A. (2015). Durability of concrete incorporating crushed brick as coarse aggregate. *Construction and Building Materials, 94*, 426–432.

Adesanya, D. (1996). Evaluation of blended cement mortar, concrete and stabilized earth made from ordinary Portland cement and corn cob ash. *Construction and Building Materials, 10*(6), 451–456.

Adesanya, D., & Raheem, A. (2009). A study of the workability and compressive strength characteristics of corn cob ash blended cement concrete. *Construction and Building Materials, 23*(1), 311–317.

Adesanya, D., & Raheem, A. (2010). A study of the permeability and acid attack of corn cob ash blended cements. *Construction and Building Materials, 24*(3), 403–409.

Akcaozoglu S., Atis C. D., & Akcaozoglu, K. (2010). An investigation on the use of shredded waste PET bottles as aggregate in lightweight concrete. *Waste Manage (Oxford), 30*(2), 285–290.

Akhtaruzzaman, A. A., & Hasnat, A. (1983). Properties of concrete using crushed brick as aggregates. *Concrete International, 5*(2), 58–63.

Akinmusuru, J. O. (1991). Potential beneficial uses of steel slag wastes for civil engineering purposes. *Resources, Conservation and Recycling, 5*(1), 73–80.

Ambily, P., Umarani, C., Ravisankar, K., Prem., P. R., Bharatkumar, B., & Nagesh R, I. (2015). Studies on ultra high performance concrete incorporating copper slag as fine aggregate. *Jornal of Construction and Building Materials, 77*, 233–240.

Anderson, D. J., Smith, S. T., & Au, F. T. K. (2016). Mechanical properties of concrete utilising waste ceramic as coarse aggregate. *Construction and Building Materials, 117*, 20–28. doi:http://dx.doi.org/10.1016/j.conbuildmat.2016.04.153

Antiohos, S. K., Papadakis, V. G., & Tsimas, S. (2014). Rice husk ash (RHA) effectiveness in cement and concrete as a function of reactive silica and fineness. *Cement and Concrete Research, 61–62*, 20–27.

Araghi, H. J., Nikbin, I. M., Reskati, S. R., Rahmani, E., & Allahyari, H. (2015). An experimental investigation on the erosion resistance of concrete containing various PET particles percentages against sulfuric acid attack. *Construction and Building Materials, 77*, 461-471.

Ay, N., & Unal, M. (2000). The use of waste ground ceramic in cement production. *Cement and Concrete Research, 30*(3), 497–499.

Bilodeau, A., & Malhotra, V. (1992). *Concrete incorporating high volumes of ASTM Class F fly ashes: Mechanical properties and resistance to deicing salt scaling and to chloride-ion penetration*. Turkey.

Bravo, M., & Brito, J. (2012). Concrete made with used tyre aggregates: Durability-Related performance. *Journal of Cleaner Production, 25*, 42–50.

Brooks, J., Cabrera, J., & Megat Johari, M. (1998). Factors affecting the autogenous shrinkage of silica fume high-strength concrete. *Paper presented at the International Workshop on Autogenous Shrinkage of Concrete*. Hiroshima: Japan Concrete Institute.

Cartuxo, F., De Brito, J., Jimenez, J., & Ledesma, E. (2015). Rheological behaviour of concrete made with fine recycled concrete aggregates – Influence of the superplasticizer. *Construction and Building Materials, 89*, 36–47.

Cazacliu, B., &Ventura, A. (2010). Technical and environmental effects of concrete production: Dry batch versus central mixed plant. *Journal of Cleaner Production, 18*(13), 1320–1327.

Chung-Ho, H., Shu-ken, L., Chao-Shun, C., & How-Ji, C. (2013). Mix proportions and mechanical properties of concrete containing very high-volume of class F fly ash. *Construction and Building Materials, 46*, 71–78.

Cota, F. P., Melo, C. C. D., Panzera, T. H., Araújo, A. G., Borges, P. H. R., & Scarpa, F. (2015). Mechanical properties and ASR evaluation of concrete tiles with waste glass aggregate. *Sustainable Cities and Society, 16*, 49-56.

Day, R. L. (1990). Strength, durability and creep of fly-ash concrete part II serviceability and durability of construction materials. *Paper presented at the First Materials Engineering Congress*. Denver, USA.

de Castro, S., & de Brito, J. (2013). Evaluation of the durability of concrete made with crushed glass aggregates. *Journal of Cleaner Production, 41*, 7–14.

Demirboğa, R., & Gül, R. (2006). Production of high strength concrete by use of industrial by-products. *Building and Environment, 41*(8), 1124–1127.

Frias Rojas, M., & Cabrera, J. (2002). The effect of temperature on the hydration rate and stability of the hydration pastes of MK-lime-water systems. *Cement and Concrete Research, 32*(1), 133–138.

Ganjian, E., & Pouya, H. (2005). Effect of magnesium and sulfate ions on durability of silica fume blended mixes exposed to the seawater tidal zone. *Cement and Concrete Research, 35*(7), 1332–1343.

Gesoğlu, M., Güneyisi, E., Khoshnaw, G., & İpek, S. (2014a). Abrasion and freezing-thawing resistance of pervious concretes containing waste rubbers. *Construction and Building Materials, 73*, 19–24.

Gesoğlu, M., Güneyisi, E., Khoshnaw, G., & İpek, S. (2014b). Investigating properties of pervious concretes containing waste tire rubbers. *Construction and Building Materials, 63*, 206–213.

Ghailan, A. H. (2005). *Modified concrete by using a waste material as a coarse aggregate*. Paper presented at the Construction Research Congress.

Gonzalez, A., & Etxeberria, M. (2014). Properties of high performance concrete made with recycled fine ceramic and coarse mixed aggregates. *Construction and Building Materials, 68*, 618–26.

Guerra, I., Vivar, I., Liamas, B., Juan, A., & Moran, J. (2009). Eco-efficient concretes: The effects of using recycled ceramic material from sanitary installations on the mechanical properties of concrete. *Waste Manage (Oxford), 29*, 643–646.

Gunasekaran, K. (2008). Lightweight concrete using coconut shells as aggregate. *Paper presented at the International Conference on Advances in Concrete and Construction.* India: ICACC.

Gunasekaran, K., Annadurai, R., & Kumar, P. (2012). Long term study on compressive and bond strength of coconut shell aggregate concrete. *Construction and Building Materials, 28*(1), 208-215.

Gunasekaran, K., Annadurai, R., & Kumar, P. (2013). Study on reinforced lightweight coconut shell concrete beam behavior under flexure. *Materials and Design, 46*, 157–167.

Gunasekaran, K., Kumar, P., & Lakshmipathy, M. (2010). Compatibility studies on the coconut shell cement composites. *Indian Journal of Indian Concrete Institute, 11*(1), 27–31.

Gunasekaran, K., Kumar, P., & Lakshmipathy, M. (2011a). Mechanical and bond properties of coconut shell concrete. *Construction and building materials, 25*(1), 92–98.

Gunasekaran, K., Kumar, P., & Lakshmipathy, M. (2011b). Study on properties of coconut shell as an aggregate for concrete. *Indian Journal of Indian Concrete Institute, 12*(2), 27–33.

Gupta, T., Chaudhary, S., & Sharma, R. (2014). Assessment of mechanical and durability properties of concrete containing waste rubber tire as fine aggregate. *Construction and Building Materials, 73*, 562–574.

Hanehara, S., Tomosawa, F., Kobayakawa, M., & Hwang, K. (2001). Effects of water/powder ratio, mixing ratio of fly ash, and curing temperature on pozzolanic reaction of fly ash in cement paste. *Cement and Concrete Research, 31*(1), 31–39.

Han-Seung, L., & Wang, X. Y. (2016). Evaluation of compressive strength development and carbonation depth of high volume slag-blended concrete. *Construction and Building Materials, 124*, 45–54.

Haque, M., Langan, B., & Ward, M. (1984). High fly ash concrete. *ACI Material Journal, 81*(1), 54–60.

Heidari, A., & Tavakoli, D. (2010). Investigated concrete with silica fume and pozzolan tile waste. *Sharif Civil Engineering, 30*, 125–133.

Huang, H., Gao, X., Wang, H., & Ye, H. (2017). Influence of rice husk ash on strength and permeability of ultra-high performance concrete. *Construction and Building Materials, 149*, 621–628. doi:http://dx.doi.org/10.1016/j.conbuildmat.2017.05.155

Islam, G. M. S., Rahman, M. H., & Kazi, N. (2017). Waste glass powder as partial replacement of cement for sustainable concrete practice. *International Journal of Sustainable Built Environment, 6*(1), 37–44.

Ismail, Z., & Al-Hashmi, E. (2008). Reuse of waste iron as a partial replacement of sand in concret. *Waste Management, 28*, 2048–2053.

Jeong, G. (2011). Processing and properties of recycled aggregate concrete. *Paper presented at the MSc dissertation in Civil Engineering.*

Kanojia, A., & Jain, S. K. (2017). Performance of coconut shell as coarse aggregate in concrete. *Construction and Building Materials, 140*, 150–156.

Köksal, F., Altun, F., Yiğit, İ., & Şahin, Y. (2008). Combined effect of silica fume and steel fiber on the mechanical properties of high strength concretes. *Construction and Building Materials, 22*(8), 1874–1880.

Le, H. T., & Ludwig, H. M. (2016). Effect of rice husk ash and other mineral admixtures on properties of self-compacting high performance concrete. *Materials and Design, 89*, 156–166.

Letelier, V., Tarela, E., & Moriconi, G. (2017). Mechanical properties of concretes with recycled aggregates and waste brick powder as cement replacement. *Procedia Engineering, 171*, 627–632.

Lin, K., Wu, H., Shie, J., Hwang, C., & Cheng, A. (2010). Recycling waste brick from construction and demolition of buildings as pozzolanic materials. *Waste Manage Res, 28*(7), 653–659.

Ling, T. C., & Poon, C. S. (2012). A comparative study on the feasible use of recycled beverage and CRT funnel glass as fine aggregate in cement mortar. *Journal of Cleaner Production, 29–30*, 46–52.

Lopez, V., Llamas, B., Juan, A., Moran, J. M., & Guerra, I. (2007). Eco-efficient concretes: Impact of the use of white ceramic powder on the mechanical properties of concrete. *Biosystems Engineering, 96*(4), 559–564.

Ma, H. (2013). *Multi-scale modeling of the microstructure and transport properties of contemporary concrete.* (PhD dissertation0. The Hong Kong University of Science and Technology, Hong Kong.

Malhotra, V. (1990). Durability of concrete incorporating high-volume of low calcium (ASTM Class F) fly ash. *Cement and Concrete Composites, 12*(4), 271–277.

Mayer, C., & Baxter, S. (1997). *Use of recycled glass for concrete masonry blocks* (97-15). Final Report, Elsevier, New York.

Mayer, C., & Baxter, S. (1998). *Use of recycled glass and fly ash for precast concrete* (98-18). Report No. NYSERDA 98-18 (4292-IABR-IA-96) to New York StateEnergy Research and Development Authority, Department of Civil Engi-neering Mechanics, Columbia University, New York.

Mazloom, M., Ramezanianpour, A., & Brooks, J. (2004). Effect of silica fume on mechanical properties of high-strength concrete. *Cement and Concrete Composites, 26*(4), 347–357.

Medina, C., Banfill, P., de Rojas, M. S., & Frias, M. (2013). Rheological and calorimetric behaviour of cements blended with containing ceramic sanitary ware and construction/demiltion waste. *Construction and Building Materials, 40*, 822–831.

Medina, C., Frías, M., & de Rojas, M. S. (2012). Microstructure and properties of recycled concretes using ceramic sanitary ware industry waste as coarse aggregate. *Construction and Building Materials, 31*, 112–118.

Medina, C., de Rojas, M. S., & Frías, M. (2012). Reuse of sanitary ceramic wastes as coarse aggregate in eco-efficient concretes. *Cement and Concrete Composites, 34*(1), 48–54.

Mehta, P. (1992). Rice husk ash – A unique supplementary cementing material. In *Proceedings of the international symposium on advances in concrete technology* (pp. 113–22). Athens, Greece.

Merlet, J., & Pimienta, P. (1993). Mechanical and physical-chemical properties of concrete produced with coarse and fine recycled aggregates. *Paper presented at the Demolition and Reuse of Concrete and Masonry*.

Naik, T., & Ramme, B. (1987). Setting and hardening of high fly ash content concrete. *Paper presented at the American Coal Ash Association's 8th International Coal Ash Utilization Symposium*. Washington, USA.

Naik, T., & Ramme, B. (1989). High strength concrete containing large quantities of fly ash. *ACI Material Journal, 86*(2), 111–117.

Olanipekun, E., Olusola, K., & Ata, O. (2006). A comparative study of concrete properties using coconut shell and palm kernel shell as coarse aggregates. *Build Environ, 41*, 297–301.

Olivares, F., & Barluenga, G. (2004). Fire performance of recycled rubber-filled high strength concrete. *Cement and Concrete Research, 34*, 109–117.

Pelisser, F., Zavarise, N., Longo, T., & Bernarin, A. (2011). Concrete made with recycled tire rubber: Effect of alkaline activation and silica fume addition. *Journal of Cleaner Production, 19*(6-7), 757–763.

Portella, K., Joukoski, A., Franck, R., & Derksen, R. (2006). Secondary recycling of electrical insulator porcelain waste in Portland concrete structures: Determination of the performance under accelerated aging. *Cerâmica, 52*, 155–167.

Poutos, K. H., Alani, A. M., Walden, P. J., & Sangha, C. M. (2006). Relative temperature changes within concrete made with glass aggregates. *Construction and Building Materials, 22*(4), 557–565.

Rahmani, E., Dehestani, M., Beygi, M. H. A., Allahyari, H., & Nikbin, I. M. (2013). On the mechanical properties of concrete containing waste PET particles. *Construction and Building Materials, 47*, 1302–1308.

Rai, A., Prabakar, J., Raju, C. B., & Morchalle, R. K. (2002). Metallurgical slag as a component in blended cement. *Construction and Building Materials, 16*, 489–494.

Saikia, N., & De Brito, J. (2004). Mechanical properties and abrasion behaviour of concrete containing shredded PET bottle waste as a partial substitution of natural aggregate. *Construction and Building Materials, 52*, 236–244.

Schwarz, N. (2007). *Evaluating the performance of fine glass powder as a cement replacement material in concrete.* Report. Clarkson University.

Schwarz, N., Cam, H., & Neithalath, N. (2008). Influence of a fine glass powder on the durability characteristics of concrete and its comparison to fly ash. *Cement and Concrete Composites, 30*, 486–496.

Schwarz, N., & Neithalath, N. (2007). Quantifying the cementing efficiency of fine glass powder and its comparison to fly ash. *Paper presented at the first International Conference on Recent Advances in Concrete Technology.* Washington DC.

Schwarz, N., & Neithalath, N. (2008). Influence of a fine glass powder on cement hydration: Comparison to fly ash and modeling the degree of hydration. *Cement Concrete Research, 38*, 429–436.

Sehaj, S., Arun, S., & B., R. (2004). Pullout behavior of polypropylene fibers from cementitious matrix. *Cement Concrete Research, 34*(10), 1919–1925.

Shafigh, P., Mahmud, H., Jumaat, M., & Zargar, M. (2014). Agricultural wastes as aggregate in concrete mixtures – A review. *Construction and Building Materials, 53*, 110–117.

Shang, M. J. (2000). High strength concrete containing natural pozzolan and silica fume. *Cement and Concrete Composites, 22*(6), 399–406.

Shao, Y., Lefort, T., Moras, S., & Rodriguez, D. (1999). Studies on concrete containing ground waste glass. *Concrete and Cement Research, 30*(1), 91–100.

Shayan, A., & Xu, A. (2006). Performance of glass powder as a pozzolanic material in concrete: A field trial on concrete slabs. *Cement Concrete Research, 36*, 457–468.

Shi, C., Wu, Y., Riefler, C., & Wang, H. (2005). Characteristics and pozzolanic reactivity of glass powders. *Cement Concrete Research, 35*, 987–993.

Siamardi, K., & Vahedi, M. (2008). Laboratory evaluation of structural lightweight concrete application of almond shell. *Paper presented at the Fourth Conference of Civil Engineering Students.*

Siddique, R. (2004). Performance characteristics of high-volume Class F fly ash concrete. *Cement Concrete Research, 34*(3), 487–493.

Sobolev, K. (2004). The development of a new method for the proportioning of high-performance concrete mixtures. *Cement and Concrete Composites, 26*(7), 901–907.

Sohrabi, M., & Karbalaie, M. (2011). An experimental study on compressive strength of concrete containing crumb rubber. *International Journal of Civil & Environmental IJCEE, 11*(3), 24–28.

Tanyildizi, H., & Coskun, A. (2008). Performance of lightweight concrete with silica fume after high temperature. *Construction and Building Materials, 22*(10), 2124–2129.

Tavakoli, D., & Heidari, A. (2013). Properties of concrete incorporating silica fume and nano-SiO2. *Indian Journal of Science and Technology, 6*(1), 3946–3950.

Tavakoli, D., Heidari, A., & Behfrouz, B. (2012). Using waste materials in cement and concrete way towards sustainable development. *Paper presented at the First International Cement Industry, Energy and Environment Conference.* Tehran University, Iran.

Tavakoli, D., Heidari, A., & Hayati Pilehrood, S. (2014). Properties of concrete made with waste clay brick as sand incorporating nano SiO2. *Indian Journal of Science and Technology, 7*(12), 1899–1905.

Tavakoli, D., Heidari, A., & Karimian, M. (2013). Properties of concretes produced with waste ceramic tile aggregate. *Asian Journal of Civil Engineering (BHRC), 14*(3), 369–382.

Tay, J., Show, K., Hong, S., Chien, C., & Lee, D. (2003). Thermal stabilization of iron-rich sludge for high strength aggregates. *Journal of Materials in Civil Engineering, 15*(6), 577–585.

Thomas, B. S., & Gupta, R. C. (2016). A comprehensive review on the applications of waste tire rubber in cement concrete. *Renewable and Sustainable Energy Reviews, 54*, 1323–1333.

Toledo F, R. D., Gonçalves, J. P., Americano, B. B., & Fairbairn, E. M. R. (2007). Potential for use of crushed waste calcined-clay brick as a supplementary cementitious material in Brazil. *Cement and Concrete Research, 37*(9), 1357-1365.

Topcu, I. B., & Canbaz, M. (2004). Properties of concrete containing waste glass. *Cement Concrete Research, 34*, 267–274.

Torgal, F., & Jalali, S. (2010). Compressive strength and durability properties of ceramic wastes based concrete. *Construction and Building Materials, 24*, 832–838.

Vaitkevičius, V., Šerelis, E., & Hilbig, H. (2014). The effect of glass powder on the microstructure of ultra high performance concrete. *Construction and Building Materials, 68*, 102–109.

Wang, C. C., & Wang, H. Y. (2017). Assessment of the compressive strength of recycled waste LCD glass concrete using the ultrasonic pulse velocity. *Construction and Building Materials, 137*, 345-353. doi:http://dx.doi.org/10.1016/j.conbuildmat.2017.01.117

Wang, D., Zhou, X., Meng, Y., & Chen, Z. (2017). Durability of concrete containing fly ash and silica fume against combined freezing-thawing and sulfate attack. *Construction and Building Materials, 147*, 398–406.

Won, J. P., Jang, C. I., Lee, S. W., Lee, S. J., & HY, K. (2009). Long-term performance of recycled PET fibre-reinforced cement composites. *Construction and Building Materials, 24*, 660–650.

Yilmaz, A., & Degirmenci, N. (2009). Possibility of using waste tire rubber and fly ash with Portland cement as construction materials. *Waste Manage, 29*, 1541–1546.

Yogendran, V., Langan, B., Haque, M., & Ward, M. (1987). Silica fume in high-strength concrete. *Materials Journal, 84*(2), 124–129.

Yun-Wang, C., Dae-joong, M., Jee-Seung, C., & Sun-Kyu, C. (2005). Effects of waste PET bottles aggregate on the properties of concrete. *Journal of Cement and Concrete Research, 35,* 776–81.

Zega, C., & di Maio, A. (2011). Use of recycled fine aggregate in concrete with durable requirements. *Waste Manage, 31*, 2336–2340.

Review Article

On Modelling Parallel Programmes for Static Mapping: A Comparative Study

Sina Zangbari Koohi[1], Nor Asilah Wati Abdul Hamid[1]*, Mohamed Othman[1] and Gafurjan Ibragimov[2]

[1]*Department of Communication Technology and Network, Faculty of Computer Science and Information Technology, University Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

[2]*Department of Mathematics and Institute for Mathematical Research, Faculty of Science Universiti Putra Malaysia, University Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

Heterogeneous parallel architecture (HPA) are inherently more complicated than their homogeneous counterpart. HPAs allow composition of conventional processors, with specialised processors that target particular types of task. However, this makes mapping and scheduling even more complicated and difficult in parallel applications. Therefore, it is crucial to use a robust modelling approach that can capture all the critical characteristics of the application and facilitate the achieving of optimal mapping. In this study, we perform a concise theoretical analysis as well as a comparison of the existing modelling approaches of parallel applications. The theoretical perspective includes both formal concepts and mathematical definitions based on existing scholarly literature. The important characteristics, success factors and challenges of these modelling approaches have been compared and categorised. The results of the theoretical analysis and comparisons show that the existing modelling approaches still need improvement in parallel application modelling in many aspects such as covered metrics and heterogeneity of processors and networks. Moreover, the results assist us to introduce a new approach, which improves the quality of mapping by taking heterogeneity in action and covering more metrics that help to justify the results in a more accurate way.

Keywords: Heterogeneous parallel architectures, mapping, parallel application modelling, scheduling

INTRODUCTION

In the last two decades, many distributed high-performance computers with thousands or millions of processing units have been in use. The emphasis on distributed and parallel

computers is on shorter execution time, decreasing energy consumption, and reducing idle time of resources. By means of efficient mapping and scheduling, a substantial improvement in these issues can be achieved (Lastovetsky & Manumachu, 2017; Tekinerdogan & Arkin, 2012). Mapping problems is the assignment of the processors to processes and communication links to data elements that are exchanged between the processes. Each processor in a distributed architecture usually manages more than one process in a single application. Managing the order of execution of these processes on each processor is known as the scheduling problem. These two crucial problems in parallel computing fall into NP-hard categories (Kumar, Grama, Gupta, & Karypis, 1994; Lewis & El-Rewini, 1992; Sarkar, 1989; Ullman, 1975).

Mapping and scheduling problems becomes more complex when it involves heterogeneous parallel architecture because this architecture is more complicated than homogeneous parallel architecture (Rico-Gallego, Lastovetsky, & Diaz-Martin, 2017). The first step in solving mapping and scheduling problems lies in the modelling phase. In computer science, to represent real entities, such as processing architecture, computational models are used. A computational model is a simplified version of these entities, where crucial characteristics are captured and implementation details are ignored (Leopold, 2001). The information captured from the programme to be executed is modelled by programme graphs or hyper-graphs.

There are several modelling approaches studied in the literature to capture the behaviour of parallel applications such as PRAM (Kumar et al., 1994), LogP (Culler et al., 1996), BSP (Bulk-Synchronous Parallel) (Valiant, 1990), TIG (Task Interaction Graph) (Long & Clarke, 1989), TPG (Task Precedence Graph) (Kasahara & Narita, 1985), TTIG (Task Temporal Interaction Graph) (Roig, Ripoll, & Guirado, 2007; Roig, Ripoll, Senar, Guirado, & Luque, 2000, 2002), TTIGHa (Temporal Task Interaction Graph in Heterogeneous Architecture) (De Giusti, Chichizola, Naiouf, Ripoll, & De Giusti, 2007), MPAHA (Model on Parallel Algorithms on Heterogeneous Architectures) (De Giusti, Naiouf, Chichizola, Luque Fadón, & De Giusti, 2009) and the hypergraph-based model proposed in the UMPa scheduling algorithm (Deveci, 2015; Deveci, Kaya, Ucar, & Catalyurek, 2015).

The varieties in modelling approaches motivated this research, which focused on an analysis and comparative study of frameworks, similarities, differences, characteristics and principles. The comparison will assist in determining the suitability, success factors and challenges of each modelling approach.

The remaining sections are organised as follows. In Section 2, each model is discussed in detail and its mathematical structure is shown. In Section 3, a comparative study of these models is presented and discussed. In Section 4, the new model is proposed as a future work and in Section 5, the study is concluded.

MODELLING STRUCTURES

In this section, modelling approaches are introduced. In the next section, a comparative study is presented based on these modelling approaches.

Task Interaction Graph (TIG)

The Task Interaction Graph (TIG) model isolates an application into the maximum number of sequential blocks (tasks) that are connected by edges. These edges represent the interactions between the blocks. The TIG model uses an undirected graph to abstract the application into the model. This model is an undirected graph $G = (V, E, c, w)$ (Long & Clarke, 1989), where:

- V is a set of nodes. Each node in this set represents a task T_i.

- E is a set of edges. Each edge in this set represents communication between tasks.

- $c(T_i)$ is a positive cost associated with task, T_i. This cost represents the computation time of the task, T_i.

- $w(T_i, T_j)$ is a non-negative weight associated to the edge between T_i and T_j. This weight represents the total communication volume between two edges.

To illustrate the model, the sample application exhibited in Figure 1 is considered. This sample application consists of three tasks communicating with each other. Each curve in this figure is a task. The corresponding pseudo-code of this sample application is given in Figure 2.
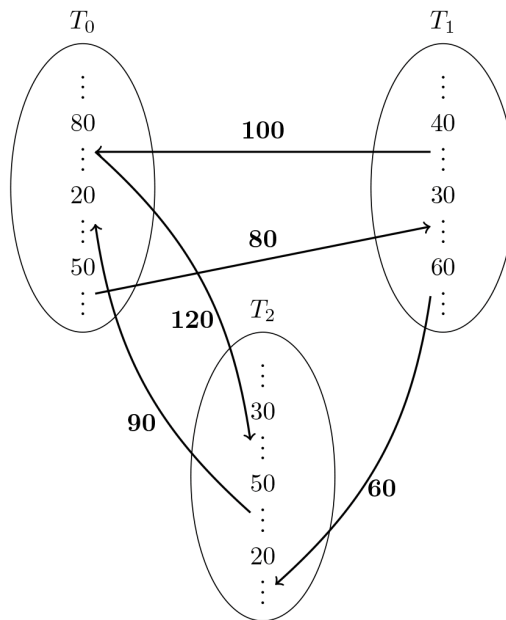
Figure 1. Temporal flow graph of sample application

Pseudo-code 1 a sample application

```
 1: if Task = T_0 then
 2:     ... BC with CT=80 ...
 3:     RECV(d_1, 100, T_1);
 4:     SND(d_2, 120, T_2);
 5:     ... BC with CT=20 ...
 6:     RECV(d_3, 90, T_2);
 7:     ... BC with CT=50 ...
 8:     SND(d_4, 80, T_1);
 9: end if
10: if Task = T_1 then
11:     ... BC with CT=40 ...
12:     SND(d_1, 100, T_0);
13:     ... BC with CT=30 ...
14:     RECV(d_4, 80, T_0);
15:     ... BC with CT=60 ...
16:     SND(d_5, 60, T_2);
17: end if
18: if Task = T_2 then
19:     ... BC with CT=30 ...
20:     RECV(d_2, 120, T_0);
21:     ... BC with CT=50 ...
22:     SND(d_3, 90, T_0);
23:     ... BC with CT=20 ...
24:     RECV(d_5, 60, T_1);
25: end if
```

Figure 2. Pseudo-code 1 sample application

In this pseudo-code, BC stands for block of computation and CT stands for computation time, while $\mathrm{SND}(d, n, T_i)$ is a communication command that sends the data element d, with volume n, to the task T_i. $RECV(d, n, T_i)$ is another communication command that receives data element d with volume n from task T_i.

Holding the above example, the TIG model for this application is formed. Figure 3 illustrates the TIG model of the sample application.
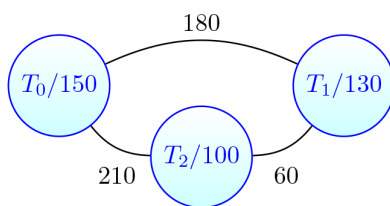
Figure 3. TIG graph for sample application

In the early years of heterogeneous parallel computing, this model was used by many researchers and in different studies, such as by El-Rewini, Lewis and Ali (1994), Hui and Chanson (1997), Kalinov and Klimov (2005), Phinjaroenphan and Bevinakoppa (2004), Sadayappan, Ercal and Ramanujam (1990), and Sanyal and Das (2005). However, with the growth of parallel computers and the appearance of more complex parallel applications, the drawbacks of this model were exposed.

Task Precedence Graph (TPG)

Liu, Shi, Lu and Mao (2007) proposed a new model named the Task Precedence Graph (TPG). The TPG is a refined version of the TIG model that attempts to capture the precedence relations between tasks. The authors used a directed graph to model the parallel applications and their communications and relations. Besides the details captured in the model, the TIG model records the predecessor and successor relations in its graph.

The TPG is a directed graph $G = (PV, PE$ (Kasahara & Narita, 1985), where:

- PV is a set of nodes. Each node pv_i represents a task T_i in the application.
- PE is a set of edges. Each edge e_{ij} represents a direct drive relation between two communicating nodes pv_i and pv_j.

$$PE = \{e_{ij} | e_{ij} = Edge(pv_i, pv_j); pv_j \in Successor(pv_i); i, j \in \{0, 1, \dots, n\}\}$$

In this model, each edge shows both communication volume and the precedence relation between the tasks. Like TIG, in this model, each node has a cost representing its computation time and each edge has a weight, representing its communication volume. Considering the sample application illustrated in Figure 1, the TPG model for this application is shown in Figure 4.
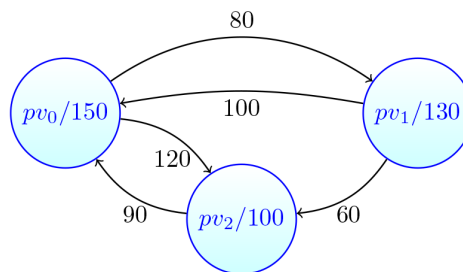
Figure 4. TPG graph for sample application

Task Temporal Interaction Graph (TTIG)

Roig et al. (2007, 2000, 2002) proposed a new model named Task Temporal Interaction Graph (TTIG), which is a refined combination of the TIG and TPG (Roig et al., 2000). In this model, a new parameter, the degree of parallelism, is proposed. This new parameter adds the ability to model recording the potential parallelism for communicating tasks with an arbitrary task interaction pattern (Roig et al., 2007).

The TTIG model for parallel applications is defined as a directed graph $G = (V, E)$ with three maps $w, c, p,$ where:

- V is a set of nodes. Each node in this set represents a task T_i.

- E is a set of directed edges. Each edge in this set represents a temporal relation between two nodes.

- $w: V \to N$ is a function that assigns a non-negative computation time $w(T_i)$ to each task in V.

- $c: V \to N$ is a function that assigns a non-negative communication cost to each edge. $c(T_i, T_j)$ is the total volume of messages being transferred between T_i and T_j.

- $p: E \to [0,1]$ is a function that assigns a normalised index to each edge. Without considering any communication cost or any other dependencies, $p(T_i, T_j)$ is the maximum degree of parallelism that two tasks T_i and T_j can obtain during parallel execution.

To set the TTIG graph for an application, four different steps should be taken:
1. Acquiring the Temporal Flow Graph
2. Calculating the task execution and communication costs
3. Calculating the degree of parallelism
4. Forming the Temporal Task Interaction Graph (TTIG)

For more details about the degree of parallelism and the TTIG model, refer to Roig et al. (2007, 2000, 2002).

Based on the previously explained steps, the TTIG graph for the sample application introduced in Figure 1 will appear as given below (Figure 5).
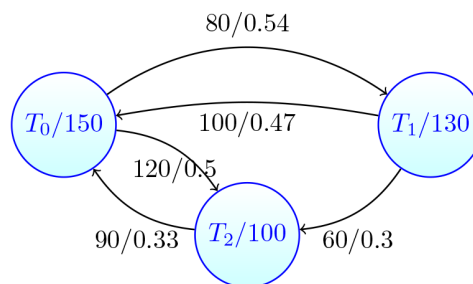
Figure 5. TTIG graph for sample application

Temporal Task Interaction Graph in Heterogeneous Architectures (TTIGHa)

Temporal Task Interaction Graph in Heterogeneous Architectures (TTIGHa) is a refined version of the TTIG model, where the heterogeneity of processors and heterogeneity of communication media in a distributed architecture are taken into account.

- Formally, TTIGHa is a directed graph $G = (V, E, T_p, T_c)$, where:

- V is a set of nodes. Each node in this set represents a task T_i.

- E is a set of directed edges. Each edges in this set represents communication between tasks.

- T_p is a set of processors. For each type of processor in the target distributed machine, there is an element in this set exhibiting the corresponding processor type.

- T_c is a set of communication media. For each type of communication medium in the target distributed machine, there is an element in this set exhibiting corresponding medium type, startup time and transfer time for one byte of the data element.

This model gathers more information about the application. First, the model determines the execution time of each task on every type of processor. $w_p(T_i)$ is the execution time of the task T_i on processor type p. A $\#m \times \#m$ dimension matrix named C is built for each edge (T_i, T_j), where $\#m$ is the number of processor types in the distributed architecture ($\#m = |T_p|$). For each edge $(T_i, T_j) \in E$, the value of $C_{sd}(T_i, T_j)$ is the time needed for transferring one byte of data from task $T_i \in V$ on processor type $s \in T_p$ to task $T_j \in V$ on processor type $d \in T_p$. A $\#m \times \#m$ dimension matrix named P is formed, where $P_{sd}(T_i, T_j)$ is the degree of parallelism between tasks $T_i \in V$ on processor type $s \in T_p$ and $T_j \in V$ on processor type $d \in T_p$. For more information about adjusting these matrices, refer to De Giusti et al. (2007).

After the formal definition of the TTIGHa model, the sample application presented in Figure 1 and Pseudo-code 1 is modelled using this approach. As this model needs information regarding the distributed architecture that the application will run on, a distributed machine with two types of processor, p_1 and p_2, is considered, where the execution speed of processor p_2 is half that of processor p_1, i.e. if task t takes e seconds to run on processor p_1, it will take $2 \times e$ seconds on processor p_2. Furthermore, it is considered that this machine utilises only one type of communication medium with a startup time of 1 second and communication speed of 0.2 seconds per byte. Considering this distributed machine, the TTIGHa model for the sample application illustrated in Figure 1 is as shown in Figure 6.
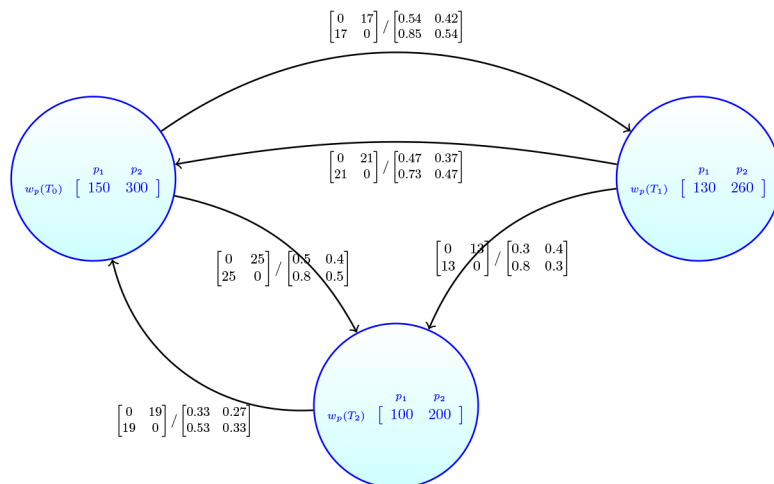
Figure 6. TTIGHa graph for sample application

Model of Parallel Algorithms on Heterogeneous Architectures (MPAHA)

The model of Parallel Algorithms in Heterogeneous Architectures (MPAHA) takes the heterogeneity of processors and networks into account, but in a clearer and simpler way. This model tries to arrange the information captured from the application in a systematic way.

According to the definition, the MPAHA is a directed graph $G = (V, E)$, where:

- V is a set of nodes. Each node in this set represents a task T_i.

- E is a set of edges. Each edge in this set represents communication between tasks.

In this model, besides the tasks and the communication between them, two other aspects of the application are recorded in the graph. The first one is the execution time of the computation phases (in this model called subtasks) for each task on different types of processor. The other one is the communication volume between computation phases (subtasks). Each task $T_i \in V$ consists of multiple subtasks I_i, where there is no communication within each subtask i.e. communication takes place between subtasks. The subtasks for each task $V_i(I_j, p)$ is defined as the computation time of the subtask I_j in task $T_i \in V$ on processor type p. For each task $T_i \in V$, the corresponding matrix $V_i(I, p)$ is formed as assigned.

The other parameter is the communication volume. $E_{ij}(I_m, I_n)$ is set as the communication volume between the subtask I_m in the task T_i ($I_m \in T_i$) to subtask I_n in task T_j ($I_n \in T_j$). Then, the matrix $E(I_m, I_n)$ is assigned to the communication E_{ij} (De Giusti et al., 2007). As in previous approaches, the sample application in Figure 1 is modelled using this modelling method. The result is shown in Figure 7.
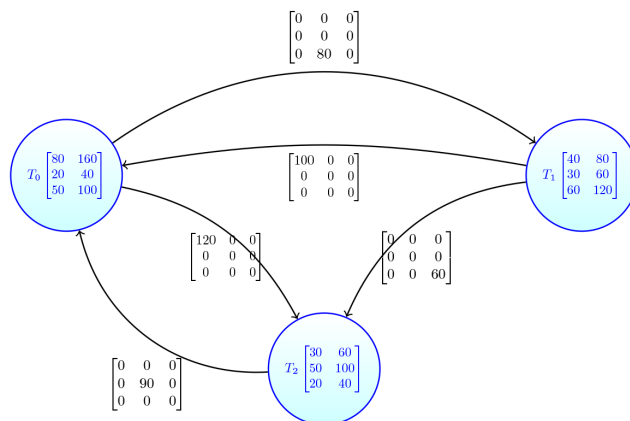
Figure 7. MPAHA graph for sample application

Hypergraph Based Modelling: UMPa

All the models discussed in the previous sections use the graph theory as their basic tool to capture the parallel application's details. However, the Hypergraph Based Modelling: UMPa proposed by Deveci (2015) and Deveci et al. (2015) uses a new structure named hypergraph. Hypergraph-based modelling approaches are more flexible in modelling parallel heterogeneous structures than graph versions.

UMPa utilises a directed hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where:

- $\mathcal{V}$ is a set of nodes. Each node in this set represents a task $t_i \in \mathcal{T}$ or data exchange in communication $d_i \in \mathcal{D}$.

- $\mathcal{N}$ is a set of nets. Each net in this set connects the task that starts the communication $t \in \mathcal{T}$ (producer), the data element being exchanged $d_i \in \mathcal{D}$ and the task receiving the data element $t_j \in \mathcal{T}$ (consumer).

A weight to each node and a cost to each net are assigned. The weight $w[v_i]$ for the node $v_i \in \mathcal{V}$ is the computation time of the task $t_i \in \mathcal{T}$ if this node denotes a task and 0 if this node denotes a data element:

$$w[v_i] = \begin{cases} exec(t_i) & v_i = t_i \in \mathcal{T} \\ 0 & v_i = d_i \in \mathcal{D} \end{cases}$$

where, $exec(t_i)$ is the computation time of the task t_i.

This weight has two functions. While it shows the computation time for each task, it also serves as a factor to distinguish between task nodes and communication nodes.

Another function, c, which assigns a cost $c[n_i]$ to each net $n_i \in \mathcal{N}$, provides the value of $c[n_i]$ for each communication between nodes equal to the communication volume of the data element being exchanged.

$$c[n_i] = size(d_i)$$

where, $size(d_i)$ is the communication volume of the data element $d_i \in \mathcal{D}$ (Deveci, 2015; Deveci et al., 2015). The model hypergraph for the sample application in Figure 1 using the UMPa approach is shown in Figure 8.
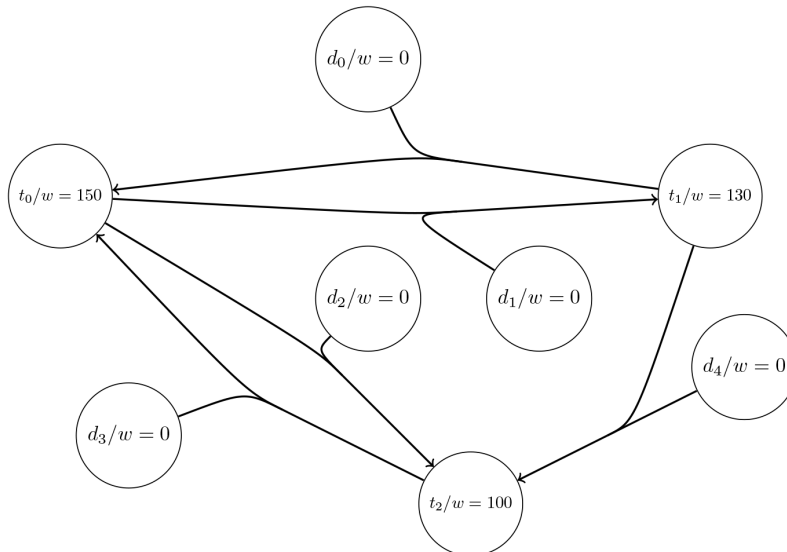
Figure 8. UMPa graph for sample application

COMPARATIVE STUDY

The modelling approaches discussed in the previous section established a set of numerous features to cover different characteristics and principles in modelling parallel applications. Due to this fact, all their possible characteristics cannot be simply listed in this study, but a group of basic characteristics and principles can be provided to distinguish between the modelling approaches. This group is divided into four dimensions: 1- temporal behaviour; 2- mathematical structure; 3- heterogeneity, and; 4- metrics. In this section these dimensions are discussed.

Temporal Behavioural

Temporal behaviour such as, "the output X within the next T time must be produced by a given system" or "the produced output must be sent in no bounded delivery time" allow for the solving of the problem of mapping and scheduling, making it more sufficient. The main objective in modelling an application is the minimisation of resource usage such as execution time (Roig et al., 2000). Inattention to temporal behaviour or task precedence relation in modelling causes tasks to be trapped in wait conditions. Waiting for a task for a message from another task to be received imposes an overhead and increases the total execution time of the application, and this reduces the performance of the mapping (Liu et al., 2007). This characteristic is not supported by all modelling approaches. Table 1 indicates the approaches that do support this feature.

Table 1
Temporal behaviour

Model	Temporal Behaviour
TIG	
TPG	✓
TTIG	✓
TTIGHa	✓
MPAHA	✓
UMPa	

Mathematical Structure

Each model employs a mathematical structure that helps to extract information from the subject application. Each structure has specific features and can capture certain types of information. In case of mapping, the common structures are graphs and hypergraphs. Moreover, each graph or hypergraph can be directed or undirected. Each modelling approach tries to capture essential details of the target application as much as possible in an unambiguous arrangement. However,

the structure being used should be able to handle the details captured. Overall, it has been shown that hypergraphs are more suitable structures in parallel application modelling and are able to capture more details than graph structures (Deveci, 2015; Deveci et al., 2015). In Table 2, the mathematical structures of the models are summed up.

Table 2
Models' mathematical structures

Models	Graph		Hyper-graph
	Undirected	Directed	
TIG	✓		
TPG		✓	
TTIG		✓	
TTIGHa		✓	
MPAHA		✓	
UMPa			✓

Heterogeneity

According to the structure being used and the weights and costs assigned to it, models can embrace some levels of heterogeneity. In modern parallel systems, two types of heterogeneity are common: processor heterogeneity and network heterogeneity. Processor heterogeneity means that a task can have a different execution time on different processors in the machine. Moreover, network heterogeneity indicates that a data block can have different transmission times on varied media that are connected to the processors. Failure to capture the heterogeneity of the target structure leads to inefficient scheduling and mapping, which reduces the overall execution performance (Asaadi, Khaldi, & Chapman, 2016; Xie, Zeng, Xiao, Li, & Li, 2017). Table 3 presents the models' ability to apprehend the heterogeneity of the machine.

Table 3
Models' ability to capture heterogeneity

Models	Heterogeneity	
	Processor	Network
TIG		
TPG		
TTIG		
TTIGHa	✓	✓
MPAHA	✓	✓
UMPa	✓	

Metrics. The final goal of modelling approaches is to map the subject application to the machine. Mapping has better quality if the produced scheduling decreases the resource usage on the target machine in comparison with a different scheduling. By defining a relevant cost function, the mapping problem transforms into an optimisation problem. The cost function for an optimisation problem specifies some constraints in the optimiser being optimised. The modelling approach should provide for the evaluation of these constraints. Metrics provided by the modelling approach correspond to the constraints of the optimisation cost function. The incompetence of the modelling approach in supporting metrics limits the optimiser's ability to produce efficient scheduling plans (Deveci, 2015). To measure the quality of the mapping, there are various metrics that have to be examined, such as total communication volume, total number of messages being sent, maximum number of messages being sent and total volume of messages being sent. The modelling scheme being used determines which metrics can be used. The modelling structure and dependent information prepare the mapper for calculating some of these metrics.

Table 4 summarises the ability of the models in calculating different metrics. In this table, the metrics shown are total communication volume (TCV), total number of messages being sent (or received) (TSM), maximum number of messages being sent (or received) (MMS) and total volume of messages being sent (or received) (TSV). The definitions and characteristics of modelling approaches are collected in the previous sections. According to this, there is no general modelling approach that can capture all required principles and characteristics in modelling parallel applications. The following is a brief discussion of our observation.

Table 4
Models' ability to provide metrics

Models	TCV	TSM	MMS	TSV
TIG	✓			
TPG	✓			
TTIG	✓			
TTIGHa	✓			
MPAHA	✓			
UMPa	✓	✓	✓	

The first and simplest modelling approach discussed is the Task Interaction Graph (TIG). Besides the broad usage of this model, the most significant flaw of this model is the lack of attention to the temporal behaviour of the applications. In this model, there is no tool or parameter to capture the temporal behaviour of the parallel programmes. Therefore, for example, if this model runs tasks t_0 and t_1 concurrently on different processors, and task t_2 needs some data produced by task t_1 at the end of its execution, then task t_2 has to wait; this waiting time is an overhead in computation resources. Furthermore, this model does not reflect the heterogeneity of the processors or networks. All these debilities make this model an inappropriate choice for big, modern distributed machines.

Nevertheless, different experiments show that this model could yield good performance in course-grain applications where a lot of computation is performed between small communication events (Ahmad, He, & Liou, 2002; Bishop, Kelliher, & Irwin, 1999; Censor, Gordon, & Gordon, 2001; Lam & Suen, 1995; Roig et al., 2007). For SPMD applications, where a single application is applied to different data sets concurrently, this model yields good performance (Karypis, Schloegel, & Kumar, 2003; Roig et al., 2002).

The other model discussed is the TPG model. This model attempts to refine the TIG model to capture the temporal behaviour of the parallel applications. It was used by many researchers such as Ali and El-Rewini (1993), Barbosa, Morais, Nobrega and Monteiro (2005), Bouvry, de Kergommeaux and Trystram (1995), Gil, Hernández, Rodriguez, Mauri and Radeva (2006), Hwang, Chow, Anger and Lee (1989), Kitajima, Tron and Plateau (1993), Kwok and Ahmad (1999, 1996), Ohtaki, Takahashi, Boku and Sato (2004), Topcuoglu, Hariri and Wu (2002), Xie and Qin (2005) and Yang and Gerasoulis (1994). These researchers showed that this model performed well for tightly coupled applications, where most of the communication events were at the beginning or at the end of the tasks. Although to model other applications it is possible to break tasks into smaller tasks where communication events take place at the beginning or end, this action causes TPG to build a very big graph of the application, which in turn, increases the complexity of the model and reduces its performance in a large scale (Roig et al., 2002). Moreover, like the TIG, this model does not seize any parameters related to heterogeneity of architecture, making it incompetent for use in model applications for modern clusters.

The TTIG model is the combination of two previously discussed models and it captures the temporal behaviour of parallel applications using a newly defined metric called degree of parallelism. This model is used in many practical algorithms as shown by Guirado, Roig and Ripoll (2013), Kang, He and Wei (2013), Upadhyaya and Rajan (2015), and Yang, Guang, Säntti and Plosila (2013). The results obtained by these authors confirmed that this model was suitable for applications for which the degree of parallelism was more than 0.5 (Roig et al., 2007, 2000, 2002). Accordingly, it is better to use previous models instead of TTIG in other applications that have a degree of parallelism more than 0.5. Additionally, this model does not weigh any parameter regarding heterogeneity of the distributed architectures; this makes it unsuitable for use in new distributed clusters.

None of the models discussed reflects the heterogeneity of the distributed architecture. The next model to be discussed, TTIGHa, is a revision of the TTIG model, which tries to capture heterogeneity within the architecture. Therefore, this model needs to be updated with the exact details of the architecture in which the parallel application is going to be executed (De Giusti et al., 2007). Some studies employ this model in their modelling phase and utilise the TTIGHa to model parallel applications (De Giusti, Chichizola, Naiouf, & De Giusti, 2008; De Giusti, Chichizola, Naiouf, & De Giusti, 2008). However, since this model is closely dependent on the machine's architecture, its structure is complex and the implementations if this model cannot be portable. In this model, the user needs to know the exact details of the architecture; however, obtaining this information is difficult or impossible in some contexts.

The MPAHA is a complete model that attempts to capture the benefits of all the previous models and propose a model that can be applied in different situations. This model captures the temporal behaviour of the applications and heterogeneity of the architecture without the

need to know the exact details of the hardware. Successful implementations of this model such as shown by De Giusti et al. (2009), and De Giusti, Chichizola, Naiouf, De Giusti and Luque (2010) reflect its good performance. However, the weakness of this model is that the application that is modelled has a lot of small volume communication in addition to some large volume communication. Since this method admits the volume of communication, in these circumstances, it induces overheads that reduce performance. In mapping that uses this model, the scheduler is able to optimise large-size communication when several such events are events are entered into the application.

The last model discussed uses the UMPa method and is called UMPa as well. Unlike the other models, this method uses the hypergraph instead of the graphs to model parallel applications. It has been shown that hypergraphs are more suitable than graphs for capturing the parallel application's characteristics (Karypis & Kumar, 2000). Although this is a new proposed method, different studies have been done on it and various implementations are done using this model (Balci & Akgüller, 2014; Schlag et al., 2015; Shahid, Raza, & Sajid, 2015). This model tries to cover many features of parallel applications such as temporal behaviour and heterogeneous architecture. The implementation of this method attests to its good performance and scalability (Balci & Akgüller, 2014; Deveci, 2015; Deveci et al., 2015; Schlag et al., 2015; Shahid et al., 2015). Nonetheless, the model has a weakness; tasks need to collect data, which are produced with more than one task. In parallel applications where the data element has more than one producer, this model cannot be used (Deveci, 2015; Deveci et al., 2015).

FUTURE WORK, A NEW MODELLING APPROACH

The main result of our observations is that the existing modelling approaches do not include all principles and characteristics needed in modelling parallel applications. In modern parallel and distributed computing, there are numerous situations, and improvements are to these models to abstract the applications.

In this section, a new modelling schema will be introduced that attempts to overcome the drawbacks of the modelling approaches. This model is an improved version of the UMPa, which was introduced in the previous section. It has been shown that hypergraphs are better structures for capturing the structure of parallel applications (Trifunovic & Knottenbelt, 2004). This model benefits from the hypergraphs along with the other measures. The detailed definition of this model is as follows. However, this is only an introduction of the model; other researchers and authors may use this as a basis for future work. In the future, the complete version of this modelling along with experiments will be provided.

Definition

In this model, for each parallel application, there are two sets:
1. Set of the tasks $\mathcal{T}$
2. Set of the volume of data elements being exchanged $\mathcal{D}$

Distributed machines on which the application is going to be executed have two sets:

1. Set of processors $\mathcal{P} = \{p_1, p_2, \ldots, p_i\}$ representing the type of each processor

2. Set of communication links $\mathcal{M} = \{m_1, m_2, \ldots, m_i\}$ representing the communication delay of each link in transferring 1KB of data.

Using sets $\mathcal{P}$ and $\mathcal{M}$, a symmetric matrix $\mathcal{S}$ with size $|\mathcal{P} \times \mathcal{P}|$ is defined for the distributed machine as shown below:

$$
\mathcal{S} = \begin{bmatrix}
ct_{11} & ct_{12} & ct_{13} & \cdots & ct_{1|\mathcal{P}|} \\
ct_{21} & ct_{22} & ct_{23} & \cdots & ct_{2|\mathcal{P}|} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
ct_{|\mathcal{P}|1} & ct_{|\mathcal{P}|2} & ct_{|\mathcal{P}|3} & \cdots & ct_{|\mathcal{P}||\mathcal{P}|}
\end{bmatrix}
$$

where, ct_{ij} represents the communication time of 1KB data between processor p_i and processor p_j. Clearly, $ct_{ij} = 0$ if and only if $i = j$ as there is no communication delay within any processor and for every two processors, p_i and p_j, $ct_{ij} = ct_{ji}$.

Using the sets and matrix defined above, a directed hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{A})$ is formed, where:

● $\mathcal{V}$ is the set of nodes. Each node v_i in this set represents a task $t_i \in \mathcal{T}$.

● $\mathcal{A}$ is the set of hyperarcs. Each hyperarc a_i in this set represents a data element $d_i \in \mathcal{D}$. The pins inside each hyperarc a_i ($pins[a_i]$) consist of tasks that produce the data d_i and the tasks that consume this data. Hyperarc $a_i \in \mathcal{A}$ is an ordered pair (Pr_i, Cn_i), where Pr_i and Cn_i are disjointed non-empty subsets of $\mathcal{V}$. Pr_i, which is the origin of a_i ($Pr_i = org(a_i)$), is a set of the producers of d_i. Moreover, Cn_i, which is the destination of a_i ($Cn_i = dest(a_i)$), is a set of the consumers of d_i. The hyperarc's flow is from $org(a_i)$ to $dest(a_i) = pins[a_i] \setminus org(a_i)$. (For more information on the directed hypergraph and its notations, refer to Gallo, Longo, Pallottino and Nguyen (1993))

Each node has a weight and each hyperarc has a cost. Each node v_i has a row vector $w_{v_i} = \begin{bmatrix} et_{i_{p_1}} & et_{i_{p_2}} & \cdots & st_{i_{p_{|\mathcal{P}|}}} \end{bmatrix}$ assigned to it as its weight. Each element of this vector $et_{i_{p_i}}$ represents the execution time of the task $t_i \in \mathcal{T}$ on processor $p_j \in \mathcal{P}$. Each hyperarc a_i has a cost $c[a_i]$ assigned to it that shows the volume of the data element $d_i \in \mathcal{D}$ ($c[a_i] = vol(d_i)$).

Modelling

In this part, a sample application presented in Figure 1 is modelled using the proposed modelling approach. Since this model needs information about the distributed architecture that the parallel application is going to be executed on, a four-processor machine consisting of two types of processor connecting by two types of communication medium is considered. Moreover, it is considered that the execution speed of the type two processor is half that of the type one and the communication delay of the type one medium is half that of the type two medium. The sample structure of the considered architecture is shown in Figure 9. Clearly, when there are different media connecting two processors, the maximum communication delay is assigned to the link.
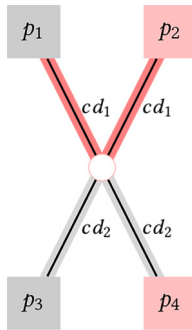
Figure 9. A schematic representation of a simple parallel machine

The preliminary sets are as below:

$$\begin{aligned}
\mathcal{T} &= \{t_1, t_2, t_3\} \\
\mathcal{D} &= \{d_1, d_2, d_3, d_4, d_5\} \\
\mathcal{P} &= \{p_{type_1}, p_{type_2}, p_{type_1}, p_{type_2}\} \\
\mathcal{M} &= \{cd_1, cd_2\}
\end{aligned}$$

where, cd_i represents the communication delay of the type i medium for 1KB of data and $cd_2 > cd_1$.

$$S = \begin{bmatrix}
0 & cd_1 & cd_2 & cd_2 \\
cd_1 & 0 & cd_2 & cd_2 \\
cd_2 & cd_2 & 0 & cd_2 \\
cd_2 & cd_2 & cd_2 & 0
\end{bmatrix}$$

The model for this sample application is a directed hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{A})$, where:

- $\mathcal{V} = \{v_1, v_2, v_3\}$
- $\mathcal{A} = \{a_0, a_1, a_2, a_3, a_4\}$

The presentation of this model for this application is shown in Figure 10.
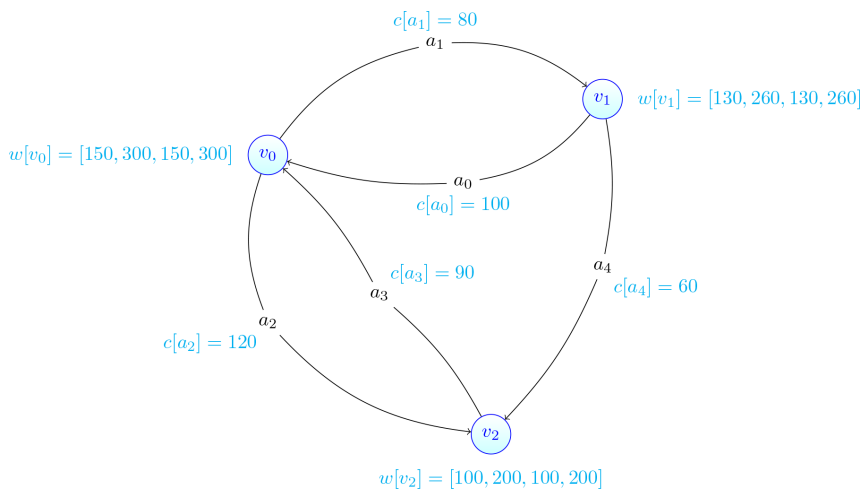
Figure 10. A graphical representation of the sample parallel application modelled using the proposed new modelling approach

CONCLUSION

This review afforded a general study of the modelling schemes proposed in the literature to enhance the understanding of unclear concepts by indicating the common principles and characteristics, similarities and differences and limitation and gap analysis of the modelling approaches. In order to help readers to make a wise selection between the models according to what they need, they are classified and compared from different aspects. The comparison framework is applied from theoretical and practical viewpoints that would recommend the most suitable model(s) according to the provisions of the machine and application. To support the conclusion drawn, four tables have been added as a summary of the evaluation. The main result of this study is that there is no one general purpose modelling approach that can capture all the principles and characteristics in modelling parallel applications. The study also identified the limitations of each modelling approach.

Finally, a new modelling approach that will solve the previous drawbacks by tolerating more types of application, allowing more metrics to measure and improving the accuracy of the mapping is briefly introduced. Extensive investigating, studying and implementing of the proposed model will be our future work.

ACKNOWLEDGEMENT

REFERENCES

Ahmad, I., He, Y., & Liou, M. L. (2002). Video compression with parallel processing. *Parallel Computing*, *28*(7), 1039–1078.

Ali, H., & El-Rewini, H. (1993). Task allocation in distributed systems: A split graph model. *Journal of Combinatorial Mathematics and Combinatorial Computing*, *14*(1993), 15–32.

Asaadi, H. R., Khaldi, D., & Chapman, B. (2016). A comparative survey of the HPC and big data paradigms: Analysis and experiments. In *Proceedings – IEEE International Conference on Cluster Computing, ICCC* (pp. 423–432). IEEE. https://doi.org/10.1109/CLUSTER.2016.21

Balci, M. A., & Akgüller, Ö. (2014a). Average weakly hyperedge domination number for a hypergraph and actor-network application. *International Journal of Modeling and Optimization, 4*(5), 346.

Barbosa, J., Morais, C., Nobrega, R., & Monteiro, A. P. (2005). Static scheduling of dependent parallel tasks on heterogeneous clusters. In *Cluster Computing, 2005. IEEE International* (pp. 1–8). IEEE.

Bishop, B., Kelliher, T. P., & Irwin, M. J. (1999). A detailed analysis of MediaBench. In *Signal Processing Systems, 1999. SiPS 99.* 1999 IEEE Workshop (pp. 448–455). IEEE.

Bouvry, P., de Kergommeaux, J. C., & Trystram, D. (1995). Efficient solutions for mapping parallel programs. *EURO-PAR'95 Parallel Processing* (pp. 379–390). Springer.

Censor, Y., Gordon, D., & Gordon, R. (2001). Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems. *Parallel Computing*, *27*(6), 777–808.

Culler, D. E., Karp, R. M., Patterson, D., Sahay, A., Santos, E. E., Schauser, K. E., … & von Eicken, T. (1996). LogP: A practical model of parallel computation. *Commun. ACM*, *39*(11), 78–85. https://doi.org/10.1145/240455.240477

De Giusti, L. C., Chichizola, F., Naiouf, M., & De Giusti, A. E. (2008). Robustness analysis for the method of assignment MATEHa. *Journal of Computer Science and Technology*, *8*(1), 1-7.

De Giusti, L. C., Chichizola, F., Naiouf, M., Ripoll, A., & De Giusti, A. E. (2007). A model for the automatic mapping of tasks to processors in heterogeneous multi-cluster architectures. *Journal of Computer Science and Technology*, *7*(1), 39-44.

De Giusti, L. C., Naiouf, M., Chichizola, F., Luque Fadón, E., & De Giusti, A. E. (2009). Dynamic scheduling in heterogeneous multiprocessor architectures. In *XV Congreso Argentino de Ciencias de la Computación* (pp. 221-230).

De Giusti, L., Chichizola, F., Naiouf, M., & De Giusti, A. (2008). Mapping tasks to processors in heterogeneous multiprocessor architectures: The MATEHa Algorithm. In *Chilean Computer Science Society, 2008. SCCC '08.* International Conference (pp. 85–91). https://doi.org/10.1109/SCCC.2008.11

De Giusti, L., Chichizola, F., Naiouf, M., De Giusti, A., & Luque, E. (2010). Automatic mapping tasks to cores-evaluating amtha algorithm in multicore architectures. *International Journal of Computer Science Issues, 7*(2), 1-6.

Deveci, M. (2015). *Load-Balancing and task mapping for exascale systems*. The Ohio State University.

Deveci, M., Kaya, K., Ucar, B., & Catalyurek, mit V. (2015). Hypergraph partitioning for multiple communication cost metrics: Model and methods. *Journal of Parallel and Distributed Computing*, *77*, 69–83. https://doi.org/10.1016/j.jpdc.2014.12.002

El-Rewini, H., Lewis, T. G., & Ali, H. H. (1994). *Task scheduling in parallel and distributed systems*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Gallo, G., Longo, G., Pallottino, S., & Nguyen, S. (1993). Directed hypergraphs and applications. *Discrete Applied Mathematics*, *42*(2–3), 177–201.

Gil, D., Hernández, A., Rodriguez, O., Mauri, J., & Radeva, P. (2006). Statistical strategy for anisotropic adventitia modelling in IVUS. *Transactions on Medical Imaging, IEEE 25*(6), 768–778.

Guirado, F., Roig, C., & Ripoll, A. (2013). Enhancing throughput for streaming applications running on cluster systems. *Journal of Parallel and Distributed Computing*, *73*(8), 1092–1105.

Hui, C. C., & Chanson, S. T. (1997). Allocating task interaction graphs to processors in heterogeneous networks. *IEEE Transactions on Parallel and Distributed Systems, 8*(9), 908–925. https://doi.org/10.1109/71.615437

Hwang, J. J., Chow, Y. C., Anger, F. D., & Lee, C. Y. (1989). Scheduling precedence graphs in systems with interprocessor communication times. *SIAM Journal on Computing*, *18*(2), 244–257.

Kalinov, A., & Klimov, S. (2005). Optimal mapping of a parallel application processes onto heterogeneous platform. In *19th IEEE International Proceedings on Parallel and Distributed Processing Symposium, 2005* (pp. 123b–123b). IEEE. https://doi.org/10.1109/IPDPS.2005.310

Kang, Q., He, H., & Wei, J. (2013). An effective iterated greedy algorithm for reliability-oriented task allocation in distributed computing systems. *Journal of Parallel and Distributed Computing*, *73*(8), 1106–1115.

Karypis, G., & Kumar, V. (2000). Multilevel k-way hypergraph partitioning. *VLSI Design*, *11*(3), 285–300.

Karypis, G., Schloegel, K., & Kumar, V. (2003). *ParMeTiS: Parallel graph partitioning and sparse matrix ordering library–Version 3.1*. University of Minnesota.

Kasahara, H., & Narita, S. (1985). Practical multiprocessor scheduling algorithms for efficient parallel processing. *Systems and Computers in Japan*, *16*(2), 11–19. https://doi.org/10.1002/scj.4690160202

Kitajima, J. P., Tron, C., & Plateau, B. (1993). Environments and tools for parallel scientific computing. In J. J. Dongarra & B. Tourancheau (Eds.), *Environments and tools for parallel Scientific computing* (pp. 213–228). Amsterdam, The Netherlands: Elsevier Science Publishers BV. Retrieved from http://dl.acm.org/citation.cfm?id=165125.165280

Kumar, V., Grama, A., Gupta, A., & Karypis, G. (1994). *Introduction to parallel computing: Design and analysis of algorithms*. Redwood City, CA, USA: Benjamin-Cummings Publishing Co., Inc.

Kwok, Y. K., & Ahmad, I. (1996). Dynamic critical-path scheduling: An effective technique for allocating task graphs to multiprocessors. *Transactions on Parallel and Distributed Systems, IEEE7*(5), 506–521.

Kwok, Y. K., & Ahmad, I. (1999). Benchmarking and comparison of the task graph scheduling algorithms. *Journal of Parallel and Distributed Computing*, *59*(3), 381–422. https://doi.org/http://dx.doi.org/10.1006/jpdc.1999.1578

Lam, L., & Suen, C. Y. (1995). An evaluation of parallel thinning algorithms for character recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (9), 914–919.

Lastovetsky, A., & Manumachu, R. R. (2017). New Model-Based Methods and Algorithms for Performance and Energy Optimization of Data Parallel Applications on Homogeneous Multicore Clusters. *IEEE Transactions on Parallel and Distributed Systems, 28*(4), 1119–1133.

Leopold, C. (2001). *Parallel and distributed computing: A survey of models, paradigms and approaches*. New York, NY, USA: John Wiley & Sons, Inc.

Lewis, T. G., & El-Rewini, H. (1992). *Introduction to parallel computing*. New York: Prentice-Hall.

Liu, X., Shi, H., Lu, Q., & Mao, Z. (2007). Visual task-driven based on task precedence graph for collaborative design. In *11th International Conference on Computer Supported Cooperative Work in Design, 2007. CSCWD 2007.* (pp. 246–251). IEEE. https://doi.org/10.1109/CSCWD.2007.4281442

Long, D. L., & Clarke, L. A. (1989). Task interaction graphs for concurrency analysis. *Proceedings of the 11th International Conference on Software Engineering* (pp. 44–52). New York, NY, USA: ACM. https://doi.org/10.1145/74587.74592

Ohtaki, Y., Takahashi, D., Boku, T., & Sato, M. (2004). Parallel implementation of Strassen's matrix multiplication algorithm for heterogeneous clusters. In *18th International Proceedings on Parallel and Distributed Processing Symposium, 2004.* (p. 112). IEEE.

Phinjaroenphan, P., & Bevinakoppa, S. (2004). A novel algorithm for mapping parallel applications in computational grid environments. In *Proceedings of Seventh International Conference on High Performance Computing and Grid in Asia Pacific Region, 2004.* (pp. 347–350). IEEE. https://doi.org/10.1109/HPCASIA.2004.1324056

Rico-Gallego, A. J., Lastovetsky, A., & Diaz-Martin, J. C. (2017). Model-Based estimation of the communication cost of hybrid data-parallel applications on heterogeneous clusters. *IEEE Transactions on Parallel and Distributed Systems*, *9219*(99), 1–1. https://doi.org/10.1109/TPDS.2017.2715809

Roig, C., Ripoll, A., & Guirado, F. (2007). A new task graph model for mapping message passing applications. *Transactions on Parallel and Distributed Systems, IEEE 18*(12), 1740–1753. https://doi.org/10.1109/TPDS.2007.1117

Roig, C., Ripoll, A., Senar, M. A., Guirado, F., & Luque, E. (2000). Modelling message-passing programs for static mapping. In *Proceedings of 8th Euromicro Workshop Parallel and Distributed Processing, 2000.* (pp. 229–236). IEEE. https://doi.org/10.1109/EMPDP.2000.823416

Roig, C., Ripoll, A., Senar, M. A., Guirado, F., & Luque, E. (2002). A new model for static mapping of parallel applications with task and data parallelism. In *Proceedings of International Parallel and Distributed Processing Symposium., IPDPS 2002, Abstracts and CD-ROM* (p. 8). IEEE. https://doi.org/10.1109/IPDPS.2002.1015586

Sadayappan, P., Ercal, F., & Ramanujam, J. (1990). Cluster partitioning approaches to mapping parallel programs onto a hypercube. *Parallel Computing*, *13*(1), 1–16. https://doi.org/http://dx.doi.org/10.1016/0167-8191(90)90115-P

Sanyal, S., & Das, S. K. (2005). MaTCH : Mapping data-parallel tasks on a heterogeneous computing platform using the cross-entropy heuristic. In *19th IEEE International Proceedings on Parallel and Distributed Processing Symposium, 2005* (p. 64b–64b). IEEE. https://doi.org/10.1109/IPDPS.2005.274

Sarkar, V. (1989). *Partitioning and Scheduling parallel programs for multiprocessors*. Cambridge, MA, USA: MIT Press.

Schlag, S., Henne, V., Heuer, T., Meyerhenke, H., Sanders, P., & Schulz, C. (2015). k-way hypergraph partitioning via n-level recursive bisection. In *2016 Proceedings of the Eighteenth Workshop on Algorithm Engineering and Experiments (ALENEX)* (pp. 53-67). Society for Industrial and Applied Mathematics.

Shahid, M., Raza, Z., & Sajid, M. (2015). Level based batch scheduling strategy with idle slot reduction under DAG constraints for computational grid. *Journal of Systems and Software*, *108*, 110–133.

Tekinerdogan, B., & Arkin, E. (2012). *Architecture framework for modeling the deployment of parallel applications on parallel computing platforms*. In *3rd International Conference on Model-Driven Engineering and Software Development (MODELSWARD), 2015* (pp. 185-192). IEEE.

Topcuoglu, H., Hariri, S., & Wu, M. (2002). Performance-effective and low-complexity task scheduling for heterogeneous computing. *IEEE Transactions on Parallel and Distributed Systems, 13*(3), 260–274.

Trifunovic, A., & Knottenbelt, W. J. (2004). A parallel algorithm for multilevel k-way hypergraph partitioning. In *Third International Workshop on Parallel and Distributed Computing, 2004. Third International Symposium on/Algorithms, Models and Tools for Parallel Computing on Heterogeneous Networks, 2004.* (pp. 114–121). IEEE.

Ullman, J. D. (1975). NP-complete scheduling problems. *Journal of Computer and System sciences*, *10*(3), 384–393. https://doi.org/10.1016/S0022-0000(75)80008-0

Upadhyaya, G., & Rajan, H. (2015). Effectively mapping linguistic abstractions for message-passing concurrency to threads on the java virtual machine. *Proceedings of the 2015 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications* (pp. 840–859). New York, NY, USA: ACM. https://doi.org/10.1145/2814270.2814289

Valiant, L. G. (1990). A bridging model for parallel computation. *Communications of the ACM*, *33*(8), 103–111. https://doi.org/10.1145/79173.79181

Xie, G., Zeng, G., Xiao, X., Li, R., & Li, K. (2017). Energy-efficient scheduling algorithms for real-time parallel applications on heterogeneous distributed embedded systems. *IEEE Transactions on Parallel and Distributed Systems, 28*(12), 3426-3442.

Xie, T., & Qin, X. (2005). A new allocation scheme for parallel applications with deadline and security constraints on clusters. In *IEEE International on Cluster Computing, 2005* (pp. 1-10). IEEE.

Yang, B., Guang, L., Säntti, T., & Plosila, J. (2013). Mapping multiple applications with unbounded and bounded number of cores on many-core networks-on-chip. *Microprocessors and Microsystems, 37*(4), 460–471.

Yang, T., & Gerasoulis, A. (1994). DSC: Scheduling parallel tasks on an unbounded number of processors. *IEEE Transactions on Parallel and Distributed Systems, 5*(9), 951–967. https://doi.org/10.1109/71.308533

Review Article

Design of a New Safe Operating Product: Review, Classification and Opportunities

Fatimazahra Guini[1]*, Abdellah El Barkany[1], Abdelouahhab Jabri[1] and El Hassan Irhirane[2]

[1]*Mechanical Engineering Laboratory, Faculty of Science and Techniques, Sidi Mohammed Ben Abdellah University, B.P. 2202 - Route d'Imouzzer – FES, Morocco*

[2]*National Schools of Applied Sciences, Cadi Ayyad University, B.P.575- Guéliz-Boulevard Abdelkrim Al Khattabi – Marrakech - Morocco*

ABSTRACT

The paper presents a literature review on the problem of safety in the design of a new product. To do this, approaches of design and domains that have been integrated into it are firstly presented. Next, diverse methods for predicting and evaluating the safety properties are defined. Then, different approaches that have been developed to solve the problem are discussed, and a synthesis table is given that regroups all the work carried out, allowing a comparative analysis leading to the identification of future research tracks.

Keywords: Availability, design, evaluation, maintainability, product, reliability, security

INTRODUCTION

The decisions taken at the preliminary design of a new product determines 70% of the future costs of the product's life cycle (Salomone, 1995). This phase is, therefore, risky. To manufacture a new product, the company must manage well the risks involved throughout this process. The risk involved in design, according to Sarbacker and Ishii (Sarbacker et al., 1997), results from exposure to a chance of failure due to a lack of understanding of the critical parameters of a decision problem and/or lack of information on some variables relevant to the design problem. Improper decision-making in the design can lead to product failure, security risks, manufacturing problems, waste etc. To reduce design risk, the designer must make advanced decisions that allow easy and reliable comparison between concepts with

regard to design objectives. It is, therefore, necessary to be able to evaluate several concepts and take into account all the constraints of the preliminary design problem. Evaluating the safety of a product during design involves assessment of its reliability, availability, maintainability and security. This article describes the basic tools for the prediction of a safe product as well as the evaluation indicators of this characteristic of safety. The second part deals with the various approaches to resolution, the research methodology adopted and a summary table regrouping all the works already considered in the literature.

Design Process

A product is traditionally designed from a specification according to needs expressed by a customer or a market study. According to Chandrasekaran(Chandrasekaran, 1990), the design process is an activity that solves problems whose input data include the list of functions that the object to be designed must accomplish, a set of constraints that must be satisfied, a set of predefined components and a set of relations connecting these various components. The solution of this problem is chosen from a set of alternatives that follows different evaluation criteria (performance, cost etc.). The literature gives several approaches and design methodologies; the well-known ones are the systematic approach developed by Pahl and Beitz(Pahl & Beitz, 1996), the axiomatic approach described by Suh(Suh, 2001), the ANSI approach EIA-632(ANSI, 1998) and the integrated engineering approach, also known as integrated design or concurrent engineering by Prasad(Prasad, 1997). In all these cases, there is a division into fairly identical phases that stems from the need to provide validation steps; each phase ends with choices and decisions. Integrated engineering is characterised by two main principles: simultaneity and integration. It avoids costly reworks associated with the late recognition of constraints specific to activities that are downstream of the product's life cycle (Bourgeois, 1997). This approach was developed and is often used by designers (See Figure 1).

Design for X (DFX). Designers are guided by several concepts, but these concepts have their own constraints, which may be related to assembly, manufacturing, reliability, environment etc. The concept that presents constraints for designers is discussed in this paper, referred to as 'X'.

In the late 1970s, Boothroyd and Dewhurst(Boothroyd, Murch, & Poli, 1978) and Stoll(Stoll, 1988) developed the design concept for assembling (DFA), which took into account assembling constraints during design phases. Inspired by the DFA, Stoll developed the design concept for manufacturing (DFM) to consider design objectives and manufacturing constraints simultaneously. Kobe(Kobe, 1990) combined the DFA and DFM to form a design for the concept of manufacture and assembly (DFMA). The use of the DFM and the DFA has given tremendous benefits, such as simplifying products, reducing manufacturing costs, improving quality and reducing time to market. Over time, researchers began to think about integrating not only assembly and fabrication but also disassembly and recycling (Zussman, Kriwet, & Seligar, 1994), environment (Korpalski, 1996), quality (Taguchi, 1993), maintainability (Moss, 1985) and reliability (Ireson, Coombs, & Moss, 1988).

Safety in design

Dependability in design is called Defect Science, Risk Analysis or FMDS. It is defined as the ability to maintain the proper functioning of a product and its constituents over time, throughout its life cycle. Assessing the safety of a product during design involves assessment of its reliability, maintainability, availability and safety.

Conducting a safety study when designing a new product has become the practice for industry, given the direct impact on the product and its environment. As a result, this process has been standardised in different fields. Evaluation of the design of products for safety compliance is done by IEC 61508 (International Electrotechnical Commission, 2000) and its derivatives that deal with the safety of electrical/electronic and programmable electronic systems, the ARP-4754 (ARP4754, S. A. E. 1996) standard for the development of aeronautical systems and the ARP-4761, which complements the ARP-4754 (SAE International S-18 Committee, 1996) standard.

Reliability. Reliability is a characteristic of the system expressed by the probability that it performs the function for which it was designed, under given conditions and for a given period of time (NF X 060-010 - AFNOR [1991]). It is the probability that an entity can perform a required function under given conditions during a given time interval [t1, t2]. It is characterised by four criteria: (i) Lifetime probability density f (t), (ii) Reliability at time t r (t), (iii) Lifetime distribution function f (t), and (iv) Failure rate λ (t). The analysis of these criteria by time interval is not easily manipulated for the calculations (derivation and integration), so researchers use mathematical functions such as the probability density functions. These laws may be discrete (Poisson's law) or continuous (normal law, exponential, Weibull). In this article, we are interested in predictive reliability, which is the reliability estimated during the design process, based on a mathematical model and the estimated reliability of the components. Without reliability data, it is difficult to assess reliability. The data sources can be exploited during the design phase, depending on the components integrated into the solution. The components can be separated into two groups (Menye, 2009): i) Identical or similar components that have already been used in other products, whose data are derived from existing system histories, and ii) New components, specially designed for the solution adopted during development. Fail rates can be evaluated either by similarity or by an approach that allows to estimate the parameters influencing the reliability of the product.

The predictive reliability assessment indicator is a probability function, so the analysis of this reliability is carried out using probabilistic approaches, taking into account the dispersion of the design variables. Among the approaches used is the constraint-resistance approach (Lewis, 1987; Murty et al., 1995), which is based on the probabilistic comparison of resistance and constraint, which are considered random variables. The method of first-/second-order approximation (FORM/SORM) is the most efficient in terms of the ratio of computational time and response accuracy; it begins by searching for the design point in variable space, followed by the approximation of boundary-state function by a first-order (FORM) or second-order Taylor expansion (Ouakki, 2011). Another method is the Monte Carlo simulation method, which is a numerical simulation technique based on the calculation of the performance function of a sample representative of the population. This method is very efficient, but it takes up significant computational time (Curbach et al., 2004).

To identify the factors that influence the reliability of a product, the designer may use analytical tools such as: (i) Failure trees and the Ishikawa diagram, which aim to determine component failures that can lead to adverse events (ii) The FMEA product (Bertsche & Lechner, 1999), which is a method of risk prevention during design with the goal of reducing failures and modifying them at a very early stage (iii) The quality house (QFD) (Akao & Mazur, 2003), which allows the targeting of factors necessary to satisfy the expectations of the client.

Availability. Availability is defined by the ability of a component to be able to perform its function under given conditions at a given instance or during a given time interval [T1, T2]. It is expressed by the probability that the system fulfils its required function under certain conditions (AFNOR, 1977). Unlike reliability, the calculation of availability is not limited to non-repairable systems. Calculation of the availability of non-repairable systems is equivalent to the calculation of their reliability. It is expressed as the rate of failure and repair of system components.

Maintainability. According to the NF X 60-010 standards (AFNOR, 1991). maintenance includes all activities intended to maintain or restore an asset in a given state or given security conditions in order to perform a required function. Maintainability is characterised by the probability M (t) that the entity E is in a state at time t to perform its functions, knowing that the entity failed at time 0 (Demri, 2009). The maintainability forecasting approaches available in the literature can be classified into two categories: i) Extrapolation of the maintainability index (Tarelko, 1995), whose forecast is based on feedback from similar products, and ii) The summation of times (Coulibaly, Houssin, & Mutel, 2008; Zwingmann, 2005).

Security. Security is defined as the absence of conditions that guide the impact of a hazard when using a product. The absence of security implies the presence of risk. It is, therefore, necessary to manage risk at the earliest stages of design to obtain a secure product.

Risk management makes it possible to plan, predict and implement the actions to be taken into account in order to avoid risk. The main stages of risk management are (Guillerm, 2011):

a. Risk identification: This is the collection of all types of risk. Possible sources of information for this collection may be consultations, archives or project briefs.

b. Risk assessment and classification: The assessment of the probability and severity of the risks identified in the first phase can reduce their criticality.

c. Risk treatment or acceptance: This step consists of treating risks with high criticality. Acceptance of these risks is defined by the standards according to the activity of the company and the commercial aspects.

d. Consequence analysis and follow-up: Re-examination of the elements impacted by the treatment of the risks to check if new risks have appeared.

After governing against all types of risk, the designer is in a position to choose concepts that have no impact of use or those that have an acceptable degree of risk. Safety indicators can help in this decision-making. Coulibaly et al. (2008) proposed a risk assessment indicator that can be used in the design phase. This indicator is the product of a risk factor FR_{is}, which indicates the existence or not of a risk, and it takes the value 0 in the absence of risk and provides a risk index IR_{is} that allows to qualify and quantify the risk. If $FR_{is} = 1$, the designer

must change its parameters to cancel the value of this flag. If this is not possible, the designer must determine the risk index IR_{is}. Sadeghi (Sadeghi, Mathieu, Tricot, & Al Bassit, 2015) developed risk assessment indicator based on that by Coulibaly, and proposed a more effective indicator of risk that can be applied throughout the design phase.

Methods of Resolution

Several resolution approaches have been developed to address the various safety parameters. Methods of solving safety problems can be classified into two classes, graphical methods and heuristics.

Graphical methods. In the literature, the validity of the safety properties requires the verification of the four categories of elementary graphical conditions i.e. the conditions of connectivity, link, distance and full coupling. Chart blocks have been widely used in the literature to solve this problem, but this tool is not sufficient for validating all the conditions. The conditions of connectivity, link and distance are verifiable by a directed graph. This type of graph represents system variables by vertices and their relationships by arcs. The condition of the complete coupling is verifiable by graphs. A graph is composed of two sets of vertices representing the variables of the system connected by arcs that represent the relations between these variables. The most used graphs for solving the security problem are the Bayesian networks, dynamical Bayesian networks, Markov chains and Petri networks.

Heuristics. A heuristic is an approximated algorithm allowing to identify in time at least one quick, not necessarily optimal, realisable solution. The use of a heuristic is efficient for computing an approximate solution of a given problem. There are three types of heuristics: Constructive heuristics, which generate solutions from an initial solution by trying to gradually add elements to them until a complete solution is obtained; Heuristics with local improvements that start with a complete solution and repetitively try to improve this solution by exploring its neighbourhood, and; Meta-Heuristics, which represent optimisation algorithms that solve difficult optimisation problems for which no more efficient conventional methods are known. They generally represent iterative stochastic algorithms that progress towards an overall optimum. Figure 2 presents the different classes of meta-heuristics.

RESEARCH METHODOLOGY

In order to develop a general bibliography covering various tools and methods for evaluating the safety of new products during design, several databases were examined, including DOAJ, Open J-Gate, IEEE Xplore, Science Direct, Web of Knowledge, Taylor and Francis Online, Springer, Elsevier, IOP science and Scopus. In order to ensure a targeted search, appropriate keywords were chosen to make it possible to enclose problems such as conception, manufacturing, product, security, reliability, maintainability, security, availability, evaluation and decision support. Several combinations of keywords were used to make the results more suitable. The literature is distributed in a good proportion in scientific articles, articles from industry professionals and a few doctoral theses.

LITERATURE REVIEW OF THE SAFETY PROBLEM

Since safety as a topic has attracted much attention given its importance, the number of publications has been gradually increasing over the last few decades (Figure 3). Table 1 summarises all the abbreviations of the graphical and heuristic methods used in this paper. Table 2 summarises the work carried out and the methods of problem solving and provides the names of the authors of the methods.

Analysis and Discussion

The analysis shown in Table 2 highlights that the majority of the work focused on reliability studies. Graphical methods were used to evaluate a single criterion, unlike heuristics. However, it can be seen that graphical methods were used more in the literature than heuristics. The most used graphical method is BN and GA for meta-heuristics. Figure 4 shows that 62.96% of the publications treated reliability problems, 17.59% dealt with the subject of maintainability (disassembly), 9.25% with availability and 10.18% took into account the problem of security.

In terms of resolution methods, the analysis revealed that the BN method represented a percentage of 39.13%, the BN hybrid, 5.79%, the PN method, 8.69% share, the MC, 20.28% and NN, 4.34%. The GA appeared in 15.94% of the documents. The ACO method was used in 2.89% of the documents, while TS and SA were seen in 1.44% of the documents. Figure 5 shows the percentage of publications by types of method relevant to the problem of safety.

CONCLUSION AND PERSPECTIVES

This article represents a general synthesis on ensuring the safe operation of new products from the design stage. We first presented the estimation and prediction tools required and then the methods and approaches available in the literature related to modelling the behaviour of multi-component products. The analysis showed that researchers are most focused on the study of reliability, and have been so since the 1990s. The majority of the problems addressed in this field were solved using modelling by Bayesian networks. If multi-criteria problems are solved, most of them are solved by genetic algorithms. From this perspective, future work will focus on solving safety problems by coupling all its characteristics.
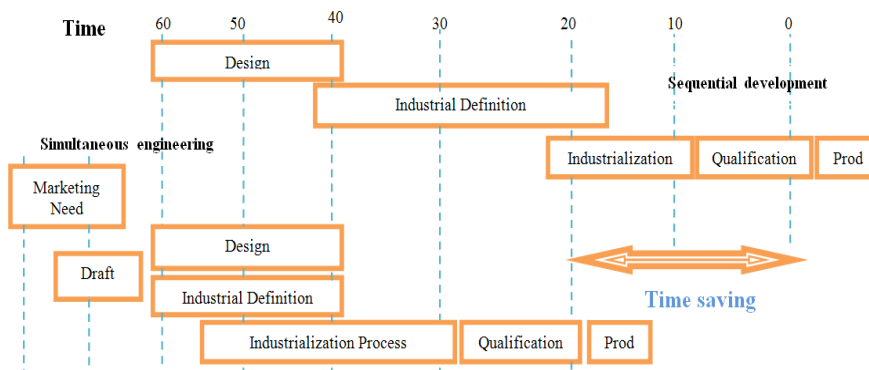
Figure 1. Difference between sequential and simultaneous engineering
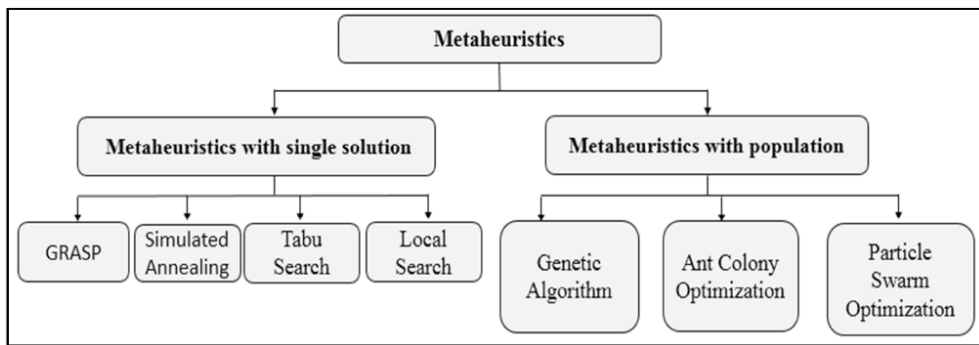
Figure 2. Different approaches to meta-heuristics
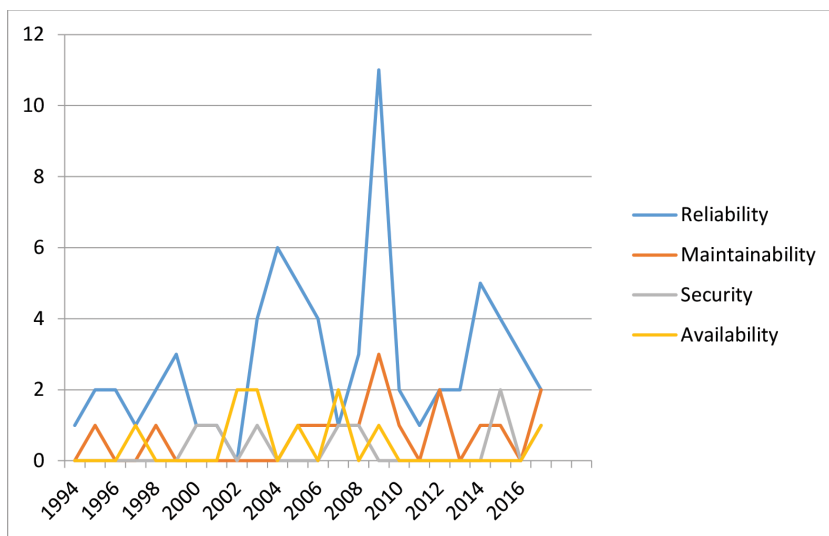
Figure 3. The evolution of publications on the problem
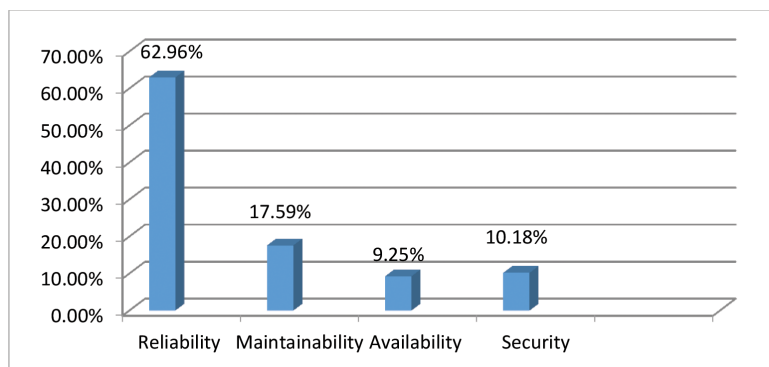
Figure 4. Percentage of safety characteristics treated in the literature
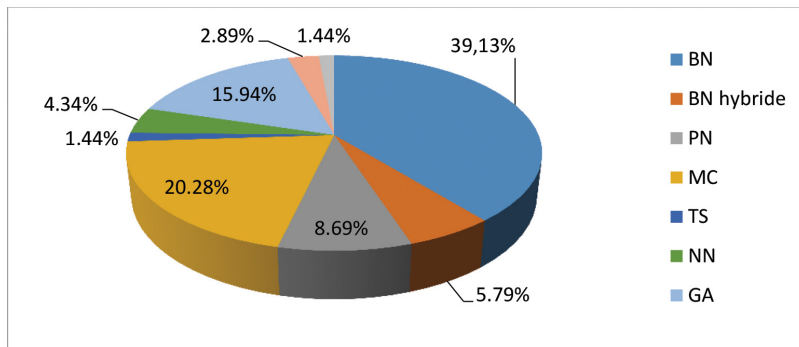
Figure 5. Percentage of methods used

Table 1
Abbreviations of the graphical and heuristic methods

Abbreviation	Approach
BN	Bayesian Network
PN	Petri Network
MC	Markov Chain
GA	Genetic Algorithm
TS	Tabu Search
NN	Neural networks
ACO	Ant Colony Optimisation
SA	Simulated Annealing

Table 2
Summary of works addressing safety issues

Reference	Problem	Approach
Hulting et al. (1994); Tang et al. (1997); Torres-Toledano et al. (1998); Portinale et al. (1999); Castillo et al. (1999); Weber et al. (2003); Guérin et al. (2003); Weber et al. (2004); Graves et al. (2005); Boudali et al. (2005a); Boudali et al. (2005b); Montani et al. (2006); Weber et al. (2006); Simon et al. (2008); Simon et al.(2009); Portinale et al. (2010); Doguc et al. (2009); Guo et al. (2013); Su et al. (2014); Cai et al.(2015); Weber et al. (2015); Mi et al. (2016)	Reliability	BN
Bouissou et al. (1999); Gulvanessian et al. (2001); Qien, (2001); Welch et al. (2000)	Security	
Boudali et al. (2006); Neil et al. (2008); Marquez et al. (2010); Langseth et al. (2009)	Reliability	BN hybrid

Table 2 (*continue*)

Reference	Problem	Approach
Schneeweiss (2001); Volovoi (2004); Charki et al. (2009); Demri (2009)	Reliability	
		PN
Moore et al. (1998); Tang et al. (2006)	Maintainability	
Giraud (2006); Zwingmann (2005); Zaghar et al. (2014); Guilani et al. (2014); Wan et al. (2016)	Reliability	
Timashev et al. (2015); Li et al. (2017); Guilani et al. (2014); Karami-Horestani et al. (2014); Wan et al. (2016)	Reliability and maintainability	MC
Amiri et al. (2007); Sadaghiani et al. (2009); Kayedpour et al. (2017)	Availability	
Tan, (1997)	Reliability and availability	
Bland (1998)	Reliability	TS
Ricotti et al. (1999); Xu et al. (2003)	Reliability	
		NN
Rajpal et al. (2005)	Reliability, availability and maintainability	
Taboada et al. (2008); Rathod et al. (2013); Safari (2012); Coit et al. (1996a); Coit et al. (1996a)	Reliability	
Johansson et al. (2017)	Reliability and security	
Papadopoulos et al. (2005)	Security	GA
Elegbede et al. (2003); Fiori de Castro et al. (2003); Chiang et al. (2007)	Availability	
Liang et al. (2004); Ouiddir et al. (2004)	Reliability	ACO
Chiang et al. (2007)	Availability	SA

REFERENCES

Akao, Y., & Mazur, G. H. (2003). The leading edge in QFD: Past, present and future. *International Journal of Quality and Reliability Management*, *20*(1), 20–35.

Amiri, M., & Ghassemi-Tari, F. (2007). A methodology for analyzing the transient availability and survivability of a system with repairable components. *Applied Mathematics and Computation, 184*(2), 300–307.

ANSI, X. (1998). *9.63. Public key cryptography for the financial sevices industry: Elliptic curve key agreement and transport protocols, draft*. ANSI.

ARP4754, S. A. E. (1996). *Certification considerations for highly-integrated or complex aircraft systems*. SAE, Warrendale, PA.

Bertsche, B., & Lechner, G. (1999). Zuverlässigkeitsanalyse eines Getriebes (Beispiel) [Reliability analysis of a gearbox (example)]. In *Zuverlässigkeit im Maschinenbau* (pp. 104-115). Springer Berlin Heidelberg.

Bland, J. A. (1998). Structural design optimization with reliability constraints using tabu search. *Engineering Optimization*, *30*(1), 55–74.

Boothroyd, G., Murch, L. E., & Poli, C. R. (1978). *Handbook of feeding and orienting techniques for small parts*. Department of Mechanical Engineering, University of Massachusetts, Amherst, Mass., USA.

Boudali, H., & Dugan, J. B. (2005a, January). A new Bayesian network approach to solve dynamic fault trees. In *Annual Proceedings of Reliability and Maintainability Symposium, 2005* (pp. 451–456). IEEE.

Boudali, H., & Dugan, J. B. (2005b). A discrete-time Bayesian network reliability modeling and analysis framework. *Reliability Engineering and System Safety*, *87*(3), 337–349.

Boudali, H., & Dugan, J. B. (2006). A continuous-time Bayesian network reliability modeling, and analysis framework. *IEEE transactions on reliability*, *55*(1), 86–97.

Bouissou, M., Martin, F., & Ourghanlian, A. (1999, January). Assessment of a safety-critical system including software: a Bayesian belief network for evidence sources. In *Annual Proceedings of Reliability and Maintainability Symposium, 1999* (pp. 142–150). IEEE.

Bourgeois, J. P. (1997). *Gestion de projet [Project management]*. Ed. Techniques Ingénieur.

Cai, B., Liu, Y., Ma, Y., Huang, L., & Liu, Z. (2015). A framework for the reliability evaluation of grid-connected photovoltaic systems in the presence of intermittent faults. *Energy*, *93*, 1308–1320.

Castillo, E., Sarabia, J. M., Solares, C., & Gómez, P. (1999). Uncertainty analyses in fault trees and Bayesian networks using FORM/SORM methods. *Reliability Engineering & System Safety*, *65*(1), 29–40.

Chandrasekaran, B. (1990). Design problem solving: A task analysis. *AI Magazine*, *11*(4), 59.

Charki, A., Demri, A., Guerin, F., & Bigaud, D. (2009). Mechatronic system reliability evaluation using Petri networks and PHI2 method. *Reliability, Risk and Safety: Theory and Applications*, *221*, 1587–1601.

Chiang, C. H., & Chen, L. H. (2007). Availability allocation and multi-objective optimization for parallel-series systems. *European Journal of Operational Research*, *180*(3), 1231–1244.

Coit, D. W., & Smith, A. E. (1996a). Penalty guided genetic search for reliability design optimization. *Computers and Industrial Engineering*, *30*(4), 895–904.

Coit, D. W., & Smith, A. E. (1996b). Reliability optimization of series – Parallel systems using a genetic algorithm. *IEEE Transactions on Reliability*, *45*(2), 254–260.

Coulibaly, A., Houssin, R., & Mutel, B. (2008). Maintainability and safety indicators at design stage for mechanical products. *Computers in Industry*, *59*(5), 438–449.

Curbach, M., & Proske, D. (2004). Application of Monte-Carlo-Simulation in structural reliability. In C. Spitzer, U. Schmocker & V. N. Dang (Eds.), *Probabilistic Safety Assessment and Management* (pp. 437-443). Springer, London.

Demri, A. (2009). *Contribution à l'évaluation de la fiabilité d'un système mécatronique par modélisation fonctionnelle et dysfonctionnelle [Contribution to the evaluation of the reliability of a mechatronic system by functional and dysfunctional modeling]*. (Doctoral dissertation). Université d'Angers.

Doguc, O., & Ramirez-Marquez, J. E. (2009). A generic method for estimating system reliability using Bayesian networks. *Reliability Engineering and System Safety*, *94*(2), 542–550.

Elegbede, C., & Adjallah, K. (2003). Availability allocation to repairable systems with genetic algorithms: A multi-objective formulation. *Reliability Engineering and System Safety*, *82*(3), 319–330.

Fiori de Castro, H., & Lucchesi Cavalca, K. (2003). Availability optimization with genetic algorithm. *International Journal of Quality and Reliability Management*, *20*(7), 847–863.

Ginting, R., & Ali, A. Y. (2016). TRIZ or DFMA combined with QFD as product design methodology: A review. *Pertanika Journal of Science and Technology, 24*(1), 1-25.

Giraud, M. (2006). Sûreté de fonctionnement des systèmes: Analyse des systèmes non réparables [Safe operation of systems: Analysis of non-repairable systems] *Techniques de l'ingénieur. Electronique, 5(E3851)*.

Graves, T. L., & Hamada, M. S. (2005). Bayesian methods for assessing system reliability: Models and computation. *Modern Statistical and Mathematical Methods in Reliability*, *10*, 41–54.

Guérin, F., Dumon, B., & Usureau, E. (2003). Reliability estimation by Bayesian method: definition of prior distribution using dependability study. *Reliability Engineering and System Safety*, *82*(3), 299–306.

Guilani, P. P., Sharifi, M., Niaki, S. T. A., & Zaretalab, A. (2014). Reliability evaluation of non-reparable three-state systems using Markov model and its comparison with the UGF and the recursive methods. *Reliability Engineering & System Safety*, *129*, 29–35.

Guillerm, R. (2011). *Intégration de la Sûreté de Fonctionnement dans les Processus d'Ingénierie Système [Integration of Dependability in Systems Engineering Processes]*. (Doctoral dissertation). Université de Toulouse.

Gulvanessian, H., & Holický, M. (2001). Determination of actions due to fire: Recent developments in Bayesian risk assessment of structures under fire. *Progress in Structural Engineering and Materials*, *3*(4), 346–352.

Guo, J., & Wilson, A. G. (2013). Bayesian methods for estimating system reliability using heterogeneous multilevel information. *Technometrics*, *55*(4), 461–472.

Hulting, F. L., & Robinson, J. A. (1994). The reliability of a series system of repairable subsystems: A Bayesian approach. *Naval Research Logistics (NRL)*, *41*(4), 483–506.

International Electrotechnical Commission. (2000). Functional safety of electrical/electronic/ programmable electronic safety related systems. *IEC 61508*.

Ireson, W. G., Coombs, C. F., & Moss, R. Y. (1996). *Handbook of reliability engineering and management*. New York, NY: McGraw-Hill Professional.

Johansson, C., Derelöv, M., & Ölvander, J. (2017). How to use an optimization-based method capable of balancing safety, reliability, and weight in an aircraft design process. *Nuclear Engineering and Technology*, *49*(2), 404–410.

Karami-Horestani, A., Golshan, M. E. H., & Hajian-Hoseinabadi, H. (2014). Reliability modeling of TCR–FC type SVC using Markov process. *International Journal of Electrical Power & Energy Systems*, *55*, 305–311.

Kayedpour, F., Amiri, M., Rafizadeh, M., & Nia, A. S. (2017). Multi-objective redundancy allocation problem for a system with repairable components considering instantaneous availability and strategy selection. *Reliability Engineering & System Safety*, *160*, 11–20.

Kobe, G. (1990). DFMA: Design for manufacture and assembly. *Automotive Industries*, (Mar), 34–38.

Korpalski, T. (1996, May). The role of the "product steward" in advancing design for environment in Hewlett-Packard's computer products organization. In *Proceedings of the 1996 IEEE International Symposium on Electronics and the Environment, 1996, ISEE-1996,* (pp. 37–41). IEEE.

Langseth, H., Nielsen, T. D., Rumí, R., & Salmerón, A. (2009). Inference in hybrid Bayesian networks. *Reliability Engineering and System Safety*, *94*(10), 1499–1509.

Lewis, E. (1994). Introduction to reliability engineering (2nd Ed.). Illinois: John Wiley.

Li, Y., Cui, L., & Lin, C. (2017). Modeling and analysis for multi-state systems with discrete-time Markov regime-switching. *Reliability Engineering and System Safety, 166*, 41-49.

Liang, Y. C., & Smith, A. E. (2004). An ant colony optimization algorithm for the redundancy allocation problem (RAP). *IEEE Transactions on Reliability*, *53*(3), 417–423.

Marquez, D., Neil, M., & Fenton, N. (2010). Improved reliability modeling using Bayesian networks and dynamic discretization. *Reliability Engineering and System Safety, 95*(4), 412–425.

Menye, J. B. (2009). *Validation de la maintenabilité et de la disponibilité en conception d'un système multi-composants [Validation of maintainability and design availability of a multi-component system]*. (Doctoral dissertation). Université Laval.

Mi, J., Li, Y. F., Yang, Y. J., Peng, W., & Huang, H. Z. (2016). Reliability assessment of complex electromechanical systems under epistemic uncertainty. *Reliability Engineering and System Safety*, *152*, 1–15.

Montani, S., Portinale, L., Bobbio, A., & Varesio, M. (2006, January). A tool for automatically translating dynamic fault trees into dynamic Bayesian networks. *Reliability and Maintainability Symposium, 2006. RAMS'06. Annual* (pp. 434–441). IEEE.

Moore, K. E., Gungor, A., & Gupta, S. M. (1998). A Petri net approach to disassembly process planning. *Computers and Industrial Engineering*, *35*(1), 165–168.

Moss, M. A. (1985). *Designing for minimum maintenance expense.* New York: Marcel Dreker Inc.

Neil, M., Tailor, M., Marquez, D., Fenton, N., & Hearty, P. (2008). Modelling dependable systems using hybrid Bayesian networks. *Reliability Engineering and System Safety*, *93*(7), 933–939.

Ouakki, W. (2011). *Prise en compte de la fiabilité en conception [Consideration of reliability in design]*. (Doctoral dissertation). Université Laval.

Ouiddir, R., Rahli, M., Meziane, R., & Zeblah, A. (2004). Ant colony optimization for new redesign problem of multi-state electrical power systems. *Journal of Electrical Engineering-Bratislava*, *55*(3/4), 57–63.

Papadopoulos, Y., & Grante, C. (2005). Evolving car designs using model-based automated safety analysis and optimisation techniques. *Journal of Systems and Software*, *76*(1), 77–89.

Pahl, G., & Beitz, W. (1996). *Engineering design-A systematic approach*. London: Springer-Verlag.

Portinale, L., & Bobbio, A. (1999, July). Bayesian networks for dependability analysis: An application to digital control reliability. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence* (pp. 551-558). Morgan Kaufmann Publishers Inc.

Portinale, L., Raiteri, D. C., & Montani, S. (2010). Supporting reliability engineers in exploiting the power of dynamic Bayesian networks. *International Journal of Approximate Reasoning, 51*(2), 179–195.

Prasad, B. (1997). *Concurrent engineering fundamentals: Integrated product development* (Vol. II). New Jersey, NJ: Saddle River.

Qien, K. (2001). A framework for the establishment of organizational risk indicators. *Reliability Engineering & System Safety*, *74*(2), 147–167.

Rajpal, P. S., Shishodia, K. S., & Sekhon, G. S. (2006). An artificial neural network for modeling reliability, availability and maintainability of a repairable system. *Reliability Engineering and System Safety*, *91*(7), 809–819.

Rathod, V., Yadav, O. P., Rathore, A., & Jain, R. (2013). Optimizing reliability-based robust design model using multi-objective genetic algorithm. *Computers and Industrial Engineering*, *66*(2), 301–310.

Ricotti, M. E., & Zio, E. (1999). Neural network approach to sensitivity and uncertainty analysis. *Reliability Engineering and System Safety*, *64*(1), 59–71.

Sadaghiani, J. S., & Mahtasshami, A. (2009). A methodology for analyzing the transient availability and survivability of a system with every combination of components by using fault tree. *Journal of Applied Sciences*, *9*(6), 1074–1081.

Sadeghi, L., Mathieu, L., Tricot, N., & Al Bassit, L. (2015). Developing a safety indicator to measure the safety level during design for safety. *Safety Science*, *80*, 252–263.

SAE International S-18 Committee. (1996). *ARP4761 guidelines and methods for conducting the safety assessment process on civil airborne system and equipment.* Warrendale, Pennsylvania: Society of Automotive Engineers.

Safari, J. (2012). Multi-objective reliability optimization of series-parallel systems with a choice of redundancy strategies. *Reliability Engineering and System Safety*, *108*, 10–20.

Salomone, T. A. (1995). *What every engineer should know about concurrent engineering*. New York, NY: M. Dekker.

Sarbacker, S. D., & Ishii, K. (1997, September). A framework for evaluating risk in innovative product development. In *Proceedings of the ASME DETC97: ASME Design Engineering Technical Conference* (pp. 1-10). Sacramento, CA.

Schneeweiss, W. G. (2001). Tutorial: Petri nets as a graphical description medium for many reliability scenarios. *IEEE Transactions on Reliability*, *50*(2), 159–164.

Simon, C., Weber, P., & Evsukoff, A. (2008). Bayesian networks inference algorithm to implement Dempster Shafer theory in reliability analysis. *Reliability Engineering and System Safety*, *93*(7), 950–963.

Simon, C., & Weber, P. (2009). Evidential networks for reliability analysis and performance evaluation of systems with imprecise knowledge. *IEEE Transactions on Reliability*, *58*(1), 69–87.

Su, C., & Fu, Y. Q. (2014). Reliability assessment for wind turbines considering the influence of wind speed using Bayesian network. *Eksploatacja i Niezawodność*, *16*(1), 1–8.

Suh, N. P. (2001). *Axiomatic design: Advances and applications* (the Oxford series on advanced manufacturing). New York, NY: Oxford University Press.

Stoll, H. W. (1988). Design for manufacture. *Manufacturing Engineering*, *100*(1), 67–73.

Taboada, H. A., Espiritu, J. F., & Coit, D. W. (2008). MOMS-GA: A multi-objective multi-state genetic algorithm for system reliability optimization design problems. *IEEE Transactions on Reliability*, *57*(1), 182–191.

Taguchi, G. (1993). Taguchi methods Ð cutting the cost of quality. *Quality testing, February*, 84-89.

Tan, Z. B. (1997). Reliability and availability analysis of two-unit warm standby microcomputer systems with self-reset function and repair facility. *Microelectronics Reliability*, *37*(8), 1251–1253.

Tang, J., Tang, K., & Moskowitz, H. (1997). Exact Bayesian estimation of system reliability from component test data. *Naval Research Logistics (NRL)*, *44*(1), 127–146.

Tang, Y., Zhou, M., & Gao, M. (2006). Fuzzy-Petri-net-based disassembly planning considering human factors. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *36*(4), 718–726.

Tarelko, W. (1995). Control model of maintainability level. *Reliability Engineering & System Safety*, *47*(2), 85–91.

Timashev, S. A., & Bushinskaya, A. V. (2015). Markov approach to early diagnostics, reliability assessment, residual life and optimal maintenance of pipeline systems. *Structural Safety*, *56*, 68–79.

Torres-Toledano, J., & Sucar, L. (1998). Bayesian networks for reliability analysis of complex systems. *Progress in Artificial Intelligence – IBERAMIA 98*, 465–465.

Volovoi, V. (2004). Modeling of system reliability Petri nets with aging tokens. *Reliability Engineering and System Safety*, *84*(2), 149–161.

Wan, Y., Huang, H., Das, D., & Pecht, M. (2016). Thermal reliability prediction and analysis for high-density electronic systems based on the Markov process. *Microelectronics Reliability*, *56*, 182–188.

Weber, P., & Jouffe, L. (2003). Reliability modelling with dynamic Bayesian networks. *IFAC Proceedings Volumes*, *36*(5), 57–62.

Weber, P., & Jouffe, L. (2006). Complex system reliability modelling with dynamic object-oriented Bayesian networks (DOOBN). *Reliability Engineering and System Safety*, *91*(2), 149–162.

Weber, P., Munteanu, P., & Jouffe, L. (2004). Dynamic Bayesian networks modelling the dependability of systems with degradations and exogenous constraints. *IFAC Proceedings Volumes, 37*(4), 207–212.

Weber, P., & Simon, C. (2015, March). Méthode de modélisation par Réseaux Bayésien de la fiabilité d'un système multi-états [Bayesian Networks Modeling Method of Reliability of a Multi-State System]. *In 11ème Congrès International Pluridisciplinaire en Qualité, Sûreté de Fonctionnement et Développement Durable, QUALITA 2015.*

Welch, R., & Thelen, T. (2000). Dynamic reliability analysis in an operational context: the Bayesian network perspective. In C. Smidts, J. Devooght & P. E. Labeau (Eds), *Dynamic Reliability: Future Directions* (pp. 277-307). University of Maryland, College Park, MD.

Xu, K., Xie, M., Tang, L. C., & Ho, S. L. (2003). Application of neural networks in forecasting engine systems reliability. *Applied Soft Computing, 2(4), 255–268.*

Zaghar, H., Sallaou, M., & Chaâba, A. (2012). Preliminary design support by integrating a reliability analysis for wind turbine. *Energy and Power Engineering*, *4*(04), 233.

Zussman, E., Kriwet, A., & Seliger, G. (1994). Disassembly-oriented assessment methodology to support design for recycling. *CIRP Annals-Manufacturing Technology, 43*(1), 9–14.

Zwingmann, X. (2005). *Modèle d'évaluation de la fiabilité et de la maintenabilité au stade de la conception [Model for assessing reliability and maintainability at the design stage].* Université Laval.

Review Article

Space Medicine – The Next Frontier for the Heart?

Balasingam, M.

Department of Medicine, Hospital Kajang, Jalan Semenyih, 43000 Kajang, Selangor, Malaysia

ABSTRACT

Researchers have in recent years pointed to microgravity as presenting a unique opportunity for better disease prevention and treatments. Spaceflight can induce many changes in human physiological systems. In particular, the cardiovascular system is especially affected by spaceflight due to changes at the cellular level. Endothelial cells are very sensitive to microgravity. Morphological and functional changes in endothelial cells have been extensively studied since they are believed to be the source of many cardiovascular diseases. Studies have also shown that endothelial cells play a key role in angiogenesis, which can be stimulated in a clinostat-induced microgravity environment. This is a review of studies, based on different research approaches, on human umbilical vein endothelial cells. The myriad molecular cascades and signalling pathways involving gene regulation, proteins, inflammatory response activation, alteration of endothelial behaviour, and cell senescence are highlighted. Age-related disorders experienced on earth are very similar to the changes induced in space by microgravity. As we seek solutions to medical problems, the most innovative and beneficial at present are in space medicines and therapies.

Keywords: Cardiovascular system, cell senescence, endothelial cells, microgravity, molecular cascade, space medicine

BACKGROUND

Research in Microgravity

The term "space research" has become synonymous with space exploration, especially in the field of medicine. Exomedicine refers to research and development of medical solutions in the microgravity environment of space for applications on earth, most of which was conducted on the International Space Station (ISS). Space expeditions can induce many changes in human physiological systems.

The cardiovascular system is affected by space flight, with changes manifesting as cardiac dysrhythmias, cardiac atrophy, orthostatic intolerance, and reduced aerobic capacity (Crawford-Young, 2006). Research in this area will provide knowledge for scientists to develop novel therapies and technologies for the treatment of humans on long-duration space flight missions, as well as for the management of age-related cardiac disorders on earth.

All cells are sensitive to applied or cell-generated mechanical forces, and their transduction into biochemical signals governs many physiological processes. Endothelial dysfunction is a common feature of age-related diseases (Collins, 2011; Vasto, Candore, Balisteri, Caruso & Colonna Romano, 2007). Additionally, pathological conditions, such as cardiovascular deconditioning, are also observed in space. The most likely reasons for these disorders may be the alterations in mechanotransduction (Wu, Fannin, Rice, Wang, & Blough, 2011; Vernikos & Schneider, 2010) since endothelial cells are very sensitive to mechanical forces (Chien, 2007; Chiu & Chien, 2011; Hsiai, Blackman, & Jo, 2010; Kliche, Jeggle, Pavenstadt, & Oberleithner, 2011; Lu & Kassab, 2011; Tzima, Irani-Tehrani, Kiosses, Dejana & Schtulz, 2005; Hirase & Node, 2011; Chien, Li, & Shyy, 1998). Indeed, endothelial cells, which cover the entire inner surface of blood vessels and play a crucial role in maintaining the functional integrity of the vascular wall, are continuously exposed to various hemodynamic forces to which they respond with significant changes in gene expression and protein networks (Mammoto, Mammoto, & Ingber, 2010). Exposure to microgravity triggers all these changes, and endothelial dysfunction is the common denominator. Researchers have studied the endothelial cells lining vessels and conducted deep gene expression and protein analysis. Differentially expressed genes in a microgravity environment promote oxidative stress which, when paired with an increased production of inflammatory proteins, can lead to cardiovascular disease and accelerated cell aging. Therefore, endothelial dysfunction is an important factor for microgravity induced cardiovascular deconditioning.

In this paper, the author reviews current research on endothelial cells in cardiovascular medicine in a microgravity environment, as well as future interpretations and medical therapies.

Endothelial Cell Dysfunction

Endothelial dysfunction in a microgravity environment has been researched extensively in the past (Versari, Longinotti, Barenghi, Maier & Bradamante, 2013). Versari et al., 2013 sought to elucidate the mechanism responsible for endothelial cell dysfunction in space using a genome-wide approach (cDNA microarray) that, together with the evaluation of the cytokine release, may provide a full picture.

In this study, researchers participated in the progress 40 P mission with spaceflight of human umbilical vein endothelial cells (HUVECs), an integrated experiment (SPHINX), which consisted of 12 inflight and 12 ground-based control modules that lasted for 10 days (Figure 1).
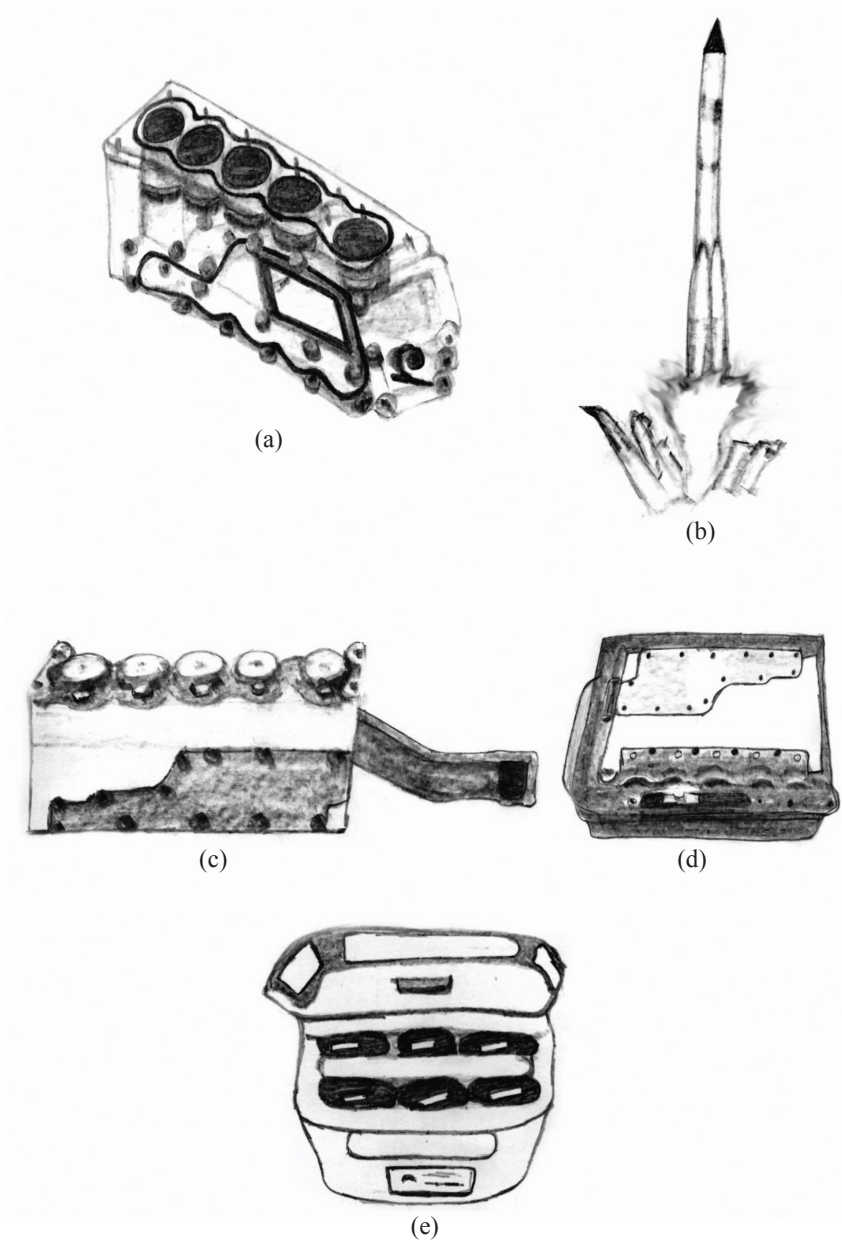
Figure 1. SPHINX experimental set-up. (a) SPHINX experiment unit. (b) Launch of the rocket carrying experiment unit. (c) SPHINX experiment unit close-up view. (d) SPHINX hardware. (e) Soft pouch to transport experimental units

Post-flight microarray analysis revealed 1023 significantly modulated genes, the majority of which were involved in cell adhesion, oxidative phosphorylation, stress responses, cell cycle, and apoptosis. A thioredoxin-interacting protein was the most upregulated, and heat shock proteins 70 and 90 were the most downregulated. Ion channels, mitochondrial oxidative

phosphorylation, and focal adhesion were widely affected. Analysis of cytokines detected in the culture media indicated a significant increase in secretion of interleukin-1α and interleukin 1β. Furthermore, nitric oxide (NO) levels were not affected. In other words, cells exposed to microgravity differentially expressed 1023 genes that altered endothelial behaviour, and promoted oxidative stress, inflammatory responses, and cell senescence.

Molecular and cellular characterisation

Kapitonova et al. (2012) conducted a study on the structures and functions of endothelial cells after a space mission using electron and laser confocal microscopy, image analysis, and MTT metabolic testing. The endothelial cells changed significantly (proliferative activity, size, contours, shape, distribution of mitochondria and microtubules) compared with control cells on Earth. These changes indicated cytoskeleton injuries and impaired barrier function of the cells, which presumably contributed to the development of endothelial dysfunction.

Additionally, morphological and functional differences were observed between endothelial cells originating from large or small vessels and between cells derived from various endothelial beds. These differences were described in terms of different responses to growth factors, expression of unique gene sets under the control of organ-specific regulatory elements, and different conditioning by the extracellular milieu (Monahan-Earley, Dvorak & Aird, 2013; Nolan et al., 2013).

Studies on the response of microvascular endothelial cells to simulated microgravity revealed important differences with macrovascular endothelial cells. No studies have been conducted using microvascular endothelial cells in real space. This led to a ground study by Balsamo et al. (2014) to define the biological and engineering requirements required to investigate cellular and molecular responses of human microvascular endothelial cells exposed to real space flight. This response was evaluated by characterising variations in gene transcription and epigenetic modifications as well as induction of genome damage and cell senescence caused by microgravity.

A recent study showed that treatment with artificial hypergravity reversed the microgravity effects induced in a model of HUVECs, thus showing potential for treating cardiovascular disease by manipulating endothelial cell phenotypes using gravitational forces (Grenon, Jeanne, Aguado-Zuniga, Conte & Hughes-Fulford, 2013). This will further assist in the identification of molecular and cellular pathways activated by space microgravity in microvascular endothelial cells, which can help in the prevention and treatment of cardiovascular disorders.

Effect on cytoskeleton

Because endothelial cells play an important role in the structure and function of the vascular wall and cardiovascular deconditioning has been observed in astronauts. Carlsson, Bertilaccio, Ballabio and Maier (2003) showed that microgravity reversibly stimulated endothelial cell growth. This effect was correlated with the overexpression of heat shock protein 70 and down-regulation of interleukin 1 alpha, a potent inhibitor of endothelial cell growth, and was also

implicated in the promotion of senescence. In addition, gravitationally unloaded endothelial cells rapidly remodelled their cytoskeleton and, after a few days, markedly down-regulated actin through a transcriptional mechanism (Carlsson et al., 2003).

Radiation effect

With advances in technological development, more manned space missions to Mars or prolonged stays on the Moon have been considered. Astronauts may be exposed to ionising radiation, cosmic rays, and solar and trapped radiation in the Van Allen belts. Space radiation can give rise to complex tissue damage that limits cellular repair and recovery (Belli, Sapora & Tabocchini, 2002; Nelson 2016).

Previous studies had examined whether mortality rates due to cardiovascular disease, cancer, accidents, and all other causes of death differed in 1) astronauts who never flew orbital missions in space, 2) astronauts who flew only in low Earth orbit (LEO), and 3) Apollo lunar astronauts, the only humans to have travelled beyond Earth's magnetosphere (Delp, Charvat, Limoli, Globus & Ghosh, 2016). The results revealed no differences in cardiovascular disease mortality rates between non-flight (9%) and LEO (11%) astronauts. However, the cardiovascular mortality rate among Apollo lunar astronauts (43%) was 4–5-fold higher than in non-flight and LEO astronauts. The long-term effect of simulated weightlessness and space-relevant total-body irradiation on vascular responsiveness in mice was also evaluated. The results showed that space-relevant irradiation induced sustained vascular endothelial cell dysfunction. Such impairment is known to lead to occlusive artery disease, and is an important risk factor for cardiovascular disease among astronauts exposed to deep space radiation.

DISCUSSION

Through SPHINX, researchers were able to study aboard the ISS the effects of microgravity on modulating the behaviour of HUVECs (Wu, Fannin, Rice, Wang & Blough, 2011; Vernikos & Schneider, 2010). This is important because microgravity rapidly promotes alterations that are similar to the ones induced by age-related diseases (Biolo, Heer, Narici, & Strollo, 2003; Nichols, Zhang, & Wen, 2006; Wang, 1999). Data showed that HUVECs were affected by spaceflight and responded by utilising mechanosensors responsible for sensing alterations in flow patterns, even though there was no flow in their environment.

Mechanotransduction is a rapidly expanding area of research because conversion of mechanical forces into biochemical signals in cells governs many physiological processes in the human body (Hoffman, Grashof, & Schwartz, 2011). It is widely recognised that defects in mechanotransduction signalling can contribute to human diseases such as immune system disorders, muscular dystrophies, and cardiomyopathies. In the specific case of endothelial cells, several mechanosensors have been proposed, from cell-to-cell adhesion and glycocalyx molecules, to ion channels, integrins, G protein-coupled receptors, and the cytoskeleton (Hsiai, Blackman & Jo, 2010). In space research, the mechanotransduction of flow in various cell types as well as in yeast is mainly observed in the cytoskeleton, which is considered as the main mechanosensor (Hughes-Fulford, 2003; Walther, Bechler, Muller, Hunzinger, & Cogoli, 1996).

Among the 1023 genes studied on the mission, the most upregulated was *TXNIP*. It is an oxidoreductase expressed in endothelial cells that controls the cell's redox status and regulates cell growth, migration, angiogenesis, and apoptosis. The TXNIP plays a key role in age-related and cardiovascular disorders since it is a critical sensor of biomechanical stress (Salminen, Ojala, Kaarniranta, & Kauppinen, 2012; Wang, Nigro, Fujiwara, Yan, & Berk, 2012; Yamawaki, Pan, Lee, & Berk, 2005; Yoshioka, Schulze, Cupesi, Sylvan, & MacGillivray, 2004; Zschauer, Kunze, Jakob, Haendeler, &Altschmied, 2011). Therefore, an overexpression of *TXNIP* can contribute to the modulation of endothelial function in space.

Oxidative stress causes extensive DNA damage and the onset of senescence in endothelial cells (Tretter, Sipos & Adam-Vizi, 2004; Martinon, 2010). The DNA damage activates the p53 signalling pathway that can lead to cell survival if the damage can be repaired or apoptosis if the damage is too severe (Liu, Chen & St. Clair, 2008), which is exactly what was observed in this study. However, the controversial results on the modulation of NO synthases and the insignificant differences found in spaceflight versus NO releases in 1-g are due to adaptation of cells after 10 days in space.

Furthermore, in studies via SPHINX, researchers were able to investigate the mechanisms responsible for alterations in HUVECs in space. They demonstrated that microgravity affected HUVECs, which subsequently responded by using the molecular machinery responsible for sensing alterations in flow patterns, even though no flow was present. It also further supported the role of oxidative stress in accelerating aging and disease.

Similarly, Shi et al., (2012) investigated the effects of microgravity on the ability of endothelial cells to foster angiogenesis. Angiogenesis is an important process by which new blood vessels form from pre-existing endothelial structures (Folkman, 1995). Previous researches have shown that proliferation, migration, and morphological differentiation of endothelial cells play critical roles in angiogenesis. NO, which is synthesized locally in the endothelium, is an important mediator of blood flow control, vascular permeability, and angiogenesis (Namkoong et al., 2008; Sessa, 2009). This study employed the use of a clinostat to simulate microgravity to permit observations of tube formation, migration, and expression of endothelial nitric oxide synthase (eNOS) in HUVECs in this environment. Significant changes were observed within 24 hours of microgravity exposure including the promotion of angiogenesis among HUVECs mediated through the P13K-Akt-Enos signal pathway.

In Malaysia, studies using HUVECs culture optimisation were conducted on microcarriers within fluid processing apparatuses that were flown to the ISS (Muid, Ali, & Nawawi, 2013). Follow-up studies were conducted on live cells after spaceflight. Results revealed that IL-6 and ICAM-1 were still elevated 3 months after space flight. However, there was a downregulation of eNOS expression in revived HUVECs. These results suggest a reduced protection of cells and the surrounding vessels, and thus may lead to hastening of atherosclerosis.

There are differences between the mechanism of cardiac deconditioning and age-related changes in microgravity. The SPHINX study clearly showed that microgravity weakened and jeopardised HUVEC function. As such, new avenues for the treatment of endothelial dysfunction in microgravity are now being researched. In addition, Shi et al. (2012) showed that simulated microgravity can have a strong effect on angiogenesis through a specific signalling pathway. Research is currently in progress to investigate the precise molecular mechanisms

responsible for the effects of simulated microgravity on the P13K signalling pathway. The Malaysian study suggested that the effects of spaceflight are not necessarily immediate after astronauts have returned to earth but may be delayed. This led to new research to explore preventive measures pertaining to atherosclerosis and its complications.

Studies in microgravity revealed the importance of molecular networks controlling signalling pathways and the numerous cascades required for cell survival. This is important in regards to planning for management of heart conditions in future space missions as well as space spin-off technologies on earth such as the tele-echocardiography machine, which was modified by the European Space Agency to study astronauts' hearts in space (Knight, 2016). The technology is now being used successfully in hospitals across the world. Looking to the future, novel innovations can be made using information from these molecular roadmaps. Thus, the study of the cardiovascular system in microgravity contributes enormously to knowledge about our future cardiac health.

CONCLUSION

In summary, human life have revolved around gravity on earth, and without it, life would be a very different experience. As such, it is inevitable that we devise methods to improvise and sustain space travellers during long journeys in space while also developing new medical technologies to benefit mankind as a whole.

REFERENCES

Balsamo, M., Barravecchia, I., Mariotti, S., Merenda, A., De Cesari, C., Vukich, M., & Angeloni, D. (2014). Molecular and cellular characterization of space flight effects on microvascular endothelial cell function- preparatory work for the SFEF project. *Microgravity Science and Technology*, *26*(6), 351-363.

Belli, M., Sapora, O., & Tabocchini, M. A. (2002). Molecular targets in cellular response to ionizing radiation and implications in space radiation protection. *Radiation Research*, *43*(Suppl), S13-S19.

Biolo, G., Heer, M., Narici, M., & Strollo, F. (2003). Microgravity as a model of ageing. *Current Opinion in Clinical Nutrition and Metabolic Care, 6*(1), 31-40.

Carlsson, S. I. M., Bertilaccio, M. T. S., Ballabio, E., & Maier, J. A. M. (2003). Endothelial stress by gravitational unloading: effects on cell growth and cytoskeletal organization. *Biochimica et BiophysicaActa, Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1642*(3), 173-179.

Chien, S. (2007). Mechanotransduction and endothelial cell homeostasis: the wisdom of the cell. *American Journal of Physiology-Heart and Circulatory Physiol*ogy, *292*(3), H1209-H1224.

Chien, S., Li, S., & Shyy, J. Y. J. (1998). Effects of mechanical forces on signal transduction and gene expression in endothelial cells. *Hypertension, 31*(1), 162-169.

Chiu, J. J., & Chien, S. (2011). Effects of disturbed flow on vascular endothelium: pathophysiological basis and clinical perspectives. *Physiological Reviews, 91*(1), 327-387.

Collins, C., & Tzima, E. (2011). Hemodynamic forces in endothelial dysfunction and vascular aging. *Experimental Gerontology, 46*(2-3), 185-188.

Crawford-Young, S. J. (2006). Effects of microgravity on cell cytoskeleton and embryogenesis. *International Journal of Developmental Biology, 50*(2-3), 183-191.

Delp, M. D., Charvat, J. M., Limoli, C. L., Globus, R. K., & Ghosh, P. (2016). Apollo lunar astronauts show higher cardiovascular disease mortality: possible deep space radiation effects on the vascular endothelium. *Scientific Reports*, *6*, 29901.

Folkman, J. (1995). Angiogenesis in cancer, vascular, rheumatoid and other disease. *Nature Medicine, 1*(1), 27-31.

Grenon, S. M., Jeanne, M., Aguado-Zuniga, J., Conte, M. S., & Hughes-Fulford, M. (2013). Effects of gravitational mechanical unloading in endothelial cells: association between caveolins, inflammation and adhesion molecules. *Scientific Reports*, *3*, 1494.

Hirase, T., & Node, K. (2011). Endothelial dysfunction as a cellular mechanism for vascular failure. *American Journal of Physiology-Heart and Circulatory Physiol*ogy, *302*(3), H499-H505.

Hoffman, B. D., Grashof, C., & Schwartz, M. A. (2011). Dynamic molecular processes mediate cellular mechanotransduction. *Nature, 475*(7356), 316-323.

Hsiai, T. K., Blackman, B., & Jo, H. (2010*). Hemodynamics and Mechanobiology of Endothelium* (pp.1-246). Singapore: World Scientific.

Hughes-Fulford, M. (2003). Function of the cytoskeleton in gravisensing during spaceflight. *Advance in Space Research, 32*(8), 1585-1593.

Kapitonova, M. Y., Kuznetsov, S. L., Froemming, G. R. A., Muid, S., Nor-Ashikin, M. N. K., Otman, S., ... & Nawawi, H. (2013). Effects of space mission factors on the morphology and function of endothelial cells. *Bulletin of Experimental Biology and Medicine*, *154*(6), 796-801.

Kliche, K., Jeggle, P., Pavenstadt, H., & Oberleithner, H. (2011). Role of cellular mechanics in the function and life span of vascular endothelium. *Pflugers Archive*, *462*(2), 209-217.

Knight, H. (2016). *Sound concept: medical spin-off from ISS technology, The Engineer*. Retrieved from https://www.theengineer.co.uk/sound concept medical spin-off- from- iss- technology

Liu, B., Chen, Y., & St. Clair, D. K. (2008). ROS and p53: a versatile partnership. *Free Radical Biology and Medicine*, *44*(8), 1529-1535.

Lu, D., & Kassab, G. S. (2011). Role of shear stress and stretch in vascular mechanobiology. *Journal of Royal Society Interface, 8*(63), 1379-1385.

Mammoto, A., Mammoto, T., & Ingber, D. E. (2012). Mechanosensitive mechanisms in transcriptional regulation. *Journal of Cell Science, 125*(13), 3061-3073.

Martinon, F. (2010). Signalling by ROS drives inflammasome. European *Journal of Immunology, 40*(3), 616-619.

Monahan-Earley, R., Dvorak, A. M., & Aird, W. C. (2013). Evolutionary origins of the blood vascular system and endothelium. *Journal of Thrombosis and Heamostasis*, *11*(s1), 46-66.

Muid, S., Ali, A. M., & Nawawi, H. (2013). Interleukin-6 and Intercellular cell adhesion molecule-1 expression remains elevated in revived live endothelial cells following spaceflight. *Malaysian Journal of Pathology*, *35*(2), 165-176.

Namkoong, S., Chung, B. H., Ha, K. S., Lee, H., Kwon, Y. G., & Kim, Y. M. (2008). Microscopic technique for the detection nitric-oxide-dependent angiogenesis in an animal model. *Methods in enzymology, 441*, 393-402.

Nelson, G. (2016). Space radiation and human exposures, a primer. *Radiation Research*, *185*(4), 349-358.

Nichols, H. L., Zhang, N., & Wen, X. (2006). Proteomics and genomics of microgravity. *Physiological Genomics*, *26*(3), 163-171.

Nolan, D. J., Ginsberg, M., Israely, E., Palikuqi, B., Poulos, M. G., James, D., ... & Butler, J. M. (2013). Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration. *Developmental Cell*, *26*(2), 204-219.

Salminen, A., Ojala, J., Kaarniranta, K., & Kauppinen, A. (2012). Mitochondrial dysfunction and oxidative stress activate inflammasomes: impact on the aging process and age-related diseases. *Cellular and Molecular Life Science*, *69*(18), 1-15.

Sessa, W. C. (2009). Molecular control of blood flow and angiogenesis: role of nitric oxide. *Journal of Thrombosis and Haemostasis, 7*(s1), 35-37.

Shi, F., Wang, Y. C., Zhao, T. Z., Zhang, S., Du, T. Y., Yang, C. B., … & Sun, X. Q. (2012). Effects of simulated microgravity on human umbilical vein endothelial cell angiogenesis and role of the P13K-Akt-eNOS signal pathway. PloS ONE, *7*(7), e40365.

Tretter, L., Sipos, I., & Adam-Vizi, V. (2004). Initiation of neuronal damage by complex I deficiency and oxidative stress in Parkinson's disease. *Neurochemical Research, 29*(3), 569-577.

Tzima, E., Irani-Tehrani, M., Kiosses, W. B., Dejana, E., & Schtulz, D. A. (2005). A mechanosensory complex that mediates the endothelial cell response to fluid shear stress. *Nature, 437*(7057), 426-431.

Vasto, S., Candore, G., Balistreri, C. R., Caruso, M., & Colonna Romano, G. (2007). Inflammatory networks in ageing, age related diseases and longevity. *Mechanisms of Ageing and Development, 128*(1), 83-91.

Vernikos, J., & Schneider, V. S. (2010). Space, gravity and the physiology of aging: parallel or convergent disciplines? A mini review. *Gerontology, 56*(2), 157-166.

Versari, S., Longinotti, G., Barenghi, L., Maier, J. A. M. & Bradamante, S. (2013). The immune system in space: are we prepared? *The FASEB Journal, 27*, 4466-4475.

Walther, I., Bechler, B., Muller, O., Hunzinger, E., & Cogoli, A. (1996). Cultivation of Saccharomyces cerevisiae in a bioreactor in microgravity. *Journal of Biotechnology, 47*(2-3)*,* 113-127.

Wang, E. (1999). Age dependent atrophy and microgravity travel: what do they have in common? *The FASEB Journal, 13*(Suppl), S167-174.

Wang, X. Q., Nigro, P., Fujiwara, K., Yan, C., & Berk, B. C. (2012). Thioredoxin interacting protein promotes endothelial cell inflammation in response to disturbed flow by increasing leucocyte adhesion and repressing Kruppel-like factor 2. *Circulation Research, 110*, 560-568.

Wu, M., Fannin, J., Rice, K. M., Wang, B., & Blough, E. R. (2011). Effect of aging on cellular mechanotransduction. *Ageing Research Reviews, 10*(1), 1-15.

Yamawaki, H., Pan, S., Lee, R. T., & Berk, B. C. (2005). Fluid shear stress inhibits vascular inflammation by decreasing thioredoxin-interacting protein in endothelial cells. *Journal of Clinical Investigation, 115*(3), 733-738.

Yoshioka, J., Schulze, P. C., Cupesi, M., Sylvan, J. D., & MacGillivray, C. (2004). Thioredoxin-interacting protein controls cardiac hypertrophy through regulation of thioredoxin activity. *Circulation, 109*(21), 2581-2586.

Zschauer, T. C., Kunze, K., Jakob, S., Haendeler, J., & Altschmied, J. (2011). Oxidative stress induced degradation of thioredoxin-1 and apoptosis is inhibited by thioredoxin-1-actin interaction in endothelial cells. *Arteriosclerosis, Thrombosis and Vascular Biology, 31*(3), 650-656.

Evaluation of Single Reservoir Performance for Flood Risk Reduction Using a Developed Simulation Model: Case Study of Makhoul Reservoir

Yousif H. Al-Aqeeli[1,2]*, S. Abd Aziz[2], Badronnisa Yusuf[3] and Aimrun Wayayok[2]

[1]*Department of Dams and Water Resources Engineering, Faculty of Engineering, Mosul University, Iraq*

[2]*Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

[3]*Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

The objectives of this study were firstly, to develop a simulation model (SM) for a single reservoir to identify the standard operating policy (SOP) of a reservoir based on a monthly operating period, and secondly, to evaluate the performance of the proposed Makhoul reservoir using a Developed Simulation Model (DSM) in reducing flood risk. This reservoir is located on the River Tigris, approximately 180 km upstream of Baghdad, Iraq. The performance of the reservoir in reducing flood risk was evaluated using two designs and records of flood waves gathered over two years. The first design was the present one, while the second was developed by increasing the operational storage to its maximum, based on the digital maps of the region. The flows downstream of the reservoir were compared, with and without the reservoir in the two years in question. Four parameters resulting from the two designs were compared: storage, surface area, elevation and power. The results suggested that the reservoir would be ineffective in reducing flood risk, but it would have the ability to provide hydroelectric power using the two designs, with the new one showing better ability at doing this. The reservoir can also serve purposes such as irrigation, fish wealth development and recreation. This DSM proved its effectiveness in evaluating the performance of the single storage system used for reservoirs.

Keywords: Flood risk, Iraq, Makhoul reservoir, operation policy, River Tigris, simulation model, Tharthar Reservoir

INTRODUCTION

The hydraulic system of dams is designed to achieve many purposes such as reducing flood risk, irrigation, development of fish wealth

and recreation. The design of these hydraulic aspects depends on the observed flows of the river during a specific period. Performance evaluation of the reservoir in the design stage is necessary. One of many software programmes designed to determine and simulate the operation of reservoirs can be used, or an optimisation-simulation model can be created with the ability to evaluate the functions assigned to it.

Many studies have been conducted to determine the optimal operation policy for reservoirs in order to reduce flood risk and for other purposes. Malekmohammadi, Zahraie and Kerachian (2010) developed a new methodology for the management of two reservoirs in real-time floods to minimise flood damage in downstream rivers. The methodology developed has been applied to the Bakhtiari and Dez river-reservoir systems in the southwest of Iran. The results indicated that the proposed models can be efficiently used for flood management and real-time operation of the cascade-river reservoir systems. Liu, Guo, Liu, Chen and Li (2011) derived optimal refill rules for a multi-purpose reservoir, with the objective function of maximising benefits under the condition of flood control safety. A multi-objective refill operation model was proposed, combining flood control and preservation. China's Three Gorges Reservoir was chosen as a case study for application of the proposed model. A multi-objective simulation-optimisation approach was proposed by Richaud, Madsen, Rosbjerg, Pedersen and Ngo (2011) to operate the Hoa Binh reservoir in Vietnam through off-line rule curve optimisation coupled with on-line real-time optimisation. The reservoir operating rules were optimised by studying the trade-offs between control of flooding and hydropower production on the one hand and providing water for irrigation on the other. Bayat, Mousavi and Namin (2011) combined a particle swarm optimisation (PSO) algorithm and a simulation model of river flood routing to derive the optimal operation of river-reservoir systems under flooding conditions. In this case minimising flood damage in downstream areas was the objective function. Bishop Dam, which is a benchmark problem in the HEC-5 software, has been used to test the developed optimisation-simulation models as a case study of flood control operation. Talukdar, Deb and Srivastava (2011) considered multi-objective functioning, which has been represented by first, minimisation of downstream area inundation and second, maximisation of conservation benefits to find the optimal operation policy for reservoirs. The model developed has been used to find the optimal releases for the Sardar Sarovar Project (SSP) reservoir, a multipurpose reservoir on the Narmada River in Central India. The Reservoir Optimisation-Simulation with Sediment Evacuation (ROSSE) model was applied by Khan, Babel, Tingsanchali, Clemente and Luong (2012) with the aim of minimising irrigation deficiency in the Tarbela Reservoir, Pakistan. The simulation results of three groups of rule curves, one existing and two optimised groups, have been compared to some parameters like irrigation deficiency, power generation, sediment evacuation and flood damages. Ngoc, Hiramatsu and Harada (2014) determined the optimal rule curves of reservoir operation based on a multi-use reservoir system. Dau Tieng Reservoir was used as a case study. This reservoir, located on the upper Saigon River in southern Vietnam, is used for multiple purposes (flood control, domestic and industrial demands, flushing out salt water intrusion from the downstream area and agricultural irrigation).

In the present study, a simulation model (SM) was developed to determine a standard operating policy (SOP) in order to evaluate the performance of a single reservoir in reducing flood risk. This Developed Simulation Model (DSM) determines the decision variables that

represent the monthly releases and their counterparts of state variables, which represent the monthly storage. In this way, the model used the monthly requirements as the initial outflow, while surplus water was added to the initial outflow when a spill occurred. This DSM was applied to the Makhoul Reservoir located on the River Tigris in the northern area of Iraq; this area is in the design and planning stage. Recorded flood waves that occurred in two different years, 1968-1969 and 1987-1988, were used in the DSM to evaluate the reservoir performance in protecting Baghdad from flood risk using the SOP. The Makhoul Reservoir was evaluated twice, first using its current design and second using the new design. This new design used a new storage option, which represents the maximum operational storage. This maximum was identified based on digital maps.

METHOD

Development of Simulation Model

The simulation model of the single storage system formulated by Al-Aqeeli, Al Mohseen, Lee and Abd Aziz (2015) using Simulink in Matlab was developed to determine the SOP for the operation of the Makhoul Reservoir based on a monthly operating period. Water release using this operating policy achieves the water requirement downstream of the reservoir, and spillage will occur if the storage in the reservoir exceeds the capacity of the reservoir (Loucks, Van Beek, Stedinger, Dijkman, & Villars, 2005). This means that the total monthly outflow includes the monthly water requirement in addition to the surplus water in the reservoir, if any. In each period, the model first computes the storage according to the monthly requirement as the initial outflow as shown in the first formula of the water balance equation (Eq. 1). The monthly requirement is shown in Figure 1. If the calculated storage exceeds the capacity of the reservoir, spillage will occur, and its quantity will be added to the initial outflow as the total outflow as shown in Equation 2. This total outflow is used in the second water balance equation (Eq. 3) to recalculate the storage (Loucks et al., 2005). It is worth mentioning here that the quantity of spillage is released first from the remaining capacity of the outlets leading to the hydropower stations, if available, where the required water is released first. After that, it is released from the bottom outlets and finally, from the spillway. As is clear, this model separates the total monthly outflow to three types of release, relying on the capacity of the reservoir outlets.

$$S_{t+1} = S_t + I_t - D_t - Ev_t + Pr_t \qquad [1]$$

S_{t+1} Storage at beginning of time period (t+1)
S_t Storage at beginning of time period (t)
I_t Inflow during time period (t)
D_t Requirement during time period (t)
Ev_t Evaporation in time period (t)
Pr_t Precipitation in time period (t)
t Time period (month)

$$R_t = D_t + SP_t \tag{2}$$

R_t Total outflow during time period (t)

SP_t Spill during time period (t)

$$S_{t+1} = S_t + I_t - R_t - Ev_t + Pr_t \tag{3}$$

Note: The units used in these equations are millions of cubic metres (MCM).

Storage at the beginning of the operation (S_1) is inserted equal to the minimum operation storage of the reservoir. The Makhoul reservoir was evaluated twice: first using the present design and then using a new design. The new design used the maximum possible operational storage. This new storage was determined according to the digital maps for reservoir location.
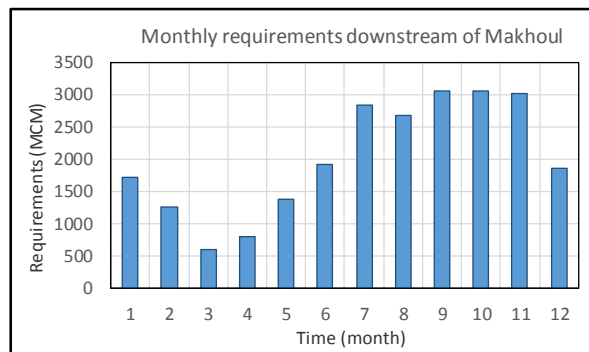
Figure 1. Monthly requirement downstream of Makhoul reservoir

The power generated per month $(P_t),$ measured by megawatt, is calculated as shown in Equation 4, assuming that the efficiency of the power plant (η) equals (80%) (Al-Aqeeli et al., 2015).

$$P_t = k * RP_t * H_t \tag{4}$$

k = Constant (0.003)

RP_t = Release from the tunnels connected to the hydroelectric station (MCM)

H_t = Average head in the time period (metre)

Brief History of the Study Area

Baghdad has suffered from frequent flooding of the River Tigris since ancient times. In the first half of the last century, several measures were taken to protect the city from flooding, including the establishment of soil barriers along the banks of the Tigris. These measures contributed to reducing the damage caused by the flood relatively, but not categorically. Another measure taken to protect Baghdad from flooding was the diversion channel opened in 1956 at

the Samarra barrage, located on the Tigris upstream of Baghdad. This barrage diverts surplus water from the Tigris to the Tharthar Reservoir during flood periods. The Tharthar Reservoir is a natural trough located between the Tigris and the River Euphrates, as shown in Figure 2. Surplus water from the Tharthar Reservoir is diverted to the Euphrates through a drainage channel. The main problem with using the Tharthar Reservoir is the high salinity of its soil (Salih, Kadim, & Qadir, 2012), which negatively affects the quality of the water diverted from it to the Euphrates. The Makhoul Reservoir was suggested for use instead of the Tharthar Reservoir, and is being constructed on the Tigris upstream of the Samarra barrage to protect Baghdad from flooding. The Makhoul Reservoir will store water released from three reservoirs, Mosul, Dokan and Bekhma, which are located upstream of its site during the flooding seasons, and manages the release of the water, among other functions. It is worth mentioning that the Mosul and Dokan Reservoirs are working effectively now, while the Bekhma Reservoir is still in the construction stage.

Figure 2. Map of the study area (UN-ESCWA and BGR, 2013)

Makhoul Reservoir: Under Planning

It was proposed that the Makhoul Reservoir be built across the Tigris River around 10 km north of the Fatha station, as shown in Figure 2. The main purpose for building the Makhoul Reservoir was to store water released from the three reservoirs, Mosul, Dokan and Bekhma,

located upstream of it, during the flooding season and for the subsequent timely release of the water when needed. The water will be used to generate hydropower and to meet requirements downstream of the reservoir within the permissible limits. The maximum capacity of the hydropower station is 316 MW. One of the major determinants related to the quantity of release from Makhoul is that the water should not exceed 12,000 m^3/sec. From this release, 9000 m^3/sec passes to the Tharthar Reservoir and the remaining 3000 m^3/sec passes through the Samarra barrage in order to protect Baghdad from the risk of flood. The elevation and volumes of storage in the Makhoul Reservoir are shown in Table 1, while Table 2 shows the capacity of the different outlets of the reservoir.

Table 1
The elevation of storage and the volumes of storage of the makhoul reservoir

Minimum level of operational storage (MASL)	140
Maximum level of operational storage (MASL)	150
Crest level (MASL)	155
Minimum operational storage of reservoir (MCM)	744
Maximum operational storage of reservoir (MCM)	2222

Table 2
The capacity of different outlets of makhoul reservoir

Tunnels of generating units (m^3/sec)	Bottom outlet (m^3/sec)	Spillway (m^3/sec)
1235	2000	8400

Storage Volume of Makhoul Reservoir

As mentioned previously, the Makhoul reservoir is currently at the planning and design stage, so its present maximum operational storage could be increased to the maximum possible. In this study, this maximum operational storage was determined. The differential between the maximum and minimum operational storage for the present design was increased by multiplying this differential by increment factors. The results were then added to the minimum operational storage while keeping the other design features unchanged. The corresponding elevation and surface areas of the storage were obtained by extrapolation using the elevation-area-storage

curve of the reservoir. In order to identify the appropriate elevation of storage in the new design of the reservoir, all of the new elevations were checked using the digital maps of the reservoir location. The space between the obtained elevation and the highest contour line without overflow was calculated. After that, the suitable elevation was identified.

Application of the DSM

The DSM was applied using two designs for the Makhoul Reservoir, the present design and the new one. The difference between the two designs is the variation in the volume of the maximum operational storage. In each design, the recorded monthly flow during two water years, 1968-1969 and 1987-1988, was adopted in the DSM; two flood waves were recorded in the two years mentioned. The water year in Iraq begins in October and ends in September the following year. The recorded flow was considered the inflow to the reservoir in the DSM. The DSM was initialised to suit both designs to determine the SOP for the reservoir. For each design, the DSM was operated independently in the two years. For each year, the performance of operating the Makhoul Reservoir using the two designs was evaluated by comparing the obtained results. These comparisons included firstly, that of the monthly storage, elevation, surface area of storage and power generation, and secondly, the flow at the Fatha Station before and after the existence of the Makhoul Reservoir. In comparing the flow, the observed flow at the Fatha Station before the existence of the Makhoul Reservoir was compared with the expected outflow of the Makhoul Reservoir when it is built. The Makhoul Reservoir will use the flow observed at the Fatha Station as its inflow.

RESULTS AND DISCUSSION

New Design of Makhoul Reservoir

The new design of the Makhoul Reservoir was identified as described in the methodology above. By increasing the amount of the differential between the maximum and minimum operational storage of the Makhoul Reservoir, three design features were identified. These features included storage, elevation and surface area, as shown in Table 3. The results were then added to the minimum operational storage while the other design features were unchanged. According to the digital maps of the location of the reservoir shown in Figure 3, the highest contour line without an overflow was identified. This contour line was 173 MASL. The space between the obtained elevation and this contour line was identified, as shown in Table 3. From this table, the appropriate elevation was chosen, and thus, the corresponding storage. The suitable elevation and storage represent the maximum possible operational elevation and storage, which were equalled to 162.4 MASL and 5178 MCM, respectively, as shown in Table 3 and Figure 3.

Table 3

Operational storage with corresponding elevation and surface area for new volume of makhoul reservoir

1	2	3	4	5	6
The differential between the maximum and minimum operational storage in the current design (MCM)	Increment factors to increase the differential between maximum and minimum operational storage	New operational storage (MCM) identified from multiplying columns 1 & 2 and then adding the results to the minimum operational storage	New operational elevation of storage (MASL) identified from the Storage-Elevation curve using column 3 (by extrapolation)	New surface areas of storage (km^2) identified from Storage-Area curve using column 3 (by extrapolation)	Space between the new operational elevation of storage identified in column 5 and the contour line 173 (m)
1478	1	2222	150.00	195.00	23
1478	1.5	2961	153.48	231.30	19.52
1478	2	3700	156.50	262.00	16.5
1478	2.5	4439	159.45	291.56	13.55
1478	3	5178	162.41	321.12	10.59
1478	3.5	5917	165.36	350.68	7.64
1478	4	6656	168.32	380.24	4.68
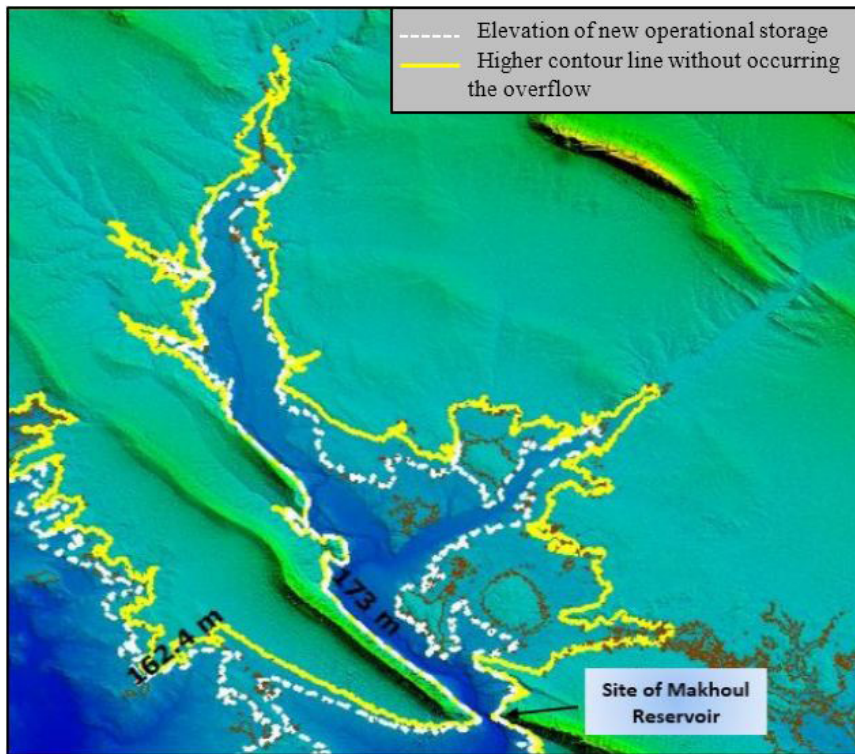
Figure 3. The higher contour line without overflow and the elevation of the new operational storage

Performance Evaluation of the Two Designs of Makhoul Reservoir

The DSM was used in order to evaluate the performance of the Makhoul Reservoir in reducing the flood risk using two designs, the present and the new designs. Recorded flow of two water years, 1968-1969 and 1987-1988, was used separately in this evaluation, when two flood waves were observed. According to the application of the DSM using the two designs (present and new designs) in the two operating years, four features of reservoir design were compared. These features included monthly storage, elevation, surface area and power, as shown in Figures 4 and 5 for the years 1968-1969 and 1987-1988, respectively. From these figures, it can be observed that the reservoir was full during most of the months of the years for both designs. This led to maximum production of hydropower as the outlets of hydroelectric power were full. In addition, the maximum operational storage and its corresponding surface area in the new design increased by 133% and 64.67%, respectively, and the maximum operational level was raised by 8.27 m. In addition, the maximum hydropower generation was increased by 42.8% using the new design.

The flow at the Fatha Station, with and without the Makhoul Reservoir, was compared in the water years 1968-1969 and 1987-1988 using the present and new designs, as shown in Figure 6 i.e. for each year, three flows at the Fatha Station were compared: the first was the recorded flow, the second, the release of the Makhoul Reservoir using the present design, and the third, the release of the Makhoul Reservoir using the new design. From these figures, it

can be seen that the inflow of the reservoir was almost equal to the outflow during most of the months of the two years. This occurred because the capacity of the reservoir was insufficient for accommodating the incoming flow of the Tigris River in the two years.
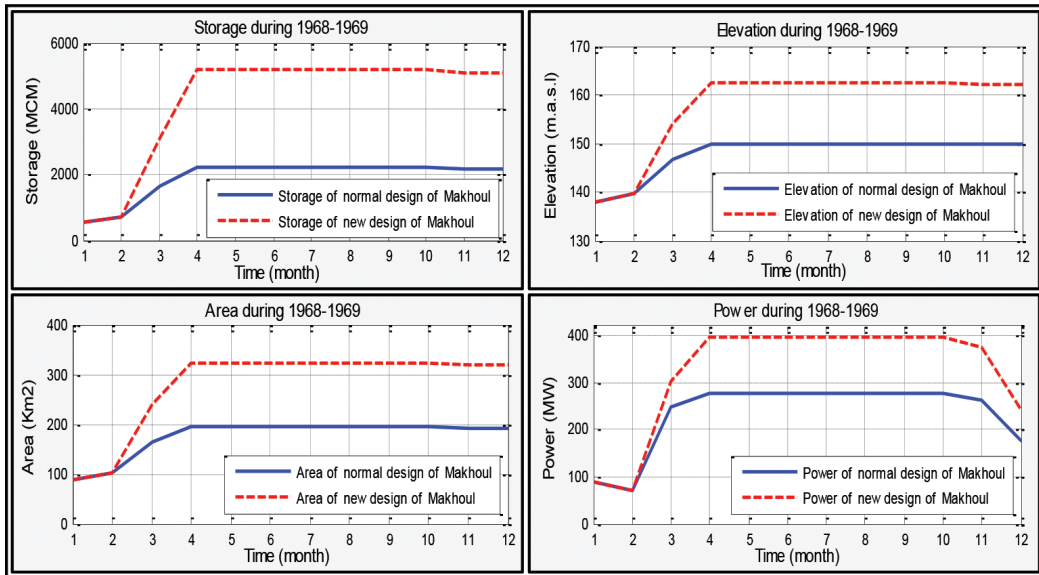
Figure 4. The monthly storage, elevation, surface area and power during the year 1968-1969, using the present and new designs
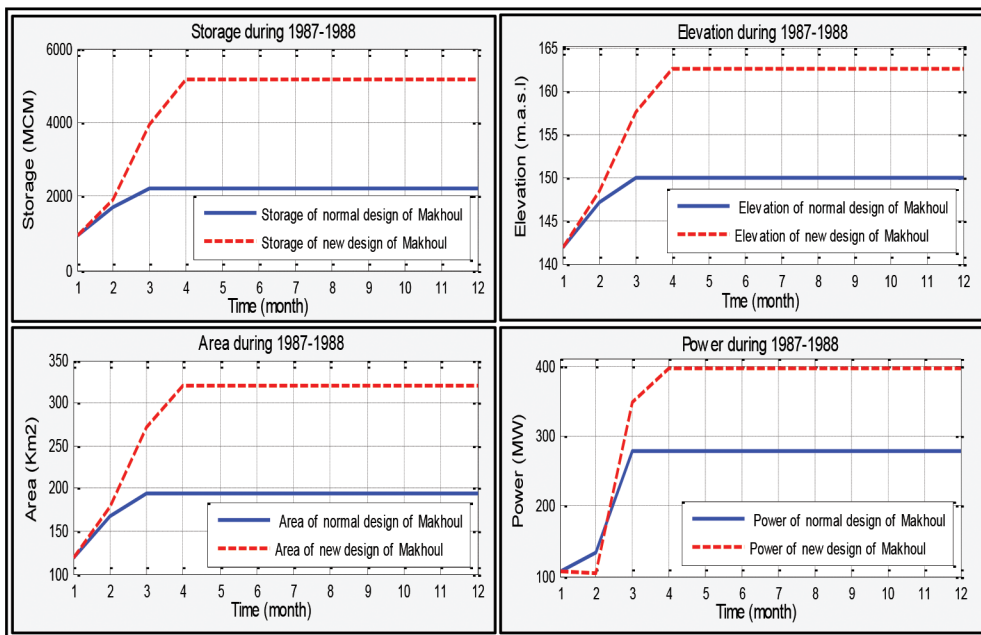
Figure 5. The monthly storage, elevation, surface area and power during the year 1987-1988, using the present and new designs
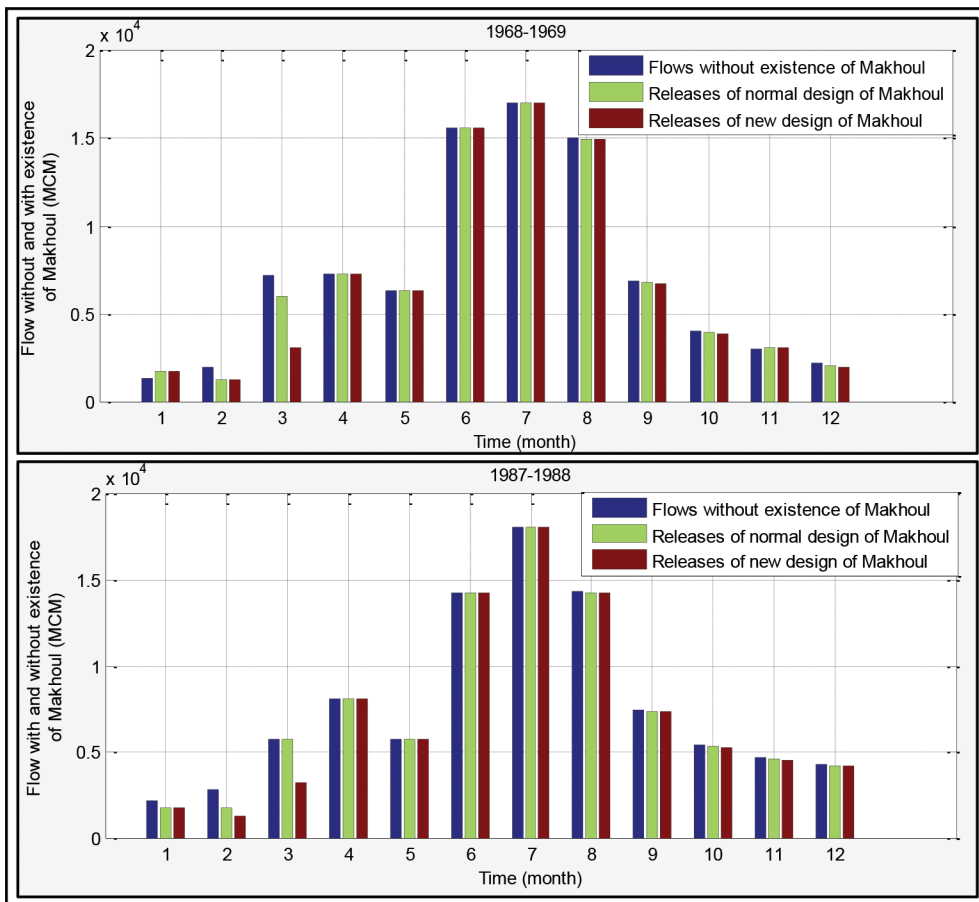
Figure 6. The monthly flow downstream of Makhoul in the years 1968-1969 and 1987-1988

CONCLUSION

In this study, a simulation model was developed to determine the SOP for the evaluation of a single reservoir by adopting two designs for the Makhoul Reservoir in two separate years of flow data. From using the two designs for the Makhoul Reservoir, it was seen that this reservoir would be ineffective should flood waves occur; thus, it would not be effective in reducing flood risk in Baghdad. Reducing flood risk in Baghdad was the main objective in constructing the reservoir. Nevertheless, some advantages were identified from using the new design for the Makhoul Reservoir, such as a rise in the capacity of the related hydroelectric power station and an increase in the elevation of storage.

However, the two designs were ineffective in reducing flood risk. The new design was preferred in terms of the benefits achieved. So, in the case of flood waves, it was clear that the water diversion canal at the Samarra barrage should be used as a precautionary measure to convert excess water at over 3000 m³/sec to the Tharthar Reservoir in order to avoid flood risk in Baghdad. As mentioned previously, it is not possible to exceed more than this amount of flow downstream of the Samarra barrage. As was evident, when using one of the two designs,

the problem of the salinity of the Tharthar Reservoir still existed; this will need resolution by other means. Therefore, the Makhoul Reservoir cannot be used as a substitute for the Tharthar Reservoir, but it can be used for other purposes, such as generating hydroelectric power, developing fishery in the area and irrigating the neighbouring lands. It can also be developed for recreation.

The new design of the reservoir achieved an increase in the volume of storage and hydropower production, but this involved loss of extra land. The maximum operational elevation was increased in the new design, leading to an increase in the height of the dam, and thus, an increase in construction costs. In this case, an economic feasibility study should be conducted for each design in order to choose the best one.

This DSM has many advantages, such as ease of inserting data and effective display of results. The results can be easily shown as values or figures. This DSM can be used initially by adopting the SOP to evaluate the performance of a single-storage system effectively. In addition, this DSM can be used with a multi-reservoir system through connection between reservoirs by taking advantage of the SIMULINK technique available in Matlab.

REFERENCES

Al-Aqeeli, Y. H., Al Mohseen, K. A., Lee, T. S., & Abd Aziz, S. (2015). Modelling monthly operation policy for the Mosul Dam, northern Iraq. *International Journal of Hydrology Science and Technology, 5*(2), 179–193.

Bayat, B., Mousavi, S. J., & Namin, M. M. (2011). Optimization-simulation for short-term reservoir operation under flooding conditions. *Journal of Water Supply: Research and Technology-AQUA, 60*(7), 434–447.

Khan, N. M., Babel, M. S., Tingsanchali, T., Clemente, R. S., & Luong, H. T. (2012). Reservoir optimization-simulation with a sediment evacuation model to minimize irrigation deficits. *Water Resources Management, 26*(11), 3173–3193.

Liu, X., Guo, S., Liu, P., Chen, L., & Li, X. (2011). Deriving optimal refill rules for multi-purpose reservoir operation. *Water Resources Management, 25*(2), 431–448.

Loucks, D. P., Van Beek, E., Stedinger, J. R., Dijkman, J. P., & Villars, M. T. (2005). *Water resources systems planning and management: An introduction to methods, models and applications*. Paris, UNESCO.

Malekmohammadi, B., Zahraie, B., & Kerachian, R. (2010). A real-time operation optimization model for flood management in river-reservoir systems. *Natural Hazards, 53*(3), 459–482.

Ngoc, T. A., Hiramatsu, K., & Harada, M. (2014). Optimizing the rule curves of multi-use reservoir operation using a genetic algorithm with a penalty strategy. *Paddy and Water Environment, 12*(1), 125–137.

Richaud, B., Madsen, H., Rosbjerg, D., Pedersen, C., & Ngo, L. (2011). Real-time optimisation of the Hoa Binh reservoir, Vietnam. *Hydrology Research, 42*(2–3), 217–228.

Salih, S. A., Kadim, L. S., & Qadir, M. (2012). Hydrochemistry as indicator to select the suitable locations for water storage in Tharthar Valley, Al-Jazira Area, Iraq. *Journal of Water Resource and Protection, 4*(08), 648–656.

Talukdar, B., Deb, D., & Srivastava, S. K. (2011). Development of multiobjective reservoir operation model for flood control benefit. In *Proceedings of the World Environmental and Water Resources Congress: Bearing Knowledge for Sustainability* (pp. 3978–3989).

UN-ESCWA & BGR. (2013). *Inventory of shared water resources in western Asia*. United Nations Economic and Social Commission for Western Asia; Bundesanstalt für Geowissenschaften und Rohstoffe, Beirut.

Demonstration of Comparison between Goat Skin and X-Ray Film Membranes on Traditional Musical Instrument Kompang

W. A. Siswanto[1] and M. Syiddiq[2]*

[1]*Department of Mechanical Engineering, Universitas Muhammadiyah Surakarta (UMS), Surakarta 57102, Indonesia*

[2]*Department of Engineering Mechanics, Universiti Tun Hussein Onn Malaysia, 86400 UTHM, Batu Pahat, Johor, Malaysia*

ABSTRACT

This paper presents a mathematical model of the traditional musical instrument, the *kompang*. In this study, a mathematical model of the *kompang* membrane is developed to simulate the vibration of the *kompang* membrane in polar coordinates by implementing the Fourier-Bessel wave function. The wave equation in polar direction is applied to provide the vibration modes of the membrane with the corresponding natural frequencies of the circular membrane. The initial and boundary conditions are determined to allow the development of numerical equation based on *kompang* membrane attachment. The mathematical model is coded in Smath for the numerical analysis as well as the plotting tool. Two *kompang* membrane cases with different membrane materials i.e. goat-skin and x-ray film are tried to test the model. The Finite Element Method (FEM) programme, Mecway, shows that the natural frequencies and the corresponding mode shapes are comparable with those from the developed model.

Keywords: Fourier-Bessel function, kompang, musical instrument modelling

INTRODUCTION

Specific traditional musical instruments and percussion instruments have been investigated and their mathematical models have been found to represent their vibration characteristics. Several tests and numerical studies have been performed in this area. Christian et al. (1984) and Rhaouti et al. (1999) reported on kettledrums, while Bertsch (2001), and Bilbao and Webb (2012) documented data on timpani. Research into musical instruments has also been done by Rossing et al. (1992) and Bilbao (2012), who did work on snare drums. Siswanto et al. (2014) reported on traditional musical instruments, providing documentation on the reproduction of the *kompang* sound and details of their analysis of the sound of the *kompang* for computer music synthesis (Ismail et al., 2006).

In this paper, the traditional percussion instrument, the *kompang*, that was used was from Johor, Malaysia. The musical instrument consists of two interacting components: a membrane and a rigid shell used for holding the membrane (See Figure 1). During *kompang* making, it is important to install the membrane to the *kompang* wood shell correctly and firmly. Incorrect installment of the membrane to the shell generates non-uniform movement across the membrane, causing inconsistency in terms of the sound (Ono et al., 2009). The *kompang* comes in various sizes, normally with a radius between 22 cm and 35 cm and a frame height between 4 cm and 6 cm (Abdullah, 2005). Continuous rigorous hitting of the membrane of the *kompang* usually can cause displacement of the *kompang* by several millimetres, that is a displacement that can be greater than the thickness of the membrane itself. It is reasonable to expect that different membranes have different sound attributes (Bank & Sujbert, 2006).

This paper aimed to provide a clear understanding of the traditional percussion musical instrument, the *kompang*, through a mathematical model of the instrument from a computational perspective. An understanding of traditional musical instruments is crucial as these instruments represent the cultural values of its users, as pointed out by Karjalainen et al. (1993), who studied the traditional musical instrument of Finland, the *kantele*. This study focusses on traditional musical instrument *kompang* for Malaysian region. In this paper, two types of *kompang* are considered: a goat-skin membrane *kompang* (see Figure 1) and an x-ray film membrane *kompang* (see Figure 2).
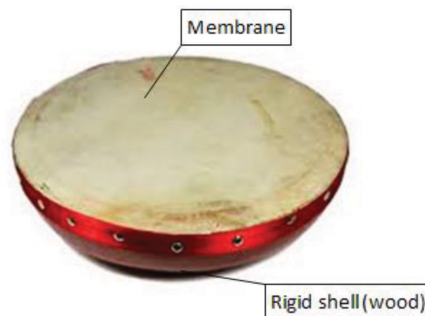
Figure 1. The traditional musical instrument, the *kompang*, using goat-skin for the membrane

Figure 2. The traditional musical instrument, the *kompang*, using x-ray film for the membrane

MATHEMATICAL DEVELOPMENT

Displacement of the circular membrane of the *kompang* during playing is due to the applied force from hitting the membrane of the instrument. The magnitude of the displacement depends on the radius of the circular membrane. Polar coordinates were selected in this study since the *kompang* membrane is circular in shape.

Forces on the *Kompang* Membrane

The forces exacted on the displaced membrane can be identified from the general mathematical equation of the membrane in polar coordinates as implemented by Morse (1948).
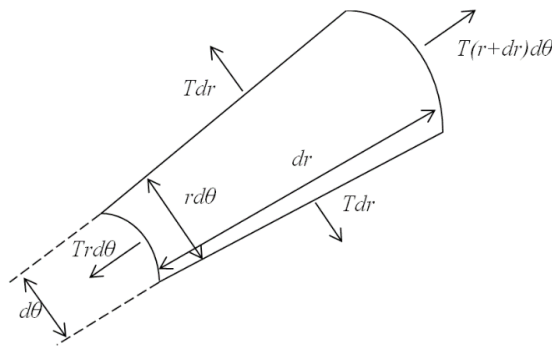
Figure 3. Forces on an element of a membrane in polar coordinates

Let us consider Figure 3. The net force due to the tensions perpendicular to the radius are given as:

$$T dr \left[\left(\frac{\partial u}{r \partial \theta} \right)_{\theta + \partial \theta} - \left(\frac{\partial u}{r \partial \theta} \right)_{\theta} \right] = \frac{T}{r^2} \frac{\partial^2 u}{\partial \theta^2} r dr d\theta \qquad (1)$$

where, r is the radius of the membrane, T is the tension of the membrane per length and θ is the angle around the axis of the membrane.

Due to the tensions parallel to the radius, the following equation is derived:

$$T d\theta \left[\left(r \frac{\partial u}{\partial r} \right)_{r + \partial r} - \left(r \frac{\partial u}{\partial r} \right)_{r} \right] = \frac{T}{r} \frac{\partial}{\partial r} \left(r \frac{\partial u}{\partial r} \right) r dr d\theta \qquad (2)$$

The Laplacian Operator

Laplacian operator is denoted by ∇^2, which can be applied in the operation to find a bulge point along the membrane surface, u, at some point (Morse, 1948). Applying the wave equation properties, we have:

$$\nabla^2 u = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} \qquad (3)$$

The equation of motion in polar coordinates at a point is:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} \tag{4}$$

converting it into a simpler form,

$$\frac{\partial^2 u}{\partial t^2} = c^2\left(\frac{\partial^2 u}{\partial r^2} + \frac{\partial u}{r\partial r^2} + \frac{\partial^2 u}{r^2\partial \theta^2}\right) \tag{5}$$

$$c = \sqrt{\frac{T}{\rho h}} \tag{6}$$

where, ρ is the membrane density and h is the membrane thickness.

Boundary Condition

The boundary conditions that were selected for the solution of the vibration modes for the circular membrane in polar coordinates were based on the model used by Nguyen et al. (2011). For a circular membrane fixed at the outer boundary, the deflection of the membrane at the radius of the membrane R (at $r = R$) is assumed to be zero; this implies that the membrane deflection decreases as the force approaches the edge of the circular boundary, expressed by the zeroing of the displacement variable, u:

$$u(R, t, \theta) = 0 \tag{7}$$

To create a numerical solution for the vibration, the initial displacement of the membrane, $P(r, \theta)$, was set to zero and the membrane movement was triggered by an initial velocity, $Q(t)$. These conditions were expressed as:

$$P(r, \theta) = 0 \quad \text{and} \quad Q(t) \neq 0 \tag{8}$$

Variable Separation

The displacement wave equation was legitimate under the condition of:

$$u = P(r, \theta)Q(t) \tag{9}$$

Embedding Equation 9 and its derivative in Equation 5, we derived:

$$\ddot{Q}P = c^2\left(\frac{\partial^2 P}{\partial r^2}Q(t) + \frac{\partial P}{r\partial r}Q(t) + \frac{\partial^2 P}{r^2\partial \theta^2}Q(t)\right) \tag{10}$$

reorganised by separation,

$$\frac{\ddot{Q}}{c^2 Q} = \frac{1}{P}\left(\frac{\partial^2 P}{\partial r^2} + \frac{\partial P}{r\partial r} + \frac{\partial^2 P}{r^2\partial \theta^2}\right) \tag{11}$$

Since both sides must be equal to a constant to generate a solution, a set of constants was drafted. Kreyszig (2011) introduced a negative constant $(-k^2)$ to satisfy the boundary conditions without being zero, as given below:

$$\frac{\ddot{Q}}{c^2 Q} = \frac{1}{P}\left(\frac{\partial^2 P}{\partial r^2} + \frac{\partial P}{r \partial r} + \frac{\partial^2 P}{r^2 \partial \theta^2}\right) = -k^2 \tag{12}$$

$$\frac{\ddot{Q}}{c^2 Q} = -k^2 \tag{13}$$

$$\frac{1}{P}\left(\frac{\partial^2 P}{\partial r^2} + \frac{\partial P}{r \partial r} + \frac{\partial^2 P}{r^2 \partial \theta^2}\right) = -k^2 \tag{14}$$

Comparing Equation 13 and Equation 14 yielded the following two differential equations:

$$\ddot{Q} + \lambda^2 Q = 0 \tag{15}$$

where,

$$\lambda = ck \tag{16}$$

$$\frac{\partial^2 P}{\partial r^2} + \frac{\partial P}{r \partial r} + \frac{\partial^2 P}{r^2 \partial \theta^2} + k^2 P = 0 \tag{17}$$

Separating the variable, $P(r, \theta)$, by using the following equation in Equation 17, we derived:

$$P = V(r)W(\theta) \tag{18}$$

where, $V(r)$ is the variable for the radius of the membrane and $W(\theta)$ is the variable for the angle around the axis of the membrane. This gave us:

$$WV'' + \frac{1}{r}WV' + \frac{1}{r^2}VW'' + k^2 VW = 0 \tag{19}$$

which was simplified as:

$$W'' + W\frac{r^2 V'' + rV' + r^2 k^2 V}{V} = 0 \tag{20}$$

Allowing:

$$-\frac{W''}{W} = \frac{r^2 V'' + rV' + r^2 k^2 V}{V} = n^2 \tag{21}$$

resulted in the following expressions:

$$W'' + n^2 W = 0 \tag{22}$$

$$r^2 V'' + rV' + (k^2 r^2 - n^2)V = 0 \tag{23}$$

Allowing $s = kr$, then $\left(k/_s\right) = \left(1/_r\right)$ and $ds/_{dr} = k$

$$\frac{dV}{dr} = \frac{dV}{ds}k \tag{24}$$

$$\frac{d^2V}{dr^2} = \frac{d^2V}{ds^2}k^2 \tag{25}$$

Using Equation 24 and 25 in Equation 23 gave us the following:

$$\frac{d^2V}{ds^2}k^2 + \frac{1}{r}\frac{dV}{ds}k + \left(k^2 - \frac{n^2}{r^2}\right)V = 0 \tag{26}$$

Substituting $r = {}^s/_k$, we got:

$$\frac{d^2V}{ds^2} + \frac{1}{s}\frac{dV}{ds} + \left(1 - \frac{n^2}{k^2 r^2}\right)V = 0 \tag{27}$$

This is Bessel's equation, as given in mathematical literature (Fletcher & Rossing, 1998). The solution of Bessel's equation was:

$$V_n(r) = J_n(s) = J_n(kr) \tag{28}$$

Equation 22 was also solved in the following method (Kreyszig, 2011):

$$W_n = \cos(n\theta) \ for \ n = 0,1,2, \ldots \tag{29}$$

Thus, using Equation 10:

$$P = J_n(kr)\cos(n\theta) \tag{30}$$

and using Equation 9:

$$u(r, t, \theta) = P(r, \theta)Q(t) = J_n(kr)\cos(n\theta)\,Q(t) \tag{31}$$

where the Eigen function $Q(t)$ can be expressed by the Fourier series as follows (Soedel, 2004):

$$Q_{mn}(t) = a_{mn}\cos(ck_{mn}t) + b_{mn}(ck_{mn}t) \ for \ m = 0,1,2, \ldots \tag{32}$$

Thus, Equation 30 can be rewritten as:

$$u_{mn}(r, t, \theta) = [a_{mn}\cos(ck_{mn}t) + b_{mn}\sin(ck_{mn}t)]J_n(k_{mn}r)\cos(n\theta) \tag{33}$$

Since, $u(R, t, \theta) = 0$

$$k_{mn} = \alpha_{mn}/R \tag{34}$$

α_{mn} = number of roots

The Fourier series coefficients are, therefore:

$$a_{mn} = \frac{2}{R^2 J_1^2(\alpha_{mn})} \int_0^R r p(r) J_0\left(\frac{\alpha_{mn}}{R} r\right) dr \tag{35}$$

$$b_{mn} = \frac{2}{c\alpha_{mn} R J_1^2(\alpha_{mn})} \int_0^R r q(r) J_0\left(\frac{\alpha_{mn}}{R} r\right) dr \tag{36}$$

and initial displacement and initial velocity are as follows, respectively:

$$p(r) = u(r, 0) \text{ and } q(r) = \frac{\partial u}{\partial t}\bigg|_{t=0} \tag{37}$$

The "α_{mn}" is defined as the *mth* positive zeros of $J_n(s)$, where m represents the nodal circle and n represents the nodal line. The positive zeros can be determined by plotting the zero-th order of the Bessel function as shown in Table 1, as found in Baricz (2010).

Table 1
Values for α_{mn}

			n	
m	1	2	3	4
0	2.4048	5.520	8.654	11.792
1	3.832	7.016	10.173	13.323
2	5.135	8.417	11.620	14.796
3	6.379	9.760	13.017	16.224

Natural Frequency

Using Equation 6 and 34 in Equation 16 provided the frequency of the membrane:

$$\lambda_{mn} = \frac{\alpha_{mn}}{R} \sqrt{\frac{T}{\rho h}} \tag{38}$$

RESULTS AND DISCUSSION

A numerical example for the first three vibrational modes where m=0 and n=1 until m=0 and n=3 is illustrated in this paper. Numerical calculation from the mathematical formula was conducted using the Smath software. The Smath software was utilised to calculate the numerical

example and to illustrate the result generated from the functions. The portrayed result in the Smath software was the value evaluated from Equation 32 for simulated 3D images of mode 01, mode 02 and mode 03, while Equation 39 provided the natural frequency, respectively. The result derived using the Smath software was then validated using the finite element analysis software, Mecway, to ensure the numerical result formulated in the Smath was correct and to determine the percentage error. For the finite element analysis in Mecway, 256 quad 8 elements were administrated with a total of 833 nodes for each of the modes simulated. The value of error between the two methods was later documented for further analysis and discussion.

The real *kompang* numeral cases were used as an example, with the radius of the circular membrane, R, given as 0.25m. Using the Hot Tack Tester V2.04.1 (see Figure 4), the tension of the membrane was determined to be about 100 N per length, which was proven to be within the range of the value provided by Salehi et al. (2014) and was assumed to be constant across both types of skin. The density of both types of skin was determined using $\rho = \frac{m}{v}$; this gave the goat-skin density as 552.905 $^{kg}/_{m^3}$ and that of the x-ray film as 1402.56 $^{kg}/_{m^3}$, therefore automatically generating the c value of 0.6014 $^{Nm^2}/_{kg}$ for the goat skin and 0.3776 $^{Nm^2}/_{kg}$ for the x-ray film.

Figure 4. The Hot Tack Tester V2.04.1 was used for determining the tension limit of the two types of membrane

The Fourier series with initial velocity of 0.5 $^{m}/_{s}$ and initial deflection of zero coefficients were then determined. A suitable initial boundary condition was important for an accurate result (Torin & Bilbao, 2013). By using these values in Equation 25 for 3D portrayal of the mode states and Equation 32 for natural frequency in the Smath software, we derived the vibrational deflection modes for mode 01, mode 02 and mode 03 with their natural frequency, as shown in Figure 4 using goat skin and Figure 5 using x-ray film membrane for the two types of *kompang*.

$\omega_{01}=150.3827$ $\qquad$ $\omega_{01}=152.0775$

$\omega_{02}=345.1961$ $\qquad$ $\omega_{02}=355.0388$

$\omega_{03}=541.1538$ $\qquad$ $\omega_{03}=574.4442$
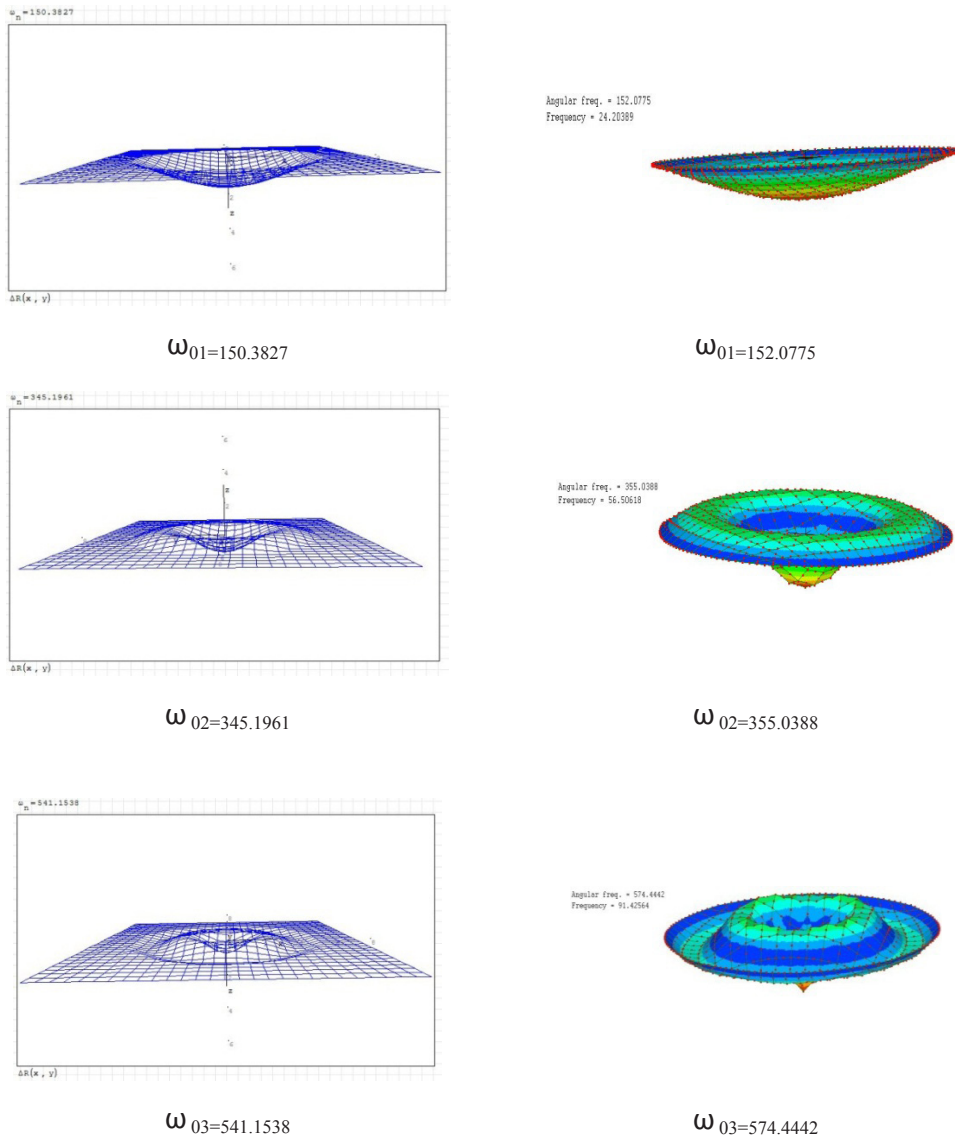
Figure 5. Results of mode 01, mode 02 and mode 03 of goat-skin *kompang* simulated in the Smath (left) and Mecway software (right)

$\omega_{01}=191.4441$ $\omega_{01}=193.6017$

$\omega_{02}=439.4505$ $\omega_{02}=451.9807$

$\omega_{03}=688.9137$ $\omega_{03}=731.294$
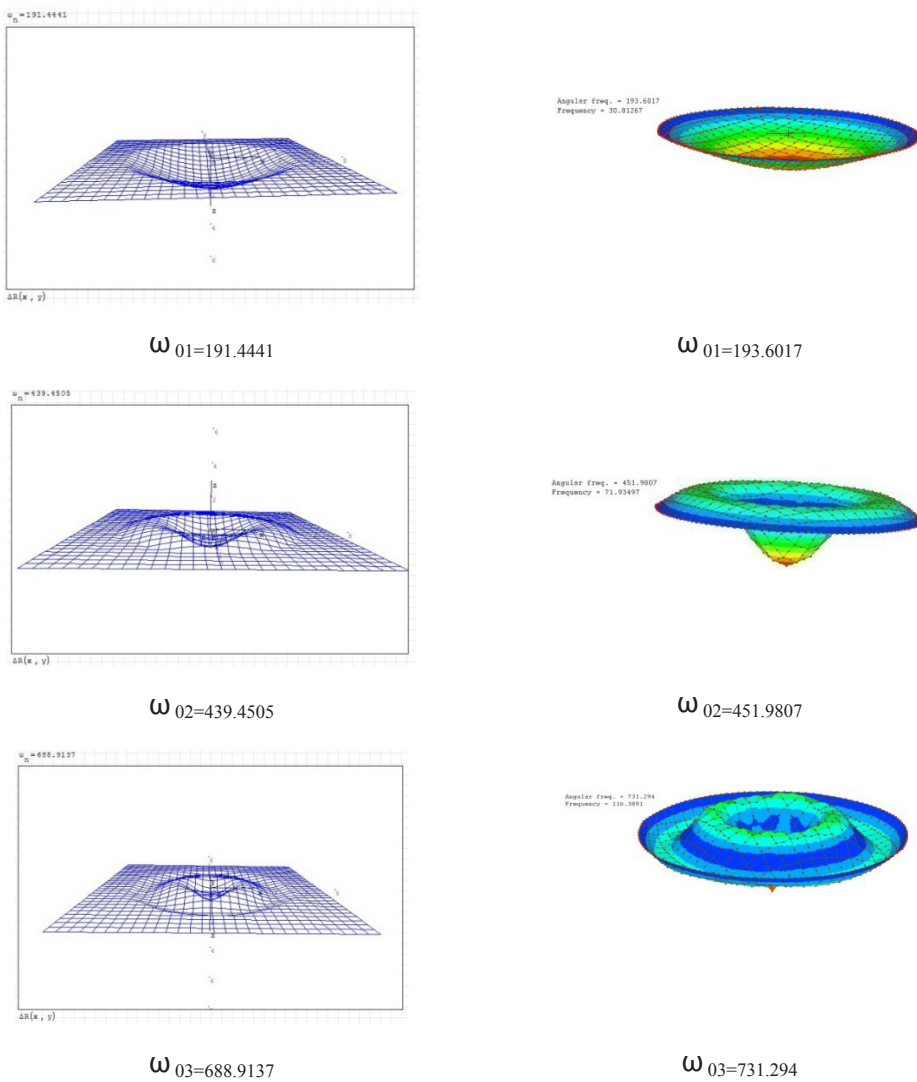
Figure 6. Results of mode 01, mode 02 and mode 03 of X-ray film *kompang* simulated in the Smath (left) and Mecway software (right)

Figure 5 and Figure 6 show the results of mode 01, mode 02 and mode 03 generated by use of the *kompang* using the numerical calculation approach provided by Smath and finite element analysis provided by Mecway for the two types of *kompang* membrane. Figure 5 shows the natural frequency results for the goat-skin membrane derived from the numerical calculation in the Smath software: 150.3827 $rad/_{sec}$ for mode 01, 345.1961 $rad/_{sec}$ for mode 02 and 541.1358 $rad/_{sec}$ for mode 03. Using Mecway provided these results: 152.0775 $rad/_{sec}$ for mode 01, 355.0388 $rad/_{sec}$ for mode 02 and 574.4442 $rad/_{sec}$ for mode 03.

Figure 6 shows the results collected for the x-ray-film *kompang*: the natural frequency value obtained was 191.4441 rad/sec for mode 01, 439.4505 rad/sec for mode 02 and 688.9137 rad/sec for mode 03 in numerical calculation in Smath. Meanwhile, the finite element analysis using Mecway showed these results: 193.6017 rad/sec for mode 01, 451.9807 rad/sec for mode 02 and 731.294 rad/sec for mode 03. These results implied that the x-ray-film *kompang* produced a slightly higher natural frequency compared with the goat-skin *kompang*. This finding was supported by the findings of Christopher and Umesh (2006), who reported that in percussion musical instruments, the higher density of the membrane results in a higher value for the natural frequency.

Percentage Error

To calculate the percentage error of value between the values generated by the Smath and Mecway software, we used mode 01 for the goat-skin membrane *kompang* as an example, as given below:

$$\frac{|approximate\ value - exact\ value|}{|exact\ value|} \times 100\% \tag{33}$$

$$\frac{|150.3827 - 152.0775|}{|152.0775|} \times 100\%$$

$$= 1.11\%$$

Table 2

Percentage error of frequency surveyed between smath and mecway software results for goat skin membrane

Mode	Goat Skin		Percentage Error (%)
	Smath	Mecway	
01	150.3827	152.0775	1.11
02	345.1961	355.0388	2.77
03	541.1538	574.4442	5.79

Table 3

Percentage error of frequency surveyed between smath and mecway software results for x-ray-film membrane

Mode	X-Ray Film		Percentage Error (%)
	Smath	Mecway	
01	191.4441	193.6017	1.11
02	439.4505	451.9807	2.77
03	688.9137	731.294	5.79

Table 2 and Table 3 shows the summarised results obtained from using the numerical mathematical calculation generated by Smath and the finite element analysis method using Mecway for both types of skin. Table 2 and Table 3 show that the frequency increased dramatically from mode 01 to mode 02 and from mode 02 to mode 03.
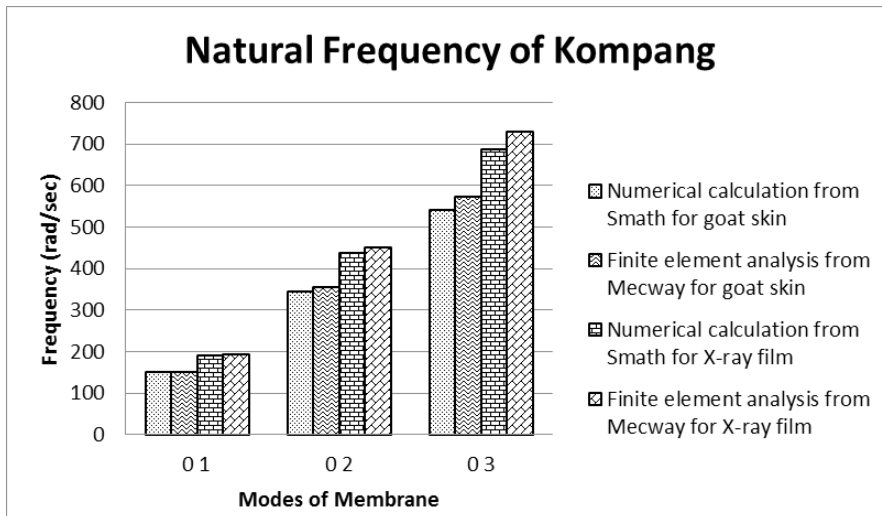
Figure 7. Chart representing the evaluated results obtained from the numerical calculation using Smath software and finite element analysis using Mecway software

Figure 7 shows the chart generated for both types of *kompang* in terms of natural frequency content with respect to mode 01, mode 02 and mode 03. From the graph plotted, it was found that the reading generated from Smath software almost overlapped the reading provided by Mecway software for mode 01 and the gap, which showed that the error widened as it proceeded from mode 02 to mode 03.

The increase in percentage error from mode 01 to mode 03 shown by the Mecway software compared with the results from the Smath software was probably due to the limited number of nodes, causing the movement to be less smooth in mode 02 and mode 03 compared with in mode 01 (refer to Figure 5 and Figure 6). Increasing the number of nodes allowed the results to be closer to the values generated by Smath since in finite element analysis, the more the number of nodes installed, the better the results obtained (Dow, 1998).

Despite this, the value of error ranging from 1.11% to 5.79% is considered to be a small error and therefore, can be neglected. This result demonstrated that the mathematical approach of Smath provided for more accurate results and was suitable for evaluating the musical instrument, the *kompang*.

CONCLUSION

To analyse the use of the type of membrane used in the traditional percussion instrument, the *kompang*, the Fourier-Bessel solution for the circular membrane vibration modes shown in this paper was used by utilising the wave equation in polar coordinates. The developed vibration modes were based on Bessel functions, with solution derivatives from the Fourier series. The solutions in the Smath were a purely mathematical approach to vibrational normal mode development using polar coordinates. The value for natural frequency provided by mathematical calculation derived from Smath were proved to be almost similar to the results generated by finite element analysis in Mecway, with a very small error, hinting that the mathematical approach was suitable and relevant for analysing the musical instrument, the *kompang*.

ACKNOWLEDGEMENT

REFERENCES

Abdullah, M. H. (2005). *Kompang: An organological and ethnomusicological study of a Malay frame drum.* (Master Thesis). Newcastle University, UK.

Bank, B., & Sujbert, L. (2006). *Physics-based sound synthesis of stringed instruments including geometric nonlinearities*. (PhD Thesis). Budapest University of Technology and Economics, Hungary.

Baricz, A. (2010). *Generalized Bessel functions of the first kind.* New York, NY: Springer.

Bertsch, M. (2001). Vibration patterns and sound analysis of the Viennese Timpani. In *Proceedings of ISMA* (pp. 1-4). Perugia, Italy.

Bilbao, S. (2012). Time domain simulation and sound synthesis for the snare drum. *The Journal of Acoustical Society of America, 131*(1), 914–925.

Bilbao, S., & Webb, C. (2012). Timpani drum synthesis in 3D on GPGPUs. In *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)* (pp. 1-8). York, United Kingdom.

Christian, R. S, Davis, R. E., Tubis, A., Anderson, C. A., Mills, R. I., & Rossing, T. D. (1984). Effects of air loading on timpani membrane vibrations. *The Journal of Acoustical Society of America*, *76*(5), 1336–1345.

Christopher, H. M. J., & Umesh, A. K. (2006). Membrane vibration experiments: An historical review and recent results. *The Journal of Sound and Vibration, 295*(3), 602–613.

Dow, J. A. T. (1998). *A unified approach to the finite element method and error analysis procedures.* United States of America, USA: Academic Press.

Fletcher, N. H., & Rossing, T. (1998). *The physics of musical instruments.* New York: Springer-Verlag.

Ismail, A., Samad, S. A., Hussain, A., Azhari, C. H., & Zainal, M. R. M. (2006). Analysis of the sound of the kompang for computer music synthesis, SCOReD 2006. In *Proceedings of 2006 4th Student Conference on Research and Development "Towards Enhancing Research Excellence in the Region* (pp 95–98). IEEE.

Karjalainen, M., Backman, J., & Polkki, J. (1993). Analysis, modeling, and real-time sound synthesis of the kantele, a traditional Finnish string instrument. In *Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing: Plenary, Special, Audio, Underwater Acoustics, VLSI, Neural Networks – Volume I* (pp. 229–232). Washington, DC: IEEE Computer Society.

Kreyszig, E. (2011). *Advanced engineering mathematics.* United States of America, USA: Jhon Wiley & Sons.

Morse, P. M. (1948). *Vibration and sound.* New York, NY: McGraw-Hill.

Nguyen, D. M., Anton, E., Allan, R. G., & Jens, G. (2011). Isogeometric shape optimization of vibrating membranes. *Journal of Computer Methods in Applied Mechanics and Engineering, 200*(13), 1343–1353.

Ono, T., Takahashi, I., Takasu, Y., Miura, Y., & Watanabe, U. (2009). Acoustic characteristics of Wadaiko (traditional Japanese drum) with wood plastic shell. *The Journal of Acoustical Science and Technology, 30*(6), 410–416.

Rhaouti, L., Chaigne, A., & Joly, P. (1999). Time-domain modeling and numerical simulation of a kettledrum. *The Journal of Acoustical Society of America, 105*(6), 3545–3562.

Rossing, T. D., Bork, I., Zhao, H., & Fystrom, D. O. (1992). Acoustics of snare drums. *The Journal of Acoustical Society of America, 92*(1), 84–94.

Salehi, M., Kadim, I., Mahgoub, O., Negahdari, Sh., & Naeeni, R. S. E. (2014). Effects of type, sex and age on goat skin and leather characteristics. *The Journal of Animal Production Science, 54*(5), 638–644.

Siswanto, W. A., Che Wahab, W. M. A., Yahya, M. N., Ismail, A. E., Nawi, I. (2014). A platform for digital reproduction sound of traditional musical instrument kompang. *Applied Mechanics and Materials, 660,* 823–827.

Soedel, W. (2004). *Vibrations of shells and plates* (3rd Ed.). United States of America, USA: CRC Press.

Torin, A., & Bilbao, S. (2013). Numerical experiments with non-linear double membrane drums. In *Proceedings of the Stockholm Musical Acoustics Conference/Sound and Music Computing Conference* (pp. 569-576). Stockholm, Sweden.

An Efficient Method for Secure 2D Image Visualization and Transmission through Chaotic Confusion and Pixel Diffusion

Gunasekaran, G.[1]* and Venkatesan, M.[2]

[1]*School of Information Technology and Engineering, VIT University, Vellore-632014, India*

[2]*School of Computer Science and Engineering, VIT University, Vellore-632014, India*

ABSTRACT

The main goal of this paper is to present an efficient method for secure 2D image visualisation and transmission through chaotic confusion and pixel diffusion methods. In recent years, the expeditious valuation of digital data exchange has become very important. The safety of any data exchange is imperative in the digital image data storage, retrieval process and communication procedure. The safety and security issues related to Internet banking and email systems necessitate text security in today's digital media. In order to provide secure transmission and retrieval of image data, pixels in an image are scrambled and replaced within the image. The proposed system uses chaotic confusion and pixel diffusion techniques for securing input images. In the confusion method, the pixels within the image are substituted with other values and in the diffusion method, the pixels are re-substituted within the input image. The confusion method and the diffusion method use the selective technique for the generation of secret keys. Improve the quality of the proposed work is utilizing various image quality parameters and error metrics such as MSE, SNR, PSNR etc. The proposed system also utilises chaos maps for the encryption and decryption of the input image.

Keywords: Confusion and diffusion, cryptography, decryption, encryption

INTRODUCTION

The two basic types of cryptography, symmetric key and asymmetric key, encode image data so that they are unreadable without a secret key to unlock the data (Singh et al., 2015; Hashim & Neamaa, 2014). Faster creation and evaluation of digital data sharing is a current need. At the present time, a more secure data transmission process is being sought as security is a very important requirement in the process of data storage and transmission. All transactions through Internet banking and online systems such as m-commerce,

e-business and e-shopping as well as the use of email systems require effective and efficient security. Digital image security is also crucial and must be effectively utilised during digital image transmission and storage; this is a crucial need in industry and research. A systematic method to develop federal data handling standards that are more powerful, adaptable, secure and fast in replacing the basic information encryption standard has been initiated by the National Institute of Standards and Technology, Gaithersburg, Maryland, United States. Digital images have a large data size and also introduce real-time problems. Current methodologies cannot be utilised to safeguard digital images or text data from illegal access, but with a few amendments to the AES method, these can be effectively safeguarded. The various encryption-decryption techniques used for security purposes are symmetric cryptosystems, DES, AES, Blowfish and RSA (Mitali & Sharma, 2014). Comparatively, symmetric cryptosystems are faster than other methods and only a system that has the key can decrypt the received data. The disadvantage of symmetric cryptosystems is that they have transportation issues; in addition, digital signatures cannot be provided. Data Encryption Standard (DES) is proven to be unaffected by all types of cryptanalysis. Since DES has a very small key size, it takes less time to be searched. In general, the AES method is resistant to almost all known attacks and is fast and flexible. However, the key required in the AES method is too large, and this makes computation complex. Compared with other methods, RSA provides better security. It is also capable of providing digital signatures. The processing of public key cryptography is very slow, and this is the disadvantage of this system. The proposed work presents an effective encryption and decryption method for digital images utilising a selective approach for securing the image data to overcome unauthorised access. It also overcomes the disadvantages that remain with the existing system.

RELATED WORK

This section discusses various works related to the field of digital transmission that uses suitable techniques for encryption and decryption of digital images. Cryptography is considered a branch of both mathematics and computer science that associates closely with information theory and coding methods, computer security, computer engineering and commerce. Many studies have been done on image encryption to meet the need for real-time safe digital image communication and transmission over the web by means of wireless communication networks. This paper focusses on efficient techniques for secure 2D image visualisation and transmission through chaotic confusion and pixel diffusion. Several previous works have been consulted for help in developing the proposed system. A fast encryption method based on vector quantisation is proposed for image security. A comparison is made by integrating diffuse codes to confuse the code concepts and to encrypt the elements of the security book using a symmetric security algorithm (Chang et al., 2001). An unreachable differential security analysis of AES-192 and AES-256 (Phan, 2004) up to sound rounds is proposed.

Selective encryption is a new technique that reduces computational complexity. Selective encryption improves a simple scalar quantised sample to illustrate the influence of the basic components, list out various essential consumer electronic equipment and use a suitable method for improving and examining selective encryption for specific compression methods

(Lookabaugh & Sicker, 2004). A study motivated and background that by the different standard and developed magnification lens rendering methods to magnify the topographic of interest, whereas compressing the outstanding volume areas without removing them left entirely are presented (Wang et al., 2005). All our lenses are accelerated on the GPU. In this method, the user accesses the obtainable screen region interactively, and is offered more regions with the best resolution and essential features. A simple technique for image security utilising an integration of block-based image transformation and encryption was proposed by Bani et al. (2008). Their proposed method provided good performance in terms of correlation and entropy calculations compared with the existing methods. A novel crypto-system for encrypting and decrypting digital images, audios and videos was presented. Associations with confusion and diffusion behaviours by utilizing chaotically coupled chaotic maps were proposed and described by Pisarchik and Zanin (2008). Their proposed scheme was more secure than existing methods in terms of computational speed, content security and improved speed of transmission between sender and receiver.

Wang et al. described cross chaotic maps that aid logistic maps and Chebyshev maps (2008). Here, every pixel in the image is randomly moved according to the encryption matrix in the process of grayscale substitution. In addition, the scrambling process is applied in the process of position permutation. The development of digital information coding based on the combination of chaotic techniques was presented by Behnia et al. (2008). Here, the combination of chaotic maps illustrates the merits of large key space and higher-end security. Also, the proposed method presented an idea for practical utilisation of security in the secure communication of personal data over the network. The proposed advanced Hill cipher secure technique was highly secure against brute force attacks when compared with original Hill cipher technique (Acharya et al., 2009). A new image coding method to support both abstract analysis and practical solutions representing high-end security was proposed by Wei-Bin and Xin (2009). The proposed method had many advantages such as support for a lot of key space, shorter running time compared with other methods and a very sensitive secret key. An advanced encryption standard to develop a secure symmetric image encryption method was presented by Kamali et al. (2010). Their work utilised row-changing alteration. Verma et al. did a comparative analysis of four commonly used techniques and presented their analysis (2011). Their work presented the possibility of faster digital transmission and a higher level of security through strong key size functions, random bit creation and Internet-based systems. On their part, Shazmeen and Prasad (2012) showed a new cryptographical approach that aids symmetric cryptography in combination of a software-based security system. The main objective was to provide better transmission security for banking customers. Their proposed work also described existing online banking issues and identified the requirement for security performance for online banking systems. Comparative performance analyses of selected symmetric cryptography methods were presented.

Selected methods such as AES, DES, 3DES, RC6, Blowfish and RC2 were demonstrated by Mathur and Kesarwani (2013). They presented experimental results related to encoding, packet size and data types. The best hybrid models for image encryption based on a genetic algorithm through chaotic function were discussed by Nichat and Sikch (2013). In their work, images were converted from plain text to cipher text with the help of the chaotic method

using a secret key and in the second step, the genetic algorithm was applied for optimising the conversion of ciphered text to the encrypted form effectively. The experimental results compared the correlation components and entropy value. Encryption and decryption methods were proposed based on a chaotic map to create a permutation matrix with two parameters to create a shuffler. The proposed method's key space was huge enough to attack all types of brute-force attacks (Bremnavas et al., 2013). The experimental results suggested that this method was a good candidate for image operation for cryptography and steganography. A selective image encryption method for medical and satellite images was presented by Panduranga and Naveenkumar (2013). The first method for selective encryption was valuable for cases where the region or location of interest is known. The second method for selective encryption was important for instances where concentrated objects appear in one image. Another selective image encoding method that contains extreme image information was described by Ullah et al. (2013). The proposed work can be applied to encrypt medical images because medical image information also requires higher-level security for image transfer through unauthorised channels. Efficient and highly secure methods are used for image encryption (Gunasekaran & Ray, 2014). The results and discussion showed that a high level of security using the image encryption method with less computational time was ensured. The study helped to increase security levels for medical and non-medical images that are illustrated. Hashim and Neamaa (2014) used the modified Elgamal cryptosystem method to support both colour and black and white images of any size stored in PNG and JPG formats for encrypting and decrypting with the help of revised cryptosystems. The modified Elgamal system provides more security than other methods. Mitali and Sharma (2014) did a comparative study of different security methods on several settings of information packets. Their paper investigated different types of encryption and decryption to provide better encryption methods with higher computational speed and to reduce energy consumption. A new scheme for image encryption and verification for digital transmission was introduced. Many chaos-based image encoding methods were presented by Thakur et al. (2014). The created hash function was embedded in the image and the whole image was encoded with a mixture of chaotic maps and substitution diffusion methods were presented. Various symmetric cryptography methods such as AES, DES, 3DES, Blowfish, RC4 and RC6 were demonstrated and reviewed by Princy (2015). Finally, the author suggested that the blowfish method was more secure compared with other cryptography methods and showed good experimental outcomes for low computational cost and a fewer number of rounds. An advanced multi-level secret sharing visualising cryptography method for encapsulating information in colour image and its respective share was illustrated as captcha to customers. Phishing websites, as well as human user attacks, can be easily monitored and resolved, as demonstrated by Priya et al. (2015). The proposed technique provided multi-level verifications and validation processing by captcha options. Two-dimensional Henon chaotic maps and skew tent maps were arranged in a scheme for effective chaos-based image encryption methods by Khan et al. (2015). Confusion and diffusion methods were applied for encryption and decryption of images; chaotic maps play a key role in permutation and diffusion machines. The method can be easily used for unbalanced images that have a different width and height as well. An efficient coding technique for encoding and decoding images with the help of symmetric key concepts were illustrated by Singh et al. (2015). The advantage is that the proposed system

used a separate key that was required for both the sender and the receiver. This cryptosystem key concept helped in sending a huge volume of messages. Other techniques for computer networking and various loopholes in current network systems were analysed by Ruman and Phaneendra (2015). The various security threads overtaken by different technologies for securing the network through cryptography and encryption methods were presented. A proposed method utilising an enhanced RSA method was proposed. It also improved the number of bits in the SSL connection for security. This method guaranteed secure service compared with other systems in terms of key creation time. An innovated image encryption method created on a Henon map and complex spatiotemporal chaos methods were presented by Zheng and Jin (2015). The work focused on pixel location permutation and pixel position shifting, which help the cryptosystem. The blowfish method developed and implemented with help of the WDDL logic pattern in a bottom-up fashion was projected by Dinesh and Suveetha (2016). A separate secure key was used for encryption and decryption by both the sender and receiver without modification of the source image and encrypted image. Various other efficient methods and techniques helped to develop secure 2D-image visualisation and digital transmission through the chaotic confusion and pixel diffusion methods. This proposed work concentrated on an encryption and decryption model using the selection approach with association with the Hundugen function for generating a secret key.

PROPOSED ENCRYPTION AND DECRYPTION METHOD

The proposed work on encrypting and decrypting digital images using the selective approach is shown in Figure 1. Security can be applied to the image data that are sent through the web or any other communication medium in the network. The proposed work ensures security so that nobody can have unauthorised access to the digital resources that need to be protected. The proposed system used the permutation combination technique to encrypt the input digital image and thereby, lock the private key. Anyone who has the key will have the privilege of accessing a particular image. In the encryption model of the image security system, the confusion technique was utilized, whereas in the decryption model, the diffusion technique (Pisarchik & Zanin, 2008; Ling et al., 2008) was utilised for its efficacy. The overall architecture had two levels of encoding: the image encryption model and the image decryption model. The image encryption model will import any type of image with a private key and encrypt the input image into a cipher code form, while the image decryption model decrypts the encrypted image and reports back to the original image. The following diagrammatical representation shows the encryption and decryption systems using the selective approach method.
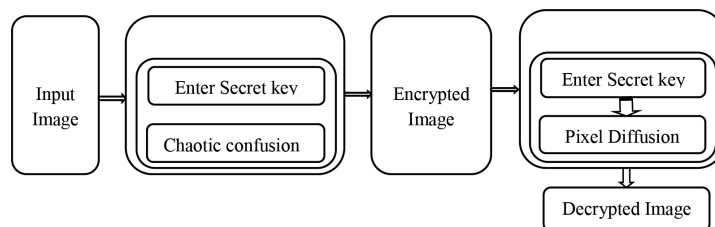

Figure 1. Overall architecture of encryption and decryption systems

The three major processes involved in the proposed method are as follows. First, the input image is read into the encryption model through the selective encryption approach. Second, the process of key generation is described; here, the primary key is generated (Acharya et al., 2009). Finally, the Hundungen process is executed to generate a Henon map by generating random numbers to decrypt the encrypted image. The Henon map is a discrete-time dynamical system that exhibits chaotic behaviour (Khan et al., 2015). Once the input image is provided, the key is generated in the developed system. The generated key acts as the password and locks the input image until the correct image is provided. Then, the key generation function is executed automatically, and all four keys K1, K2, K3 and K4 are calculated. Once the Hundungen function is executed, it manipulates the images pixel by pixel and then encrypts the image by manipulating the pixels within the image. Chaotic confusion is an encryption technique that shuffles the pixels within an image or substitutes the pixels within an image. Pixel diffusion is another encryption technique in which the pixels in an image will be replaced with the XORed values of the RGB components. The key approaches involved in the encryption and decryption are as follows.

Selective Encryption Approach

Selective encryption is a technique to save computational complexity or enable interesting new system functionality by only encrypting a portion of the image while still achieving adequate security (Lookabaugh & Sicker, 2004). The selective encryption approach connects all other models in the system and will run as the main function in the system process. Once the input image is read from the user, the respective key has to be submitted to the system to encrypt and lock the image. Then, its Red (R), Green (G) and Blue (B) values are taken, and the Hundungen function is executed to generate various keys to encrypt the image so that the original image cannot be viewed by any unauthorised users. If the image is to be decrypted, then the person who is accessing it has to provide the correct key to view the original image. The image is decrypted back again using the key-gen function, which will reverse the process from the end of the encryption and return the original image as the final one. The histogram of the image gives the number of intensity values that are present in a particular image. The histograms of the input image and the decrypted image are compared to determine the visual efficiency of the decryption. In the proposed work, the chaos-based image encryption scheme was suggested; the scheme involved two different processes, chaotic confusion and pixel diffusion (Pisarchik & Zanin, 2008; Ling et al., 2008). Chaotic confusion is the process of transforming to the cipher image from the plain image. In this proposed method, the confusion process was performed by bitXORing, the first pixel of the red channel of the image with the key, K_1, the first pixel of the green channel with K_2 and the first pixel of the blue channel with key K_3 and the other pixels were XORed.

$$R\,(1,1) = R\,(1,1)\ \text{XOR}\ K_1 \tag{1}$$

$$G\,(1,1) = G\,(1,1)\ \text{XOR}\ K_2 \tag{2}$$

$$B(1,1) = B(1,1) \text{ XOR } K_3 \qquad\qquad\qquad [3]$$

$$R(1,2) = R(1,2) \text{ XOR } K_4 \qquad\qquad\qquad [4]$$

The process is repeated until all the pixels of the R, G and B components are XORed. After the chaotic confusion process, the pixel diffusion process is executed. In this pixel diffusion process, the output bit values should depend on the input bit of values in a very difficult way. To achieve this, the pixel diffusion process was carried out in two steps i.e. horizontal diffusion and vertical diffusion.

The proposed algorithm was as follows.

Algorithm 1: Selective Approach Encryption

Start

Step 1: Read the input image, I.

Step 2: Input the private key, K, into the encryption model.

Step 3: Generate RGB components from the input image.

Step 4: Calculate Henon map, $X_{n+1} = Y_{n+1} - aX^2_n$, $Y_{n+1} = bX^2_{n,}$ and $Z_{n+1} = 1 - cY^2_{n+1} + X_n$.

Step 5: Bit-XOR the private key, K, with all the RGB components of the input image.

Step 6: Substitute with the XOR-ed pixels into one to form the cipher image.

Step 7: Encrypted image $I_{Encr\ is}$ generated.

Stop

Algorithm 2: Selective Approach Decryption

Start

Step 1: Read the encrypted image, $I_{Encr.}$.

Step 2: Generate random numbers, Rnd = $\{(X_1, Y_1, Z_1), (X_2, Y_2, Z_2)\ldots, (X_n, Y_n, Z_n)\}$

Step 3: Generate Henon map using random numbers.

Step 4: Enter the private key, K.

Step 5: Bit-XOR the random numbers, Rnd, with the original input image, I.

Step 6: Compare the results of the previous step with the results of step 3 in the encryption algorithm.

Step 7: If the two results match, re-substitute the original pixels to obtain the decrypted image, I_{Decr}.

Step 8: If the results do not match, perform step 3 to step 5; otherwise,

Stop

Hundungen function

The Henon map is utilised for image encryption in the proposed method. The Henon map is used to represent the disconnected time dynamic system, which is a mathematical concept in which a fixed rule describes the time dependence of a point in a geometric space. The Henon map is defined by the function:

$$X_{n+1} = Y_{n+1} - aX^2_n \qquad\qquad [5]$$

$$Y_{n+1} = bX^2_n \qquad\qquad [6]$$

$$Z_{n+1} = 1 - cY^2_{n+1} + X_n \qquad\qquad [7]$$

where, a, b and c are constants and X, Y and Z are pixel colours for each channel like Red, Green and Blue, respectively. This function generates a random value, and these random values are bitXORed with the original value of the image i.e. the X value will be bitXORed with the red channel pixel, Y with the green channel and Z with the blue channel. Image encryption using the chaos map includes the input image and the secret key for encrypting the plain image. The first step will be to generate the keys, which will consist of the four variables K_1, K_2, K_3, and K_4. The values of X_n, Y_n and Z_n will be generated using the Henon map function, which will be used to generate the key. The K_1 key will be generated by multiplying the values of X_n and Y_n and dividing the result by 256. The K_2 key will be generated by multiplying the values of Y_n and Z_n and dividing the result by 256, and the K_3 key will be generated by multiplying the values of Z_n and X_n and dividing the result by 256. The K4 key will be generated by taking the mod of X_n and 256. The keys will be generated by the following rules:

$$K_1 = (X_n * Y_n)/256 \qquad\qquad [8]$$

$$K_2 = (Y_n * Z_n)/256 \qquad\qquad [9]$$

$$K_3 = (Z_n * X_n)/256 \qquad\qquad [10]$$

$$K_4 = mod\,(X_n, 255) \qquad\qquad [11]$$

Key Generation Function

The keys are generated in the reverse manner from which the Hundungen function is executed because the values of the matrix at the same positions of the original image have to be reversed. Once the correct key is provided, then the values of all the respective keys are calculated and checked. Then, these keys are bitXORed to calculate the values of the matrix pixel by pixel. The processes of chaotic confusion and pixel diffusion are reversed, and each pixel value is the same as that of the original image pixel value.

RESULTS AND DISCUSSION

Several different types of digital image with a size of 256*256 are utilised to demonstrate the encryption and decryption of images in the proposed system. A sample of five different images is considered and projected here, and the quality measures and error metrics of the images are calculated to determine the efficiency of the proposed system. Table 1 shows the empirical results and the elapsed times that are obtained from the developed system. The various quality

measurements for the encrypted and decrypted images are projected from Tables 2 and 3, and the error metrics for both the encrypted and decrypted images are illustrated from Tables 4 and 5. A graphical representation of the elapsed time for processing the encrypted and decrypted images is depicted in Figure 2.

Table 1

Experimental results for input, encrypted and decrypted images
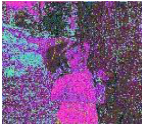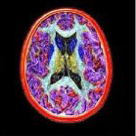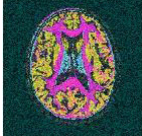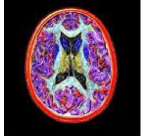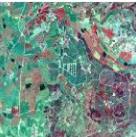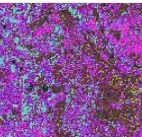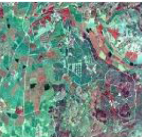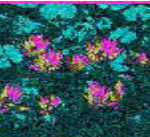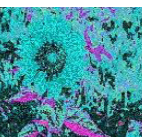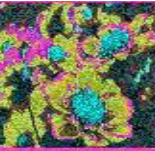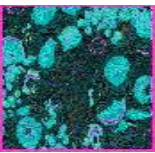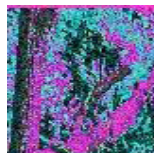
Sl. No.	Input Image	Encrypted Image	Decrypted Image	Sl. No.	Input Image	Encrypted Image	Decrypted Image
1				7			
2				8			
3				9			
4				10			
5				11			
6				12			

Table 1 (*continue*)

Sl. No.	Input Image	Encrypted Image	Decrypted Image	Sl. No.	Input Image	Encrypted Image	Decrypted Image
13				18			
14				19			
15				20			
16				21			
17				22			
23				24			

The comparison of various encryption and decryption methods is shown in Table 6 and Table 7. The data encryption standard method attained a maximum entropy of 8.25, whereas the proposed method managed to attain 7.25 and the data decryption standard method attained a maximum entropy of 8.75, whereas the proposed method managed to attain 7.75, which was the least among the other methods. Maximum entropy represents the maximum disorder and the minimum entropy represents the minimum disorder of data. Thus, the proposed system developed provides a better means of secured transmission of data.
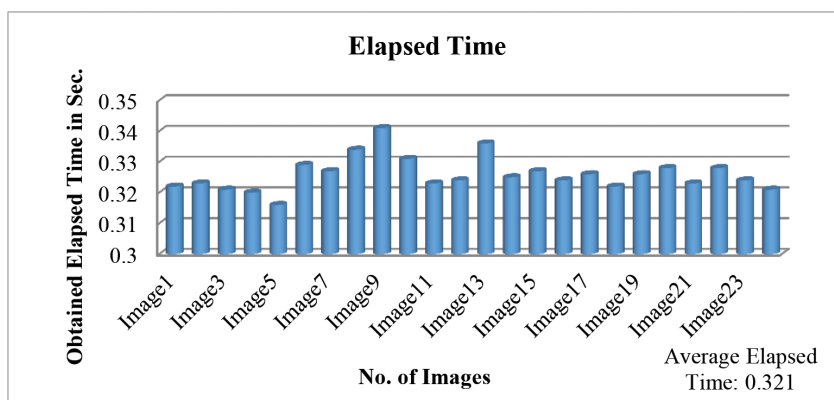
Figure 2. Graphical representation of elapsed time achieved for the input images

Table 2
Image quality measures between input image and encrypted images

Sl. No.	MSE	PSNR	MNCC	AD	SC	MD	NAE
Image 1	41.49	15.77	0.93	16.82	1.10	220	0.10
Image 2	40.23	16.04	0.98	-7.23	1.00	237	0.12
Image 3	45.85	14.90	1.00	-10.99	0.94	173	0.14
Image 4	36.92	16.78	0.94	13.54	1.07	222	0.09
Image 5	34.10	17.48	0.96	8.07	1.06	250	0.08
Image 6	42.05	15.65	0.94	3.80	1.06	224	0.12
Image 7	46.79	14.73	0.92	17.80	1.11	236	0.12
Image 8	30.56	18.42	0.98	-2.50	1.01	252	0.08
Image 9	29.93	18.60	0.98	5.97	1.02	181	0.08
Image 10	43.30	15.40	0.92	18.97	1.12	236	0.11
Image 11	33.20	17.70	0.98	-0.06	1.01	205	0.09
Image 12	52.05	13.80	0.89	23.76	1.19	245	0.14
Image 13	65.20	11.84	0.84	37.48	1.29	236	0.16
Image 14	43.86	15.29	0.91	23.48	1.16	175	0.11
Image 15	44.39	15.18	0.927	17.72	1.12	248	0.11
Image 16	82.69	9.78	0.80	46.38	1.39	247	0.21
Image 17	35.86	17.03	0.99	-1.58	0.99	220	0.10
Image 18	38.19	16.49	0.99	-5.42	0.96	198	0.11
Image 19	33.76	17.56	0.962	10.45	1.07	221	0.09
Image 20	47.28	14.63	0.926	16.29	1.12	238	0.12
Image 21	36.49	16.89	0.954	12.02	1.07	228	0.10
Image 22	46.87	14.71	0.905	19.01	1.16	217	0.12
Image 23	28.78	18.94	0.989	2.34	1.03	239	0.08
Image 24	34.38	17.40	0.95	12.89	1.08	235	0.09

MSE: Mean Squared Error; PSNR: Peak Signal-to-Noise Ratio; MNCC: M-Normalized Cross-Correlation; AD: Average Difference; SC: Structural Content: MD: Maximum Difference: NAE: Normalized Absolute Error

Table 3

Image quality measures between encrypted and decrypted images

Sl. No.	MSE	PSNR	MNCC	AD	SC	MD	NAE
Image 1	41.49	15.77	1.03	-16.81	0.90	207	0.11
Image 2	40.22	16.04	0.98	7.23	0.99	254	0.12
Image 3	45.85	14.90	0.94	10.99	1.06	253	0.12
Image 4	36.92	16.78	1.02	-13.54	0.92	180	0.10
Image 5	34.10	17.47	1.01	-8.07	0.94	182	0.08
Image 6	42.05	15.65	1.01	-3.80	0.93	202	0.12
Image 7	46.79	14.72	1.03	-17.80	0.89	190	0.13
Image 8	30.56	18.42	0.99	2.50	0.99	145	0.08
Image 9	29.93	18.60	1.00	-5.97	0.97	158	0.08
Image 10	43.30	15.40	1.04	-18.97	0.89	185	0.12
Image 11	33.20	17.70	0.99	0.06	0.98	197	0.09
Image 12	52.04	13.80	1.06	-23.76	0.84	173	0.16
Image 13	65.20	11.84	1.09	-37.48	0.77	103	0.19
Image 14	43.86	15.29	1.05	-23.47	0.86	111	0.13
Image 15	44.39	15.18	1.039	-17.72	0.89	168	0.12
Image 16	82.69	9.78	1.11	-46.38	0.71	189	0.26
Image 17	35.86	17.03	0.98	1.58	1.01	203	0.09
Image 18	38.19	16.49	0.97	5.42	1.02	178	0.10
Image 19	33.76	17.56	1.01	-10.45	0.93	164	0.09
Image 20	47.28	14.63	1.03	-16.29	0.89	191	0.13
Image 21	36.49	16.89	1.01	-12.01	0.93	189	0.10
Image 22	46.87	14.71	1.05	-19.00	0.86	188	0.14
Image 23	28.78	18.94	1.00	-2.34	0.97	170	0.07
Image 24	34.37	17.40	1.02	-12.89	0.92	163	0.09

Table 4

Error metric measurement for original images

Sl. No.	SNR	RMSE	UIQI	EME1	EME2	PCC1	PCC2	MAE
Image 1	-5.78	31.05	0.63	25.46	6.16	16704.49	19879	6.45
Image 2	-9.98	35.30	0.54	26.72	7.39	17749.23	19879	6.07
Image 3	-5.58	30.14	0.41	13.59	5.01	17035.99	19739	5.89
Image 4	-5.40	29.73	0.65	22.16	5.05	16058.24	20303	6.17
Image 5	-6.08	31.80	0.61	30.82	8.41	39404.04	44943	6.39
Image 6	-8.31	33.66	0.57	41.63	9.11	8158.30	8857	5.62
Image 7	-6.11	32.11	0.58	28.97	6.24	8009.28	9047	6.45
Image 8	-7.83	36.04	0.44	48.97	12.64	7221.16	8943	6.55

Table 4 (*continue*)

Sl. No.	SNR	RMSE	UIQI	EME1	EME2	PCC1	PCC2	MAE
Image 9	-5.95	31.13	0.41	21.20	3.27	7207.68	8735	6.55
Image 10	-5.75	30.80	0.57	31.49	10.99	7726.37	8797	6.08
Image 11	-7.00	33.41	0.70	64.75	12.08	7409.20	8669	6.47
Image 12	-6.63	33.24	0.64	36.94	8.80	7931.42	8669	6.26
Image 13	-6.86	35.18	0.46	7.62	1.98	6653.55	8834	6.42
Image 14	-6.01	31.92	0.42	9.89	2.52	7510.33	9029	6.43
Image 15	-6.30	32.87	0.57	29.29	6.21	8044.06	9063	6.59
Image 16	-7.49	37.73	0.42	15.39	3.22	11470.84	13224	6.20
Image 17	-5.76	30.82	0.67	30.46	8.94	7122.60	8548	6.26
Image 18	-5.44	29.86	0.43	19.95	6.96	7362.12	8631	6.03
Image 19	-6.88	34.89	0.69	44.47	7.28	7602.61	8924	7.49
Image 20	-6.37	32.96	0.80	49.59	7.71	8119.59	9029	6.53
Image 21	-5.57	30.24	0.54	28.28	11.23	7145.53	8599	6.20
Image 22	-6.48	32.65	0.64	44.46	10.06	8078.24	8839	6.04
Image 23	-6.32	32.72	0.53	36.34	10.00	7508.96	8651	6.14
Image 24	-5.87	31.33	0.68	34.15	6.01	7180.56	8669	6.59

SNR: Signal-to-Noise Ratio; RMSE: Root Mean Squared Error; UIQI: Universal Image Quality Index; EME1 (Original Image): Measurement of Enhancement; EME2 (Encrypted Image); PCC1 (Original Vs Noise): Pearson Correlation Coefficient; PCC2 (Original Vs Original): Pearson Correlation Coefficient; MAE: Mean Absolute Error

Table 5

Error metric measurement for encrypted images

Sl. No.	SNR	RMSE	UIQI	EME1	EME2	PCC1	PCC2	MAE
Image 1	-5.44	29.86	0.74	38.32	8.72	16135.62	19599	5.96
Image 2	-6.16	32.38	0.58	39.92	9.42	16832.39	19737	6.18
Image 3	-5.97	31.72	0.63	26.22	5.39	16856.93	19739	6.74
Image 4	-5.62	30.50	0.75	32.07	7.41	16334.77	20019	6.31
Image 5	-5.71	30.81	0.64	30.39	6.95	36983.29	44729	6.49
Image 6	-6.61	33.98	0.58	47.56	8.11	6440.08	8414	6.59
Image 7	-5.59	30.24	0.73	49.10	15.32	7188.09	8735	5.84
Image 8	-6.34	32.82	0.58	50.79	17.28	5451.67	8567	6.04
Image 9	-5.73	30.74	0.70	43.41	13.26	7593.14	8903	6.16
Image 10	-5.18	28.93	0.69	36.96	16.39	7066.68	8735	5.68
Image 11	-6.19	32.35	0.71	52.20	9.72	7117.01	8903	6.29
Image 12	-5.72	30.73	0.62	36.79	10.47	6672.96	8735	6.34
Image 13	-5.79	31.09	0.58	23.02	6.19	6562.29	9359	6.95

Table 4 (*continue*)

Sl. No.	SNR	RMSE	UIQI	EME1	EME2	PCC1	PCC2	MAE
Image 14	-5.61	30.44	0.61	28.79	7.54	6870.67	9151	6.39
Image 15	-6.13	32.20	0.73	45.07	10.47	7286.15	8988	6.68
Image 16	-5.63	30.52	0.50	26.89	6.88	6033.29	8199	6.77
Image 17	-5.88	31.29	0.78	46.00	8.69	7886.52	9221	6.37
Image 18	-5.81	31.17	0.70	29.80	5.41	7448.92	9047	6.60
Image 19	-6.33	32.99	0.70	47.99	14.47	7516.60	9239	6.76
Image 20	-5.82	31.08	0.78	50.52	9.21	7228.54	8714	6.47
Image 21	-5.56	30.176	0.72	41.24	7.77	7266.09	8857	6.06
Image 22	-5.56	30.23	0.65	32.59	9.18	6553.60	8499	6.40
Image 23	-6.03	31.66	0.69	47.98	9.34	7157.70	8839	6.01
Image 24	-5.18	29.00	0.71	38.55	5.19	6851.51	8857	5.76

Table 6

Comparison of different encryption methods with entropy value

Encryption Algorithm	Entropy Values
DES	8.2525
AES	7.8978
MAES	7.8991
Proposed Method	7.2548

Table 7

Comparison of different decryption methods with entropy value

Decryption Algorithm	Entropy Values
DES	8.7536
AES	8.3978
MAES	8.4002
Proposed Method	7.7559

CONCLUSION

An efficient system for the secure transmission of digital images was developed and presented in this paper. The chaotic confusion and pixel diffusion techniques were utilised for encryption and decryption, respectively. The effectiveness of the developed system was demonstrated based upon the elapsed time between the encryption and decryption of the digital images. A histogram representation showed that there was almost no loss of data during the encryption and decryption processes. Quality measures and error metrics were computed for all the input images, and the results were tabulated. It could be concluded that the developed system ensured that the image data could be transmitted in an effective and efficient manner.

REFERENCES

Acharya, B., Panigrahy, S. K., Patra, S. K., & Panda, G. (2009). Image encryption using advanced hill cipher algorithm. *International Journal of Recent Trends in Engineering, 1*(1), 243–247.

Bani Younes, M. A., & Jantan, A. (2008). Image encryption using block based transformation algorithm. *IAENG International Journal of Computer Science, 35*(1), 407–415.

Behnia, S., Akhshani, A., Mahmodi, H., & Akhavan, A. (2008). A novel algorithm for image encryption based on mixture of chaotic maps. *Chaos, Solitons and Fractals, 35*(2), 408–419.

Bremnavas, I., Poorna, B., & Mohamed, I. R. (2013). Secured medical image transmission using chaotic map. *Elixir Computer Science and Engineering 54*(1), 12598–12602.

Chang, C. C., Hwang, M. S., & Chen, T. S. (2001). A new encryption algorithm for image cryptosystems. *Journal of Systems and Software, 58*(2), 83–91.

Dinesh, M., & Suveetha, D. P. (2016). Real time image encryption and decryption using blowfish algorithm. *International Journal of Emerging Technology in Computer Science and Electronics (IJETCSE), 22*(2), 9–14.

Gunasekaran, G., & Ray, B. K. (2014). Encrypting and decrypting image using computer visualization techniques. *ARPN Journal of Engineering and Applied Sciences, 9*(5), 646–650.

Hashim, H. R., & Neamaa, I. A. (2014). Image encryption and decryption in a modification of ElGamal cryptosystem in MATLAB. *International Journal of Sciences: Basic and Applied Research (IJSBAR), 14*(2), 141–147.

Kamali, S. H., Shakerian, R., Hedayati, M., & Rahmani, M. (2010, August). A new modified version of advanced encryption standard based algorithm for image encryption. In *International Conference on Electronics and Information Engineering (ICEIE), 2010* (Vol. 1, pp. 141–145). IEEE.

Khan, J., Ahmad, J., & Hwang, S. O. (2015). An efficient image encryption scheme based on Henon map, skew tent map and S-Box. In *6th International Conference* on *Modeling, Simulation, and Applied Optimization (ICMSAO), 2015* (pp. 1–6). IEEE.

Lookabaugh, T., & Sicker, D. C. (2004). Selective encryption for consumer applications. *IEEE Communications Magazine, 42*(5), 124–129.

Mathur, M., & Kesarwani, A. (2013). Comparison between DES, 3DES, RC2, RC6, Blowfish and AES. In *Proceedings of National Conference on New Horizons, IT-NCNHIT* (Vol. 3, pp. 143–148). Mumbai, India.

Mitali, V. K., & Sharma, A. (2014). A survey on various cryptography techniques. *International Journal of Emerging Trends and Technology in Computer Science, 3*(4), 6, 307–312.

Nichat, S. P., & Sikchi, S. S. (2013). Image encryption using hybrid genetic algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(1), 427–432.

Panduranga, H. T., & Naveenkumar, S. K. (2013). Selective image encryption for medical and satellite images. *International Journal of Engineering and Technology (IJET), 5*(1), 115–121.

Phan, R. C. W. (2004). Impossible differential cryptanalysis of 7-round Advanced Encryption Standard (AES). *Information Processing Letters, 91*(1), 33–38.

Pisarchik, A. N., & Zanin, M. (2008). Image encryption with chaotically coupled chaotic maps. *Science Direct, Physical D: Nonlinear Phenomena, 237*(20), 2638–2648.

Princy, P. (2015). A comparison of symmetric key algorithms DES, AES, Blowfish, RC4, RC6: A survey. *International Journal of Computer Science and Engineering Technology (IJCSET), 6*(5), 328–331.

Priya, S., Malathi, M., & Jayaseelan, G. (2015) Secured internet banking using image captcha technique. *Proceedings of International Journal of Emerging Technology and Innovative Engineering, 1*(11), 1–6.

Ruman, K., & Phaneendra, H. D. (2015). Implementation of methods for transaction in secure online banking. *International Journal of Technical Research and Applications, 3*(4), 41–43.

Shazmeen, S. Y., & Prasad, S. (2012). A practical approach for secure internet banking based on cryptography. *International Journal of Scientific and Research Publications, 2*(12). 1–6.

Singh, J., Lata, K., & Ashraf, J. (2015). Image encryption & decryption with symmetric key cryptography using MATLAB. *International Journal of Current Engineering and Technology, 5*(1), 448–451.

Thakur, A., Kumar, R., Bath, A., & Sharma, J. (2014). Design of selective encryption scheme using MATLAB. *International Journal of Electrical & Electronics Engineering, 1*(1), 14–19.

Ullah, I., Iqbal, W., & Masood, A. (2013). Selective region based images encryption. In *2nd National Conference on Information Assurance (NCIA), 2013* (pp 125–128). IEEE.

Verma, O. P., Agarwal, R., Dafouti, D., & Tyagi, S. (2011, April). Performance analysis of data encryption algorithms. In *3rd International Conference on Electronics Computer Technology (ICECT), 2011* (Vol. 5, pp. 399–403). IEEE.

Wang, L., Ye, Q., Xiao, Y., Zou, Y., & Zhang, B. (2008). An image encryption scheme based on cross chaotic map. In *Image and Signal Processing* (Vol. 3*,* pp. 22-26). IEEE.

Wei-Bin, C., & Xin, Z. (2009). Image encryption algorithm based on Henon chaotic system. In *International Conference on Image Analysis and Signal Processing (IASP), 2009* (pp. 94–97). IEEE.

Zheng, Y., & Jin, J. (2015). A novel image encryption scheme based on Henon map and compound spatiotemporal chaos. *Multimedia Tools and Applications, 74*(18), 7803–7820.

SCIENCE & TECHNOLOGY

Socio-Demographic Factors Associated with Low Birth Weight in Nepal Using Imputation of Missing Determinants

Singh, U.[1], Ueranantasun, A.[1]* and Kuning, M.[1,2]

[1]*Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, 94000, Thailand*

[2]*Centre of Excellence in Mathematics, CHE, Si Ayutthaya Rd., Bangkok 10400, Thailand*

ABSTRACT

Birth weight is considered a factor that determines a baby's survival and predicts longstanding health. Hence, proper estimation of prevalence and determinants that are associated with low birth weight (LBW) is necessary for reduction of LBW. However, data on birth weight from developing countries are usually incomplete. Therefore, this study is aimed at handling missing values on determinants and identification of factors associated with LBW. The child data set from the Nepal Demographic and Health Survey (NDHS), 2011 was applied in this study. It contained information on 5240 children, but 3318 children had missing values on birth weight. Hence, in this study 1922 infants with measured birth weight were analysed. Multiple imputation was applied to handle missing values on determinants. The results revealed that only one variable was significant with LBW. Mothers residing in the Eastern and Far-Western development region were more inclined to give birth to LBW. The results obtained after handling missing values is considered less biased than assuming non-missing values or ignoring missing values. This suggests that missing values on determinants should be handled rather than ignored or assumed to be non-missing.

Keywords: Low birth weight, missing values, multiple imputation

INTRODUCTION

Weight at birth is recognised as a factor that determines infants' survival and forecasts long-term health (Barker, 2004; Wilcox, 2001). Globally, an estimate of 2.8 million infants die each year during the first 28 days of life, for which low birth weight (LBW) (< 2500 g) is a major factor (Lawn et al., 2014). LBW is also associated with morbid conditions such as stunting (Christian et al., 2013), cognitive

delay and behavioural disorders (Arcangeli et al., 2012; Frondas-Chauty et al., 2014) in children and cardio-vascular disease, hypertension and non-insulin dependent diabetes in adults (Barker, 2004). LBW still has a significant impact on public health. the World Health Organization (WHO) targets to reduce 30% of LBW globally by 2025 (WHO, 2014).

Every year, more than 20 million births are recorded globally, out of which 15.5% are LBW. A total of 95.6% of all LBW infants are born in developing countries (UNICEF & WHO, 2004). The prevalence of LBW is highest in South Asia (28%), despite the fact that more than half of the women in this region give birth at home and are unable to provide numeric birth weight of their babies (WHO, 2014). In Nepal the prevalence of LBW is 11.5% (MOHP & New ERA, 2011). Studies conducted on LBW have identified factors like mother's age at child's birth (Reichman & Teitler, 2006), mother's education (Muula et al., 2011), consumption of iron and folic acid during pregnancy, attendance of antenatal visit (ANC) during pregnancy (Khanal, Zhao, & Sauer, 2014b; Nisar & Dibley, 2014), gender of child (Sreeramareddy et al., 2011), economic status and residence (Kayode et al., 2014). Birth weight is considered the main indicator of neonatal and infant health (Almond et al., 2002). In Nepal, neonatal mortality rate is 33 per 1000 live births, which has been constant over the past five years. Nepal is a developing country (World Bank, 2015), and correct estimation of percentage of LBW and its determinants is necessary for intervention programmes for the reduction of infant and neonatal mortality. However, data on birth weight in Nepal consist of highly missing values, because two thirds of women give birth at home (MOHP & New ERA, 2011). Studies conducted in Nepal have used either the mother's opinion on birth size as a proxy to the birth weight (Khanal, Sauer, Karkee, & Zhao, 2014b) or only a sample of complete measured birth weight (Khanal et al., 2014b). However, missing values presented on determinants were not handled in the aforementioned studies. Analysis without acknowledging missing data for determinant variables and including only complete case data tends to be more biased and has less statistical power and precision (Sterne et al., 2009), and the results from studies that ignore missing data have been proven to be different and inferior to those from studies that impute missing data (Moons et al., 2006). Thus, this study aimed to handle missing data presented on determinants and identification of factors associated with LBW.

METHODS

This research was a cross-sectional study in which data were obtained from a survey. The study aimed to identify factors associated with LBW. All study variables were obtained from the survey data. LBW was considered as an outcome. Variables such as mother's age at child's birth, education, attendance of antenatal care and consumption of iron tablet during pregnancy, smoking, parity, women's decision for utilisation of health services, gender of child, birth interval, wealth index, ethnicity, cooking fuel, residence, ecological region and development region were included as determinants. Since the data contained missing values on determinants, multiple imputation was employed to handle missing data. Finally, logistic regression was used to identify factors associated with LBW.

Nepal Demographic and Health Survey (NDHS) data

This study utilised secondary data from NDHS 2011, a nationally representative cross-sectional household survey provided by New ERA under the aegis of the Ministry of Health and Population (MOHP). The most recent survey data of 2016 were not included in this study because they were not available at the study time. In this survey, multistage cluster sampling was used. In the first stage, probability proportionate to size was used to select wards from rural and sub-wards from urban areas. In the second stage, random sampling was done among select households. The survey interviewed 12,674 women aged 15 to 49 and 4121 men aged 15 to 59. In this study, child data set, which included information on children under five years old was used. Infant's birth weight was considered a binary outcome; that is, normal birth weight was taken as >=2500 g and LBW was taken as <2500 g. Conditions such as maternal factors, child factors and socio-demographic factors were considered exposure variables. Variables under maternal factors included age at child's birth, education, attendance of antenatal care and consumption of iron tablets during pregnancy, smoking, parity and women's decision for utilisation of health services. Child conditions included gender of child and birth interval. Determinants like wealth index, ethnicity, cooking fuel, residence, ecological region and development region were included in the socio-demographic factors. Here, all study variables were categorical and the categorisation of the study variables was based on previous studies that used similar Demographic and Health Survey (DHS) data sets conducted in Nepal (Khanal et al., 2014a, 2014b). A detailed explanation of the calculation of wealth index is given in Section 2.2 of the survey (MOHP & New ERA, 2011). The data set contained information on 5306 children. Twins, triplets and other multiple-birth children are more likely to be LBW (Gomella et al., 2004). Hence, all the 66 multiple births were excluded from this study. However, out of 5240 children, only 1922 infants' birth weight was available and was used in the present study. Out of 1922 infants, 417 (22%) infants contained missing values in at least one of the measured variables.

Data Management

Missing values were presented in the determinants of the samples numbering 1922. The frequency and percentage of missing values in each determinant is shown in Table 1.

Table 1
Frequency and percentage of missing data

Variable	Frequency of Missing Values	Percentage of Missing Values
ANC visit during pregnancy	273	14.2
Iron tablet consumption during pregnancy	274	14.3
Women's decision	148	7
Cooking fuel	17	< 1

There are three types of missing mechanism under which missing data occur i.e. missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (Rubin, 1976). Statistically, Little's test can be used to identify data under MCAR mechanism, but there are no clear tests to diagnose the MAR and MNAR mechanism; these can only be reasoned or hypothesised (Schafer & Graham, 2002). Under the MCAR mechanism, missing data are handled by means of imputation methods without any bias, but this type of mechanism seldom occurs (Schafer & Graham, 2002). However, under the MNAR mechanism, there exists no appropriate method to handle missing data (Little & Rubin, 2002). Most of the missing data hold MAR assumption and under this mechanism, the multiple imputation method is considered the appropriate method to handle missing values (Schafer & Graham, 2002). Even though this study did not focus on the causes of missing determinants, determination of the missing mechanisms was considered. In the present study, missing values were presented in the determinant, ANC visit. This was probably due to the perception that most mothers, especially those who live in rural areas, usually do not open up on the ANC visit during survey interviews. The direct consequence of missing values in the ANC visit was that there were missing values in the consumption of iron tablets. In the DHS, mothers belonging to the household (*de jure*) and those who did not belong to the household (non-*de jure*) were interviewed. However, questions related to cooking fuel were collected at the household level and assigned to individuals in the individual data file. Thus, a mother who was not a member of the household, would in turn lack data on cooking fuel. It was clear that missing values on the determinants were not missing due to themselves, but were missing due to other factors. Hence, missing data held MAR assumption; thus, the multiple imputation method was applied to handle missing data.

Statistical Analysis

It is suggested that the number of imputation should be at least equal or greater than the percentage of missing (White et al., 2011). Hence, multiple imputation was carried out 15 times because the highest percentage of missing data was 14.3%. All data management and analyses were carried out in R version 3.1.1 (R Core Team, 2015). For the iterations, the MICE package (van Buuren & Groothuis-oudshoorn, 2011) was deployed and for combining each imputed data set MiTools was used (Lumley, 2015). After imputation, a survey package (Lumley, 2010) was applied to each imputed data set to adjust for a complex sample design and sampling method. Survey logistic regression was carried out to identify determinants associated with LBW. The overall p-value for each variable was calculated using the Likelihood ratio test and the p-value for each subgroup was calculated using the Wald test.

Ethics

Ethical approval for conducting the NDHS survey was obtained by the Nepal Health Research Council, Nepal and the ICF Macro International Review Board in Calverton, Maryland, USA and the data analysis protocol was approved by the Curtin University Human Research Ethics Committee.

RESULTS

Preliminary Results

The overall and subgroup estimation of LBW prevalence and their 95-percent confidence intervals (95% CI) was calculated, and is presented in Table 2. The prevalence of LBW was found to be 11.5% (95% CI=10-13%). The prevalence of LBW was different in each category of factors except for the determinants such as mother's age at child's birth, consumption of iron tablets during pregnancy, gender of child, cooking fuel, ethnicity, wealth index, residence and ecological region, which were almost equal in each subgroup. The percentage of giving birth to LBW infants was highest among mothers who had primary education (14.6%) compared with the other two categories. The prevalence of having LBW infants among mothers who had attended one or more ANC visits (12.1%) and four or more ANC visits (11.4%) was close and higher than for mothers with no ANC visit (9.7%). The prevalence of giving birth to LBW babies was greater among smoker mothers (11.5%) than among non-smoking mothers (9.6%). For the variables such as parity and birth interval, the pattern of percentage of LBW in each subgroup was similar. The greatest percentage of LBW was found among the mothers who gave birth to two to three children (13%) within an interval of less than 24 months (13.1%). For the development region, the highest prevalence of having LBW infants was noticed among mothers residing in the Eastern development region (15%) and Far-Western development region (14.2%) compared with the other subgroups.

Table 2

Overall and subgroup prevalences of LBW after imputation

Variables	Birth Weight (Mean)	95% CI
Overall	11.5	9.6, 13.4
Maternal Factors		
Age at Child's Birth (Years)		
15-19	12.3	8.5, 16.0
20-24	11.4	8.7, 14.0
25-29	11.6	7.7, 15.4
≥30	10.6	5.7, 15.4
Education		
No Education	9.0	6.2, 12.1
Primary Education	14.6	10.6, 18.7
Secondary/Higher Education	11.6	8.5, 14.6
Body Mass Index (BMI)		
<18.5 (Underweight)	13.5	7.9, 18.8
18.5-23.0 (Normal)	12.3	8.5, 14.7
>23.0 (Overweight)	9.0	6.0, 11.7
ANC Visit During Pregnancy		
No Visit	9.7	1.0, 19.2
One-Three Visits	12.1	8.0, 16.0

Table 2 (*continue*)

Variables	Birth Weight (Mean)	95% CI
Four or More Visits	11.4	9.0, 13.7
Consumption of Iron Tablets During Pregnancy		
No	11.0	4.1, 18.3
Yes	11.5	9.8, 13.5
Women and Husbands Together	11.8	9.0, 14.5
Husband or Others	13.4	10.3, 16.5
Child Factors		
Birth Interval		
No Interval	12.3	10.0, 14.7
<24 months	13.1	7.1, 19.0
≥24 months	10.2	7.6, 12.7
Gender of baby		
Male	10.7	8.3, 13.1
Female	12.4	9.3, 15.4
Socio-Demographic Factors		
Cooking Fuel		
Low Polluting Fuel	10.8	7.7, 14.0
High Polluting Fuel	11.9	9.4, 14.3
Wealth Index		
Poor	11.8	8.6, 15.0
Middle	10.8	7.4, 14.3
Rich	11.6	8.8, 14.5
Ethnicity		
Relatively Advantaged	12.2	9.5, 15.0
Relatively Disadvantaged (Janjati)	10.2	7.4, 13.0
Relatively Disadvantaged (Dalit)	10.9	6.8, 15.0
Residence		
Rural	11.6	9.0, 13.2
Urban	11.4	9.2, 14.0
Ecological region		
Mountain	12.6	8.2, 14.4
Hill	11.9	8.9, 15.0
Terai	11.2	8.6, 13.8
Development Region		
Eastern	8.6	6.2, 11.1
Central	15.0	10.2, 19.7
Western	9.0	4.6, 14.1
Mid-Western	12.3	8.1, 16.5
Far-Western	14.2	9.7, 18.8

Factors Associated with LBW

Simple survey logistic regression was applied to all the studied variables as univariate analysis. The results obtained from this analysis are shown in Table 3. Even though the *p*-value calculated from Wald's test for women's decision for utilisation of health services was significant in Wald's test (*p*-value=0.032), the overall *p*-value calculated from the Likelihood ratio test was 0.105. Hence, only the development region was considered statistically significantly. Mothers residing in the Eastern development region (odds ratio [OR 1.86, 95% CI=1.12, 3.01]) and the Far-Western development region (OR 1.75, 95% CI= 1.08, 2.84) were more likely to have LBW infants. However, determinants such as mother's age at child's birth, mother's education, ANC visit and consumption of iron tablets during pregnancy, smoking, parity, birth interval, child's gender, wealth index, ethnicity, cooking fuel, residence and ecological region were insignificant.

Table 3

Unadjusted odds ratio and 95% CI of study variables

Variables	Unadjusted OR	95% CI	*p*-value
Maternal Factors			
Age at Child's Birth (Years)			*0.950*
≥30	1.00		
25-29	1.10	0.56, 2.18	0.774
20-24	1.08	0.61, 1.92	0.791
15-19	1.18	0.66, 2.10	0.579
Education			*0.105*
Secondary/Higher Education	1.00		
Primary Education	1.31	0.81, 2.12	0.267
No Education	0.77	0.49, 1.23	0.280
ANC Visit During Pregnancy			*0.894*
Four or More Visits	1.00		
One-Three Visits	1.07	0.70, 1.65	0.826
No Visit	0.82	0.27, 2.46	0.869
Consumption of Iron Tablets During Pregnancy			*0.940*
Yes	1.00		
No	0.94	0.44, 1.99	0.984
Smoke			*0.687*
No	1.00		
Yes	0.81	0.30, 2.21	0.687
Parity			*0.107*
Four and Above	1.00		
Two-Three	1.67	0.96, 2.89	0.070
One	1.33	0.72, 2.46	0.365

Table 3 (*continue*)

Variables	Unadjusted OR	95% CI	*p*-value
Women's Decision For Health Service Utilisation			*0.105*
Women	1.00		
Women and Husbands Together	1.46	0.92, 2.31	0.106
Husband or Others	1.68	1.04, 2.72	0.032*
Child Factors			
Birth Interval			*0.373*
No Interval	1.00		
<24 months	1.07	0.64, 1.82	0.790
≥ 24 months	0.81	0.58, 1.12	0.205
Gender of Baby			0.401
Male	1.00		
Female	1.18	0.80, 1.72	0.401
Socio-Demographic Factors			
Wealth Index			*0.921*
Rich	1.00		
Middle	0.92	0.58, 1.47	0.740
Poor	1.02	0.69, 1.51	0.930
Ethnicity			*0.576*
Relatively Advantaged	1.00		
Relatively Disadvantaged (Janjati)	0.82	0.55, 1.21	0.310
Relatively Disadvantaged (Dalit)	0.88	0.54, 1.42	0.599
Cooking Fuel			*0.668*
Low Polluting Fuel	1.00		
High Polluting Fuel	1.11	0.73, 1.68	0.668
Residence			*0.941*
Urban	1.00		
Rural	0.99	0.70, 1.38	0.941
Ecological Region			*0.846*
Terai	1.00		
Hill	1.07	0.73, 1.58	0.727
Mountain	1.15	0.71, 1.86	0.579
Development region			*0.039*
Central	1.00		
Eastern	1.86	1.12, 3.01	0.013*
Western	1.09	0.57, 2.07	0.794
Mid-Western	1.49	0.90, 2.44	0.119
Far-Western	1.75	1.08, 2.84	0.024*

Note: *p*-value in italic form was calculated from Likelihood ratio test, *statistically significant at 5% level.

DISCUSSION

The overall prevalence of LBW in this study was 11.5% and this finding was similar to that of a study conducted by Khanal et al. (2014b). That study also reported a prevalence of 11.5%. This is because the above-mentioned study also excluded infants with missing birth weight from the analysis. The prevalence of LBW for determinants such as mother's age at child's birth, consumption of iron tablets during pregnancy, gender of child, wealth index, ethnicity, cooking fuel, residence and ecological region were nearly equal for each category. This finding reflected that each category had the same possibility of having LBW.

The current study revealed that mothers residing in the Eastern development region and the Far-Western development region were more prone to delivering LBW infants. This is consistent with the study by Khanal et al. (2014a, 2014b). The Far-Western development region of Nepal depicts remoteness, less development of basic infrastructure and few livelihood opportunities (United Nations Nepal Information Platform, 2012). Due to this, women living in this region are deprived of nutritious food during pregnancy and are hindered from going to health facilities. As a result, there is a higher chance of having adverse outcome pregnancy outcomes.

In this study, determinants such as ANC visit and consumption of iron tablets during pregnancy were not significantly associated with LBW. This result differs from that documented by Khanal et al. (2014a, 2014b), who found the above-mentioned variables significant. This is because the aforementioned studies presumed that missing values presented on ANC visits and consumption of iron tablet during pregnancy as 'no ANC visit' and 'no consumption', respectively. Consequently, there is a greater sample for 'no ANC visit' and 'no consumption'. Hence, ANC visit and consumption of iron tablets were found significant as a factor in LBW in the above-mentioned studies.

The current study assumed that missing data were MAR and multiple imputation was considered the best method to handle missing data when the missing data held the MAR mechanism. Therefore, missing values presented on determinants were handled using multiple imputation and were not ignored. White et al. (2011) suggested that the number of imputations should be at least equal or greater than the number of missing values. Therefore, imputation was carried out 15 times because the highest percentage of missing value was 14.3%. Although the current study and the study conducted in Nepal (Khanal et al., 2014b) analysed the same subset of the sample (1,922), the results were not consistent. In the present study, missing data on variables such as ANC visit, consumption of iron tablets during pregnancy, cooking fuel and women's decision on utilisation of health services were handled through multiple imputation rather than ignored or assumed to be missing values as non-missing. Using imputation for replacing missing values does not remove bias completely, but it does reduce bias and this is preferred to ignoring bias altogether (Lumley, 2010). Therefore, handling missing values on determinants is a better option to presuming values to be not missing or ignoring it.

The limitation of this study was that infants with missing birth weight were excluded from this study. Furthermore, this study analysed secondary data; therefore, the reason for missing values was not confirmed.

CONCLUSION

The prevalence of LBW was 11.5%, which was still high, in spite of deleting infants with missing birth weight. This finding suggested that there is a need for implementation of programmes focused on reduction of LBW in Nepal, especially targeting mothers living in the Eastern and the Far-Western development region of Nepal.

The findings revealed that assuming missing values as non-missing or ignoring missing values provide different results. Therefore, it is suggested that future researchers handle missing data and not assume values are non-missing or ignore them.

ACKNOWLEDGEMENT

REFERENCES

Almond, D., Chay, K. Y., & Lee, D. S. (2002). Does low birth weight matter? Evidence from the US population of twin births. *Working Paper No. 53.* Retrieved from http://cle.berkeley.edu/working_papers.shtm1#4

Arcangeli, T., Thilaganathan, B., Hooper, R., Khan, K. S., & Bhide, A. (2012). Neurodevelopmental delay in small babies at term: A systematic review. *Ultrasound in Obstetrics and Gynecology*, *40*(3), 267–275. http://doi.org/10.1002/uog.11112

Barker, D. J. P. (2004). The developmental origins of chronic adult disease. *Acta Paediatrica*, *93*(446), 26–33. http://doi.org/10.1080/08035320410022730

Christian, P., Lee, S. E., Donahue Angel, M., Adair, L. S., Arifeen, S. E., Ashorn, P., … & Black, R. E. (2013). Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. *International Journal of Epidemiology*, *42*(5), 1340–55. http://doi.org/10.1093/ije/dyt109

Frondas-Chauty, A., Simon, L., Branger, B., Gascoin, G., Flamant, C., Ancel, P. Y., … & Rozé, J. C. (2014). Early growth and neurodevelopmental outcome in very preterm infants: impact of gender. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *99*(5), F366–F372. http://doi.org/10.1136/archdischild-2013-305464

Gomella, T. L., Cunningham M. D., Eyal F. G., & Zenk, K. E. (1999). *Neonatology: Management, procedures, on-call problems, diseases and drugs.* New York, NY: McGraw-Hill.

Kayode, G. A, Amoakoh-Coleman, M., Agyepong, I. A., Ansah, E., Grobbee, D. E., & Klipstein-Grobusch, K. (2014). Contextual risk factors for low birth weight: A multilevel analysis. *PloS One*, *9*(10), e109333. http://doi.org/10.1371/journal.pone.0109333

Khanal, V., Sauer, K., Karkee, R., & Zhao, Y. (2014a). Factors associated with small size at birth in Nepal: Further analysis of Nepal demographic and health survey 2011. *BMC Pregnancy and Childbirth*, *14*(1), 32-40. http://doi.org/10.1186/1471-2393-14-32

Khanal, V., Zhao, Y., & Sauer, K. (2014b). Role of antenatal care and iron supplementation during pregnancy in preventing low birth weight in Nepal: Comparison of national surveys 2006 and 2011. *Archives of Public Health*, *72*(1), 4-13. http://doi.org/10.1186/2049-3258-72-4

Lawn, J. E., Blencowe, H., Oza, S., You, D., Lee, A. C. C., Waiswa, P., … & Barros, A. J. D. (2014). Every newborn: Progress, priorities, and potential beyond. *The Lancet, 384*(9938), 189–205. http://doi.org/10.1016/S0140-6736(14)60496-7

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd Ed.). New York, NY: John Wiley & Sons, Inc.

Lumley, T. (2010). *Complex surveys: A Guide to analysis using R.* Hoboken, NJ: Wiley.

Lumley, T. (2015). *MiTools: Tools for multiple imputation of missing data, R version 2.3*. Retrieved from http://cran.r-project.org/web/packages/mitools/

MOHP & New ERA. (2011). *Nepal demographic and health survey (NDHS) 2011*. Kathmandu, Nepal: Ministry of Health and Population, New ERA, and ICF International, Calverton, Maryland.

Moons, K. G. M., Donders, R. A. R. T., Stijnen, T., & Harrell, F. E. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology*, *59*(10), 1092–1101. http://doi.org/10.1016/j.jclinepi.2006.01.009

Muula, A. S., Siziya, S., & Rudatsikira, E. (2011). Parity and maternal education are associated with low birth weight in Malawi. *African Health Sciences*, *11*(1), 65–71.

Nisar, Y. B., & Dibley, M. J. (2014). Antenatal iron-folic acid supplementation reduces risk of low birthweight in Pakistan: Secondary analysis of demographic and health survey 2006-2007. *Maternal and Child Nutrition, 12*(1), 85-98. http://doi.org/10.1111/mcn.12156

R Core Team. (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, Retrieved from http://www.R-project.org

Reichman, N. E., & Teitler, J. O. (2006). Paternal age as a risk factor for low birthweight. *American Journal of Public Health*, *96*(5), 862–866. http://doi.org/10.2105/AJPH.2005.066324

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. http://doi.org/10.2307/2335739

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. http://doi.org/10.1037//1082-989X.7.2.147

Sreeramareddy, C. T., Shidhaye, R. R., & Sathiakumar, N. (2011). Association between biomass fuel use and maternal report of child size at birth – An analysis of 2005-06 India demographic health survey data. *BMC Public Health*, *11*(1), 403-412. http://doi.org/10.1186/1471-2458-11-403

Sterne, J.A.C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., … & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *British Medical Journal*, *338*, b2393. http://doi.org/10.1136/bmj.b2393

UNICEF & WHO. (2004). *Low birthweight: Country, regional and global estimates*. New York, UNICEF & WHO. Retrieved from https://www.unicef.org/publications/index_24840.html

UNNIP. (2012). *Women from the mid and far west*. United Nations Nepal Information Platform. Retrieved from http://un.org.np/oneun/undaf/women_mid_far.

van Burren, S., Groothuis-oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67. http://dx.doi.org/10.18637/jss.v045.i03

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–99. http://doi.org/10.1002/sim.4067

World Health Organisation. (2014). *WHA global nutrition targets 2025; Low birth weight policy brief.* Geneva: WHO. Retrieved from http://www.who.int/nutrition/publications/globaltargets2025_policybrief_lbw/en/

Wilcox, A. J. (2001). On the importance–and the unimportance–of birthweight. *International Journal of Epidemiology*, *30*(6), 1233–1241. http://doi.org/10.1093/ije/30.6.1233

World Bank. (2015). *Country and lending groups*. Retrieved from http://data.worldbank.org/about/country-and-lending-groups.

SCIENCE & TECHNOLOGY

Production of Lentiviral Vector with Polycistronic Transcripts for Reprogramming of Mouse Fibroblast Cells

Akram Al Abbar[1], Norshariza Nordin[1,2], Siew Ching Ngai[3] and Syahril Abdullah[1,2,4]*

[1]Medical Genetics Laboratory, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[2]Genetics and Regenerative Medicine Research Centre, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[3]School of Biosciences, Faculty of Science, University of Nottingham Malaysia, Jalan Broga, 43500 Semenyih, Selangor, Malaysia

[4]Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

iPS cells were originally generated using monocistronic retroviral vectors carrying the Yamanaka factors 'OSKM'. The development of a polycistronic viral vector with OSKM linked by 2A peptides has simplified reprogramming procedure and reduced the risk of multiple proviral integrations and insertional mutagenesis. In this study, we demonstrated the production of the polycistronic lentiviral vector encoding OSKM in a single cassette without a reporter gene or drug-based selection system. Syncytia formations were clearly seen following the co-transfection of a lentiviral plasmid construct with the structural and packaging plasmids. The virion was collected at 48 hours post-transfection. Afterwards, the viral titers were measured by the expression of Sox2 protein from transduced HT1080 cells. Subsequently, Oct4 expression was successfully detected in mouse fibroblasts in the range of 5, 10 and 20 MOIs with expression of 90.7%, 97.5% and 98%, respectively. The results obtained from this study could be used as a model for the production of OSKM lentiviral vector for newcomers to cellular reprogramming research.

Keywords: 2A peptides, iPS cells, lentivirus production, OSKM, polycistronic vectors

INTRODUCTION

The first iPS cells were generated from Oct4, Sox2, Klf4, and c-Myc transcription factors or 'OSKM', which were delivered by monocistronic retroviral vectors (Takahashi & Yamanaka, 2006). Following this, interest in research into iPS cells spread rapidly, and a

number of strategies have been developed to eliminate the random transgene integration into the genome that may lead to accidental activation of proto-oncogene (Lai et al., 2011; Nordin et al., 2011). In order to reduce the risk, a polycistronic vector was developed, through which OSKM genes can be delivered in one reprogramming cassette. The mRNA of each gene is translated independently to generate separate proteins for cellular conversion (Kaji et al., 2009; Sommer et al., 2009). This strategy is able to convert mouse and human fibroblasts into iPS cells, albeit at low conversion efficiency (Carey et al., 2008). This obstacle can be resolved by the insertion of an internal ribosome entry site (IRES) in combination with multiple promoters. Unfortunately, the use of IRES-based polycistronic vectors is limited due to the substantially lower levels of downstream cistron expression (Chang et al., 2009; Gonzalez et al., 2009).

To overcome this constraint, a polycistronic vector containing 2A 'self-cleaving' peptides from foot-and-mouth disease virus (FMDV) linking the OSKM was generated to express a sufficient level of OSKM for cellular reprogramming (Carey et al., 2008; Kim et al., 2011). In addition, the 2A 'self-cleaving' polycistronic vector system could express the four proteins in near equimolar amounts *in vivo* (Szymczak et al., 2004). As yet, successful production of a polycistronic vector separated by 2A sequences without a reporter gene or drug selection in a serum-free medium has not been reported.

An HIV-based vector has been the preferred strategy for *in-vivo* gene transfer. The genome of the HIV lentiviral vector construct has been engineered to enhance its biosafety and expression profiles (Jurgens et al., 2001; Zufferey & Dull, 1998). Two of the most interesting features of this vector are its ability to transduce proliferating and non-dividing cells and to integrate its own genome into the host genome, an action that may lead to prolonged expression of the transgene (Hotta et al., 2009; Kutner, Zhang, & Reiser, 2009). InvivoGen has developed the LENTI-Smart™ lentiviral system equipped with two essential packaging plasmids, pLV-iVSV-G and pLV-HELP containing *gag*, *pol*, *rev* and *tat*, that supply structural and replication proteins *in trans* for complete virion production. These plasmids are pre-complexed with LyoVec™ transfection reagent forming lyophilizate. In addition, the transfer vector genome encoding the Yamanaka factors, pLV-OSKM, was also designed in a single polycistronic transcript separated by the 'self-cleaving' peptides.

In our study, a polycistronic lentiviral vector from InvivoGen encoding the Yamanaka factors was produced in a serum-free environment, concentrated and then evaluated for transduction efficiency by analysing the expression of the transcription factor genes. The Multiplicity of Infection (MOI) was determined using tail-tip fibroblast infection of an adult C57BL/6 mouse.

MATERIALS AND METHODS

Cell Lines and Mouse Cell Culture

Human embryonic kidney (293FT) and human fibrosarcoma (HT1080) cell lines were purchased from Invitrogen and American Type Culture Collection (ATCC), respectively. The C57BL/J mouse strain was purchased from Jackson Laboratory. The tail-tip fibroblasts mouse cells were isolated as previously described (Takahashi & Yamanaka, 2006). The experimental procedure

was approved by the Animal Care and Use Committee (ACUC) of the Faculty of Medicine and Health Sciences, Universiti Putra Malaysia (UPM/FPSK/PADS/BR-UUH/00427). Cell lines 293FT and HT1080 were cultured in the DMEM medium, whereas the mouse cells were cultured in the DMEM/F12 medium. The medium was supplemented with 10% FBS, 1× non-essential amino acid, 2 mM GlutaMAX® and 1 mM MEM sodium pyruvate. Five hundred microgram of Geneticin® was used to maintain the sterility of the 293FT cells. All the products were purchased from Invitrogen.

Lentivirus Particle Production

In a T-75 flask, 1×10^7 293FT cells were suspended in 12 ml of a medium supplemented with KnockOut® Serum Replacement (SKR) (Invitrogen). The cells were co-transfected with 12 µg polycistronic plasmid encoding murine Oct4, Sox2, Klf4 and c-Myc (pLV-OSKM) and the LENTI-Smart™ (InvivoGen) structural and packaging plasmids. After 12 hours of co-transfection, the medium was discarded, and replaced with 12 ml of fresh DMEM medium. Virus particles in the supernatant were harvested over a period of 48 h, centrifuged, filtered through low protein binding 0.4 µM filters (Millipore), concentrated at 650×g for 30 min using Amicon Utra 100 kDa filter (Millipore) and stored at -80°C in aliquots. The culture medium used was antibiotic free, while the transfection was performed in a suspension cell state.

Flow Cytometry Titration Method

Virus titration was performed using the HT1080 suspension cell state transduction procedure. The virus solution was diluted to 10^{-1}, 10^{-5} and 10^{-6} in a complete DMEM medium containing 8 µg/ml polybrene (Sigma-Aldrich). The cells (1×10^5) were transduced with the diluted virus on a 48-well gelatin coated plate at a final volume of 200 µl. Upon 4 h of incubation, the transduction medium was discarded and replaced with 500 µl of fresh culture medium, after washing with 1X PBS. At 72 h post-transduction, the positive population of Sox2-transduced cells was assessed using flow cytometry. The data were converted into numerical value using the following formula: (F x C/V) x D, where F = frequency of Sox2$^+$ cells (percentage obtained divided by 100); C = total number of cells in the well at the time of transduction; V = volume of inocula in ml; D = LV dilution (Ngai et al., 2012a).

Lentiviral-OSKM Transduction of Mouse Fibroblasts

Mouse cells (1×10^5) were transduced with 5, 10 and 20 MOI of the virus in 1 ml complete DMEM medium containing 8 µg/ml polybrene (Sigma-Aldrich) in a 12-well gelatin-coated plate. At 12 h post-transduction, the medium was aspirated and replaced with 2 ml of mouse reprogramming medium (KnockOut™ DMEM supplemented with 15% KnockOut™ Serum Replacement (SKR), nonessential amino acids 0.1 mM, Gluta-MAX™ 2 mM and 2-mercaptoethanol 1×) (Invitrogen). The percentage of Oct4-expressing cells was measured by flow cytometry and immunochemistry at day 3 post-transduction.

RT-PCR Analyses

Total RNA was harvested from the transduced HT1080 cells using RNeasy® plus kit (Qiagen). The RNA was then subjected to cDNA synthesis using the QuantiTect® reverse transcription kit (Qiagen). Complementary DNA was synthesised from nearly 1 µg of total RNA using a random primer and the reverse transcription system (Qiagen). PCR was carried out with GoTaq® flexi buffer final concentration (1×), MgCl$_2$ solution (1.5 mM), dNTP (0.2 mM), forward/reverse primers (1 µM) and GoTaq® DNA polymerase (1.25u). These products were purchased from Promega. A sample DNA (250 ng) was also used. The synthesised cDNA was amplified by 35 cycles of PCR, at 55 to 65°C annealing temperature.

The primer sequences for exogenous mouse marker genes are: Oct4 (1024 bp: F,C CCGCTAGCCTGCCTTCTCCCTCCTGTG); Oct4 (1024 bp: R,CGCGGATCCACTG TCTAGAGTTTGAATGCATGG), Sox2 (996 bp: F,CCCGCTAGCAGGTCCCATGT ATAACATGATGGA), Sox2 (996 bp: R,CGCGGATCCCAGAGAGAAGTTCGTGG CTCT), c-Myc (1436 bp: F,CCCGCTAGCGGTGACGTGGAGGAGAATCC), c-Myc (1436 bp: R, CGCGGATCCGTACTGCTAGCTTTATGCACCAGA), β Actin (150 bp: F,CCTGTCAGCAATGCCTGGGT), β Actin (150 bp: R, CCAGCCTTCCTTCTTGGGTA).

Detection of Exogenous Protein Expression

Cells were collected, gently fixed and permeabilised using Cytofix/Cytoperm (BD™). The cells were stained with primary antibodies on ice, Oct4 (1.1) mouse IgG$_{2b}$ (mouse, monoclonal), rabbit anti mouse/human Sox2 (Stemgent, USA) or relevant isotype controls for 45 min. After washing, the cells were stained with secondary antibodies, anti-mouse IgG$_{2b}$- Alexa (Fluor®488) and anti-rabbit Alexa (Fluor®488) on ice (Molecular, Probes UAS) for 30 min. The stained cells were subsequently analysed using a flow cytometer (FC500, BD Biosciences). For immunocytochemistry analysis, the transduced cells were fixed with 4% paraformaldehyde and permeabilised with 0.1% Triton X -100. This was followed by washing three times with 1X PBS and incubation with blocking buffer (10% goat serum and 0.3% BSA) (Sigma-Aldrich). The cells were incubated overnight at 4°C with Oct4 mouse IgG$_{2b}$ (mouse, monoclonal- Santa Cruz, Bio) and Klf4 (mouse anti-mouse/human- Stemgent, Cambridge) primary antibodies. Next, the cells were incubated with the secondary antibodies, anti-mouse Alexa Fluor®IgG$_{2b}$, anti-rabbit Alexa Fluor®488 and anti-mouse Cy™3 (Molecular Probes and Stemgent, USA) at room temperature. Stained cells were mounted in DAPI (4, 6-diamidino-2-phenylindole) for 10 min at room temperature and washed three times with 1X PBS before nuclear visualisation using Olympus fluorescence microscope.

Statistical Analysis

One-way ANOVA was performed to determine the statistical significance between different serial dilutions using SPSS (Version 20.0, Chicago: SPPS Inc). The value $p<0.05$ was considered as statistically significant.

RESULTS AND DISCUSSION

Syncytia Formation

As shown in Schematic 1, the four-defined mouse cellular reprogramming factors Oct4, Sox2, Klf4 and c-Myc were separated by three different types of self-cleaving 2A peptide. This report describes the production and assessment of a polycistronic lentiviral vector carrying OSKM linked with 2A peptides. To produce the virus, 293FT producer cells were co-transfected with a lentiviral vector plasmid (pLV-OSKM) and the packaging/structural plasmids (LENTI-Smart™ components). The culture-cell medium was devoid of serum during transfection in order to increase viral titer as suggested in previous studies (Ansorge et al., 2009; Broussau et al., 2008; Kuroda et al., 2011; Kuroda et al., 2009; Reiser, 2000). In addition, it has been reported that the influence of serum in the culture media can be cytotoxic as the serum interferes with the physiological state of the transduced cells (Reiser, 2000). Of note, we observed small clumps in the culture due to spontaneous breakdown of the cell membranes by the transfection reagent (data not shown). The 293FT cell line was derived from the HEK-293 cells after insertion with a plasmid (pCMVSPORT6TAg.neo) encoding the simian virus 40 (SV40) large T antigen driven by an internal cytomegalovirus (CMV) promoter (Naldini et al., 1996). The SV40 T antigen was used to induce transient expression of proteins on the helper plasmid through the incorporation of viral vector replication signals (Hotta et al., 2009).
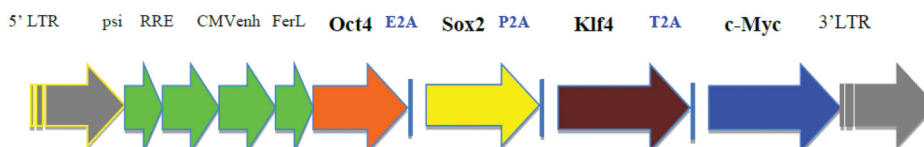
Figure 1. Schematic diagram of vector illustrating the polycistronic lentiviral backbone encoding mouse reprogramming factors Oct4, Sox2, Klf4 and c-Myc linked by 3 different types of 2A peptides, E2A (equine rhinitis A virus), P2A (porcine teschovirs-1) and T2A (*Thosea asigna* virus). Abbreviations: LTR: long terminal repeat, psi: packaging signal, RRE: rev response element, CMVenh: human cytomegalovirus enhancer/promoter, FerL: ferritin light promoter

Syncytia of fused 293FT cells were seen (Figure 2) within 24 h post-transfection due to the expression of the VSV-G envelope (Kutner, Zhang, & Resiser, 2009; Papapetrou & Sadelain, 2011). After the host 293FT genome was transiently transfected, the viral structural proteins and replication enzymes were transcribed, resulting in the production of viral particles containing the vector genome (Malim et al., 1989; Buchschacher & Wong-Staal, 2000). Subsequently, the *psi* (Ψ) packaging signal allowed packaging of viral vector RNAs into virus capsids, which bud from the producer cells together with the VSV-G glycoprotein on the cell surface into the culture supernatant (Gama-Norton et al., 2001; Buchschacher & Wong-Staal, 2000).
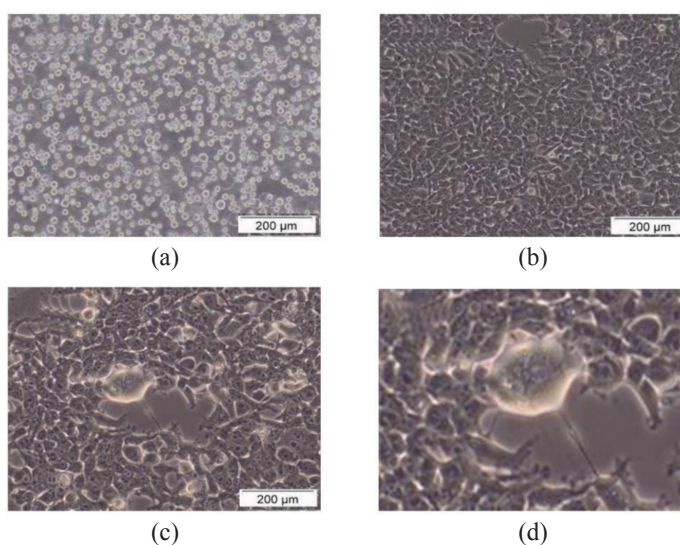
Figure 2. Co-transfection of 293FT cells with pLV-OSKM and LENTI-Smart structural and packaging plusmids. (a) 293FT suspended cell during co-transfection, (b) healthy 293FT cells, (c) fused, and large multinucleated 293FT cells at 24 h post transfection. (d) A zoomed in fused, giant cell from the original photo (c). The scale bar is 200 μm

LV-OSKM Titration

Generally, a high titer of virus was obtained when the lentivirus was collected at day 3 post-transfection (Ngai et al., 2012b). Therefore, the virion was harvested from the supernatant medium at day 3 post-transfection. It was noteworthy that 239FT cells need to be handled carefully due to their weak attachment to the surface of culture plates upon transfection (Dick et al., 2011; Tiscornia, Singer, & Verma, 2006).

Concentrating the lentiviral vector using ultracentrifugation permits higher titer of pseudotyped viral particles and better MOI (Hotta et al., 2009; Ichim et al., 2011). Hence, we concentrated virus particles using Amicon Ultra 100 kDa filter, which is a well-established technique.

To measure the viral titer, HT1080 cells were transduced with the diluted virus and flow cytometry analysis was utilised to determine the proportion of Sox2$^+$ cells at day 3 post-transduction. This indicated the efficiency of transduction when proteins are recognised by the specific antibodies (Dull et al., 1998). The cells transduced with virus solution at 10^{-1} and 10^{-5} showed a significantly high percentage of Sox2$^+$ expression (73.6% and 37.4%, respectively) compared with non-transduced cells and the 'rabbit IgG' isotype (Figure 3). Meanwhile, the cells transduced with virus diluted at 10^{-6} showed 27.5% of Sox2 expression. To avoid a multiple-copy number of integration, the infectivity of the diluted virus has to be between 5 and 30% (Papapetrou et al., 2011). Therefore, the virus diluted at 10^{-6} was chosen in this study for the lentiviral titer calculation. The Sox2 expression level of the transduced sample was statistically significant compared with the control (p<0.001).
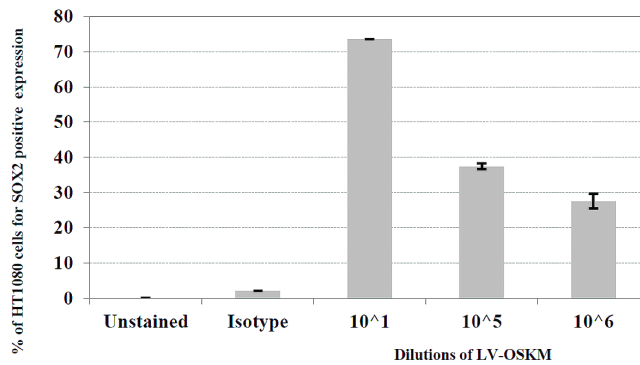
Figure 3. Titration of LV-OSKM. HT1080 cells were transduced with diluted LV-OSKM and harvested at 72 hr post-transduction. Sox2 expression was detected using flow cytometry. The cells were stained with rabbit IgG as Isotype control. Samples were prepared in duplicate and results are shown as mean ± SD. Total event of 10,000 was acquired

To complement the results obtained from cells transduced with 10^{-6} diluted virus, the expression of exogenous genes was assessed by RT-PCR. The exogenous genes (Oct4, Sox2 and c-Myc) were clearly expressed in the 10^{-6} diluted virus. The pLV-OSKM plasmid (as positive control) was used to validate the amplification process (Figure 4).
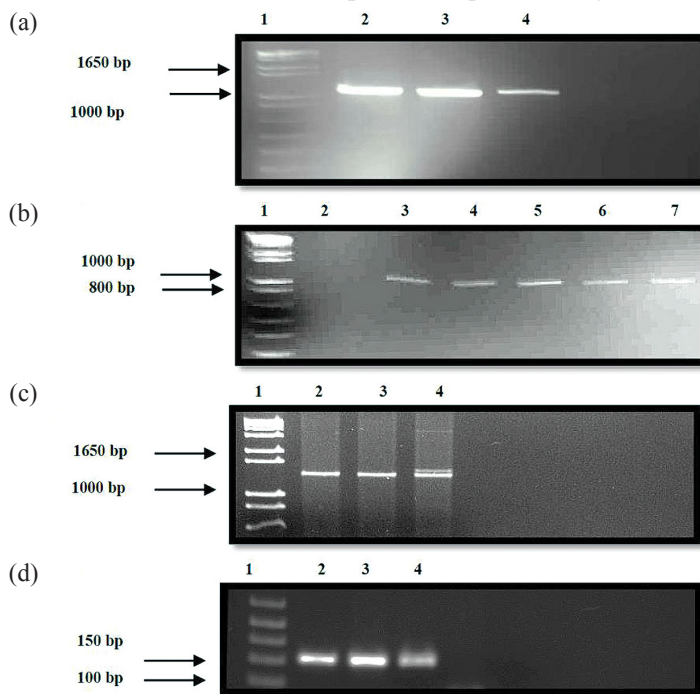
Figure 4. Exogenous gene expression of Oct4, Sox2 and c-Myc from HT1080 cells transduced with 10^{-6} diluted virus analyzed by RT-PCR. Lanes 1 in panel (a), (b) and (c) show a DNA ladder (Tracklt® 1 kb plus). Panel (a) represents Oct4 expression; Lane 2 shows pLV-OSKM used as a positive control to validate amplification of DNA fragments of 1024 bp (Oct4), lanes 3 and 4 show Oct4 expression from 2 independent samples. Panel (b) represents Sox2 expression; lanes 3 and 4 are the pLV-OSKM positive control that gave rise to an amplified DNA fragment of 995 bp (Sox2), lanes 5, 6 and 7 show Sox2 expression from 2 independent samples. Panel (c) represents c-Myc (1317 bp) and pLV-OSKM, which served as the positive control (lane 2), transduced samples (lanes 3-4). None of these genes was expressed in non-transduced HT1080 cells (as a negative control) as shown in Panel (a- lane 5), (b- lane 2), and (c- lane 5). Panel (d) lane1: 50 bp DNA ladder from GeneDriex®, while lanes 2, 3, and 4 indicate β-actin expression (150 bp)

To further authenticate the results, we detected the cells expressing Klf4 by immunocytochemistry. A high expression level of Klf4 was observed in cell nuclei transduced with 10^{-6} diluted virus (Figure 5). Klf4 is considered a major player in regulating cell growth (Garrett-Sinha et al., 1996; Katz et al., 2005). It was noteworthy that our initial attempt was to use the average value of Oct4 expression from HT1080 transduced cells to measure the viral titer. Surprisingly, there was a high percentage of control cells that had Oct4 expression. It is known that Oct4 has two isoforms: Oct4-A and Oct4-B (Araki et al., 2009; Smith et al., 2010). This could be the confounding factor for the unexpected results. To overcome the false positive expression, a mouse IgG_{2b}-specific antibody was chosen for Oct4-A specific detection.
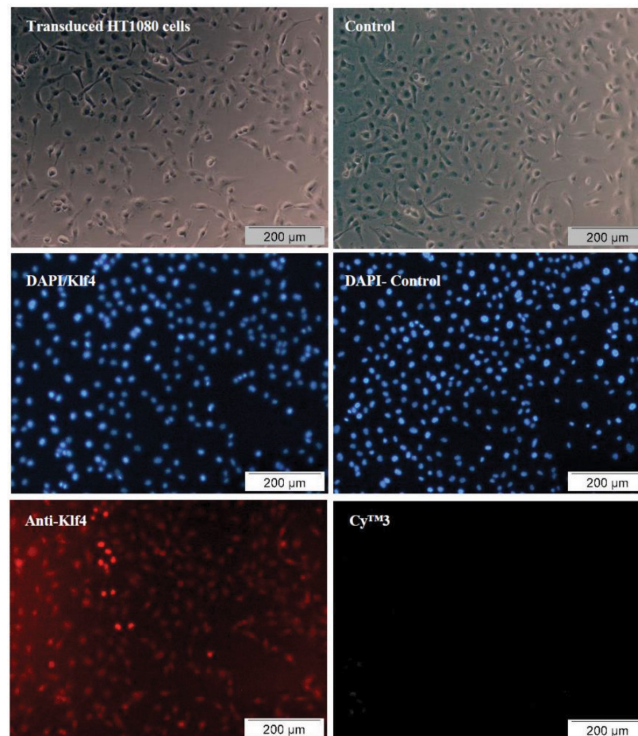
Figure 5. Expression of Klf4 in HT1080 cells transduced with 10^{-6} diluted virus. Phase-contrast and immunocytochemistry images show DAPI-stained nuclei. High expression of Klf4 in HT1080 cell nuclei following transduction with virus diluted at 10^{-6} at day 3 post-transduction. Cells were treated with Klf4 antibody at 1:200 dilution followed by a secondary Cy™3 conjugated antibody (red). Cy™3 was used as a negative control. The scale bar is 200 μm

Mouse Fibroblasts Transduction

We next sought to optimise the MOI for LV-OSKM transduced mouse fibroblasts. The cells were transduced with different MOIs (5, 10 and 20) of LV-OSKM in a serum-free medium to increase the conversion efficiency (Okada, Oka, & Yoneda, 2010). The desired MOI was

calculated using the lowest concentration of the LV-OSKM that considerably yielded the highest percentage of transduced target cells based on the expression of Oct4, which serves as a faithful indicator for the acquisition of pluripotency.

At 48 h post-transduction, the cells seemed smaller and appeared to proliferate faster compared with the non-transduced cells (data not shown). These dynamic morphological changes could be due to the induction of cell proliferation genes, histone modifications and silencing of the somatic genes expression following the expression of the transduced exogenous genes (Araki et al., 2009; Smith et al., 2010).

Expression of Oct4 protein was examined using a flow cytometer at 72 h post-transduction. Histogram results of untransduced mouse cells showed auto-fluorescence (2.5%), while the cells transduced with different MOIs of virus showed apparent Oct4 expression of 90.7% (MOI 5), 97.5% (MOI 10) and 98% (MOI 20) (Figure 6). Immunochemistry analysis was also performed and it revealed positive Oct4 markers in the nuclei of cells transduced with virus at MOI 5, although some cells showed faint cytoplasmic Oct4 protein immunoreactivity (Figure 7). In this study, the results showed that the expression level of Oct4 in cells transduced with virus at MOI 5 was favourable for reprogramming (Carey et al., 2008).
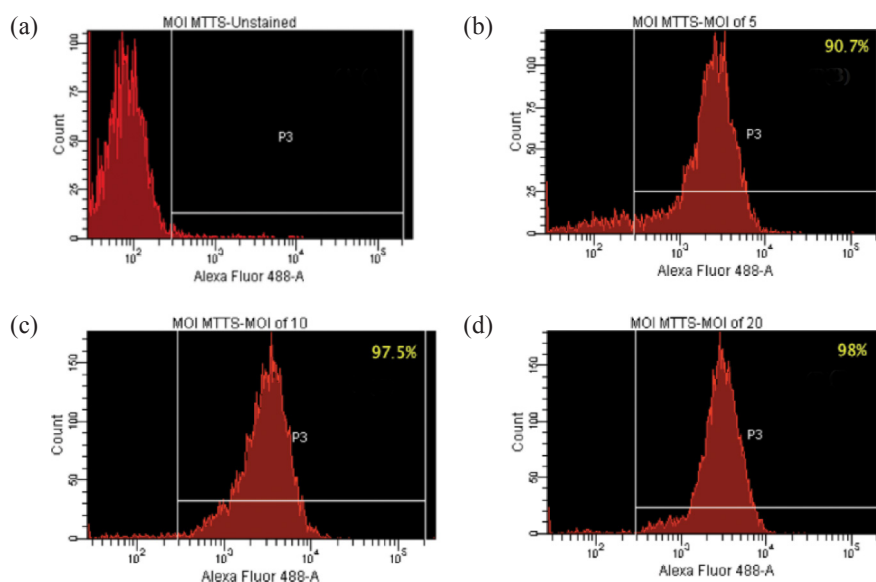
Figure 6. Representative flow cytometry histograms of transduced mouse fibroblasts at day 3 post-transduction with increasing MOI of LV-OSKM. As shown in (a), the peak was gated for negative cells (non-transduced). In contrast, peaks shifted to the right when cells transduced with virus at MOI 5 (b), 10 (c), and 20 (d) were positive for Oct4. Total event of 10,000 was acquired. Data indicated no significant difference, when 5, 10 and 20 MOI of virus was used for transduction. Hence, the lowest MOI of virus producing high expression level was selected. In this study, the most optimal MOI was 5

(a)	(b)	(c)
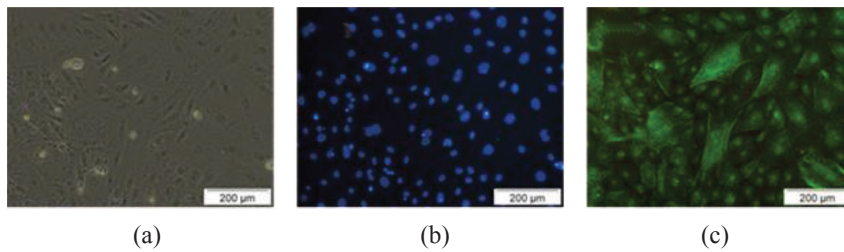
Figure 7. Immunocytochemical analysis of Oct4 expression. (a) Bright field view of mouse fibroblast cells transduced with virus at MOI 5. (b) Cells treated with DAPI staining to visualize nuclei. (c) Immunocytochemical analysis revealed high level of Oct4 in cell nuclei. Cells were incubated with Oct4 antibody (1:1000 dilution) followed by an Alex-Fluor®488 IgG$_{2b}$ conjugated antibody (green). The scale bar is 200 µm

The expression of c-Myc was also detected at MOI 5 (result not shown). It is known as 'double-edged sword' because the low c-Myc expression was sufficient for cellular reprogramming (Sridharan et al., 2009); however, the over-expression of c-Myc could lead to tumorigenicity (Gonzalez et al., 2009; Nakagawa et al., 2007). Therefore, it is crucial to regulate its expression to achieve the targeted effect without inducing tumorigenesis.

CONCLUSION

This report described the production and assessment of polycistronic lentiviral vector encoding OSKM in a single cassette devoid of a reporter gene or drug-based selection system in a serum-free medium. The level of the OSKM expression mediated by the vector could fulfil the requirement for cellular reprogramming.

ACKNOWLEDGEMENT

REFERENCES

Ansorge, S., Lanthier, S., Transfiguracion, J., Durocher, Y., Henry, O., & Kamen, A. (2009). Development of a scalable process for high-yield lentiviral vector production by transient transfection of HEK293 suspension cultures. *The Journal of Gene Medicine*, *11*(10), 868–876.

Araki, R., Jincho, Y., Hoki, Y., Nakamura, M., Tamura, C., Ando, S., & Abe, M. (2009). Conversion of ancestral fibroblasts to induced pluripotent stem cells. *Stem Cells*, *28*(2), 213–220.

Broussau, S., Jabbour, N., Lachapelle, G., Durocher, Y., Tom, R., Transfiguracion, J., & Massie, B. (2008). Inducible packaging cells for large-scale production of lentiviral vectors in serum-free suspension culture. *Molecular Therapy*, *16*(3), 500–507.

Buchschacher, G. L., & Wong-Staal, F. (2000). Development of lentiviral vector for gene therapy for human diseases. *Blood*, *95*(8), 2499–2504.

Carey, B. W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., & Jaenisch, R. (2008). Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proceedings of the National Academy of Sciences, 106*(1), 157–162.

Chang, C., Lai, Y., Pawlik, K. M., Liu, K., Sun, C., Li, C., & Townes, T. M. (2009). Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells. *Stem Cells, 27*(5), 1042–1049.

Dick, E., Matsa, E., Bispham, J., Reza, M., Guglieri, M., Staniforth, A., & Denning, C. (2011). Two new protocols to enhance the production and isolation of human induced pluripotent stem cell lines. *Stem Cell Research*, *6*(2), 158–167.

Dull, T., Zufferey, R., Kelly, M., Mandel, R., Nguyen, M., Trono, D., & Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *Journal of Virology*, *72*(11), 8463–8471.

Gama-Norton, L., Botezatu, L., Herrmann, S., Schweizer, M., Alves, P., Hauser, H., & Wirth, D. (2011). Lentivirus production is influenced by SV40 large t-antigen and chromosomal integration of the vector in HEK293 cells. *Human Gene Therapy*, *22*(10), 1269–1279.

Garrett-Sinha, L. A., Eberspaecher, H., Seldin, M. F., & Crombrugghe, B. D. (1996). A gene for a novel zinc-finger protein expressed in differentiated epithelial cells and transiently in certain mesenchymal cells. *Journal of Biological Chemistry*, *271*(49), 31384–31390.

Gonzalez, F., Monasterio, M. B., Tiscornia, G., Pulido, N. M., Vassena, R., Morera, L. B., & Belmonte, J. C. (2009). Generation of mouse-induced pluripotent stem cells by transient expression of a single nonviral polycistronic vector. *Proceedings of the National Academy of Sciences, 106*(22), 8918–8922.

Hotta, A., Cheung, A. Y., Farra, N., Garcha, K., Chang, W. Y., Pasceri, P., & Ellis, J. (2009). EOS lentiviral vector selection system for human induced pluripotent stem cells. *Nature Protocols*, *4*(2), 1828–1844.

Ichim, C. V., & Wells, R. A. (2011). Generation of high-titer viral preparations by concentration using successive rounds of ultracentrifugation. *Journal of Translational Medicine*, *9*(1), 137.

Jurgens, C. K., Young, K. R., Madden, V. J., Johnson, P. R., & Johnston, R. E. (2011). A novel self-replicating chimeric lentivirus-like particle. *Journal of Virology*, *86*(1), 246–261.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., & Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature*, *458*(7239), 771–775.

Katz, J. P., Perreault, N., Goldstein, B. G., Actman, L., Mcnally, S. R., Silberg, D. G., & Kaestner, K. H. (2005). Loss of Klf4 in mice causes altered proliferation and differentiation and precancerous changes in the adult stomach. *Gastroenterology*, *128*(4), 935–945.

Kay, M. A., Glorisos, J. C., & Naldini, L. (2001). Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutic. *Natural Medicine*, 7(1), 33–40.

Kim, J. H., Lee, S., Li, L., Park, H., Park, J., Lee, K. Y., & Choi, S. (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS ONE*, *6*(4), e18556.

Kuroda, H., Kutner, R. H., Bazan, N. G., & Reiser, J. (2009). Simplified lentivirus vector production in protein-free media using polyethylenimine-mediated transfection. *Journal of Virological Methods*, *157*(2), 113–121.

Kuroda, H., Marino, M. P., Kutner, R. H., & Reiser, J. (2011). Production of lentiviral vectors in protein-free media. *Current Protocols in Cell Biology, 50*(1), 26.8.1—26.8.13.

Kutner, R. H., Zhang, X., & Reiser, J. (2009). Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nature Protocols*, *4*(4), 495–505.

Lai, M. I., Wendy-Yeo, W. Y., Ramasamy, R., Nordin, N., Rosli, R., Veerakumarasivam, A., & Abdullah, S. (2011). Advancements in reprogramming strategies for the generation of induced pluripotent stem cells. *Journal of Assisted Reproduction and Genetics*, *28*(4), 291–301.

Malim, M. H., Hauber, J., Le S.-Y., Maizel, J. V., & Cullen, B. R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Letters to Nature*, *338*(6212), 254–257.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., & Yamanaka, S. (2007). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnology, 26*(1), 101–106.

Naldini, L., Blomer, U., Gage, F., Trono, D., & Verma, I. (1996). Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. *Proceedings of the National Academy of Sciences*, *93*(21), 11382–11388.

Ngai, S. C., Rosli, R., Nordin, N., Veerakumarasivam, A., & Abdullh, S. (2012a). Lentivirus vector driven by polybiquitin C promoter without woodchuck posttranscriptional regulatory element and central polypurine tract generates low level and short-lived reporter gene expression. *Gene, 498*(2)*,* 231–236.

Ngai, S. C., Ramasamy, R., & Abdullah, S. (2012b). Production of lentivirus carrying green fluorescent protein with different promoters for in vitro gene transfer. *Pertanika Journal of Science and Technology*, *20*(2), 269–281.

Nordin, N., Lai, M. I., Veerakumarasivam, A., Ramasamy, R., Abdullah, S., Wendy-Yeo, W. Y., & Rosli, R. (2011). Induced pluripotent stem cell: History, properties and potential applications. *Medical Journal Malaysia*, *66*(1), 35–40.

Okada, M., Oka, M., & Yoneda, Y. (2010). Effective culture conditions for the induction of pluripotent stem cells. *Biochimica et Biophysica Acta (BBA) – General Subjects*, *1800*(9), 956–963.

Papapetrou, E. P., & Sadelain, M. (2011). Derivation of genetically modified human pluripotent stem cells with integrated transgenes at unique mapped genomic sites. *Nature Protocols*, *6*(9), 1274–1289.

Reiser, J. (2000). Production and concentration of pseudotyped HIV-1-based gene transfer vectors. *Gene Therapy*, *7*(11), 910–913.

Smith, Z. D., Nachman, I., Regev, A., & Meissner, A. (2010). Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nature Biotechnology*, *28*(5), 521–526.

Sommer, C. A., Stadtfeld, M., Murphy, G. J., Hochedlinger, K., Kotton, D. N., & Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells*, *27*(3), 543–549.

Sridharan, R., Tchieu, J., Mason, M. J., Yachechko, R., Kuoy, E., Horvath, S., & Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. *Cell*, *136*(2), 364–377.

Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F., & Vignali, D. A. (2004). Corrigendum: Correction of multi-gene deficiency *in vivo* using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature Biotechnology, 22*(6), 760–760.

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676.

Tiscornia, G., Singer, O., & Verma, I. M. (2006). Production and purification of lentiviral vectors. *Nature Protocols*, *1*(1), 241–245.

Zufferey R., & Dull, T. (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *Journal of Virology, 72*(12), 9873–9880.

Increased Endothelial Progenitor Cells with Age and Grade of Malignancy in Astrocytic Glioma Patients

Priscilla Das[1,2]*, Nyi Nyi Naing[3], Nadiah Wan-Arfah[3], KON Noorjan[4], Yee Cheng Kueh[1] and Kantha Rasalingam[5]

[1]Unit of Biostatistics and Research Methodology, Department of Psychiatry, School of Medical Sciences, Universiti Sains Malaysia, 16150 USM, Kubang Kerian, Kelantan, Malaysia

[2]Faculty of Health Sciences, Asia Metropolitan University, G-8, Jalan Kemacahaya 11, Taman Kemacahaya, Batu 9, 43200 Cheras, Selangor Darul Ehsan, Malaysia

[3]Institute for Community (Health) Development (i-CODE), Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia

[4]Departmentof Psychiatry, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[5]Department of Neuroscience, Hospital Kuala Lumpur, 50586 Jalan Pahang, Kuala Lumpur, Malaysia

ABSTRACT

Astrocytic gliomas are the most common primary brain tumours that originated from human glial cells. The tumours rely upon endothelial progenitor cells (EPCs) for neoangiogenesis. This study aimed to investigate the association between tissue resident EPCs in a brain tumour and normal adjacent tissue in relation to age and grade of astrocytic glioma. Astrocytic glioma patients (n=22), grade I to grade IV were consented from Hospital Universiti Sains Malaysia. Brain tumour tissue and normal adjacent brain tissue samples were obtained from each patient during surgery. The EPCs were stained with CD133[+] and VEGFR-2[+] markers. The tissue residents EPCs for each sample were determined using the immunofluorescence microscopy method. The age of the patients increased by disease severity in the following order (Grade I: 21.33±20.79 years) < (Grade II: 46.50±0.707 years) < (Grade III: 47.38±11.95 years) < (Grade IV: 48.44 ±10.66 years). The EPCs in brain tumour correlated significantly with the age of the patients with positive correlation (Spearman's rho correlation test, r=0.52; p=0.013). The tissue resident EPCs in the brain tumour (median=0.40, IqR=0.59) were significantly higher compared with the adjacent normal brain (median=0.067, IqR=0.29) (Wilcoxon Signed-Rank Test, Z stat=-3.587, p<0.001). Higher tissue resident EPCs were found in high grade (III & IV) glioma compared with EPCs in low grade (I & II) glioma (median=0.61, IqR=0.70 vs. median=0.26, IqR=0.30; z=-1.763 p=0.078). This study showed increased EPCs with age and

E-mail addresses:
daspriscilla@yahoo.com (Priscilla Das)
syedhatimnoor@gmail.com (Nyi Nyi Naing)
wanwaj_85@gmail.com (Nadiah Wan-Arfah)
tnnjkk99@gmail.com (KON Noorjan)
yckueh@gmail.com (Yee Cheng Kueh)
kr_dr@hotmail.com (Kantha Rasalingam)
*Corresponding Author

grade of malignancy in astrocytic glioma patients. Therefore, targeting EPCs in gliomas based on tumour grade malignancy and age of the patients might be useful in effective treatment of astrocytic glioma.

Keywords: Astrocytomas, brain tumours, endothelial progenitor cells, glioma

INTRODUCTION

According to WHO classification of central nervous system tumours, astrocytomas are typically classified as pilocytic (Grade I – less aggressive), diffuse (Grade II), anaplastic (Grade III) or glioblastoma multiforme (Grade IV – most aggressive) in order of increasing anaplasia (Okada et al., 2009). Glioblastoma multiforme is the most common type of astrocytoma and it is a highly invasive and almost uniformly fatal tumour. Glioblastoma multiforme is among the most highly vasculariaed of all malignancies and relies upon angiogenesis for growth and histological progression (Vredenburgh et al., 2009). The Grade IV gliomas are cytologically malignant, mitotically active and necrosis-prone neoplasms and they are related typically with formation of rapid pre- and post-operative disease. A propensity for craniospinal dissemination and infiltration of surrounding tissue characterises Grade IV tumours. Such tumours also have endothelial proliferation with apparent multi-layering of endothelium and glomeruloid microvascular proliferation (Louis et al., 2007).

Grade III tumours (anaplastic astrocytoma) are commonly found with lesions and histological malignancy, including nuclear atypia and brisk mitotic activity. The treatment options for Grade III glioma are either by radiation or chemotherapy or a combination of both. Diffusely infiltrative astrocytic tumours with cytological atypia is a characteristic of Grade II tumours (diffuse astrocytoma). The Grade II *neoplasms* are generally *infiltrative and* always recur despite low-level proliferation. They also have the tendency to develop into higher grades i.e. from diffuse astrocytoma (low grade) that proliferates into anaplastic astrocytoma and glioblastoma. The Grade I gliomas are assigned into more circumscribed pilocytic astrocytoma. They carry properties of low proliferative potential of lesions and chances of curability by surgical resection (Louis et al., 2007).

The endothelial progenitor cell (EPC) acts as an angiogenic precursor. *EPCs* probably have a great *potential* in pathophysiology and treatment of brain cancer (Gao et al., 2010). The number of EPCs that are found in patients with glioma varies by grade of the malignancy and age of the patients. A *previous study reported* that patient with glioma Grade IV and Grade III who have undergone treatment (radiotherapy or chemotherapy) showed significant reduction of circulating EPCs compared with healthy controls (Corsini et al., 2012). In addition, the age of the patients also affects tumour pathophysiology in astrocytic gliomas (Rebetz et al., 2008). Therefore, targeting EPCs in gliomas by considering the grade and age of the patients might be beneficial in effective antiangiogenic treatment and will serve as potential new targets or vectors for adjuvant therapy (Yu et al., 2010).

Significant reduction of EPCs showed impaired and delayed growth in tumours followed by reduction in tumour vessel density. The study showed that the inhibition of EPC might block the neoangiogenesis process particularly in the growing tumours after incomplete surgical removal of primary lesion or treatment (Nolan et al., 2007). During vascular injury, circulating host-derived endothelial cells are highly demanded for neoangiogenesis (Briasoulis et al., 2011; Leone et al., 2009; Yin et al., 2010). Moreover, with advancing age, it is likely that oxidative cellular damage accumulates in EPCs and deteriorates its function (He et al., 2009). Therefore, it is reasonable to *believe that* the number of EPCs and age determine the severity of astrocytic glioma patients. In this study, the associations between tissue resident EPCs and age and grade of glioma were investigated.

METHOD

Patient Recruitment

Astrocytic glioma patients (n=22) were consented from the Hospital of Universiti Sains Malaysia (HUSM). Only cases that confirmed with WHO grading (I to IV) of astrocytic glioma were included in the study. Ethical approval to conduct this study was obtained from the Human Research Ethics Committee, Universiti Sains Malaysia (FWA Reg No: 00007718; IRB Reg. No: 00004494).

Tissue Analysis for EPCs

Microsurgical specimens of brain tumour and adjacent normal brain tissue were obtained from each patient for the analyses of tissue resident EPCs. Tissues were analysed using immunofluorescence staining. The total number of tissue resident EPCs in the tumour and adjacent normal brain tissue specimens were characterised using immunofluorescence microscopy.

In the IHC technique, tissue biopsies were fixed in 10% paraformaldehyde and processed for histology and immunohistochemistry analyses. The antibodies that were used in this study included CD133 (clone ACC133/1) and vascular endothelial growth factor-2 (VEGFR-2) to characterise the tissue resident EPCs (Hilbe et al., 2004; Rafat N., 2010; Toshner M., 2009).

Slide Preparation

Tissue biopsy was fixed in paraformaldehyde 10% and three sections from each brain tissue sample was cut at an interval length of about 6.5 mm with a thickness of 4 mm for brain tumour tissues, while normal brain tissues were adjusted. The tissues were processed consecutively in an automated tissue processor as follows:

a.   80% ethanol for 1 h
b.   95% ethanol for 1 h
c.   95% ethanol for 1 h
d.   Absolute ethanol for 1 h
e.   Absolute ethanol for 1 h
f.   Absolute ethanol for 1 h

g. Xyelene for 1 h
h. Xyelene for 1 h
i. Xyelene for 1 h
j. Warmed paraffin for 2 h
k. Warmed paraffin for 2 h

The paraffin embedded tissue blocks were trimmed and sectioned with a microtome to obtain a thickness of 3 μm of the tissue section for immunofluorescence staining. The ribbons of sectioned tissue were floated in a 38^0C water bath and 'fished' onto a slide. The slides were put on a 60^0C hot plate or slide warmer.

Procedure for Histological Assessment

The tissue was deparaffinised with two changes of xylene, followed by xylene 1:1 with ethanol and rehydration with two changes of absolute, 95% and 70% ethanol. The tissue was then rinsed under running cold water from a tap. The flow of staining is shown below:
a. Xylene for 3 min
b. Xylene for 3 min
c. Xylene 1:1 with ethanol for 3 min
d. Absolute ethanol for 3 min
e. Absolute ethanol for 3 min
f. 95% ethanol for 3 min
g. 70% ethanol for 3 min
h. Rinse with tap water for 5 min

Immunofluorescence Staining

Enough drops of 0.1% Triton X-100/PBS were added to the tissue slide for 10 min and the tissue was washed three times with 1X PBS. A unit of 0.5% PBS/BSA was added to the tissue slide for 5 min. The tissue sections were then stained with PE-conjugated anti-human CD133 (1:15 dilution) and FITC-conjugated anti-human VEGFR-2 (1:200 dilution) to identify the tissue resident EPCs. The tissue sections were incubated overnight at 4 °C in a dark, humid incubation chamber. After the incubation, the tissue sections were washed three times with 1x PBS and counterstaining was performed with 4',6-diamidino-2-phenylindole (DAPI) for 30 min. Finally, the slides were washed three times with 1X PBS and mounted in Prolong antifade mounting reagent from Life Technologies and assessed using BX41 Olympus microscopy at 200X magnification. Excitation in the ultraviolet (330-385 nm), blue (460-490 nm) and green (510-550 nm) is used in this BX41 Olympus microscope. The ultraviolet filter was used to identify the cell nuclei stained *blue* with *DAPI,* the blue filter used to detect the FITC-VEGFR2+ marker that is reflected in green and the green filter for the PE-CD133 marker in red. Both the reflected light and sample fluorescence are viewed through and recorded at the microscope. A Nikon Coolpix 5.1 megapixel camera as well as a USB camera with image capture software were used to record the images. The images were captured and merged to review the expression of the markers.

Immunocytochemical Scoring

About 24 field images were captured of the brain-tumour sample and another 10 to 24 images of the adjacent normal brain per patient. Therefore, the total area assessed of both the tumour and adjacent normal brain for 22 patients was about 606.50 mm² for tissue resident EPCs. The images were captured at 200X magnification. The counts were expressed as the average of all fields examined. The percentage of EPCs in the tissues was analysed using Image J software version 1.45 s.

Statistical Analysis

The Mann-Whitney test was applied for comparing the independent samples and the Wilcoxon Signed-Rank test was applied for comparing related samples. The Spearmann-ranked correlation coefficient was applied to determine correlation. Statistical significance was determined at $p < 0.05$ and SPSS software version 22.0 was used in the study.

RESULTS

A total of 22 astrocytic glioma patients were included in this study. The clinical characteristics of the patients are displayed in Table 1. Patients with various types of glioma were enrolled in the study; the types included pilocytic astrocytoma, diffuse fibrillary astrocytoma, diffuse astrocytoma, anaplastic astrocytoma, anaplastic oligodendroglioma, anaplastic ependymoma, anaplastic gemistocytic astrocytoma, glioblastoma with oligodendroglioma component, gliosarcoma and glioblastoma multiformae.

Table 1
Clinical characteristics of respondents

Characteristics	n (%)
Age	
≤30 years	2 (9.1)
31-40 years	6 (27.3)
41-50 years	7 (31.8)
51-60 years	5 (22.7)
>60 years	2 (9.1)
Gender	
Male	16 (70.0)
Female	6 (30.0)
Astrocytic glioma diagnosis	
Glioblastoma multiformae WHO grade IV	9 (40.9)
Anaplastic WHO grade III	8 (36.4)
Diffuse WHO grade II	2 (09.1)
Pilocytic WHO grade I	3 (13.6)

The age of the patients increased by disease severity in the following order: Grade I: 21.33±20.79 years < Grade II: 46.50±0.707 years < Grade III: 47.38±11.95 years < Grade IV: 48.44 ±10.66 years). The EPCs in the brain tumour correlated significantly with the age of the patients with positive association (Spearman's rho correlation test, r=0.52; p=0.013; Figure 1). No correlation was found in adjacent normal brain tissue with the age of the patients (Spearman's rho correlation test, r=- 0.051; p=0.820).
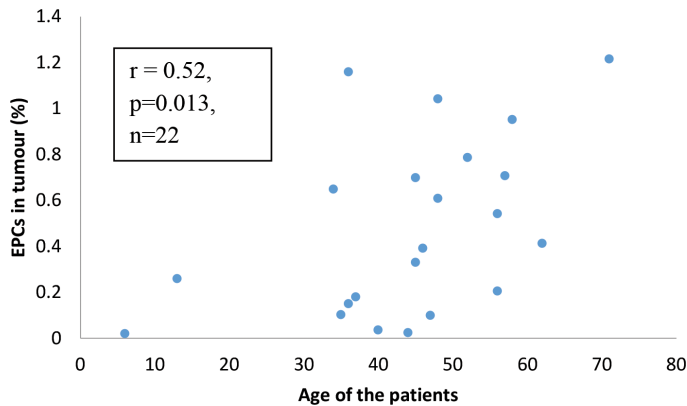
Figure 1. Correlation between tissue resident EPCs in brain tumour and age of the patients (Spearman's rho correlation test, r=0.52; p=0.013, n=22)

The tissue resident EPCs were significantly higher in the brain tumour (median=0.40, IqR=0.59) compared to the adjacent normal brain (median=0.067, IqR=0.29) (Wilcoxon Signed-Rank Test, Z stat=-3.587, p<0.001); Table II. The higher tissue resident EPCs found in Grade III and IV gliomas were compared with the low-grade (Grade I & Grade II) gliomas; however, the p-value was not significant (median=0.61, IqR=0.70 vs. median=0.26, IqR=0.30; z=-1.763 p=0.078; see Table 3). The detection of tissue resident EPCs CD133[+]/VEGFR2[+] in the brain tumours by grade of astrocytic glioma is shown in Figure 2.

Table 2

Comparison of median epcs in the brain tumour and normal adjacent brain of patients

	Median (IqR)		Z stat[a]	p-value
	Brain Tumour	Normal Adjacent Brain		
Median EPCs	0.40 (0.59)	0.067 (0.29)	-3.587	p<0.001

[a]Wilcoxon Signed-Rank Test

Table 3
Comparison of median epcs in high-grade and low-grade glioma patients

	Median (IqR)		Z stat[a]	p-value
	High Grade	Low Grade		
EPCs	0.61 (0.70)	0.26 (0.30)	-1.763	0.078

[a]Mann-Whitney Test

Tissue	DAPI	CD133+	VEGFR2+	MERGED
Grade IV				
Grade III				
Grade II				
Grade I				
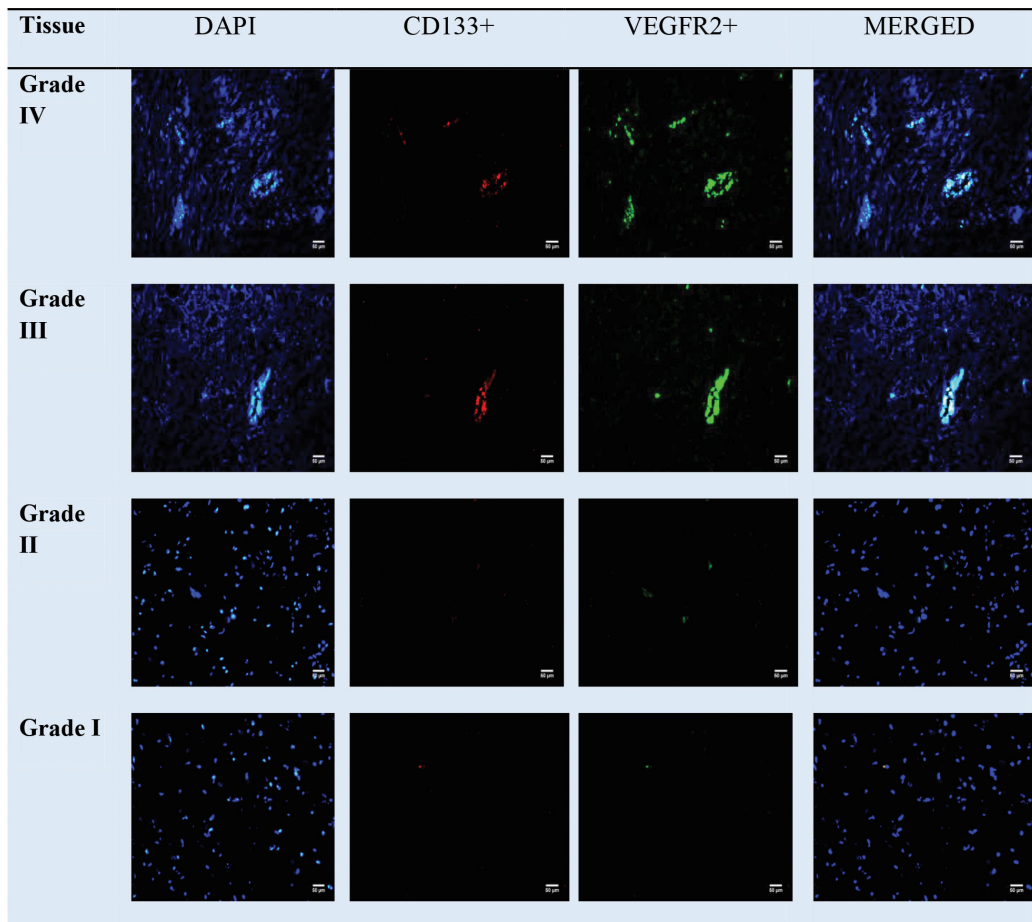
Figure 2. Detection of endothelial progenitor cells (EPCs) in tissue glioma of the patient

DISCUSSION

The World Health Organization classified astrocytic tumours in four stages. Stage I includes: Subependymal giant cell astrocytoma and Pilocytic astrocytoma; Stage II: Pilomyxoid astrocytoma, Diffuse astrocytoma and Pleomorphic xanthoastrocytoma; Stage III: Anaplastic

astrocytoma; and Stage IV: Glioblastoma, Giant cell gliobastoma and Gliosarcoma (Louis et al., 2007). Gliomas are classified as high (III and IV) and low (I and II) grade. The high-grade gliomas have more characteristics of anaplastics. They also contain immature astrocytes, oligodendrocytes or both. The low-grade gliomas are characterised by well differentiated astrocytes or oligodendrocytes lineage and are commonly diagnosed in younger patients. They have the ability to progress into anaplastic gliomas or secondary glioblastomas. However, the high-grade gliomas are mostly detected *de novo* among older patients, apart from clinically identified history (Rebetz et al., 2008). In another study it was also proven that patients with advanced age are more prone to be diagnosed with brain tumour (Zainal et al., 2006). A similar result was found in the current study, where it was found that Grade IV gliomas were diagnosed among the older patients and the less severe Grade I gliomas were identified among the younger patients.

EPC were characterised using two markers, CD133 and VEGFR-2 (KDR), that were found to be localised in structure of the capillary of the solid tumour and they promoted vasculogenesis. Microvascular density was assessed using the CD31 marker and it was found to be correlated significantly with the tumour grade as it was found to be the highest in Grade III tumour patients. CD31 is expressed on activated and non-activated endothelial cells (Hilbe et al., 2004). Previous research findings reported that the endothelial cells present most in anaplastic astrocytomas compared to low-grade astrocytomas and glioblastomas (Strik et al., 2001). Moreover, comparing the glioblastoma and low grade astrocytoma, it was found that the glioblastoma multiformae had more expression of thymidine phosphorylase, which is responsible for secreting the vascular endothelial growth factor for angiogenesis in the tumour (Yao et al., 2001), and this might explain why the higher-grade gliomas were found to have more tissue resident EPCs compared with the lower-grade gliomas in the current research findings.

The morphology of diffuse astrocytomas showed well differentiated neoplastic astrocytes in a microcystic tumour matrix, having moderate cellularity and nuclear atypia but without mitotic activity. The anaplastic astrocytomas showed diffuse increase in cellularity and nuclear atypia with presence of mitotic activity. The glioblastomas, on the other hand, were poorly differentiated, highly anaplastic and pleomorphic tumours with increased nuclear atypia and high mitotic activity. There were also prominent vascular proliferations and necrosis as well as parenchymal invasion in the anaplastic astrocytomas. The proliferation rate of the tumour determined using the Ki-67 marker was noted to be significantly increased when the low-grade (Grade II) and high-grade tumours (Grade III and Grade IV) were compared; however, no significant increment was seen between Grade III and Grade IV tumours (Stanca et al., 2012). Significant increment in the Ki67 expression with the increase in the age of the patients was also found in another study and an association between nestin and the Ki67 proliferation marker with grade of the tumours were found (Osama et al., 2010). These findings support the current research, which has reported increased EPCs in gliomas associated with the increase in the grade of malignancy and the age of the patients.

The population of endothelial cells, astrocytes, neural stem/progenitor cells, mesenchymal stem cells, pericytes and microglia/macrophage were abundantly present in the glioblastomas compared with in the normal brain parenchyma (Golebiewska et al., 2013). Malignant gliomas are known to infiltrate into adjacent normal healthy brain tissue, and this finding has been

seen in animal models using BT4C malignant gliomas. BT4C gliomas are known to be highly angiogenic as the tumour vessels metastasise to distant parts of the brain after anti-angiogenic treatment (Wirth, 2012). This finding supported that of the current study, which found that tissue resident EPCs were significantly higher in the brain tumour compared with in the adjacent normal brain.

Patients with Grade IV and Grade III gliomas who had undergone treatment (radiotherapy or chemotherapy) showed a significant reduction in circulating EPCs compared with in the healthy control. The number of EPCs in patients not undergoing treatment for Grade IV and Grade III gliomas had higher EPCs compared with the healthy control. Patients before chemotherapy and after surgery showed increment in the number of EPCs compared with the healthy patients. Chemotherapy decreases VEGF levels significantly in patients (Corsini et al., 2012). Therefore, chemotherapy based on tumour grade and age of the patient is important, and might serve as a potential indicator for improved and effective antiangiogenic treatment for gliomas.

LIMITATION

Determining the EPCs especially in the astrocytic glioma patients was a challenging task since EPCs are rare cells. Therefore, this factor should be considered when interpreting the current results.

CONCLUSION

We observed in this study that EPCs increased with age and grade of malignancy in astrocytic glioma patients. Therefore, our findings suggest that the antiangiogenic drugs given based on age and grade of tumour might be useful for preventing tumour growth in glioma patients.

ACKNOWLEDGEMENT

REFERENCES

Briasoulis, A., Tousoulis, D., Antoniades, C., Papageorgiou, N., & Stefanadis, C. (2011). The role of endothelial progenitor cells in vascular repair after arterial injury and atherosclerotic plaque development. *Cardiovascular Therapeutics*, *29*(2), 125–139. doi: 10.1111/j.1755-5922.2009.00131.x

Corsini, E., Ciusani, E., Gaviani, P., Silvani, A., Canazza, A., Bernardi, G., … & Salmaggi, A. (2012). Decrease in circulating endothelial progenitor cells in treated glioma patients. *Journal of Neurooncol, 108*(1), 123–129.

Fava, G., Ruini, C., & Rafanelli, C. (2004). Psychometric theory is an obstacle to the progress of clinical research. *Psychother Psychosom, 73*(3)*,* 145–148.

Gao, P., Chen, Y., Lawton, M. T., Barbaro, N. M., Yang, G. Y., Su, H., … & Young, W. L. (2010). Evidence of endothelial progenitor cells in the human brain and spinal cord arteriovenous malformations. *Neurosurgery, 67*(4), 1029-1035.

He, T., Joyner, M. J., & Katusic, Z. S. (2009). Aging decreases expression and activity of glutathione peroxidase-1 in human endothelial progenitor cells. *Microvascular research, 78*(3), 447-452.

Hilbe, W., Dirnhofer, S., Oberwasserlechner, F., Schmid, T., Gunsilius, E., Hilbe, G., … & Kahler, C. M. (2004). CD133 positive endothelial progenitor cells contribute to the tumour vasculature in non-small cell lung cancer. *Journal of clinical pathology, 57*(9), 965–969.

Leone, A. M., Valgimigli, M., Giannico, M. B., Zaccone, V., Perfetti, M., D'amario, D., ... & Crea, F. (2009). From bone marrow to the arterial wall: the ongoing tale of endothelial progenitor cells. *European Heart Journal, 30*(8), 890-899.

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., … & Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathologica, 114*(2), 97–109.

Nolan, D. J., Ciarrocchi, A., Mellick, A. S., Jaggi, J. S., Bambino, K., Gupta, S., … & Benezra, R. (2007). Bone marrow-derived endothelial progenitor cells are a major determinant of nascent tumor neovascularization. *Genes and Development, 21*(12), 1546–1558.

Okada, H., Kohanbash, G., Zhu, X., Kastenhuber, E. R., Hoji, A., Ueda, R., & Fujita, M. (2009). Immunotherapeutic approaches for glioma. *Critical Reviews in Immunology, 29*(1), 1–42.

Osama, S. A., Bassma, M. E., & Amro, A. (2010). Characterization of cancer stem cells in patients with brain astrocytomas: A clinico-pathological and immunohistochemical study. *Alexandria Journal of Medicine, 46*(4), 357-363.

Rafat, N., Beck, G., Schulte, J., Tuettenberg, J., & Vajkoczy, P. (2010). Circulating endothelial progenitor cells in malignant gliomas. *Journal of neurosurgery, 112*(1)*,* 43–49.

Rebetz, J., Tian, D., Persson, A., Widegren, B., Salford, L. G., Englund, E., … & Fan, X. (2008). Glial progenitor-like phenotype in low-grade glioma and enhanced CD133-expression and neuronal lineage differentiation potential in high-grade glioma. *PLoS ONE, 3*(4), e1936.

Stanca, D., Craltoiu, S., Tudorica, V., Albu, C., Alexandru, O., Pirscooveanu, D., … & Zaharia, C. (2012). An immunohistological study of the presence of inflammatory cells in malignant brain tumors. *Current Health Sciences Journal, 38*(3), 113-117.

Strik, H., Deininger, M., Frank, B., Schluesener, H., & Meyermann, R. (2001). Galectin-3: Cellular distribution and correlation with WHO-grade in human gliomas. *Journal of Neuro-Oncology, 53*(1), 13–20.

Toshner, M., V. R., & Southwood, M. (2009). Evidence of dysfunction of endothelial progenitors in pulmonary arterial hypertension. *American journal of respiratory and critical care medicine, 180*(8), 780–787.

Vredenburgh, J. J., & Cloughesy, T. F. (2009). *Update on antiangiogenic therapy for advanced malignant glioma* (pp. 1-5). Cambridge, MA: The Angiogenesis Foundation.

Yao, Y., Kubota, T., Sato, K., & Kitai, R. (2001). Macrophage infiltration-associated thymidine phosphorylase expression correlates with increased microvessel density and poor prognosis in astrocytic tumors. *Clinical Cancer Research, 7*(12), 4021–4026.

Yin, Y., Zhao, X., Fang, Y., Yu, S., Zhao, J., Song, M., & Huang, L. (2010). SDF-1α involved in mobilization and recruitment of endothelial progenitor cells after arterial injury in mice. *Cardiovascular Pathology, 19*(4), 218–227. doi: http://dx.doi.org/10.1016/j.carpath.2009.04.002

Yu, D. C., Chen, J., Sun, X. T., Zhuang, L. Y., Jiang, C. P., & Ding, Y. T. (2010). Mechanism of endothelial progenitor cell recruitment into neo-vessels in adjacent non-tumor tissues in hepatocellular carcinoma. *BMC Cancer, 10*(1), 435-444.

Zainal, A. O., Zainudin, M. A., & Nor Saleha, I. T. (2006). *Malaysian cancer statistics data and figure peninsular Malaysia. 2006: 8*. Retrieved from http://www.moh.gov.my/v/c_report?mode=public., 8.

Synthesis, Characterisation and Biological Activities of Ru(III), Mo(V), Cd(II), Zn(II) and Cu(II) Complexes Containing a Novel Nitrogen-Sulphur Macrocyclic Schiff Base Derived from Glyoxal

Chah, C. K.[1], Ravoof, T. B. S. A.[1]* and Veerakumarasivam, A.[2,3]

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

[2]Department of Obstetrics and Gynaecology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[3]Medical Genetics Laboratory, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

A novel nitrogen-sulphur macrocyclic Schiff base, 4,11,20,27-tetrathioxo-3,12,19,28-tetrathia-5,6,9,10,21,22,25,26-octaazatricyclo[28.2.2.2^{14,17}]hexatriaconta 1(33),6,8,14(36),15,17(35),22,24,30(34),31-decaene-2,13,18,29-tetraone (TGSB) derived from terephthaloyl-bis-dithiocarbazate (TDTC) and glyoxal (ethane-1,2-dione) is synthesised via condensation. Metal complexes are formed by reacting the Schiff base with various metal salts such as Ru(III), Mo(V), Cd(II), Zn(II) and Cu(II). The complexes are expected to have a general formula of M_2L or M_3L with a square planar or square pyramidal geometry. These compounds were characterised by various physico-chemical and spectroscopic techniques. From the data, it is concluded that the azomethine nitrogen atom and the thiolate sulphur atom from the ligand are bonded to the metal ion. In the IR spectra of the complexes, the presence of the C=N band in the region of 1600 cm^{-1} indicates the successful formation of the Schiff base. The structures of the Schiff base and metal complexes are confirmed via FT-IR, GC-MS and NMR spectroscopic analysis. The magnetic susceptibility measurements, electronic spectral data and molar conductivity analysis support the desired geometry of the complexes. The Schiff base and its metal complexes are evaluated for their biological activities against the invasive human bladder carcinoma cell line (EJ-28) and the minimum-invasive human bladder carcinoma cell line (RT-112). The RuTGSB and CdTGSB complexes showed selective activity against RT-112.

Keywords: Biological activities, bladder cancer, complexes, dithiocarbazate, glyoxal, macrocyclic Schiff base

INTRODUCTION

Ligands are very promising compounds from the view point of coordination due to their ability for complexation (Singh et al., 2009). Hence, they can react with metals to produce metal complexes that often have better biological properties and applications such as anticancer agents and antimicrobial agents (Azarkish et al., 2012; Mewis et al., 2010; Hossain et al., 1996; Chandra et al., 2008; Mohamed et al., 2009). Dithiocarbazate, $NH_2NHCS_2^-$ and its substituted derivatives, especially ligands with nitrogen and sulphur as donor atoms, have been of great interest to researchers over the past few decades. Schiff bases are prepared through condensation of dithiocarbazate derivatives with various aldehydes and ketones.

The term 'macrocycle' is defined as a cyclic macromolecule or a cyclic compound with nine or more members (Constable, 1999). Macrocyclic Schiff bases act as ligands that are formed by reacting dithiocarbazate with various dicarbonyl compounds. Macrocyclic ligands are of interest because of their unique coordination chemistry (Aqra, 1999). The cytotoxicity of the complexes is higher than that of the ligand; this implies an increase in antitumor activity with coordination. The macrocycle ring enables a molecule to achieve a degree of structural pre-organisation, such that key functional groups can interact across extended binding sites in proteins without a major entropic loss on binding (Driggers et al., 2008). Macrocyclic structures appear as promising polydentate ligands, complexones and ionophores for obtaining magneto-contrast compounds, extragents and analytical reagents as well as compounds with potentially high pharmacological and biological activities.

The macrocyclic Schiff base in this work was synthesised by reacting dithiocarbazate with glyoxal via condensation. In coordination chemistry, the functionally substituted Schiff bases bearing additional donor groups represent the most important class of heteropolydentate ligands capable of forming polynuclear complexes with transition metals (Borisova et al., 2007). The synthesis of macrocyclic ligands still remains challenging for coordination chemists due to unexpected complexations (Xie et al., 2008).

Recently, researchers have been interested in the synthesis of new metal-based anticancer drugs with minimal side effects. This is because several serious side effects from the treatment will decrease the efficacy of the commercial drugs used for anticancer treatment. Hence, a novel macrocylic Schiff base and metal complexes are reported in this research to evaluate their structural properties and their potential as anticancer drugs. The geometry of the synthesised compounds is proposed to have a planar structure so that they can easily interact with deoxyribonucleic acid (DNA) in the DNA binding studies.

MATERIALS AND METHOD

All the chemicals and solvents were of analytical reagent grade, purchased from Merck, Sigma Aldrich, BDH or Fluka and used without further purification. The melting points of the ligand and metal complexes were determined by the electrothermal IA9100 digital melting point apparatus. Infrared spectra were obtained with a Perkin-Elmer 100 Series FT-IR spectrophotometer (4000-280 cm^{-1}) using KBr pellets. ^1H and ^{13}C NMR spectra of the ligand were recorded via a JOEL JNM 500 MHz spectrometer using deuterated chloroform as the

solvent and TMS as the internal reference. The ^{1}H spectrum was measured from $\delta_H = 0$ to 14 ppm, while the ^{13}C spectrum was from $\delta_C = 0$ to 250 ppm. The mass spectrum (GC-MS) of the ligand was recorded by a Shimadzu GC-MS QP5050A spectrometer using an electron ioniser with the voltage of 70 eV. Carbon, hydrogen and nitrogen elemental analysis was carried out using the LECO CHNS-932 analyser with a temperature of 1000 K. Metal content in the complexes was obtained by a Perkin Elmer emission plasma 1000, inductively coupled plasma optical emission spectrometer. Three different concentrations were prepared according to the absorption range. Magnetic susceptibilities of the complexes were measured by the Gouy method at room temperature using the Sherwood scientific magnetic susceptibility balance and distilled water as calibrant. The molar conductance of 10^{-3} M solutions of the complexes in dimethyl sulfoxide (DMSO) was measured at room temperature. A Jenway 4310 conductivity meter and a dip-type cell with platinised electrode were used in these measurements. The electronic spectra were obtained using the Shimadzu UV-Vis 160A Spectrophotometer over a range of 200-1000 nm. Complex solutions of molar conductance 10^{-3} M, 10^{-4} M and 10^{-5} M were prepared by dissolving the substances in DMSO and then measuring the solutions using a quartz cuvette.

Synthesis of Terephthaloyl-bis-dithiocarbazate (TDTC)

Hydrazine hydrate (6.3 mL, 0.2 mol) was added dropwise into a solution of KOH (11.2 g, 0.2 mol) in ethanol (70 mL) followed by carbon disulphide (12.1 mL, 0.2 mol) and terephthaloyl dichloride (20.3 g, 0.1 mol) in warm ethanol (200 mL). The mixture was stirred to reduce the volume to half and the white precipitate that was formed was filtered and then recrystallised from absolute ethanol (Figure 1). The yield was 87%, m.p. 128-129°C.
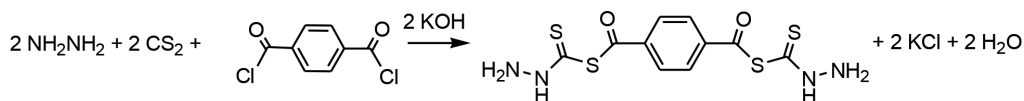
Figure 1. Synthesis of terephthaloyl-bis-dithiocarbazate (TDTC)

Synthesis of the Schiff Base, 4,11,20,27-tetrathioxo-3,12,19,28-tetrathia-5,6,9,10,21,22,25,26-octaazatricyclo[28.2.2.2^{14,17}]hexatriaconta-1(33),6,8,14(36),15,17(35),22,24,30(34),31-decaene-2,13,18,29-tetraone (TGSB)

Glyoxal (0.9 mL, 0.02 mol) was added to a solution of TDTC (6.9 g, 0.02 mol) in warm absolute ethanol (120 mL). The mixture was refluxed for 8 h and the yellow precipitate that formed was recrystallised from absolute ethanol (Figure 2). The yield was 52%, m.p. 137-138°C.
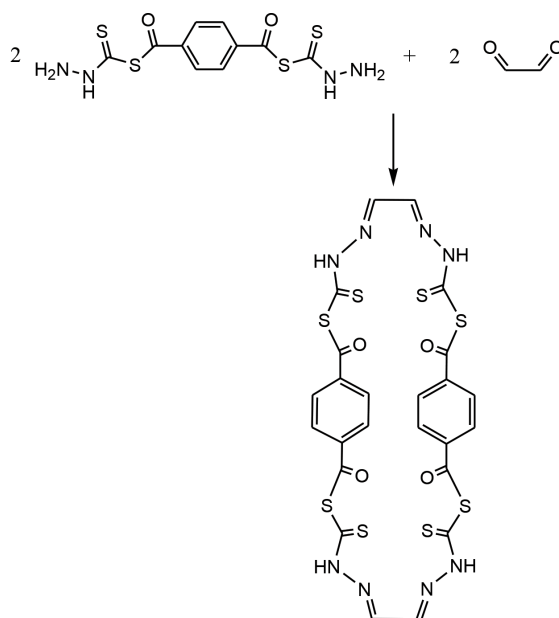
Figure 2. Synthesis of Schiff base,4,11,20,27-tetrathioxo-3,12,19,28-tetrathia-5,6,9,10,21,22,25,26-octaazatricyclo[28.2.2.2^{14,17}]hexatriaconta 1(33),6,8,14(36),15,17(35),22,24,30(34),31-decaene-2,13,18,29-tetraone (TGSB)

Synthesis of Macrocyclic Schiff Base Complexes

Macrocyclic complexes, M_2L or M_3L, were synthesised via the reaction between TGSB and metal acetates [Cu(II), Zn(II) and Cd(II)] or metal chlorides [Mo(V) and Ru(III)] in a molar ratio of 1:2. A solution of metal acetate or metal chloride (2 mmol) in ethanol (20 mL) was added to a solution of ligand (1 mmol) in absolute ethanol (40 mL). The mixture was refluxed for 16 h. The precipitate that formed was filtered off, washed with cold ethanol and then diethyl ether and finally air-dried.

Cytotoxicity Studies

Two bladder cancer cell lines, EJ-28 and RT-112, were used to study the anticancer properties of the compounds and were obtained from the National Cancer Institute, U.S.A. The cells were cultured in an RPMI-1640 medium supplemented with 10% fetal bovine serum. During plating, phosphate buffered saline was used to remove dead cells. Trypsin was used for cell detachment. A haemacytometer and trypan blue were used during cell counting. One hundred microlitres of mixture of medium and cells were titrated into 96 well plates using a multichannel pipette and incubated for 24 h with 5% CO_2 and 37°C. For treatment, six different concentrations (1 µM, 0.5 µM, 0.4 µM, 0.3 µM, 0.2 µM and 0.1 µM for active compounds; 5 µM, 4 µM, 3 µM, 2 µM, 1 µM and 0.5 µM for moderately active compounds) of compounds were used to determine the IC_{50} values that were the minimum concentration of the drug to kill the cancer cells by 50%. DMSO was used as the control. Anticancer activity or cytotoxicity was

determined using the microtitration of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay (Mosmann, 1983). MTT was reduced to purple formazan in living cells caused by mitochondrial reductase. Enzyme-linked immunosorbent assay (ELISA) was used to determine the anticancer activity of the compounds.

RESULTS AND DISCUSSION

A macrocyclic Schiff base, TGSB, and five of its metal complexes [Ru(III), Mo(V), Cd(II), Zn(II), and Cu(II)] were synthesised and fully characterised by physico-chemical and spectroscopic techniques. The macrocyclic Schiff base, TGSB, was synthesised by a [2+2] template condensation reaction of terephthaloyl-bis-dithiocarbazate and glyoxal in a ratio of 2:2 (Figure 2). All the synthesised compounds were stable in air and at room temperature. TGSB was soluble in DMSO, chloroform and absolute ethanol but insoluble in water. All the synthesised complexes were soluble in DMSO, DMF and chloroform but insoluble in water, absolute ethanol and methanol. The complexes were only slightly soluble in acetonitrile and ethyl acetate.

The melting point obtained for the TGSB was 137-138ºC and the melting point obtained for all the complexes was greater than 300ºC except for CdTGSB, which had a melting point of 262-263ºC. The synthesised metal complexes had higher melting points compared with the corresponding ligand (Ali et al., 2006); hence, the complexes were more stable compared with the ligand. The melting points for all the compounds were sharp (over a 1 or 2ºC range), indicating that the compounds were relatively pure and free from impurities. Unfortunately, repeated recrystallisation only produced crystals that were unsuitable for single-crystal X-ray structure determination. The differences between the experimental and calculated values for the elemental analysis were within the experimental errors (Table 1). RuTGSB, MoTGSB and CuTGSB were expected to be dinuclear while ZnTGSB and CdTGSB were expected to be trinuclear.

Table 1

Physical data, elemental analyses, magnetic moment and molar conductance of the compounds

Compound	Colour	Melting Point (ºC)	Yield (%)	μ_{eff} (B.M.)	Λ_m ($\Omega^{-1}cm^2mol^{-1}$)	Found (Calculated %) C	H	N	M
TDTC	White	128-130	77	-	-	35.1 (34.7)	2.9 (2.9)	16.4 (16.3)	-
TGSB	Yellow	137-138	52	-	-	39.6 (39.1)	2.3 (2.2)	15.7 (15.2)	-
RuTGSB	Black	>300	71	Dia	46.3	25.6 (25.1)	1.0 (1.1)	9.3 (9.8)	17.1 (17.6)
MoTGSB	Dark Blue	>300	56	Dia	6.96	25.0 (25.3)	1.2 (1.1)	9.8 (9.9)	16.0 (16.9)
CdTGSB	Brown	262-263	58	Dia	14.5	27.7 (28.3)	1.7 (1.7)	10.0 (9.4)	27.9 (28.3)
ZnTGSB	Dark Brown	>300	65	Dia	16.3	32.7 (32.1)	1.7 (1.9)	11.2 (10.7)	19.4 (18.7)
CuTGSB	Grey	>300	69	2.07	36.9	35.8 (35.1)	2.0 (2.2)	10.7 (10.2)	10.9 (11.6)

Dia - Diamagnetic

From physico-chemical and spectroscopy data, the atoms bound to the central metal in RuTGSB were two azomethine nitrogen atoms, two thiolate sulphur atoms and a Cl⁻ ion. For MoTGSB, the donor atoms involved in the complexation were two azomethine nitrogen atoms, two thiolate sulphur atoms, and three Cl⁻ ions, whereas for the ZnTGSB and CdTGSB complexes, the binding modes towards the central metal were two azomethine nitrogen atoms, two thiolate sulphur atoms and an oxygen atom from the CH_3COO^- ion. For CuTGSB, the donor atoms involved in the complexation were two azomethine nitrogen atoms and two thiolate sulphur atoms.

Hence, the proposed geometry for the RuTGSB, CdTGSB and ZnTGSB complexes was square pyramidal, while MoTGSB was expected to be pentagonal bipyramidal and CuTGSB had a square planar geometry. Long-distance binding is favoured for central large metal ions but not for smaller metal ions like Cu(II), Ni(II), Zn(II) ions generally (Temel et al., 2006; Morales et al., 2001). Hence, the structure of CuTGSB and ZnTGSB were kinetically and thermodynamically stable compared with that of MoTGSB, RuTGSB and CdTGSB.

FT-IR spectral analysis

The IR absorption bands of the macrocyclic Schiff base and its metal complexes are summarised in Table 2. The presence of the v(C=N) band at 1542 cm⁻¹ for TGSB indicated the successful formation of the Schiff base. During the complexation with metal salts, shifting of the v(C=N) band to higher wavenumbers was observed (West et al., 1998). This was due to the coordination of the azomethine nitrogen atom with the central metal ion when complexation occurred (Mohan et al., 1985). The v(N-H) stretching of the primary amine was found around 3432 cm⁻¹ in TGSB. On complexation, disappearance of the v(N-H) bands in the IR spectra of the metal complexes suggested deprotonation of the N-H because of bonding (Prasad et al., 2007). A proton was lost through complexation, leading to the formation of a covalent bond between the nitrogen donor atom of the ligand and the metal complex. The v(N-N) stretching for the Schiff base was observed at 1110 cm⁻¹. The v(N-N) stretching shifted to the right in the IR spectra of the metal complexes (Hossain et al., 1996). The IR spectrum also exhibited the presence of v(CSS) bands, which were observed at 877 cm⁻¹. The v(CSS) stretching shifted to higher wavenumbers in the IR spectra of the metal complexes. This was due to the reduction in the repulsion between the lone pairs of electrons on the sulphur atoms as a result of the coordination through the thiolate sulphur atom (Crouse et al., 2004).

Table 2

IR and electronic spectra data of the compounds

Compound	IR (cm⁻¹)				λ_{max} (nm)
	N-H	C=N	N-N	CSS	
TDTC	3422 w	-	1104 s	875 m	-
TGSB	3432 w	1542 w	1110 s	877 m	263, 298
RuTGSB	-	1569 m	1084 s	894 m	252, 303
MoTGSB	-	1563 w	1084 s	892 m	262, 294, 418
CdTGSB	-	1567 s	1079 m	887 m	264, 310, 420
ZnTGSB	-	1571 m	1079 s	897 m	248, 267, 304
CuTGSB	-	1564 m	1073 s	894 m	260, 291, 418, 578

w– weak intensity

m– medium intensity

s – strong intensity

NMR and GC-MS spectral analysis

The structure of the macrocyclic Schiff base, TGSB, was confirmed using ¹H NMR. The NMR spectrum data of TGSB are summarised in Table 3. TGSB contains a benzyl group para-substituted, which exhibited bands at δ_H = 8.24 to 8.05 ppm, and this chemical shift was assigned as the aromatic hydrogen bands. The resonance at δ_H = 7.27 ppm was attributed to the solvent, CDCl₃ (Gottlieb et al., 1997). Lastly, the band at δ_H = 4.07 ppm was assigned to =C-H, found at higher fields with the most highly shielded type of proton. The structure was further supported by ¹³C NMR, where the spectrum showed a weak singlet signal at δ_H = 166.35 ppm, which was ascribed to the azomethine carbon, C=N (Keypour et al., 2007). The doubly substituted benzyl showed bands at δ_H = 133.99 to 129.62 ppm. These bands appeared at the aromatic carbon region as expected (Keypour et al., 2008). Another significant signal was observed at δ_H = 77.37 ppm, which was assigned to the solvent CDCl₃.

Table 3

NMR spectral data of TGSB

Compound	¹H (ppm)			¹³C (ppm)	
	C-H Ar.	CDCl₃	=C-H	C=N	C=C Ar.
TGSB	8.05-8.24	7.27	4.07	166.35	129.62-133.99

Ar. - Aromatic

The mass spectra showed a molecular ion peak at *m/z* 346 for TDTC and this matched the molecular formula and molecular weight expected for TDTC, C₁₀H₁₀N₄O₂S₄. For TGSB, the molecular ion peak at *m/z* 736 matched the molecular formula and molecular weight expected,

$C_{24}H_{16}N_8O_4S_8$. The tallest peak in the mass spectrum had the most abundant ion formed in the ionisation chamber (Pavia et al., 2001). From the mass spectra, the base peak occurred at *m/z* 163, which was the most stable ion fragment, and corresponding to $C_8H_3O_2S^+$. The possible fragments of TGSB are shown in Figure 4.
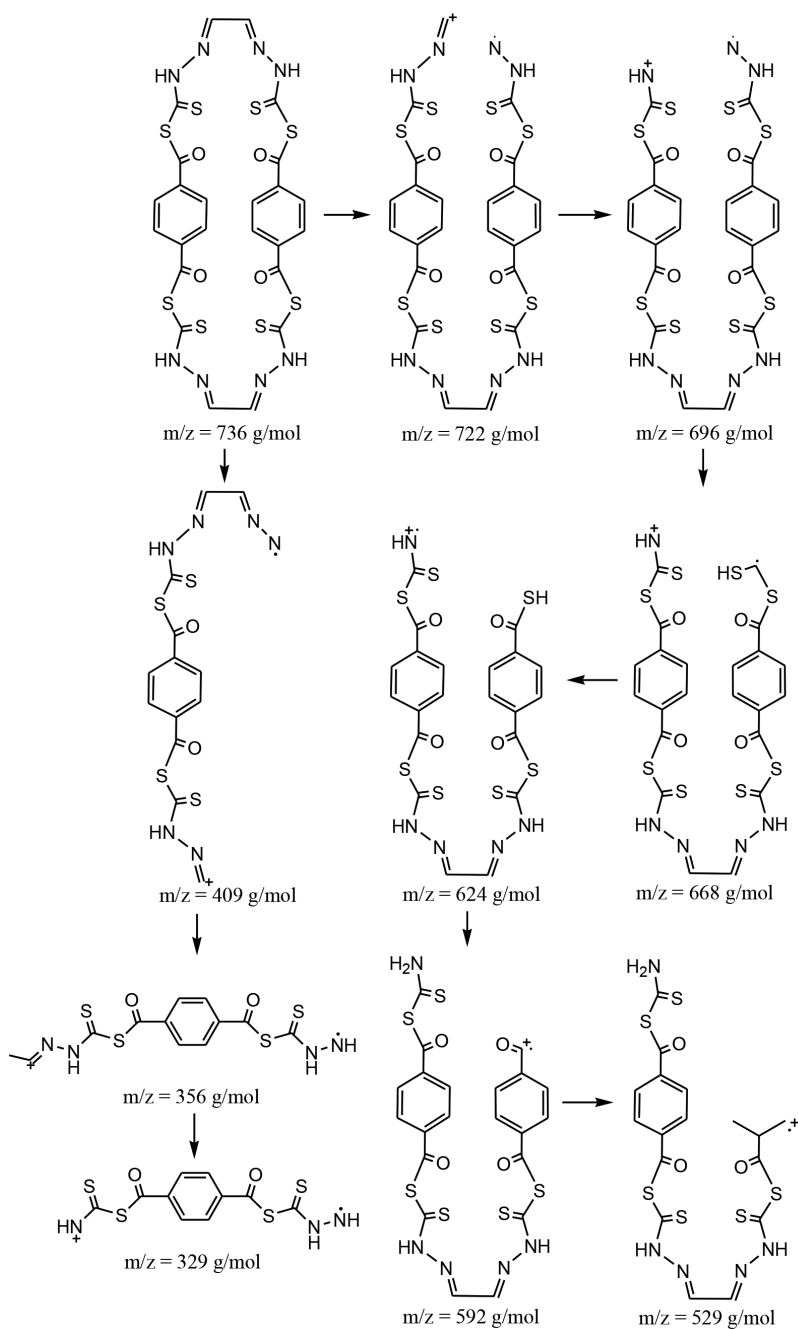
Figure 4. Fragmentation pattern of TGSB

Magnetic Susceptibility, Molar Conductivity and Electronic Spectral Analysis

The magnetic moment and molar conductance are shown in Table 1 and the electronic spectra data are shown in Table 2. In the magnetic susceptibility analysis, metal complexes are subjected to an external magnetic field and only the paramagnetic compounds are susceptible to this field due to the presence of unpaired electrons. At room temperature, the magnetic susceptibility analysis showed that all the complexes obtained except CuTGSB had diamagnetic properties. CuTGSB exhibited paramagnetism with magnetic moments of 2.07 B.M., indicating the presence of one unpaired electron (Geary, 1971). The slightly higher magnetic moment obtained was due to the distorted structure of the complex. A planar complex should have magnetic moments of 1.73 B.M. for Cu(II) ion (Ali et al., 1977).

From molar conductivity measurements, MoTGSB, CdTGSB and ZnTGSB were covalently bonded in the inner sphere and were non-electrolytes as the molar conductance values obtained were lower than 30 $\Omega^{-1}cm^2mol^{-1}$ in DMSO (Ilhan et al., 2008; Temel et al., 2008). RuTGSB and CuTGSB had higher conductance values, indicating their electrolytic nature. These high conductance values were due to the presence of other anions such as CH_3COO^- and Cl^- in the outer sphere. The presence of the Cl^- anion could be supported by a Cl^- test with $AgNO_3$ solution via halide abstraction (Abou-Hussein et al., 2012). Formation of the white precipitate, AgCl, happened when $AgNO_3$ reacted with RuTGSB, indicating the presence of the Cl^- anion on the outer sphere lattice. For MoTGSB, the Cl^- anion was only coordinated with the metal ion due to the absence of the white precipitate, AgCl.

In the electronic spectral analysis, four different transitions were observed in the complexes, namely $\pi \rightarrow \pi^*$, $n \rightarrow \pi^*$ intra-ligand transitions, ligand to metal charge transfer (LMCT) transition and $d \rightarrow d$ transitions. All the metal complexes showed an intense absorption band in the far UV region from 248 to 264 nm, which was assigned to the $\pi \rightarrow \pi^*$ transition (Ilhan, 2008). A moderately intense band in the region of 291 and 310 nm was assigned to the $n \rightarrow \pi^*$ transition. Bands between 418 and 420 nm were assigned as LMCT transition (Kalia et al., 2011) due to the sulphur donor atom to metal centre charge transfer transitions. Presence of this band further proved that the metal complexes were coordinated to sulphur through the thiolo sulphur atom (Crouse et al., 2004). Lastly, the lowest energy band at 578 nm was assigned to a $d \rightarrow d$ transition or $^1B_{1g}$ - ^1A_g transition for CuTGSB due to the presence of one unpaired electron consistent with a square-planar geometry (Raman et al., 2011). The proposed geometries for RuTGSB, CdTGSB and ZnTGSB were hence, square pyramidal and MoTGSB was expected to be pentagonal bipyramidal (Sayin, 2014). The proposed structures for all the synthesised complexes are shown in Figure 3 (a-e).
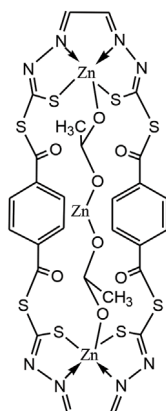
Figure 3(a). Proposed structure of the Zn(II) complex of 4,11,20,27-tetrathioxo-3,12,19,28-tetrathia-5,6,9,10,21,22,25,26-octaazatricyclo[28.2.2.2[14,17]]hexatriaconta 1(33),6,8,14(36),15,17(35),22,24,30(34),31-decaene-2,13,18,29-tetraone (ZnTGSB)
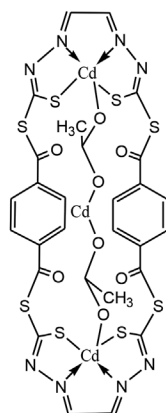
Figure 3(b). Proposed structure of the Cd(II) complex of 4,11,20,27-tetrathioxo-3,12,19,28-tetrathia-5,6,9,10,21,22,25,26-octaazatricyclo[28.2.2.2[14,17]]hexatriaconta 1(33),6,8,14(36),15,17(35),22,24,30(34),31-decaene-2,13,18,29-tetraone (CdTGSB)
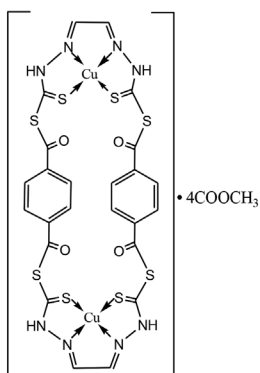
Figure 3(c). Proposed structure of the Cu(II) complex of 4,11,20,27-tetrathioxo-3,12,19,28-tetrathia-5,6,9,10,21,22,25,26-octaazatricyclo[28.2.2.2[14,17]]hexatriaconta 1(33),6,8,14(36),15,17(35),22,24,30(34),31-decaene-2,13,18,29-tetraone (CuTGSB)
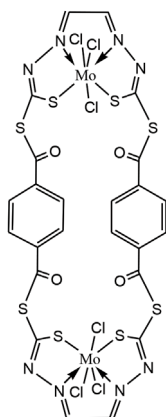
Figure 3(d). Proposed structure of the Mo(V) complex of 4,11,20,27-tetrathioxo-3,12,19,28-tetrathia-5,6,9,10,21,22,25,26-octaazatricyclo[28.2.2.2^{14,17}]hexatriaconta 1(33),6,8,14(36),15,17(35),22,24,30(34),31-decaene-2,13,18,29-tetraone (MoTGSB)
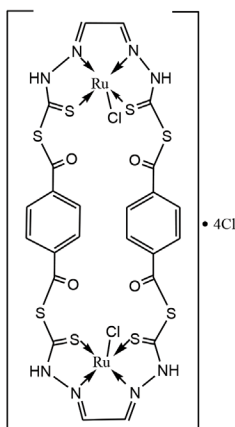
Figure 3(e). Proposed structure of the Ru(III) complex of 4,11,20,27-tetrathioxo-3,12,19,28-tetrathia-5,6,9,10,21,22,25,26-octaazatricyclo[28.2.2.2^{14,17}]hexatriaconta 1(33),6,8,14(36),15,17(35),22,24,30(34),31-decaene-2,13,18,29-tetraone (RuTGSB)

Cytotoxicity Studies

All the synthesised complexes were tested for their cytotoxic properties *in vitro* against two different bladder cancer cell lines, RT-112 and EJ-28. An invasive cell line is not native or fixed to a specific location and has a tendency to spread uncontrollably to a degree believed to cause damage to the environment. It has been found in previous studies that metal complexes of nitrogen-sulphur donor atoms were active against different cancer cell lines such as leukemic and breast cancer cell lines (Ali et al., 1974; Tarafder et al., 2001a; Tarafder et al., 2001b).

IC$_{50}$ values were determined using the MTT method (Mosmann, 1983) and the results are shown in Table 4. IC$_{50}$ is the inhibitory concentration at 50% i.e. the concentration of drug that inhibited the growth of cancer cells by 50%. IC$_{50}$ values less than 0.5 µM

indicated that the complex was strongly active, whereas IC_{50} values of 0.5-5.0 µM and more than 5.0 µM indicated that the complex was moderately active and inactive, respectively.

Table 4

IC_{50} data for the compounds against RT-112 and EJ-28 cell lines

Compound	IC_{50} (µM)	
	RT-112	EJ-28
TGSB	>5	>5
RuTGSB	0.47	>5
MoTGSB	>5	>5
CdTGSB	3.99	>5
ZnTGSB	>5	>5
CuTGSB	>5	>5

EJ-28 - Invasive human bladder carcinoma cell line

RT-112 - Minimum-invasive human bladder carcinoma cell line

TGSB was inactive (IC_{50}>5 µM) against both the RT-112 and EJ-28 cell lines. All the complexes were inactive (IC_{50}>5 µM) against EJ-28. This may be due to the huge and non-planar macrocyclic structure of the complexes. Interestingly, for RT-112, RuTGSB was strongly active with an IC_{50} value of 0.47 µM and CdTGSB was moderately active at an IC_{50} of 3.99 µM. MoTGSB, ZnTGSB and CuTGSB were inactive against RT-112 cells. The relative effectiveness of the anti-cancer activities showed that the different complexes with different metal ions acted against the same cell with different mechanisms, while a complex such as RuTGSB acted against different cells with a different mechanism of action. These results suggested that complexes can act as selective and specific agents for antitumor activities (Du et al., 2011). It has been suggested that chelation or coordination on complexation will reduce the polarity of the metal ion (Raman et al., 2009); thus, it can increase the lipophilic behaviour of the complexes, and this will enhance the permeation through the lipid layer of the membrane in the cells.

CONCLUSION

A new macrocyclic Schiff base and five new metal complexes containing the Schiff base were synthesised and characterised via various physico-chemical and spectroscopic techniques. The proposed geometries for RuTGSB, CdTGSB and ZnTGSB were hence, square pyramidal, while MoTGSB was expected to be pentagonal bipyramidal. CuTGSB obeyed square planar geometry. None of the complexes were active against EJ-28 bladder cell lines but RuTGSB was strongly active and CdTGSB was moderately active against RT-112 bladder cell lines, indicating that the complexes were selective and specific in action.

ACKNOWLEDGEMENT

REFERENCES

Abou-Hussein, A. A. A., & Linert, W. (2012). Synthesis, spectroscopic and biological activities studies of acyclic and macrocyclic mono and binuclear metal complexes containing a hard-soft Schiff base. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 95*, 596–609.

Ali, M. A., Mizra, A. H., Bujang, F. H., Hamid, M. H. S. A., & Bernhadt, P. V. (2006). Synthesis, characterization and X-ray crystallographic structural study of copper(II) and nickel(II) complexes of the 2-quinoline carboxaldehyde Schiff base of *S*-methyldithiocarbazate (Hqaldsme). *Polyhedron, 25*(17), 3245–3252.

Ali, M. A., & Livingstone, S. E. (1974). Metal complexes of sulphur-nitrogen chelating agents. *Coordination Chemistry Reviews, 13*(2-3), 101–132.

Ali, M. A., & Tarafder, M. T. H. (1977). Metal complexes of sulphur and nitrogen-containing ligands: Complexes of s-benzyldithiocarbazate and a Schiff base formed by its condensation with pyridine-2-carboxaldehyde. *Journal of Inorganic and Nuclear Chemistry, 39*(10), 1785–1791.

Aqra, F. M. A. M. (1999). Transition metal ions as cores in the construction of an unprecedented large macrocycle. *Transition Metal Chemistry, 24*(1), 71–73.

Azarkish, M., & Sedaghat, T. (2012). Synthesis, spectral studies, thermal behavior, and antibacterial activity of Ni(II), Cu(II), and Zn(II) complexes with an ONO tridentate Schiff base. *Chinese Chemical Letters, 23*(9), 1063–1066.

Borisova, N. E., Reshetova, M. D., & Ustynyuk, Y. A. (2007). Metal-free methods in the synthesis of macrocyclic Schiff bases. *Chemical Reviews, 107*(1), 46–79.

Chandra, S., & Pundir, M. (2008). Spectroscopic characterization of chromium(III), manganese(II) and nickel(II) complexes with a nitrogen donor tetradentate, 12-membered azamacrocyclic ligand. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 69*(1), 1–7.

Constable, E. C. (1999). *Coordination chemistry of macrocyclic compounds*. New York: Oxford University Press Inc.

Crouse, K A., Chew, K. B., & Tarafder, M. T. H. (2004). Synthesis, characterization and bio-activity of *S*-2-picolyldithiocarbazate (S2PDTC), some of its Schiff bases and their Ni(II) complexes and X-ray structure of *S*-2-picolyl-β-*N*-(2-acetylpyrrole)dithiocarbazate. *Polyhedron, 23*(1), 161–168.

Driggers, E. M., Hale, S. P., Lee, J., & Terrett, N. K. (2008). The exploration of macrocycles for drug discovery – An underexploited structural class. *Nature Reviews Drug Discovery, 7*(7), 608–624.

Du, D., Jiang, Z., Liu, C., & Sakho, A. M. (2011). Macrocyclic organotin(IV) carboxylates based on benzenedicarboxylic acid derivatives: Syntheses, crystal structures and antitumor activities. *Journal of Organometallic Chemistry, 696*(13), 2549–2558.

Geary, W. J. (1971). The use of conductivity measurements in organic solvents for the characterisation of coordination compounds. *Coordination Chemistry Reviews, 7*(1), 81–122.

Gottlieb, H. E., Kotlyar, V., & Nudelman, A. (1997). NMR chemical shift of common laboratory solvents as trace impurities. *The Journal of Organic Chemistry, 62*(21), 7512–7515.

Hossain, M. E., Alam, M. N., Ali, M. A., & Nazimuddin, M. (1996). The synthesis, characterization and bioactivities of some copper(II) complexes of the 2-acetylpyridine Schiff bases of s-methyl- and s-benzyldithiocarbazate, and the x-ray crystal structure of the nitrato(s-benzyl-β-n-(2-acetylpyridyl) methylenedithiocarbazato)copper(II) complex. *Polyhedron, 15*(5-6), 973–980.

Ilhan, S. (2008). Preparation and characterization of binuclear Cu(II) complexes derived from diamines and dialdehydes. *Journal of Coordination Chemistry, 61*(18), 2884–2895.

Ilhan, S., Temel, H., Sunkur, M., & Tegin, I. (2008). Synthesis, structural characterization of new macrocyclic Schiff base derived from, 1, 6-bis-(2-formylphenyl) hexane and 2, 6-diaminopyridine and its metal complexes. *Indian Journal of Chemistry, 47A*(4), 560–564.

Kalia, S. B., Lumba, K., & Sankhyan, P. (2011). Magnetic and spectral studies on nickel (II) and copper (II) dithiocarbazates derived from isoniazid. *Journal of Coordination Chemistry, 64*(7), 1216–1228.

Keypour, H., Azadbakht, R., & Khavasi, H. (2008). Synthesis and characterization of three Cd(II) Schiff-base macrocyclic N_3O_2 complexes. *Polyhedron, 27*(2), 648–654.

Keypour, H., Goudarziafshar, H., Brisdon, A. K., & Pritchard, R. G. (2007). New macrocyclic Schiff base complexes incorporating a phenanthroline unit: Part 1; Template synthesis of three cadmium(II) complexes and crystal structure, NMR and ab initio studies. *Inorganica Chimica Acta, 360*(7), 2298–2306.

Mewis, R. E., & Archibald, S. J. (2010). Biomedical applications of macrocyclic ligand complexes. *Coordination Chemistry Reviews, 254*(15-16), 1686–1712.

Mohamed, G. G., Omar, M. M., & Ibrahim, A. A. (2009). Biological activity studies on metal complexes of novel tridentate Schiff base ligand. Spectroscopic and thermal characterization. *European Journal of Medicinal Chemistry, 44*(12), 4801–4812.

Mohan, M., Sharma, P., & Jha, N. K. (1985). Metal(II) chelates of 4-methyl-5-amino-1-formylisoquinoline thiosemicarbazone: Their preparation, characterization and antitumouractivity. *Inorganica Chimica Acta, 106*(4), 197–201.

Morales, R. G. E., Jara, G. P., & Vargas, V. (2001). Ultraviolet absorption bands and electronic charge transfers of salicylideneanilines in singlet excited states. *Spectroscopy Letters, 34*(1), 1–12.

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunological Methods, 65*(1-2), 55–63.

Pavia, D. L., Lampman, G. M., & Kriz, G. S. (2001). *Introduction to spectroscopy.* Pacific Grove, California: Thomson Learning Inc.

Prasad, R. N., Mathur, M., & Upadhyay, A. (2007). Synthesis and spectroscopic studies of Cr(III), Fe(III) and Co(II) complexes of hexaazamacrocycles. *Journal of the Indian Chemical Society, 84*(12), 1202–1204.

Raman, N., & Joseph, J. (2009). Synthesis, spectral characterization and antimicrobial activity of macrocyclic Schiff-base copper(II) complexes containing polycrystalline nanosized grains. *Journal of Coordination Chemistry, 62*(7), 1162–1171.

Raman, N., Raja, S. J., & Sakthivel, A. (2009). Transition metal complexes with Schiff-base ligands: 4-aminoantipyrine based derivatives – A review. *Journal of Coordination Chemistry, 62*(5), 691–709.

Sayin, K. (2014). Theoretical spectroscopic study of seven zinc(II) complex with macrocyclic Schiff-base ligand. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 133*, 348–356.

Singh, M., Aggarwal, V., Singh, U. P., & Singh, N. K. (2009). Synthesis, characterization and spectroscopic studies of a new ligand [N′-(2-methoxybenzoyl)hydrazinecarbodithioate] ethyl ester and its Mn(II) and Cd(II) complexes: X-ray structural study of Mn(II) complex. *Polyhedron, 28*(1)*,* 107–112.

Tarafder, M. T. H., Kasbollah, A., Crouse, K. A., Ali, A. M., Yamin, B. M., & Fun, H. K. (2001a). Synthesis and characterization of Zn(II) and Cd(II) complexes of S-benzyl-β-N-(2-pyridyl) methylenedithiocarbazate (HNNS): Bioactivity of the HNNS Schiff base and its Zn(II), Cu(II) and Cd(II) complexes and the X-ray structure of the [Zn(NNS)$_2$] complex. *Polyhedron, 20*(18), 2363–2370.

Tarafder, M. T. H., Saravanan, N., Crouse, K. A., & Ali, A. M. (2001b). Coordination chemistry and biological activity of nickel(II) and copper(II) ion complexes with nitrogen–sulphur donor ligands derived from S-benzyldithiocarbazate (SBDTC). *Transition Metal Chemistry, 26*(6), 613–618.

Temel, H., Alp, H., Ilhan, S., & Ziyadanogullari, B. (2008). Spectroscopic and extraction studies of new transition metal complexes with N, N′-bis (2-aminothiophenol)-1, 4-bis (2-carboxaldehydephenoxy) butane. *Journal of Coordination Chemistry, 61*(7), 1146–1156.

Temel, H., Ilhan, S., Aslanoglu, M., Kilic, A., & Tas, E. (2006). Synthesis, spectroscopic and electrochemical studies of novel transition metal complexes with quadridentate Schiff base. *Journal of the Chinese Chemical Society, 53*(5), 1027–1031.

West, D. X., El-Sawaf, A. K., & Bain, G. A. (1997). Metal complexes of N(4)-substituted analogues of the antiviral drug methisazone {1-methylisatin thiosemicarbazone}. *Transition Metal Chemistry, 23*(1), 1–6.

Xie, Y. S., Pan, X. H., Zhao, B. X., & Liu, J. T. (2008). Synthesis, structure characterization and preliminary biological evaluation of novel 5-alkyl-2-ferrocenyl-6,7-dihydropyrazolo[1,5-*a*]pyrazin-4(5*H*)-one derivatives. *Journal of Organometallic Chemistry, 693*(7), 1367–1374.

SUPPLEMENTARY DATA
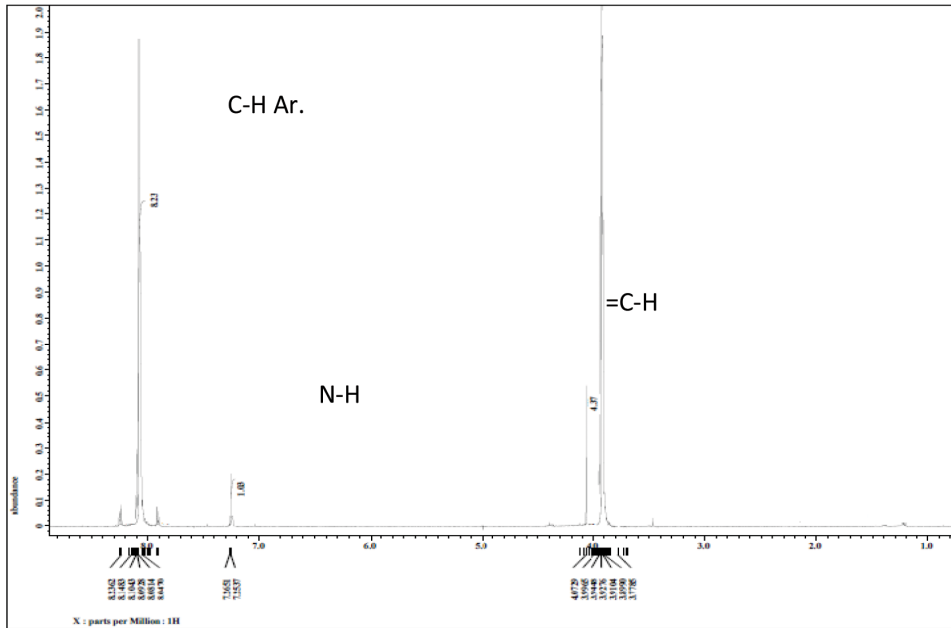
Supplement 1. ^{1}H NMR spectrum of TGSB
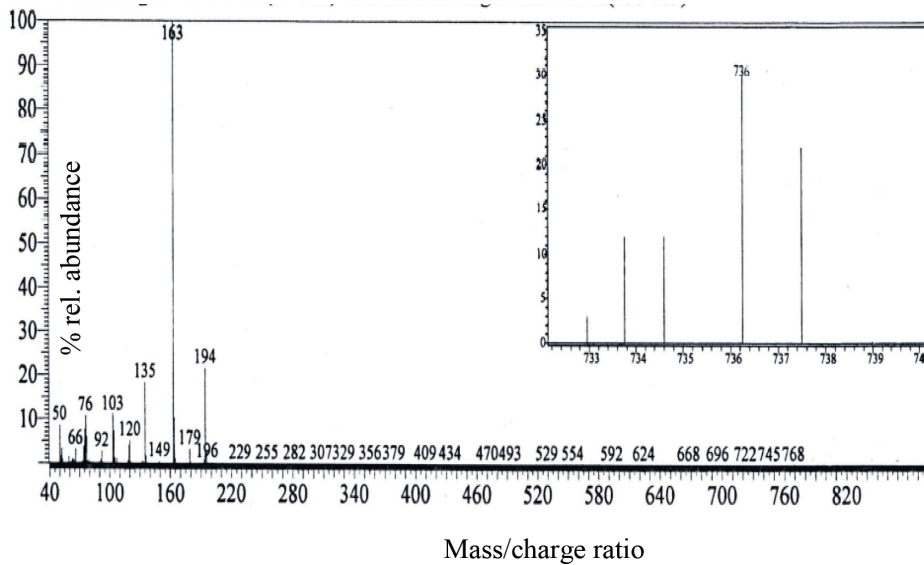
Supplement 2. Mass spectrum of TGSB
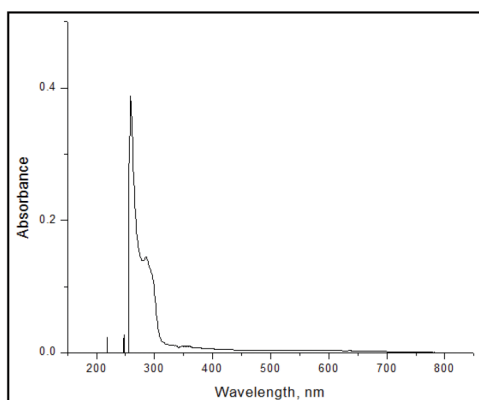
S 3a. UV/Vis spectrum of TGSB.
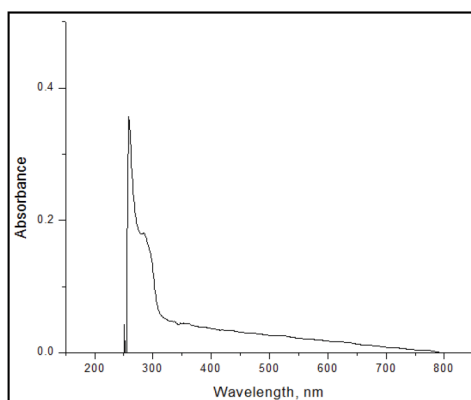
S 3b. UV/Vis spectrum of RuTGSB.
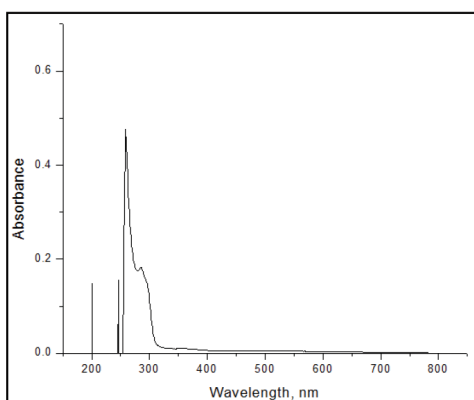
S 3c. UV/Vis spectrum of MoTGSB.
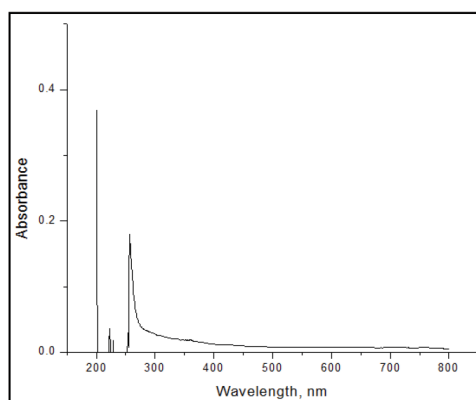
S 3d. UV/Vis spectrum of CdTGSB.
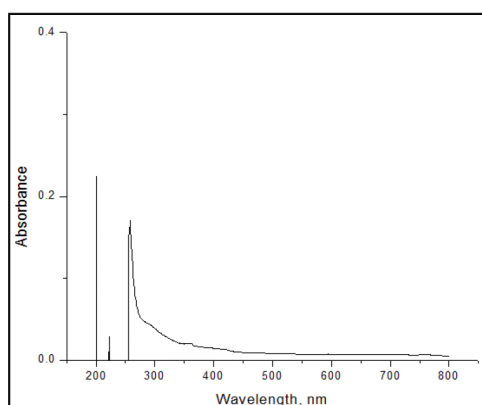
S 3e. UV/Vis spectrum of ZnTGSB.
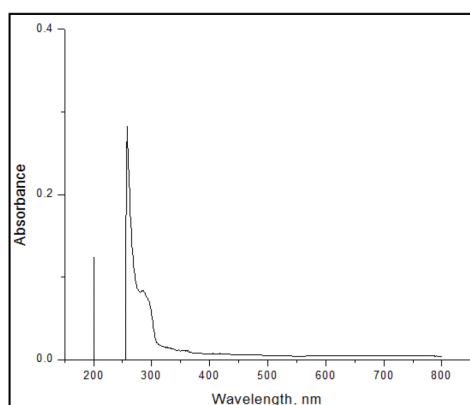
S 3f. UV/Vis spectrum of CuTGSB.

SCIENCE & TECHNOLOGY

The Effect of Pre-processing and Testing Methods on Online Kannada Handwriting Recognition: Studies Using Signal Processing and Statistical Techniques

S. Ramya* and Kumara Shama

*Department of Electronics and Communication Engineering, Manipal Institute of Technology,
Manipal Academy of Higher Education, Manipal 576104, Karnataka, India*

ABSTRACT

Pre-processing and testing methodology plays a significant role in online handwritten character recognition. Although many researchers have proposed several pre-processing and testing methods, the effect of these techniques on the recognition and comparisons among them are ignored. In this work, experiments were conducted to analyse the effect of various pre-processing and testing methods on Kannada handwritten data. The focus of the present work is to statistically quantify the effect on recognition time and accuracy through experiments using different pre-processing methods on online handwritten data processed by the Support Vector Machine (SVM). The performance of the SVM is also compared with various other training and testing methodology. The performance of the online handwriting recognition system is affected dramatically by the various pre-processing and testing methods. Stratified tenfold cross validation showed better performance for the Kannada handwritten dataset.

Keywords: Bootstrapping, cross-validation, down sampling, normalisation, online handwriting recognition, resampling, single partition testing, smoothing

INTRODUCTION

New technology like the pen-based interface plays a significant role in human-computer interaction. The challenge with using pen-enabled interfaces is to recognise user handwriting. Handwriting Recognition (HR) is a field of pattern recognition that aims to assign objects to a set of categories (Plamondon et al., 2000). In Online Handwriting Recognition (OHR) systems, handwritten data are captured as a sequence of coordinate locations during the writing process, known as digital inking. Online handwritten data are diversified due to individual writing styles and contexts

(Tappert et al., 1990). To achieve robust recognition, handwritten data are subjected to pre-processing. Pre-processing is one of the important stages in handwriting analysis; it addresses the methodology of elimination of imperfections, data reduction, normalisation and resampling (Biadsy et al., 2011). Vuori et al. (2001) applied down sampling by two to reduce processing time followed by normalisation. For smoothing, a five-tap Gaussian filter was used (Namboodiri et al., 2004; Deepu et al., 2004; Joshi et al., 2005), followed by normalisation and resampling being performed. A moving average filter (Prasad et al., 2010; Babu et al., 2007) with a span of 3 has also been used for smoothing data. Elimination of duplicate points, a necessary procedure for accurate analysis, was emphasised by Bahlmann (Bahlmann et al., 2004). To make a stroke equi-spaced along the arc length, a resampling size of 60 was used by Prashanth et al. (2007); a resampling size of 30 has also been reported (Prasad et al., 2007; Babu et al., 2007).

Many researchers have proposed various methodology for pre-processing to achieve better recognition accuracy. The widely implemented pre-processing techniques in literature are smoothing, resampling and normalisation. Down sampling and duplicate point removal are incorporated to reduce data content. Though these are common pre-processing methods, variation in the type of smoothing techniques with various window spans and resampling with different resample size are also used. It is essential to meticulously select the pre-processing methods for the given data set to enhance the performance of the classifier for character recognition. However, there is less effort into investigating the effect of these pre-processing methods in terms of character recognition accuracy and processing time. Therefore, efficacy experiments are conducted on the pre-processing techniques, namely, smoothing, normalisation, resampling, down sampling and removal of redundant points. Various combinations of pre-processing methods are tested and their effect on recognition accuracy and processing time with a single classifier for the Kannada handwritten data set is statistically quantified.

Testing and training is yet another critical phase of the recognition system that provides the recognition estimation of the classifier. In the field of online handwriting recognition, generally, experiments are conducted either with a single partition or K-fold cross validation (Namboodiri et al., 2004; Bharath et al., 2009). Studies on various testing and training methods, namely Holdout, K-fold cross validation (with or without stratification), random sub-sampling and bootstrapping have been done (Kohavi et al., 1995; Claudia et al., 2005; Kim, 2009) in various fields of research. Each training method behaves differently and this influences the bias and variance in the error. For estimating the accuracy of the classifier, low bias and low variance in error are preferable as they indicate the classification stability of the classifier. Experiments are conducted to compare these five types of methodology with recognition accuracy, bias and variance in recognition error as the parameters of comparison.

Our work focusses on the design of experiments on pre-processing and testing methodology on Kannada handwritten data as a case study and analyses their influence on the recognition system.

METHODS

Online Kannada Handwriting Recognition System

The block diagram of online Kannada handwriting recognition system is as shown in Figure 1. Kannada handwritten characters with handwriting variations in size, translation, speed and noise were collected and stored in the database. Kannada characters from the database were input to the pre-processing module. The pre-processing module did the standardisation of script to handle the variations in handwriting. Features, namely, normalised horizontal and vertical coordinates, tangent angular differences of consecutive points, writing directions and curvature were extracted from pre-processed data and used to train and test the classifier. The performance of the classifier was evaluated for the recognition accuracy and the processing time with different combinations of pre-processing and testing methodology.

Figure 1. Block diagram of online Kannada handwriting recognition system

Data Acquisition

Since the standardised online Kannada handwritten character database is not available publicly, a database was created. The script of Kannada language is syllabic in nature. It evolved from the Kadamba script (Ghosh et al., 2010). It is one of the official languages of India and the state language of Karnataka. The language has 16 vowels (Swaragalu) and 34 consonants (Vyanjanagalu) (Kunte et al., 2007), which constitute the basic Kannada character set as shown in Figure 2. To collect the Kannada character data, a graphical user interface was developed in visual C# using the Genius Mouse Pen i608X as an input device. Native users of Kannada from different age groups ranging from 10 years to 60 years of age were selected. A total of 150 data samples of each Kannada vowel and consonant was collected. Figure 2(c) shows the sample data of online Kannada handwritten characters. The work was focussed on the basic Kannada character set; hence, the dataset was restricted to 50 character classes and it can be extended to consonant-vowel modifier combinations by pre-segmentation (Prasad et.al 2009). The horizontal and vertical coordinates of pen trace with a time stamp were acquired and saved into a text file with a unique file ID. The database included characters with varying sizes and speeds of writing.
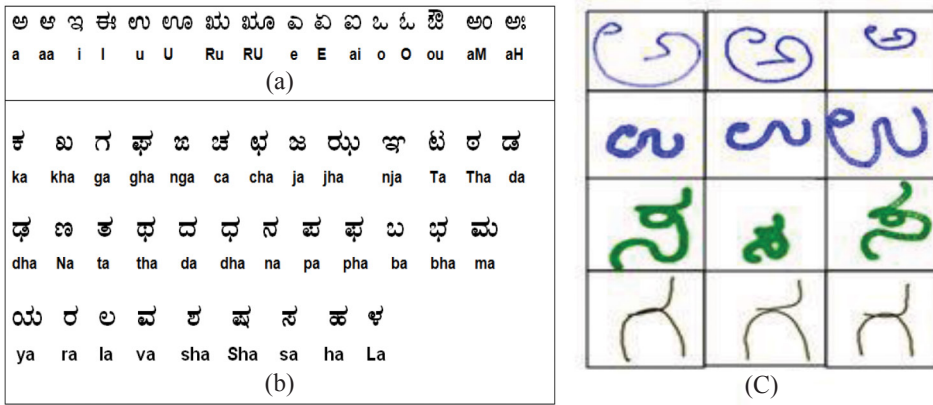
Figure 2. Kannada character set (a) Vowels (b) Consonants (c) Sample data

Pre-Processing

The raw data collected from the user are usually noisy and inconsistent. To obtain efficient classification, these data are pre-processed. In this work, the pre-processing methods, namely, normalisation, resampling, smoothing and down sampling as well as the removal of duplicated points are implemented.

Normalisation. The size and location of the input handwritten text differ from user to user. The goal of normalisation is to make data invariant to the size (Li et al., 1997). Normalised horizontal and vertical coordinates, $X_{norm(i)}$ and $Y_{norm(i)}$ are calculated using Equations 1 and 2.

$$X_{norm(i)} = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \tag{1}$$

$$Y_{norm(i)} = \frac{(Y_i - Y_{min})}{(Y_{max} - Y_{min})} \tag{2}$$

where, (X_{max}, Y_{max}) and (X_{min}, Y_{min}) are the maximum and minimum horizontal and vertical coordinates of a character, allowing the data to be within the bounding box [0,1].

Resampling. The raw data captured by the digitiser are sampled uniformly in time. Resampling is performed to obtain data that are uniformly sampled along the pen trace to make the data invariant to writing speed. The original data points are replaced with regularly spaced points along the arc length using the following piecewise linear interpolation algorithm.

Algorithm: Resampling Algorithm
1. Find the Euclidean distance between the consecutive points for all the points.
2. Find the cumulative distance (D), which gives the total length of the pen trajectory.
3. Find the required spacing by dividing the distance D by the number of intervals required (Resampling size)
4. Replace the original points with a calculated set to obtain equi-spaced points.

Smoothing. While writing on the pen tablet, erratic pen movement causes shaking; this introduces noise into the signal. In smoothing, the data points of a signal are modified such that adjacent point coordinate values vary smoothly. From the literature survey, it was found that researchers generally implemented the Gaussian filter or Moving Average filter with different window spans. Gaussian smoothing computes the average of neighbouring points weighted by the Gaussian kernel. A Gaussian kernel is in the shape of a normal distribution curve (Harris et al., 1978). The input data are convolved with the Gaussian window span of N, resulting in a smoothing of data. Using Equation 3, the coefficients of a Gaussian window of size N were computed in this study.

$$W(n) = e^{-\frac{1}{2}\left(\alpha\frac{n}{N/2}\right)^2} \qquad [3]$$

where, n lies $-(N-1)/2 \leq n < (N-1)/2$. The relationship between the standard deviation σ and the Gaussian probability density function is given by Equation 4.

$$\sigma = \frac{N}{2\alpha} \qquad [4]$$

In a moving average filter, each data point is replaced with the mean of the adjacent data points defined within the span (Jaeger et al., 2001). The smoothening response is given by Equation 5.

$$ys_{(i)} = \frac{1}{(2M+1)}\left(y_{(i+M)} + y_{(i+M-1)} + \cdots + y_{(i-M)}\right) \qquad [5]$$

where, $ys_{(i)}$ is the resultant smoothed value for the ith data point, with M as the number of neighbouring data points on either side of $ys_{(i)}$ and $(2M+1)$ as the filter span. In this study experiments were conducted to evaluate the performance of these filters on Kannada OHR.

Removal of duplicate points. A point is said to be a duplicate point if it has the same horizontal and vertical coordinate values as that of its adjacent point. Duplicate points result when the writing process is slower than the digitiser sampling frequency. These data are redundant points; hence, they are removed from the captured ink data. Thresholding of the horizontal and vertical coordinates' differences to zero is implemented to remove duplicate points.

Down sampling. Down sampling the data will reduce the number of horizontal and vertical coordinates in a character stroke, which will reduce the pre-processing time. Experiments were conducted in this study with down sampling by factor 2. Hence, every alternative sample was selected.

Feature Extraction and Character Recognition

Pre-processed data are the input used in the feature extraction module. The feature extraction module captures the essential characteristics of a stroke to help in recognition. Feature selection is an important task that affects the performance of the classifier (Namboodiri et al., 2004). Kannada characters are curvilinear in nature, similar to Telugu and Tamil script. Writing

direction and curvature at a point show a more promising result for Telugu and Tamil (Bharath et al., 2009; Toselli et al., 2007; Prashanth et al., 2007). Other features like normalised horizontal and vertical coordinates are also widely used in Indic script (Bharath et al., 2009). Therefore, these features were selected in this work for the online Kannada handwriting recognition.

Writing direction at a given point (x (i), y (i)) is well described using the cosine and sine angle given by the following equations.

$$\cos \propto (i) = \frac{\Delta x(i)}{\Delta s(i)} \tag{6}$$

$$\sin \propto (i) = \frac{\Delta y(i)}{\Delta s(i)} \tag{7}$$

where, **Δ**s (i), **Δ**x (i) and **Δ**y (i) are defined by Equations 8, 9 and 10, respectively.

$$\Delta s(i) = \sqrt{\Delta x^2(i) + \Delta y^2(i)} \tag{8}$$

$$\Delta x(i) = x(i-1) - x(i+1) \tag{9}$$

$$\Delta y(i) = y(i-1) - y(i+1) \tag{10}$$

The curvature at a point can be calculated as:

$$\cos \beta(i) = \cos \alpha(i-1) \cos \alpha(i+1) + \sin \alpha(i-1) \sin \alpha(i+1) \tag{11}$$

$$\sin \beta(i) = \cos \alpha(i-1) \sin \alpha(i+1) - \sin \alpha(i-1) \cos \alpha(i+1) \tag{12}$$

These features are used to train the classifier. The Support Vector Machine (SVM) is widely used in object classification and recognition (Cortes et al., 1995), content-based pattern recognition, handwriting recognition, biometrics etc. SVM is based on supervised learning that analyses data and classifies the data. Many works on OHR using the SVM classifier for

different languages (Bahlmann et al., 2002) including Kannada (Rampalli et al., 2011) have been reported. Hence, the SVM classifier with RBF kernel was implemented in this study and tested for Kannada handwritten characters.

Experimental Design

A flow chart of the implementation methodology to evaluate the effect of different pre-processing methods is given in Figure 3. Numbers 1 to 4 indicate test cases. Test Case 1 focusses on performance evaluation of different resampling sizes on the recognition system. The normalised data were resampled by the following sizes.

Resample size = [10, 20, 30, 40, 50, 60, 70, 80, 90]

The normalisation and resampling methods are considered the basic pre-processing module. This module was present in all the test cases. Test Case 2 evaluated the impact of smoothing algorithms on recognition. The performance of the moving average filter and the Gaussian filter with window spans 3, 5 and 7 was tested, and their performance was compared based on recognition time and accuracy. Test Case 3 focused on down sampling the data by half, while Test Case 4 focused on analysing the effect of removing duplicate points before the basic pre-processing module.

The recognition accuracy of the classifier depends on the data used to train and test the classifier. Error estimation is a challenging task in handwriting recognition, where mainly small sample sizes with large features exist. Selecting the appropriate training and testing method is essential for evaluating the stability of the classifier. Test Case 5 compared the performance of the classifier for the following different training and testing methods.

In the single partition training-testing method, also known as 'Holdout', a pre-defined percentage of data was used to train and test the classifier. The single partition training-testing method with 70% of the data for training and 30% of the data for testing was implemented and tested. In random subsampling, the holdout procedure was repeated a number of times and the average accuracy was estimated (Kohavi et al., 1995). In this study, the holdout procedure was repeated 100 times for fair comparison (Kim et al., 2009).

In the tenfold cross-validation method, N data samples were divided into 10 groups. Then the classifier was trained with nine groups and tested with one cluster. Training and testing was repeated 10 times, with each cluster used once as a test set. Final performance estimation was done by averaging the 10 results. In stratified cross validation, the folds were divided to contain approximately equal proportions of character class.

The bootstrap is another method for accuracy estimation (Efron et al., 1993). In this method, the dataset is sampled with a replacement. When the data are drawn with a replacement, the particular sample may occur more than once in the bootstrap set. Then, the remaining data serve as the test data set.
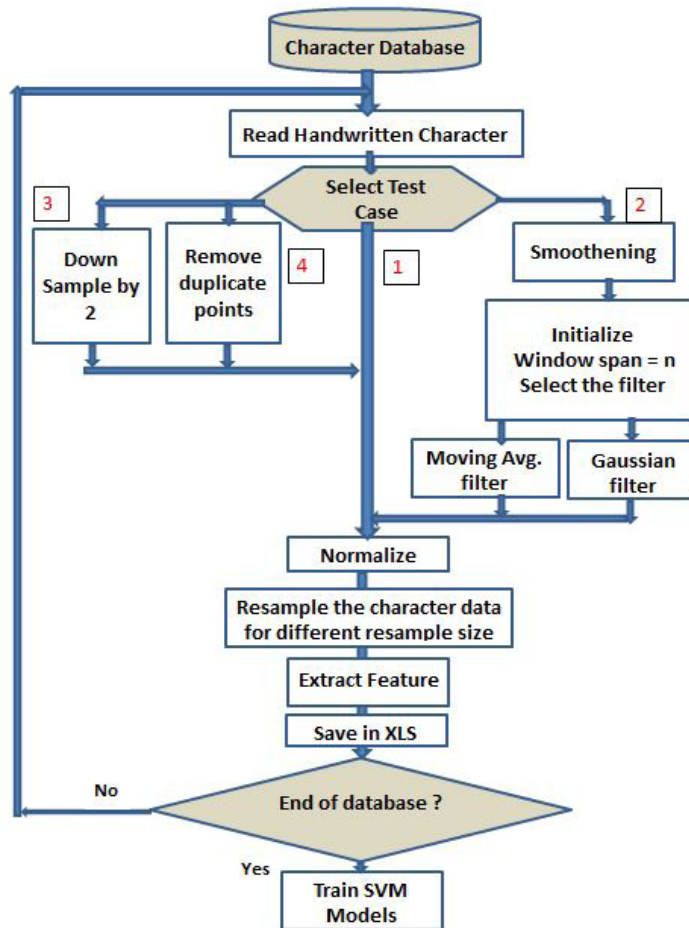
Figure 3. Flow chart of design of experiments to evaluate the effect of different pre-processing methods

Test Cases 1 to 4 were tested and evaluated using the single partition method. In Test Case 5, these results were compared with five different testing and training methods for recognition accuracy, error bias and variance to quantify the better testing-training methodology for character recognition.

RESULTS AND DISCUSSION

The effects of various pre-processing methodology were analysed and statistically evaluated. Overall recognition accuracy and processing time were parameters used for performance evaluation.

Overall Recognition Accuracy is the number of correct predictions made divided by the total number of predictions. It is defined as:

$$\text{Overall RecognitionAccuracy} = \sum_{i=0}^{n} \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i} \qquad [13]$$

Where,

- n is the number of samples
- True Positives (TP): Number of positive examples, labelled correctly
- False Positives (FP): Number of negative examples, labelled incorrectly
- True Negatives (TN): Number of negative examples, labelled correctly
- False Negatives (FN): Number of positive examples, labelled incorrectly

Effect of Resampling Size

To overcome variations in writing speed, the data were resampled to a fixed size, using the linear interpolation method. Resampling of data with different resampling sizes is shown in Figure 4. In some cases, like the resampling size of 10 as seen in Figure 4(b), the data get distorted, which may lead to misclassification in the recognition module. Therefore, selecting the optimum resampling size is a critical criterion for online handwriting recognition.
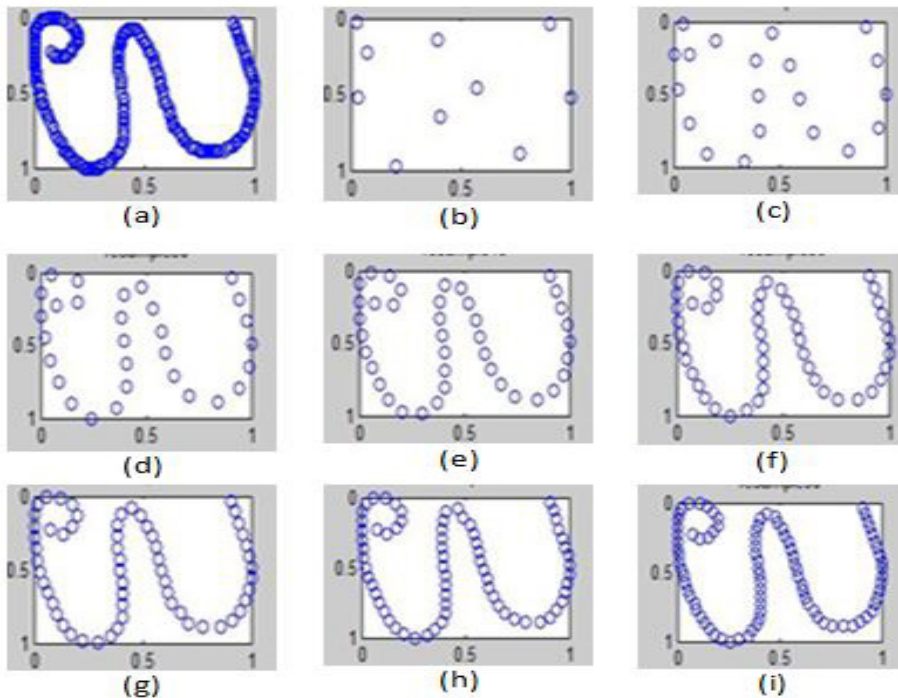
Figure 4. Effect of resampling on character shape (a) Character shape before resampling (b) – (i) Character shape for resample size 10, 20, 30, 40, 50, 60, 70, 90, respectively
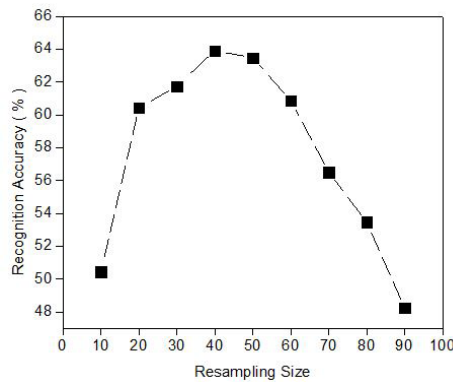
Figure 5(a). Recognition accuracy for different resampling sizes

The effect of resampling size on recognition accuracy is as depicted in Figure 5(a). When the resampling size is increased from 10 to 40, recognition accuracy improves but reduces as the resampling size reaches beyond 40. This is due to over-fitting of the features. For lower resampling sizes, recognition accuracy is low due to the distortion of the character shape, as seen in Figure 4. The resampling size of 40 exhibits the highest recognition rate of 63.91% for the Kannada dataset.

To quantify the effect of resampling size on recognition accuracy, one variable t-test was applied with the following hypothesis 'Ha' for 95% confidence.

Ha = There is no effect caused by the resampling size on recognition accuracy.

Table 1

Results of one variable t-test on resampling size

Mean	St Dev	SE Mean	T Value	P Value
57.68	5.78	1.93	29.96	0.000

Table 1 presents the results of the one variable t-test on resampling size. The mean of the data was 57.68 with a standard deviation (St Dev) of 5.78 and standard error means (SE mean) of 1.93. The t value was 29.96 with a p value of 0. Since $p<0.05$, proving that there was a significance variation in recognition accuracy with resampling size.
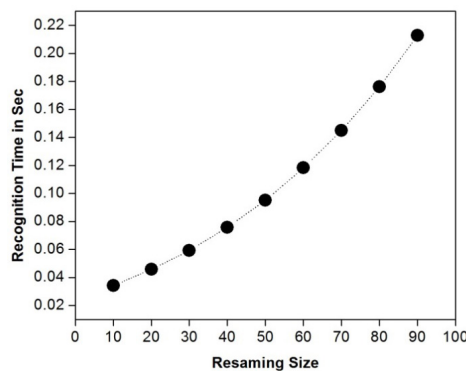
Figure 5(b). Recognition time required for different resampling sizes

Recognition time was plotted against resampling size, as shown in Figure 5(b). As the resampling size increased, it was obvious that the data samples also increased, which in turn increased the feature set length. Increase in the feature set length contributed to the increase in computation. Hence, the recognition time varied in almost linear direction with the resampling size. Although processing time increased with the resampling size, recognition accuracy did not increase with the resampling size as seen in Graph 5(a). Hence, empirical testing was required to obtain the optimum resampling size for the given dataset in general and the Kannada character set in particular. These experiments and statistical quantification signified that resampling size affected the performance of the classifier.

Effect of Smoothing

Due to shaking in the digitiser or due to writing style, noise can be introduced in the handwritten data. Smoothing is carried out to overcome such noise. The moving average filter and Gaussian filter of window sizes 3, 5, and 7 were implemented and tested. The recognition accuracy of the classifier after the smoothing with different filters is depicted in Graph 6(a) and 6(b). The horizontal axis represents the resampling size whereas the vertical axis represents the recognition performance. Character identification accuracy was around 50% to 70% for the different resampling sizes for both the smoothing methods.

	10	20	30	40	50	60	70	80	90
■ Window Span 3	51.30	60.87	66.52	66.09	68.26	63.48	63.91	66.09	60.43
▨ Window Span 5	52.61	60.87	63.04	66.52	65.22	63.48	64.35	61.74	61.74
▨ Window Span 7	54.78	62.17	60.00	66.52	63.48	66.52	63.91	63.48	66.52
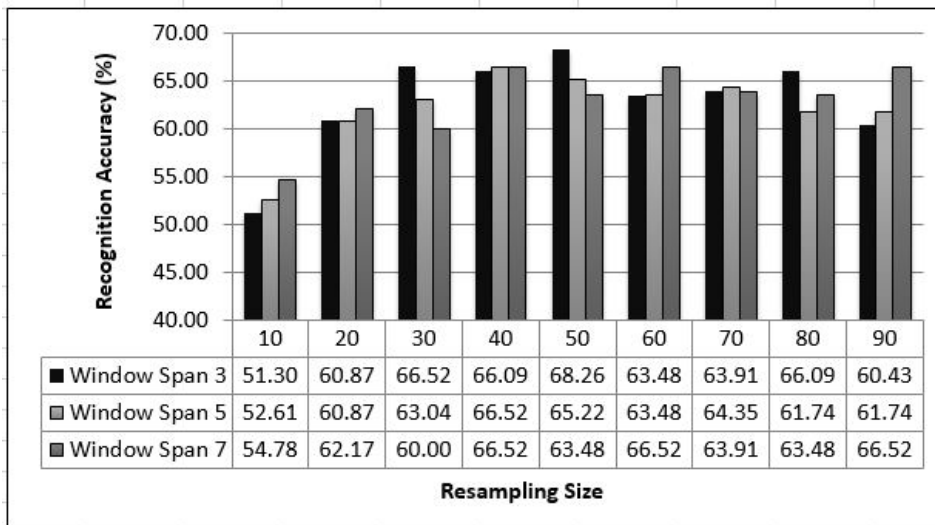
Figure 6(a). Effect of moving average filter with different window spans on recognition accuracy

	10	20	30	40	50	60	70	80	90
■ Window Span 3	50.43	57.83	59.13	62.91	62.61	61.74	59.13	57.83	52.61
▫ Window Span 5	51.30	55.65	62.61	59.57	61.30	61.74	62.61	60.00	60.87
▨ Window Span 7	50.43	55.65	62.61	62.17	63.48	63.91	61.74	63.48	58.70
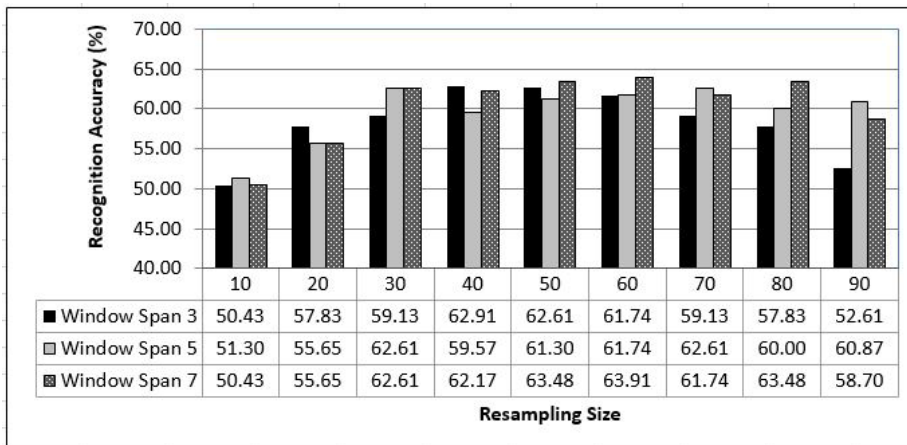
Figure 6(b). Effect of Gaussian filter with different window spans on recognition accuracy

The moving average filter with window size 3 exhibited the maximum recognition rate of around 68.26% for the resampling size of 50, whereas the Gaussian filter responded to a maximum of 63.91% for the window size 7 at the resampling size 60. The recognition rate was higher for the resampling size 40-60 range for both the smoothing filters.

Figure 7 shows the plot of recognition time required for different smoothing techniques. The processing time increases more or less in linear manner with the resampling size. The average time difference between the Gaussian filter (Gauss) with window span (WS) 7 and the moving average (mov Avg) filter with window span 3 response was 3 ms, while the Gaussian filter was faster than the moving average filter.
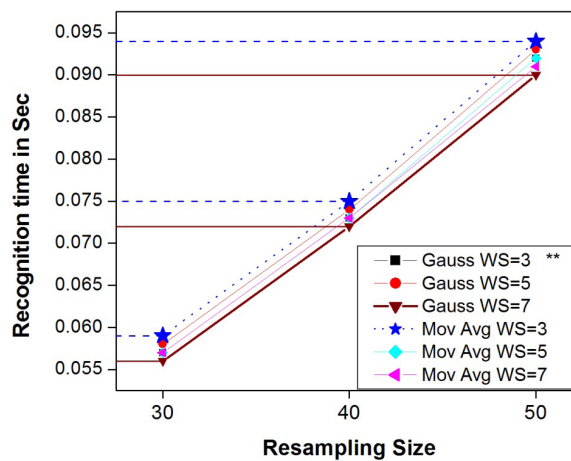
Figure 7. Recognition time variation for the different smoothing techniques
** Gauss - Gaussian filter, WS - Window Span, Mov Avg – Moving Average Filter

Effect of Removal of Duplicate Points and Down Sampling

Due to the speed variation in online handwriting, duplicate points or repetitive points were present in the data samples. Removing the duplicate points removed variation in the shape of the characters, but the length of the strokes was reduced. The number of duplicate points in our data was found to be in the range of 1 to 4 points.
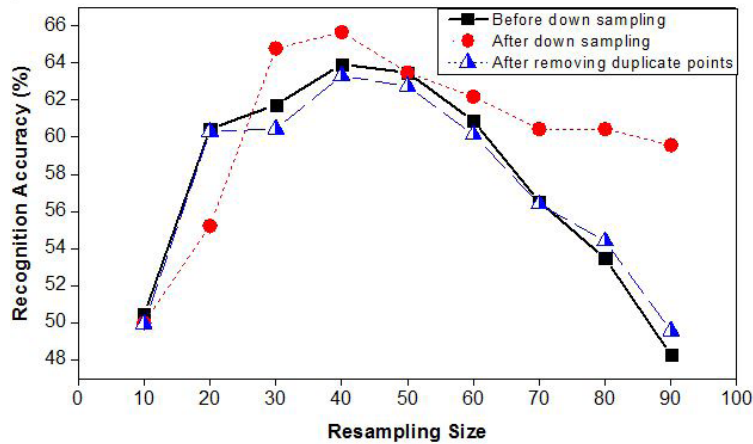
Figure 8(a). Response of recognition accuracy with and without down sampling and duplicate point removal technique
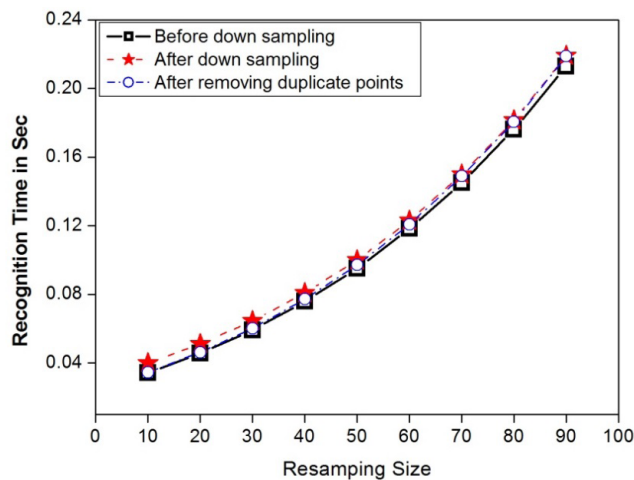
Figure 8(b). Recognition time distribution for down sampling and duplicate point removal technique

Down sampling is a method that reduces the data sample size. Down sampling by two sizes was implemented; hence, every alternative sample which reduced the signal size was selected. Figure 8(a) shows the plot of the recognition accuracy variations, before and after removing duplicate points and down sampling. From the plot, it can be seen that the performance of the

recogniser showed improvement after down sampling. However, there was little change in the recognition accuracy with the duplicate point removal method; this also happened with the processing time plot in Figure 8(b). Hence, there was no significant effect in removing duplicate points on the performance of the classifier. The processing time for down sampling was long because the data had to be down-sampled, followed by normalisation and resampling, which takes time.

Table 2
Results of paired t-test for recognition accuracy of different pre-processing methods

Pre-Processing Methods	Mean	St. Dev	SE Mean	Mean Diff	t-Value	p-Value
Removal of duplicate points	57.25	5.38	1.79	-0.435	-0.56	0.591
Down sampling	60.19	4.93	1.64	2.51	1.61	0.147
Smoothing with Gauss filter of window span 3	58.36	4.44	1.48	0.676	0.77	0.464
Smoothing with Gauss filter of window span 5	59.52	3.74	1.25	1.84	0.97	0.360
Smoothing with Gauss filter of window span 7	60.24	4.55	1.52	2.56	1.49	0.175
Smoothing with Moving Avg. filter of window span 3	63.00	5.10	1.70	5.31	3.50	0.008
Smoothing with Moving Avg. filter of window span 5	62.17	4.01	1.34	4.49	3.08	0.015
Smoothing with Moving Avg. filter of window span 7	63.04	3.80	1.27	5.36	2.66	0.029

To understand the significance of all these pre-processing methods, the paired t-test was applied. The recognition results with and without the pre-processing were compared for 95% confidence. Table 2 represents the results of various pre-processing methods. The following observation can be derived from Table 2. The highest mean of 63.04 was shown for smoothing with the moving average filter of window span 7. However, removal of duplicate points depicted the maximum deviation of 5.38. The t-value was the highest for a moving average filter with a window span of 3, indicating that the effect was significant compared to the other pre-processing methods. The minimum t-value of -0.56 was observed for the removal of duplicate points, signifying it was least significant, which was also indicated by the p value, >0.05. The p-value column indicates that the moving average filter had more significance than the Gaussian filter as the p-value was comparatively low.

Results of Different Testing and Training Methodology

Experiments with five different training-testing methods, namely, single partition random subsampling, tenfold cross-validation with and without stratification and the bootstrap methods were implemented. The classifiers tested for the moving average filter with a span of 3 and the Gaussian filter with the window size 7 because these two methods yielded the highest recognition rate in Test Case 2. The resampling size of 40, 50 and 60 were selected as these resampling sizes exhibited overall better recognition rates in all the above test cases. Table 3 lists the various pre-processing methods that were adopted to estimate the performance of the training and testing methods.

Table 3
List of different pre-processing methods used for the performance estimation of the testing and training methodology

Pre-Processing Method	Smoothing Filter	Window Span	Resampling Size
P1	Moving Average	3	40
P2	Moving Average	3	50
P3	Moving Average	3	60
P4	Gaussian	7	40
P5	Gaussian	7	50
P6	Gaussian	7	60
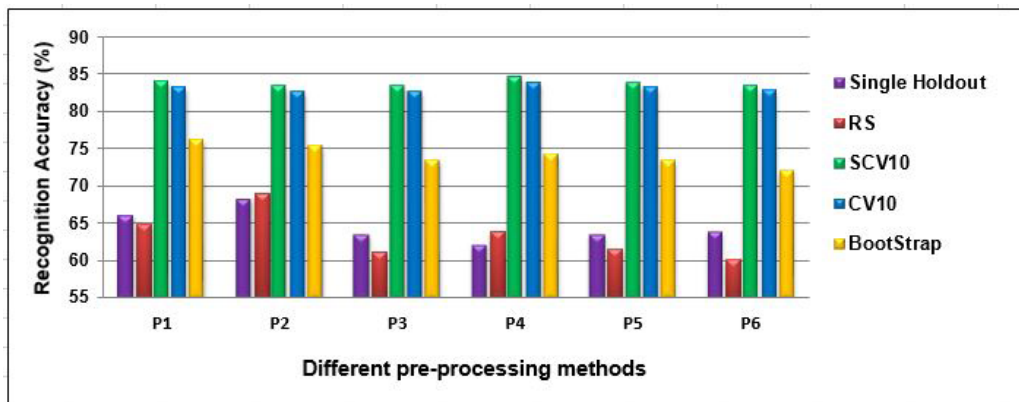
Figure 9. Comparison of recognition accuracy with the different testing and training methods

The performance of the training-testing methods is shown in Figure 9. The stratified tenfold cross validation method (SCV10) showed a better recognition result than the other estimation methods. Random subsampling (RS) showed poor recognition accuracy compared with the

other methods. Bootstrapping performance was better than the single holdout estimation. The computational time of single holdout was faster as it did involve repeated training as with the other methodology. Overall, there was a 20% increase in recognition accuracy using the stratified tenfold cross validation method compared to the single partition method for the same SVM classifier with RBF kernel. However, the recognition accuracy obtained by the training and testing method depends on the classifier used along with features drawn from the dataset.
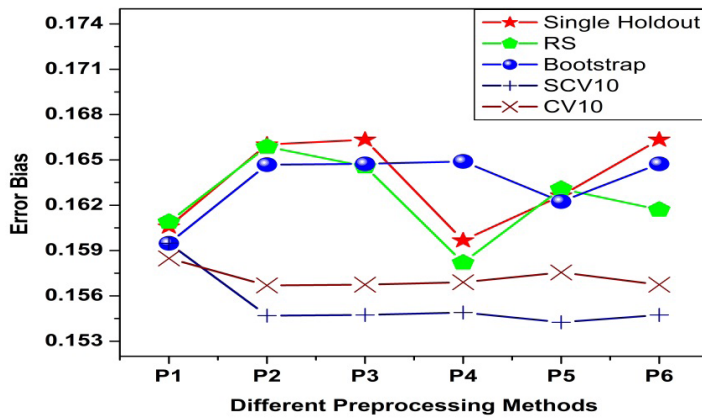
Figure 10(a). Error bias plot for the different training-testing methodologies with the pre-processing methods
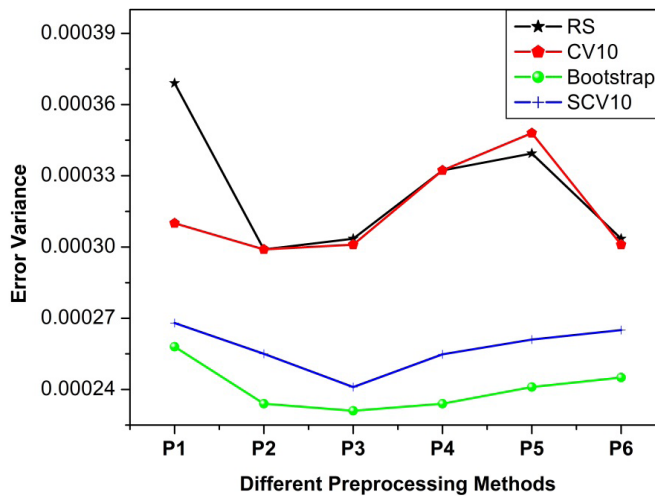
Figure 10(b). Error variance plot for the different training-testing methodologies with the pre-processing methods

The error bias and error variance plots for the different training and testing methods are shown in Graphs 10(a) and (b). The bootstrap method showed greater bias in some of the pre-processing methods. The tenfold cross validation and stratified cross validation methods showed stable bias for the different pre-processing methods compared with the other training-testing methods. Stratification reduced the variance of the cross validation as seen in the Graph 10(b). The random subsampling method had relatively high variance and bias for all the pre-processing methods. The bootstrap method had the least variance.

Our results indicated that cross validation with stratification performed better in terms of recognition, error bias and error variance compared with the tenfold cross validation method. Bootstrap estimation had low variance but showed large bias for some of the pre-processing methods. Overall, stratified cross validation performance is recommended for online Kannada handwriting recognition.

CONCLUSION

A computational analysis was carried out on the online Kannada handwriting recognition system with different pre-processing and testing methodology for a specific classifier, namely SVM, for the feature set consisting of normalised horizontal and vertical coordinates, writing direction and curvature at a point. Based on the results, the following conclusions could be drawn.

Recognition accuracy significantly depends on resampling size. The resampling size of 40 was found to be better for the Kannada character dataset used in this study. The recognition time showed more variation and was less linear with the resampling size.

Smoothing using the moving average and Gaussian filters was implemented. A moving average filter with a window span of 3 and Gaussian filter with a window span of 7 provided better recognition results. The Gaussian filter with window span 7 was 3 ms faster than the moving average filter.

Down sampling the data by a factor of 2 showed improved recognition accuracy. Removal of duplicate points had the least significance in the recognition results.

The stratified tenfold cross validation method performed better compared with the other training-testing methods. The recognition accuracy of the classifier improved from 63% to 83% when compared with the single partition method. However, the recognition accuracy obtained by the training and testing method depends on the classifier used as well as the features drawn from the dataset.

These observations will benefit researchers in choosing appropriate pre-processing and testing methodology for online handwriting recognition. Though these conclusions are drawn based on the performance of a single classifier, a comparative study of several different classifiers will provide a more generalised insight into the effect of pre-processing methods. The recognition accuracy can be improved further by selecting more features and tuning the classifiers.

ACKNOWLEDGEMENT

REFERENCES

Babu, V. J., Prasanth, L., Sharma, R. R., & Bharath, A. (2007). HMM-based online handwriting recognition system for Telugu symbols. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition* (pp. 63–67). IEEE.

Bahlmann, C., Haasdonk, B., & Burkhardt, H. (2002). Online handwriting recognition with support vector machines – A kernel approach. In *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition* (pp. 49–54). IEEE.

Bahlmann, C., & Burkhardt, H. (2004). The writer independent online handwriting recognition system frog on hand and cluster generative statistical dynamic time warping. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(3), 299–310.

Bharath, A., & Madhvanath, S. (2009). Online handwriting recognition for Indic scripts. *Guide to OCR for Indic Scripts* (pp. 209–234). Springer, London.

Biadsy, F., Saabni, R., & El-Sana, J. (2011). Segmentation-Free Online Arabic Handwriting Recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, *25*(07), 1009–1033.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine learning, 20*(3), 273-297.

Deepu, V., Madhvanath, S., & Ramakrishnan, A. G. (2004). Principal component analysis for online handwritten character recognition. In *Proceedings of the 17th International Conference on Pattern Recognition* (pp. 327–330). IEEE.

Ghosh, D., Dube, T., & Shivaprasad, A. (2010). Script recognition – A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(12), 2142–2161.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE, 66*(1), 51-83.

Jaeger, S., Manke, S., Reichert, J., & Waibel, A. (2001). Online handwriting recognition: The NPen ++ recognizer. *International Journal on Document Analysis and Recognition*, *3*(3), 169–180.

Joshi, N., Sita, G., Ramakrishnan, A. G., Deepu, V., & Madhvanath, S. (2005). Machine recognition of online handwritten Devanagari characters. In *Proceeding of the Eighth International Conference on Document Analysis and Recognition* (pp. 1156–1160). IEEE.

Kunte, R. S., & Samuel, R. S. (2007). A simple and efficient optical character recognition system for basic symbols in printed Kannada text. *Sadhana*, *32*(5), 521-533.

Li, X., & Yeung, D. Y. (1997). Online handwritten alphanumeric character recognition using dominant points in strokes. *Pattern Recognition*, *30*(1), 31–44.

Namboodiri, A. M., & Jain, A. K. (2004). Online handwritten script recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(1), 124–130.

Plamondon, R., & Srihari, S. N. (2000). Online and off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 63–84.

Prasad, M. M., Sukumar, M., & Ramakrishnan, A. G. (2009, July). Divide and conquer technique in online handwritten Kannada character recognition. In *Proceedings of the International Workshop on Multilingual OCR* (pp. 11–16). ACM.

Prasad, M. M., Sukumar, M., & Ramakrishnan, A. G. (2010). Orthogonal LDA in PCA transformed subspace. In *Proceedings of the International Conference on Frontiers in Handwriting Recognition (ICFHR)* (pp. 172–175). IEEE.

Prasanth, L., Babu, V. J., Sharma, R. R., & Rao, G. V. P., & Mandalapu, D. (2007 September). Elastic matching of online handwritten Tamil and Telugu scripts using local features. In *Ninth International Conference on Document Analysis and Recognition, 2007* (Vol. 2, pp. 1028-1032). IEEE.

Rampalli, R., & Ramakrishnan, A. G. (2011). Fusion of complementary online and offline strategies for recognition of handwritten Kannada characters. *Journal of Universal Computer Science*, *17*(1), 81–93.

Sin, B. K., & Kim, J. H. (1997). Ligature modeling for online cursive script recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(6), 623–633.

Tappert, C. C., Suen, C. Y., & Wakahara, T. (1990). The state of the art in online handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(8), 787–808.

Toselli, A. H., Pastor, M., & Vidal, E. (2007, June). On-Line handwriting recognition system for Tamil handwritten characters. In *Proceedings of the 3rd Iberian Conference on Pattern Recognition and Image Analysis, Part I* (pp. 370–377). Springer, Berlin, Heidelberg.

Vuori, V., Laaksonen, J., Oja, E., & Kangas, J. (2001). Experiments with adaptation strategies for a prototype-based recognition system for isolated handwritten characters. *International Journal on Document Analysis and Recognition*, *3*(3), 150–159.

Statistical Modelling of Daily Rainfall Variability Patterns in Australia

Bright Emmanuel Owusu[1,2]* and Nittaya McNeil[1]

[1]*Department of Mathematics and computer Science, Faculty of Science and Technology, Prince of Songkla University, Mueang Pattani, 94000, Thailand*

[2]*Department of Information and Communication Technology/Mathematics, Faculty of Science and Technology, Presbyterian University College, Ghana*

ABSTRACT

The present study uses statistical methods, specifically multiple regression (MR) and gamma generalised linear models (Gamma GLM), for the modelling and analysis of period rainfall variability in Australia during the period 1950-2013. The data from 92 observational stations were collected from the Australian Bureau of Meteorology. Factor analysis was first used to group the 92 stations into eight rainfall regions. The consecutive five-day mean rainfall in a year (period) in each region is then modelled using MR and the Gamma GLM. The results show that the models fit the data quite well in all the regions, but the MR model did better than the Gamma GLM in some of the regions. The MR models revealed three rainfall groupings, and each group had diverse rainfall patterns and trends. Significant decreasing annual rainfall trend was revealed in the southwest and the north southeast regions. In contrast, significant increasing annual rainfall trends were found in the north and the northwest regions.

Keywords: Australia, factor analysis, gamma generalised linear models, multiple regression, rainfall

INTRODUCTION

Increasing severe weather conditions such as flood and drought in many areas of the globe have been attributed to climate change (O'Gorman & Schneider, 2009). Intensified extreme rainfall events are not the only concerns of climate change. Other features of rainfall such as annual mean, occurrence and number of events or the time of the year in which they occur and the sequence and duration of these events can also be affected (Kumar, 2013). Consistent water sources and supply are essential, particularly to the increasing population of human society. They can be rigorously affected by long-term changes in rainfall. The fluctuations in the spatial distribution of rainfall will increase

differences between dry and wet areas (Held & Soden, 2006; Allan et al., 2010). Therefore, modelling rainfall variability is essential for the determination of the various types of life in the wet tropical regions and mainly for production in dry areas in some countries.

Numerous authors have studied daily rainfall variability and patterns using various statistical techniques. One of the simplest models considered in the analysis and modelling of rainfall is MR. Mekanik, Imteaz, Gato-Trinidad and Elmahdi (2013) applied MR and Artificial Neural Networks (ANN) to estimate long-term seasonal spring rainfall in Victoria. The ANN models revealed greater correlation relative to the MR models, showing that ANN models explained the pattern and trend of rainfall well compared with the MR models.

In contrast, Jeong, St-Hilaire, Ouarda and Gachon (2012) evaluated MR, robust regression, ridge regression and ANN models to establish suitable transfer functions in statistical downscaling models for daily maximum and minimum temperature and daily precipitation occurrence and amounts. They discovered that the monthly MR, annual ANN and annual MR models performed better than the robust and ridge regression models based on the modified Akaike information criterion. In regression modelling, observation must be independent and normally distributed; this assumption may not be true for most time series data such as rainfall. Rainfall data are highly skewed even after aggregation, and in modelling such data, transformation is normally applied. Meng, Zhang, Dai and Yan (2007) applied a logarithmic transformation on monthly rainfall to attain normality. However, application of this transformation to the summer monsoon rainfall amounts in Asia by Mooley (1973) revealed poor results.

Alternatively, transformation of the data to normality can be overlooked and it can directly be modelled using non-normal distributions. In the modelling of periods of non-zero rainfall amount, some authors have revealed that the Gamma GLM fits the data quite well (Coe & Stern, 1982; Stern & Coe, 1984; Kenabatho et al., 2012). Rainfall data between adjacent observational stations are spatially correlated, and numerous methods can be used to manage this concern. Factor analysis has currently gained attention in climate research to describe the variability of correlated observations with possibly fewer unobserved variables called factors. This method has been used in studies such as analysis of the variability of temperature (McNeil & Chooprateep, 2014; McNeil & Chirtkiatsakul, 2016; Wanishsakpong & McNeil, 2016 ;), solar radiation (Cheung et al., 2015) and rainfall (Wickramagamage, 2010). It was applied to put the period rainfall of the meteorological stations into groups. This study uses MR and Gamma GLM models in describing the variability and patterns of five-daily rainfall observations in Australia during the period 1950 to 2013.

MATERIALS AND METHOD

Daily accumulated rainfall from 92 stations throughout Australia for the period between 1950 and 2013 were acquired from the Australian Bureau of Meteorology. The data were analysed using statistical methods. The period rainfall for each station was computed to minimise the skewness and thus, reduce the number of observations to 73 for each year after removing data for 29 February. Data for leap years were excluded to obtain equal observations for all years. Moreover, the period rainfall was deemed appropriate to use because some of the stations had

data on aggregation for as many as five days. The computed mean was based on the number present in cases where data were missing. After computing the five-day average rainfall, only one of the stations had more than 6% (7.3%) missing data. The remaining stations had mostly less than 1% missing data. However, after finding the five-day average region-wise, only one region had 0.2% missing data, and it was imputed by linear regression in fitting the MR model and GLM in fitting the Gamma GLM models. Factor analysis was initially used to put all the stations into groups, and this was done by conducting maximum likelihood factor analysis on the correlation matrix of the period rainfall for each station. The model with p factors can be written as:

$$y_q = \mu_q + \sum_{m=1}^{p} \lambda_q^{(m)} \phi^{(m)} \qquad [1]$$

Where, μ_q, is the mean over the five days for each station, λ_q is the factor loading and ϕ is the common factor. The factors were recognised via their loading values. Promax, an oblique rotation method, was applied to control the loading (Venables & Ripley, 2002). A station is assigned to a factor if its loading is at least 0.33 (Hair et al., 1998).

The MR and Gamma GLM models were then fit to the non-zero period rainfall observations in the groups revealed by the factor analysis. These methods used the mean period rainfall observations from each region as the dependent variable and the year and period factors as the independent variables in fitting the models. Brief accounts of these methods are given below.

Let μ_{kst} be the mean period rainfall observation for each region k ($k = 1,...8$) at period s and year t, then the MR model in this paper is

$$h(\mu_{kst}) = \mu_{0k} + \sum_{i=1}^{64} \eta_{ik} x_{ti} \sum_{j=1}^{73} \beta_{jk} w_{sj} + h(\mu_{t(s-1)k}) + \varepsilon_{kst} \qquad [2]$$

where, $h(.)$ represents the dependent variable fourth root transformation, which is defined as $h(\mu_{tsk}) = \sqrt[4]{\mu_{tsk}}$, ε_{tsk}, which are random error terms that are normally distributed, having a mean 0 and variance σ^2. x_{ti} is the predictor for relating years t and year i, w_{sj} is the predictor for relating periods s and period j. Thus, x_{ti} becomes 1 if $t = i$ and 0 elsewhere and w_{sj} becomes 1 if $s = j$ and 0 elsewhere. The year 1950 and period 1 are the starting year and period, respectively. For region k, μ_{0k} denotes the mean period rainfall, which is the starting year and period, η_{ik} is the year i effect and β_{jk} the period j effect. Fit values $h(\mu_{kst})$ from the model [2] have to be transformed backwards to give fit values of μ_{kst}. Thus, the backward-transformed value can be written as μ_{kst}. Let $\hat{\mu}$ be the fit value for μ and $h(\mu)$ the above defined transformation. Then $\hat{\mu}$ can be estimated from the fit value $h(\hat{\mu})$ as $\hat{\mu} = (h(\hat{\mu}))^4$.

The probability density function of the Gamma distribution is normally written as:

$$f(x \mid \alpha, \gamma) = \frac{1}{\Gamma(\alpha)} \gamma^{\lambda} x^{\lambda-1} e^{-\gamma x}, x, \alpha, \gamma > 0 \qquad [3]$$

where, α and γ are the shape and the scale parameters, respectively and x is the non-zero period rainfall amount and α is assumed to be constant throughout the dataset for each

region. If μ_s is the mean rainfall amount for period s related to predictors ξ_s then the mean period rainfall amount is estimated by Equation 4, where τ_0 is constant and τ's and ω are slopes.

$$\ln(\mu_s) = \tau_0 + \tau_1\xi_1 + \tau_2\xi_2 + ... + \tau_k\xi_k + \omega\mu_{s-1} \qquad [4]$$

R programming language is used for the data analysis and graphical displays (R Development Core Team, 2015).

RESULTS AND DISCUSSION

The factor model categorised the 92 stations into eight groups (factors) via the factor loading. Table 1 gives the loading of the factors organised in descending order within each factor and the uniqueness values excluding the mixed factors.

Table 1
Factor loading of the 92 stations with the identified dominating factors in bold

Stations	Factor Loading								Uniqueness
	F1	F2	F3	F4	F5	F6	F7	F8	
15085	**0.828**								0.381
15086	**0.810**								0.401
15015	**0.788**								0.464
37043	**0.744**			0.115	0.163	-0.132			0.435
14815	**0.731**	-0.145				0.101		0.109	0.405
15005	**0.728**				0.127	-0.106			0.510
2012	**0.703**				-0.119	0.205			0.400
14902	**0.690**	-0.145			-0.103		0.159	0.105	0.309
29008	**0.620**			0.119	0.116	-0.117	0.196		0.489
38003	**0.602**	0.111		0.192	0.156	-0.146			0.516
15590	**0.584**	0.316		0.106			-0.214	-0.107	0.512
3027	**0.562**	-0.114			-0.118	0.286			0.489
14015	**0.512**	-0.118			-0.219	0.107	0.205	0.104	0.349
38020	**0.347**	0.260				-0.114			0.746
14042	**0.499**	-0.112		-0.122	-0.178	0.107	0.325	0.122	0.280
18012	-0.204	**0.804**	0.148						0.353
16005		0.804							0.461
18002	-0.234	**0.775**				0.106			0.449
16031		**0.746**	-0.126			0.127			0.536
16000		**0.739**	-0.124						0.533
18091	-0.205	**0.712**	0.221						0.386

Table 1 (*continue*)

Stations	Factor Loading								Uniqueness
	F1	F2	F3	F4	F5	F6	F7	F8	
11003	-0.205	**0.655**				0.212			0.616
24003		**0.650**	0.176	-0.107				0.254	0.364
17031	0.198	**0.646**	-0.101						0.520
20017		0.593						0.187	0.566
16047	0.317	0.518					-0.161		0.604
46043		0.434		0.163				0.321	0.495
48115		0.346		0.249				0.259	0.578
98003			0.886						0.188
90015			0.874						0.210
91034	0.107		0.860				-0.102		0.293
91033	0.111		0.772						0.387
90059	-0.103	0.228	**0.755**						0.232
99005			0.752					0.207	0.382
94027		-0.184	**0.714**				-0.124	-0.162	0.557
97006		-0.174	**0.688**		-0.121		-0.126	-0.279	0.549
89002		0.197	**0.661**					0.290	0.263
92018			0.444					0.253	0.728
35029				0.806					0.414
41100				0.803				0.152	0.395
43020				0.792					0.377
54023				0.742			-0.173	0.297	0.373
48031				0.721				0.249	0.357
35065			0.115	**0.682**			0.298	-0.128	0.411
39020				0.580			0.344		0.534
40141	-0.145		-0.124	**0.559**			0.207		0.594
59000		-0.118	-0.254	**0.552**		0.109		0.199	0.520
44181	0.144	0.225		0.380		-0.112		0.168	0.569
10620		-0.169			0.918				0.282
8230		-0.126			**0.849**	0.137			0.333
8004					**0.809**	0.102			0.336
9518					**0.775**	-0.156			0.299
12011					**0.736**	0.265			0.420
9594		0.165			0.554				0.563
7059						0.724			0.511
7103		0.113				0.718			0.514

Bright Emmanuel Owusu and Nittaya McNeil

Table 1 (*continue*)

Stations	Factor Loading								Uniqueness
	F1	F2	F3	F4	F5	F6	F7	F8	
12108		0.171				0.674			0.538
13002		0.183			-0.153	**0.637**			0.588
4006	0.175				-0.156	**0.628**			0.536
7049					0.267	**0.615**			0.526
5001	0.134				-0.116	**0.601**			0.572
5008						0.594			0.630
6029					0.129	**0.554**	0.107		0.651
4019	0.290				-0.101	**0.487**			0.615
33001		0.128		0.152			0.741		0.394
33013				0.284			**0.711**	-0.118	0.401
32001	0.163	0.125	-0.108				0.698		0.353
31062	0.181	0.129	-0.150	-0.201			0.645		0.471
27042	0.328			-0.132	-0.188		0.582		0.308
27005	0.261			-0.139			0.477		0.401
73025		0.146	0.218	0.124				**0.592**	0.352
69018			-0.193	0.268				**0.542**	0.565
84016			0.114	0.116				**0.530**	0.661
80044		0.333	0.329					**0.455**	0.348

The eight factors were made up of the following regions of Australia: Factor 1 (F1), the north (N), Factor 2, (F2) the central south (CS), Factor 3, (F3) the south southeast (SSE), Factor 4, (F4) the central east (CE), Factor 5, (F5) the southwest (SW), Factor 6, (F6) the northwest (NW), Factor 7, the northeast (NE) and Factor 8, (F8) the north southeast (NSE). Figure 1 shows the stations as grouped by factor analysis. The classification reveals the spatial rainfall distribution of the 92 stations used in the study. A similar classification was uncovered by factor analysis (Cheung et al., 2015) in the analysis of spatial and temporal patterns of solar absorption by clouds and the modelling of daily maximum temperature in Australia by Wanishsakpong and McNeil (2016).

Further concern for the modelling of time series observation is the dependency on response observation, which contravenes independent error assumption. Figure 2 is the autocorrelation function (ACF) that is used to evaluate the autocorrelation of mean period rainfall. Some of the sample lag values (the black vertical lines) are beyond the 95% confidence interval line (the horizontal dotted lines), which indicates significant dependency among the response variables. The dependency among the period rainfall in each region was minimised by adding the AR(1) term (Chatfield, 1996) to both the MR and Gamma GLM models (Figure 3). Most of the sample lag values are within the 95% confidence interval, indicating a drastic reduction of serial autocorrelation in the models. Also, multicollinearity was assessed using the variance inflation factor (VIF), which is defined as $1 / (1-R^2)$, where

R^2 is the coefficient of determination. The response variables have been noted to have multicollinearity if VIF >10 (Lin, 2008). The values of the VIF were below 5 in all the regions, which revealed no indication of multicollinearity.

The pattern of period rainfall amount for all the eight regions as modelled by Equations 2 and 4 are shown in Figure 4. Both models showed an obvious seasonal periodic pattern, and this is clearly seen especially in the N and NE regions, where over 52% of the variations were explained by the models (coefficient of determination of 0.54 and 0.52, respectively). Moreover, the fluctuations and their timing in the data were well displayed by the models as well as the 64 years' mean. However, the models underestimated the magnitude of the fluctuations. Relatively, there is not much difference between the models from both methods in the remaining regions as shown in Figure 4. The eight factor regions display diverse rainfall variability and trends for all the fit models.
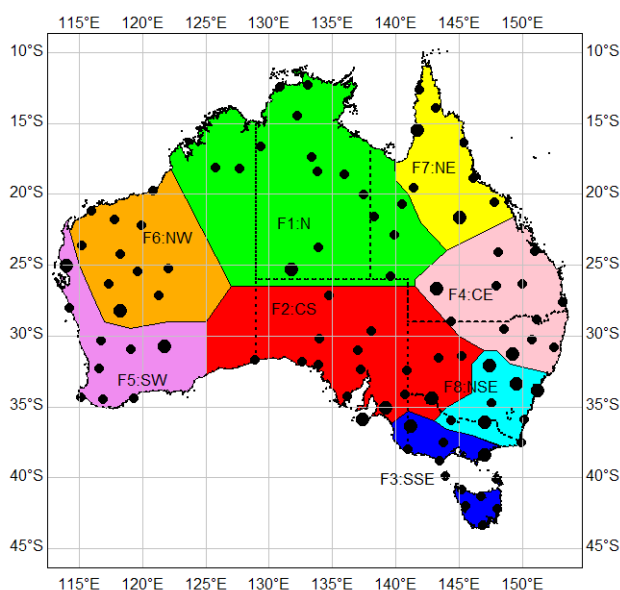
Figure 1. The locations of the stations used in the study and the groups (colours) revealed by factor analysis. The large points represent mixed factors. The state boundaries are also included (dotted lines)

The N and NE, which is in the tropical region of Australia, have similar period rainfall variation as in the NW and the CE. The pattern in CS and NSE was also similar to that in the SW and SSE. In general, the NE region relative to the other regions had the highest amount of period rainfall in the seasonal span, which can be due to a substantial number of temporal meteorological conditions such as fronts and low pressure systems, which possibly can disturb rainfall in this area. The area is additionally more susceptible to severe weather systems such as tropical cyclones and a great deal of convective clouds entrenched, resulting in greater proportion of cloud cover relative to other regions (Cheung et al., 2015).

The models were assessed by plotting residuals alongside the theoretical quantiles (Figure 5). The figure shows that the residual plots had low variations with respect to the expected line of best fit, indicating that both models fit the data quite well in the factor regions except for
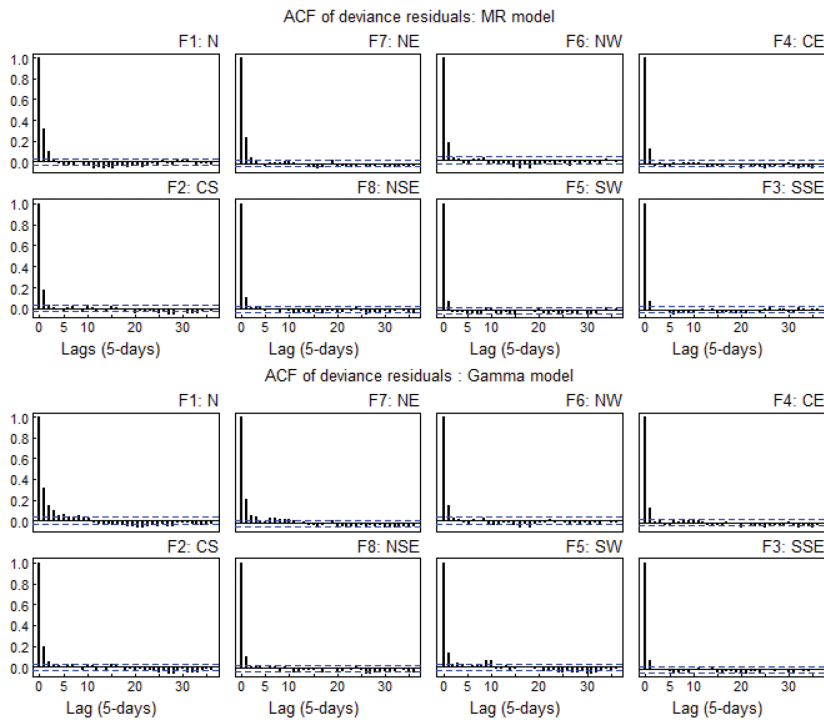
Figure 2. The autocorrelation function for the models. The top and bottom panels relate to before minimising serial correlation
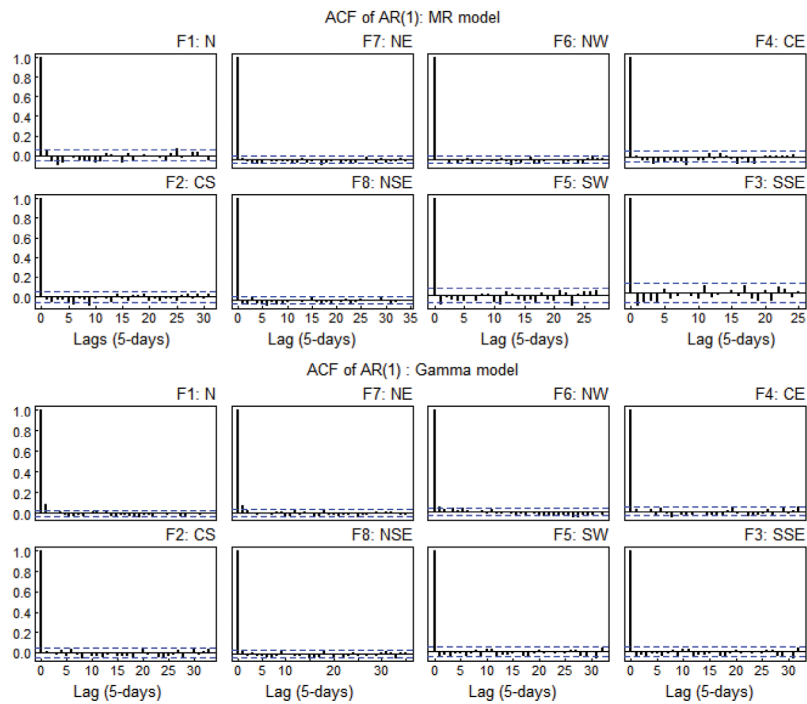
Figure 3. The autocorrelation function for the models. The top and bottom panels relate to after minimising serial correlation

little departures at the upper and lower tails of some of the models. The MR model seemed to perform better than the Gamma GLM model in the N, NE and SW regions. Analysis of the deviance residuals and the fit values of the models did not show any patterns and outliers.

Further, the MR was used to model the 64 years' mean period to show the monthly and annual rainfall patterns for all the regions. This model was applied because it is one of the simplest methods to describe monthly rainfall patterns (Hughes & Saunders, 2002; Oettli & Camberlin, 2005). It also fit the period mean rainfall better than the Gamma GLM models in most of the regions.
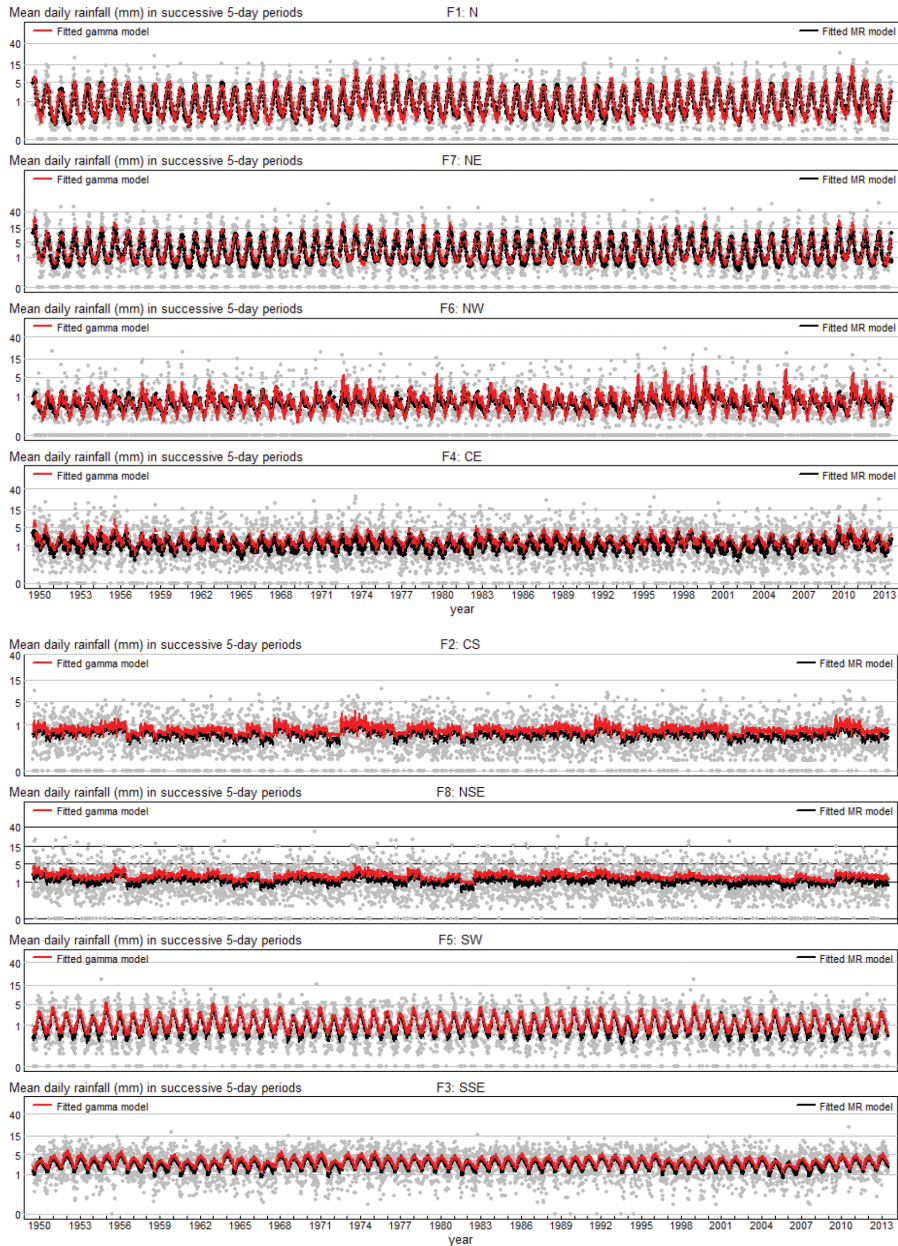
Figure 4. The plots of period rainfall and their estimated mean patterns in eight regions during 1950-2013. The grey dots represent the observed period's rainfall, while the black and red curves represent their estimated mean from models 2 and 4, respectively
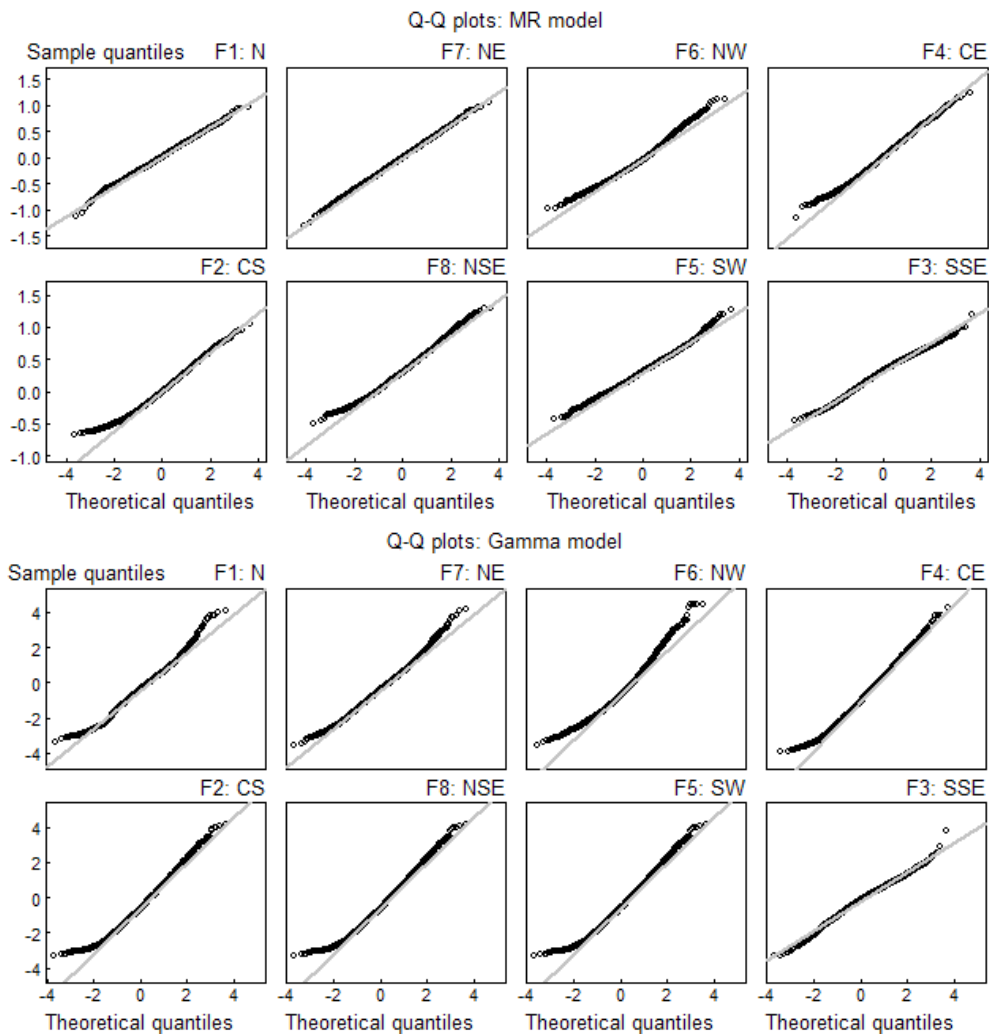
Figure 5. The residual quantile-quantile (Q-Q) plots for the models. The top graph shows that of the MR (response variable is fourth root transformed) and the bottom shows the Gamma. Both models were filtered with AR(1) term

Figure 6 shows the results of the estimated rainfall variations for the 64 years for all the regions. Analysis of the graph indicated three main categories regarding the period's rainfall estimates. The models in the N, NE, CE and NW regions can be grouped as one category (first group) of items possessing similar patterns, while the models in the CS and NSE can be grouped together (second group).
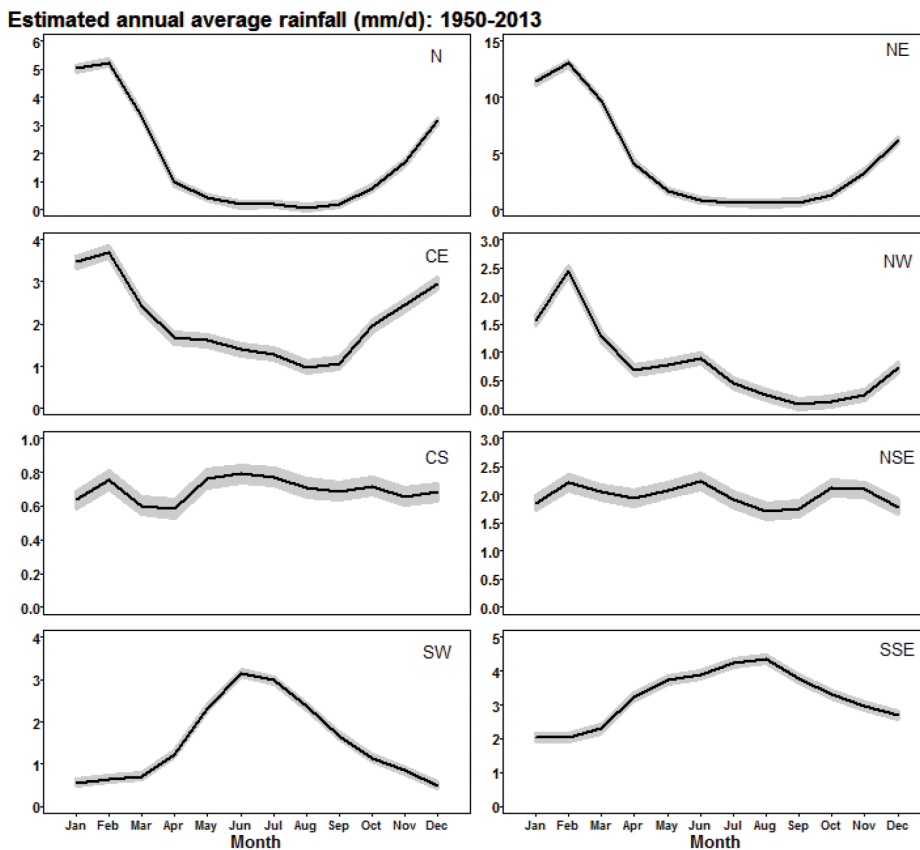
Figure 6. The estimated monthly rainfall (line) and the 95% confidence interval (shaded) in each month of a year in the factor regions

The SW and SSE also form a group (final group) with similar patterns. In the first group, the period rainfall rises sharply from January to March, reaches its maximum between this period, then decreases steadily to its minimum between August and September. The period rainfall observed in the second group is quite stable throughout the year (it varies from 0.6 to 2.2 mm d^{-1}). The largest estimated period rainfall in these regions occurred between May and July, but the minimum was observed during March and April in the CS and occurred between August and October in the NSE. On the other hand, the final group received their minimum period rainfall between January and February. The period rainfall increased steadily from February until it attained the largest between June and August. Generally, the estimated period rainfall is different in each region: the NE, which comprises equatorial and tropical regions, has the highest in the Austral summer during the active part of the monsoon, while the mid-latitude region of SSE has the most during the Austral winter.

Figure 7 shows the results of the estimated inter-annual rainfall pattern. In all eight regions, these patterns fluctuated between 0.3 and 7.1 mm d^{-1} and attained the lowest value between 1950 and 1975 in the NW. However, the highest value occurred during 1950 and 1960 in the

NE. In general, the highest annual rainfall values in most of the regions also occurred between 1950 and 1970 except in the N, CS and NW regions, where it occurred between 1970 and 2012. The estimated annual rainfall amount was low, especially in CS and the NW relative to the remaining regions.
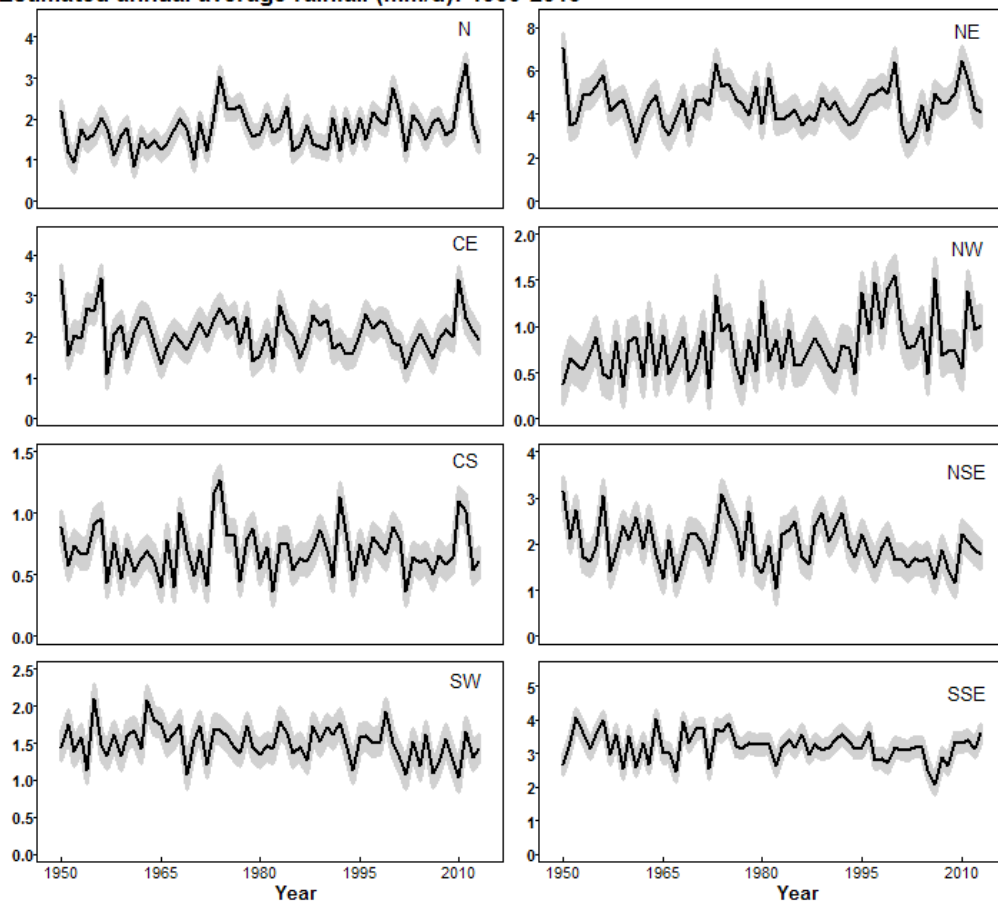
Figure 7. The estimated annual rainfall (line) and the 95% confidence interval (shaded) for factor regions from 1950-2013

Examination of the 64 years' annual rainfall time series by linear regression indicated negative and positive slopes for the regions (Table 2). Analysis of Table 2 revealed that the CE, CS, NSE, SW and SSE regions were characterised by negative slopes, which indicated decreasing trends. Significant decreasing trends were evident in the NSE and SW regions. The IOCI (2002) also reported long-term decline in winter (May-October) rainfall of 15-20% since the 1970s in the southwestern corner of Western Australia. Average spring rainfall during 1997-2006 in the southeast of Australia was revealed to be below average and resulted in the

problem of recurring drought in this part of Australia (Murphy & Timbal, 2008). Hope et al. (2010) revealed similar results in the analysis of temporal relationship concerning rainfall variability in the SW and southeast of Australia.

Timbal and Fawcett (2013) also observed unique shortfalls in rainfall at southeastern Australia. The reported rainfall deficit was evident in pre-winter and early winter rainfall. These declining trends may be partly due to changes in large-scale atmospheric circulation (Nicholls, 2006; Timbal & Fawcett, 2013) including a pole-ward movement of the westerly winds and increasing atmospheric surface pressure. Changes in anthropogenic greenhouse gases and ozone levels (Delworth & Zeng, 2014) and expansion of the Southern Hemisphere Hadley cell (Post et al., 2014) can affect decreasing rainfall trends in these regions.

Table 2
Use of linear regression for analysis of annual rainfall trends

Region	N	NE	CE	NW	CS	NSE	SW	SSE
Slope	0.0076	0.0065	-0.0029	0.0001	-0.0001	-0.0075	-0.0035	-0.0050
p-value	0.0302	0.8757	0.4725	0.0026	0.7898	0.0235	0.0444	0.0964

In contrast, significant increasing annual rainfall trends were evident in the N and NW regions of Australia, while a non-significant increasing trend at 95% confidence level was evident in the NE region (Table 2). Unlike the N and the NW regions, the NE region is mostly liable to intense weather conditions, for instance, tropical cyclones that bring a convergence of convective clouds. This explains the greater part of cloud cover in the N and NW regions (Cheung et al., 2015). It could also explain part of the increasing rainfall trend in this region since cloud formation is related to rainfall.

CONCLUSION

This study used statistical models in describing the period rainfall in Australia during 1950-2013. It is based on 64 years' daily accumulated observation obtained from 92 stations distributed over Australia. Factor analyses identified eight factors, which corresponded to eight geographical rainfall regions. The eight regions include the north, northeast, central east, northwest, central south, north southeast, southwest and south southeast. The rainfall from each region exhibits diverse overall means with various seasonal variations and were explored using Gamma generalised linear and multiple regression models.

These methods were applied to model the period rainfall obtained from each of the regions. The period rainfall observations in successive periods were serially correlated, and this was minimised using the AR(1) technique in fitting the models. Both models fit the data quite well in all the regions, particularly in the tropical region (lower latitudes) where the atmospheric internal variability is small relative to the forced change. However, the multiple regression models did better than the Gamma generalised linear model in the north, northeast and the southwest regions.

The estimated period and annual rainfall means by the multiple regression models revealed three different groups, and each of the groups showed diverse monthly or annual patterns. In general, the eight identified regions of the factor analysis showed three main periods of rainfall category, with each having a distinctive dominating pattern. The understanding and knowledge of these patterns are essential for planning the rainfall resources in numerous regions of Australia. The northeast received the highest annual rainfall followed by the south southeast. In contrast, central south and the northwest received low rainfall with a similar pattern during this period. The decreasing annual rainfall observed in the southern regions may be due to a reduction in Austral autumn and winter rainfall over the southern parts of Australia, particularly South Western Australia.

ACKNOWLEDGEMENT

REFERENCES

Allan, R. P., Soden, B. J., John, V. O., Ingram, W., & Good, P. (2010). Current changes in tropical precipitation. *Environmental Research Letters*, *5*(2), 1–7.

Chatfield, C. (1996). *The analysis of time series*. Melbourne, Australia: Chapman & Hall.

Cheung, K. K., Chooprateep, S., & Ma, J. (2015). Spatial and temporal patterns of solar absorption by clouds in Australia as revealed by exploratory factor analysis. *Solar Energy, 111*, 53–67.

Coe, R., & Stern, R. D. (1982). Fitting models to daily rainfall data. *Journal of Applied Meteorology*, *21*(7), 1024–1031.

Delworth, T. L., & Zeng, F. (2014). Regional rainfall decline in Australia attributed to anthropogenic greenhouse gases and ozone levels. *Nature Geoscience*, *7*(8), 583–587.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis* (5th ed.). New York, NY: Prentice Hall International.

Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of Climate, 19*(21), 5686–5699.

IOCI. (2002). Climate variability and change in south west Western Australia. *Indian Ocean Climate Initiative Panel.* Perth.

Hope, P., Timbal, B., & Fawcett, R. (2010). Associations between rainfall variability in the southwest and southeast of Australia and their evolution through time. *International Journal of Climatology*, *30*(9), 1360–1371.

Jeong, D. I., St-Hilaire, A., Ouarda, T. B. M. J., & Gachon, P. (2012). Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over Canada. *Stochastic Environmental Research and Risk Assessment*, *26*(5), 633–653.

Kenabatho, P. K., McIntyre, N. R., Chandler, R. E., & Wheater, H. S. (2012). Stochastic simulation of rainfall in the semi-arid Limpopo basin, Botswana. *International Journal of Climatology*, *32*(7), 1113–1127.

Kumar, P. (2013). Hydrology: Seasonal rain changes. *Nature Climate Change*, *3*(9), 783–784.

Lin, F. J. (2008). Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Quality and Quantity*, *42*(3), 417–426.

Lloyd-Hughes, B., & Saunders, M. A. (2002). Seasonal prediction of European spring precipitation from El Niño-Southern oscillation and local sea-surface temperatures. *International Journal of Climatology, 22*(1), 1–14.

McNeil, N., & Chirtkiatsakul, B. (2016). Statistical models for the pattern of sea surface temperature in the North Atlantic during 1973-2008. *International Journal of Climatology, 36*(11), 3856-3863.

McNeil, N., & Chooprateep, S. (2014). Modeling sea surface temperatures of the North Atlantic Ocean. *Theoretical and Applied Climatology*, *116*(1–2), 11–17.

Mekanik, F., Imteaz, M. A., Gato-Trinidad, S., & Elmahdi, A. (2013). Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology, 503*, 11–21.

Meng, W. G., Zhang, Y. X., Dai, G. F., & Yan, J. H. (2007). The formation and development of a heavy rainfall mesoscale convective system along southern China coastal area. *Journal of Tropical Meteorology*, *23*(6), 521-529.

Mooley, D. A. (1973). Gamma distribution probability model for Asian summer monsoon monthly rainfall. *Monthly Weather Review*, *101*(2), 160–176.

Murphy, B. F., & Timbal, B. (2008). A review of recent climate variability and climate change in southeastern Australia. *International Journal of Climatology, 28*(7), 859–879.

Nicholls, N. (2006). Detecting and attributing Australian climate change: A review. *Australian Meteorological Magazine*, *55*(3), 199–211.

Oettli, P., & Camberlin, P. (2005). Influence of topography on monthly rainfall distribution over East Africa. *Climate Research, 28*(3), 199–212.

O'Gorman, P. A., & Schneider, T. (2009). The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences*, *106*(35), 14773–14777.

Post, D. A., Timbal, B., Chiew, F. H., Hendon, H. H., Nguyen, H., & Moran, R. (2014). Decrease in southeastern Australian water availability linked to ongoing Hadley cell expansion. *Earth's Future, 2*(4), 231–238.

R Development Core Team. (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org

Stern, R. D., & Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society. Series A (General), 147*(1), 1–34.

Timbal, B., Arblaster, J., Braganza, K., Fernandez, E., Hendon, H., Murphy, B., & Wheeler, M. (2010). *Understanding the anthropogenic nature of the observed rainfall decline across South Eastern Australia*. Centre for Australian Weather and Climate Research.

Timbal, B., & Fawcett, R. (2013). A historical perspective on southeastern Australian rainfall since 1865 using the instrumental record. *Journal of Climate*, *26*(4), 1112–1129.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th Ed.). Queensland: Springer.

Wanishsakpong, W., & McNeil, N. (2016). Modelling of daily maximum temperatures over Australia from 1970 to 2012. *Meteorological Applications*, *23*(1), 115–122.

Wickramagamage, P. (2010). Seasonality and spatial pattern of rainfall of Sri Lanka: Exploratory factor analysis. *International Journal of Climatology*, *30*(8), 1235–1245.

Simulation and Optimisation of Bioethanol Purification using Extractive Distillation with Additive Solvent

S. M. Anisuzzaman[1,2]*, D. Krishnaiah[2], A. Bono[2], F. A. Lahin[2], E. Suali[2] and I. A. Z. Zuyyin[2]

[1]*Energy Research Unit (ERU), Universiti Malaysia Sabah, 88400 UMS, Kota Kinabalu, Sabah, Malaysia*

[2]*Chemical Engineering Programme, Faculty of Engineering, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia*

ABSTRACT

In this study, simulation and optimisation of the purification of bioethanol from an azeotropic mixture was done using the Aspen HYSYS and the Response Surface Methodology (RSM), respectively, to achieve an acceptable bioethanol content with minimal energy use. The objective of this study is to develop the simulation process of bioethanol production from a fermentation effluent. Additionally, the effects of parameters such as solvent temperature, number of entrainer feed stage, mass flow rate and third components of the process for production of bioethanol were studied. As bioethanol is a product of biofuel production, the main challenge facing bioethanol production is the separation of high purity ethanol. However, the separation of ethanol and water can be achieved with the addition of a suitable solvent such as 1,3-butylene glycol (13C4Diol), mixture 13C4Diol and ethylene glycol (EGlycol) and mixture 13C4Diol and glycol ethyl ether (DEG) in the extractive distillation process. For the 13C4Diol mixture, the temperature of entrainer is 90°C with 1500 kg/hr of entrainer rate, while the number of entrainer feed stage is one. The optimum conditions for mixture 13C4Diol and EGlycol require a temperature of entrainer of 90.77°C with an entrainer rate of 1500 kg/hr, while the number of entrainer feed stage is one. Lastly, for optimum conditions for the mixture 13C4Diol and DEG, the temperature of entrainer should be 90°C with an entrainer rate of 1564.04 kg/hr, while the number of entrainer feed stage is one. This study shows that process simulation and optimisation can enhance the removal of water from an azeotropic mixture.

INTRODUCTION

The high demand of fuel by a population is predicted to increase by 25% in the next 20 years, with most of the growth in countries

with emerging economies. Due to this significant energy demand, alternative energy sources are required to sustain future needs. Biofuel is one such alternative. Bioethanol is a biofuel that is commonly used nowadays. According to Szulczyk et al. (2010), worldwide bioethanol demand has grown rapidly due to government mandates such as the Energy Independence and Security Act of 2007 and environmental regulations forbidding the use of *methyl tert-butyl ether* (MTBE) as fuel oxygenate. Bioethanol production represented 4% of total gasoline consumed around the world (2007), and worldwide production may increase to 125 billion litres by 2020 (Balat & Balat, 2009; Flatch et al., 2015).

Distillation is the most common and recognised industrial purification technique for ethanol. Extractive distillation is the partial vaporisation process that occurs in the presence of a miscible entrainer that alters the relative volatilities of the components present in the mixture to be separated (García-Herreros & Gómez, 2011). Extractive distillation is used more often than azeotropic distillation. It is because extractive distillation produces low energy consumption and flexible selection of solvent. Extractive distillation of bioethanol purification is used with ethylene glycol or tetraethylene glycol as entrainer. Ravagnani et al. (2010) pointed out the toxicity of ethylene glycol. However, reports indicated that tetraethylene glycol as entainer had high energy consumption. Gil et al. (2012) reported a design of the extractive distillation process to produce ethanol using glycerol as entrainer. This study hoped to show that glycerol can be used in the production of high purity ethanol by taking advantage of its low cost and high availability. According to Segovia-Hernandez et al. (2014), the analysis of the study reported that the theoretical control properties of complex distillation sequences can be improved by using glycerol as entrainer. Besides that, Bauer and Hulteberg (2013) claimed that the use of glycerol as entractive agent can be increased for the foreseeable future due to the availability at low cost of this chemical compound as co-product of bioethanol production. Zhao et al. (2006) reported the use of several ionic liquids, including 1-butyl-3-methylimidazolium bromide and 1-butyl-3-methylimidazolium chloride for the separation of the mixture, ethanol-water. This is because ionic liquids are considered green solvents that have more advantages, for example, low vapour pressure, low toxicity and high decomposition temperature. Ionic liquids are better in the distillation process when using high concentrations, but ionic liquids are very expensive in contrast to ethylene. In the literature, simulation tools such as Aspen Plus® simulator version 11.1 and Aspen Hysis were used for bioethanol separation (Gil et al., 2008; Pla-Franco et al., 2014)

The aim of this study was to optimise bioethanol purification via additive solvent in a multicomponent distillation system. The third additive that was used to optimise bioethanol production was a mixture of 1,3-butylene glycol (13C4Diol), mixture 13C4Diol and ethylene glycol (EGlycol) and mixture 13C4Diol and glycol ethyl ether (DEG). The process was designed using Aspen HYSYS to study the effect of the entrainer in distillation and to investigate the effect of parameters to optimise the purification of bioethanol.

MATERIALS AND METHOD

Materials

In the extractive distillation process, 13C4Diol, mixture 13C4Diol and EGlycol, mixture 13C4Diol and DEG were used as solvents. Tables 1 and 2 show the process parameters and main components of the fermentor effluent, respectively.

Table 1
Process parameters during the ethanol dehydration experiments

Parameter	Unit	Variation range
Temperature	ºC	90-110
Pressure	kPa	101.3
Mass flow rate	kg/hr	1500-3400
Number of stages	tray	1-20
Product ethanol	wt%	80-100

Source: (Niemistö et al., 2013)

Table 2
Main components in fermenter effluent

Component	Composition, %
Ethanol	10
Water	83
Sucrose	4
Carbon	3

Source: (Batista et al., 2012)

This mixture was selected based on the literature, in which most of the components were yeast, sucrose, ethanol and water (Habaki et al., 2015; Langston et al., 2005; Navarrete-Contreras et al., 2014). Theoretically, carbohydrate can be converted to 75-95% with about 10-16% of the final ethanol concentration entering into the distillation column. By taking the composition of dissolved carbohydrate from molasses and 10% of ethanol production, presumably, the water content would be 83% and 25% and unconverted sugar consisting of sucrose, glucose and fructose would be 4%. The remaining percentage was assumed to be the amount of yeast. The components were set according to their name in HYSYS except for yeast, which was named as carbon. Carbon serves the same function as yeast that would settle in the hydrocyclone.

HYSYS Simulation

The suggested process consisted of the removal of solids using hydrocyclone, separation of non-volatile material using a separator and distillation using a distillation column.

Hydrocyclone. The fermentation state contains a carbon particle, which was removed using hydrocyclone. The fermentation component was fed into the hydrocyclone in order to remove the solid carbon particle as solid particles cannot remain in extractive distillation. The outlet of the upper stream was in liquid form and the outlet of the bottom stream was the solid carbon particle.

Separator. The sucrose had to be separated from ethanol and water to ensure the impurities would not affect the final product. The component was fed into a two-phase separator to separate the liquid and vapour components in the fermented mixture. The inlet stream of the fermented mixture passed through the valve and heater to maintain the pressure drop and increase the temperature of the fermented mixture before entering the two-phase separator and the fermented mixture consisted of vapour and liquid. The components in the fermented stream evaporated due to the processing and was released together with the vapour stream. The fermented mixture was collected at the bottom stream.

Distillation column. In order to achieve a high concentration of ethanol, an ethanol-water mixture was added to the purification process during the extractive distillation stage. The third chemical component was also added in this process. Three units of the distillation column were used and simulated. The first distillation column was used to separate the water and solvent from the ethanol. However, since there was still a small amount of water and solvent, the second and third distillation columns were used in order to achieve ethanol concentration higher than 99% mol. For the second and third columns, another solvent stream was added. The temperature of the solvent that was produced at the bottom of the distillation column was reduced using three units of coolers and then mixed in order to recycle it back to the fed solvent stream.

Response Surface Methodology (RSM)

The HYSYS design simulation data was used for fitting the model to find the best polynomial equation. This data was analysed using Design Expert version 7.0.0. The three main analytical steps were analysis of variance (ANOVA), a regression analysis and the plotting of a response surface. These steps were performed to find an optimal condition for the yield of bioethanol production. The experiment data that was obtained using the optimal conditions established from the mathematical model developed were used as the validating set and these were compared with the predicted values. The fitted quadratic response model is given as:

$$Y = b_0 + \sum_{i=1}^{k} b_i X_i + \sum_{i=1}^{k} b_{ii} X_i^2 + \sum_{i<j}^{k} b_{ij} X_i X_j + e \tag{1}$$

where, Y = response variable, which is yield of bioethanol, b_0 = is the intercept value, b_i ($i=1,2\ldots k$) is the first-order model coefficient, b_{ij} = the interaction effect, b_{ii} = the quadratic coefficients of X_i, X_i and X_j = the input variable that influenced the response variable, e = the random error. An effect that exceeded the vertical line ($p=0.05$) may be considered significant.

RESULTS AND DISCUSSION

Data Output

Table 3, 4 and 5 show the removal of solids, non-volatile compounds and operating condition composition in the fractional distillation process, respectively.

Table 3 shows that the amount of solid particles removed from the fermenter effluent was 4% (w/w). It was completely removed from the mixture because the other equipment would not have been able to run if there had been solid particles.

Table 3
Removal of solids

Stream	Temperature, °C	Pressure, kPa	Mass Flow Rate, kg/h	Composition
Feed	25	101.3	1200	0.03
Out 1	25	101.2	5527	0.00
Out 2	25	101.2	2763	0.04

Table 4 shows that the amount of non-volatile compound decreased from 4 to 13% (w/w). Before entering the two-phase separator, the fermenter was heated to 100°C to form a two-phase mixture. The vapour phases consisted of an azeotropic mixture, while the liquid phase consisted of a sucrose. Since sucrose is a non-volatile compound, it cannot enter the distillation process. Therefore, it had to be removed completely.

Table 4
Non-volatile compound

Stream	Temperature, °C	Pressure, kPa	Mass Flow Rate, kg/h	Composition
Feed	100	101.2	5527	0.04
Out 1	100	101.2	2425	0.00
Out 2	100	101.2	3102	0.13

Table 5 shows the results of the fractional distillation process; the amount of water and ethanol was 86% (w/w) and 14% (w/w), respectively. In this process, the most volatile component, ethanol, was concentrated to a greater degree in the vapour, left in the liquid, while the water flows downward through the column as the bottom product of the column.

Table 5
Operating condition and composition in fractional distillation process

Stream	Temperature, °C	Pressure, kPa	Mass Flow Rate, kg/h	Composition	
				Ethanol	Water
Feed	78.1	101.2	2425	0.14	0.86
Out 1	78.61	99.3	15.7	0.60	0.40
Out 2	84.27	100.3	2409	0.14	0.86

Optimisation

The flowsheet for the distillation sequence simulated in Aspen Plus is shown in Figure 1. This distillation contains two distillation columns, the extractive distillation column and the entrainer recovery column. In this process, the bottom product from the extractive distillation column contained ethanol, with $p<0.05$ (excluding separating agent). The entrainer recovered high purity ethanol through the second column. The recovery column separated the azeotropic mixture from the entrainer and recycled it back to the initial feed.
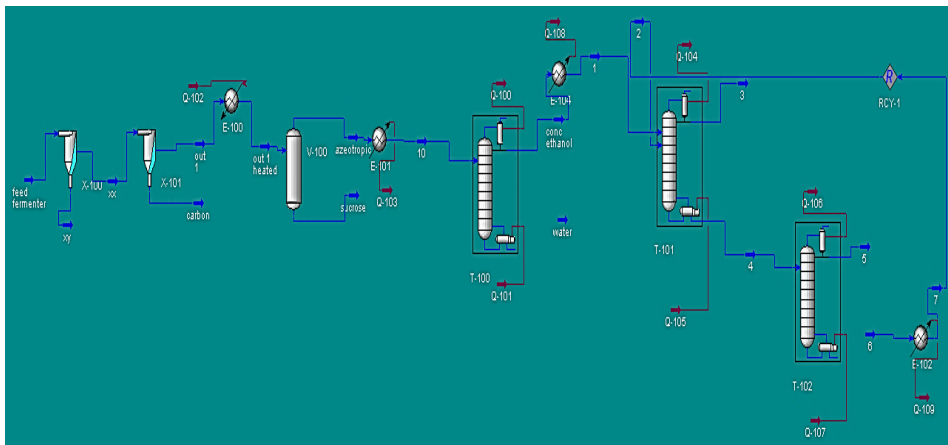
Figure 1. Hysys process simulation

Table 6 shows the material balance for the three separating agents that were used for bioethanol production. The three separating agents underwent the same process flow but gave a different balance. This is because their properties were not the same. Therefore, they adjusted their condition to produce more bioethanol and were efficient in water removal.

Table 6

Material balance for extractive distillation process

Entrainer: 13C4Diol							
Stream	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Temperature, °C	25.00	90.00	78.20	176.50	84.80	182.20	90.00
Pressure, kPa	99.30	101.30	99.00	99.10	101.30	101.30	101.30
Ethanol, kmol/hr	0.60	0.01	0.86	0.02	0.43	0.01	0.01
Water, kmol/hr	0.40	0.03	0.12	0.03	0.31	0.03	0.03
13C4Diol	-	0.96	0.03	0.95	0.18	0.96	0.96
Mass Flow Rate, kg/hr	15.70	3039	5.00	3050	17.26	3032	3032
Entrainer: 13C4Diol and EGlycol							
Stream	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Temperature, °C	25.00	90.00	78.23	178.00	85.94	188.50	90.00
Pressure, kPa	99.30	101.30	99.00	99.10	101.30	101.30	101.30
Ethanol, kmol/hr	0.60	0.06	0.87	0.01	0.38	0.01	0.01
Water, kmol/hr	0.40	0.02	0.10	0.03	0.41	0.02	0.02
13C4Diol, kmol/hr	-	0.71	0.02	0.70	0.11	0.71	0.71
Eglycol, kmol/hr	-	0.26	0.06	0.26	0.10	0.26	0.26
Mass Flow Rate, kg/hr	15.70	1564	6.00	1574	17.24	1556	1556

Table 6 (*continue*)

Stream	1	2	3	4	5	6	7
Temperature, °C	25.00	90.77	77.74	202.60	84.52	241.80	90.77
Pressure, kPa	99.30	101.30	99.00	99.10	101.30	101.30	101.30
Ethanol, kmol/hr	0.60	-	0.89	0.01	0.42	-	-
Water, kmol/hr	0.40	-	0.10	0.01	0.42	-	-
13C4Diol, kmol/hr	-	0.04	0.00	0.04	0.04	0.04	0.04
DEG, kmol/hr	-	0.96	0.01	0.94	0.12	0.96	0.96
Mass Flow Rate, kg/hr	15.70	1500	5	1511	17.30	1493	1493

Entrainer: 13C4Diol and DEG (table header)

Effect of Separating Agent

To analyse the effect of the separating agent in water removal for an azeotropic mixture, the process of extractive distillation was simulated in Aspen Properties v7.1, developed by AspenTech. Table 7 shows that 13C4Diol, 13C4Diol and EGlycol and 13C4Diol and DEG were used as separating agents in the extractive distillation process. The three agents yielded more than 80% of ethanol compared to other separating agents such glycerol and ethylene glycol, which yielded more than 98%. This study investigated other separating agents that could be used in the extractive distillation process. Even though the separating agent could not give release a high amount ethanol, it was highly efficient in water removal. The yield of bioethanol was higher at 88.49% (w/w) with the 13C4Diol and DEG as separating agents. Others separating agents like 13C4Diol had 84.86% (w/w) yield, while 13C4Diol and Eglycol yielded 85.60% (w/w) of bioethanol. The values do not show a significant difference in bioethanol production. This can be improved by adding a suitable component to the mixture.

Table 7

Yield of bioethanol in different separating agent

Condition	Temperature, °C	No. of Feed of Entrainer	Mass Flow Rate, kg/hr	Yield of Ethanol, %
Entrainer: 13C4Diol				
Maximum	108.56	5	3400	77.27
Optimum	90.00	1	1500	84.86
Minimum	90.00	1	3039.29	85.32
Entrainer: 13C4Diol and EGlycol				
Maximum	110.00	20	3398.12	76.38
Optimum	90.00	1	1564.04	85.60
Minimum	90.00	1	1564.04	85.60
Entrainer: 13C4Diol and DEG				
Maximum	108.36	20	3399.99	74.89
Optimum	90.77	1	1500.02	88.49
Minimum	90.00	1	1500	88.49

Effect of Parameters on Separating Agent

In order to obtain higher bioethanol production in the extractive distillation process, the parameter analysis was carried out. The analysed parameters were:

1) Temperature of entrainer
2) Number of entrainer feed stage
3) Molar flow rate of entrainer

The feed conditions of the binary mixture and the entrainer are given in Table 8.

Table 8

Parameter of simulation process

Stream	Feed of azeotropic mixture	13C4Diol	13C4Diol and EGlycol	13C4Diol and DEG
Temperature, ^{o}C	25.00	90.00	90.00	90.77
Pressure, kPa	99.30	101.30	101.30	101.30
Ethanol, kmol/hr	0.60	0.01	0.01	-
Water, kmol/hr	0.40	0.03	0.02	-
13C4Diol, kmol/hr	-	0.96	0.71	0.04
Eglycol, kmol/hr	-	-	0.26	-
DEG, kmol/hr	-	-	-	0.96
Mass Flow Rate, kg/hr	15.70	3039	1564	1500

Effect of Temperature in Entrainer

Temperature of entrainer has an important effect on the distillate composition and the reboiler energy consumption. Several authors have recommended considering temperature as a design variable and operating 5-15°C below the top temperature of the extractive distillation column (Mulia-Soto & Flores-Tlacuahuac, 2011). It can be observed that using a high entrainer feed temperature yielded less bioethanol (Figure 2) because higher temperature demands a high reflux ratio to reach a specified separation. The increase in temperature causes water present in this stage to vapourise, thus increasing the content of water in the distillate and decreasing its purity. By using the standard reflux ratio, a higher yield is produced in optimum temperature. The least energy demand corresponds to a low entrainer feed temperature and a low reflux ratio.

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)

Figure 2. Effect of temperature and no. of feed stage on yield of ethanol (a) 13C4Diol (b) 13C4Diol and EGlycol (c) 13C4Diol and DEG

Effect of Mass Flow Rate

The effect in molar flow rate is different for pure and mixture solvents. Figure 3 shows that pure components like 13C4Diol have a high molar flow rate for high efficiency in order to remove water from ethanol, while mixture components have a lower mass flow rate to obtain high ethanol production. This indicates that mixture components have a lower mass flow rate because the mixture components react together to remove water from ethanol. Pure components have a high mass flow rate because they have to react by themselves in the extractive distillation process.

(a) (b) (c)
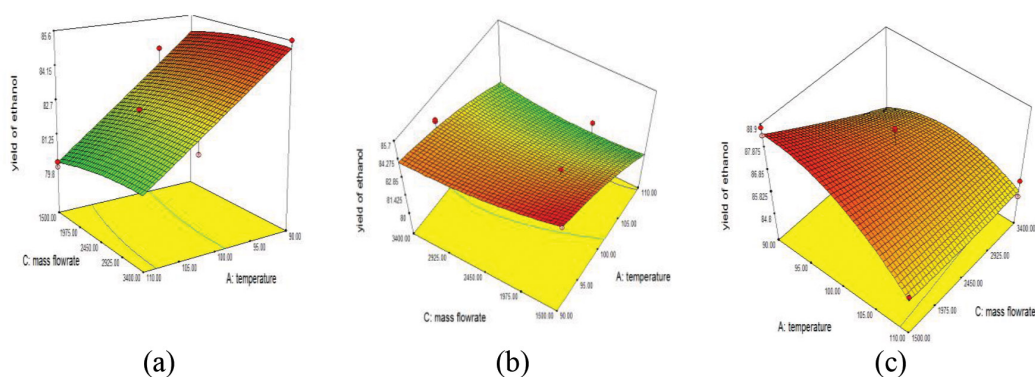
Figure 3. Effect of temperature and mass flow rate on yield of ethanol (a) 13C4Diol (b) 13C4Diol and EGlycol (c) 13C4Diol and DEG

Effect of Number of Stages for Feed Entrainer

Another parameter affecting the yield is the number of feed entrainers in the stages. The distillation column is operated at semi-batch fashion when the azeotropic mixture is introduced in the reboiler and the entrainer feed is continuous in other stages. Figure 4 presents the result for different feed stages and it can be seen in comparison to the yield of bioethanol production. For the specified number of stages, there is a best mixture feed stage, at which a higher yield of ethanol can be obtained. The separation is improved as more ethanol can be obtained in a longer period because the entrainer feed is close to the top of the distillation column and has greater contact with the ethanol-water mixture. However, when the feed stage is close to the bottom, the numbers of stages in the stripping section are not enough to extract the ethanol from an azeotropic mixture, and the ethanol in the bottom product does not achieve the specific value. This can cause a decrease in the composition distilled according to the mass balance. The reboiler has the same propose as the molar composition of ethanol in the distillate, so that the energy consumption increases when high ethanol concentration in the distillate is obtained.
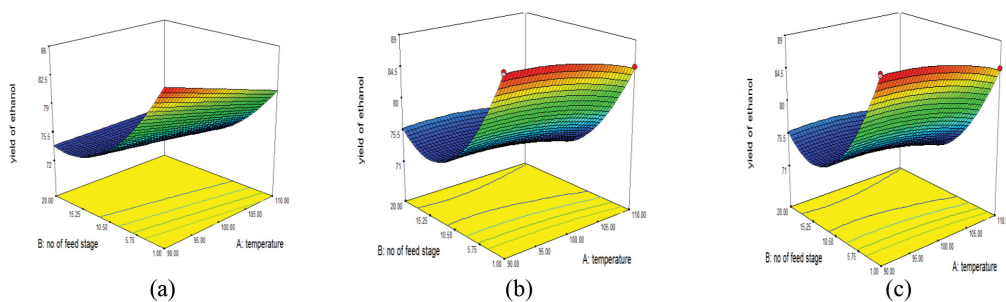
| (a) | (b) | (c) |
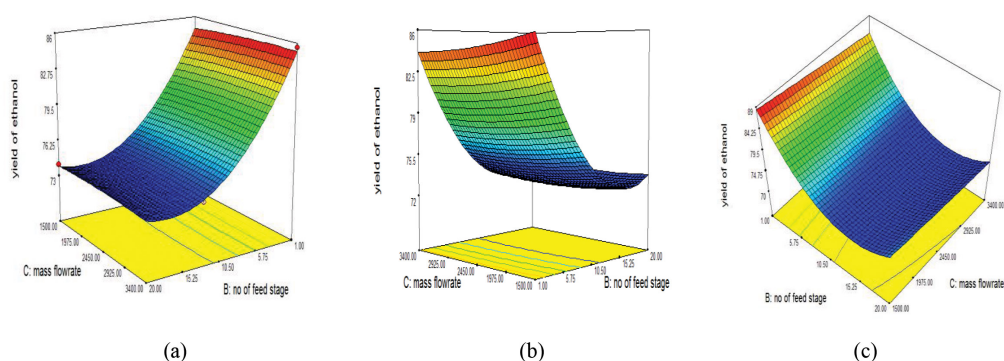
Figure 4. Effect of mass flow rate and no. of feed stage on yield of ethanol (a) 13C4Diol (b) 13C4Diol and EGlycol (c) 13C4Diol and DEG

CONCLUSION

In this study, the primary focus was to study the optimal operating parameters for production of bioethanol process. 13C4Diol, 13C4Diol and EGlycol and 13C4Diol and DEG were proposed as entrainers for the separation of an azeotropic mixture by extractive distillation with three columns to produce bioethanol. The operation process was investigated by simulation using the Aspen HYSYS v 7.0.0 software. In order to purify the bioethanol using a fermenter effluent, other compounds such as solid particles had to be removed besides water and ethanol. Therefore, an RSM design was employed to analyse the process variable, including the temperature of entrainer, number of entrainer feed stages and molar flowrate of entrainer. The optimum conditions for entrainer were identified. For the 13C4Diol, the temperature of the entrainer was 90°C with 1500 kg/hr as the entrainer flow rate, while the number of entrainer feed stage was one. Next, the optimum conditions for a mixture of 13C4Diol and EGlycol, the temperature of the entrainer was 90.77°C, with 1500 kg/hr as the entrainer rate and the number of entrainer feed stage was one. Lastly, the optimum conditions for a mixture of 13C4Diol and DEG needed a temperature of 90°C for the entrainer, with 1564.04 kg/hr as the entrainer rate, while the number of entrainer feed stage was one.

REFERENCES

Balat, M., & Balat, H. (2009). Recent trends in global production and utilization of bio-ethanol fuel. *Applied Energy*, *86*(11), 2273–2282.

Batista, F. R. M., Follegatti-Romero, L. A., Bessa, L. C. B. A., & Meirelles, A. J. A. (2012). Computational simulation applied to the investigation of industrial plants for bioethanol distillation. *Computers and Chemical Engineering*, *46*, 1–16.

Bauer, F., & Hulteberg, C. (2013). Is there a future in glycerol as a feedstock in the production of biofuels and biochemicals? *Biofuels, Bioproducts and Biorefining, 7*(1), 43–51.

Flach, B., Lieberz, S., Rondon, M., Williams, B., & Teiken C. (2015). *EU Biofuels Annual 2015* (pp. 1-42). Global Agriculture Information Network.

García-Herreros P., & Gomez, J. M. (2011). Optimization of the design and operation of an extractive distillation system for the production of fuel grade ethanol using glycerol as entrainer. *Industrial and Engineering Chemistry Research, 50*(7), 3977–3985.

Gil, I. D., Gómez, J. M., & Rodríguez, G. (2012). Control of an extractive distillation process to dehydrate ethanol using glycerol as entrainer. *Computers and Chemical Engineering*, *39*, 129–142.

Gil, I. D., Uyazan, A. M., Aguilar, J. L., Rodríguez, G., & Caicedo, L. A. (2008). Separation of ethanol and water by extractive distillation with salt and solvent as entrainer: Process simulation. *Brazilian Journal of Chemical Engineering, 35(1)*, 207–215.

Habaki, H., Hu, H., & Egashira, R. (2015). Liquid-liquid equilibrium extraction of ethanol with mixed solvent for bioethanol concentration. *Chinese Journal of Chemical Engineering, 24*(2), 253–258.

Langston, P., Hilal, N., Shingfield, S., & Webb, S. (2005). Simulation and optimisation of extractive distillation with water as solvent. *Chemical Engineering and Processing: Process Intensification*, *44*(3), 345–351.

Mulia-Soto, J. F., & Flores-Tlacuahuac, A. (2011). Modeling, simulation and control of an internally heat integrated pressure-swing distillation process for bioethanol separation. *Computers & Chemical Engineering*, *35*(8), 1532–1546.

Navarrete-Contreras, S., Sánchez-Ibarra, M., Barroso-Muñoz, F. O., Hernández, S., & Castro-Montoya, A. J. (2014). Use of glycerol as entrainer in the dehydration of bioethanol using extractive batch distillation: Simulation and experimental studies. *Chemical Engineering and Processing: Process Intensification*, *77*, 38–41.

Niemistö, J., Pasanen, A., Hirvelä, K., Myllykoski, L., Muurinen, E., & Keiski, R. L. (2013). Pilot study of bioethanol dehydration with polyvinyl alcohol membranes. *Journal of Membrane Science*, *447*, 119–127.

Pla-Franco, J., Lladosa , E., Loras S., & Montón , J. B. (2014) Thermodynamic analysis and process simulation of ethanol dehydration via heterogeneous azeotropic distillation. *Industrial and Engineering Chemistry Research, 53*(14), 6084–6093.

Segovia-Hernandez, J. G., Vázquez-Ojeda, M., Gómez-Castro, F. I., Ramírez-Márquez, C., Errico, M., Tronci, S., & Rong, B.-G. (2014). Process control analysis for intensified bioethanol separation systems. *Chemical Engineering and Processing: Process Intensification*, *75*, 119–125.

Szulczyk, K. R., McCarl, B. A., & Cornforth, G. (2010). Market penetration of ethanol. *Renewable and Sustainable Energy Reviews*, *14*(1), 394–403.

Ravagnani, M. A. S. S., Reis, M. H. M., Filho, R. M., & Wolf-Maciel, M. R. (2010). Anhydrous ethanol production by extractive distillation: A solvent case study. *Process Safety and Environmental Protection*, *88*(1), 67–73.

Zhao, J., Dong, C. C., Li, C. X., Meng, H., & Wang, Z. H. (2006). Isobaric vapor–liquid equilibria for ethanol-water system containing different ionic liquids at atmospheric pressure. *Fluid Phase Equilibria*, *242*(2), 147–153.

SCIENCE & TECHNOLOGY

PERTANIKA JOURNALS

24-Hour Discharge Post Laparoscopic Ovarian Cystectomy: A Feasibility Study

Wan Ahmad Hazim[1], Nur Hidayah Aeshah Ng[1] and Salleha Khalid[2]*

[1]*Department of Obstetrics & Gynaecology, Hospital Putrajaya, Wilayah Persekutuan Putrajaya, Malaysia*

[2]*Faculty of Medicine and Health Sciences, University Sains Islam Malaysia, 55100 USIM, Kuala Lumpur, Malaysia*

ABSTRACT

Minimally invasive surgery results in faster recovery. The objective of this study is to identify criteria for the feasibility of 24-hour discharge post laparoscopic ovarian cystectomy. This is a prospective cross-sectional study that was carried out at the Obstetrics and Gynaecology Department in Putrajaya Hospital between 1 January and 31 December, 2016. The inclusion criteria were: age between 15 and 45, no comorbidities, no family history of malignancy, BMI of less than 30, mass size less than 18 weeks, single uninoculated simple cyst and no ascites. The exclusion criteria were post-menopause women, known medical illness, family history of malignancy, mass size more than 18 weeks, multiloculated or bilateral ovarian cyst, presence of solid area within the cyst and ascites. The sample size was calculated to be 14. A total of 16 participants were identified. Results showed that using the Visual Analogue Score (VAS), the mean pain score post operatively in the first six hours, 24 hours, 48 hours, two weeks and three months were 3.67, 2.57, 0.5, 0 and 0, respectively. Two of the subjects experienced post-operative nausea and vomiting, one had urinary tract infection and one had minor bleeding from the surgical site. All the participants were discharged within 24 hours post-operatively. There was no readmission. In conclusion, 24-hour discharge post laparoscopic cystectomy is safe and feasible. Factors determining the success must be adhered to closely to ensure a good and satisfactory outcome. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Keywords: Laparoscopic cystectomy, 24-hour discharge

INTRODUCTION

Ovarian cyst is common among the reproductive age group of women. It may be both symptomatic or asymptomatic (National Institue of Health Census Development Conference Statment, 1994). Six percent of

5000 healthy women in a study reported by Campbell et al. (1989) had detectable adnexal masses on transabdominal ultrasound. Of these, 90% were cystic in nature, with most cases diagnosed as benign cysts (Royal College of Obstetricians & Gynaecologist and British Society for Gynaecologic Endoscopy, 2011). Ovarian cysts were the fourth most common gynecologic cause of hospital admission according to a late 1980s study by Grimes and Hughes (Grimes & Hughes, 1989). Most cysts spontaneously resolve, while some will persist. The persistent ovarian cysts are most likely to be surgically managed. It has always been the standard practice to do an ovarian cystectomy. Previously, this was done via laparotomy. Being a major surgical procedure, the patient usually stays for at least two or more days prior to discharge. Recovery usually takes about four to six weeks.

With the advancement of minimally invasive surgery, surgery can be done in comparable time to conventional laparotomy. The advantages are enormous, as access is minimal and there is less pain, minimal blood loss, less risk of post-operative adhesion formation, shorter hospitalisation, fast recovery and earlier return to normal activity. Within two weeks, the patient can return to work. Hence, the laparoscopic approach has been regarded as the gold standard approach in managing women with ovarian cysts (Royal College of Obstetricians & Gynaecologist and British Society for Gynaecologic Endoscopy, 2011). Putrajaya Hospital has been practising laparoscopic surgery since 2000 even though this technique was established in the early1990s. Since then the laparoscopic approach has been the preferred approach in more than 90% of cases dealing with benign ovarian cysts.

The objective of this paper was to look into the criteria and feasibility of 24-hour discharge post laparoscopic ovarian cystectomy. It is a pilot study as there is no reference to 24-hour discharge post laparoscopic ovarian cystectomy in the literature. Therefore, cross reference to daycare laparoscopic cholecystectomy was done.

METHOD

This was a prospective cross-sectional study that was carried out in the Obstetrics and Gynaecology Department in Putrajaya Hospital. Based on the inclusion and exclusion criteria, women who were referred to the Obstetrics and Gynaecology Department at Putrajaya Hospital from 1 January till 31 December, 2016 for further management of ovarian cysts were recruited as part of this cross-sectional study. The following were the inclusion and exclusion criteria:

1. Inclusion criteria
    a. All patients with confirmed diagnosis of ovarian cyst
    b. Pre-menopausal patient (age 15 to 45)
    c. No known medical illness
    d. No family history of malignancy
    e. Body Mass Index (BMI) of less than 30
    f. Mass size less than 18 weeks
    g. Single uniloculated simple cyst
    h. Has no ascites

2. Exclusion criteria
    a. Presence of solid area within the cyst
    b. Women who did not consent to take part in this study
    c. Operation lasted more than two hours
    d. Intra-operative findings of severe adhesions
    e. Other clinical or intra-operative features suspicious of malignancy (such as papillary projections and increased vascularity on imaging, solid area found intra-operatively)

Tumour markers, including Ca 125, were not requested. Patients selected were pre-menopause with ultrasound features that only included patients with uniloculated cyst with no solid areas or ascites. Taking tumour markers would also have had significant cost implication in conducting the study. In view of this, risk for malignancy index was not calculated.

Sample size was determined by using the following formulae (Daniel, 1999):

$$N = (Z*\sigma/\Delta)^2 = (1.96*1.9/1)^2 = (3.724)^2 = 13.86$$

where, N = number of sample size; Z = statistic for a level of confidence (1.96); σ = standard deviation [4]; Δ = precision (1)

The sample size was 14 patients in order to estimate the mean with a precision of 1 unit. We decided to take 16 patients with an additional 20% for anticipated non- response cases.

Procedure

All patients who reported to the Obstetrics and Gynaecology Department at Putrajaya Hospital for ovarian cysts who fulfilled the inclusion and exclusion criteria were enrolled into the study. There was no comparison group in this study. The duration of subject participation was 12 months.

Those who agree to participate in the study were required to sign the consent form provided during consultation at the outpatient clinic. Patients who were under the age of 18 years old were required to have their parents or guardians give written consent for the surgery on their behalf. This was in accordance to the requirements of the Malaysian Medical Council Guideline when obtaining consent from minors (Malaysian Medical Council, 2016). The patients were counselled thoroughly. Participation in this study was based on voluntary basis. Patients were free not to participate or to withdraw from the study at any point of time without jeopardising the care given. If they had doubts or questions or needed more time to think about the surgery, a second appointment was arranged accordingly.

An anaesthetist reviewed all the patients within the four weeks before surgery and assigned the ASA physical status score grade. Patients were given an information booklet outlining the procedure, potential problems and details of perioperative care.

The procedure was performed in the morning by a well-trained surgeon with an assistant. Patients were admitted to the ward one day before the surgery. Surgery was done as per routine.

Post operatively, patients were shifted to the post-operative recovery room and maintained on intravenous fluids for four hours post-surgery. The patients were assessed at regular intervals by a member of the surgical team and attending nurse for post-operative complaints and vital signs. Analgesic and antiemetic were given for pain and nausea or vomiting as required.

After four hours, the operating surgeon along with the anaesthesiologist evaluated each patient for consciousness level, vital signs, pain, nausea and vomiting. They were encouraged to sit up, drink and go to the toilet under supervision. Pain was assessed on the 0-10 Visual Analogue Scale (VAS): 0-4 (mild pain), 5-7 (moderate) and 8-10 (severe pain) (Vaughan, Gurusamy, & Davidson, 2013). Patients who fulfilled all the following criteria were discharged within 24 hours:

1. The surgeon did not anticipate that there would be any problems stemming from the operation.
2. Intra-operative blood loss was less than 500cc
3. Duration of surgery less than two hours
4. No intra-operative visceral injuries
5. Stable vital signs
6. Patient able to understand instructions and can ambulate
7. Patient relieved of nausea, vomiting and pain
8. Able to tolerate liquids and void urine
9. No bleeding from surgical sites
10. Patient feels comfortable and is ready to go home willingly

It was decided that patients would not be discharged within 24 hours if:
1. There were conversion to open cystectomy
2. Discharge criteria were not met
3. Unexpected medical problems or complications attributed to the surgery arose

The patients were given tablet Celecoxib acid 200 mg orally two times daily. They were provided with the hospital phone number and advised to contact the hospital if required or to report to the 24-hour emergency department if necessary. They were contacted the next morning at 48 hours to assess their general well-being, pain, discomfort, nausea, vomiting or any other side effects attributable to the anaesthesia or surgery. Patients with a pain score of more than 5 over 10 during follow-up were asked to return to the hospital immediately for further assessment of any complications. All the patients were given another appointment to visit the gynecology outpatient clinic after two weeks and three months to assess any early and late complications. Early complications are defined as any complication that arises within two weeks and late complications, within three months of surgery.

The patients' details were made known only to the primary investigator. All the results were confidential and all data extracted from this research for use in reports did not, under any circumstances, contain names or identifying characteristics. The duration and means of storage and archival of medical records and study data was per standard regulation i.e. a period of seven years. The study data will be destroyed after the period. No payment was made for taking part and patient did not have to pay for the study treatment and procedures.

The data obtained was entered into the Statistical Package for the Social Sciences (SPSS) version 15 for Windows. The data was analysed descriptively.

The ethical issue in this study was that the subjects were discharged earlier than the routine practice. As such, they were at risk of experiencing early complications at home that would go unnoticed but which would have routinely been recognised had they still been warded. This risk was managed by terminating any study subject who developed complications intra-operatively or who had to return to hospital for complications within 48 hours. This study was approved by the Medical Research and Ethics Committee of Malaysia (NMRR-14-1129-21265).

RESULTS

A total of 16 patients with ovarian cyst were recruited for this study. The age range was from 25 to 38 years old, with the mean age of 32. Nine were referred from health clinics and seven were referred by private general practitioners. All of the recruited patients had mass size per abdomen of less than 18 weeks. Eight cases were done by consultants and the rest were done by specialists. The mean operation time by consultants and specialists were 86.7 and 75.11 minutes, respectively. Seven cases had mild to moderate adhesions intra-operatively. Adhesionolysis was done successfully with no complications. A total of 15 (93.7%) patients had blood loss of less than 100 ml and one patient had blood loss of 400 ml. Table 1 summarises these findings. During the post-operative assessment, all the patients were orientated to time, place and person, ability to ambulate and ability to tolerate orally and none developed shoulder tip pain. We evaluated the post-operative pain score using the Visual Analogue Score (VAS) at the first six hours, 24 hours, 48 hours, two weeks and three months post-operatively. The results are shown in Table 2.

Table 1
Summary of research outcome

Criteria	Mean (SD)	N (%)
Age	31.6 (3.58)	
Operative time (minutes)		
Overall	80 (28.9)	
Consultant	86.7 (28.6)	
Specialist	75.11 (29.7)	
Surgeon		
Consultant		7 (43.8)
Specialist		9 (56.3)
Adhesion		11 (68.8)
Spillage		10 (62.5)
Estimated blood loss		
<100ml		15 (93.75)
>100ml		1 (6.25)

Table 2

Pain-score chart using the visual analogue score (vas) against time

Duration Post-Operation	Mean Pain Score (Visual Analogue Score)
Immediate	1.5
1 hour	1.8
2 hours	2.3
3 hours	2.58
4 hours	3.33
5 hours	3.58
6 hours	3.67
24 hours	2.57
48 hours	0.5
2 weeks	0
3 months	0

None of the patients developed early post-operative complications such as emphysema and late post-operative complications such as incisional hernia or wound infection. However, one patient had urinary tract infection, two experienced nausea and vomiting and one had minimal surgical site bleeding that was stopped by direct compression. Mean systole, diastole blood pressure and pulse rate at the time of arrival at ward from the operation theatre, 0 h, 1 h, 2 h, 3 h, 4 h, 5 h and 6 h were normal throughout, as shown in Figures 1 and 2.
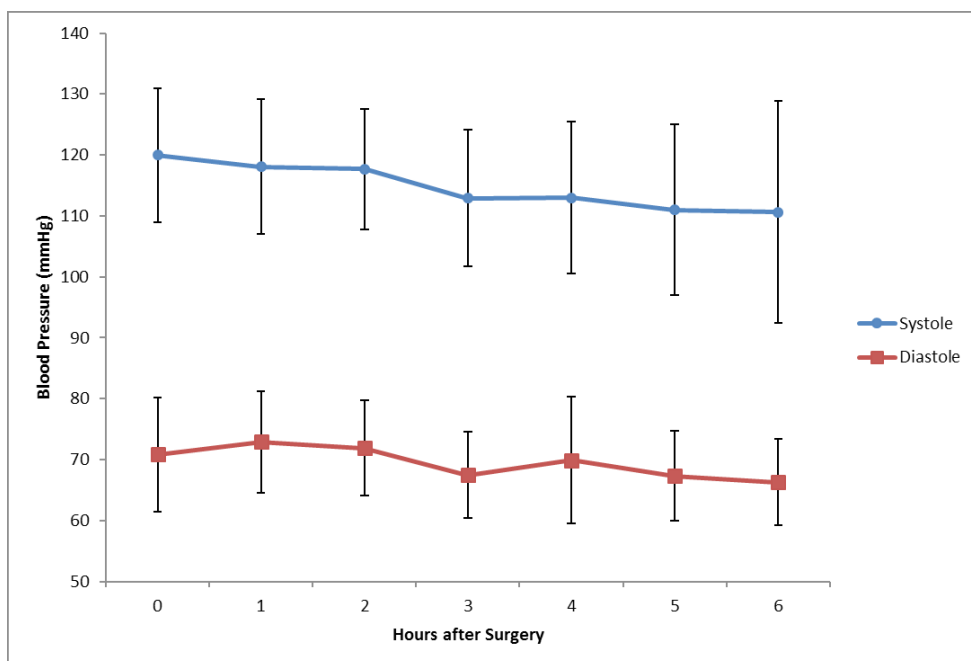
Figure 1. Post-operative (first 6 hours) mean systole and diastole blood pressure (mmHg)
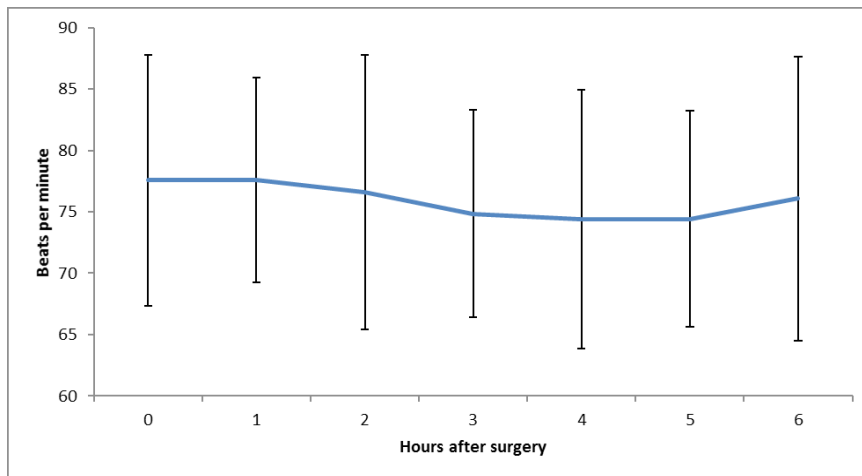
Figure 2. Post-operative (first 6 hours) mean pulse rate (bpm)

DISCUSSION

The feasibility of 24-hour discharge post laparoscopic cholecystectomy was demonstrated in this study to be safe (Kaman, Iqbal, Bukhal, Dahiya, & Singh, 2011). However, to the best of the authors' knowledge no similar data were available on 24-hour discharge laparoscopic cystectomy. In this study, 24-hour discharge was demonstrated to be safe, feasible and cost effective when applied carefully to patients who fulfilled a set of criteria (Kaman et al., 2011). In our series of laparoscopic cystectomy, we used the standard three or four-port technique using 5-mm instrument ports for dissection as well as for the camera. The main finding demonstrates that no critical complications occurred throughout the period of this study. We recorded the pain score at within six hours after surgery and discharge was proceeded with once the criteria were fulfilled. There was an initial increasing trend in pain score during the first 6 h due to the weaning effect of the general anaesthesia. This improved subsequently as the patients took an oral analgesic, with a mean pain score of less than 4 at 24 h. This shows that laparoscopy itself is associated with better pain control and early recovery. Initial follow-up was via telephone interview within 48 hours, then two weeks and finally three months (outpatient review) post-operatively. None of the patients had prolonged pain episodes after 48 hours post-operative in this series. The absence of readmission in this study indicated the safety and feasibility of 24-hour discharge post laparoscopic cystectomy.

Among other outcomes were post-operative nausea and vomiting. Post-operative nausea and vomiting (PONV) is unpleasant and exhausting for patients. It prolongs recovery time and delays patients' discharge, leading to increased hospital cost. The aetiology of PONV after laparoscopic cystectomy is not entirely clear. The intra-operative use of isoflurane and fentanyl, nitrous oxide, carbon dioxide insufflations, stretching of the peritoneum and increased blood pressure in the peritoneal cavity and post-operative administration of opioids was thought to contribute to PONV (Gan et al., 2003). Two patients developed PONV. Nevertheless, PONV was managed well and it did not delay the timing of discharge. The longer operative time taken

by the consultants was the result of more difficult surgery being needed. There was, however, no significant difference in the operative time between consultants and specialists (p=0.408). Four specialists were involved in this study. The experience range was from a minimum of two to eight years' experience in laparoscopy. The three consultants had more than 10 years' experience in laparoscopic surgery.

Psychologically, earlier discharge saw the advantage of better pain control, perception of quick recovery and self-confidence. This series exhibited that discharge within 24 hours of laparoscopic cystectomy was safe. It can be performed safely in any hospital equipped with trained supporting staff and surgeons. Careful selection of patients and good team work with the anaesthetist during induction and post-operative recovery enhance the good outcome. A strict criterion for discharge is an important requirement for ensuring that the safety of the patient will not be compromised. For a patient to be discharged the next day after the operation, certain patient selection criteria must be met, such as premorbid well patients with no medical illness, no family history of malignancy, normal body mass index, mass size less than 18 weeks- no intra-operative complications, not requiring post-operative blood transfusion, surgery duration of less than two hours, stable post-operative vital signs and pain score less than 4 using VAS.

Discharge with 24 hours post laparoscopic cystectomy is feasible, as reduced post-operative pain reduces requirement of an analgesic, has a better cosmetic effect in the long term and reduces the risk of post-operative infection compared to in the case of a laparotomy. The length of hospital stay required is significantly shorter with laparoscopic surgery. Most patients can return to their everyday lives much sooner than after open surgery.

The strength of this study was that it was a prospective study. Patient selection was strictly adhered to, resulting in a comparable group of patients. The limitation of this study includes the small sampling size and the fact that it was done in one laparoscopic centre rather than in several. Discharge within 24 hours in the case of simple laparoscopic cystectomy may not be applicable in other centres that may have a different operating procedure and a different set of patients and expertise.

CONCLUSION

Discharge within 24 hours of laparoscopic cystectomy is safe and feasible in carefully selected patients with no intra-operative complications and a post-operative VAS less than 4. Factors determining success must be adhered to closely to ensure a good and satisfactory outcome.

REFERENCES

Campbell, S., Bhan, V., Royston, P., Whitehead, M. I., & Collins, W. P. (1989). Transabdominal ultrasound screening for early ovarian cancer. *British Medical Journal, 299*(6712), 1363-1367.

Daniel, W. (1999). *Biostatistics: A foundation for analysis in the health sciences* (7th ed.). New York: John Wiley & Sons.

Gan, T. M. (2003). Consensus guidelines for managing postoperative nausea and vomiting. *Anesthesia and Analgesia, 97*(1), 62–71.

Grimes, D. H. (1989). Use of multiphasic oral contraceptives and hospitalizations of women with functional ovarian cysts in the United States. *Obstetrics and Gynecology*, 73(6), 1037–9.

Kaman, L., Iqbal, J., Bukhal, I., Dahiya, D., & Singh, R. (2011). Day care laparoscopic cholecystectomy: Next standard of care for gall stone disease. Gastroenterology research, 4(6), 257-261.

Magrina, J. (2002). Complications of laparoscopic surgery. *Clinical Obstetrics and Gynecology, 45*(2), 469–80.

Malaysian Medical Council. (2016). *Guideline on consent.* Kuala Lumpur: Malaysian Medical Council.

NIHCDCS. (1994). Ovarian cancer: Screening, treatment and follow-up. National Institue of Health Census Development Conference Statment. *Gynecol Oncol, 55*(3pt 2), S4–14.

Royal College of Obstetricians & Gynaecologist and British Society for Gynaecologic Endoscopy. (2011). Management of suspected ovarian masses in premenopausal women. *Green Top Guideline, No. 62*.

Seltzer, V. D. (1995). Ovarian cancer: Screening, treatment, and follow-up. *JAMA*, *273*(6), 491–497.

Vaughan, J., Gurusamy, K. S., & Davidson, B. R. (2013). Day-surgery versus overnight stay surgery for laparoscopic cholecystectomy. *The Cochrane Library.*

Yuen, P. M., Yu, K. M., Yip, S. K., Lau, W. C., Rogers, M. S., & Chang, A. (1997). A randomized prospective study of laparoscopy and laparotomy in the management of benign ovarian masses. *American Journal of Obstetrics and Gynecology, 177*(1), 109–14.

Performance Comparison of Classification Algorithms for Medical Diagnosis

Anju Jain*, Saroj Ratnoo and Dinesh Kumar

Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar-125001, India

ABSTRACT

Knowledge extraction from medical datasets is a challenging task. Medical datasets are known for their complexity in terms of noise, missing values and imbalanced class distribution. Classification algorithms can assist medical experts in disease diagnosis provided that a rigorous and methodological evaluation of classification models is applied by selecting appropriate sampling techniques, performance metrics and statistical tests. An ad hoc approach in this regard can result in unexpectedly high misclassification rates, which may prove very costly in terms of people's health and lives. In this paper, we illustrate a methodology to evaluate and compare multiple classification algorithms on multiple medical datasets. The example experiment is conducted by applying five well-known machine learning algorithms i.e. the Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Ant Colony Optimisation (ACO) and Genetic Algorithm (GA) for discovering classification models for disease diagnosis from 11 publicly available medical datasets from the UCI machine learning laboratory. We conclude through a stepwise evaluation process that the performance of the Random Forest classifier was significantly better in diagnosing various diseases. The paper also addresses the issue of class imbalance and non-uniform misclassification costs, usually prevalent in datasets for disease diagnosis.

Keywords: Classification algorithms, machine learning, medical diagnosis, performance evaluation

INTRODUCTION

Machine learning, a sub-discipline in the field of Artificial Intelligence, involves the study and design of algorithms that can learn from data (Brink & Richards, 2013). Machine learning is mainly useful in cases where deterministic solutions are not available. Medical diagnosis is one of the important activities in the field of medicine because early and accurate diagnosis helps people recover faster and saves the cost of prolonged treatment. Research on automated knowledge

extraction from medical data has been growing rapidly (Esfandiari, Babavalian, Moghadam, & Tabar, 2014). However, extracting accurate and comprehensive knowledge from enormous medical data is a challenging task. Medical datasets essentially have missing values and class imbalance that consequently influence the accuracy and comprehensibility of classification algorithms used for disease diagnosis.

Classification algorithms are being widely used in various medical applications related to the predictive domain (Çınar, Engin, Engin, & Ateşçi, 2009; Azar & El-Metwally, 2012). A classification algorithm works in two phases: i). The training phase learns the classifier from the training data set; and ii) The classification/prediction phase uses the model to predict the instances unseen during the training phase. The performance of a classification algorithm is analysed on a test set.

A proper way of evaluating classification algorithms for medical diagnosis is of crucial importance. Though plenty of research has been done on evaluating the performance of classifiers, yet it has been observed that most researchers do not follow a rigorous approach for evaluating classification algorithms (Japkowicz & Shah, 2014; Vanaja & Rameshkumar, 2015). A common approach for performance evaluation of classification algorithms is based on their accuracy or error rates over a large number of datasets using k-fold cross validation (CV). Subsequently, the results are averaged and the paired t-test is applied to test statistical significance in the observed performances of different classification algorithms. There are several issues with this approach such as error estimation, aggregation of the results, selection of evaluation metrics and statistical significance testing. The most widely used evaluation metric for classification algorithm is accuracy/error rate. However, it may be misleading to use accuracy as the lone performance criterion, particularly, in the presence of class imbalance and non-uniform misclassification costs; this is most often the case with medical datasets for disease diagnosis. At the same time, the most widely used test of statistical significance, the t-test, is certainly not suitable in cases where such testing is to be performed over multiple classifiers and datasets.

The aim of this research was to illustrate proper evaluation of classification algorithms on datasets related to disease diagnosis in terms of choice of evaluation methods, performance metrics and appropriate statistical analysis. A comparison of two classification algorithms over a single dataset is simple and can be carried over by applying t-test for performance metric values obtained over the tenfold cross validation sampling technique. Comparing multiple classifiers on multiple datasets is relatively a more complex problem that needs the attention of the people working in the field of computational medical diagnosis. In this paper, we demonstrate the complete process of the evaluation of multiple classifiers on multiple medical datasets. We have divided the whole process into five basic components: i) Selection of datasets and handling missing values in datasets ii) Choice of learning algorithms; iii) Performance measures of interest; iv) Error estimation/Sampling method; and v) Application of statistical tests and Dealing with non-uniform misclassification costs.

The rest of the paper is organised as follows: The next section describes the related work. Section 3 depicts the proposed classifier evaluation framework for comparing multiple classification algorithms applied to multiple datasets for medical diagnosis and datasets used

in the proposed work. We report the experimental setup with the results of our experimentation and address the issue of non-uniform misclassification cost in Section 4. Finally, we conclude the paper in the last section.

RELATED WORK

It is often emphasised that researchers should use appropriate evaluation methods, performance metrics and statistical analysis for evaluating classifiers. However, many a time, classification algorithms are evaluated and compared solely on the basis of point estimates of accuracy or error rates. This lack of proper understanding and appreciation of the context in which the different components of the evaluation framework operate may lead to misinterpretation of the evaluation outcomes (Japkowicz & Shah, 2014) and may have adverse influence in fields like medical diagnosis due to high misclassification costs. In this direction, researchers Tanwani, Afridi, Shafiq and Farooq (2009) have shown that the nature of the given dataset plays an important role in the classification accuracy of algorithms. They have provided 10 generic guidelines that can help researchers of the biomedical classification community to select an appropriate classifier for a given problem. Danjuma and Osofisan (2014) have underlined that since clinical data are noisy and imbalanced, appropriate sampling techniques and evaluation metrics should be used to estimate the performance of classifiers and subsequently, the results should further be subjected to rigorous assessment before applying them to real life clinical decisions. Further, it has also been highlighted that the quality of clinical decision support systems depends on the ability of researchers to carefully choose, tune and apply machine learning algorithms to clinical data (Danjuma, 2015; Bal et al., 2014).

The accuracy estimates of many classification algorithms lose their credibility in the presence of missing values in the datasets. The classification accuracy is predominantly affected by the presence of missing feature values because most of the classification algorithms cannot deal with these automatically (Marlin, 2008; Howell, 2007). Therefore, handling missing attribute values is a common concern in the field of data mining and knowledge discovery from medical datasets.

Ferri et al. (2009) have expressed concerns about what aspects of classifiers' performance need to be evaluated (Ferri, Hernández-Orallo, & Modroiu, 2009). The authors have emphasised that accuracy as a single performance evaluation metric is simple in its implementation, but less efficient in covering various aspects of the evaluation process and it should be supplemented with a Receiver Operating Curve (ROC). The ROC curve has a complex implementation, but it is more efficient and gives meaningful interpretations in medical domain (Esfandiari et al., 2014; Fawcett, 2006; Yousef, Wagner, & Loew, 2006; Prati, Batista, & Monard, 2011).

Some authors have raised objections to using the Null Hypothesis Significance Testing (NHST) in the context of machine learning algorithms (Demsar, 2008; Drummond, 2006; Ioannidis, 2005). Demsar (2008) propagated that NHST does not tell us all that we need to know and what many researchers believe it conveys. The test computes the (conditional) probability of certain statistics and says nothing about the hypothesis. Drummond (2006) also argued that rather than being confirmatory, the experiments should remain exploratory. Nevertheless, statistical testing does remain relevant for comparing classification algorithms

provided that appropriate statistical tests are applied. Many researchers use repeated paired t-tests for comparing more than two classification algorithms on multiple datasets. However, the t-test is only appropriate for comparing two classifiers on a single dataset. Further, Salzberg (1997) has suggested that because of the large number of experimental runs, a significance level of 0.05 most often used for t-tests is not stringent enough and the inferences drawn may have been obtained by chance (Hubbard & Lindsay, 2008). Such problems are well addressed by using the ANOVA and the Friedman tests. In addition, Demsar (2008) has compared the ANOVA and Friedman tests and he concluded that the Friedman test, being non-parametric, is safer than ANOVA for comparing multiple classifiers over multiple datasets since it does not assume normal distribution for the values of attributes and there is less chance of committing a type I error. Such non-parametric tests can be applied to classification accuracies, error ratios or any other performance metric of classification models.

Most of the medical datasets suffer from the problems of class imbalance and asymmetric misclassification cost. Sampling and cost-sensitive learning are the two commonly used methods to address these problems (Kotsiantis, Kanellopoulos, & Pintelas, 2006; Weiss, McCarthy, & Zabar, 2007; Ling & Sheng, 2011). Sampling includes under-sampling, over-sampling and a combination of the two. Cost-sensitive learning includes methods like MetaCost, CostSensitiveClassifier and Empirical Thresholding. Cost-sensitive learning has many times outperformed sampling methods (Japkowicz & Stephen, 2002). MetaCost, a general method for making a classifier cost sensitive, produces large execution cost reduction compared to the sampling techniques (Domingos, 1999; Kim, Choi, Kim, & Suh, 2012).

MATERIALS AND METHOD

Proposed Classifier Evaluation Framework for Medical Diagnosis

There are a few important steps for evaluating the performance of data mining algorithms. These steps depend on the type of the data mining task under consideration, underlying datasets and the purpose of the evaluation. Figure 1 shows the proposed classifier evaluation framework. This section describes the proposed framework for evaluating multiple classification algorithms on multiple datasets for the purpose of medical diagnosis.

1. Selection of datasets: This is not independent of other components of the proposed model since it affects the other components like sampling method, selection of performance measures and statistical tests. Therefore, keeping the purpose of the research in mind, a variety of medical datasets are chosen from the UCI repository. These cover a broad spectrum of variability in data characteristics such as dimensionality and class distribution. The class distribution is specifically important for the field of disease diagnosis.

2. Handling missing values in datasets: Accuracy achieved by any classification algorithm is inversely proportional to the amount of noise and missing values present in medical datasets. The removal of the examples with missing values belonging to the minority class can significantly affect the performance of a classification algorithm. Therefore, replacing missing values instead of removing the examples that contain missing values

may be more beneficial. In this work, a non-parametric method, called 'missForest', was used to replace missing values. This method was found to be suitable for replacing missing values because it can cope with mixed-type variables, nonlinear relations and high dimensionality of datasets (Stekhoven & Bühlmann, 2012). The algorithm is based on the Random Forest technique. For each variable, the 'missForest' algorithm fits a random forest on the observed values and then predicts the missing values.

3. Choice of Learning Algorithms: Among various algorithms in data modelling for disease diagnosis, a decision tree is well known for its simplicity and interpretability (Chang & Chen, 2009; Azar & El-Metwally, 2012). Other efficient algorithms such as the Support Vector Machine (SVM) and Random Forests have also become popular (Esfandiari et al., 2014; Hassanien & Kim, 2012; Jin, Tang, & Zhang, 2007) in the medical domain. The use of evolutionary and swarm intelligence algorithms have also gained popularity for discovering knowledge for medical diagnosis (Peña-Reyes & Sipper, 2000; Parpinelli, Lopes, & Freitas, 2002; Chen et al., 2011; Fidelis, Lopes, & Freitas, 2000; Freitas, 2013; Kumar, 2011). Five algorithms were selected for the evaluation exercise. These algorithms included SVM, the Decision Tree (DT) and the Random Forest (RF) as decision-based classifiers. The other two algorithms, General Algorithm (GA) and Ant Colony Optimisation (ACO), are nature-inspired classifiers. An algorithm known as the Pittsburgh Genetic Interval Rule Learning Algorithm (PGIRLA) was included in the GA-based classification algorithm (Corcoran & Sen, 1994). This GA follows the Pittsburgh approach, which is considered better than the Michigan approach because it evaluates the classification model as a whole. The most preliminary ACO algorithm for classification rule mining is the Ant Miner (Parpinelli et al., 2002). Several variants of this algorithm have been proposed with advanced features. An algorithm, Advanced Ant Miner Plus-C, was also included in this work. This algorithm is known for having discovered simple and small rule lists. This improves comprehensibility of the knowledge discovered (Parpinelli et al., 2002). The choice of classification algorithms is not an issue of this study. Researchers can choose classification algorithms of their interest.

4. Performance measures of interest: The selection of appropriate performance measures is essential in the domain of medical diagnosis. In case of imbalanced datasets, the use of accuracy as the lone performance evaluation metric does not give the real picture of how good a classifier is in discerning one class from the other (Chawla, 2005). For example, let us consider a dataset related to diagnosing cancer. Assume that this data set contains 90% of the examples with the negative occurrence of cancer and only 10% of the examples with positive occurrence of cancer. An inconsequential classifier can decide to always predict the majority class (non-cancerous class) for all data examples in the test set. Such a trivial classifier is expected to achieve an overall accuracy of 90 percent. This high accuracy figure may look impressive, but it is misleading because this classifier achieves perfect predictions for the majority class only (non-cancerous class) and no correct predictions at all for the minority class (cancerous class). All of the 10% of people who actually have cancer will be predicted as having no cancer, and this is a bad scenario for medical diagnosis. This example clearly indicates that

accuracy in itself is not an adequate measure to use for the datasets with a highly skewed class distribution. In medical applications, the sensitivity and specificity of a classifier matter more than its overall accuracy. Sensitivity measures the capability of a classifier to predict a diseased person correctly and specificity measures the capability of a classifier to diagnose a healthy person correctly. Therefore, we should use a performance metric that accounts for sensitivity as well as specificity (SE*SP). Another metric used in the medical domain is Youden's index (J), which is computed as (sensitivity+specificity-1) (Newby, Freitas & Ghafourian, 2013). One more simple measure of interest could be F-measure that takes account of precision and recall. The formula for F-measure is given below.

$$F - measure = \frac{2 \times \text{Pr}\,ecision \times \text{Re}\,call}{\text{Pr}\,ecision + \text{Re}\,call}$$

An additional important performance metric for classification problems related to medical diagnosis is the area under the Receiver Operating Curve (ROC). This measure can characterise the behaviour of a classifier over the full operating range and is more robust than accuracy in class imbalanced situations. Finally, we propose to use Sensitivity (SE), Specificity (SP), SE*SP, F-measure and ROC as the performance metrics to compare classification algorithms for medical diagnosis.

5. Error estimation/sampling method: Selecting the appropriate error estimation and re-sampling method is important while working with medical datasets. A stratified tenfold cross validation is suitable for medical diagnosis because of the presence of skewed class distribution. It ensures that the data distribution is respected in the training and testing sets created at every fold.

6. Application of statistical tests: The last step is to select appropriate tests for assessing the statistical significance of the difference in performance of the classification algorithms. For comparing the performances of multiple classifiers on multiple domains, the one-way repeated ANOVA measure and Friedman tests fit the bill. Due to lack of strong assumption about the distribution of underlying medical data, the non-parametric Friedman test is more appropriate. The null hypothesis of the Friedman test is that there is no significant difference in the performance of the classification algorithms (Japkowicz & Shah, 2014). Further, in case the Friedman test returns affirmative results in terms of statistical significance i.e. it rejects the NULL hypothesis that there is no significant difference in the classification algorithms, the post-hoc Nemenyi test is applied to identify the significantly different pairs of classifiers.

7. Dealing with non-uniform misclassification costs: The datasets for disease diagnosis often have a class imbalance. For example, people who are diagnosed positive for cancer will be far fewer than those who are diagnosed negative for the disease. This introduces a learning bias for the majority class and will have a poor prediction rate for diagnosing the disease. The additional issue is of non-uniform misclassification costs, which means that the type of misclassifications have unequal costs associated with them. In the domain of disease diagnosis, we consider the occurrence of cancer as the positive class (minority class in the data) and non-occurrence as the negative class

(majority class in the data). Although false positive and false negative, both types of errors have some cost associated with them, but false negative errors are considered costlier than false positive errors in disease diagnosis i.e. it is costlier to send those patients home who actually had the cancer but were diagnosed otherwise. The false negative error rate tends to increase in the presence of the class imbalance scenario. The performance metric, accuracy, does not account for unequal misclassification costs. We can address the problem either by using sampling or cost-sensitive learning techniques. The cost-sensitive learning techniques have been proven more effective in dealing with non-uniform misclassification costs than the sampling methods (Ling & Sheng, 2011). Therefore, we used a cost-sensitive learning method known as MetaCost to deal with non-uniform classification costs.

Datasets Used

This study used five classification algorithms on 11 medical datasets (Table 1) from the UCI Machine Learning repository. All the datasets were two-class datasets, except for the heart dataset, and are numeric in nature. The heart dataset, a multi-class dataset, is modified to obtain a two-class problem by defining the joint of two or more classes as negative and one class as positive. The datasets considered are partitioned using the tenfold cross validation (10-fcv) procedure. The main characteristics of these datasets are summarised in Table 1.

RESULTS AND DISCUSSION

This study applied the Decision Tree (J48), Random Forest and Support Vector Machine in WEKA (Bouckaert et al., 2002) and PGIRLA and Advanced_Ant_Miner_Plus-' in KEEL (Alcalá-Fdez et al., 2008) library. The entire process of applying classification algorithms on datasets is shown in Figure 2. Missing values in the dataset, if any, were replaced prior to running classification algorithms using the 'missForest' package in statistical software R, which was also used for applying the statistical tests (Venables & Smith, 2008). Since the purpose of this experiment was not to tune the parameters for optimised performance, all the classification algorithms were run with their default parameters as such. The parameters of these algorithms are presented in Table 2.

Table 3 shows the results obtained for the accuracy of the chosen algorithms and a classifier known as ZeroR. The ZeroR classifier has no predictive power and is useful for determining a baseline performance as a benchmark for other classification methods. It ignores all the predictors and predicts every instance in the test data to belong to the majority class (Witten, Frank, & Hall, 2011). It is noteworthy that in case of four datasets highlighted in bold font, ZeroR is comparable with the rest of the classifiers. The accuracy achieved by SVM classifier is almost the same as that of ZeroR classifier except for the Mammographic Mass dataset. This shows that accuracy does not necessarily suffice in its classical form for evaluating classifiers.

Table 4 shows the results obtained for the ROC, Specificity (SP), Sensitivity (SE), SP*SE and F-measure for classifiers J48, RF, SVM, GA and ACO. The values of specificity are significantly greater than the values of sensitivity for almost all the classifiers and datasets. This shows the bias of classification algorithms towards the majority class. This difference is more visible for the datasets (heart, hepatitis, WBCD, ILPD and thoracic) with high class imbalance. This fact highlights the need for addressing the issue of class imbalance.

A comparison of values of sensitivity for different datasets with all the five classifiers is shown in Figure 3. Figure 3 shows that the performance of SVM is the worst when the value of sensitivity is zero or very low for all the datasets except for the mammographic mass dataset. This means that the SVM fails to make correct predictions for diseased people. The figure also depicts that heart and thoracic are the most difficult datasets to classify. None of the algorithms work well on these datasets.
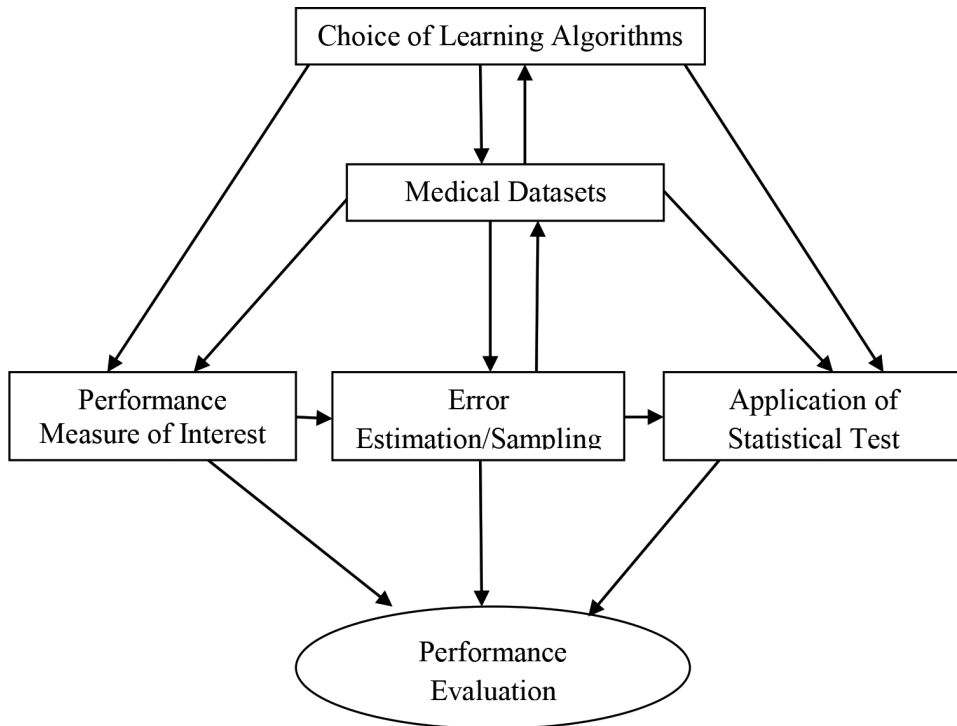
Figure 1. Proposed classifier evaluation framework for medical diagnosis
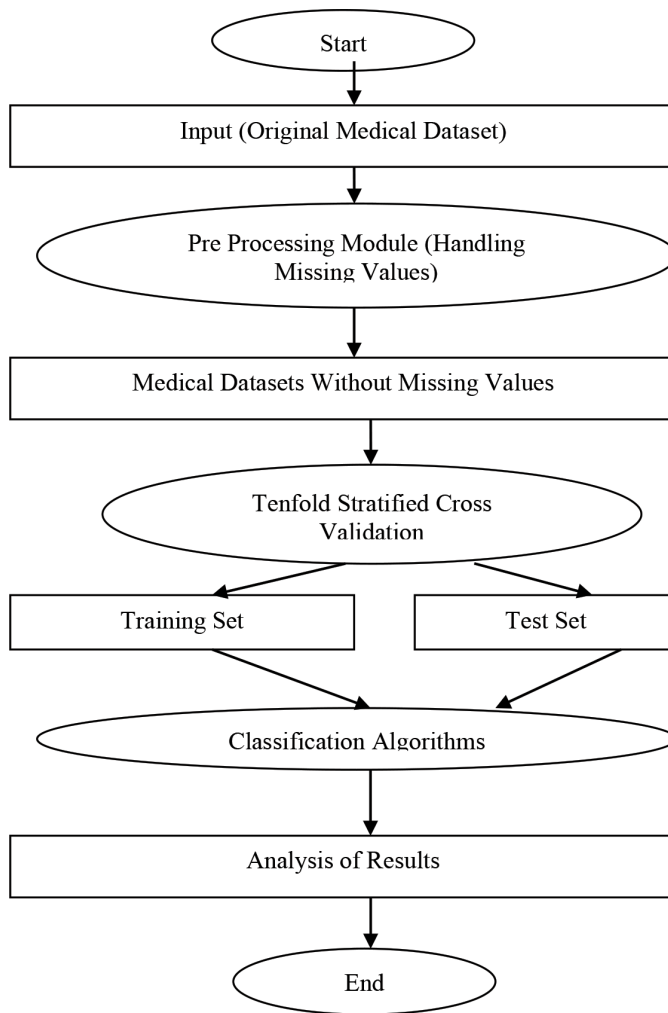
Figure 2. Flowchart of applying classification algorithms on medical datasets

Table 1
Datasets used in the experiment

Datasets	#Instances	#Attributes	#Classes	Class distribution ratio Majority: Minority
Heart (HE)	303	14	5	Remainder:19
Hepatitis (HEP)	155	20	2	79:21
Chronic Kidney Disease (CKD)	400	25	2	63:37
WBCD	699	11	2	66:34
Bupa Liver Disorder (BLD)	345	7	2	58:42
Pima Indian Diabetes (PID)	768	9	2	65:35
ILPD	583	11	2	72:28
Thoracic Surgery (TS)	470	17	2	85:15
Statlog Heart (SH)	270	13	2	56:44
Mammographic Mass Data (MMD)	961	6	2	54:46
WPBC	198	34	2	76:24

Table 2
Parameters considered for the classifiers

Algorithms	Parameters
J48	Confidence factor = 0.25, numFolds = 3
RF	numTrees = 100, Number of random features considered = 5
SVM	KernelType = radial basis function
GA	No. of generations = 1000, Pop-size = 61, Crossover probability = 0.7, Mutation probability = 0.5
ACO	No. of ants = 1000, Max_Pheromone = 0.99, Min_Pheromone = 0.1

Table 3
Results obtained for accuracy

Datasets	J48	RF	SVM	GA	ACO	ZeroR
HE	**52.8**	**58.41**	**54.12**	**53.97**	**56.0**	**54.12**
HEP	89.03	89.67	79.35	79.41	85.3	79.35
CKD	99.5	99.75	62.5	93.25	98.75	62.5
WBCD	94.27	96.7	66.38	92.69	94.54	65.5
BLD	68.69	73.33	59.42	64.9	43.48	57.97
PID	73.82	74.34	65.1	73.31	75.0	65.10
ILPD	**68.78**	**70.15**	**72.38**	**69.44**	**68.45**	**71.35**
TS	**84.46**	**84.04**	**85.1**	**84.89**	**84.0**	**85.10**
SH	76.66	81.85	55.92	70	81.48	55.55
MMD	84.7	80.12	81.37	80.41	82.75	53.6
WPBC	**76.26**	**80.81**	**76.26**	**67.9**	**63.68**	**76.26**
Average Accuracy	78.99727	80.83364	68.9	75.47	75.76636	66.03

Table 4
Result obtained for *ROC, SP, SE, SP*SE, F-measure for J48,RF,SVM,GA and ACO classifiers*

DATASETS	Results for various performance measures for Decision Tree classifier (J48)					Results for various performance measures for Random Forest classifier				
	ROC	SP	SE	SP * SE	F-measure	ROC	SP	SE	SP * SE	F-measure
HE	0.773	0.85	0.11	0.09	0.1	0.9	0.9	0.135	0.122	0.15
HEP	0.797	0.943	0.688	0.64	0.721	0.897	0.951	0.688	0.654	0.733
CKD	0.999	0.992	1	0.99	0.993	1	0.996	1	0.996	0.997
WBCD	0.933	0.956	0.917	0.87	0.917	0.992	0.965	0.971	0.93	0.953
BLD	0.665	0.8	0.531	0.424	0.588	0.764	0.815	0.621	0.5	0.662
PID	0.751	0.814	0.597	0.48	0.614	0.815	0.836	0.571	0.47	0.608
ILPD	0.674	0.829	0.335	0.27	0.674	0.742	0.861	0.305	0.26	0.37
TS	0.502	0.99	0.014	0.013	0.027	0.642	0.983	0.029	0.02	0.051

Table 4 (*continue*)

DATASETS	Results for various performance measures for Decision Tree classifier (J48)					Results for various performance measures for Random Forest classifier				
	ROC	SP	SE	SP * SE	F-measure	ROC	SP	SE	SP * SE	F-measure
SH	0.744	0.793	0.733	0.58	0.736	0.891	0.867	0.758	0.65	0.788
MMD	0.864	0.849	0.845	0.71	0.836	0.877	0.812	0.789	0.64	0.786
WPBC	0.58	0.861	0.447	0.38	0.472	0.691	0.993	0.213	0.21	0.345
AVERAGE	0.752	0.88	0.565	0.495	0.607	0.8373	0.907	0.552	0.495	0.585

DATASETS	Results for various performance measures for SVM Classifier					Results for various performance measures for GA classifier (PGIRLA)				
	ROC	SP	SE	SP * SE	F-measure	ROC	SP	SE	SP * SE	F-measure
HE	0.5	1	0	0	0	0.5	0.925	0.07	0.06	0.09
HEP	0.5	1	0	0	0	0.5	1	0	0	0
CKD	0.5	1	0	0	0	0.904	0.924	0.93	0.85	0.904
WBCD	0.512	1	0.025	0.025	0.049	0.95	0.98	0.28	0.27	0.42
BLD	0.518	0.995	0.041	0.04	0.079	0.59	0.805	0.29	0.23	0.372
PID	0.5	1	0	0	0	0.7	0.83	0.54	0.44	0.58
ILPD	0.518	1	0.036	0.036	0.069	0.5	0.95	0.03	0.02	0.05
TS	0.5	1	0	0	0	0.4	0.99	0.02	0.019	0.03
SH	0.504	1	0.008	0.008	0.017	0.67	0.84	0.51	0.428	0.59
MMD	0.813	0.822	0.804	0.66	0.8	0.81	0.85	0.79	0.67	0.8
WPBC	0.5	1	0	0	0	0.3	1	0	0	0
AVERAGE	0.533	0.983	0.083	0.069	0.09	0.6313	0.918	0.314	0.2715	0.348

| DATASETS	Results for various performance measures for ACO Classifier (Advanced Ant Miner Plus-C)				
	ROC	SP	SE	SP * SE	F-measure
HE	0.56	0.9	0.05	0.045	0.05
HEP	0.65	0.91	0.62	0.5642	0.62
CKD	0.98	1	0.96	0.96	0.97
WBCD	0.93	0.96	0.9	0.864	0.91
BLD	0.5	0.095	0.9	0.085	0.56
PID	0.7	0.85	0.56	0.476	0.61
ILPD	0.64	0.72	0.58	0.4176	0.5
TS	0.5	0.98	0.05	0.049	0.08
SH	0.76	0.89	0.71	0.63	0.76
MMD	0.82	0.89	0.81	0.72	0.81
WPBC	0.59	0.6	0.72	0.43	0.48
AVERAGE	0.693	0.8	0.623	0.476	0.577
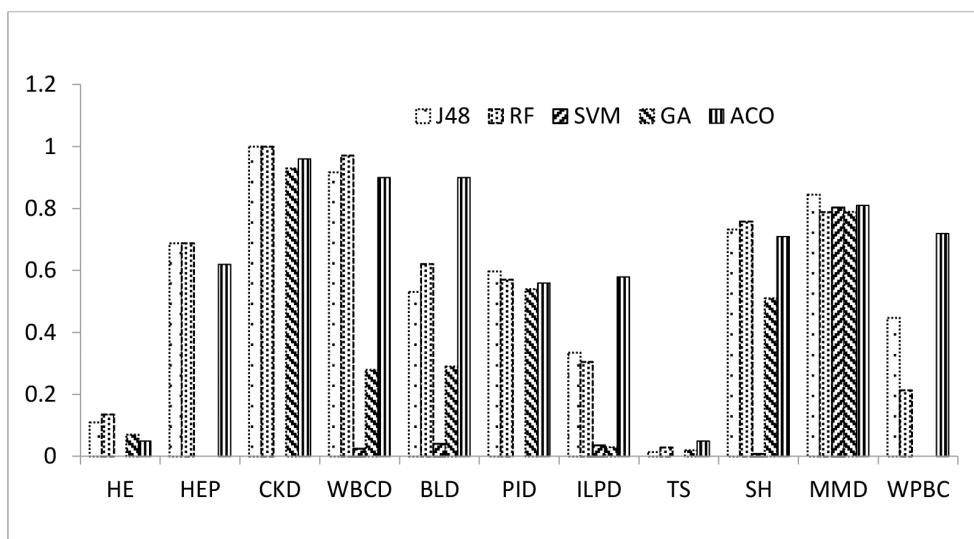
Figure 3. Comparison of sensitivity (SE) of five classifiers

One of the common techniques practised for an overall comparison of classification algorithms is to aggregate their performance over several datasets. Figure 4 shows the average values for ROC, SP*SE, F-measure and SE. The graph shows that the Random Forest classification algorithm is superior to all other classifiers in terms of the aggregated performance over all the datasets. Out of the selected classifiers, Random Forest has the highest average ROC (83%), SP*SE (49%), F-measure (59%) and SE (55%), followed by the Decision Tree and ACO classifiers. However, this is not sufficient and does not rule out the possibility that the result may have been obtained by chance. Therefore, we need to apply appropriate statistical tests.

This study applied the Friedman test, a non-parametric statistical test for comparing multiple algorithms on multiple domains, to see if indeed the performance of the Random Forest classifier on different domains is significantly better than that of the other classifiers. The test results are summarised in Table 5. The results show that the Random Forest Algorithm is significantly better than the rest of the classification algorithms on the performance metrics of ROC area, SP*SE, F-measure and SE at a significance level of 5% and degree of freedom, which equals 4. The results of the Friedman test agree on all metrics of the performance measures in this study but it may not be true for similar experimental studies with different combinations of classification algorithms and datasets.

This study also applied the post-hoc Nemenyi test to make a pair-wise comparison by using the Random Forest as the control classifier. The resulting Q values computed for the Nemenyi test for discovering the pair-wise differences between the performance of RF and other classifiers with respect to the metrics of ROC, SP*SE, F-measure and SE are given in Table 6.

The value of (q_α) (for a degree of freedom that equals 40) computes to 2.68 for a significance level of $\alpha = 0.05$. The null hypothesis are rejected only if the respective Q value for J48, SVM, GA and ACO exceed the value of q_α. One can conclude from the Q values given above that the RF classification algorithm was significantly better than SVM, GA and ACO classifiers

individually. Further, RF was significantly better than J48 on the basis of the metrics SE*SP and the ROC area. However, the NULL hypothesis could not be rejected for the RF and J48 classifiers on the basis of the F-measure and SE.
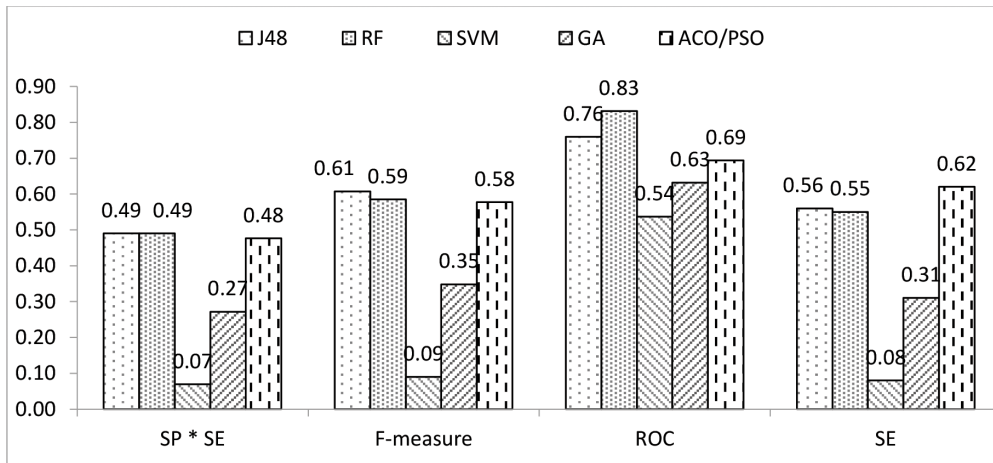

Figure 4. A comparison of aggregate performances of five classifiers of 11 medical datasets

Table 5
Summary of the results of the Friedman test

Domains	Friedman chi-squared value	Degree of freedom	p-value	Null hypothesis rejected/accepted
ROC	34.930	4	0.000000479	Rejected
SP*SE	25.981	4	0.00007695	Rejected
F-measure	27.410	4	0.000016422	Rejected
SE	27.037	4	0.000019539	Rejected

Table 6
Nemenyi test results

	ROC Area	SP*SE	F-measure	SE	Turkey's critical value (q_α)	Null hypothesis rejected/accepted
$Q_{J48, RF}$	21.64	3.73	0	-1.49		Rejected for ROC and SP*SE, Not rejected for F-measure and SE
$Q_{SVM, RF}$	56.71	45.52	44.02	41.79	2.68	Rejected
$Q_{GA, RF}$	50.74	32.08	32.08	31.34		Rejected
$Q_{ACO, RF}$	23.5	4.47	4.47	2.98		Rejected

Cost Sensitive Analysis

The issue of class imbalance and unequal misclassification costs is itself a very wide topic and needs a separate study. Thus, we have left a detailed study on the topic as a future scope and addressed the issue in a very limited way in this paper.

One of the important techniques to deal with unbalanced class medical datasets is to increase the cost of misclassification of minority class. There are two misclassification errors: False Positive (FP) errors and False Negative (FN) errors. In case of medical datasets, an FP error amounts to diagnosing a healthy person as diseased, whereas an FN error pertains to diagnosing a diseased person as healthy. Often, the datasets for disease diagnosis are biased towards the non-diseased class. Therefore, of these two misclassification errors, more cost needs to be assigned to FN errors because the class representing diseased people is the minority class.

The objective of a cost-sensitive approach is to build a classifier with the lowest misclassification cost among all the examples in a test dataset. We can reduce certain types of error (either FP or FN) by assigning a higher cost for such misclassifications (FN in our case). We have used a cost-sensitive learning method known as MetaCost to demonstrate how to take account of non-uniform misclassification costs. The MetaCost algorithm takes in misclassification costs usually determined by the subject experts. Assigning a high misclassification cost to FN errors will result in reducing the FN errors and increasing the sensitivity of a classification model. Table 6 depicts the results of using MetaCost in combination with the Decision Tree (J48) and Random Forest on some of the datasets with highly imbalanced or skewed class distributions.

It is clear from Table 6 that sensitivity increases and specificity decreases with the increasing misclassification cost for FN errors. These findings show that using a cost-sensitive technique can overcome the bias of a classifier towards the majority class up to some extent i.e. the tradeoff between specificity and sensitivity can be adjusted according to requirements. It is pertinent to observe from Table 7 that even after applying MetaCost only 20% and 15% of the diseased class instances of the thoracic surgery dataset (a highly class-imbalanced dataset) were classified correctly by the J48 and RF classification algorithms, respectively. This indicates the need for more research to correct the problem of class imbalance.

Figure 5 depicts the trend for sensitivity, specificity and accuracy with the increasing misclassification cost of FN errors for J48 and RF classifiers for two datasets with high class imbalance. The figure shows that sensitivity increases and specificity and accuracy decrease with the increasing misclassification costs for FN errors. The important conclusion is that a classifier with slightly low accuracy is not essentially a bad classifier for medical diagnosis because we are interested in the tradeoff between sensitivity and specificity. A classifier with a slightly low accuracy and high sensitivity rate may be considered better for medical diagnosis. The balance of tradeoff between FN and FP errors is finally an expert's decision.

CONCLUSION

The application of machine learning algorithms can only succeed and be useful in fields like medical or fault diagnosis if the domain is well understood and a proper evaluation framework is practised. The common approach for the performance evaluation of

classification algorithms for medical diagnosis has many issues such as selection of re-sampling techniques, choice of appropriate evaluation metrics and statistical significance testing. Since as a part of experimental design these choices influence the performance of classifiers, we have to be extremely careful in designing experiments for selecting classification algorithms for a sensitive field like medical diagnosis. In this study, we proposed and demonstrated a performance evaluation framework for comparing multiple classification algorithms across multiple datasets for the purpose of medical diagnosis. This approach will help researchers in sensitive fields like medical diagnosis, where misclassification costs are high, to adopt an appropriate approach for evaluating classification algorithms.

The nature of medical data makes it difficult to quickly select, tune and apply classification algorithms to disease diagnosis. The classification algorithms for disease diagnosis must specifically take care of missing values and skewed class distributions in medical datasets. Some techniques for replacing missing values need to be used and classifiers' bias towards the majority class needs to be taken care of either through sampling techniques or cost-sensitive analysis.

Accuracy, as the sole criteria for measuring the performance of classification algorithms in the domain of disease diagnosis, can be misleading in the presence of skewed class distributions. Keeping in view such circumstances, we need to focus on other criteria like sensitivity (rate of diseased persons classified correctly), specificity (rate of healthy persons classified correctly) and ROC. In domains like medical diagnosis, it is important for a classifier to perform well at discovering positive instances (belonging to the diseased class, which is actually the minority class) than to have an overall high accuracy rate. A classifier for medical diagnosis should keep the false negative rate (number of instances of diseased persons misclassified as healthy) to be as close to zero as reasonably possible. An expert's opinion is important for fixing the tradeoff between sensitivity and specificity. An expert may very well decide to have minimal FN errors. This may increase FP errors (healthy persons misclassified as diseased); however, these may be eliminated by conducting further tests on these instances.

Table 7

A cost-sensitive analysis using metacost with j48 and rf classifiers on four datasets

Cost FP:FN	Datasets	J48				RF			
		SE	SP	SE*SP	Accuracy	SE	SP	SE*SP	Accuracy
1:1	Hepatitis	0.781	0.951	0.742	91.61	0.656	0.959	0.629	89.67
	ILPD	0.347	0.839	0.291	69.81	0.275	0.858	0.235	69.12
	Thoracic Surgery	0.014	0.983	0.013	83.82	0	0.988	0	84.04
	WPBC	0.383	0.848	0.324	73.73	0.191	0.993	0.189	80.30
1:2	Hepatitis	0.75	0.951	0.713	90.96	0.719	0.902	0.648	86.45
	ILPD	0.467	0.755	0.352	67.23	0.443	0.781	0.345	68.43
	Thoracic Surgery	0.057	0.94	0.053	80.85	0.029	0.978	0.028	83.61
	WPBC	0.447	0.795	0.355	71.21	0.255	0.993	0.253	81.81

Table 7 (*continue*)

| Cost FP:FN	Datasets	J48				RF			
		SE	SP	SE*SP	Accuracy	SE	SP	SE*SP	Accuracy
	Hepatitis	0.813	0.935	0.76	90.96	0.844	0.902	0.761	89.03
	ILPD	0.563	0.661	0.372	63.29	0.665	0.704	0.461	69.29
1:3	**Thoracic Surgery**	0.071	0.903	0.064	77.87	0.043	0.95	0.04	81.48
	WPBC	0.468	0.788	0.368	71.21	0.277	0.947	0.262	78.78
	Hepatitis	0.844	0.894	0.754	88.38	0.844	0.878	0.741	87.09
	ILPD	0.689	0.647	0.445	65.86	0.772	0.637	0.491	67.58
1:4	**Thoracic Surgery**	0.129	0.845	0.109	73.82	0.114	0.93	0.106	80.85
	WPBC	0.426	0.662	0.282	60.60	0.362	0.834	0.301	72.22
	Hepatitis	0.813	0.821	0.667	81.93	0.875	0.846	0.740	85.16
	ILPD	0.671	0.635	0.426	64.49	0.832	0.567	0.471	64.32
1:5	**Thoracic Surgery**	0.2	0.783	0.156	69.57	0.157	0.895	0.140	78.51
	WPBC	0.596	0.629	0.374	62.12	0.426	0.722	0.307	65.15
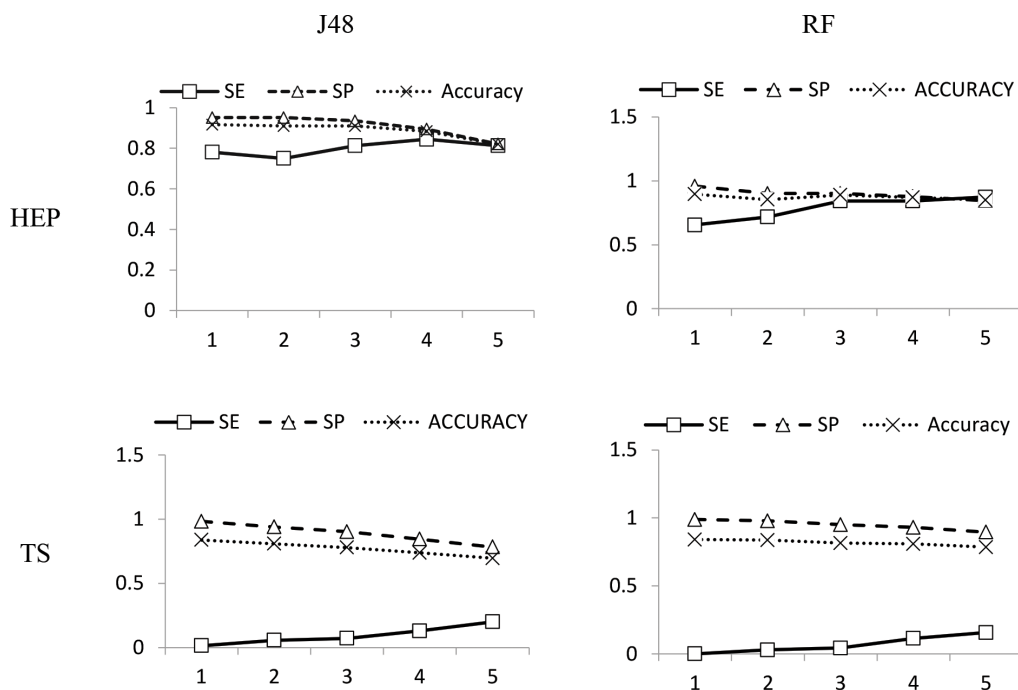
Figure 5. Change in sensitivity, specificity and accuracy with increasing misclassification cost for datasets using J48 and RF classifiers

REFERENCES

Alcalá-Fdez, J., Sánchez, L., García, S., del Jesus, M. J., Ventura, S., Garrell, J. M., … & Herrera, F. (2008). KEEL: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing, 13*(3), 307–318.

Azar, A. T., & El-Metwally, S. M. (2012). Decision tree classifiers for automated medical diagnosis. N*eural Computing and Applications, 23*(7-8), 2387–2403.

Bal, M., Amasyali, M. F., Sever, H., Kose, G., Demirhan, A., Bal, M., … & Demirhan, A. (2014). Performance evaluation of the machine learning algorithms used in inference mechanism of a medical decision support system. *The Scientific World Journal, 2014,* 1-15.

Bouckaert, R. R., Frank, E., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2002). WEKA Manual for Version 3-7-8. *University of Waikato, New Zealand*, 588-595.

Brink, H., & Richards, J. W. (2013). *Real-World machine learning* (7th Ed.). United States of America, USA: Manning Publications.

Chang, C. L., & Chen, C. H. (2009). Applying decision tree and neural network to increase quality of dermatologic diagnosis. E*xpert Systems with Applications, 36*(2, Part 2), 4035–4041.

Chawla, N. V. (2005). Data mining for imbalanced datasets: An overview. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 853–867). Springer US.

Chen, H. L., Yang, B., Wang, G., Wang, S. J., Liu, J., & Liu, D. Y. (2011). Support vector machine based diagnostic system for breast cancer using swarm intelligence. *Journal of Medical Systems, 36*(4), 2505–2519.

Çınar, M., Engin, M., Engin, E. Z., & Ateşçi, Y. (2009). Early prostate cancer diagnosis by using artificial neural networks and support vector machines. *Expert Systems with Applications, 36*(3), 6357–6361.

Corcoran, A. L., & Sen, S. (1994). Using real-valued genetic algorithms to evolve rule sets for classification. In *IEEE Conference on Evolutionary Computation Proceedings* (pp. 120–124). IEEE.

Danjuma, K. J., & Osofisan, A. O. (2014). Evaluation of predictive data mining algorithms in erythemato-squamous disease diagnosis. *International Journal of Computer Science Issues, 11*(6), 85–94.

Danjuma, K. J. (2015). Performance evaluation of machine learning algorithms in post-operative life expectancy in the lung cancer patients. I*nternational Journal of Computer Science Issues, 12*(2), 1-11.

Demsar, J. (2008). On the appropriateness of statistical tests in machine learning. In *Proceedings of ICML'08*. Association for Computing Machinery.

Domingos, P. (1999). MetaCost: A general method for making classifiers cost-sensitive. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 155–164). ACM.

Drummond, C. (2006). Machine learning an experimental science (revisited). In *AAAI'06 Workshop on Evaluation Methods for Machine Learning* (pp. 1–5).

Esfandiari, N., Babavalian, M. R., Moghadam, A.-M. E., & Tabar, V. K. (2014). Knowledge discovery in medicine: Current issues and future trend. *Expert Systems with Applications, 41*(9), 4434–4463.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters, 27*(8), 861–874.

Ferri, C., Hernández-Orallo, J., & Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters, 30*(1), 27–38.

Fidelis, M. V., Lopes, H. S., & Freitas, A. A. (2000). Discovering comprehensible classification rules with a genetic algorithm. In Evolutionary Computation, 2000. *Proceedings of the 2000 Congress on* (Vol. 1, pp. 805–810). IEEE.

Freitas, A. A. (2013). *Data mining and knowledge discovery with evolutionary algorithms*. New York, NY: Springer Science and Business Media.

Hassanien, A. E., & Kim, T. (2012). Breast cancer MRI diagnosis approach using support vector machine and pulse coupled neural networks. *Journal of Applied Logic, 10*(4), 277–284.

Howell, D. C. (2007). The treatment of missing data. *The SAGE handbook of social science methodology* (pp. 212–226). Great Britain: SAGE Publications Ltd.

Hubbard, R., & Lindsay, R. M. (2008). Why P values are not a useful measure of evidence in statistical significance testing. *Theory and Psychology, 18*(1), 69–88.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Med, 2*(8), e124.

Japkowicz, N., & Shah, M. (2014). *Evaluating learning algorithms: A classification perspective* (1st Ed.). United States of America, USA: Cambridge University Press.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis, 6*(5), 429–449.

Jin, B., Tang, Y. C., & Zhang, Y.-Q. (2007). Support vector machines with genetic fuzzy feature transformation for biomedical data classification. *Information Sciences, 177*(2), 476–489.

Kim, J., Choi, K., Kim, G., & Suh, Y. (2012). Classification cost: An empirical comparison among traditional classifier, cost-sensitive classifier, and metacost. *Expert Systems with Applications, 39*(4), 4013–4019.

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering, 30*(1), 25–36.

Kumar, D. (2011). A Genetic algorithm with entropy based probabilistic initialization and memory for automated rule mining. In *Advances in Computer science and information technology* (pp. 604–613). Springer, Berlin, Heidelberg.

Ling, C. X., & Sheng, V. S. (2011). Cost-sensitive learning and the class imbalance problem. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 231–235). Springer US.

Marlin, B. (2008). Missing data problems in machine learning. (Doctor of Philosophy in Dept of Computer Science). University of Toronto, Canada.

Newby, D., Freitas, A. A., & Ghafourian, T. (2013). Coping with unbalanced class data sets in oral absorption models. *Journal of Chemical Information and Modeling, 53*(2), 461–474.

Parpinelli, R. S., Lopes, H. S., & Freitas, A. A. (2002). Data mining with an ant colony optimization algorithm. I*EEE Transactions on Evolutionary Computation, 6*(4), 321–332.

Peña-Reyes, C. A., & Sipper, M. (2000). Evolutionary computation in medicine: An overview. *Artificial Intelligence in Medicine, 19*(1), 1–23.

Prati, R. C., Batista, G. E. A. P. A., & Monard, M. C. (2011). A survey on graphical methods for classification predictive performance evaluation. *IEEE Transactions on Knowledge and Data Engineering, 23*(11), 1601–1618.

Salzberg, S. L. (1997). On comparing classifiers: pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery, 1*(3), 317–328.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest - Nonparametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112–118.

Tanwani, A. K., Afridi, J., Shafiq, M. Z., & Farooq, M. (2009). Guidelines to select machine learning scheme for classification of biomedical datasets. In *Proceedings of the 7th European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics* (pp. 128–139). Berlin, Heidelberg: Springer-Verlag.

Vanaja, S., & Rameshkumar, K. (2015). Performance analysis of classification algorithms on medical diagnoses-a survey. *Journal of Computer Science, 11*(1), 30–52.

Venables, W. N., Smith, D. M., & R Development Core Team. (2008). *An Introduction to R.* United Kingdom, UK: Network Theory Ltd.

Weiss, G. M., McCarthy, K., & Zabar, B. (2007). Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs? In R. Stahlbock, S. F. Crone, & S. Lessmann (Eds.), *Proceedings of the 2007 International Conference on Data Mining, DMIN 2007* (pp. 35–41). CSREA Press.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd Ed.). Burlington, MA: Morgan Kaufmann.

Yousef, W. A., Wagner, R. F., & Loew, M. H. (2006). Assessing classifiers from two independent data sets using ROC analysis: A nonparametric approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(11), 1809–1817.

Estimating the Economic Impact of Climate Change on Agricultural Water Management Indicators

Hayat Lionboui[1]*, Tarik Benabdelouahab[1], Fouad Elame[1], Aziz Hasib[2] and Abdelali Boulli[2]

[1]*National Institute of Agricultural Research, PB 415 RP, Hassan II Avenue, Rabat, Morocco*

[2]*Faculty of Science and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco*

ABSTRACT

Climate change has been strongly dealt with in the literature on the economy in recent decades. In many semi-arid regions of the world, issues related to climate change including the scarcity of water resources are crucial for consideration in overcoming obstacles to development. The value of water increases and presents itself as a quantitative and qualitative constraint to its domestic, industrial and, particularly, agricultural use. Given the multi-dimensional and multi-scalar nature of climate change, this research proposes an integrated agronomic, economic and hydrologic modelling for the Tadla sub-basin in Morocco to assess the potential economic impacts of climate change on agricultural water management in this region. This tool permits to predict inter-annual variations of selected socio-economic water management indicators from 2014 to 2050. The simulated climate scenarios include the RCP (Representative Concentration Pathway) 4.5 and RCP 8.5 scenario, taking into account the main crops, water resources and the socio-economic context of the study area. The results show that climate change can have an impact in reducing the total agricultural profit calculated across the Tadla sub-basin (by about -0.34% and -0.44% each year according to RCP 4.5 and RCP 8.5 climate scenarios, respectively) and in increasing irrigation water shadow price progressively. Overall, this research helps to improve understanding of the potential economic impact of climate change and to guide stakeholders in their choices in terms of future adaptation policies.

Keywords: Agricultural profit, climate change, economic impact, RCPs 4.5 & 8.5, water management

INTRODUCTION

Climate change is a socio-economic and environmental phenomenon that challenges and concerns the international political and scientific communities, given its growing impact on global food security. In

Mediterranean countries, water resources are severely affected by climate change and so, the effects become more noticeable and more expensive (Iglesias et al., 2011).

Located in Africa, with a significant Mediterranean coast, Morocco is a country highly vulnerable to extreme weather and climate change (Niang et al., 2014). The country's hydrological context is mainly influenced by a spatial and temporal heterogeneity of renewable water volumes and by their scarcity, about 22 billion (m³/year) (CESE, 2014), the equivalent of 660 (m³/pers.an). Considering the water scarcity indicator, defined by Falkenmark (1989) as the volume of renewable water per capita, Morocco is facing a chronic water shortage, which can be aggravated by climate change. For this, the country was early committed to the international regime of climate change. It adopted the Framework Convention of the United Nations about Climate Change (UNFCCC) in 1992, which marks the basis for defining the means for worldwide sustainable development. With this convention and the Kyoto Protocol, various initiatives and actions have been implemented. However, climate change remains a major challenge facing the country because it is difficult to identify all the negative consequences of this phenomenon on ecosystems and humans.

In fact, one of the most affected activities, in this regard, is agriculture. Agriculture is considered a strategic sector for the socio-economic development of Morocco in terms of its contribution to the Gross Domestic Product (GDP), its role in employment (about 80% of rural employment) and its contribution to external trade (Toumi, 2008). However, the sector remains highly dependent on rainfall and, therefore, on climatic conditions. In Morocco, water availability is one of the main limiting factors in achieving good yields. A set of irrigated areas in the country was equipped with the means to improve and secure better crop production. Irrigated agriculture occupies only 15% of the cultivated area (about 1.5 million ha) in the country, but accounts for about 45% of the agricultural Gross Domestic Product and 75% of agricultural exports, depending on the season. According to Ministry of Agriculture reports, these statistics are more interesting when production in rain-fed areas is severely affected during dry years.

In spite of the large amounts of consumed irrigation water, crop yields in irrigated areas remain unstable from one season to another due to fluctuating water availability (Benabdelouahab et al., 2016). Therefore, climate insecurity may cause costly food insecurity for the country, leading to the need for massive imports in dry years, and drastically affects the living conditions of the rural population, with fallouts on other economic activities. Furthermore, productivity per irrigated hectare has not yet reached the potential for certain farms, now 69% in terms of economic efficiency on average, and the volume of water (m³) is not sufficiently valued by high-value crops (Lionboui et al., 2016).

In this situation, a considerable effort in scientific research must be undertaken to support current agricultural policies in order to assess and control the economic impact of climate change on agricultural water management. Several models have been adopted to identify and assess the damage caused by climate change and to support and assist adaptation policies. However, most of this research has been based on biophysical modelling that focusses on one specific dimension of climate change, such as the hydrological dimension (Immerzeel et al., 2010; Nepal, 2016) or the agronomic dimension (Moriondo et al., 2010; Ventrella et al., 2012). For the economic dimension, the models used for assessing the economic impact can be

classified into two main types, namely, the general equilibrium and partial equilibrium models. The first model considers the economy as a complete system of interdependent elements. The second is based on the analysis of part of the whole economy as a single market or a subset of markets. Owing to the sophistication involved in their use, the general equilibrium models are generally unsuitable for developing countries (Baby et al., 2014) such as Morocco. Therefore, the partial equilibrium models are the most appropriate in this case for assessing the economic impact of climate change on agriculture (Mishra & Sahu, 2014). The literature mentions, on the one hand, models based on the Ricardian approach (Mendelsohn et al., 1994) or hedonic pricing method; however, in the economic theory, this method is considered more appropriate for consumer goods than for production factors (Martin & Vaitkeviciute, 2016). On the other hand, we find models based on agronomic-economic and/or hydrologic approach (D'Agostino et al., 2014; Esteve et al., 2015; Hurd & Coonrod, 2012) that are based on controlled experiments in the field or laboratory settings that we have adopted for this research.

The proposed model was developed to predict inter-annual variations of selected socio-economic water management indicators at the Tadla sub-basin level from 2014 to 2050. In this perspective, this dynamic agronomic-economic-hydrologic model allows analysis of potential economic impact of climate change on agricultural water management while taking into account the development projects scheduled in the context of current policies. Therefore, this research was intended to provide applied tools for decision-makers to implement new approaches to adopt the most appropriate agricultural policies that may be able to reduce the negative impact of climate change.

MATERIALS AND METHODS

Agricultural Water Resources and Irrigation Schemes of the Study Area

The Tadla sub-basin is characterised by semi-arid to arid climate, with a dry season from April to October and a wet season from November to March. It covers three agricultural territorial units (ATUs): the Tadla plain (ATU 1), a rain-fed agricultural area that uses private groundwater pumping (ATU 2) and the Dir unit (ATU 3). Created in the 1940s, the irrigated perimeter of this sub-basin was among the first large irrigation schemes in the country. The landscape is a plain located in central Morocco (32°23′ N latitude; 6°31′W longitude; 445 m above sea level). This plain covers about 100,000 ha and is characterised by flat topography. It is characterised by semi-arid climate, with about 300 mm average annual precipitation over the period of 1970-2010, with a high inter-annual variation ranging from 130 to 600 mm in the same period (ABHOER, 2012). The average temperature is 18°C, with a maximum of 38°C in August and a minimum of 3.5°C in January (ORMVAT, 2014).

The irrigated perimeter of Tadla is divided into two sub-schemes by the Oum-Er-Rbia river, flowing from the Middle-Atlas Mountains (east) to the Atlantic Ocean (west) (Figure 1). Irrigation water used in the Tadla perimeter comes mainly from surface water (87.1% of the total amount of irrigation water consumed in 2009/2010). Two dams, Ahmed-Al-Hansali (750 mm³) and Bin-El-Ouidane (1.5 billion m³), supply irrigation water to the Tadla perimeter, besides groundwater pumping.
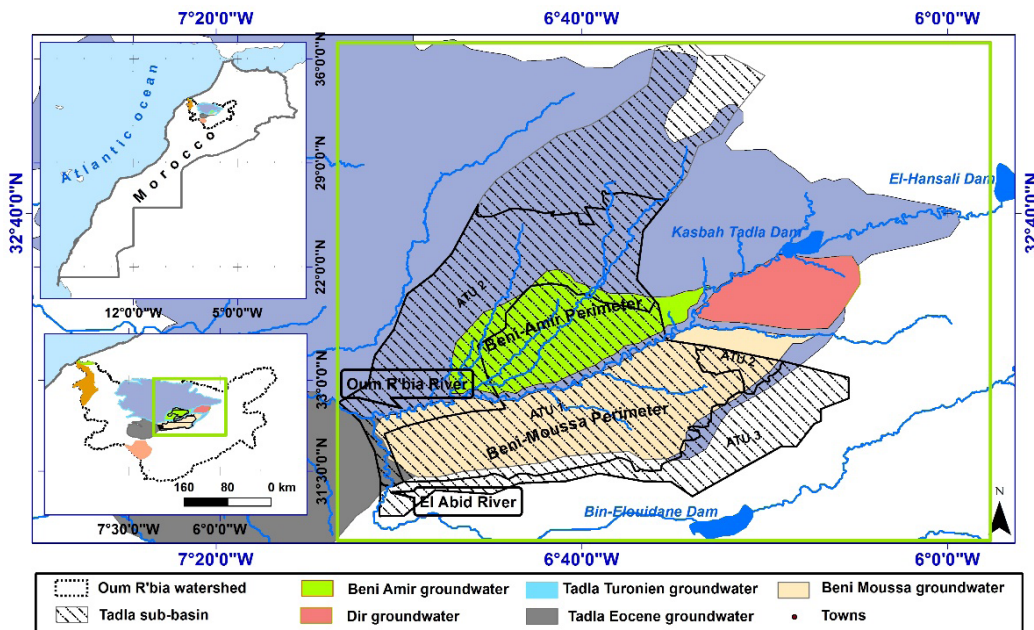
Figure 1. Location of the Tadla sub-basin

The overexploitation of groundwater has led to a lowering of piezometric levels. However, the amount of groundwater use has increased in these last few years due to the frequent recorded droughts. Groundwater used for agricultural purposes in Tadla comes from Beni-Moussa and Beni Amir groundwater and the Turonian deep water table.

Model and Data

In a context of climate change, the main objective of this study is to develop an agronomic, economic and hydrologic model in order to help decision-makers to choose the most appropriate adaptation measures for the country's context. This model is based on the simulation of water flows, equilibrium equations of water supply and use, water flows at different river nodes and the allocation of water resources (Lionboui et al., 2014). It is programmed in GAMS (Brooke et al., 1998) and was resolved using the non-linear solver CONOPT.

In addition to reflecting the dynamics of interactions between the different components i.e. hydrologic, agronomic and economic components, this tool allows the assessment of the potential economic impacts of climate change on agricultural water management under various climate scenarios (RCPs 4.5 and 8.5), taking into account the major crops, water resources and the socio-economic context of the region (Figure 2).
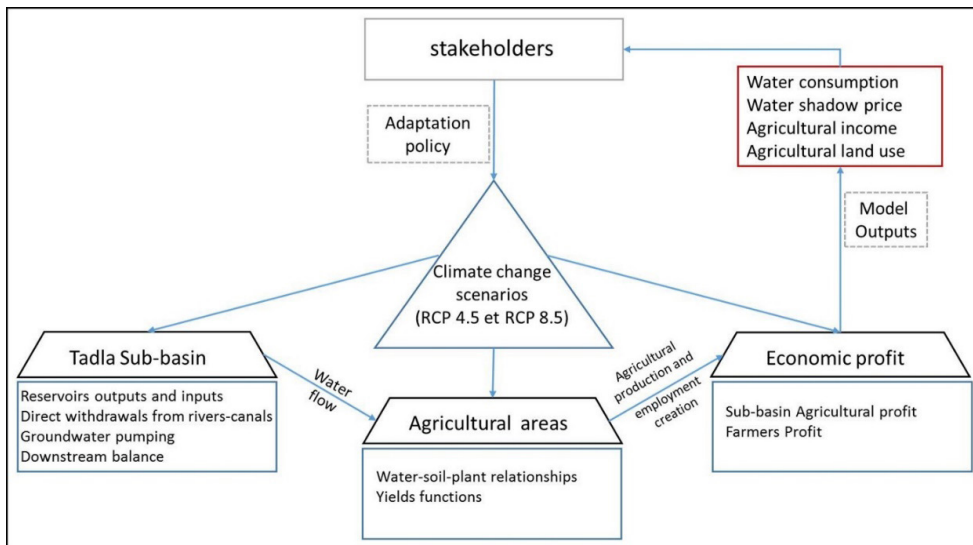
Figure 2. Tadla sub-basin model linkage

The proposed model is a non-linear optimisation model, which, given various constraints, maximises an objective function that can be value-added, or any other function reflecting the preferences and choices of decisions-makers at the sub-basin level. Once the objective function (Eq.1) and the functions of specified constraints was identified, the calibration of the model was obtained by using positive mathematical programming (Howitt, 1995) as follows:

$$Max\ VA = \sum_A \left(\sum_{IM} VA\_CAL_{A,IM} \right) \tag{1}$$

'VA' is the agricultural net value-added factor at the sub-basin level and 'VA_CAL' is the net value-added per agricultural area and per irrigation mode after calibration. 'A' is the agricultural area, referring to the sub-unit of the commune in order to differentiate between the existing irrigation modes. The net value-added per agricultural area was calculated from the output generated by the agricultural production and the labour cost, minus the producing and amortisation costs.

The model optimises the objective function on the basis of the observed values. Thus, the Earth availability constraint was also taken into account in the model. 'CU_AREA' represents the sum of cultivated areas by agricultural area 'A' and did not exceed the available area of arable land 'AV_AREA' in the Tadla sub-basin (Eq. 2):

$$AV\_AREA \geq \sum_{A,IM,Cr} CU\_AREA_{A,IM,Cr} \tag{2}$$

The initial land-use constraint refers to the part of each crop in the cultivated land by agricultural area and by irrigation mode. The marginal value of this constraint was used in the model calibration (Eq. 3):

$$CROP\_PT_{A,IM,Cr} = \frac{CU\_AREA_{A,IM,Cr}}{\sum_{Cr} CU\_AREA_{A,IM,Cr}} \tag{3}$$

'CROP_PT' defines the variable that determines the part of each crop in the total cultivated area and 'CU_AREA' is the parameter that indicates the area occupied by each crop in the reference year (2013-2014).

In the Tadla sub-basin, the water quantity used for irrigation ('TOT_WAT') refers to water coming from reservoirs and other surface sources ('SF_WAT') by period (months) 'Pd' and water pumped ('GD_WAT) from aquifers 'Aq' (Eq.4):

$$\sum_{Cr} TOT\_WAT_{A,IM,Cr,Pd} = SF\_WAT_{A,IM,Pd} + \sum_{Aq} GD\_WAT_{A,IM,Aq,Pd} \tag{4}$$

In this model, crop-yield functions are designed as non-linear approximations of the ratio between the maximum and actual evapotranspiration according to the definition proposed by FAO, derived from the Penman Monteith method (Allen et al., 1998; Samuelsson et al., 2015; Xin-e et al., 2015), which makes the yield depend on the water demand per hectare.

The observed data used in this study included, firstly, agronomic parameters, such as yield per crop, production factor requirements, crop areas, effective rainfall, maximum evapotranspiration, and the crop-yield response coefficient. Data relating to technical and hydrological parameters like loss rate of agricultural water, water demand in relation to agricultural area and farm type, regulated volume, evaporation and maximum volume of reservoirs, gradient, depth, permeability and storage coefficient of each aquifer were also considered by the model. Finally, socio-economic parameters were included for each agricultural area such as selling price of agricultural products, agricultural production input prices, selling price of irrigation water and farm economic efficiency rates.

In order to reflect the complexity of operating and valuation conditions of irrigation water, the proposed model was designed to be a recursive dynamic type that runs over 37 years (from 2014 to 2050). The period was limited because of the absence of agricultural policy and economic data forecasts beyond the year 2050 (Elbadraoui & Berdai, 2011; Lahlimi Alami, 2015; Sinan et al., 2009). The dynamic aspect of the model was done by defining a period of years (37 years) and a loop. This loop would make iterations, taking into account the parameters and variables that may be influenced by climate change and that are introduced in the set representing the year. We used two parameters (rainfall and temperature) for climate data from the recent results of experiments based on regional and international climate models that were available from the National Meteorology Directorate in Morocco (DMN), namely, the MIROC-ESM (Model for Interdisciplinary Research on Climate-Earth System Model) and CanESM2 (Canadian Earth System Model, 2nd generation). The results of these models show a decrease in rainfall for the two climate scenarios (RCPs 4.5 and 8.5) during the period covered by this study (Figure 3).
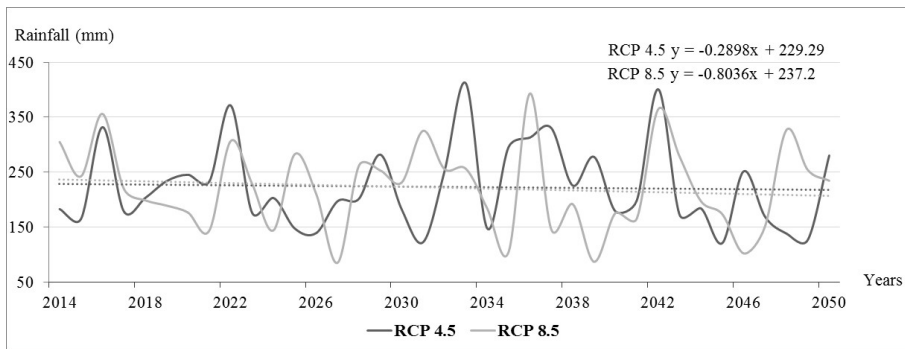
Figure 3. Annual rainfall trends in the Tadla sub-basin

These models leave little ambiguity about climate risks associated with increasing greenhouse gasses (Le Treut & Jancovici, 2004). Simulations conducted in this research involve the main crops in the study area under two climate scenarios, RCP 4.5 and 8.5.

As for reservoirs, water remaining in the first year (end of August) was used as the initial water quantity in the second year (beginning of September). Some other data were taken from agricultural development project reports, including the rate of conversion to water-saving systems established through the National Programme of Irrigation Water Economy (NPIWE) projects and projects focussing on crop expansion. The reference year chosen for this research was 2013-2014 from September to August (crop year) when the rainfall was about 410 (mm) according to data collected by the National Meteorology Directorate.

RESULTS AND DISCUSSION

Agricultural Profit

The average agricultural profit calculated for the Tadla sub-basin was 7486.57 MAD[1] per hectare. During the simulated years, this value showed a different evolution according to the climate scenarios, RCPs 4.5 and 8.5. The results of predicting this evolution are shown in Figure 4.
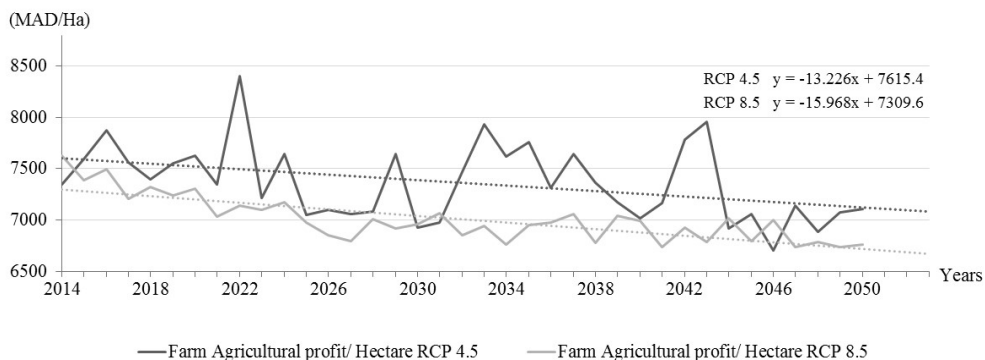

Figure 4. Predicted changes in agricultural profit according to the climate change scenarios, RCP 4.5 and RCP 8.5

[1] At the time of writing (mid-February 2017), MAD1.00 was approximately equivalent to £0.07, $US0.11 and €0.09.

The results showed that the agricultural profit per hectare in the study area was likely to experience a slight decline over the years decrease by an average of -13.23 MAD/Ha and -15.97 MAD/ha each year according to the RCP 4.5 and RCP 8.5 climate scenarios, respectively. The total agricultural profit calculated across the Tadla sub-basin was 2094 mil MAD. It also showed a declining trend (Figure 5).
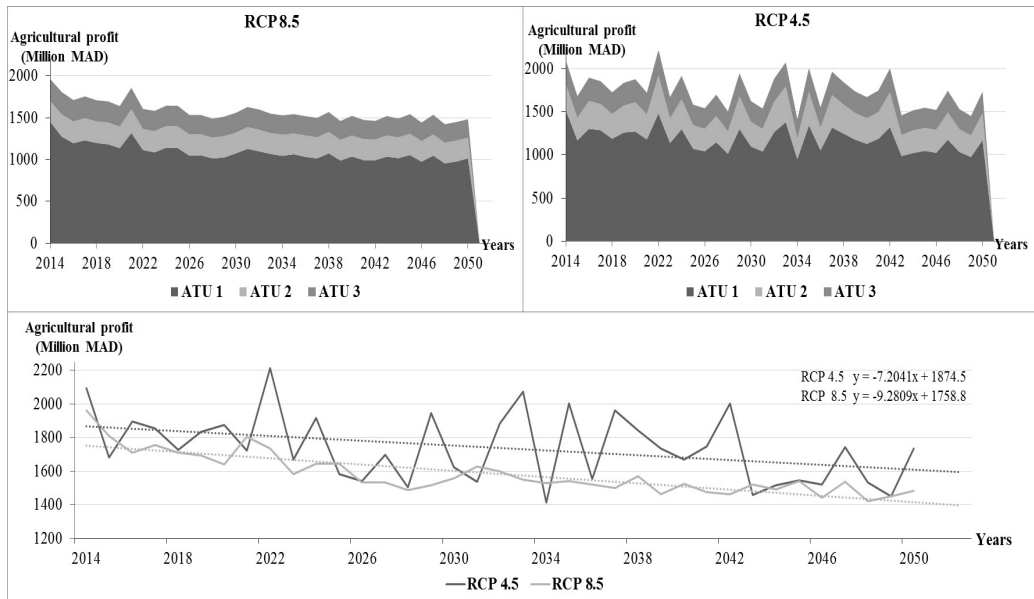

Figure 5. Predicted changes in total farm profit according to climate change scenarios, RCP 4.5 and RCP 8.5

This decline was quantified at about -7.2 million MAD (-0.34 %) per year according to RCP 4.5 and -9.28 million MAD (-0.44 %) according to RCP 8.5. This is mainly caused by the decrease in the cultivated areas that may have been affected by reducing irrigation water supplies predicted during the study period. In addition, we noticed existing differences between the agricultural territorial units and the evolution of their agricultural profit. The decline of its value was more pronounced at the first agricultural territorial unit (ATU1) because of the irregularity of surface water supplies through the irrigation system channels coming from the Ahmed El Hansali and Bine Elouidane dams. For the two other agricultural territorial units, (ATU2) and (ATU3), the decline was less important. In these agricultural territorial units, encouraged by the availability of groundwater throughout the year, farmers are diversifying and intensifying their agriculture.

Irrigation Water Shadow Price

Irrigation water shadow price is defined as the marginal increase in the value of the objective function (agricultural profit) if water availability is increased by an additional cubic metre.

This shadow price reflects the scarcity of water resources in contrast with the financial price. Thus, it represents one of the most significant results calculated in this study, which intended to assess the change in its value through the years under two climate scenarios (RCP 4.5 and RCP 8.5) (Figure 6).
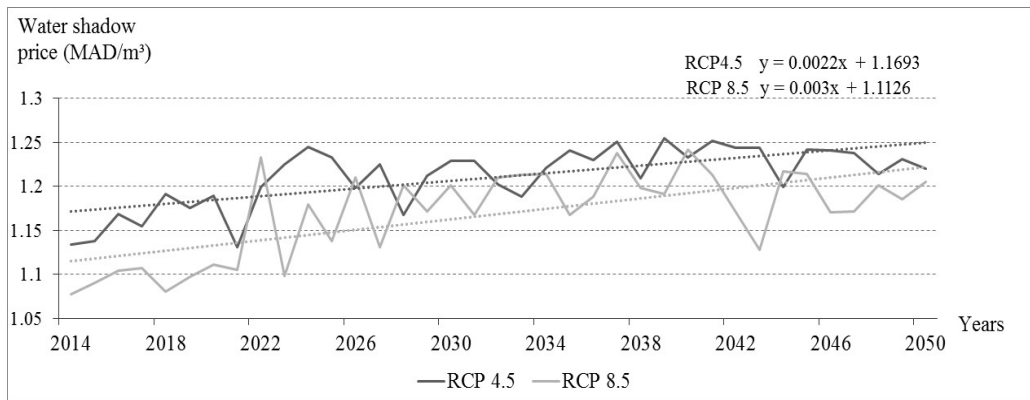

Figure 6. Predicted changes in the irrigation water shadow price according to the climate change scenarios, RCP 4.5 and RCP 8.5

The average shadow price of irrigation water calculated after calibration for the Tadla sub-basin was 1.13 MAD/m³ for the climate scenario RCP 4.5 and 1.08 MAD/m³ for RCP 8.5. These values were much higher than the selling price of irrigation water in the study area, which is 0.32 MAD/m³.

During the simulated years, the irrigation water shadow price was predicted to increase by 0.0022 MAD/m³ (+0.20 %) and 0.003 MAD/m³ (+0.28 %) per year according to the climate change scenarios RCP 4.5 and RCP 8.5, respectively. It could, therefore, reach a maximum value of 1.26 MAD/ m³ in 2039. This increase confirmed the results obtained by Heidecke and Heckelei (2010) in a relatively similar context of water availability. This result was linked to the irregularity of surface water supplies through the irrigation channels system coming from the Ahmed El Hansali and Bine Elouidane dams in the Tadla sub-basin, making irrigation water a production-limiting factor. In addition, it can be explained by the expected programmed intensification projects and by the orientation of farmers towards high-value crops.

Water Consumption

The total consumption of surface water calculated across the Tadla sub-basin was about 784 Mm³ on average in the reference year (2013/2014). The other part of irrigation water consumption is provided by groundwater, with a volume of 295 Mm³. The change in irrigation water consumption during the simulated years according to the climate change scenarios, RCP 4.5 and RCP 8.5, are shown in Figure 7.
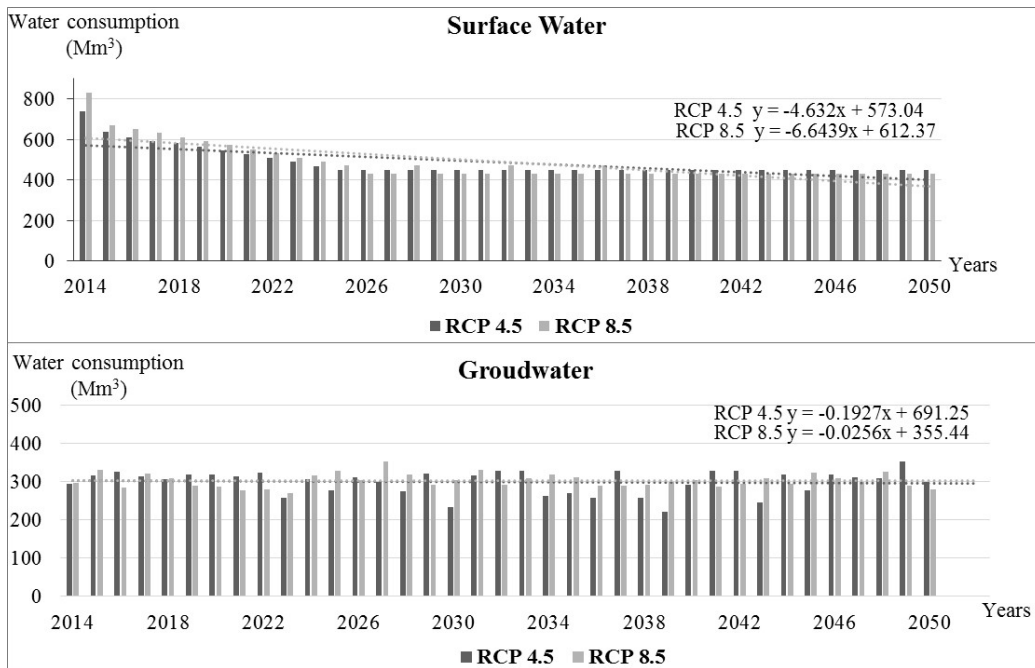
Figure 7. Predicted changes in total water consumption according to the climate change scenarios RCP 4.5 and RCP 8.5

From 2014 until 2050, the total consumption of irrigation water is likely to experience a significant decrease, particularly in the case of surface water. The decline is estimated at -4.63 mm³ per year according to the climate scenario RCP 4.5 and -6.64 mm³ for RCP 8.5. This is essentially linked to the reduction of cultivated areas caused by the expected decrease in surface water resources expected during the simulated period. Moreover, in order to cope with the expected reduction of surface water inflows in the Tadla sub-basin, current agricultural policies have proposed a set of measures, such as providing grants and launching extension programmes to encourage the adoption of water-saving systems.

Groundwater consumption will tend to stabilise over the years. Thus, the results show a slight decrease of -0.19 mm³ and of -0.03 mm³ per year according to the climate change scenarios RCP 4.5 and RCP 8.5, respectively. Indeed, encouraged by the availability of groundwater throughout the year, farmers tend to diversify and intensify their agriculture. Consequently, they reserve more areas to crops that offer good profit margins, whatever their water consumption.

Agricultural Land Use

The availability of climate change information will bring different agricultural stakeholders in the region to seek the best way to limit the negative consequences of this global phenomenon. In the Tadla sub-basin, cropping plans will vary during the simulated years. These changes are presented in Figure 8 through shares of major crops in the region according to the climate change scenarios, RCP 4.5 and RCP 8.5.
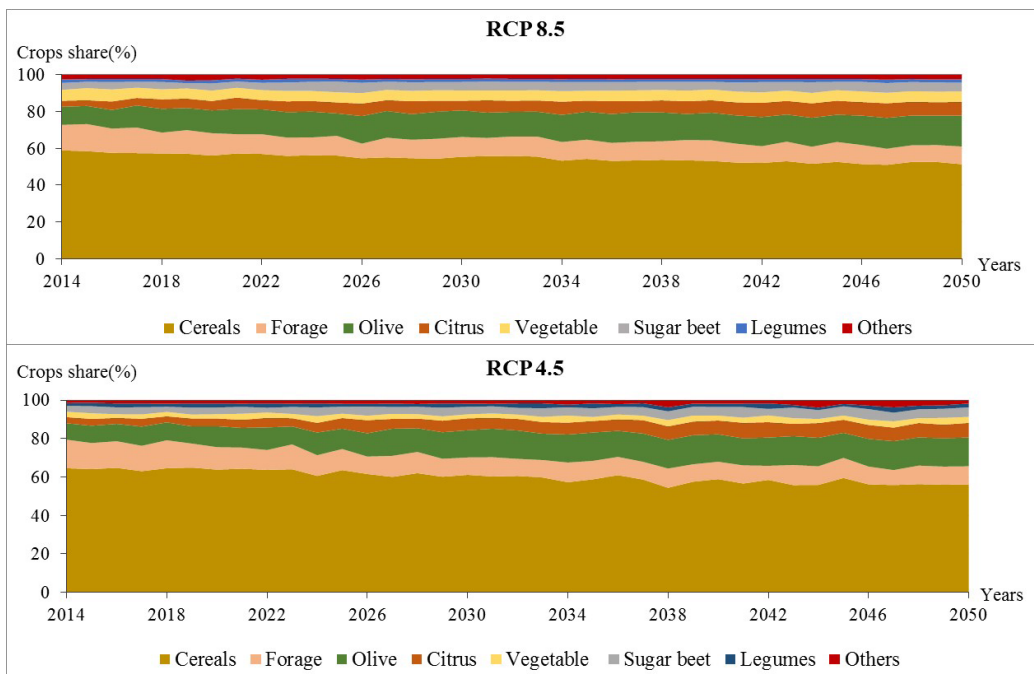
Figure 8. Predicted changes in the crops according to the climate change scenarios, RCP 4.5 and RCP 8.5

In the Tadla sub-basin, the reduction in water inflows will lead farmers to adapt to this new situation and thus, to optimise their choices by opting for crops that require less water, but offer good margins, in order to maximise their profits. Thus, an increase in areas allocated to tree crops, sugar beet and vegetables will be recognised at the expense of cereal and forage crops.

CONCLUSION

According to the two climate scenarios simulated in this study, RCP 4.5 and RCP 8.5, it was found that climate change may severely affect the agriculture sector in Tadla region. The total agricultural profit will have a declining trend (-0.34% per year, in the case of the climate scenarios, RCP 4.5 and -0.44%, for the RCP 8.5). This will be mainly due to the reduction in cultivated areas caused by the declining availability of water inflows during the simulated years. On its part, the irrigation water shadow price is predicted to increase by 0.20% and by 0.28% each year according to the RCP 4.5 and RCP 8.5 climate scenarios, respectively. This result will occur mainly because of the irregularity of surface water inflows, making irrigation water a limiting factor of production. The risk for farmers differs according to the technology used, the mode of access to irrigation water and the spatial location of agricultural land in the sub-basin. The climate change will also affect the irrigation water consumption that will undergo a significant decrease over the years. This can be explained by the reduction in land since many areas are left uncultivated during unfavourable years. The reduction of cultivated areas will lead farmers to optimise their choice by opting for crops that require less water and make significant margins to maximise their profit.

Finally, this study offers a scientific approach and a useful decision tool for improving the understanding of potential climate change impact in order to help policy makers to select the most appropriate adaptation policies for the Tadla sub-basin.

ACKNOWLEDGEMENT

REFERENCES

ABHOER, Oum Er Rbia Watershed Agency. (2012). *Study of the master plan for integrated water resources management in the Oum Er Rbia basin and the Atlantic coastal basins*. Ministry delegated to the Minister of Energy, Mines, Water and Environment - in charge of Water, Rabat, Morocco.

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration – Guidelines for computing crop water requirements. *FAO Irrigation and Drainage Paper 56*. Rome: Food and Agriculture Organization.

Baby, S., Yang, D., Mishra, D., & Sahu, N. C. (2014). Economic impact of climate change on agriculture sector of coastal Odisha. In *5th International Conference on Environmental Science and Development – ICESD* (Vol. 10, pp. 241-245).

Benabdelouahab, T., Balaghi, R., Hadria, R., Lionboui, H., Djaby, B., & Tychon, B. (2016). Testing aquacrop to simulate durum wheat yield and schedule irrigation in a semi-arid irrigated perimeter in Morocco. *Irrigation and Drainage, 65*(5)*,* 631–641.

Brooke, A., Kendrik, D., & Wilson, A. (1998). *GAMS: A user's guide*. Redwood City, Calif., USA: Scientific Press.

CESE. (2014). *Governance through integrated management of water resources in Morocco: fundamental lever for sustainable development*. Economic, Social and Environmental Council, Rabat, Morocco.

D'Agostino, D. R., Scardigno, A., Lamaddalena, N., & El Chami, D. (2014). Sensitivity analysis of coupled hydro-economic models: quantifying climate change uncertainty for decision-making. *Water Resources Management, 28*(12)*,* 4303–4318.

Elbadraoui, M. H., & Berdai, M. (2011). *Adaptation of the water-energy system to climate change: National study - Morocco*. Regional Activity Center, PNUE/PAM.

Esteve, P., Varela-Ortega, C., Blanco-Gutiérrez, I., & Downing, T. E. (2015). A hydro-economic model for the assessment of climate change impacts and adaptation in irrigated agriculture. *Ecological Economics, 120,* 49–58.

Falkenmark, M. (1989). The massive water scarcity threatening Africa – Why isn't it being addressed? *Ambio, 18*(2)*,* 112–118.

Heidecke, C., & Heckelei, T. (2010). Impacts of changing water inflow distributions on irrigation and farm income along the Drâa River in Morocco. *Agricultural Economics, 41*(2)*,* 135–149.

Howitt, R. E. (1995). Positive mathematical programming. *American Journal of Agriculture Economics, 77*(2)*,* 329–342.

Hurd, B. H., & Coonrod, J. (2012). Hydro-economic consequences of climate change in the upper Rio Grande. *Climate Research, 53*(2)*,* 103–118.

Iglesias, A., Garrote, L., Diz, A., Schlickenrieder, J., & Martin-Carrasco, F. (2011). Re-thinking water policy priorities in the Mediterranean region in view of climate change. *Environmental Science & Policy, 14*(7)*,* 744–757.

Immerzeel, W. W., van Beek, L. P. H., & Bierkens, M. F. P. (2010). Climate change will affect the Asian water towers. *Science, 328*(5984), 1382-1385.

Lahlimi, A. A. (2015*). Morocco between millennium development goals and sustainable development : achievements and challenges*. Rabat, Morocco: High Commission for Planning.

Le Treut, H., & Jancovici, J. M. (2004). *Greenhouse effect: Are we going to change the climate?* Scientific Research National Center, Polytechnic School, Pierre and Marie Curie University.

Lionboui, H., Benabdelouahab, T., Hasib, A., & Boulli, A. (2016). Analysis of farms performance using different sources of irrigation water: A case study in a semi-arid area. *International Journal of Agricultural Management and Development, 6,* 145–154.

Lionboui, H., Fadlaoui, A., Elame, F., & Benabdelouahab, T. (2014). Water pricing impact on the economic valuation of water resources. *International Journal of Education and Research, 2*(6), 147–166.

Martin, E., & Vaitkeviciute, J. (2016). Measuring the impact of climate change on Côte-d'Or agriculture. Économie rurale, 355, 21-48.

Mendelsohn, R., Nordhaus, W., & Shaw, D. (1994). The impact of global warming on agriculture: A Ricardian analysis. *American Economic Review of Agricultural Economics, 84,* 753–71.

Mishra, D., & Sahu, N. C. (2014). Economic impact of climate change on agriculture sector of Coastal Odisha. *APCBEE Procedia, 10,* 241–245.

Moriondo, M., Bindi, M., Kundzewicz, Z. W., Szwed, M., Chorynski, A., Matczak, P., … & Wreford, A. (2010). Impact and adaptation opportunities for European agriculture in response to climatic change and variability. *Mitigation and Adaptation Strategies for Global Change, 15,* 657–679.

Nepal, S. (2016). Impacts of climate change on the hydrological regime of the Koshi river basin in the Himalayan region. *Journal of Hydro-Environment Research, 10,* 76–89.

Niang, I., Ruppel, O. C., Abdrabo, M. A., Essel, A., Lennard, C., Padgham, J., & Urquhart, P. (2014). *Climate Change 2014: Impacts, Adaptation and Vulnerability*. Section B: Regional Aspects. Contribution of Working Group II to the Fifth Report. Cambridge University Press Cambridge, UK and New York, NY, USA, 1199-1265.

ORMVAT. (2014). *The Tadla region monograph*. Agricultural Development Office, Tadla, Morocco.

Samuelsson, P., Gollvik, S., Jansson, C., Kupiainen, M., Kourzeneva, E., & van de Berg, W. J. (2015). The surface processes of the Rossby Centre regional atmospheric climate model (RCA4). *Meteorologi, 157*, 42.

Sinan, M., Boussetta, M., & El Rherari, A. (2009). *Climate change: Causes and consequences on climate and water resources*. HTE Review, 142.

Toumi, L. (2008). *The New Agricultural Strategy in Morocco, Green Plan: The Keys to Success*. Government document, Rabat, Morocco.

Ventrella, D., Charfeddine, M., Moriondo, M., Rinaldi, M., & Bindi, M. (2012). Agronomic adaptation strategies under climate change for winter durum wheat and tomato in southern Italy: Irrigation and nitrogen fertilization. *Regional Environmental Change, 12,* 407–419.

Xin-e, T., Hua, C., Chong-yu, X., Yu-kun, H., & Meng-xuan, J. (2015). Analysis and prediction of reference evapotranspiration with climate change in Xiangjiang River Basin, China. *Water Science and Engineering, 8,* 273–281.

Integrated Rainwater Drainage System for Groundwater Improvement and Economic Benefit

Manisha D. Desai* and Jayantilal N. Patel

Department of Civil Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, India

ABSTRACT

Groundwater is a precious resource, both for yield and quality. This paper demonstrates the imaginative, progressive and innovative approaches of simulated renovations giving substantial advantages to the general public. Utilising groundwater to stimulate wells, in conjunction with a rainwater conservation framework, is proposed in this review. The unique plan and cost correlations as well as groundwater quality change are examined for the procedures of the tempest water seepage framework with and without reviving wells. The region chosen for study is Bhimrad, a new urban centre in Surat City in Gujarat State, India. Two rainwater structures are considered: (1) a traditional rainwater water drainage system excluding groundwater revive wells, and (2) a rainwater drainage system including groundwater revive wells. This paper shows that due to determination of the optimal diameter of revive wells, cost saving in the modified system is possible, along with improvement in groundwater contamination. The use of a modified rainwater drainage system would economically benefit the SMC (Surat Municipal Corporation) by conserving 25.43% of their reserve funds.

Keywords: Economic benefit, groundwater contamination, optimal diameter, rainwater conservation

INTRODUCTION

The establishment and design of common assets are a fundamental part of managing the economy for the success of human life (Schirmer, 2013). As indicated by Kellagher (2008), the 21st century has a new way of dealing with configuring rainwater seepage, utilising maintainability pointers to demonstrate and to gauge framework execution instead of outline criteria to meet a base level of administration. In previous research, the majority of software and models were developed based on basic rainfall-runoff relationship. Many urban runoff models have been developed and verified. Research today focusses on sustainable development

of rainwater management approaches to reduce the damage caused by impervious cover and to maximise the infiltration rate by using techniques like LID, BMP and WSUS.

Following this concept, research and surveys were conducted for sustainable planning of rainwater drainage systems in the selected research area, Bhimrad in Surat, India, and a plan was proposed for the new urban area. A novel integrated approach rainwater drainage system conjunction with groundwater recharge wells is used in the present research.

The objectives of this research are:
(1) To study a traditional rainwater drainage system that does not use a revive well;
(2) To modify planning of a rainwater drainage system that uses a revive well;
(3) To provide an economic solution for the rainwater drainage system of Bhimrad, and;
(4) To present a quality prediction system for groundwater in the future using an artificial revive well.

METHODOLOGY

Study Sites

Surat City, settled on the banks of the Tapi Basin near the estuary of the Arabian Sea, is the eighth greatest city of India and is a leading business centre, considering its valuable diamond and textile endeavours. The drainage systems established in Surat City were notably poor in the past, and during the months of heavy rain, various areas in Surat City experienced transient flooding and blockage of rainwater. Precipitation from June to October is dynamic, with a typical reading of 1143 mm. Surat City experienced uncontrollable floods in 1998, 2006 and late 2013. A modified rainwater drainage system should be provided in order to reduce water logging and flooding issues. Also, revived groundwater would meet the future requirements of new urban areas that are developed in Surat City. This paper reviews the issues relevant to this topic and provides choices derived from research for methods to update living conditions for the general public located in urban zones such as Bhimrad. It also provides configurations to decrease the level of the water table to enhance groundwater contamination at a lower cost.

Method of Designing a Traditional Rainwater Drainage System Using GIS

Firstly, a traditional rainwater system was designed as part of this study, with medium flow and maximum discharge of each watershed. These were determined using rational methods. The hydrological response was significantly influenced by land cover and the changes in land cover due to time affected floods were assessed (Nutchanart, 2011). It is important to decide the correct rate for impenetrable surfaces and land when utilising cover mapping in ARC GIS (Boulos, 2005). This decision is based on several GIS applications that incorporate technologies to utilise effective management for urban rainwater collection systems (Bryant, 2000). On that premise, the catchment zone of every drainage and pipe system was outlined in light of superimposing the current shape of the review zone in ARC GIS and DEM (Digital Elevation Model) by utilising the hydrology application in GIS, which naturally creates stream lines, and the introduction of the channels and catchment regions of each proposed depletion was

surveyed for adequacy. For the engineering design of various components, this traditional rainwater drainage system was prepared (CPHEEO, 2013). The peak streams can be utilised using Manning's formula to acquire the measurements for each depletion section (Harpalani, 2013). This is an iterative procedure to guarantee deplete measurements, and subsequently, it can withstand the peak overflow computed.

As the surface of drains deteriorates with the passage of time, a roughness coefficient was measured for the design period, assuming reasonable conditions in drains. The roughness coefficient `n' was believed to be similar for every channel and was given a value of 0.013 for the outline.

Method of Designing a Rainwater Drainage System Combined with a Revive Well

A sustainable rainwater drainage system combined with groundwater revive wells was proposed and discussed by Nolan (2006), including factors influencing groundwater recharging. Rainwater drainage system designs are based on engineering analysis, which takes into consideration artificial groundwater revive, runoff rates, pipe-flow capacity, hydraulic grade lines and discussions about experimental investigations for particle breakage using natural sand (Hattamleh, 2013). Following this, dirt attributes (penetrability) are completed with molecule estimate dissemination, with the assistance of a hydrometer test (Carrier III, 2003). The century-old Hazen formula was used to gauge permeability on the basis of soil particle size (D_{10}). Hydrogeological and geophysical parameters for more profound groundwater assets were recorded with the assistance of the electric resistivity technique in the examination zone (D. R. Kumar, 2012). Soil parameters were established by determining the steady-state infiltration rate using a single-ring infiltrometer and double-ring infiltrometer (Neris, 2012). Outlining of the casing pipes, which drive water into the aquifer, is a critical component of the bore well. Its distance across is kept shorter than the width of the bore well to suit gravel pressing in the annular space between the casing pipe and the bore well.

The drag of the revive well is made 5 cm bigger than that of the energise pipe, facilitating the lowering of the pipe. Thus, for a casing of size 20 cm, a minimum bore of 25 cm is necessary. The measure of the gravel pack should be a thickness twice that of the traverse of the gravel. The size of a tube well drilled with a reverse rotary gear should be about 30 cm in diameter. On the screen, the section is 20 cm; therefore, the thickness of the gravel pack will be 5 cm up to the base of the pipe. The design of the well screen is the most fundamental piece of the revive well. Small or no enhancement in well efficiency results in open areas greater than 25%, whereas efficiency falls rapidly as the open area decreases to 10%. Hence, it may be concluded that other than slotted pipes it is desirable to provide an open area of about 20% for well screens (Athens, 1970).

The separation crosswise should ensure that the extent of the opening is available on the screen for a stream of water. Entrance velocity of the screen should not be more than 3 m/sec to prevent clogging of the screen. In areas where adequate sand thickness is not available, a minimum velocity of 5 cm/sec is permitted. On the off-chance that an event of the homogeneous water table aquifer emerges, the last one third of the aquifer is screened. In the case of the homogeneous artesian aquifer, 75-90% of the thickness of the water-bearing sand

ought to be utilised for screening. For an aquifer thickness of less than 8 m, a screening of 75% is satisfactory. No less than 30 cm of aquifer profundity at the top and base of the screen ought to be left unscreened to protect against a mistake if there should be an occurrence of the arrangement of the revive well screen amid the establishment of the opening. To prevent clogging, the minimum length of the well screen for a non-gravel pack well is determined by:

$$H = \frac{Q_0}{A_0 V_\epsilon}$$ [1]

In the above equation, 'H' is the minimum length of well screen in metres, 'Q_o' is the maximum expected discharge capacity of well screen in m^3/min, 'A_o' is taken as the effective open area per metre length of the well screen, m^2, and 'V_e' is the entrance velocity at the screen, m/min.

This equation is used to compute the length of the screen in a gravel pack. The normal estimation of the permeability of the aquifer and the gravel pack is utilised to decide the passageway speed of the screen. The length of the screen provided on a revive well depends on the thickness of the aquifer available. When the permissible value of entrance velocities is greater than the result of excessive recharging, the well is a failure due to the design of the revive well screen (Walton, 1970). A design of a well screen chamber is provided to check the fine soil particles moving in the rainwater flowing into the drain. The water, after screening, is allowed to enter the revive well pipe (Kamra, 2012). In the sand-based rainwater filtration system for groundwater revive wells, the screen chamber includes different layers of sand, gravel and brickbats to remove suspended impurities from the rainwater (S. K. Kumar, 2012).

Cost Analysis

The procedure has been discussed above; correlations of the cost appraisals are now made, in light of the necessity for a new urban range in Surat City as T. P. - 42 and 43 – Bhimrad. The quantity of material was calculated and the SOR (Schedule of Rate) of (GWSSB-SOR, 2014-2015) was used.

Method for Groundwater Quality Analysis

For water quality index, groundwater quality parameters were analysed (Dohare, 2014). Groundwater samples were collected from seven locations in the study area during the pre-monsoon and the post-monsoon period of 2015. The collected samples were tested in the laboratory and analysed for concentration of different parameters in water quality for drinking purposes. To acquire data from previous years, different interpolation techniques of GIS were used, as suggested by Garnero (2013). The method includes information about artificial revive wells followed by quantifying the change in chemical composition of groundwater for the study area (Sharma, 2008). In conjunction with improving the groundwater quality with an artificial revive well, it is very much necessary to analyse the behaviour of the aquifer during this recharging process (Sharma, 2011). Future prediction of groundwater quality was determined using GIS software along with a dilution equation, as follows:

$$C_3 = \frac{C_1 V_1 + C_2 V_2}{V_1 + V_2}$$ [2]

where, C_3 is the changed concentration of a particular groundwater quality parameter in a water sample after artificial recharging and V_1 is the quantity of water available in the well. This amount of water is added to the well through artificial revive, and C_2 is the grouping of fresh water (rainwater) parameters. An ERDAS Model was prepared in GIS software to determine groundwater quality dilution from 2015-2050. Future prediction of groundwater quality was made from 2015 to 2050 and change in groundwater quality assessed to arrive at the Groundwater Quality Index (GWQI) using GIS (Gorai, 2013). This Groundwater Quality Index can be computed using the ERDAS Modeller in GIS, as it is a powerful tool for modelling water quality.

RESULTS AND DISCUSSION

Traditional Rainwater Drainage System

To compute impervious surface area with the help of land use and land cover mapping in GIS. To make impenetrable land cover, a Google map of the review territory (Bhimrad) was accessed through ArcMap's implicit base guidework. The streets, structures, waterways, springs, water bodies and other existing areas utilised were found by digitising the Google map. A land use impact boundary shape file was created for the study area, shown in Figure 1.
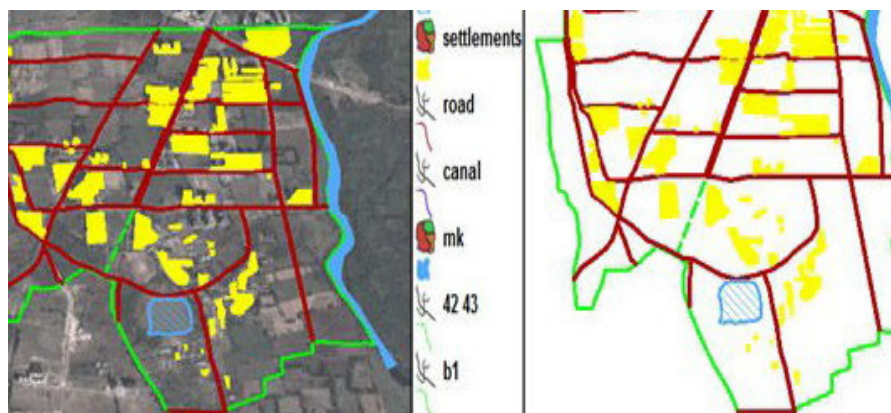

Figure 1. Digitising shape file of impervious land cover

Rainwater drainage network in GIS. Catchments and pipe networks were delineated based on their flow direction, flow accumulation, stream link, conditional stream, stream order and stream shape, as shown in the figure below, and are all derived from a digital elevation model in Arc GIS as per Figure 2. Since rainwater runoff from inverse sides of a stream can encounter diverse conditions, catchments were additionally subdivided by the seepage organise layer,

enabling arrival on each side to be analysed independently in reasonable investigations. Detailed drawings of rainwater drainage networks with catchments are shown in Figure 3. Flow direction is based on a digital elevation model Figure 2 with stream locations imposed on it.
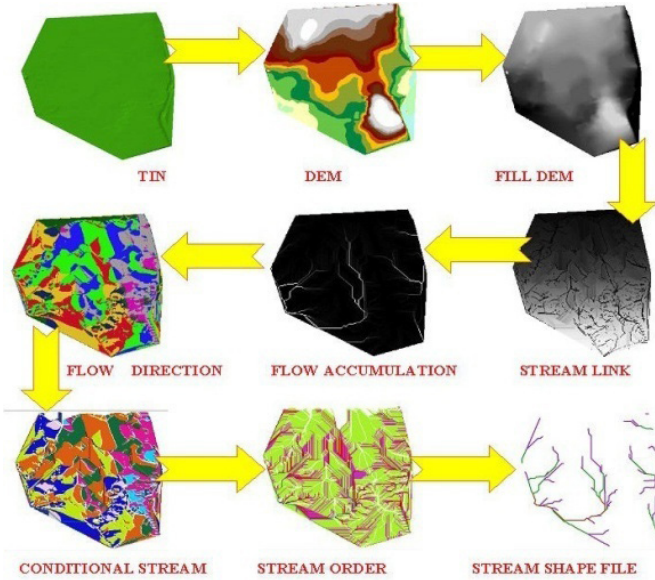

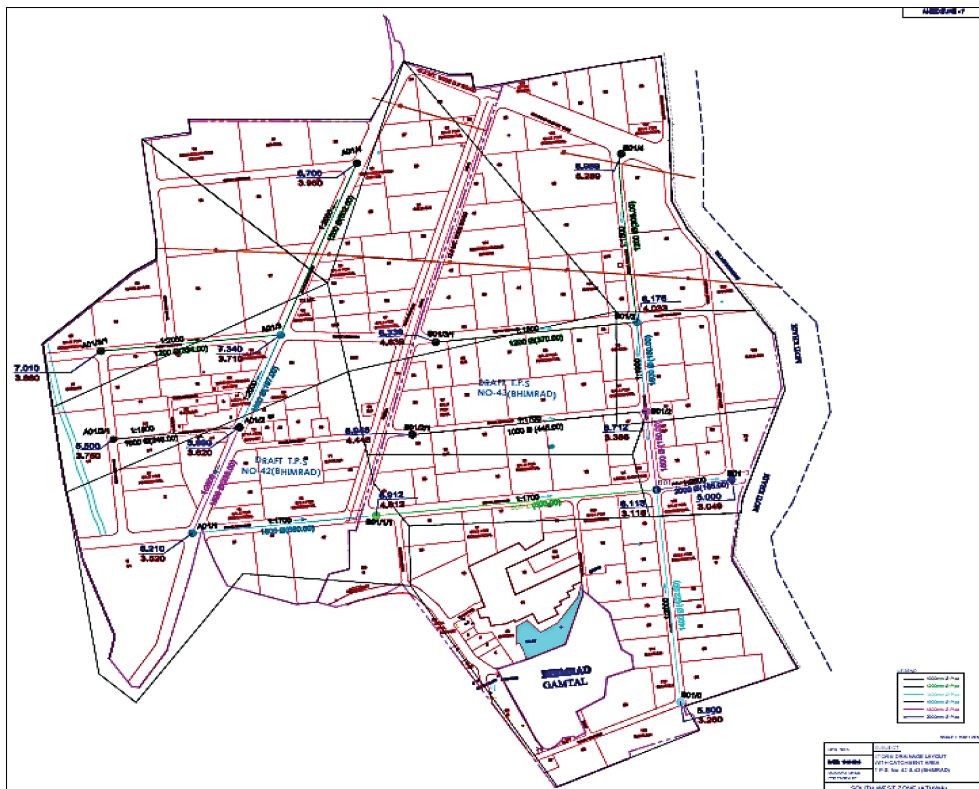*Figure 2*. DEM and hydrology application results in ArcGIS

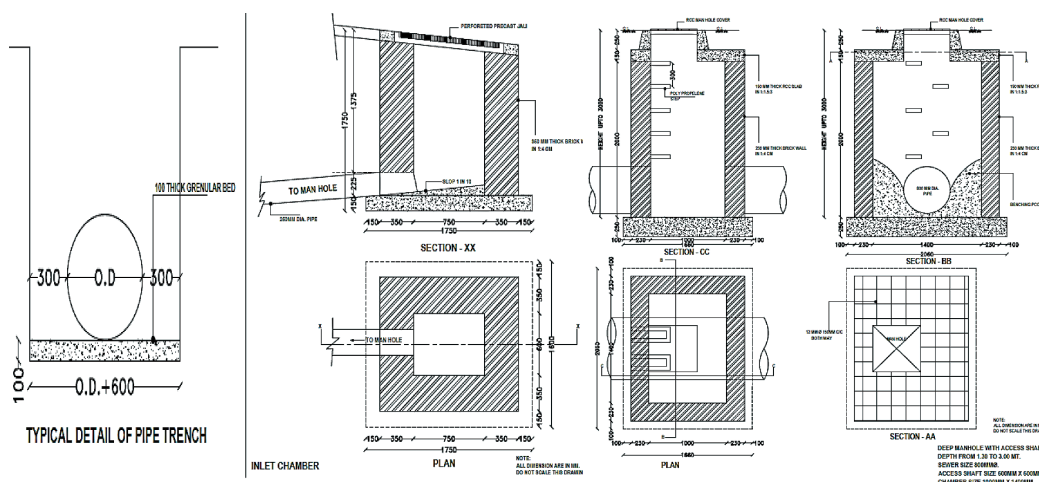*Figure 3(a)*. Storm water drainage network with detail

Figure 3(b). Structural details of Storm water drainage system

Engineering design of traditional rainwater drainage system. The layout of the storm drainage system is formulated in such a way that a maximum area can be drained out. The proposed drains are generally laid along roads, as marked on the town-planning scheme of this area (Basu, 2012). To facilitate the inlet chamber connection, the drains are provided with the minimum initial depth, depending on the site conditions. The diameter of the drains varies from 800 mm to 2000 mm. Thereafter, RCC double ducts of 2.2 m X 1.2 m to 3.5 m X 1.2 m sizes are proposed, shown in Figure 3(a) and Figure 3(b). The depth of drains at the outfall is kept above R.L. 3.0 m to minimise back flooding from the creek into the drain, based on studies estimating flooding and its control by section modification in Mithi Creek at Surat (Jariwala, 2012). The design of the traditional rainwater drainage system is given in Table 1.

Table 1

Design of conventional storm drainage system for TP-42 & 43 – Bhimrad, Surat, India

Section		Cum. Length of Drain Section	Total Increment	Estimated Time of Concentration (t_c)				Average Rainfall Intensity from Graph (I_c)	Runoff Coefficient from Graph (C)	Actual Runoff (q)	Profile Shape & Size		Slope of Drain	Flow Full (Q) = A.V	Full Flow Velocity (v)
U/S Side	D/S Side			Eq. 100% Imp. Area	Time of Inlet (t_i)	Time of Flow (t_f)	(t_c)				Pipe Dia Provided	No of Pipes			
		(mt)	hectare	hectare	(minutes)	(minutes)	(minutes)	(mm/hr)		(m³/s)	mm		1 in L	(m³/s)	m/s
AO1/4	AO1/3	502	98381	60996	0.00	9.301	9.301	99	0.473	0.80	1000	1	1400	0.76	0.96
AO1/3/1	AO1/3	335	120677	74820	0.00	6.198	6.198	119	0.315	0.78	1200	1	2000	1.03	0.91
AO1/3	AO1/2	699	219058	135816	9.30	12.942	22.243	66	0.599	1.50	1200	1	1400	1.23	1.09
AO1/2/1	AO1/2	245	99817	61886	0.00	4.542	4.542	143	0.210	0.52	1000	1	1700	0.69	0.88
AO1/2	AO1/1	934	318875	197702	22.24	17.295	39.538	51	0.677	1.91	1200	1	1200	1.33	1.18
AO1/1	BO1/1/1	1264	539806	334680	39.54	23.411	62.949	42	0.744	2.88	1200	2	1800	2.17	0.96
BO1/1/1	BO1/1	1764	605770	375577	62.95	32.671	95.620	34	0.804	2.89	1200	2	2000	2.06	0.91
BO1/4	BO1/3	305	148101	91822	0.00	5.648	5.648	129	0.263	0.87	1000	1	1200	0.82	1.04
BO1/3/1	BO1/3	370	170514	105719	0.00	6.852	6.852	119	0.315	1.10	1200	1	1800	1.09	0.96
BO1/3	BO1/2	550	382985	237451	6.85	10.185	17.037	75	0.569	2.80	1200	1	600	1.88	1.66
BO1/2/1	BO1/2	445	114679	71101	0.00	8.241	8.241	105	0.420	0.87	1000	1	1200	0.82	1.04
BO1/2	BO1/1	700	497664	308552	17.04	12.963	30.000	58	0.642	3.18	1200	1	450	2.17	1.92
BO1/0	BO1/1	422	261922	162392	0.00	7.815	7.815	111	0.368	1.84	1200	1	1200	1.33	1.18
BO1/1	BO1	1929	1409404	873831	95.62	35.726	131.346	30	0.837	6.05	1200	2	500	4.12	1.82

Design of Rainwater Drainage System with a Revive Well

Results of geological strata for groundwater level with the help of electric resistivity experiment. In this exploration, an unconfined aquifer was planned as a revive well for which Wenner's technique for electrical resistivity was used. The actual geological strata of a bore well of the study area were classified as per IS 1498: 1970 (Reaffirmed, 1997). The electrical resistivity method was extremely useful for determining the average condition of different strata up to a depth of 30 m or more. The depth of groundwater can be anticipated by utilising this test for the depth of the aquifer and the type of aquifer (confined or unconfined). Winner's method for conducting an electrical resistivity survey was launched. Five resistive zones were distinctly outlined (Point 1: N- S; Point 2: N-S; Point 2: E-W; Point 2: NE-SW; Point 3: N-S). The results were computed for electric resistivity from Figure 4 (Winner's Method IS-3043), and determined for a deeper geological condition. It was not possible to drill a bore well up to 30-40 m. Alongside this investigation, estimations were made for a depth of up to 80 m for the geographical strata and unconfined aquifer for a revive of groundwater. The overview obtained from the electric resistivity test was utilised for areas of various strata in groundwater. Glacial till was defined as an unconsolidated, heterogeneous mixture of clay, sand, pebbles, cobbles and boulders, as shown in Table 2. Thus, the electrical resistivity method was useful for obtaining a soil profile with reasonably reliable results.
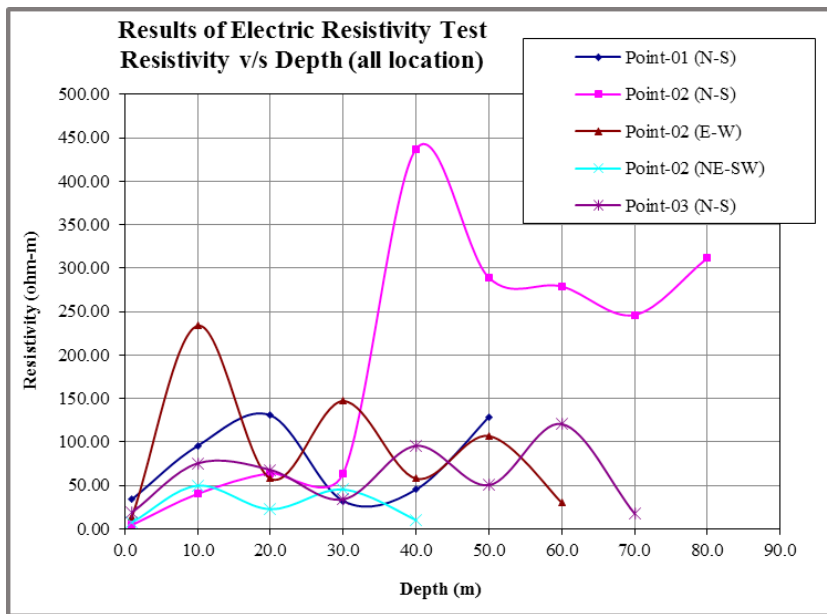

Figure 4. Results of electric resistivity test (all location)

Table 2
Identification of soil from resistivity

Sr. No.	Depth from GL	Electrical Resistivity (Ohm.m.)	Visual Identification of Soil
1.	1.0	4.40	Clay (Saturated) and Silt
2.	10.0	40.21	Clay (Saturated) and Silt
3.	20.0	64.09	Clay (Saturated) and Silt
4.	30.0	64.09	Clay (Saturated) and Silt
5.	40.0	437.31	Glacial Till (Clayey Sand, Pebbles, Cobble, & Boulders)
6.	50.0	289.03	Glacial Till (Clayey Sand, Pebbles, Cobble, & Boulders)
7.	60.0	278.97	Glacial Till (Clayey Sand, Pebbles, Cobble, & Boulders)
8.	70.0	246.30	Glacial Till (Clayey Sand, Pebbles, Cobbles, & Boulders)
9.	80.0	311.65	Glacial Till (Clayey Sand, Pebbles, Cobbles, & Boulders)

Result of value of permeability using Hazen's formula with the help of hydrometer test of soil. By checking the test samples, a summary of grain size distribution parameters could be provided under a dominant soil type (clay, silt, sand). The typical curves are shown in Figure 5, Figure 6 and Figure 7. Unlike for the most common soil type, gradation curves for layers with higher silt and clay content exhibit different shapes from the dominant soil. The finest 10% of the material from testing at various depths of the borehole is derived from this graph. The after-effect of R.L. -9.00 to -9.45 m and the estimation of D_{10} of 0.000039 were acquired. Another two-depth test indicates a rate better than 10% and appears in Figure 5. Three samples of different depths were analysed and the results are shown in Tables 3a, 3b and 3c. The different depths were R.L. -1.5 to 1.95, R.L. -4.5 to -4.95 and R.L. -7.0 to -7.45, respectively in which R.L. -4.50 to -4.95 had 10% finer materials than did the other samples. The other two samples had a particle size not more than 10%. The value of D_{10} taken from R.L. -4.50 to -4.95 is 0.000116 is shown in Figure 6. The third borehole was provided on the opposite side of the creek, located in the northern side of the study area. A hydrometer test investigation was done from three distinct depths of the borehole opening, the depths of which were about R.L.4.5 to -4.95, R.L. -7.5 to -7.95 and R.L. -13.0 to -13.45. After carrying out the analysis and calculations, the graph of particle size to percentage finer was plotted, and 10%

finer particles of the value D_{10} was marked and used in Hazen's formula to find the hydraulic conductivity. From the examination, two D_{10} qualities were obtained, one from the top strata at about R.L. -4.5 to -4.95 of D_{10} around 0.000027 and another from the depth of R.L. -13.0 to -13.45 of D_{10} at around 0.000802, shown in Figure 7.
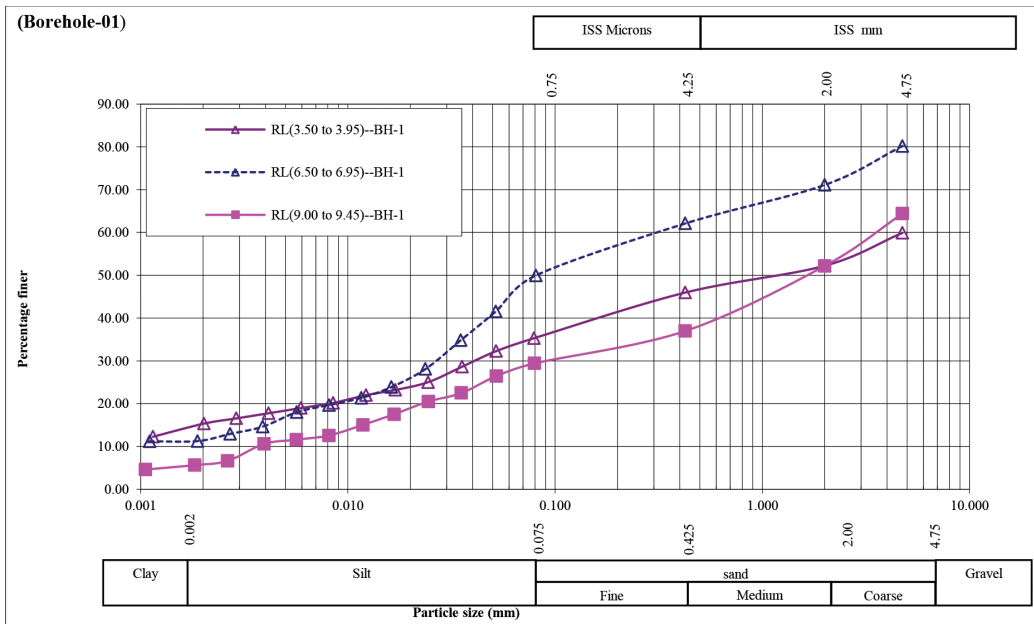

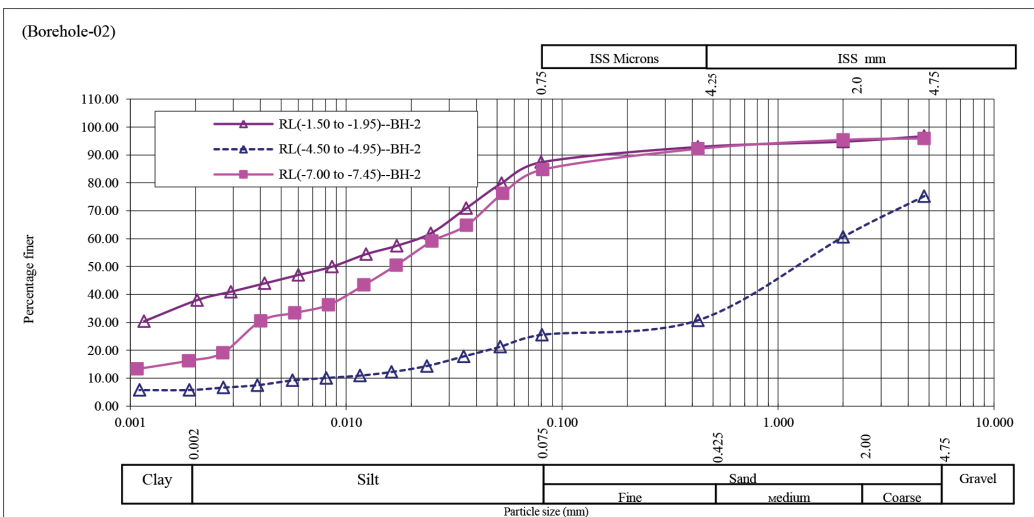*Figure 5*. Particle size distribution curve (sieve and hydrometer analysis – Borehole no. 1)

Figure 6. Particle size distribution curve (sieve and hydrometer analysis – Borehole no. 2)
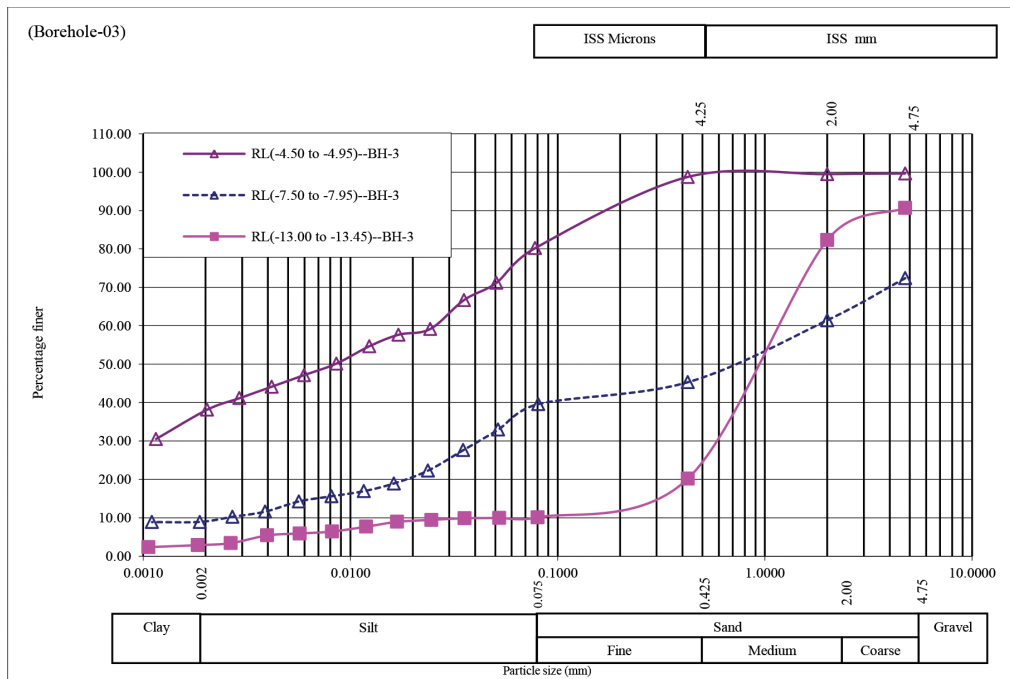
Figure 7. Particle size distribution curve (sieve and hydrometer analysis – Borehole no. 3)

Estimate of Permeability Using the Hazen Formula

From the above results, three different boreholes were found to have a 10% finer grain particle size of D_{10} value, conducted at a different level below the ground. A total of four numbers of D_{10} values were obtained, from which two D_{10} values were of the third borehole. A permeability test was conducted (k) using the Hazen formula from the particle size distribution test (Hazen, 1892; Vuković 7 Siro, 1992; Doing, 2007). From the above outcomes, a high penetrability consequence of 7.29 x 10^{-8} at the top strata of R.L. -4.5 to -4.95 was obtained, from which the separation of the creek was vast. The same borehole depth of R.L. was about -13.00 to -13.45, and had less porosity after the effects of 6.43 x 10^{-5} than different areas taken for a decision to release the revive well (Patel & Desai, 2011).

Result of infiltration rate using the single- and double-ring infiltrometer test. Calculation by Horton's method (for single-ring infiltrometer)

$$F = f_c + (f_o - f_c)e^{-kt} \qquad [3]$$

Initially, the results showed the infiltration rate of a single-ring infiltrometer rate as 26.3 cm/hr and a double-ring infiltrometer rate of 21.5 cm/hr. After an hour, it demonstrated estimations of around 3.29 cm/hr and 1.82 cm/hr individually, and after two and a half hours, it indicated rates of around 1.34 cm/hr and 1.10 cm/hr. Based on the outcomes appearing in Table 3 and Figure 8, and the literature on this topic, it can be inferred that the reviewed region has a great penetration rate, which can help for artificial groundwater revive in unconfined aquifer (Patel, 2011).

Table 3
Results of horton's equations for single-ring infiltrometer and double-ring infiltrometer

Time	Single-Ring Infiltrometer	Double-Ring Infiltrometer	Result by Horton's Equation for SRI	Result by Horton's Equation for DRI
0	26.3	21.5	26.3	21.5
0.25	14.7	12.3	14.58242065	9.951805745
0.5	9.3	7.9	8.356907931	4.940905145
0.75	5.8	3.8	5.049312785	2.766615011
1	3.4	1.9	3.291997982	1.82316433
1.25	2.9	1.2	2.358342205	1.413789714
1.5	2.2	1.1	1.862293854	1.236157137
1.75	1.6	1.1	1.59874494	1.159080222
2	1.3	1.1	1.458722238	1.12563562
2.25	1.3	1.1	1.384328621	1.111123604
2.5	1.3	1.1	1.344803529	1.104826666
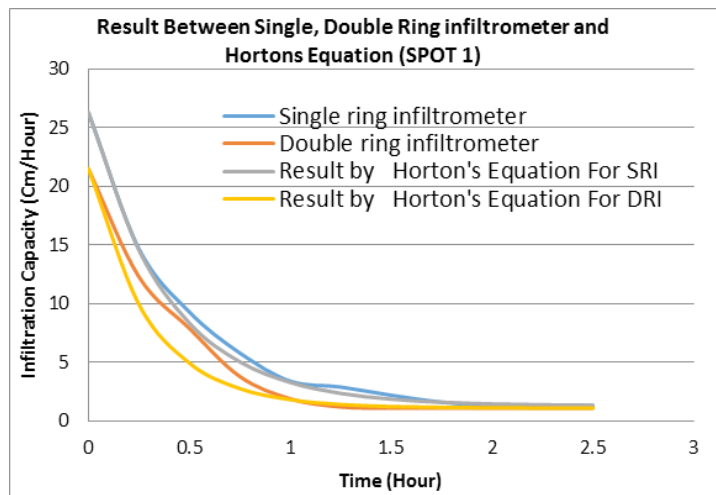
Figure 8. Graph of infiltration rate by SRI and DRI

Results of modified design of the rainwater drainage system with a revive well. In this exploration, distinctive geotechnical parameters and their effect on artificial groundwater revive methods utilised for urban foci are viewed (Patel, 2011). The configuration is for an unconfined aquifer as a revive well, with an abutting rainwater drainage framework. From a

past study on electric resistivity exploration, the unconfined aquifer depth h_2 was taken as 35 m and depth of h_1 was taken as 20 m. The hydrometer test determines the value of D_{10} in grain size analysis and permeability $k = 6.43 \times 10^{-5}$ m/sec is determined from the Hazen formula. Radius of influence R was taken as 100 m, with a calculating revive rate Q and sample calculation of Artificial Revive as Unconfined Aquifer:

$$Q = \pi k \frac{(h_2 - h_1)}{ln(\frac{r_2}{r_1})}$$ [4]

where, 'Q' is the rate of water entering the revive well, 'h_2' is the depth of water in the revive well above the impervious stratum, 'h_1' is the depth of the water table in an unconfined aquifer, 'r_2' is the radius of influence and 'r_1' is the radius of the well.

For all different sizes (diameters) of revive wells, the results of revive rate Q are shown in Table 4. These different sizes (diameter) of revive wells are used to determine the optimum size diameter of revive wells that can be more beneficial than others. Based on these outcomes, when the distance across the revive well expands, the revive rate likewise increases. By having two revive wells (with and without rainwater drainage frameworks) in every catchment area of the rainwater drainage frameworks, the revive rate is expanded. The rainwater drainage system is modified as, due to the provision of a revive well, the actual runoff of the rainwater drainage system of the catchment area decreased. In the outline, two revive wells in every catchment territory of rainwater depletion was given. One well was provided on the upstream side of each rainwater drain and another one on the perpendicular road where the rainwater drain was not available. Modified design of the rainwater drainage system was as per Table 5-6 and Figure 9. A detailed cross-section of the revive well, which provides a rainwater drainage system, is shown in Figure 10.

Table 4
Results of recharge rate for different diameters of well

Sr. No.	Diameter of Recharge Well m	Radius of Recharge Well (R) m	Recharge Rate (Q) m^3/sec
1.	0.15	0.075	0.0231565
2.	0.225	0.1125	0.0245395
3.	0.25	0.125	0.024926
4.	0.30	0.15	0.025625
5.	0.35	0.175	0.026247
6.	0.45	0.225	0.027329
7.	0.50	0.25	0.0278095
8.	0.60	0.3	0.0286825

Table 5

Modified design of the storm water drainage system with recharge well (rw dia 15 cm-30 cm)

Section		Length of Drain	Actual Runoff (q)	Old Diameter of Pipe (D)	15 cm dia			22.5 cm dia			25 cm dia			30 cm dia		
U/S Node	D/S Node				Well Discharge (Qw)	Cum. Discharge (q')	Pipe Dia Provided	Well Discharge (Qw)	Cum. Discharge (q')	Pipe Dia Provided	Well Discharge (Qw)	Cum. Discharge (q')	Pipe Dia provided	Well Discharge (Qw)	Cum. Discharge (q')	Pipe Dia Provided
			(m³/s)	(mm)	(m³/s)	(m³/s)	(mm)	(m³/s)	(m³/s)	(mm)	(m³/s)	(m³/s)	(mm)	(m³/s)	(m³/s)	(mm)
AOI/4	AOI/3	502.27	0.795	1000	0.069469	0.7258	800	0.0736	0.7216	800	0.0748	0.7205	800	0.0769	0.7184	800
AOI/3/1	AOI/3	334.68	0.780	1200	0.069469	0.7110	800	0.0736	0.7069	800	0.0748	0.7057	800	0.0769	0.7036	800
AOI/3	AOI/2	196.58	1.501	1200	0.069469	1.2925	1000	0.0736	1.2800	1000	0.0748	1.2766	1000	0.0769	1.2703	1000
AOI/2/1	AOI/2	245.25	0.517	1000	0.069469	0.4471	600	0.0736	0.4429	600	0.0748	0.4418	600	0.0769	0.4397	600
AOI/2	AOI/1	235.06	1.909	1200	0.069469	1.4917	1000	0.0736	1.4668	1000	0.0748	1.4599	1000	0.0769	1.4473	1000
AOI/1	BOI/1/1	330.31	2.883	1200*2	0.069469	2.4666	800	0.0736	2.4417	800	0.0748	2.4348	800	0.0769	2.4222	800
BOI/1/1	**BOI/1**	500.00	2.885	1200*2	0.069469	2.3990	800	0.0736	2.3700	800	0.0748	2.3618	800	0.0769	2.3472	800
BOI/4	BOI/3	305.00	0.866	1000	0.069469	0.7896	800	0.0736	0.7928	800	0.0748	0.7917	800	0.0769	0.7896	800
BOI/3/1	BOI/3	370.00	1.103	1200	0.069469	1.0333	1000	0.0736	1.0292	1000	0.0748	1.0280	1000	0.0769	1.0259	800
BOI/3	BOI/2	180.00	2.801	1200	0.069469	2.5923	1200	0.0736	2.5799	1200	0.0748	2.5764	1200	0.0769	2.5701	1200
BOI/2/1	**BOI/1**	445.00	0.869	1000	0.069469	0.7994	800	0.0736	0.7952	800	0.0748	0.7941	800	0.0769	0.7920	800
BOI/2	**BOI/1**	150.00	3.180	1200	0.069469	2.8322	1200	0.0736	2.8115	1200	0.0748	2.8057	1200	0.0769	2.7952	1200
BOI/0	**BOI/1**	422.03	1.844	1200	0.069469	1.7744	1000	0.0736	1.7703	1000	0.0748	1.7691	1000	0.0769	1.7670	1000
BOI/1	BO1	165.00	6.046	1200*2	0.069469	5.0737	1000	0.0736	5.0156	1000	0.0748	4.9993	1000	0.0769	4.9700	1000

Table 6

Modified design of the storm water drainage system with recharge well (rw dia 35 cm-60 cm)

| Section		Length of Drain	Actual Runoff (q)	Old Diameter of Pipe (D)	35 cm dia			45 cm dia			50 cm dia			60 cm dia		
U/S Node	D/S Node				Well Discharge (Qw)	Cum. Discharge (q')	Pipe Dia Provided	Well Discharge (Qw)	Cum. Discharge (q')	Pipe Dia Provided	Well Discharge (Qw)	Cum. Discharge (q')	Pipe Dia Provided	Well Discharge (Qw)	Cum. Discharge (q')	Pipe Dia Provided
			(m³/s.)	(mm)	(m³/s)	(m³/s)	(mm)	(m³/s)	(m³/s)	(mm)	(m³/s)	(m³/s)	(mm)	(m³/s)	(m³/s)	(mm)
AOI/4	AOI/3	502.27	0.795	1000	0.0787	0.7165	800	0.0820	0.7133	800	0.0834	0.7118	800	0.08605	0.7092	800
AOI/3/1	AOI/3	334.68	0.780	1200	0.0787	0.7018	800	0.0820	0.6985	800	0.0834	0.6971	800	0.08605	0.6944	800
AOI/3	AOI/2	196.58	1.501	1200	0.0787	1.2647	1000	0.0820	1.2549	1000	0.0834	1.2506	1000	0.08605	1.2428	1000
AOI/2/1	AOI/2	245.25	0.517	1000	0.0787	0.4378	600	0.0820	0.4345	600	0.0834	0.4331	600	0.08605	0.4305	600
AOI/2	AOI/1	235.06	1.909	1200	0.0787	1.4361	1000	0.0820	1.4166	1000	0.0834	1.4080	1000	0.08605	1.3923	1000
AOI/1	BOI/1/1	330.31	2.883	1200*2	0.0787	2.4110	800	0.0820	2.3915	800	0.0834	2.3829	800	0.08605	2.3672	800
BOI/1/1	**BOI/1**	500.00	2.885	1200*2	0.0787	2.3341	800	0.0820	2.3114	800	0.0834	2.3013	800	0.08605	2.2829	800
BOI/4	BOI/3	305.00	0.866	1000	0.0787	0.7877	800	0.0820	0.7845	800	0.0834	0.7830	800	0.08605	0.7804	800
BOI/3/1	BOI/3	370.00	1.103	1200	0.0787	1.0241	800	0.0820	1.0208	800	0.0834	1.0194	800	0.08605	1.0168	800
BOI/3	BOI/2	180.00	2.801	1200	0.0787	2.5645	1200	0.0820	2.5547	1200	0.0834	2.5504	1200	0.08605	2.5426	1200
BOI/2/1	BOI/2	445.00	0.869	1000	0.0787	0.7901	800	0.0820	0.7869	800	0.0834	0.7854	800	0.08605	0.7828	800
BOI/2	**BOI/1**	150.00	3.180	1200	0.0787	2.7859	1200	0.0820	2.7696	1200	0.0834	2.7624	1200	0.08605	2.7493	1200
BOI/0	**BOI/1**	422.03	1.844	1200	0.0787	1.7652	1000	0.0820	1.7619	1000	0.0834	1.7605	1000	0.08605	1.7578	1000
BOI/1	BO1	165.00	6.046	1200*2	0.0787	4.9439	1000	0.08199	4.8984	1000	0.083429	4.8782	1000	0.08605	4.8416	1000
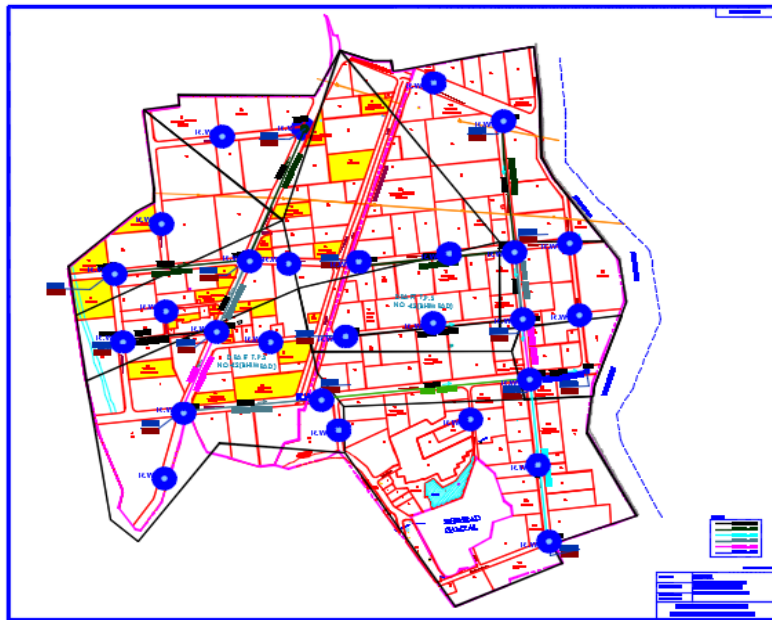
Figure 9. Plan of modified design of the storm water drainage system with recharge well
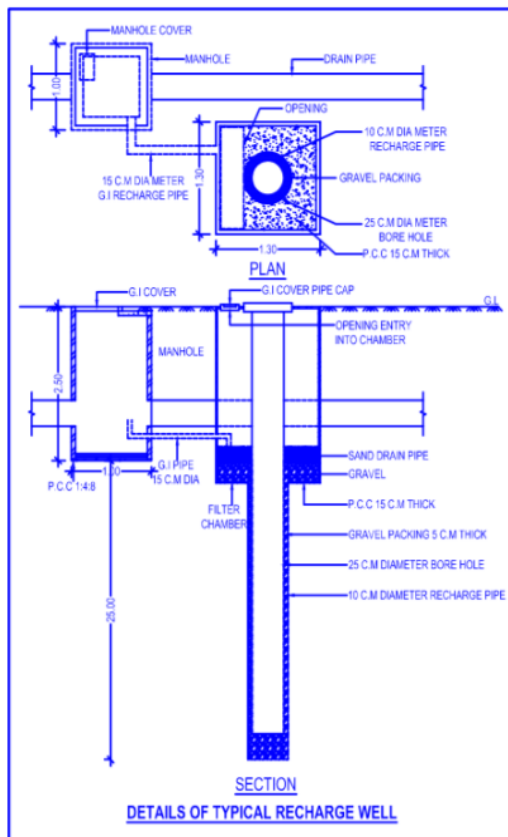

Figure 10. C/S of recharge well with storm water drainage system

Cost Analysis and Economic Aspect

This study attempted to conserve water through the use of groundwater revive and diminish the cost of providing for a rainwater drainage framework; in this way, the study achieved numerous advantages at the same time (Wolf, 2015). The different costs, for example, were:

(1) Total cost of revive wells and net cost of the rainwater drainage framework with groundwater revive wells;

(2) Total cost of rainwater drainage system without groundwater revive wells; and

(3) Total cost of the rainwater drainage system with groundwater revive, excluding revive wells.

The above figure indicates the variation in cost for different diameters of groundwater revive wells. According to Figure 11, it is clear that the cost of rainwater drainage frameworks stays steady, as there are no new building developments. The cost of groundwater revive wells increases with the increase in the diameter of the revive wells.
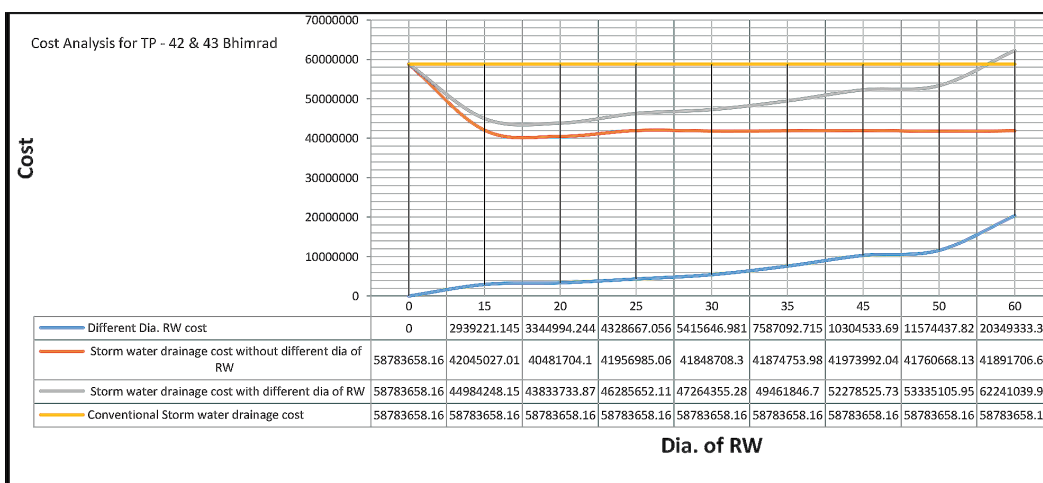
Figure 11. Cost analysis for TP - 42 & 43 Bhimrad

As seen in Figure 12, the sparing of the cost of the rainwater drainage framework with groundwater revive shifts from 9.5% to 25.5%, when contrasted with the regular rainwater drainage framework for the reviewed territory. The highest value of percentage savings falls between 25% and 26%, when the diameter of the revive well is kept at 20 cm.
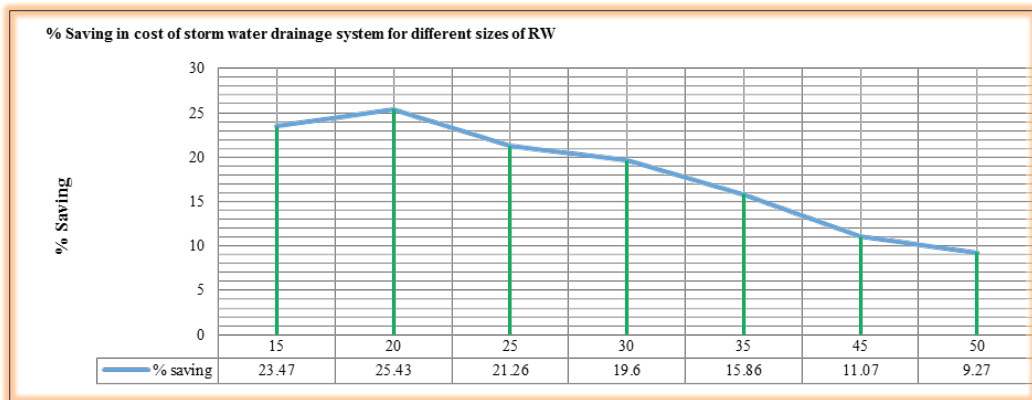
Figure 12. Percentage saving in cost of storm water drainage system for different sizes of RW

Future Groundwater Quality Prediction by Artificial Revive

Results of different parameter tests for groundwater quality in laboratory (Pre-monsoon & post-monsoon 2015).   For groundwater quality assessment, experimental work was conducted during the pre-monsoon and post-monsoon period of 2015; its results are shown in Figure 13. As per the record of the Collector's Office, Surat, the highest rainfall for Surat City was 1400 mm as of October 31, 2015. The concentrations of each parameter test results were changed for the pre-monsoon and post-monsoon periods. When an artificial revive was provided, it proved beneficial for the groundwater quality, thereby making it both usable and sustainable for future water demands of rapid increases in urbanisation.

	EC			TDS(mg/l)			pH			Chlorine			Hardness			Alkalinity		
	Standard value	Pre monsoon value	Post monsoon value	Standard value	Pre monsoon value	Post monsoon value	Standard value	Pre monsoon value	Post monsoon value	Standard value	Pre monsoon value	Post monsoon value	Standard value	Pre monsoon value	Post monsoon value	Standard value	Pre monsoon value	Post monsoon value
Aakash Era	750	4865	4723	500	4513	4486	7	7.74	7.68	0.2	0.2	0.22	300	450	430	120	440	390
Halpati Niwas	750	4583	4478	500	3416	3187	7	8.25	8.2	0.2	0.4	0.37	300	580	570	120	150	140
Aakash Evenue	750	4320	4219	500	3127	2956	7	8.11	8.05	0.2	0.3	0.24	300	520	510	120	160	160
Ashirwad Enclave (Near Khadi)	750	4249	4015	500	2896	2619	7	7.99	7.84	0.2	0.3	0.1	300	810	790	120	430	240
Santosh Height (Near Khadi)	750	4106	3998	500	3910	3649	7	7.92	7.62	0.2	0.1	0.11	300	740	710	120	480	450
Aastha Flats (Near Khadi)	750	4598	4476	500	4126	3973	7	8.1	8	0.2	0.4	0.3	300	470	460	120	520	480
Swagat Residency (Near Khadi)	750	4320	4213	500	3189	2646	7	7.47	7.59	0.2	0.3	0.23	300	490	450	120	440	310

Figure 13. Results of pre-post monsoon sample test results – 2015

Result of acquiring past-year groundwater quality data using interpolation techniques in GIS. Groundwater quality data, relating to previous years, at separate locations of Surat City, were collected from the GWSSB. The GWSSB is the government body that regularly checks and monitors yearly groundwater quality for different important parameters for domestic use.

Thereafter, all collected data were sorted by year, according to pre-monsoon and post-monsoon seasons for all locations adjoining the study area; finally, all sample point locations near the study area were marked on the map of Surat City and converted into shape files by attribute, with location name of GWQ parameter concentrations. The programme generated a point data of selected locations by digitisation.

A point location map from the previous year's data was generated in ARC GIS with the help of different techniques of interpolation to extract data from maps. Three interpolation techniques, IDW, Kriging and Natural Neighbour, were used to acquire the previous years' GWQ concentration data relating to four different parameters i.e. pH, total hardness (TH), total dissolved solid (TDS) and chlorides (CL) for a validation check of the interpolation map values with original values and for values of each parameter concentration observed by the GWSSB in 2006. These are shown in Figure 14 and Figure 15. The results were analysed and comparisons of all three interpolation methods of the GWQ parameter were made and matched with their original values of the IDW interpolation technique. The results for pH were matched, showing a value of 99.14% for IDW; 98.77% for Kriging; and 100% for Natural Neighbour. When measuring TH, TDS, and CL, all parameters using the Natural Neighbour method gave the nearest values when compared with original values, showing, therefore, that the Natural Neighbour technique was the most suitable for acquiring previous years' missing data of GWQ for an area.

	map value	orginal value	difference	percentage	map value	orginal value	difference	percentage
			pH				TH	
IDW	8.04	8.11	0.07	99.13686806	169.395	168	-1.395	100.8303571
KRIGING	8.01	8.11	0.1	98.76695438	184.19	168	-16.19	109.6369048
NATURAL NEIGHBOUR	8.11	8.11	0	100	189.691	168	-21.691	112.9113095
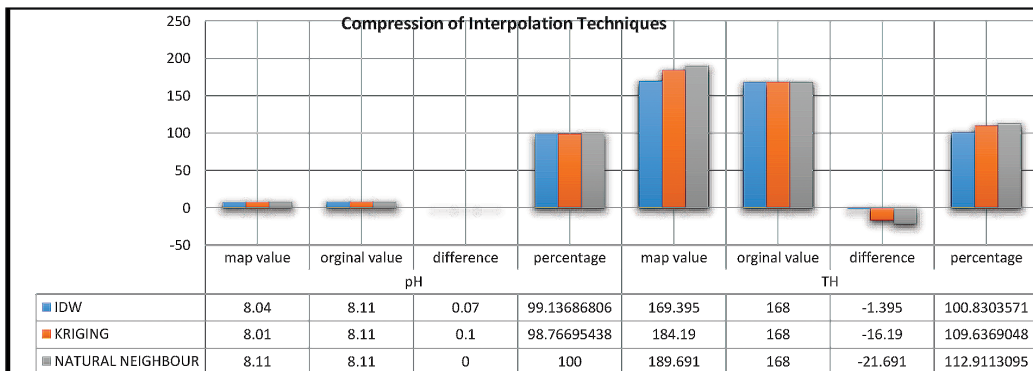
Figure 14. Comparative graph of interpolation techniques of pH and TH

	map value	orginal value	difference	percentage	map value	orginal value	difference	percentage
			TDS				CL	
IDW	439.432	628.58	189.148	69.90868306	51	124.53	73.53	40.95398699
KRIGING	524.666	628.58	103.914	83.4684527	58.6	124.53	65.93	47.05693407
NATURAL NEIGHBOUR	622.281	628.58	6.299	98.99790003	121.611	124.53	2.919	97.65598651
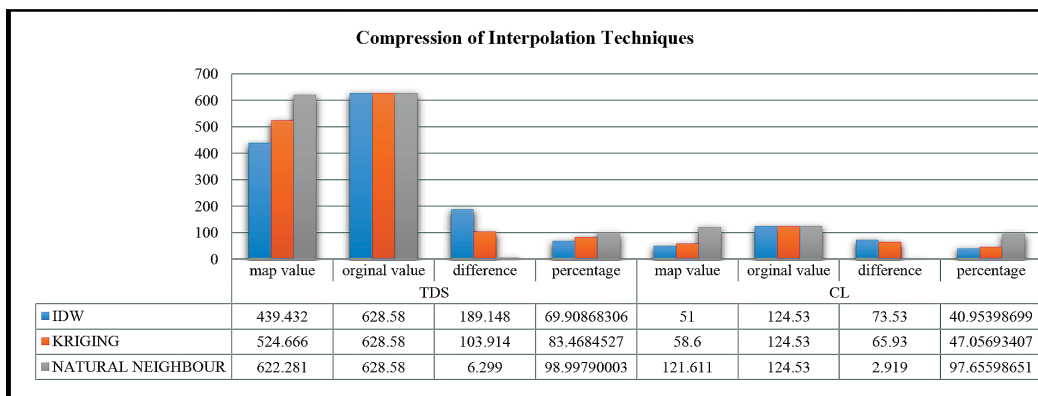
Figure 15. Comparative graph of interpolation techniques of TDS and CL

The extent of coverage between this new polygon and the underlying polygons is utilised as the weights. Groundwater quality data are extracted from seven points of the study area by the GIS interpolation method of Natural Neighbour with its map of groundwater quality. The previous years of 2004, 2006, 2008, 2011, 2012, 2013 pre-monsoon and post-monsoon, and the 2014 pre-monsoon groundwater quality parameters, like pH, electric conductivity, TH, TDS and CL, were extracted. Past-year data can be used for analysis of groundwater quality of an area. From an extracted map of the study area, contrast colour and raster datasets with pixel values were generated. Colour can help render each raster dataset as a single seamless image. In Natural Neighbour interpolation, the map randomly selected seven sample data set pixel values of the study area, and acquired the missing data of the past year.

Groundwater quality index was computed in ERDAS programming of GIS, utilised for getting a GWQI outline. Based on the analysis, GWQ was decreasing year by year with the increase in urbanisation in the area. As future water demands are not likely to be met by SMC, the study provided the design of groundwater for domestic purposes. Based upon the graphical analysis shown in Figure 16, post-monsoon GWQ was increasing faster than pre-monsoon GWQ, with natural infiltration of rainwater. Post-monsoon GWQI was improving with artificial revive wells, thereby proving that the artificial revive wells were beneficial.
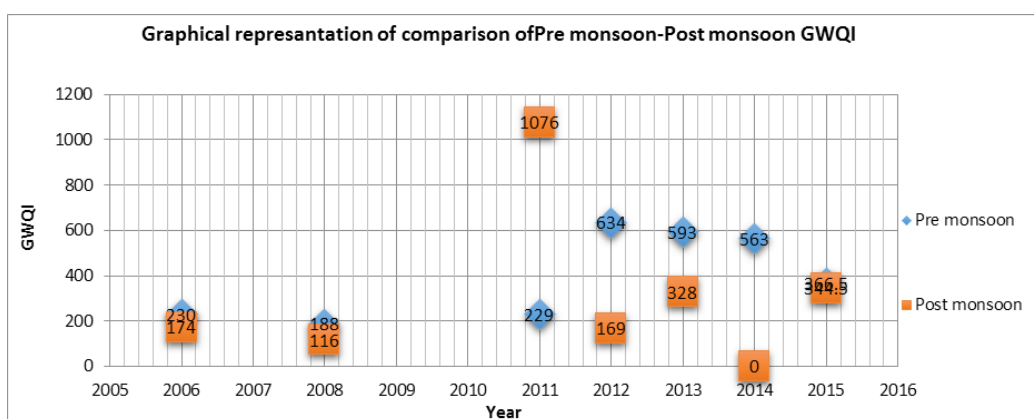
Figure 16. Graphical representation of comparison of pre-monsoon-post-monsoon GWQI

Results of the reduction in concentration of various parameters under various dilution conditions for future GWQ in GIS. Groundwater quality improvement predictions, after the use of artificial revive wells, from 2015 to 2050, were determined with the help of GIS software using the dilution equation (Sharma & Patel, 2010). In the model, an initial concentration map was taken as an input parameter for the first stage. In the second-stage formula, dilution was used to determine changes in concentration after the use of artificial revive wells in the form of a concentration map. From the map, well distribution in seven location results was established. This procedure was carried out for each groundwater quality parameter, including pH, electrical conductivity, total dissolved solids, chlorine, total hardness and alkalinity. The information found comprised groundwater quality and contamination parameters, and a correlation was made with the underlying fixation and focus after dilution. A change in the reduction of actual

concentrations containing compounds of each parameter generally results after a dilution of groundwater quality concentrations. The percentage of change in actual concentrations was determined for each parameter for well distribution in the seven locations, and an improvement of the groundwater quality year by year (up to 2050) was predicted. There was a 4% reduction in concentration for the year 2015, showing pH reduction up to 1.5-2.5%, EC reduction up to 11-12%, TDS reduction up to 12-13%, Cl reduction up to 8-9%, TH reduction up to 10-12% and alkalinity reduction up to 7-8%. These results showed that artificial revive can improve groundwater quality.

Results of the yearly GWQI (Groundwater Quality Index) (2015-2050) in GIS. Groundwater quality in the study area was determined in pre-monsoon and post-monsoon seasons for the next 36 years (2015-2050), and the summarised statistical outcome of observed analysis is shown in Figure 17. The results showed that GWQI had improved every year, with the effect of dilution in the study area. In 2015, GWQI pre-monsoon was 357 and in 2050, GWQI pre-monsoon was 69.4.

Year wise Mean Dilution and GWQI

Year	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050
Postmonsoon	298.9	267.6	240.5	217	196.8	179.3	164.1	151.1	139.8	130	121.5	114.2	107.9	102.4	97.72	93.64	90.12	87.07	84.43	82.15	80.18	78.48	78.48	75.73	74.63	73.68	72.86	72.14	71.53	71	70.54	70.14	69.8	69.5	69.25	69.02
Premonsoon	312	278.9	250.3	225.5	204.1	185.6	169.6	155.8	143.8	133.5	124.6	116.9	110.2	104.4	99.43	95.12	91.39	88.17	85.38	82.97	80.89	79.09	76.19	75.03	74.02	73.15	72.4	71.75	71.19	70.71	70.29	69.92	69.61	69.34	69.1	

Figure 17. Year-wise mean dilution and GWQI

CONCLUSION

The development of new techniques for rainwater drainage systems, in conjunction with groundwater recharging, is advantageous in many ways. Different diameters of revive wells were used to study their economic advantages and to calculate all diameters as the revive rates with artificial revive well and pipe diameter of rainwater drain are inversely proportional to each other, that is, with the increase in revive rates enabled by the use of artificial revive well, the pipe diameter of rainwater drains decreased, and vice-versa. As discussed earlier, cost savings from 9-10% to 25-26% were observed for the different sizes of revive wells, compared with traditional rainwater drainage systems. The optimal pipe diameter considered in the study was 20 cm. The total cost of the rainwater drainage system without a groundwater revive well was INR5.5 cr to 6.0 cr, whereas the total cost of a rainwater drainage system with a groundwater revive well of diameter 20 cm was INR4.0 cr to 4.5 cr, which indicated savings of 25-26%.

Artificial revive wells, with and without rainwater drainage networks, have already been proposed for the research area and have proven to be helpful in satisfying future demands for water. Dilution results for future groundwater quality, including percentage reduction

in concentration of 2015, showed 2-12% for different parameters. After artificial revive, concentrations of groundwater quality parameters were reduced and quality improved with time. The results showed that GWQI was improving each year through the effect of the dilution technique in the study area, enhancing groundwater contamination.

In this study, an attempt was made in the direction of conserving water through groundwater revive as well as reducing the cost of rainwater drainage systems, thus achieving two-fold benefits. This research included the design of a rainwater drainage system combined with groundwater revive wells to save water as well as to decrease the total cost of the traditional rainwater system.

ACKNOWLEDGEMENT

REFERENCES

Basu, N. B. (2012). New life for Kolkata's sewers. *Journal of Civil Engineering, ASCE, 82*(7/8), 68–75.

Boulos, P. F. (2005). An innovative geocentric decision support solution to comprehensive planning, design, operation, and management of urban drainage systems. In *World Environmental and Water Resource Congress 2006: Examining the Confluence of Environmental and Water Concerns* (pp. 1-10). Omaha, Nebraska, United States.

Bryant, D. S. (2000). GIS tools for proactive urban watershed management. *Water Environment Federation, (WEF) – ASCE/EWRI,* 1–15.

Carrier III, W. (2003). Goodbye, Hazen; hello, Kozeny-Carman. *Journal of Geotechnical and Geoenvironmental Engineering, 129*(11), 1054–1056.

CPHEEO, M. (2013). Central public health and environmental engineering. *Manual on Sewerage and Sewage Treatment Organization*. New Delhi, India: Ministry of Urban Development.

Dohare, D. D. (2014). Analysis of groundwater quality parameters: A review. *Research Journal of Engineering Sciences (ISSN 2278 – 9472), 3*(5), 26–31.

Garnero, G. A. (2013). Comparisons between different interpolation techniques. *International Journal of the Photogrammetry – Remote Sensing and Spatial Information Sciences, XL-5W, 3*, 139–144.

Gorai, A. K. (2013). Spatial distribution analysis of groundwater quality index using GIS: A case study of Ranchi municipal corporation (RMC) area. *Journal of Geoinfor Geostat: An Overview, 1*(2), 1–11.

GWSSB-SOR. (2014-2015). Gujarat water supply and sewerage board. *Schedule of Rates*. Gandhinagar, India: Gujarat Water Supply and Sewerage Board.

Harpalani, P. D. (2013). Analysis of rainfall data and design of rainwater drainage system in an urban area. *Journal of Global Research Analysis, 2*(4), 100–104.

Hattamleh, O. H. (2013). The consequence of particle crushing in engineering properties of granular materials. *International Journal of Geosciences, 4,* 1055–1060.

Jariwala, R. C. A. D. (2012). The estimation of flood and its control by section modification in Mithi and Kankara tributaries at Surat. *International Journal of Engineering Research and Applications (IJERA-2248-9622), 2*(4), 862–867.

Kamra, S. K. (2012). Evaluation of sand-based rainwater filtration system for groundwater revive wells. *Journal of Current Science, 103*(4), 395–404.

Kellagher, R. A.-C. (2008). Sustainability criteria for the design of rainwater drainage systems for the 21st century. *11th International Conference on Urban Drainage.* Edinburgh, Scotland, UK.

Kumar, D. R. (2012). *Hydrogeological and geophysical study for deeper groundwater resource in quartzitic hard rock ridge region from 2D resistivity data.* Hyderabad, India: CSIR-National Geophysical Research Institute.

Kumar, S. K. (2012). Evaluation of sand-based rainwater filtration system for groundwater revive wells. *Journal of Current Science, 103*(4), 395–404.

Neris, J. J. (2012). Vegetation and land-use effects on soil properties and water infiltration of Andisols in. *Catena (Elsevier), 98,* 55–62.

Nutchanart, S. A. (2011). Effects of land cover on runoff coefficient. *Journal of Hydrology, 410*(3-4), 226–238.

Patel, P. D. (2010). Artificial groundwater revive field study: Site characterization and test results. *International Journal of Advanced Engineering Technology, 1*(2), 150–164.

Patel, P. D. (2011). Evaluation of hydraulic conductivity for the estimation of artificial groundwater revive rate. *Middle-East Journal of Scientific Research*, *10*(5), 638–646. ISSN: 1990-9233.

Patel P. D. M. (2011). Geotechnical parameters impact on artificial groundwater recharging technique for urban centers. *Journal of Water Resource and Protection, 3,* 275–282.

Schirmer, M. L. (2013). Current research in urban hydrogeology – A review. *Journal of Advances in Water Resources, 51,* 280–291.

Sharma, N. D. (2008). Development of software module for groundwater revive. *World Applied Sciences Journal*, *3*(6), 963–968. ISSN 1818-4952.

Sharma, N. D. (2011). Experimental study of groundwater quality improvement by recharging with rainwater. *International Journal of Civil Engineering and Technology*, *2*(1), 10–16. (IJCIET- ISSN 0976 – 6308).

Walton, W. (1970). *Groundwater resource evaluation*. New York, NY: McGraw-Hill Inc.

Wolf, D. F. D. A. (2015). Whole life costs and benefits of sustainable urban drainage systems in Dunfermline, Scotland. *Journal of American Society of Civil Engineers*, *7902*(5), 419–427. (ASCE-ISBN: 978-0-7844).

PERTANIKA JOURNALS

Mathematical Modelling and Residual Life Prediction of an Aluminium Electrolytic Capacitor

Bhargava, C.[1]*#, Banga, V. K.[2] and Singh, Y.[3]

[1]*Department of Electronics and Communication Engineering, IKG Punjab Technical University, Kapurthala 144603, Punjab, India*

[2]*Office of Principal, Amritsar college of Engineering and Technology, Amritsar 143001, Punjab, India*

[3]*Department of Electrical Engineering, Harcourt Butler Technological University, Kanpur 208301, Uttar Pradesh, India*

ABSTRACT

An electrolytic capacitor is a widely used, passive component in electronics and the manufacturing industry. But it is known for its low reliability when used in critical applications and environments. The prediction of the residual life of a capacitor is a challenging issue. It is beneficial for users as they will be ready to replace a component when it is faulty. This will save them replacement and repairing costs. This paper focusses on the mathematical modelling and behaviour analysis of electrolytic capacitors related to various electrical parameters and environmental stresses. A novel method is proposed to access the impact of humidity on the health of a capacitor. A capacitor's end of life is explored using the design of experiments approach. After calculating the lifetime of a capacitor, an expert system is modelled that can predict the sudden crash of the capacitor before it actually fails, using various statistical and analytical techniques. A comparison was done using the experimental calculated lifetime and proposed analytical method outcome. The proposed corrective model of lifetime calculation proves 92.76% accuracy. This shortfall is due to the fact that some electrical and environmental variables change at a pace that is too fast for the prediction system to respond.

Keywords: Accelerated testing, electrolytic capacitor, failure analysis, military handbook, reliability, Taguchi approach

INTRODUCTION

As thousands of electronic components are integrated on a single chip, the reliability of the chip may be compromised, and this poses a problem for manufacturers as well as users. Manufacturers of electronic components prepare a datasheet listing the lifetime of each

component that is produced. However, electronic products often have to perform in extreme environmental conditions or within varying electrical parameters that can cause their sudden failure. If a component fails before a certain period, the entire system can break down. This can lead to unnecessary costs incurred for replacement and repair of the failed system.

To ensure that electronic products are reliable and useful throughout their lifetime, different environmental stress tests as well as electrical parameter variation tests need to be carried out. The main cause of failure/degradation of electrolytic capacitors has been found to be the loss of electrolytes (Bengt, 1995). Kececioglu (2002) suggested that thermal stress causes a change in electrochemical reactions, resulting in evaporation of electrolytes.

Reduction in the weight of an electrolytic capacitor is the main indicator of electrolytic loss. Dehbi, Wondrak, Ousten and Danto (2002) suggested that not only temperature, but voltage and ripple current also play an important role in determining residual life. Saha, Goebel and Christophersen (2009) compared various algorithms for capacitor life calculation. This paper emphasises the calculation of an electrolytic capacitor's life based on various physics of failure laws. Based on the design of experiments (DOE) approach, the four-factor four-level design was selected. This design gives 16 combinations of input variables. Capacitor life was calculated using various laws. Then on the same set of combinational variables, the accelerated life testing method was used. Capacitor life was calculated using the Arrhenius law. The experimental and theoretical results were compared and analysed for error.

Mathematical Modelling of an Electrolytic Capacitor

Reliability has been a critical issue in the manufacture of electrolytic capacitors from design to development to marketing to real-time use. The key factor for success in the electronics industry is reliability of products. Sankaran, Rees and Avant (1997) suggested failure prediction techniques for capacitors. Different prediction models have also been considered based on components' parameters, stress factors and historical failure data etc. Bhargava, Banga and Singh (2014) reviewed the various reliability techniques for failure prediction of electronic components.

Empirical (Standard) Model

The empirical methods or standard prediction models are based on historical data of component failure and experience of the user. The standard set in the military handbook MILHDBK-217F and the telecommunication handbook, TELECORDIA, among others, is widely used for reliability estimation. The assumption for such models is that system failure is integrally associated with the component's failure. As suggested by Elbert and March-Force (1989), various environmental factors are considered for reliability calculation.

MILHDBK-217F is very well known as a reliability prediction model in military applications. The failure rate using this military handbook can be calculated as:

$$\lambda = \sum_{i=1}^{n} (\lambda ref, i * \pi S * \pi T * \pi E * \pi Q * \pi A) \qquad [1]$$

where,

λref = Failure rate under reference conditions

πS = Stress factor

πT = Temperature factor

πE = Environmental factor

πQ = Quality factor

πA = Adjustment factor

However, because the military handbook's data were outdated, a revised version was launched, but this too met with some dissatisfaction. A telecommunication company then launched TELECORDIA SR-332 to address the infant mortality rate as well as the steady state useful operating life failure rate. On their part, Held and Fritz (2009) discussed various empirical standards for performance calculation of capacitors. The difference between both methods is summarised in Table 1.

Table 1

Comparison between MILHDBK and TELECORDIA

MILHDBK-217F	TELECORDIA SR-332
Acts as a reliability prediction model in telecommunication industry	Acts as a reliability prediction model in telecommunication industry
Military handbook uses failure in time (FIT) per million hours	Uses failure in time (FIT) per billion hours
It has 14 environment classifications (three ground, eight air, one space, two sea)	It has six environment classifications (four ground, one air, one space)
A specific SMT device model	No specific SMT device model
Handles lower number of gate count IC	Handles higher number of gate count IC
Provides only steady state useful life failure rate	Provides a measure of infant mortality rate and steady state useful operating life failure rate

Analytical Models for Life Prediction

Empirical models, which are based on historical or user data, are not up-dated. In contrast to empirical methods, there are several other parameters on which life of a system depends. These models analyse the acceleration factor, a constant multiplier for different stress levels. Then the effect of this stress level acceleration factor is studied on the life of the product, as claimed by the product's datasheet. Stevens and Dapo (2002) reviewed the service life of an aluminium electrolytic capacitor and its lifetime calculation.

Effect of Temperature on Lifetime

As proposed by Whitman (2012), de-rating of a capacitor's life and its performance degradation accelerates with increased temperature. As the temperature increases, the capacitor initiates a change in chemical reactions because an electrolytic capacitor is an electrochemical device. Kötz, Hahn and Gallay (2006) considered temperature a critical factor for evaluating the performance of capacitors. It has been assumed that every 10 degrees of rise of temperature degrades the capacitor's life by a factor of 2. So, increase in temperature furthermore reduces capacitance. The expected life using this method can be calculated as Equation [2]. Ma and Wang (2005) suggested various fault diagnosis techniques for electrolytic capacitors.

$$Lifetime\ (Expected) = Lifetime(D) \times A_T \tag{2}$$

$$Lifetime\ (E) = Lifetime(D) \times 2^{\frac{(Tm-Ta)}{10}} \tag{3}$$

where,
Lifetime (D) = lifetime declared by manufacturer in datasheet i.e. 2,000 hr
A_T = Temperature acceleration factor=$2^{[(Ta-Tm)/10]}$
Tm (k) = Maximum temperature i.e. 155°C=155+273=428K
Ta (k) = Accelerated temperature i.e. 105°C=105+273=378K

Effect of Voltage on Lifetime

As suggested by Naikan and Rathore (2016), the degradation of capacitor life does not depend on only temperature acceleration, but also on the rated and operated voltage. The capacitor life is longer if it is used within the rated voltage limit. Kulkarni, Biswas, Celaya and Goebel (2011) analysed the effect of electrical parameters on capacitor performance. The expected life using this methodology was calculated as Equation [5], which is a modified version of Equation [2]:

$$Lifetime\ (Expected) = Lifetime(D) \times A_T \times A_V \tag{4}$$

$$Lifetime\ (E) = Lifetime(D) \times 2^{\frac{(Tm-Ta)}{10}} \times \left(\frac{Va}{Vm}\right)^{-n} \tag{5}$$

where,
Va = Applied voltage=9V
Vm = Rated voltage=16V
A_V = Voltage acceleration factor= $(Vo/Vn)^{-N}$
N = constant=1 for radial capacitors

Effect of Ripple Current on Lifetime

Joule's law states that when ripple current is applied to an electrolytic capacitor, the capacitor generates heat. Due to heat generation, electrolytes evaporate. As electrolytes evaporate, indicated by weight loss, capacitance tends to decrease. Heat rise causes a reduction in life as

compare to Arrhenius law. Let the acceleration factor for ripple current be A_I. The expected lifetime also has influence on the ripple current acceleration factor. Parler Jr and Dubilier (2004b) suggested various multipliers for measuring the performance of a capacitor.

$$Lifetime\ (Expected) = Lifetime(D) \times A_T \times A_V \times A_I \qquad [6]$$

$$Lifetime\ (E) = Lifetime(D) \times 2^{\frac{(Tm-Ta)}{10}} \times \left(\frac{Va}{Vm}\right)^{-n} \times Ki^{A\frac{\Delta T}{10K}} \qquad [7]$$

where, $A_I = K_I^{[1-(Ia/Io)2]*[\Delta To/10K]}$
K_I = Ripple current multiplier = 2
Ia = Applied and ripple current
Io = Maximum ripple current
ΔTo = Increase in core temperature

Corrective Analytical Model for Behaviour Analysis

As the literature suggests, while calculating the lifetime of an electrolytic capacitor, temperature, voltage and ripple current needed to be considered as they play an important impact on the lifetime of a capacitor. However, there are also several other electrical and environmental parameters that have an impact on degradation and life estimation of an electrolytic capacitor (Parler Jr & Dubilier, 2004a).

Impact of Humidity on Lifetime

Humidity plays an important role in capacitor reliability. It causes electrolytic corrosion due to moisture present in the capacitor. Moisture causes a decrease in the value of capacitance and subsequently affects the life of the capacitor. As suggested by Sinnadurai (1974), the humidity acceleration factor can be calculated using Equation [8], which is modified from Peck's law. It has been found that if a capacitor has been used more than the rated humidity, the capacitor's life is degraded at an exponential rate. After taking temperature, voltage and humidity into consideration, the equation formed is Equation [9] below.

$$Lifetime\ (Expected) = Lifetime(D) \times A_T \times A_V \times A_I \times A_H \qquad [8]$$

$$Lifetime\ (E) = Lifetime(D) \times 2^{\frac{(Tm-Ta)}{10}} \times \left(\frac{Va}{Vm}\right)^{-n} \times Ki^{A\frac{\Delta T}{10K}} \times \exp\{C \times ((RHa)^n - (RHm)^n)\} \qquad [9]$$

where,
C = Humidity constant = 0.00044
RHs = Rated relative humidity
RHa = Applied relative humidity = 30
A_H = Humidity acceleration factor = $[\exp(C*((RH_S)^n-(RH_O)^n)]$
n = Humidity activation exponent = 2

Clearly, the life of a capacitor depends not only temperature, voltage and current, but also on humidity, as shown above in Equation [9].

Experimental Approach for Lifetime Calculation

Kalaiselvan and Rao (2016) explored the practical use of capacitors and found several factors that increase/reduce the value of their performance. To observe capacitor performance under various conditions for a couple of hours can be a tedious job. Accelerated life testing is the method by which a capacitor is placed under stressed conditions and then measured for output. For example, using the Arrhenius equation, a Sterile Barrier System that is subjected to 40 days of Accelerated Aging at +55° has similar ageing characteristics as a one-year-old Real Time sample. So, accelerated ageing test data are accepted by most regulation bodies.

METHOD AND MATERIALS

In experimental setup, we placed a capacitor, covered with sand on a hot plate. Environmental and electrical stresses were introduced and lifetime was observed using the Arrhenius equation (Huang et al., 2017). A hole of 0.9 mm was drilled on the capacitor to accelerate the heating process. Ahmad, Islam and Salam (2006) used the Weibull distribution to analyse accelerated life testing response.

Figure 1. Experimental setup for accelerated life testing

Failure in time (FIT) (per 10^9 h) was calculated using the Arrhenius equation of acceleration as given below:

$$FIT(\lambda) = \frac{F}{D \times T \times A_F} \tag{10}$$

where,
F = Number of failures
D = Number of devices tested
T = Test hours
A_F = Acceleration factor $= \exp^{Ea/k(1/T}\text{use}^{-1/T}\text{test})$
E_a = Activation energy (ev)
K = Boltzmann constant
T_{use} = use temperature, T_{test} = test temperature as per datasheet

After calculating the failure rate, the MTBF corresponding to FIT was calculated as:

$$MTBF \ (hours) = \frac{1}{\lambda} \times 10^9 \qquad\qquad [11]$$

The calculated life using accelerated life testing method is summarised in Table 2.

DOE Approach: Taguchi Method

Design of experiments (DOE) is the systematic approach to finding the relationship between process variables and response of system. The Taguchi approach is a systematic means of designing, conducting and analysing experiments that are of great significance in quality planning. The experiments are conducted based on the Taguchi approach and tool life values obtained. The use of the Taguchi approach leads to a significant economic advantage, resulting in reduction of material and machining time required for expedition due to the reduced set of experiments. The steps in designing, conducting and analysing experiments are as follows:

1. Selection of factors for the study: Temperature, ripple current, voltage and humidity.
2. Selection of the number of levels for the factors; four factors at four levels.
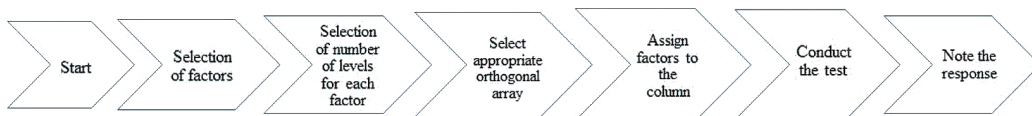
Figure 2. Flow chart of Taguchi approach

Table 2
Process variables and their levels

Process Variables	Units	Notation	Limits			
			1 (Low)	2 (Medium)	3 (High)	4 (Extremely High)
Temperature	°C	t	87	94	101	108
Ripple current	mA	r	26	28	30	32
Humidity	Rh	r	80	82	84	86
Voltage	V	v	5.8	6	6.2	6.4

3. Selection of appropriate orthogonal array. The L_{16} orthogonal array was selected for performing experiments using the Taguchi approach.

Table 3
L16 orthogonal taguchi approach

Run	Factors			
	Temperature	Ripple current	Humidity	Voltage
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	1	4	4	4
5	2	1	2	3
6	2	2	1	4
7	2	3	4	1
8	2	4	3	2
9	3	1	3	4
10	3	2	4	3
11	3	3	1	2
12	3	4	2	1
13	4	1	4	2
14	4	2	3	1
15	4	3	2	4
16	4	4	1	3

4. Assignment of factors to columns. To calculate the life of the capacitor, the values were assigned to all the factors, divided among four levels. Then, the acceleration factors were calculated for all the factors. Afterwards, the total lifetime of the electrolytic capacitor was calculated using Equation (10). The values of factors, as seen in Table 2 were inserted in Table 3. After inserting the values of the factors, acceleration factors corresponding to those factor values were calculated using Equations (3-11). Then, the lifetime of the capacitor was analysed.

5. Conduct of the test. Life calculation experiments were conducted as per the design matrix of Table 3. The different combinations of process variables and factors were conducted. The extremely high temperature corresponded to the accelerated stress factors and using the Arrhenius equation, the output life could be calculated. Then the analytical and experimental life has been compared.

Statistical Model of Lifetime Prediction

The regression technique was used to estimate the relationship of a response variable to a set of predictor variables to develop a mathematical model to predict the tool life of the capacitor. The relationship between the independent process parameter variables and capacitor life can be represented by the following mathematical model:

$$CL(Lifetime) = C(T^l \times R^m \times V^n \times H^o)$$ [12]

where, CL is the measure of response (capacitor life); T, R,V and H the values of the process parameters i.e. temperature, ripple current, voltage, frequency and humidity respectively; C is a model constant; and l, m, n, and o are model parameters.

The above Equation (14) can be represented in linear form as under:

$$\ln CL = \ln C + l \ln T + m \ln R + n \ln V + o \ln H \qquad [13]$$

Equation (15) can be further modified as given below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \qquad [14]$$

where, Y is the capacitor life on logarithmic scale; X1, X2, X3, X4 and X5 are logarithmic transformations of temperature, ripple current, voltage, frequency and humidity (i.e. X1=ln T; X2=ln R; X3=ln V; X4=ln H; β0=ln C; β1=l; β2=m; β3=n; β4=o). The values of β0, β1, β2, β3, β4 and β5 represent the regression coefficients to be determined. In this study, a multiple regression model using the linear regression technique was developed to predict tool life based on experimental values. The coefficients for the regression model were determined using Minitab statistical software.

RESULTS AND DISCUSSION

Using the design of experiment (DOE) approach, the L16 design matrix was generated. Life was calculated using the analytical method, statistical method and experimental method. Then, the different techniques were compared and the most accurate one was chosen.

Analytical Life Calculation

Using analytical life calculation models, the values of factors assigned in the design matrix and then the values of acceleration factors were calculated and substituted in Equation [8] to find the lifetime of the capacitor.

Table 4
Lifetime calculation using the analytical method

| Run | Factors (Actual form) | | | | Lifetime (hours) Using Analytical Method |
	Temperature (°C)	Ripple Current (mA)	Humidity (%Rh)	Voltage (Volts)	
1	87	26	80	5.8	8220.05
2	87	28	82	6.0	7648.26
3	87	30	84	6.2	7110.02
4	87	32	86	6.4	6639.03
5	94	26	82	6.2	4079.17
6	94	28	80	6.4	5087.09
7	94	30	86	5.8	4027.87
8	94	32	84	6.0	5077.32
9	101	26	84	6.4	2102.44
10	101	28	86	6.2	2070.36
11	101	30	80	6.0	3716.00
12	101	32	82	5.8	3684.14
13	108	26	86	6.0	1185.15
14	108	28	84	5.8	1569.27
15	108	30	82	6.4	1827.66
16	108	32	80	6.2	2438.30

Statistical Life Calculation

Using Minitab 18.1 software, regression analysis was done. The t-test was used to test the accuracy of the model. Accuracy was measured by the fact that the variance value should be minimal and the R^2 value should be higher.

Regression equation=-10465 + 2097.8 Temperature + 1840 Current 2653 Humidity
+ 5639 Voltage   [15]

Table 5
Analysis of regression model

S	R-sq	R-sq (adj)	R-sq (pred)
208.913	99.39%	99.16%	98.76%
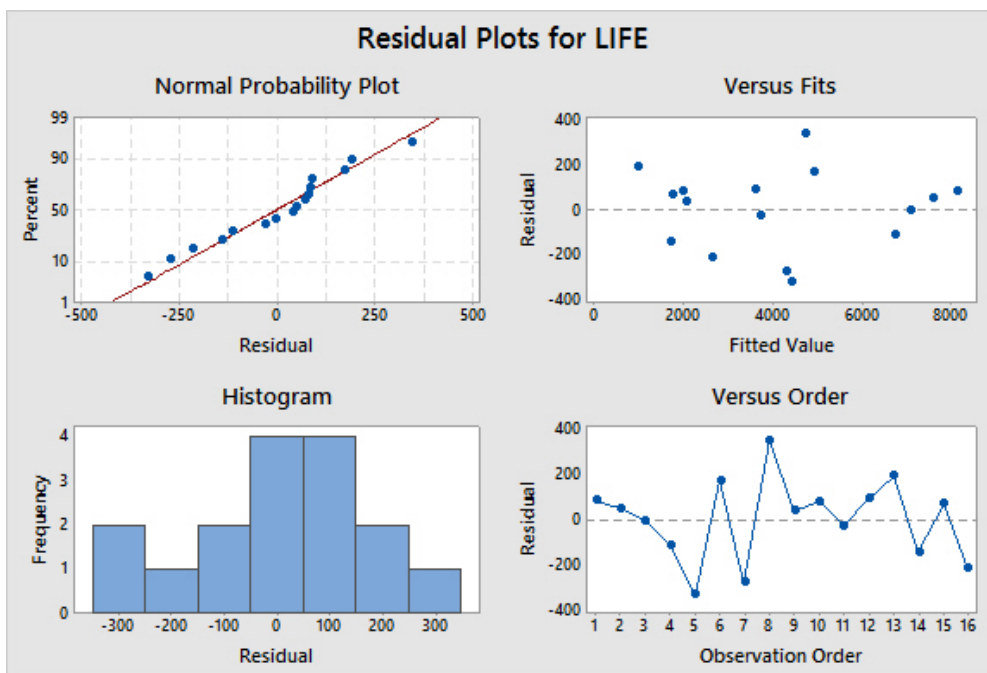
Figure 3. Regression outputs using Minitab software

Using Equation [15], life was calculated using the regression equation shown in Table 6.

Table 6

Lifetime calculation using regression method

Run	Factors (Actual Form)				Lifetime (hours) Using Regression	Error (%)= ([Analytical-Regression]/ Analytical)*100
	Temperature (^{O}C)	Ripple Current (mA)	Humidity (%Rh)	Voltage (Volts)		
1	87	26	80	5.8	8138.07	1.00
2	87	28	82	6.0	7598.97	0.64
3	87	30	84	6.2	7114.96	-0.07
4	87	32	86	6.4	6752.90	-1.72
5	94	26	82	6.2	4409.64	-8.10
6	94	28	80	6.4	4915.83	3.37
7	94	30	86	5.8	4303.34	-6.84
8	94	32	84	6.0	4731.87	6.80
9	101	26	84	6.4	2064.21	1.82
10	101	28	86	6.2	1990.80	3.84
11	101	30	80	6.0	3746.03	-0.81
12	101	32	82	5.8	3594.96	2.42
13	108	26	86	6.0	993.76	16.15
14	108	28	84	5.8	1710.95	-9.03
15	108	30	82	6.4	1758.78	3.77
16	108	32	80	6.2	2652.99	-8.80
				Average Error (%)		**0.28**

Experimental Life Calculation

Gulbrandsen, Arnold, Caswell and Cartmill (2014) analysed the lifetime of electrolytic capacitors using practical methods. In their experimental setup, different levels of stress were introduced to the component i.e. temperature, voltage, current and humidity. Twenty capacitors were put on a hotplate and covered with sand. The initial measured capacitance, ESR and weight of the individual capacitors showed changes in characteristics during heating. Those were measured again. If capacitance decreased by 20% or ESR increased by 100% and weight decreased, then the capacitor was said to be failed. Lifetime was calculated using the Arrhenius law as per Equation [10-11].

The extract lifetime, using the experimental method, is summarised in Table 7. The analytically calculated life is compared with the experimental value and the percentage error was calculated. Jano and Pitica (2012), and Ashburn and Skamser (2008) used accelerated life testing for exploring de-rating factors in the electrolytic capacitor.

Table 7
Error analysis of analytically calculated lifetime

Run	Lifetime (Experimentally) (hours) "A"	Lifetime Analytically (hours) "B"	Lifetime Statistically (hours) "C"	Error (%)= [(A-B)/A] *100	Error (%)= [(A-C)/A] *100
1	8,875.76	8,220.05	8,138.07	7.39	8.31
2	8,289.17	7,648.26	7,598.97	7.73	8.33
3	7,654.65	7,110.02	7,114.96	7.12	7.05
4	7,042.39	6,639.03	6,752.90	5.73	4.11
5	5,514.56	4,079.17	4,409.64	26.03	20.04
6	5,929.18	5,087.09	4,915.83	14.20	17.09
7	3,915.28	4,027.87	4,303.34	-2.88	-9.91
8	5,965.11	5,077.32	4,731.87	14.88	20.67
9	2,156.27	2,102.44	2,064.21	2.50	4.27
10	2,015.98	2,070.36	1,990.80	-2.70	1.25
11	3,946.79	3,716.00	3,746.03	5.85	5.09
12	3,741.43	3,684.14	3,594.96	1.53	3.91
13	1,315.29	1,185.15	993.76	9.89	24.45
14	1,662.12	1,569.27	1,710.95	5.59	-2.94
15	2,177.59	1,827.66	1,758.78	16.07	19.23
16	2,365.62	2,438.30	2,652.99	-3.07	-12.15

Comparison Between Experimental and Analytical Models

The lifetime of an electrolytic capacitor was calculated using the accelerated life testing method as well as the proposed model of analytical testing. The average error between analytical life and experimental calculated life, in percentage, was 7.24%. So, the proposed analytical method gives an accuracy of measure of 92.76%. The graphical analysis of lifetime calculation using both models, for various conditions as suggested by the Taguchi method is shown in Figure 3.
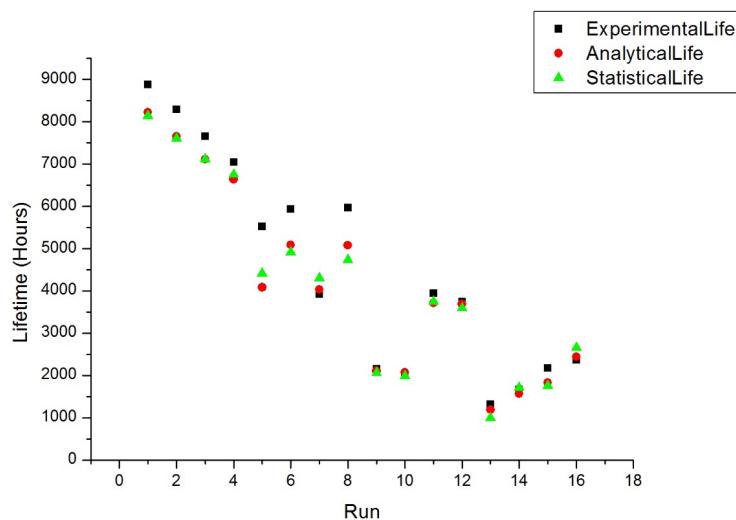
Figure 4. Graphical analysis of lifetime calculation

It has been seen that after considering the humidity factor and its effect on the performance of an electrolytic capacitor, the calculated lifetime (hours) was close to the value of lifetime (hours) that was received using the accelerated life testing method. Sometimes, it is not feasible to carry out experiments over long periods. In such cases, the analytical method can be used for failure prediction of an electrolytic capacitor.

CONCLUSION

The mathematical modelling of electrolytic capacitors was explored in this paper. The analysis of the performance of a capacitor with respect to various environmental conditions and electrical parameters were investigated. A new model was proposed for calculating residual life of an electrolytic capacitor; it considered the effect of humidity on performance and condition of the capacitor. Using the Taguchi method, experiments were designed and performed. Using the Minitab 18.1 software, the statistical method provided a regression analysis that had an average error rate of 0.28%. The corrected model of the analytical method stipulated that mean life was 4535.45 hr in comparison with the experimentally calculated mean life, which was. 4155.13 hr. After comparing the life time of the capacitor using the experimental and analytical methods, an error analysis was conducted and the error was calculated. The proposed corrective model had an error rate of 7.24% and accuracy of 92.76%, which validated the mathematical model.

REFERENCES

Ahmad, N., Islam, A., & Salam, A. (2006). Analysis of optimal accelerated life test plans for periodic inspection: The case of exponentiated Weibull failure model. *International Journal of Quality and Reliability Management, 23*(8), 1019–1046.

Ashburn, T., & Skamser, D. (2008). Highly accelerated testing of capacitors for medical applications. *Paper presented at the Proceedings of the 5th SMTA Medical Electronics Symposium.* Anaheim, California.

Bengt, A. (1995). Electrolytic capacitors theory and applications. *RIFA Electrolytic Capacitors*.

Bhargava, C., Banga, V. K., & Singh, Y. (2014). Failure prediction and health prognostics of electronic components: A review. *Paper presented at the Engineering and Computational Sciences (RAECS)*, 2014 Recent Advances in.

Dehbi, A., Wondrak, W., Ousten, Y., & Danto, Y. (2002). High temperature reliability testing of aluminum and tantalum electrolytic capacitors. *Microelectronics Reliability, 42*(6), 835–840.

Elbert, H. F., & March-Force, G. (1989). Environmental testing facility for electronic components. *U.S. Patent No. 4,871,965*. Washington, DC: U.S. Patent and Trademark Office.

Gulbrandsen, S., Arnold, J., Caswell, G., & Cartmill, K. (2014). Comparison of aluminum electrolytic capacitor lifetimes using accelerated life testing. *Paper presented at the International Symposium on Microelectronics*.

Held, M., & Fritz, K. (2009). Comparison and evaluation of newest failure rate prediction models: FIDES and RIAC 217Plus. *Microelectronics Reliability, 49*(9), 967–971.

Huang, X., Denprasert, P. M., Zhou, L., Vest, A. N., Kohan, S., & Loeb, G. E. (2017). Accelerated life-test methods and results for implantable electronic devices with adhesive encapsulation. *Biomedical Microdevices, 19*(3), 46-57.

Jano, R., & Pitica, D. (2012). Accelerated ageing tests of aluminum electrolytic capacitors for evaluating lifetime prediction models. *Acta Technica Napocensis, 53*(2), 36-41.

Kalaiselvan, C., & Rao, L. B. (2016). Accelerated life testing of nano ceramic capacitors and capacitor test boards using non-parametric method. *Measurement, 88,* 58–65.

Kececioglu, D. (2002). *Reliability and life testing handbook* (Vol. 2). United States of America, USA: DEStech Publications, Inc.

Kötz, R., Hahn, M., & Gallay, R. (2006). Temperature behavior and impedance fundamentals of supercapacitors. *Journal of Power Sources, 154*(2), 550–555.

Kulkarni, C., Biswas, G., Celaya, J., & Goebel, K. (2011). *Prognostic modeling and experimental techniques for electrolytic capacitor health monitoring*. In *8th International Workshop on Structural Health Monitoring 2011 (IWSHM)*. Stanford University, Stanford, CA.

Ma, H., & Wang, L. (2005). Fault diagnosis and failure prediction of aluminum electrolytic capacitors in power electronic converters. *Paper presented at the Industrial Electronics Society, 2005. IECON 2005*. 31st Annual Conference of IEEE.

Naikan, V., & Rathore, A. (2016). Accelerated temperature and voltage life tests on aluminium electrolytic capacitors: A DOE approach. *International Journal of Quality and Reliability Management, 33*(1), 120–139.

Parler Jr, S. G., & Dubilier, P. C. (2004a). Deriving life multipliers for electrolytic capacitors. *IEEE Power Electronics Society Newsletter, 16*(1), 11–12.

Parler Jr, S. G., & Dubilier, P. C. (2004b). Reliability of CDE aluminum electrolytic capacitors. *Application Note, Cornell-Dubilier Electronics*. Retrieved from http://www.cde. com/tech/reliability. pdf

Saha, B., Goebel, K., & Christophersen, J. (2009). Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control, 31*(3–4), 293–308.

Sankaran, V., Rees, F., & Avant, C. (1997). *E*lectrolytic capacitor life testing and prediction. *Paper presented at the Industry Applications Conference, 1997. Thirty-Second IAS Annual Meeting, IAS'97*, Conference Record of the 1997 IEEE.

Sinnadurai, F. (1974). The accelerated ageing of plastic encapsulated semiconductor devices in environments containing a high vapour pressure of water. *Microelectronics Reliability, 13*(1), 23–27.

Stevens, J. L., & Dapo, R. F. (2002). Electrolytic capacitor cover method and materials for the manufacture thereof. *U.S. Patent No. 6,370,016*. Washington, DC: U.S. Patent and Trademark Office.

Whitman, C. S. (2012). Impact of ambient temperature set point deviation on Arrhenius estimates. *Microelectronics Reliability, 52*(1), 2–8.

Comparison Extraction of Peanut Skin between CO_2 Supercritical Fluid Extraction and Soxhlet Extraction in Term of Oil Yield and Catechin

Nicky Rahmana Putra[1], Mohd Azizi Che Yunus[1,2]*, Muhammad Syafiq Hazwan Ruslan[3], Zuhaili Idham[2] and Fadillah Nur Idrus[2]

[1]Centre of Lipid Engineering and Applied Research (CLEAR),
Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia

[2]Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia

[3]Department of Chemical Engineering, Universiti Teknologi Petronas, 32610 UTP, Bandar Seri Iskandar, Perak, Malaysia

ABSTRACT

The objective of this study was to compare the extraction of peanut skin using supercritical carbon dioxide extraction and soxhlet extraction in terms of oil yield and catechin. For supercritical fluid extraction, the temperatures used were 40 and 70°C, while pressure used was 10 and 30 MPa, the flow rate was CO_2 3 mL/min, and the concentration of co-solvents was 0 and 5%. Meanwhile, for soxhlet extraction, the extraction time was 6 hr with ethanol, hexane and water as the solvents. The results showed that soxhlet extraction gave the highest yield of extract (36.282%) using ethanol as solvent as compared with supercritical CO_2 extraction (15.47%) at pressure 30 MPa, temperature 70°C and 5% concentration of co-solvent. This study reveals that the extracts from SC-CO_2 extraction yielded the highest amount of catechin (208.73 µg/g sample) compared with that yielded in the soxhlet extraction (42.24 73 µg/g sample) with distilled water as a solvent analysed with High Performance Liquid Chromatography (HPLC).

Keywords: Catechin, peanut skin, soxhlet extraction, supercritical fluid extraction

INTRODUCTION

Peanuts are usually served as a side dish in Asian countries, especially Malaysia and Indonesia. Approximately 0.74 million metric tons of peanuts are globally produced and the quantity is still increasing (Sobolev & Cole, 1999). Peanut skin is usually peeled off the

seed because peanut skin gives an astringent taste to dishes. Peanut skin is commonly used for forage in villages. In the peanut butter industry, peanut skin is removed from peanuts because it reduces the quality of peanut butter in terms of taste. Peanut skin is a renewable raw material; it contains antioxidant compounds that protect human health and enhance a sustainable environment (Hoang, Dostálová, Pudil, & Pokorný, 2008).

Although peanut skin is a product waste, it is a rich source of antioxidants such as phenolics, flavonoids and tannins. Peanut skin contains procyanidins, epicatechin, oleic acid and catechin (Hoang, Apostolova, Dostalova, Pudil, & Pokorny, 2008; Nepote, Grosso, & Guzman, 2002; Sobolev & Cole, 1999; Yu et al., 2006). In this study, catechin was detected in the extract of peanut skin. Most antioxidant activities have been associated with reduced risk of cardiovascular diseases, antidiabetic indications and anti-inflammatory effects; antioxidants also help in the prevention of cancers (Nepote et al., 2002).

Carbon dioxide supercritical fluid extraction (SFE) is a green technology that was developed to extract oil, antioxidants and bioactive compounds from plants and herbs. This method has been used by many researchers to effectively extract bioactive compounds from plants and herbs such as lycopene from tomato skin (Kassama, Shi, & Mittal, 2008) and djencolic acid from *Pithecellobium Jiringan* (Jack) prain seeds (Yunus et al., 2013). Carbon dioxide supercritical fluid extraction is beneficial as it allows the extract to retain high purity of solute content, keeps it free of organic solvents and allows easy separation of the extract from the solvent. Its important advantage is that it provides high amounts of bioactive compounds that are available in the extract because supercritical carbon dioxide uses low critical temperature and pressure, ensuring that bioactive compounds are not easily degraded. Based on the temperature and pressure, the solubility of supercritical carbon dioxide as the solvent can be manipulated in order to obtain the maximum yield of the selected compound (Yunus et al., 2012). The presence of ethanol is needed to enhance the polarity of supercritical carbon dioxide as a solvent because ethanol can extract two kinds of compound, polar and nonpolar, found in the solute but carbon dioxide can only extract nonpolar compounds (Trabelsi et al., 2016).

Soxhlet extraction is a conventional method to extract compounds from herbs and plants. The benefit of soxhlet extraction is that this process exhausts the extraction process, providing a high amount of yield compared with other methods. However, soxhlet extraction yields only low quality extracts because compounds in the extracts are degraded due to the high temperature needed for this extraction process (De Castro & García-Ayuso, 1998). In this research, the aim of the study was to compare the extraction of peanut skin using supercritical fluid extraction and soxhlet extraction in order to obtain the maximum amount of oil and catechin.

MATERIALS AND METHOD

Materials

Raw peanut skin was obtained from G-Tachfood Industries Sdn Bhd, Johor Bahru, Malaysia. The peanut skin was dried in an oven at 60°C for 4 hr and then blended into powder, classified as 425 µm size of particle with shaving process, stored in a plastic sample and placed in the

freezer until it was ready to be used. Catechin was purchased from Sigma-Aldrich (St. Louis, MO). Denatured ethanol 95%, absolute ethanol 99.86%, N-hexane and distilled water were purchased from Fisher Scientific (Atlanta, GA).

Soxhlet Extraction

Soxhlet extraction was carried out to compare the extraction performance with SC-CO$_2$ extraction. Denatured ethanol 95%, N-hexane and distilled water were used to extract the peanut skin. A volume of 100 mL of solvent was placed in a thimble that was put into the soxhlet apparatus containing 5.0+0.005 g of peanut skin powder. The extraction process was done for 6 hr at a temperature based on the boiling point of each solvent; for N-hexane this was 68°C, for denatured ethanol it was 78°C and for water it was 100°C. Vacuum-drying temperature was set at 40°C and pressure was set at 80 mBar in order to avoid degradation of bioactive compounds (Danlami, Zaini, Arsad, & Yunus, 2015).

Supercritical Carbon Dioxide Extraction

The extraction process was performed at temperature 40 and 70°C, with a pressure of 10 and 30 MPa, co-solvent concentration of 5% (Vethanol/Vtotal) and CO$_2$ flow rate of 3 ml/min. The extraction time of this method was 180 min. The chiller temperature was set at 6°C, while the heater on the back pressure regulator (Jasco BP 2080 Plus Automated BPR) was set at 50°C. Next, 5 ±0.005 g of peanut skin was placed in an extraction vessel. Then, liquid CO$_2$ was continuously pumped from the CO$_2$ tank into the system with a supercritical pump at a flow rate of 3 mL/min; ethanol 98.86% as modifier was also pumped. The extracted oil was collected in a vial and recorded every 30 min of the extraction process. After each extraction process, the extract obtained was sealed and stored at 4°C to prevent any possible degradation (Yunus et al., 2013).

Analysis of Catechin by High Performance Liquid Chromatograph (HPLC)

The analysis of catechin was developed by Chang and Wu (2011). High performance liquid chromatogram (HPLC) with ultraviolet-visible detection (Perkin Elmer Series 200, Connecticut, United States of America) was used to identify catechin. The column temperature controller was set at 30°C and the detection of wavelength was set at 210 nm. The injection volume and solvent flow rate was set at 10 µL and 1.0 mL/min, respectively. The RP C18) Merck, Darmsract, Germany – LiChrosper ® 1100 NH$_2$ column with C$_{18}$ guard column was used, while HPLC grade methanol (A) and 0.5% ortho-phosphoric acid in water were used in the mobile phase. The programme was set up as follows: A:B (20:80 v/v for minute 0 to 5, linear gradient to 24/76 v/v at minute 5 to 7, hold at 24:76 v/v at minute 7 to 10 and back 20:80 v/v minute 10 to 15).
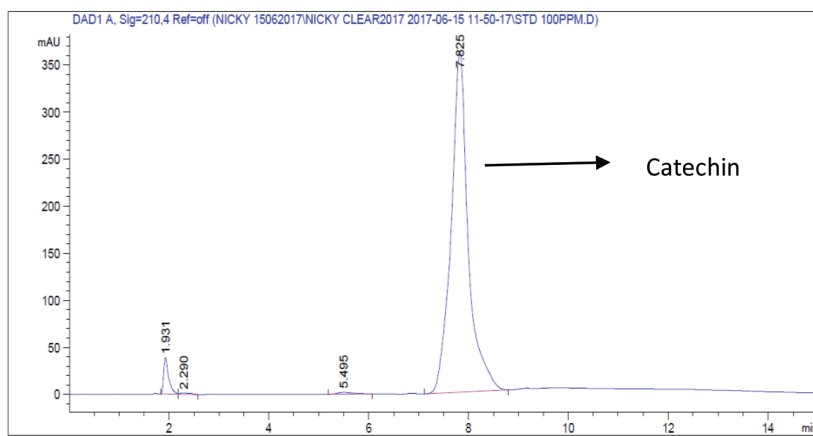
Figure 1. Peak area of catechin standard at concentration of 100 ppm
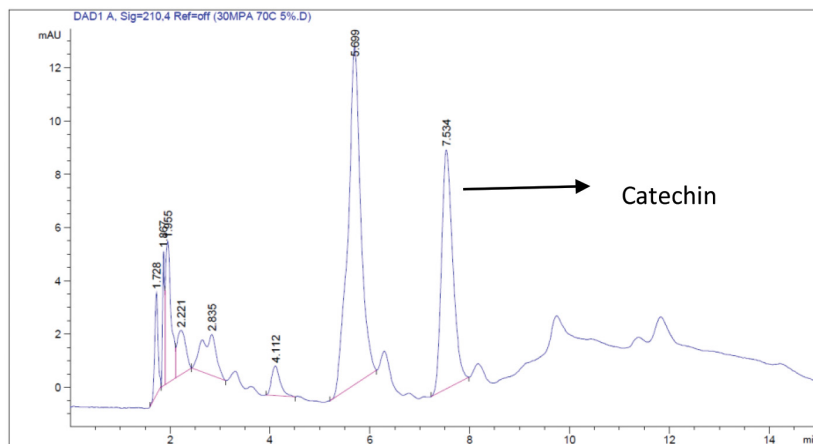
Figure 2. Peak area of catechin at 30 MPa, 70°C and 5% ethanol
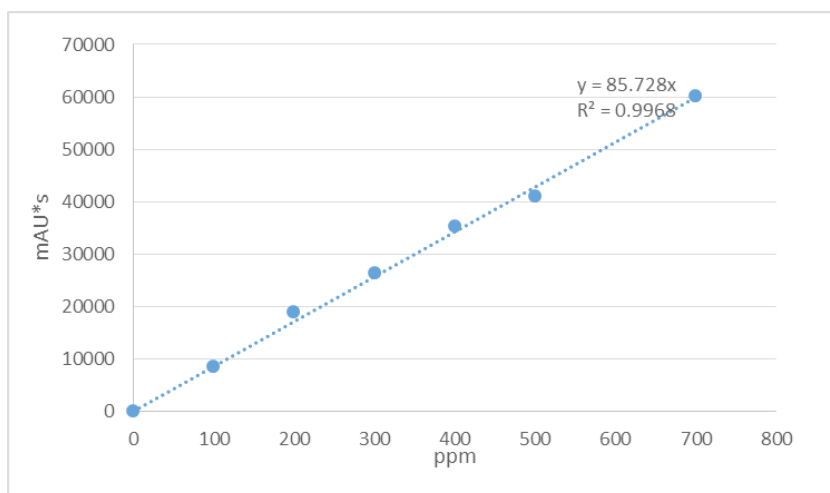
Figure 3. Calibration curve of the catechin standard

Calculation of Yield Extract

The extract yield was calculated using Equation (1):

$$\text{Extract Yield (\%)} = \frac{W_x}{W_{xy}} \times 100\,\% \tag{1}$$

where, W_x is mass of dry extract in gramme and W_{xy} is mass of sample in gramme. Based on the calculation of the extract, the extract yielded from using soxhlet extraction and supercritical carbon dioxide was compared.

Quantification of Catechin by High Performance Liquid Chromatograph (HPLC)

The slope on the calibration curve was used to denote quantification of catechin. The catechin peak area of the sample was substituted in the equation of calibration of the catechin standard curve. The equation of calibration standard is written as follows:

$$\text{Peak Area (mAU*s)} = 85.73*x \tag{2}$$

where, x is the concentration of catechin (mg/g_{sample}).

RESULTS AND DISCUSSION

Soxhlet Extraction

In this study, 5 ±0.005 g of peanut skin was used in soxhlet extraction, with 100 mL of N-hexane, denatured ethanol and distilled water used as solvent. The extraction time was 6 hr.
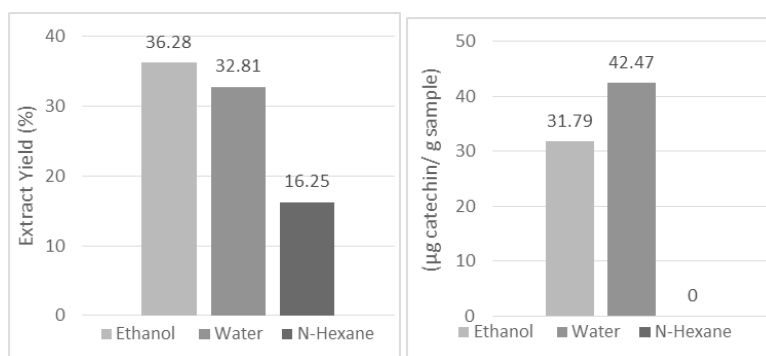
Figure 4. Extract yield (%) of soxhlet extraction using N-hexane, denatured ethanol and distilled water

The results obtained using the soxhlet extraction method are reported in Figure 4. Generally, every solvent used gives a high percentage yield in the range between 16.25% and 37%. Figure 1 shows that ethanol gave the highest amount of extract yield (36.28 ± 3.4%) compared with N-hexane and distilled water due to the polarity of the solvent. Ethanol is a bipolar solvent that can extract and make a bond between polar and nonpolar compounds in a solute (Mandana et

al., 2012). As a result, the extract yield from using water (32.81 ± 2.3%) was lower than from using denatured ethanol. Based on the polarity, polarity of water as solvent was higher than that of denatured ethanol and hexane, but the extract yield of water was not much more than ethanol and catechin in the peanut skin extract obtained from using distilled water (42.47 μg catechin/g sample). The amount was higher than the amount of extract obtained from using ethanol (31.79 μg catechin/g sample). N-hexane had the smallest extract yield (16.25 ± 1.3%); in addition, catechin was not extracted from the peanut skin because N-hexane is a nonpolar solvent. Polar solvents give higher results in the extraction of plants and herbs than nonpolar solvents because most of the components in plants and herbs are polar compounds; this creates difficulties for N-hexane when it is used to extract bioactive compounds in the solute (Saim, Dean, Abdullah, & Zakaria, 1997). Furthermore, peanut skin contains high oil content, thus, ethanol as a bipolar compound could extract the oil in the solute, but water as a pure polar solvent could not extract the oil in the solute. Therefore, ethanol could give a higher percentage of extract compared with water. However, water gave the higher concentration of catechin in the extract because water and catechin are polar compounds that bond easily lift catechin from the solute. This finding is similar to that reported in the extraction of *Artocarpus heterophyllus* L using various solvents. Ethanol as solvent gave the highest percentage of antioxidant activity (43%), total flavonoids (381.4 mgQE /g dry extract) and total phenolic compound (79 mg GAE/g dry extract) compared with N-hexane as solvent (Daud, Fatanah, Abdullah, & Ahmad, 2017). Siilarly, extraction of blackberry with ethanol as solvent led to high antioxidant activity and phenolic compound extraction (Wajs-Bonikowska et al., 2017).

Supercritical Carbon Dioxide Extraction

Figure 5(a) illustrates the percentage of yield of SC-CO_2 extraction with ethanol as co-solvent in different pressure conditions of 10 and 30 MPa, temperature of 40 and 70°C, flow rate of CO_2 3 mL/min and co-solvent concentration of 5% ($V_{ethanol}/V_{total}$). The highest extracted yield obtained (15.528%) was at the highest pressure of 30 MPa, temperature 70 °C and 5% concentration of co-solvent ethanol. Meanwhile, the lowest yield extract obtained (6.692%) was at the lowest pressure, 30 MPa, temperature 70 °C and did not use ethanol as co-solvent. In this work, an increase in pressure increased the extract due to density and solubility. Increasing pressure will increase the density of carbon dioxide and the solubility of solvent. Increasing density enhances the amount of carbon dioxide in the solvent, and this helps to create a bond between the solvent and the extract in the solute as increasing the solubility of the solvent can lift the extract. This condition is similar to that needed for extraction of *Pithecellobium Jiringan* (Jack) prain seeds with supercritical carbon dioxide as an increase in the pressure and temperature increased the solubility and yield extract (Yunus et al., 2013).
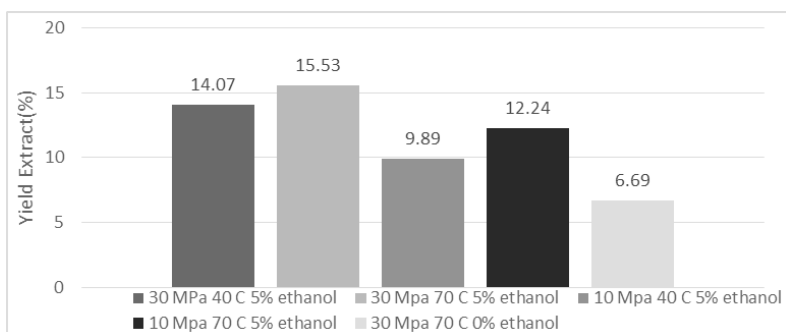
(a)
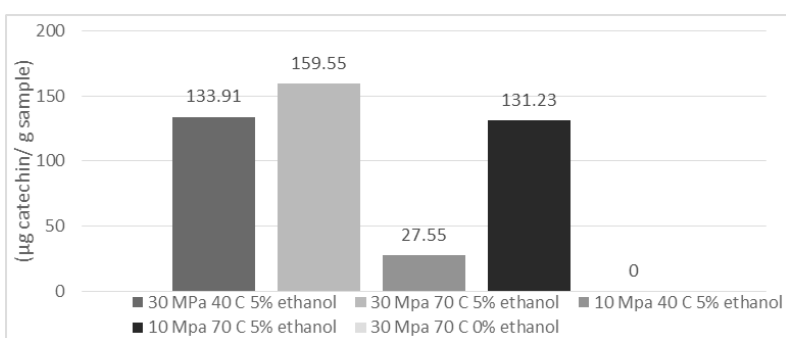
(b)

Figure 5(a), (b). Yield extract and concentration of catechin of supercritical carbon dioxide extraction with co-solvent and without co-solvent ethanol

Increasing the temperature from 40°C to 70°C at the low pressure of 10 MPa and high pressure of 30 MPa increased the yield and concentration of catechin in the extract due to the solute vapour pressure, which contributed to the increase of the mass transfer between the extract and solvent. Without ethanol as co-solvent, the amount of yield was lower as ethanol is needed to enhance the solvating power of the solvent and the polarity of the solvent (Mohd-Setapar, Yian, Yunus, Muhamad, & Zaini, 2013). Increasing the polarity gave the solvent power to extract polar and nonpolar compounds in the solute, while ethanol enhanced the porosity

of the solute, helping the solvent to penetrate the solute to lift the extract. Supercritical carbon dioxide was not adequate for penetrating the inner solute of peanut skin, but ethanol as modifier encouraged the supercritical carbon dioxide to absorb the extract from the solute (Baumann, Ferrante, Deeg, & Bräuchle, 2001). The additional ethanol in the supercritical carbon dioxide caused the swelling of the matrix, increasing the internal volume and the surface area between the solute and the solvents (Machmudah, Shotipruk, Goto, Sasaki, & Hirose, 2006). Without a co-solvent, the supercritical carbon dioxide was unsuccessful at extracting the catechin due to differences in polarity. Catechin is a highly polar compound, but carbon dioxide is a nonpolar solvent; hence, there was no interaction between the catechin and the supercritical carbon dioxide.

Comparison Between Supercritical Carbon Dioxide Extraction and Soxhlet Extraction

The overall total percentage of oil yield obtained from optimum conditions of supercritical carbon dioxide was compared with that obtained from using soxhlet extraction with different solvents, namely, water, ethanol and N-hexane. The conditions of the supercritical carbon dioxide and co-solvent ethanol were selected at higher conditions (30 MPa, flow rate of CO_2 3 mL/min, co-solvent concentration of 0 and 5% and temperature of 40 and 70°C) in order to compare the performance of using the supercritical carbon dioxide method and other methods. The highest overall extraction of oil yield was 36.28% using soxhlet extraction with ethanol as the co-solvent, followed by water (32.82%), N-hexane (16.25%) and CO_2 supercritical fluid extraction (15.52%) at pressure 30 MPa, temperature 70°C and 5% rate of co-solvent ethanol, respectively. Finally, the lowest oil yield, 16.25%, was obtained from using N-hexane as the solvent in the soxhlet extraction as presented in Figure 6(a). The soxhlet extraction with ethanol as a co-solvent generated higher oil yield than other extraction media. This is because most of the compounds in peanut skin are polar compounds such as procyanidin, catechin and epicatechin (Yu, Ahmedna, Goktepe, & Dai, 2006). Ethanol and water are polar molecules that easily interact and extract compounds from peanut skin.
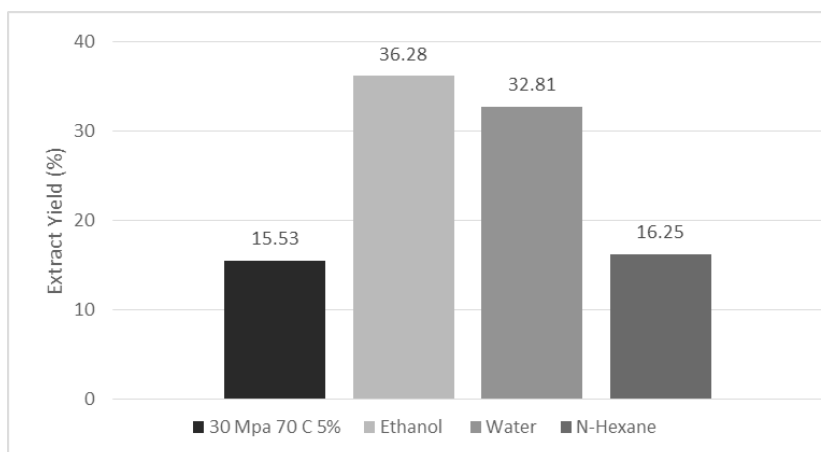
(a)
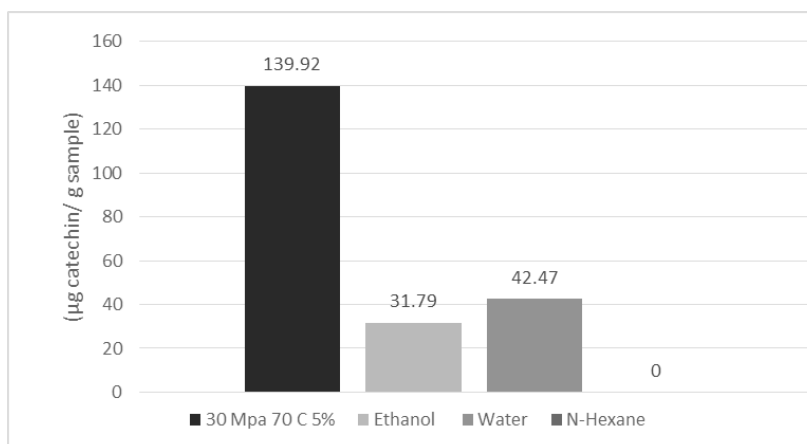
(b)

Figure 6(a),(b). Comparison of extract yield (%) and concentration of catechin in the extract using supercritical carbon dioxide and soxhlet extraction with various solvents

Although extraction of peanut skin using soxhlet extraction gave the maximum yield extract, the concentration of catechin in the extract was lower than the concentration of catechin in the extract obtained using supercritical fluid extraction as the temperature of extraction using soxhlet was relatively high compared with that used in supercritical fluid extraction. High temperatures lead to the degradation of bioactive compounds in extracts and also compromises the antioxidant activity of extracts (Hasmida et al., 2015). Bioactive compounds, especially flavonoids, are sensitive to high temperatures. This finding is similar to that obtained in previous research, which showed that catechin had degraded at below 100°C due to the decarboxylation of benzoid acid process in the structural molecules of catechin (Khuwijitjaru et al., 2014). Supercritical carbon dioxide extraction was preferable because it needed less solvent and a shorter extraction time. Moreover, the quality of the extract obtained using supercritical carbon dioxide was better than that obtained from soxhlet extraction.

CONCLUSION

In conclusion, the maximum peanut skin extract came from soxhlet extraction (36.28%) compared with using supercritical carbon dioxide extraction (15.53%). Although soxhlet extraction yielded a more extract (31.79 µg catechin/g sample), the bioactive compounds derived from this extraction were lower in quality than those extracted using supercritical carbon dioxide (139.92 µg catechin/g sample). Carbon dioxide supercritical fluid extraction was preferred in this work due to its benefits of shorter time of extraction, lower amount of solvent used and higher amount of catechin extracted. Both methods used ethanol as the extraction solvent, thus, toxicity of solvent should be ruled out.

ACKNOWLEDGEMENT

REFERENCES

Baumann, R., Ferrante, C., Deeg, F., & Bräuchle, C. (2001). Solvation dynamics of nile blue in ethanol confined in porous sol-gel glasses. *Journal of Chemical Physics, 114*(13), 5781–5791.

Chang, C. L., & Wu, R. T. (2011). Quantification of (+)-catechin and (−)-epicatechin in coconut water by LC-MS. *Food Chemistry, 126*(2), 710–717.

Danlami, J. M., Zaini, M. A. A., Arsad, A., & Yunus, M. A. C. (2015). Solubility assessment of castor (Ricinus communis L) oil in supercritical CO_2 at different temperatures and pressures under dynamic conditions. *Industrial Crops and Products, 76*, 34–40.

Daud, M. N. H., Fatanah, D. N., Abdullah, N., & Ahmad, R. (2017). Evaluation of antioxidant potential of *Artocarpus heterophyllus* L. J33 variety fruit waste from different extraction methods and identification of phenolic constituents by LCMS. *Food Chemistry, 232*, 621–632.

De Castro, M. L., & Garcıa-Ayuso, L. (1998). Soxhlet extraction of solid materials: An outdated technique with a promising innovative future. *Analytica chimica acta, 369*(1), 1–10.

Hasmida, M., Liza, M., Nur Syukriah, A., Harisun, Y., Mohd Azizi, C., & Fadzilah Adibah, A. (2015). Total phenolic content and antioxidant activity of quercus infectoria galls using supercritical CO_2 extraction technique and its comparison with soxhlet extraction. *Pertanika Journal of Science & Technology, 23*(2), 287-295.

Hoang, V. H., Apostolova, P., Dostalova, J., Pudil, F., & Pokorny, J. (2008). Antioxidant activity of peanut skin extracts from conventional and high-oleic peanuts. *Czech Journal of Food Sciences, 26*(6), 447–457.

Hoang, V. V., P. A., Dostálová, J., Pudil, F., & Pokorný, J. (2008). Antioxidant activity of peanut skin extracts from conventional and high-oleic peanuts. *Czech Journal of Food Sciences, 26*(6), 447–457.

Kassama, L. S., Shi, J., & Mittal, G. S. (2008). Optimization of supercritical fluid extraction of lycopene from tomato skin with central composite rotatable design model. *Separation and Purification Technology, 60*(3), 278–284.

Khuwijitjaru, P., Plernjit, J., Suaylam, B., Samuhaseneetoo, S., Pongsawatmanit, R., & Adachi, S. (2014). Degradation kinetics of some phenolic compounds in subcritical water and radical scavenging activity of their degradation products. *Canadian Journal of Chemical Engineering, 92*(5), 810–815.

Mandana, B., Russly, A., Farah, S., Noranizan, M., Zaidul, I., & Ali, G. (2012). Antioxidant activity of winter melon (Benincasa Hispida) seeds using conventional soxhlet extraction technique. *International Food Research Journal, 19*(1), 229-234.

Mohd-Setapar, S. H., Yian, L. N., Yunus, M. A. C., Muhamad, I.-I., & Zaini, M. A. A. (2013). Extraction of rubber (Hevea brasiliensis) seeds oil using supercritical carbon dioxide. *Journal of Biobased Materials and Bioenergy, 7*(2), 213–218.

Nepote, V., Grosso, N. R., & Guzman, C. (2002). Extraction of antioxidant components from peanut skins. *Grasas y aceites, 53*(4), 391–395.

Saim, N. a., Dean, J. R., Abdullah, M. P., & Zakaria, Z. (1997). Extraction of polycyclic aromatic hydrocarbons from contaminated soil using soxhlet extraction, pressurised and atmospheric microwave-assisted extraction, supercritical fluid extraction and accelerated solvent extraction. *Journal of Chromatography A, 791*(1), 361–366.

Sobolev, V. S., & Cole, R. J. (1999). Trans-Resveratrol content in commercial peanuts and peanut products. *Journal of Agricultural and Food Chemistry, 47*(4), 1435–1439.

Trabelsi, D., Aydi, A., Zibetti, A. W., Della Porta, G., Scognamiglio, M., Cricchio, V Mainar, A. M. (2016). Supercritical extraction from Citrus Aurantium amara peels using CO_2 with ethanol as co-solvent. *Journal of Supercritical Fluids, 117*, 33–39.

Wajs-Bonikowska, A., Stobiecka, A., Bonikowski, R., Krajewska, A., Sikora, M., & Kula, J. (2017). A comparative study on composition and antioxidant activities of supercritical carbon dioxide, hexane and ethanol extracts from blackberry (Rubus fruticosus) growing in Poland. *Journal of the Science of Food and Agriculture, 97*(11), 3576-3583.

Yu, J., Ahmedna, M., Goktepe, I., & Dai, J. (2006). Peanut skin procyanidins: Composition and antioxidant activities as affected by processing. *Journal of Food Composition and Analysis, 19*(4), 364–371.

Yunus, M., Hasan, M., Othman, N., Mohd-Septapar, S., Ahmad-Zaini, M., Idham, Z., & Zhari, S. (2013). Effect of particle size on the oil yield and catechin compound using accelerated solvent extraction. *Jurnal Teknologi (Sciences & Engineering), 60*, 21–25.

Yunus, M. C., Arsad, N., Zhari, S., Idham, Z., Setapar, S., & Mustapha, A. (2013). Effect of supercritical carbon dioxide condition on oil yield and solubility of *Pithecellobium Jiringan* (Jack) prain seeds. *Jurnal Teknologi, 60*, 45–50.

Watermelon Radioprotection against Radiation Stress in Mice

Md Saad, W. M.[1]*, Mohd Nor, N. A.[1], Abdul Razak, H. R.[2], Mat Salleh, M. S.[1], Abdul Rasid, A.[1] and Hanifah, R.[1]

[1]*Department of Medical Laboratory Technology, Faculty of Health Sciences, Universiti Teknologi MARA, UiTM Puncak Alam, 42300 Kuala Selangor, Selangor, Malaysia*

[2]*Department of Medical Imaging, Faculty of Health Sciences, Universiti Teknologi MARA, UiTM Puncak Alam, 42300 Kuala Selangor, Selangor, Malaysia*

ABSTRACT

Ionising radiation (IR) has been extensively used as therapy and diagnostic modality to detect abnormalities inside a human body. Interaction between IR and cells can lead to production of free radicals. This study aims to evaluate radioprotective properties of 50% watermelon juice against low dose ionising radiation (LDIR)-induced stress in mice lung and liver tissues following 14 days of juice supplementation. Eighteen (18) ICR mice were randomly divided into three groups, negative control (Cx), radiation (Rx) and treatment group (Tx). The Cx group was treated with normal diet and filtered water while the Rx group was given a normal diet, filtered water and irradiated with 100 µGy x-ray; Tx group was fed a normal diet, 50% watermelon juice and irradiated with 100 µGy x-ray. After 14 days, level of superoxide dismutase (SOD), reduced glutathione (GSH) and malondialdehyde (MDA) in lung and liver tissues were evaluated. The SOD inhibition activity revealed a significant decrease in Rx and Tx compared with Cx ($p<0.001$). The Rx and Tx showed significant reduction in GSH level compared with Cx ($p<0.001$) respectively. The MDA levels for lung tissues revealed a statistically significant value between Tx and Cx ($p<0.05$). The results suggest that 14 days supplementation of 50% watermelon juice was insufficient to foster radioprotective properties against LDIR-induced stress.

Keywords: Glutathione (GSH), low dose ionizing radiation (LDIR), malondialdehyde (MDA), oxidative stress, superoxide dismutase (SOD)

INTRODUCTION

Ionising radiation (IR) is a radiation energy that is capable of removing tightly bound electrons from the atomic orbital during an interaction, thus ionising the atom. According to a study. The IR exhibits wide beneficial applications in medicine involving the

management of a wide range of tumours, along with chemotherapy and surgery (Mohammad, Mohamed, Zakaria, Abdul Razak, & Saad, 2014). Despite its efficacy in killing tumour cells, radiation suffers from several major drawbacks like damage to surrounding healthy tissue and deoxyribonucleic acid (DNA) to patients and radiotherapy workers (Mansour, 2013). In a low dose radiation exposure, oxidative stress may induce modification to redox balance after a certain exposure period. Studies showed that acute exposure to IR may contribute to water radiolysis, thereby generating reactive oxygen species (ROS) (Devasagayam et al., 2004; Meydan, Gursel, Bilgici, Can, & Ozbek, 2011).

The ROS including superoxide anion (O_2-), hydroxyl radical ($OH-$) and hydrogen peroxide (H_2O_2) are produced by aerobic organism (Stice et al., 2015). Overproduction of ROS may cause oxidative damage to cellular macromolecules such as DNA, lipids, and proteins, but they are also beneficial in regulating homeostasis at cellular level in normal healthy tissues (Eltahawy, Abunour, & Elsonbaty, 2012; Mansour, 2013; Jagetia & Ravikiran, 2014; Saada, Rezk, Eltahawy, & Wiley, 2010; Stice et al., 2015). In a normal physiological condition, antioxidant will counterbalance the excessive amount of ROS via neutralisation. Imbalance between antioxidant and ROS would lead to oxidative stress and several biological consequences (Freitinger, Zölzer, Beránek, & Racek, 2012).

An antioxidant is a molecule that has the ability to scavenge free radicals and protect body from oxidative damage. Antioxidant system can be classified to enzymatic antioxidant (e.g. superoxide dismutase, catalase and glutathione peroxidase) and non-enzymatic antioxidants (e.g. glutathione, ascorbic acid, tocopherol and carotene). Those biosubstances may be assessed as markers for oxidative stress due to their participation in cellular defence system. According to a study, fruits and vegetables can provide the best antioxidant properties against the development of chronic disease (Asita & Molise, 2011).

However, there is lack of information about the radioprotective properties of watermelon as antioxidant against radiation-induced oxidative stress. Hence, the properties of watermelon juice against low dose radiation-induced oxidative stress are yet to be fully characterised. The maximum likelihood method proposed by Shang, may not be suitable for this study (Shang, 2011). Thus, this study was designed to evaluate radioprotective properties of local watermelon juice against low-dose ionising radiation (LDIR)-induced oxidative stress in lung and liver tissues of mice.

MATERIALS AND METHODS

Chemicals

OxiSelect ™ Superoxide Dismutase Activity Assay kit, OxiSelect ™ Glutathione (GSSG/GSH) Assay kit and OxiSelect ™ TBARS Assay kit (MDA Quantification) were purchased from Cell Biolabs, Inc. Other chemicals used throughout the study were purchased from Sigma-Alderich.

50% Watermelon Juice Preparation

Locally harvested watermelon (Citrullus Lanatus (Thunb) Matsum and Nakai) were purchased from the Selangor Fruit Valley's dealer in Saujana Utama, Selangor. The fruits were washed and peeled and the flesh cut into smaller pieces and blended using fruit juice extractor (Pensonic Classic Series Juice Extractor: PJ-67S). The pure fruit juice was then diluted with filtered tap water in ratio 1:1 (v/v) to obtain a 50% watermelon juice. The juice was freshly prepared twice a day for the mice's consumption.

Ethics Approval

Animal study was approved by the UiTM Committee of Animal Research and Ethics (UiTM CARE), UiTM, Puncak Alam, Selangor, with strict compliance to the guidelines. All procedures were performed under diethyl ether anaesthesia.

Animal Handling and Study Design

Eighteen (18) four-week-old male ICR mice weighing about 30g each, were obtained from Laboratory Animal Facility and Management (LAFAM), Faculty of Pharmacy, UiTM Puncak Alam, Selangor. The mice were placed in cages with corncob bedding in Animal Holding Room (FSK 1,5 at 6th floor) equipped with ventilation system, temperature of $25 + 2^oC$ and 70 air changes per hour (ACH) pressure. The mice underwent 2 weeks acclimatisation period and fed normal pellet diet with filtered tap water ad libitum. For study design, the mice were randomly placed in three groups that consisted of six mice per group, and divided into negative control group (Cx), radiation group (Rx) and treatment group (Tx). Animals from Rx and Tx were given filtered tap water and 50% watermelon juice for 14 days respectively, and exposed to 100μGy on day 15. All animals were euthanised by cervical dislocation and their liver and lung tissues were immediately harvested and kept at -80^oC for further analyses.

Superoxide Dismutase (SOD) Inhibition Activity Assay

The tissue sample was homogenised in 5 ml cold 1X Lysis Buffer per gram tissue using mortar and pestle. The sample was centrifuged at 12,000 rpm for 10 minutes. The tissue lysate supernatant was collected and assayed directly for SOD activity assay. Absorbance was read at 490 nm using POLARstar Omega Microplate Reader. The SOD inhibition activity assay was determined by inhibition percentage of the SOD activity.

Total Glutathione (GSH) Assay

One-gram tissue sample was washed with isotonic saline solution of 1X PBS with 0.16 mg/ml heparin to prevent coagulation. The tissue was blot-dried and weighed before 5% ice-cold MPA was added and homogenised using mortar and pestle. The homogenate was centrifuged at 12,000 rpm for 15 minutes at 4^oC and supernatant was collected and assayed directly for the

GSSG level. The POLARstar Omega Microplate Reader was set at 405 nm and the absorbance reading was set at one-minute interval for 10 minutes. The concentration of total GSH in samples was demonstrated by comparison to standard curve.

Lipid Peroxidation Product, MDA Assay

One-gram tissue was resuspended in 50 to 100 mg/ml PBS containing 1X butylated hydroxytoluene (BHT) and homogenised on ice using mortar and pestle. The lysate was spun at 10,000 rpm for five minutes. The supernatant was collected and assayed directly to determine TBARS level. The interaction of MDA in samples with thiobarbituric acid (TBA) at $95^{\circ}C$ were measured spectrophotometrically at 532 nm with POLARstar Omega Microplate Reader and determined by MDA standard curve.

Statistical Analysis

Data obtained was analysed by analysis of variance (ANOVA) test, followed by *post hoc* Tukey test using social science statistical package (SPSS) version 21.0. The differences were considered significant when p value was less than 0.05 ($p < 0.05$). An effective computer technology like 3D gait signatures computed from 3D data could not be implied due to several limitations such as cost, facilities and lack of expertise.

RESULTS

General Examination

After irradiation with total body x-ray, the animals were observed for behavioural and physical changes. However, no significant physical changes were observed. Gross examination for liver and lung tissues revealed no pathological defects.

SOD Inhibition Activity in Lung and Liver Tissues

Figure 1 shows SOD inhibition activity in lung and liver tissues of mice. In lung tissues, there was no statistical difference observed in pair wise comparisons. However, there was an increasing trend of SOD inhibition activity between groups. The lung Tx showed highest inhibition percentage of SOD activity (54.0064 ± 2.6362 %) compared with Rx (51.1560 ± 1.5651 %) and Cx (48.4603 ± 1.2417 %). In liver tissues, results revealed that SOD inhibition activity was significantly decreased in Rx and Tx when compared with Cx *(p<0.001)*. The mean percentage of SOD inhibition activity in liver Cx was 48.29 ± 2.74 %, Rx was 24.03 ± 3.64 % and Tx was 31.53 ± 2.97 %. There was an increasing trend of SOD inhibition activity in liver Tx compared with Rx; however, there was no statistical difference observed in the pair wise comparison.
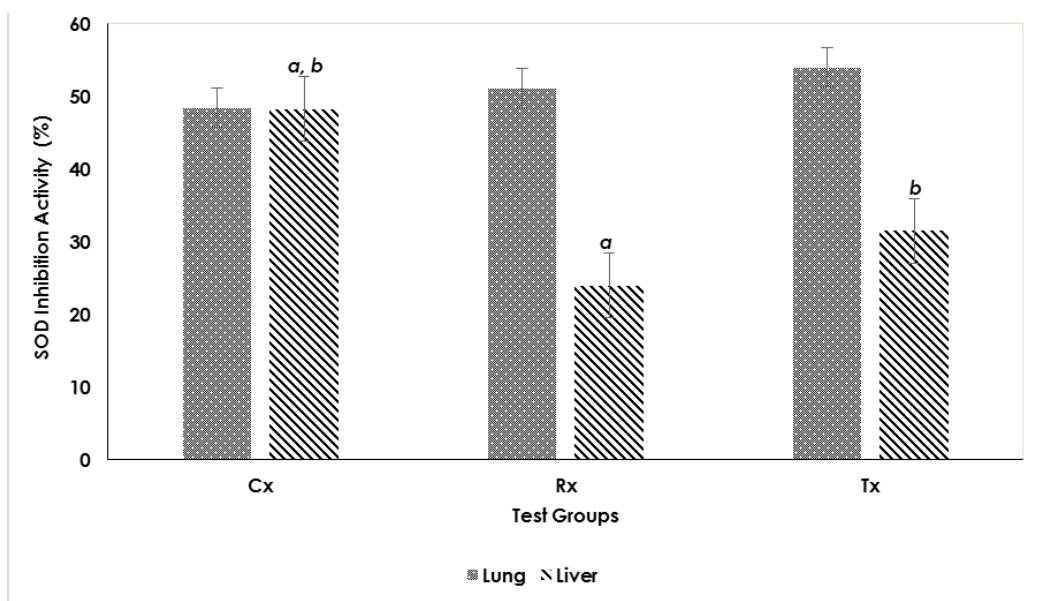
Figure 1. Determination of SOD inhibition activity in lung and liver tissues. The result represents the SOD inhibition activity (%) of Cx, Rx and Tx test groups in lung and liver tissues. Values were expressed as mean ± SEM (n=6)

[a] Indicates significant differences between Rx and Cx of liver tissues (*p<0.001*)
[b] Indicates significant differences between Tx and Cx of liver tissues (*p<0.001*)

Total Glutathione (GSH) in Lung and Liver Tissues

Figure 2 refers to the mean value of total GSH level in lung and liver tissues measured in micromolar (µM). In lung tissues, exposure to LDIR in Rx (0.5286 ± 0.0046 µM) significantly decreased the total GSH level when compared with Cx (3.7106 ± 0.0847 µM) with $p < 0.001$. Similar trend was observed in lung Tx (0.4482 ± 0.8149 µM) when compared with Cx ($p < 0.001$). GSH level in lung Tx was depleted compared with Rx; however, no significant differences were observed. In liver tissues, total GSH level showed a significant reduction of total GSH level in Rx (0.70 ± 0.04 µM) and Tx (0.73 ± 0.04 µM) when compared with control group (4.62 ± 0.29 µM) with $p<0.001$. The GSH level in liver Tx was higher compared with Rx; however, no significant differences were observed.
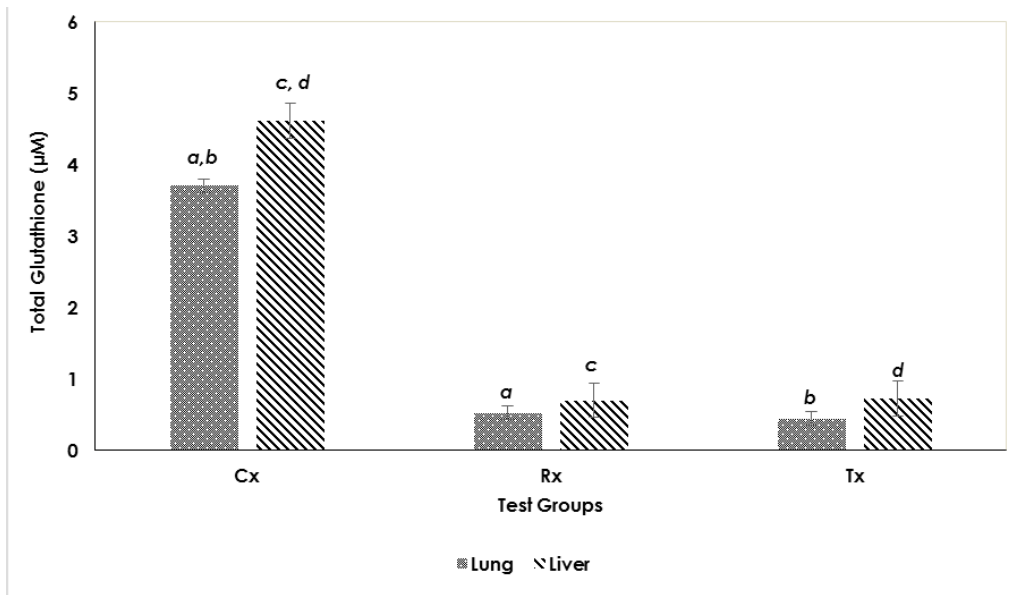
Figure 2. Determination of total glutathione (GSH) in lung and liver tissues. The result represents the total glutathione (GSH) of Cx, Rx and Tx test groups in lung and liver tissues. Values were expressed as mean ± SEM (n=6)

[a] *Indicates significant differences between Rx and Cx of lung tissues (p< 0.001)*
[b] *Indicates significant differences between Tx and Cx of lung tissues (p< 0.001)*
[c] *Indicates significant differences between Rx and Cx of liver tissues(p<0.001)*
[d] *Indicates significant differences between Tx and Cx of liver tissues (p<0.001)*

Lipid Peroxidation Product, MDA in Lung and Liver Tissues

Determination of MDA level in tissues could reflect the lipid peroxidation occurrence because it is the main product for oxidation process. Figure 3 shows that the level of MDA in lung Tx (23.6853 ± 0.4387 µM) was significantly increased compared with Cx (16.7081 ± 1.4351 µM) with $p<0.05$. There was an increasing trend of MDA level in lung Tx compared with Rx, and between Rx and Cx. However, there was no statistical difference observed in the pair wise comparisons. In liver tissues, there was no statistical difference observed in the pair wise comparisons. However, there was an increasing trend of MDA level in Tx (16.58 ± 1.00 µM) and Rx (17.38 ± 0.97 µM) when compared with Cx (16.13 ± 1.58 µM).
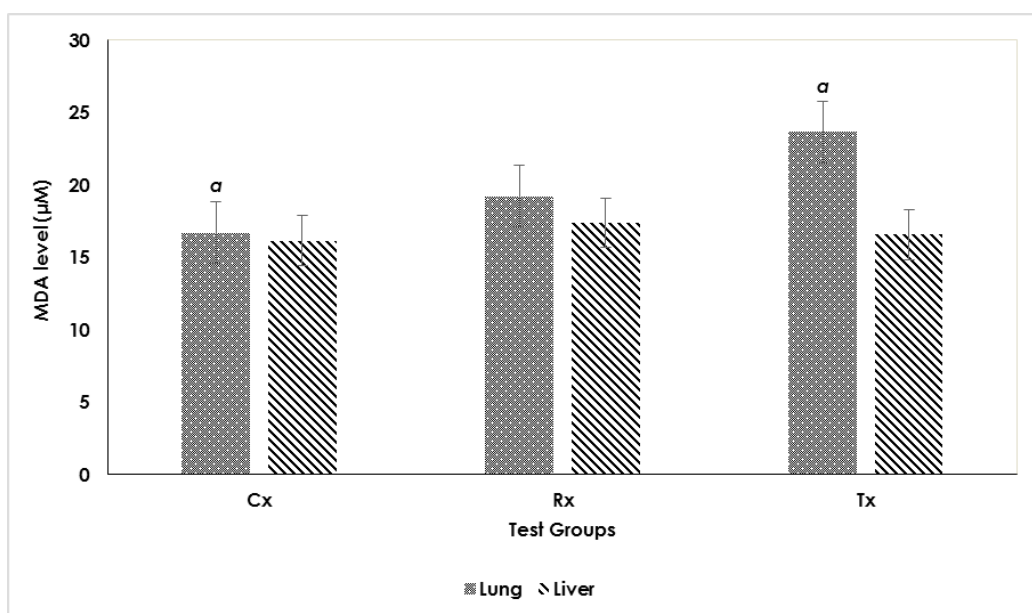
Figure 3. Determination of lipid peroxidation product, MDA in lung and liver tissues. The result represents the MDA level of Cx, Rx and Tx test groups in lung and liver tissues. Values were expressed as mean ± SEM (n=6)

[a] *Indicates significant differences between Tx and Cx of lung tissues (p< 0.05)*

DISCUSSION

The SOD as the major antioxidant enzyme, removes superoxide radicals in different cellular compartment by dismutation to oxygen and hydrogen peroxide (Kinnula & Crapo, 2003; Mohammad et al., 2014). Thus, assessment of SOD inhibition activities may reveal oxidative stress in living organisms exposed to LDIR. The determination of SOD inhibition activity in lung and liver tissues had been carried out to evaluate the level of antioxidant enzymes activity (Pan et al., 2012). In lung tissues, an increasing trend observed in SOD inhibition activities of Rx and Tx compared with Cx with no statistical significance. In liver tissues, results revealed that SOD inhibition activity were significantly decreased in Rx and Tx when compared with Cx *(p*<0.001). Results may suggest the overproduction of ROS by the LDIR exposure has taken place in both tissues but liver was more affected. The oxidative stress and redox imbalance caused by 100 µGy radiation in Rx might be well tolerated by the lung tissues of the mice by the regulation of homeostasis. The cells' response and regulation ability towards redox imbalance might vary between different organs. According to Kinula and Crapo, the lungs are directly exposed to higher oxygen concentrations than most other tissues and SOD role in protecting lung tissue has been confirmed by transgenic and knockout animal studies (Kinnula & Crapo, 2003).

Apart from SOD inhibition activities, the assessment of GSH level was one of the most reliable biomarker in evaluating oxidative stress (Sikder et al., 2013). The GSH is non-enzymatic antioxidant which become an indicator of cell health because depletion of GSH level may indicate the susceptibility to oxidant attack (Ran et al., 2014). As stated by Meydan et al., liver has highest concentration of GSH compared with other tissues (Meydan et al., 2011). In this study, exposure to LDIR in both tissues significantly decreased total GSH level in Rx and Tx when compared with Cx ($p<0.001$). In liver tissues, Tx showed an increasing trend of GSH level against Rx while lung tissue showed a contradictory effect with no statistical difference observed in the pair wise comparisons. This may indicate that exposure to LDIR significantly elevated the oxidative stress in lung and liver tissues as depicted by significant reduction of total GSH level. This was supported by a study by Meydan et al., who conducted a nearly similar study to determine the protective effect of lycopene against radiation-induced hepatic toxicity in mice and discovered that radiation group exposed to single-fraction 8Gy abdominopelvic radiotherapy significantly reduced the GSH level compared with control group (Meydan et al., 2011). In the present study, the observed decrease in GSH level in Rx during radiation exposure was mainly due to high usage of GSH at that particular period of time, to neutralise ROS in the lung and liver tissues. However, 14 days supplementation of 50% watermelon juice was not able to ameliorate the oxidative damage cause by radiation exposure, with no significant difference observed between Tx and Rx for both tissues. This was contrary to the findings of Mohammad et al. that postulated the radioprotective effect of 50% watermelon juice mitigate the GSH level of 28 days treatment group compared with 100 µGy radiation group in mice's brain, liver and lung tissues (Mohammad et al., 2014). The different findings between the two studies were clearly determined by different treatment period of the 50% watermelon juice. This suggested that 14 days pre-treatment of 50% watermelon juice was not enough to achieve a statistical significance in GSH level when compared with the radiation group.

Evaluation of MDA level in lung and liver tissues reflected the occurrence of lipid peroxidation (Saada, Said, Meky, & Azime, 2009). In lung tissues, MDA level in Tx was significantly elevated compared to Cx ($p<0.05$). There was an increasing trend of MDA level in lung Tx compared with Rx, and between Rx and Cx however, there was no statistical difference observed in the pair wise comparisons. In liver tissues, there was no statistical difference observed within any pair wise comparisons; however, MDA level increased in Tx and Rx compared with Cx. The inconsistent trends of MDA level in both tissues may postulate that the GSH had already quenched most of the ROS in the pathway that contributed to less oxidation of lipid in both tissues. This hypothesis could be supported by significant reduction in GSH level of Tx and Rx when compared with control group for both tissues. The difference between the mean values of Tx compared with Rx, in both lung and liver tissues were not statistically significances, indicating 14 days supplementation of 50% watermelon juice was insufficient to encounter lipid peroxidation. This finding was consistent with that of Saada et al. which revealed an insignificant reduction of MDA level in rat's heart and pancreatic tissues administered with grape seed extract by gavage for 14 days, following 5 Gy whole body gamma radiation (Saada et al., 2009). This may support the finding of present study indicating 14 days supplementation of 50% watermelon juice had insufficient radioprotective effect on MDA level compared with Rx to conclude that natural product pre-treatment should be longer in order to mitigate the MDA levels prior to exposure to radiation.

CONCLUSION

The present study showed that 100 µGy of LDIR are able to cause significant oxidative damage on lung and liver tissues, determined by several biomarkers. The supplementation of 50% watermelon juice for 14 days was insufficient to protect lung and liver from LDIR-induced oxidative stress at cellular level.

ACKNOWLEDGEMENT

REFERENCES

Asita, A. O., & Molise, T. (2011). Antimutagenic effects of red apple and watermelon juices on cyclophosphamide-induced genotoxicity in mice. *African Journal of Biotechnology*, *10*(77), 17763–17768.

Devasagayam, T. P. A., Tilak, J., Boloor, K. K., Sane, K. S., Ghaskadbi, S. S., & Lele, R. D. (2004). Free radicals and antioxidants in human health: current status and future prospects. *The Journal of the Association of Physicians of India*, *52*(794804), 794–804. http://doi.org/10.1016/S0300-483X(03)00149-5

Eltahawy, N., Abunour, S., & Elsonbaty, S. (2012). Effectiveness of Punica granatum Juice in Ameliorating Oxidative Damage and Ultrastructural Changes in Paneth Cells of Rat Intestine. *Journal of Pharmacy and Bioallied Sciences, 4*(2), 25–31.

Jagetia, G. C., & Ravikiran, P. B. (2014). Radioprotective Potential of Nigella Sativa Extract in Swiss Albino Mice Exposed to Whole Body γ-Radiation. *Alternative and Integrative Medicine*, *03*(4), 1–18. http://doi.org/10.4172/2327-5162.1000168

Kinnula, V. L., & Crapo, J. D. (2003). State of the Art Superoxide Dismutases in the Lung and Human Lung Diseases. *American journal of respiratory and critical care medicine, 167*(12), 1600–1619. http://doi.org/10.1164/rccm.200212-1479SO

Mansour, H. H. (2013). Protective effect of ginseng against gamma-irradiation-induced oxidative stress and endothelial dysfunction in rats. *EXCLI Journal, 12*, 766–777.

Meydan, D., Gursel, B., Bilgici, B., Can, B., & Ozbek, N. (2011). Protective effect of lycopene against radiation-induced hepatic toxicity in rats. *The Journal of International Medical Research*, *39*, 1239–52. http://doi.org/10.1177/147323001103900412

Mohammad, M. K. A., Mohamed, M. I., Zakaria, A. M., Abdul Razak, H. R., & Saad, W. M. M. (2014). Watermelon (Citrullus lanatus (Thunb.) Matsum. and Nakai) juice modulates oxidative damage induced by low dose x-ray in mice. *BioMed Research International*, *2014*, 1-6. http://doi.org/10.1155/2014/512834

Pan, J., Su, Y., Hou, X., He, H., Liu, S., Wu, J., & Rao, P. (2012). Protective effect of recombinant protein SOD-TAT on radiation-induced lung injury in mice. *Life Sciences*, *91*(3-4), 89–93.

Ran, Y., Wang, R., Gao, Q., Jia, Q., Hasan, M., Awan, M. U. F., … & Qing, H. (2014). Dragon's blood and its extracts attenuate radiation-induced oxidative stress in mice. *Journal of Radiation Research*, *55*, 699–706. http://doi.org/10.1093/jrr/rru013

Saada, H. N., Rezk, R. G., Eltahawy, N. A., & Wiley, J. (2010). Lycopene Protects the Structure of the Small Intestine against Gamma-Radiation-induced Oxidative Stress. *Phytotherapy research, 208*(November 2009), 204–208.

Saada, H. N., Said, U. Z., Meky, N. H., & Azime, A. S. A. E. (2009). Grape seed extract vitis vinifera protects against radiation-induced oxidative damage and metabolic disorders in rats. *Phytotherapy Research*, *23*, 434–438. http://doi.org/10.1002/ptr.2684

Shang, Y. (2011). Likelihood Estimation for Stochastic Epidemics with Heterogeneous Mixing Populations. *International Journal of Mathematical and Computational Sciences, 5*(7), 1041–1045.

Sikder, K., Sinha, M., Das, N., Das, D. K. R., Datta, S., & Dey, S. (2013). Moringa oleifera leaf extract prevents in vitro oxidative DNA damage. *Asian Journal of Pharmaceutical and Clinical Research*, *6*, 157–161.

Skalická, Z. F., Zölzer, F., Beránek, L., & Racek, J. (2012). Indicators of oxidative stress after ionizing and/or non-ionizing radiation: superoxid dismutase and malondialdehyde. *Journal of photochemistry and photobiology B: biology, 117*, 111-114. http://doi.org/10.1016/j.jphotobiol.2012.08.009

Stice, C. P., Liu, C., Aizawa, K., Greenberg, A. S., Ausman, L. M., & Wang, X. D. (2015). Dietary tomato powder inhibits alcohol-induced hepatic injury by suppressing cytochrome p450 2E1 induction in rodent models. *Archives of Biochemistry and Biophysics*, *572*, 81–88. http://doi.org/10.1016/j.abb.2015.01.004

SCIENCE & TECHNOLOGY

PERTANIKA JOURNALS

Utilisation of PET-CT in Oesophageal Cancer Management: A Clinician's Perspective

Abdul Razak, H. R.[1]*, Azmi, N. A.[1,2] and Vinjamuri, S.[3]

[1]*Department of Medical Imaging, Faculty of Health Sciences, Universiti Teknologi MARA, UiTM Puncak Alam, 42300, Selangor, Malaysia*

[2]*Diagnostic Imaging and Radiotherapy Programme, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 UKM, Kuala Lumpur, Malaysia*

[3]*Department of Nuclear Medicine, Royal Liverpool University Hospital, Liverpool, United Kingdom*

ABSTRACT

Positron emission tomography-computed tomography (PET-CT) is a hybrid imaging modality that plays a crucial role in detecting and managing oesophageal cancer. However, not much is known about the clinicians' perspective on its usage for oesophageal cancer. The aim of this study is to know the perspective of clinicians on the utilisation of PET-CT imaging for oesophageal cancer patients. A total of 73 clinicians with multidisciplinary clinical specialties for oesophageal cancer management were interviewed. All these clinicians had completed a survey consisting of 31 questions on; practicality, clinical efficacy and cost-effectiveness of PET-CT. The survey used Likert-scale to evaluate the responses. In terms of PET-CT practicality and clinical efficacy 39.7% - 43.8% and 47.9% - 83.5% of the sample respectively responded positively while in with regards to cost effectiveness, there was a significant difference from being neutral to having no opinion in 6 out of 9 questions. It was clear from the study that PET-CT has positive impact in the management of oesophageal cancer patients. However, issues related to expertise, availability, staffing and bureaucracy need to be addressed to improve competency and quality of services.

Keywords: Clinical efficacy, cost effectiveness, PET-CT, practicality

INTRODUCTION

The process of cancer management inclusive of diagnosing, staging, planning and monitoring treatment relies heavily on anatomic imaging using computed tomography (CT) or magnetic resonance imaging (MRI). Anatomic imaging generally has a high sensitivity of up to 47% in detecting clear structural changes but

inaccurate for differentiating abnormality as either malignant or benign (Barber et al., 2012; Royal College of Radiologist, 2005). PET-CT is a hybrid imaging modality that combines PET and CT technologies in a single scanner. PET images provide information about physiological functions of the body such as cell metabolism, while CT images provide anatomical detail of the patient (Antoch et al., 2003; Heinrich et al., 2005). Development of PET-CT scanner has a big impact in the field of nuclear medicine, particularly in staging cancer (Barber et al., 2012; Briggs, Chowdhury, Lodge, & Scarsbrook, 2011; Lardinois et al., 2003; Yap et al., 2001).

The PET-CT scan is an important diagnostic imaging technique which requires a continuous evaluation to satisfy the needs of clinicians, physicians, radiologists, and other decision makers. The most biggest advantage of PET-CT compared with other modalities is its ability to locate lesion accurately in regard to relationships with neighbouring structures and the presence of local invasion (Bar-Shalom et al., 2003). In terms of its application in oncology, physiological information have been proven to be beneficial for localising complex anatomy (El-Hariri, Gouhar, & Refat, 2012).

New technologies continue to emerge and require health care management to prioritise scarce resources. Reimbursement bodies increasingly require evidence not only for clinical efficacy but also for economic efficiency in order to make rational decisions. Introduction of new technologies in nuclear medicine like PET-CT should be evaluated through clinical efficacy and cost effectiveness. In addition, expenses incurred in the management of patients should be worth the results (Schreyögg et al., 2010). Multiple studies have shown a positive impact of PET-CT in detecting and staging oesophageal cancer (Antoch et al., 2003; Antoch et al., 2004; Bar-Shalom et al., 2003). However, only few researches focused on the clinicians' opinions. This was the first survey conducted in North West to assess the perception of clinicians who have had an important role in the management of patients with oesophageal cancer.

MATERIALS AND METHODS

Population

A total of 496 clinicians from various disciplines having experience in oesophageal cancer management were interviewed to gather their opinions and perceptions on the impact of PET-CT on clinical management of patients in 38 hospitals in North West, UK.

Survey

This is cross-sectional survey using a convenience sampling method. The approval from the North West research ethics board and Royal Liverpool Broadgreen University Hospital Trust (RLBUHT) ethics committee were obtained.

Topic

The survey questionnaire was based on three main topics: practicality, clinical efficacy and cost-effectiveness of PET-CT. It used a 5-point Likert scale, with higher scores representing

a positive impact, ranging from 1 (strong disagreement with the statement) to 5 (strong agreement with statement). Completed questionnaires were then analysed using descriptive statistical analysis.

RESULTS

A total of 134 clinicians responded with a return rate of 27%; 46% of the respondents who had no experience or current clinical practice related to oesophageal cancer management were excluded. The final sample size was 73 respondents (a response rate of 54%). The non-reply form was categorised as missing. The summary of the results is shown in Figure 1 below:

Number	Questions	0 No Answer	1 Strongly Disagree / Disagree	2	3 Neutral / No opinion	4 Agree / Strongly Agree	5
Practicality							
Q1	PET-CT studies are easily available to be included in any stage of patient management whenever needed.	1.4%	42.4%		12.3%	43.8%	
Q2	PET-CT studies enable the physician to avoid any tests or procedures that are usually will be included as part of the management?	4.1%	35.6%		20.5%	39.7%	
Q3	If PET-CT were not available, would you have done some type of alternative assessment in most of the cases?	1.4%	26%		28.8%	43.8%	
Clinical Efficacy							
Q4	PET-CT is useful for clinical studies for the detection, staging in clinical management of oesophageal cancer.	1.4%	5.4%		9.6%	83.5%	
Q5	PET-CT is useful for clinical studies for follow up on disease progression in clinical management of oesophageal cancer.	1.4%	17.8%		32.9%	47.9%	
Q6	PET-CT studies altered the initial course of clinical management for oesophageal cancer.	1.4%	9.5%		12.3%	76.7%	
Cost Effectiveness & Service Delivery							
Q7	PET-CT is cost-effective because the modality is precise for staging disease, improve treatment plans and avoid unnecessary treatments for oesophageal cancer	1.4%	9.5%		26%	63.1%	
The only drawback to include PET-CT study in patient management is the limitation to its availability due to: (Tick one or more):							
Q8a	Financial Issues	5.5%	23.2%		17.8%	53.4%	
Q8b	Availability of Isotopes	4.1%	28.7%		13.7%	53.4%	
Q8c	Access to Equipment	8.2%	28.8%		43.8%	19.1%	
(Q9a-Q9e)	Shortage of Staff						
Q9a	• PET-CT Technician	8.2%	21.9%		57.5%	12.3%	
Q9b	• Reporting Staff (Radiologist, Nuclear Medicine Specialist, etc)	6.8%	30.1%		39.7%	23.3%	
Q9c	• End-user staff (Oncologist, Surgeon, etc)	8.2%	34.2%		46.6%	11%	
Q9d	Training for staff	8.2%	28.8%		45.2%	17.8%	
Q9e	Bureaucracy (e.g.: Long waiting list)	9.6%	24.6%		41.1%	24.6%	

Figure 1. Survey shows positive agreements for practicality (min = 39.7%) and clinical efficacy (min = 47.9%) and neutral opinion for cost effectiveness (min = 39.7%)

In 9 out of 15 questions, most of the participants agreed with to the role of PET-CT in cancer management (agree or strongly agree), ranging from 39.7% up to 83.5%. Only 2 questions elicited negative responses (disagree or strongly disagree), ranging from 35.6% to 42.4%. On the topic of practicality and clinical efficacy, the majority of respondents agreed

with the statement (agree or strongly agree) and the lowest percentage of response was ranging from 1.4% to 4.1%. Trend rate analysis for cost- effectiveness was found to be an uncommon finding in comparison to other topics. The majority of responses related to t the availability of PET-CT (6 out of 9 questions) skewed towards neutral or no opinion at all (39.7% - 57.5%) in contrast to questions related in other topics as in practicality and clinical efficacy.

DISCUSSION

Practicality

The majority of clinicians agreed on the on practicality of PET-CT scans. Despite increased demands for PET-CT scans for patients in North West, the practicality for application in clinical practice is still restricted due to the limited availability of this modality within the studied area.

The answers to Question 2 and 3 indicate that PET-CT is generally considered as an additional imaging modality rather than as an alternative to other modalities, thereby it may lead to increased cost. The PET-CT availability has become a key main concern for physicians as most of them suggested that PET-CT studies are not easily accessible to be included at any stage of the oesophageal cancer patient's management. Most of the respondents disagreed that they will forgo any test with the present of PET-CT. This finding indicate that the PET-CT is not the main choice for decision makers to be included as a first line diagnostic test in the clinical management. An introduction of new technologies for diagnostic purpose should be able to prove its ability to give a positive impact on disease management compared with current pathways before it can be considered to be implemented into clinical practice. This new technology should be more clinically effective, fast and cheaper compared with current modalities in practice (Vernon et al., 2008).

The availability of PET-CT in relation to clinical demands is closely related to the high cost associated with it (Antoch et al., 2003). The PET-CT machine cost between £1 million to £2.5 million and there are yearly maintenance costs as well. This prohibitive cost makes it is quite impossible for it to be installed in every hospital, thus requiring proper planning (Royal College of Radiologist, 2005).

Clinical efficacy

For the clinical efficacy topic, this survey has indicated the level of knowledge and understanding of the participants on the current clinical guidelines and research findings. As PET-CT could change the diagnosis that has been confirmed by other tests or screening; it might affect the appropriateness of treatment option and alter the initial course of the clinical management. As a decision maker, clinicians depend on the diagnostic imaging modalities among other tests to decide on the best treatments to be prescribed. Clinicians may offer different treatment option for patient with additional information from PET-CT results.

This integrated imaging modality has an increasingly valuable role in many clinical fields such as cardiology, neurology and oncology with increased sensitivity and specificity compared to other dedicated PET, CT (Antoch et al., 2004; El-Hariri et al., 2012). A positive

agreement with the clinical efficacy topic may be due to the fact that most of the respondents have continually improved their knowledge on latest innovation or intervention introduced in clinical practice nationwide. Improvement of knowledge from training and reading materials could enrich their judgement to have a consensus that this new modality is valuable to be included in oesophageal cancer management.

Cost effectiveness

Potential savings are associated with PET-CT as a result of avoiding additional imaging examinations or invasive procedures and by helping clinicians make the optimum treatment decisions (Saif, Syrigos, Tzannou, & Makrilia, 2010). The benefits of structured cost-effectiveness evaluation include a better utilisation of limited healthcare resources, prevention of diffusion of unproven tests into widespread clinical practice, and withdrawal of diagnostic procedures that cannot be shown to be of value. The current world financial state deems limited financial availability, in which any new procedures or interventions goes beyond its clinical effectiveness before it can be implemented into clinical practice.

There are significant differences in comparison to other topics based on the majority of neutral or no opinion responses. This may be due to the fact that this topic is not something within respondent responsibilities as their expertise is more towards assessment on clinical application and practicality of the imaging procedure. It is in agreement with various studies in which clinicians viewed their clinical responsibilities patient management as taking precedence over strategic management role (Dopson & Stewart, 1993; Dopson, Stewart, & Templeton, 1989; Thomas & Dunkerley, 1999). Implementation of interdisciplinary involvement in strategic management and planning is essential to ensure continuous development of institutions.

Introduction of new technology aside from the need for new staff also brings together the issues of expertise and training of staff assigned to ensure continuous competency and quality of delivered services (Royal College of Radiologist, 2005). Rapid expansion of PET-CT installation worldwide requires a growing number of personnel and specialised training for them in order to keep up with the development of this modality, maintain level of competency and services provided to the patients. Personnel involved include, but is not limited to, the radiologist, nuclear medicine specialist, radiographer, nuclear medicine technologist and specialist nurses.

CONCLUSION

Findings of this study show benefits of this integrated imaging modality may include changes in patient management. This study indicates some new directions for future research as one of the important role of clinicians is to provide the clinical implication reports so that non-clinical decision can be made related to finance and infrastructures by the administrator. This survey will be a benchmark for continuity of future surveys to measure any new intervention or diagnostic procedure in compliance with clinical guidelines and compatibility with current clinical practice. Therefore, a continuous improvement in services, facilities, staffing and overall process of management could be achieved and sustained.

This review showed clinical efficacy and practicality, which include direct facilitation of clinicians expertise knowledge and background exposure, had more positive outcomes than cost effectiveness alone. However, it is unclear whether this is a function of the format of knowledge delivery, the length of expertise experiences of the clinicians and/or the nature of the clinicians-policy maker relationship. Variability between themes in the outcomes measured was also a confounding factor. This study used exclusively cost effectiveness of PET-CT theme despite the fact that the proximal aim of study was to see the impact on cost effectiveness as a whole. Future research should aim to control for some of these confounding variables, as well as including both clinicians and policy makers, in order to gain further insight into the impact of cost effectiveness at the practical level. As previously identified, there is considerable heterogeneity in the nature of different themes evaluated in this study in terms of both clinicians experience and background of expertise. Clearly, this variation introduces multiple confounding variables when comparing results of different themes.

Development of hybrid imaging is an evolution from standalone imaging developed over the past 50 years. This new technology has revolutionised the way older and especially younger generation of clinicians perceive imaging. Hybrid images providing superimposition of radiotracer signals on a set of CT or MRI in oesophageal cancer for instance has become a much preferred visualisation tools for the clinicians.

The PET CT also has aesthetic advantages over side by side data or independent modalities in terms of increased diagnostic accuracy and clinicians' confidence in localising abnormalities. High cost and restricted availability of PET in many parts of the world has made it been generally reserved only for cases with equivocal standalone imaging results. However, in the oesophageal and other oncological cases, major contribution of FDG PET is to avoid futile procedure in patients with positive metastatic from PET results that shows false negative in other standalone imaging modalities irrespective of whether there was any equivocation of these modalities. Hybrid imaging with more precise characterisation of the nature and location of abnormalities is likely to further improve diagnostic performance and thereby, treatment selection and planning. This will not only improve patient care but also reduce cost by avoiding futile treatment interventions, despite the higher upfront cost of the imaging component of the management paradigm if PET-CT and other hybrid imaging tests were used as the primary diagnostic test. There is increasing evidence across a broad range of indications that hybrid PET-CT are usually more accurate that this modality allows when compared with side-by-side comparison of each modality acquired separately. A large and increasing number of studies published on hybrid modalities will continue to inform older and younger generations of clinicians' perspective on clinical efficacy, practicality and cost effectiveness with no doubt of turning back to stand alone modalities. The evolution of hybrid imaging technologies is essential for the future of cancer imaging.

This study gives an insight into the values of clinical, cost effectiveness and practicality from the perspective of clinicians that might influence the administrative power to spend on new modalities or procedures. However, more in-depth research is required to contribute to this study's initial findings on the clinicians' perceived views on PET-CT for oesophageal cancer. Further research exploring the linkages between the clinicians' input on the clinical impact of newer modality and its effect on the administrative spending power decision making is also worthwhile.

ACKNOWLEDGEMENT

REFERENCES

Antoch, G., Beyer, T., Freudenberg, L. S., Muller, S. P., Bockisch, A., & Debatin, J. F. (2003). PET/CT or CT/PET? A Radiologist's Perspective. *Electromedica -erlangen-, 71*(1), 64-69.

Antoch, G., Saoudi, N., Kuehl, H., Dahmen, G., Mueller, S. P., Beyer, T., … & Freudenberg, L. S. (2004). Accuracy of whole-body dual-modality fluorine-18-2-fluoro-2-deoxy-D-glucose positron emission tomography and computed tomography (FDG-PET/CT) for tumor staging in solid tumors: comparison with CT and PET. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 22*(21), 4357-4368.

Bar-Shalom, R., Yefremov, N., Guralnik, L., Gaitini, D., Frenkel, A., Kuten, A., … & Israel, O. (2003). Clinical performance of PET/CT in evaluation of cancer: additional value for diagnostic imaging and patient management. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 44*(8), 1200-1209.

Barber, T. W., Duong, C. P., Leong, T., Bressel, M., Drummond, E. G., & Hicks, R. J. (2012). 18F-FDG PET/CT has a high impact on patient management and provides powerful prognostic stratification in the primary staging of esophageal cancer: a prospective study with mature survival data. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 53*(6), 864-871.

Briggs, R. H., Chowdhury, F. U., Lodge, J. P. A., & Scarsbrook, A. F. (2011). Clinical impact of FDG PET-CT in patients with potentially operable metastatic colorectal cancer. *Clinical Radiology-Edinburgh-, 66*(12), 1167-1174.

Dopson, S., & Stewart, R. (1993). Information technology, organizational restructuring and the future of middle management. *New Tech Work Empl New Technology, Work and Employment, 8*(1), 10-20.

Dopson, S., Stewart, R., & Templeton, C. (1989). *What is happening to middle management?* Oxford: Templeton College.

El-Hariri, M. A., Gouhar, G. K., & Refat, A. M. (2012). Integrated PET/CT in the preoperative staging of lung cancer: A prospective comparison of CT, PET and integrated PET/CT. *The Egyptian Journal of Radiology and Nuclear Medicine, 43*(4), 613-621.

Heinrich, S., Goerres, G. W., Schäfer, M., Sagmeister, M., Bauerfeind, P., Pestalozzi, B. C., … & Clavien, P. A. (2005). Positron emission tomography/computed tomography influences on the management of resectable pancreatic cancer and its cost-effectiveness. *Annals of surgery, 242*(2), 235-243.

Lardinois, D., Weder, W., Hany, T. F., Kamel, E. M., Korom, S., Seifert, B., … & Steinert, H. C. (2003). Staging of NonSmall-Cell Lung Cancer with Integrated Positron-Emission Tomography and Computed Tomography. *England Journal of Medicine, 348*(25), 2500-2507.

RCR. (2005). *PET-CT in the UK : a strategy for development and integration of a leading edge technology within routine clinical practice*. London: Royal College of Radiologists.

Saif, M. W., Syrigos, K., Tzannou, I., & Makrilia, N. (2010). Role and cost effectiveness of PET/CT in management of patients with cancer. *Journal of Biology and Medicine, 83*(2), 53-65.

Schreyögg, J., Weller, J., Stargardt, T., Herrmann, K., Bluemel, C., Dechow, T., … & Buck, A. K. (2010). Cost-effectiveness of hybrid PET/CT for staging of non-small cell lung cancer. *Journal of nuclear medicine: official publication, Society of Nuclear Medicine, 51*(11), 1668-1675.

Thomas, R., & Dunkerley, D. (1999). Careering Downwards? Middle Managers' experiences in the downsized Organization. *British Journal of Management, 10*(2), 612-619.

Vernon, M. R., Maheshwari, M., Schultz, C. J., Michel, M. A., Wong, S. J., Campbell, B. H., … & Wang, D. (2008). Clinical Outcomes of Patients Receiving Integrated PET/CT-Guided Radiotherapy for Head and Neck Carcinoma. *International Journal of Radiation Oncology, Biology, Physics., 70*(3), 678-684.

Yap, C. S., Seltzer, M. A., Schiepers, C., Gambhir, S. S., Rao, J., Phelps, M. E., … & Czernin, J. (2001). Impact of whole-body 18F-FDG PET on staging and managing patients with breast cancer: the referring physician's perspective. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 42*(9), 1334-1337.

SCIENCE & TECHNOLOGY

PERTANIKA
JOURNALS

A Mono-Window Algorithm for Land Surface Temperature Estimation from Landsat 8 Thermal Infrared Sensor Data: A Case Study of the Beas River Basin, India

Gopinadh Rongali*, Ashok Kumar Keshari, Ashvani Kumar Gosain and Rakesh Khosa

Department of Civil Engineering, Indian Institute of Technology Delhi, New Delhi, Delhi 110016, India

ABSTRACT

Land surface temperature (LST) is estimated using thermal infrared remote sensing data, which record the apparent temperature of the earth's surface by measuring the radiant energy of its surface. However, it is also possible to estimate LST through satellite images and image processing software. The Landsat 8 satellite was successfully launched in 2013, with two thermal infrared bands for continuous earth observations to provide for the estimation of LST. However, the calibration notifications issued by the United States Geological Survey (USGS) indicate that the data from the Landsat 8 thermal infrared sensor (TIRS) Band 11 show large uncertainty and thus, it was suggested to use TIRS Band 10 data as a single spectral band for LST estimation. In this study, we present a mono-window (MW) algorithm for LST estimation from the Landsat 8 (Path-147 and Row-38) using TIRS Band 10 data with a 100-m resolution. Emissivity was derived with the help of the normalised difference vegetation index (NDVI) proportion of vegetation technique for which operational land imager (OLI) Bands 4 and 5 (30-m resolution) were used. The results show that the LST was higher in the regions of barren land but lower in snow-covered areas. Further, the LST results were also compared with the air temperature data and they were found to be in good agreement. The MW algorithm presented in the study could be used as an efficient method for LST estimation from the Landsat 8 TIRS Band 10 data.

Keywords: Land surface temperature, landsat 8 TIRS, mono-window algorithm, NDVI, OLI

INTRODUCTION

Remote sensing is the acquisition of information about an object from the electromagnetic spectrum without making physical contact with an object. Nowadays, advanced satellite data has widely been used in environmental and climate change

studies. Thermal bands of Landsat 8 are significant data for land surface temperature (LST) calculation. LST has been identified as a significant variable of microclimate and radiation transfer within the atmosphere (Rajendran & Mani, 2015). Further, land use/land cover (LULC) of an area is also an important factor in the estimation of LST. Natural and anthropogenic activities change the physical and biological conditions of a region, and this in turn affects the LST of that area. With a change in LST value, the local climate of an area also changes (Rajeshwari & Mani, 2014). The traditional way of surface temperature estimation, such as through monitoring by a meteorological department weather station and other public and private sector observation methods, is not feasible for all types of terrain condition; in addition, they are also time consuming. However, remote sensing satellites can provide data for any topographic and climatic condition of a region, especially distinctive local climates (microclimates) produced by different land surfaces. The National Aeronautics and Space Administration (NASA) Landsat programme provides different spatial resolution satellite images for various time periods. Landsat 8 captures two different sets of images, one from OLI with nine bands (with 30-m resolution) and the second from TIRS with two bands (Band 10 and Band 11 with 100 m resolution) that are useful in providing more accurate surface temperatures. The TIRS uses quantum well infrared photo detectors (QWIPs) to detect the long wavelengths of light emitted from the earth surface whose intensity depends on surface temperature. As surface temperatures are directly related to surface physical properties, the normalised difference vegetation index (NDVI) analysis is an ideal approach for estimating LST in the Himalayan landscape.

LITERATURE REVIEW

Land surface integrated with temperature is identified as a significant variable to study the microclimate and radiation transfer within the atmosphere. As LST is the skin temperature of surface soil-water content and vegetation cover (Rajendran & Mani, 2015), it is very important to consider in climate change studies. Srivanit, Hokao and Phonekeo (2012) used Landsat thematic mapper (TM) imagery from 1994, 2000 and 2009 to identify the thermal characteristics of the rapidly urbanised Bangkok metropolitan area by investigating the correlation between the LST and NDVI. In their study, LST and NDVI were found to be closely correlated in several LULC categories, particularly in vegetated areas. Decrease of biomass primarily triggered the impacts of urban expansion on LST. Similarly, Li et al. (2004), and Giannini, Belfiore, Parente and Santamaria (2015) suggested the use of high resolution sequence satellite data for the analysis of LST over a watershed area in Lowa, Bangkok. LST images were extracted from the Landsat 5 TM and Landsat 7 enhanced thematic mapper (ETM) thermal bands. The NDVI was derived from visible and near-infrared bands (Band 4 and Band 5) of Landsat to estimate emissivity for Landsat thermal bands. The estimated LST values were compared with ground truth data measurements. The actual difference between the LST estimated from Landsat 5 and 7 and the ground truth actual measurements was 0.98^0C and 1.47^0C, respectively. Thus, the spatial differences of LST were identified using these satellite images. In another study, Dagliyar, Avdan, Yildiz and Nefeslioglu (2015) used Landsat TIRS and OLI data to estimate

the LST variation over Thiruvananthapuram, capital city of the state of Kerala, India. In their study, potentials of semi-automatic classification plug-in integrated with the open-source geographical information system package quantum (QGIS) were utilised for image acquisition, pre-processing, land cover classification and derivation of LST from land surface. LST of the urban Erzurum was estimated using Landsat 8 OLI and TIRS having 12-bit radiometric resolution and evaluated on the basis of surface emissivity and brightness values. In order to validate the LST derived from Landsat 8 bands, kinetic surface temperature measurements acquired from the general directorate of the state meteorological service in and around the study area were incorporated. The results showed that maximum temperature difference was around 6.45^0C, while the minimum difference was around 1.86^0C. In this study also, temperature difference was observed between LULC and LST derived from Landsat 8. Furthermore, the studies by Lv and Zhou (2011), Suresh, Ajay and Mani (2016) and Ning, Gao, Meng, Xu and Gao (2017) identified the relationship between LULC change and the LST using Landsat 5 TM and ETM images. In order to derive land use classification, the object-based method was used. The retrieval of LST was followed by the use of the MW algorithm. The derived results indicated that LST was highly influenced by the LULC. Similarly, LST was also found to be positively correlated with impervious surface and vice versa. These studies strongly recommended that such a study should be applied in regions with a trend of rapid urbanisation. Recently, Landsat 5 TM and Landsat ETM for the years 1990, 2001 and 2010 were used to estimate LST using the Landsat 7 user handbook method in Devikulam *taluk*. The results showed that the mean temperature was increasing steadily (Suresh et al., 2016).

The literature concluded that land surface temperature (LST) is an important factor and it needs to be considered in climate change studies. Further, the estimation of the accuracy of the LST using advanced satellite data (i.e. Landsat 8) is very important as it affects accuracy directly. However, so far very few studies have used the advanced TIRS data. In this research, we used TIRS data from the Landsat 8 satellite to study the Beas river basin area to estimate LST.

The Study Area

The Beas river basin is located in the state of Himachal Pradesh, India (Figure 1). The study area is between 31°N and 32°N (latitude) and 77°E and 78°E (longitude) and covers an area of 5383 km², with elevation ranging from 857 m to 6582 m. The catchment area mostly contains impulsive slopes and the rocks are commonly bare. In winter, most of the river is snow and later in summer, the Beas river basin is mostly fed by snowmelt. The Beas river basin gets heavy rainfall during the monsoon season, with rain normally falling from July to late September. It collects the moisture comportment of the winds from both the Arabian Sea and the Bay of Bengal. The upper portion of the basin receives snowfall during winter.
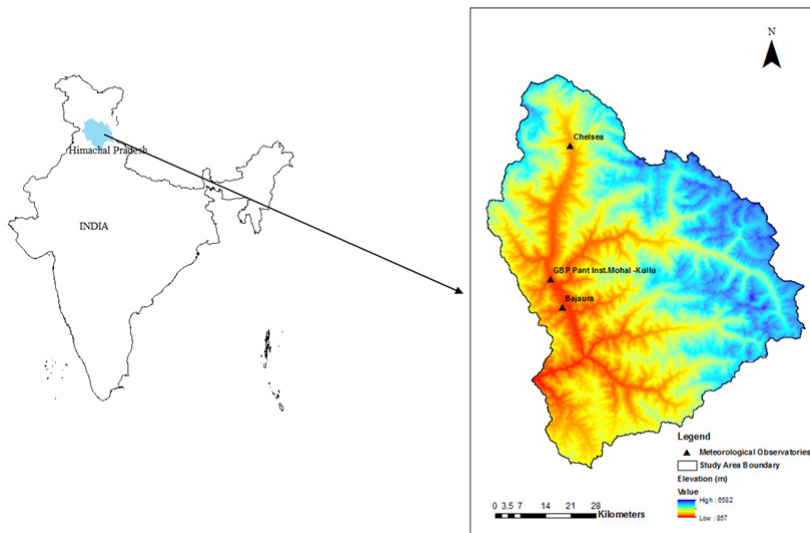
Figure 1. The Beas river basin up to Pandoh dam showing the available climate stations

METHODS AND DATA USED

In this study, Landsat 8 OLI and the TIRS image of 24 April, 2015 (Path/Row - 147/38) pertaining to the study area was used to calculate NDVI and LST. A digital elevation model (DEM) of the study area at 30 m spatial resolution, extracted from the advanced space borne thermal emission and reflection radiometer (ASTER) giving global DEM for the Beas river basin was downloaded from the USGS earth explorer data centre website (https://earthexplorer.usgs.gov/). Further, air temperature data were downloaded from the National Remote Sensing Centre's (NRSC) Meteorological and Oceanographic Satellite Data Archival Centre (MOSDAC) website (http://www.mosdac.gov.in/) for three ground stations (Chelsea, GB Pant Inst. Mohal-Kullu and Bajaura) located within the Beas river basin (Figure 1). Later, the MW algorithm method was employed to calculate the LST in the study area. Vegetation proportion calculation, emissivity calculation, LST calculation etc. were executed using the ArcGIS 10.3 software platform. A detailed description of the methodology is outlined in Figure 2.
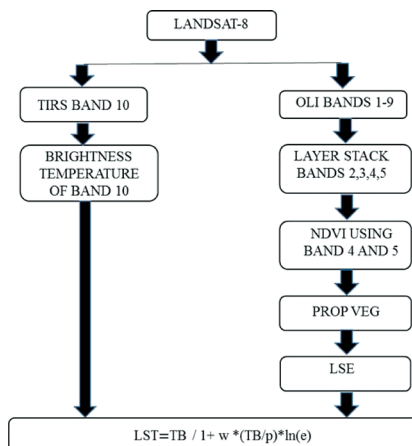
Figure 2. Flow chart of the Mono-Window algorithm

Image Acquisition and Pre-Processing

The images were already rectified to WGS-1984-UTM-Zone_43N. In the next step, the conversion of digital number (DN) to the physical measure of top of atmospheric (TOA) reflectance given in the metadata file and the thermal band at satellite brightness temperature (TB) was done. Later, the file with extension of ".MTL" was provided in the Landsat 8 image set, which contains the thermal constants needed to convert TIRS data using the satellite TB. Lastly, TIRS band data were used to convert spectral radiance to TB by processing thermal constants provided in the metadata file (Table 1 to 3).

Table 1
Landsat 8 metadata of the Beas river basin

Sensor	No. of Bands	Resolution (m)	Path/Row	Date of Acquisition
OLI	9	30	147/38	24 April, 2015
TIRS	2	100		

Table 2
K1 and K2 values

Thermal Constant	Band 10
K1	1321.08
K2	777.89

Table 3
Rescaling factor

Rescaling Factor	Band 10
M_L	0.0003342
A_L	0.1

RESULTS AND DISCUSSION

LST Calculation Using Mono-Window Algorithm for Landsat 8 Data

The primary objective of the study was to estimate land surface temperature (LST) as well as vegetation index using Landsat 8 data. The LST and NDVI obtained were for the year 2015. The Landsat 8 TIRS sensors acquired temperature data and stored this information as a digital number (DN) ranging between 0 and 255. Generally, the LST can be estimated by applying structured mathematical algorithms such as mono-window (MW), split-window (SW), single channel (SC) and multi angle (MA) algorithm. These algorithms use the TB of the TIRS band, mean and difference in land surface emissivity for estimating the LST of an area. In the present study, the MW algorithm has been used to estimate the LST. The detailed step-by-step procedure for LST calculation is given below.

$$LST = TB/1 + W * (TB/p) * \ln(e) \tag{1}$$

where,

LST = Land surface temperature (K)

TB = At satellite temperature (K)

W = Wavelength of emitted radiance $(11.5 \mu m)$

$p = h * c/s \, (1.438 * 10^{-2} \, mK)$

h = Planck's constant $(6.626 * 10^{-34} \, Js)$

s = Boltzmann constant $(1.38 * 10^{-23} \, J/K)$

c = Velocity of light $(2.998 * 10^8 \, m/s)$

p = 14380

Step 1. We converted the DN to radiance using the given formula, where $L\lambda$ is the spectral radiance at the sensor aperture $(watts /(m^2 * ster * \mu m))$

$$L\lambda = M_L Q_{cal} + A_L \tag{2}$$

where,

$L\lambda$ = TOA spectral radiance $(watts /(m^2 * ster * \mu m))$

M_L = Band specific multiplicative rescaling factor from the metadata (Radiance_mult_Band_X, where X is the band number 10)

A_L = Band specific additive rescaling factor from the metadata (Radiance_add_Band_X, where X is the band number 10)

Q_{cal} = Quantised and calibrated standard product pixel value (DN)

Here, $L\lambda$ for the study area was $L\lambda$ = 0.0003342 * Band 10 + 0.1, resulting in the value of radiance Band 10.

Step 2. We converted the band radiance (which was derived from Equation 2) to TB using the thermal constant given in metadata file (Table 2). The conversion formula is given below:

$$T = K2/l_n (K1/L\lambda + 1) - 273.15 \tag{3}$$

where,

T = At satellite brightness temperature in Kelvin (K)

$L\lambda$ = TOA spectral radiance $(watts /(m^2 * ster * \mu m))$

$K1$ = Band_specific thermal conversion from the metadata

($K1$ – Constant_Band_X, where X is the band number, 10)

$K2$ = Band_specific thermal conversion from the metadata

($K2$ – Constant_Band_X, where X is the Band number, 10)

273.15 = Convert Kelvin to °Celsius

According to Equation 3, T can be calculated for the study area as $T = 1321.08 / l_n (774.89 / Band10 radiance + 1) - 273.15$. From the formula, we can get Band 10 sattemp as an output. It shows the TB in °Celsius.

Step 3. We found the temperature of Band 10 sattemp using the cell statistics tool and we got the Band 10 sattemp as an output. Here, the minimum temperature was -14.6673°C, while the maximum temperature was 33.0782°C.

Step 4. We estimated land surface emissivity (LSE) using the given equation.

$$e = 0.004 P_V + 0.986 \qquad (4)$$

e = Emissivity
P_V = Proportion of vegetation that is calculated using the NDVI value
$NDVI$ = Normalised difference vegetation index
NDVI can be calculated in ArcGIS by applying the given formula

$$Float(Band5 - Band4) / Float(Band5 + Band4) \qquad (5)$$

where,
$Band5$ = Near infrared (NIR - 0.85–0.88 μm) Band, $Band4$ = Red Band (0.64–0.67 μm) wavelengths from the $NDVI$ result, The $NDVI_{min}$ value was $- 0.453896$ and the $NDVI_{max}$ value was 0.81123 (Figure 3). We substituted these values in Equation 6 and derived the proportion of vegetation (P_V).

$$P_V = (NDVI - NDVI_{min} / NDVI_{max} - NDVI_{min})^2 \qquad (6)$$

In ArcGIS, this can be done by applying this formula:
$Square(NDVI + 0.453896 / 0.81123 + 0.453896)$.

From this we derived the $PROPVEG$ as an output of the proportion of vegetation. Then we calculated LSE by applying the P_V value in Equation 4: $0.004 * PROPVEG + 0.986$. From this we got the LSE of the study area, which is shown in Figure 4.

Step 5. We estimated LST using Equation 1. The output such as Band 10 sattemp was substituted for BT that was in °C, and the LSE value was replaced by 'e', that is, emissivity. Equation 1 is: $LST = BT / 1 + W * (BT / p) * l_n(e)$. In ArcGIS it can be done using this formula:

$$LST = "Band10 sattemp" / 1 + "Band10" * ("Band10 sattemp" / 14380) * l_n("LSE")$$

Finally, we got the actual LST of Band 10 of the study area (Figure 5). Table 4 gives the statistics of the study area's LST. Sun elevation of 62.54660086 represents the Landsat 8 image acquisition time that was probably in the morning.

Figure 5 shows the spatial distribution of the estimated LST for the year 2015. The maximum and minimum temperature and statistics of the LST are shown in Table 4, -14.667°C and 33.078°C, respectively, with a mean average of 9.546°C. The archive meteorological

observatory data obtained from the MOSDAC was compared with the estimated LST. The satellite derived estimates for the LST and MOSDAC climate station observed temperature were comparable and validated the findings of the estimated temperature. The MOSDAC observations and satellite-derived LST were nearly in agreement with respect to temperature values in the respective year. The LST output portrayed that it varied from -14.667ºC to 33.078ºC. The highest LST values were traced in the southern plains of the study area. However, the lowest LST values were seen in the highly elevated regions that experienced snow fall.

The NDVI map (Figure 3) represents the Beas river basin on 24 April, 2015 derived from Band 5 (NIR) and Band 4 (RED) of the OLI sensor using the ArcGIS 10.3 raster calculator. The range of NDVI varied from -0.453 to 0.8112 for MW algorithm (zero for negative NDVI values using conditions). The southwestern part of the Beas river basin had the highest NDVI value, whereas the area under snow fall had a negative value (Figure 3). The value of the NDVI was later used to calculate the proportion of vegetation using Equation 6 and the LSE was calculated using Equation 4. Later, we implemented equations in the ArcGIS 10.3 raster calculator to calculate the LSE map shown in Figure 4 from the MW algorithm. The LSE of the Beas river basin ranged between 0.987 and 0.999. Highly elevated regions in the basin had more snow cover; hence, LSE was low in these regions. High LSE was found in the western and southwestern parts of the basin, whereas low LSE was noticed in the northern and eastern parts of the study area. Further, the TIRS Band 10 was used to estimate the TB in Celsius using the algorithm described in Equation 3. Similarly, equation 1 was used to estimate the LST using the raster calculator. Figure 5 represents the final LST image of the Beas river basin seen on 24 April, 2015.

LST Validation

The two major LST validation models were obtained through ground measurements or near-surface air temperature (Srivastava, Majumdar, & Bhattacharya, 2009; Li et al., 2013). The LST results comparing with the ground measurements results may have an error of up to 5°C; in the case of Srivastava et al. (2009), the accuracy of the results in some areas showed a difference of ±2°C with actual ground temperature measurements. According to Li et al. (2013), another method using the mean near-surface air temperature to verify the retrieved LST results showed that the LST retrieving error was about 0.7°C.

As such, no ground measurements of LST were available on the field; the LST obtained from Landsat 8 were compared with the air temperature observed at three stations. The comparison was made for air temperature, which is different and can sometimes result in big differences since the resolution of Landsat 8 for the used bands is 100 m for the thermal band and 30 m for the red and NIR bands. The LST was calculated and taken according to the pixel used by the respective meteorological stations. The relationship between air temperature and LST seen from Landsat 8 data, depicted in Figure 6, is in close agreement. In general, air temperature was higher than LST.

Sometimes, the differences can be great depending on weather conditions and other factors (Gallo, Hale, Tarpley, & Yu, 2011). It should also be taken into consideration that there is a 1.1-to-2-m difference between the LST and the air temperature, which means that differences in the temperatures are normal and expected.
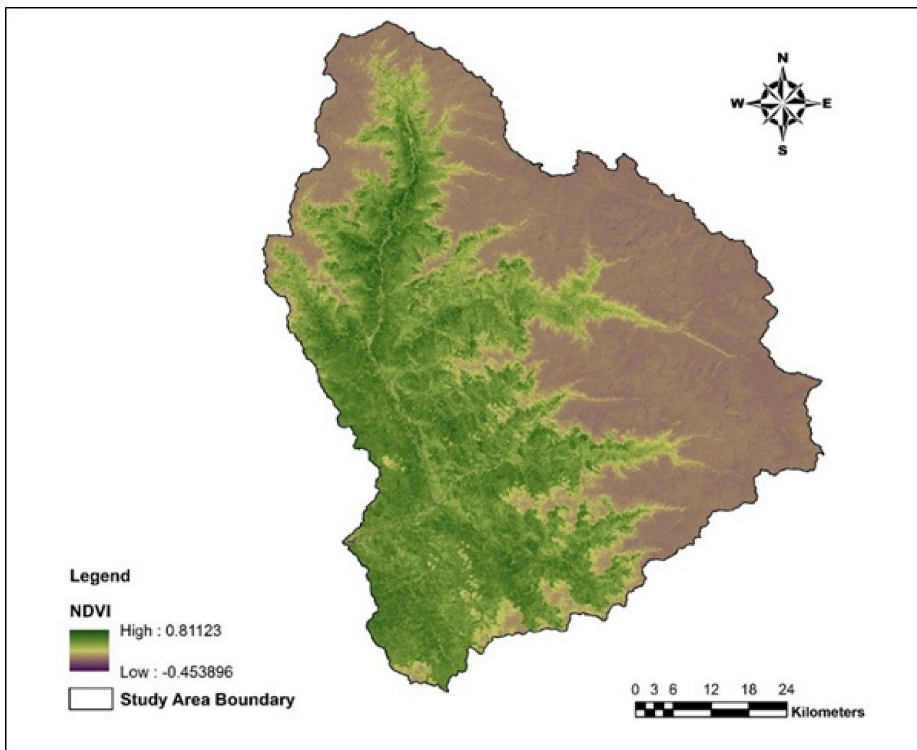
Figure 3. NDVI map of 24 April, 2015 from Mono-Window algorithm
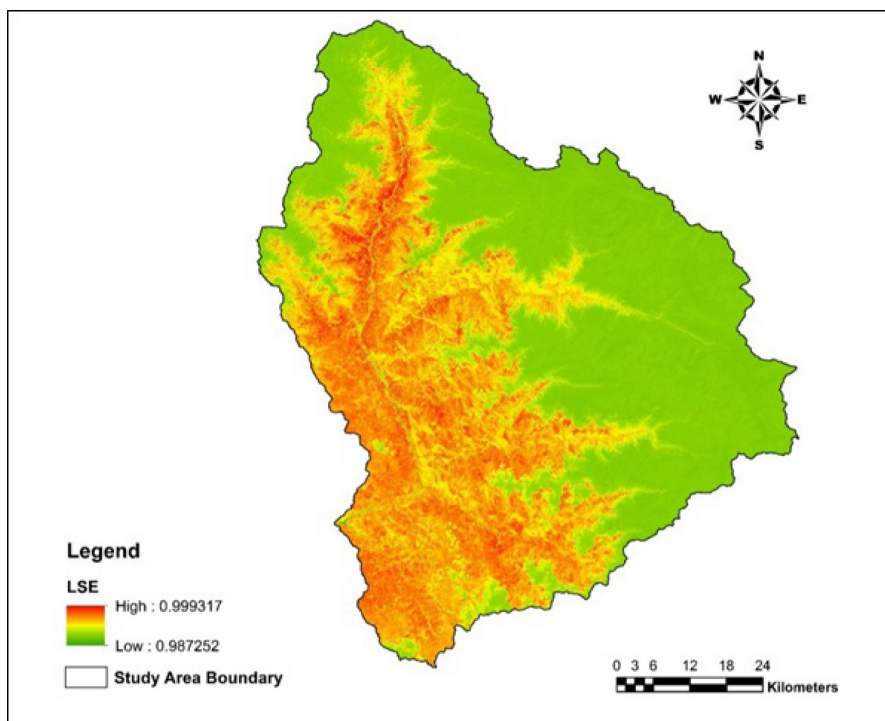
Figure 4. LSE map of 24 April, 2015 from Mono-Window algorithm
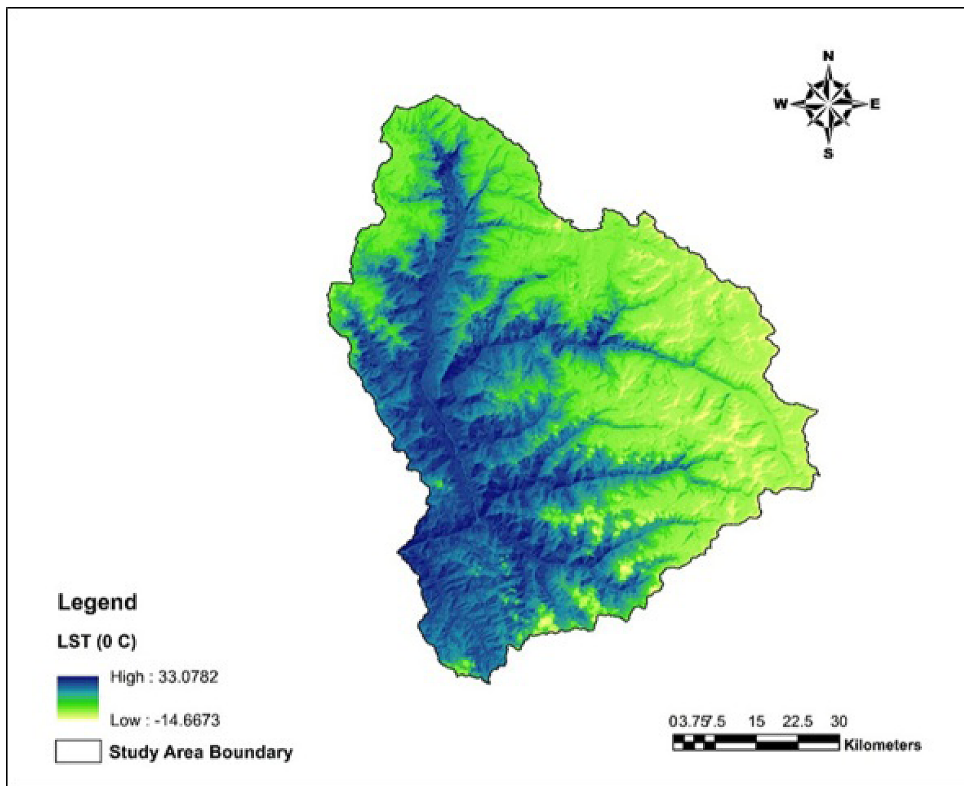
Figure 5. LST map of 24 April, 2015 from Mono-Window algorithm
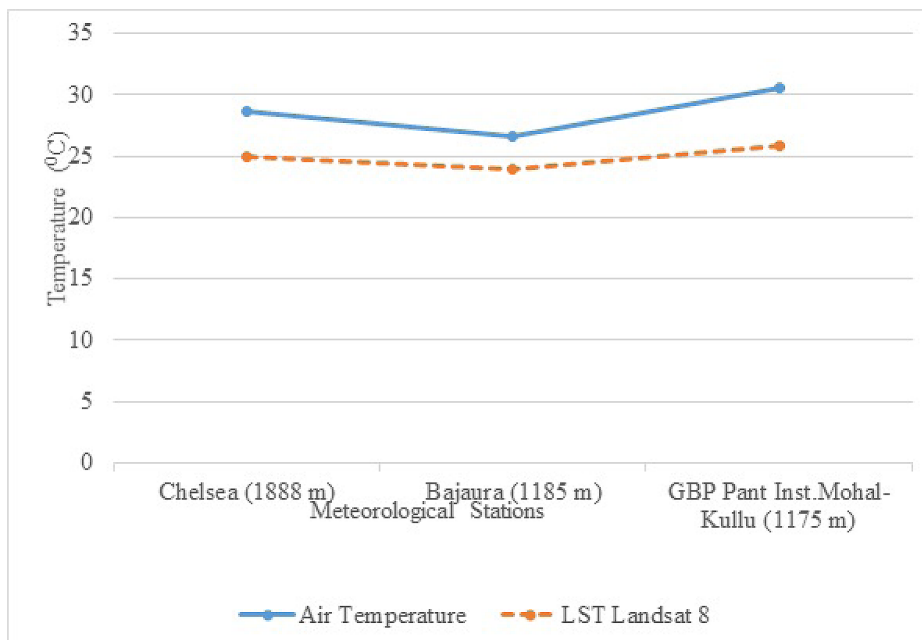
Figure 6. Comparison between air temperature data and LST values from Landsat 8 on 24 April, 2015 at three different stations

Table 4
Statistics of landsat 8 lst of Beas river basin

Statistics of Landsat 8 LST	
Sensor	TIRS
Minimum temperature	-14.667
Maximum temperature	33.078
Mean temperature	9.546
Standard deviation	10.073

CONCLUSION

In this paper, the potential of remote sensing to study temperature variation in the Beas river basin by estimating LST distribution with the help of the Landsat 8 OLI and TIRS sensor was studied. The MW algorithm method was applied to estimate the LST from the TIRS data. The analysis of the results suggested that the heat energy radiated by the earth's surface determined factors such as different types of land use, vegetation cover, soil and snow in the study area, revealing the variation in surface temperature for different surface patterns. It was also evident from the results that surface temperature variation controlled surface heat and water exchange with the atmosphere, resulting in climate change in the region. Some climatic phenomena play a minor role in temperature variation, but some play a major role. Activities such as land conversion due to rapid tourism development, ever increasing automobile carbon emission, firewood combustion from kitchens, periodical removal of firewood, for example eucalyptus, and forests replaced by settlement and restaurants etc. result in higher temperature variation. Remote sensing technology data such as Landsat 8 TIRS provides an efficient way to estimate LST. The results documented in this study can help in estimating weather phenomena such as microclimate, heat pockets and maximum temperature in vulnerable regions in the study area and also help in deciding what necessary scientific actions can be taken like reforestation, frequent checking of vehicles for pollution and reduced plastic incineration to curb temperature increase.

ACKNOWLEDGEMENT

REFERENCES

Dagliyar, A., Avdan, U., Yildiz, N. D., & Nefeslioglu, H. A. (2015). Determination of land surface temperature using Landsat 8 TIRS: A case study in Erzurum, Turkey. *Geophysical Research Abstracts, 17,* EGU2015-11007.

Gallo, K., Hale, R., Tarpley, D., & Yu, Y. (2011). Evaluation of the relationship between air and land surface temperature under clear- and cloudy-sky conditions. *Journal of Applied Meteorology and Climatology, 50*(3)*,* 767–775.

Giannini, M. B., Belfiore, O. R., Parente, C., & Santamaria, R. (2015). Land surface temperature from Landsat 5 TM images: Comparison of different methods using airborne thermal data. *Journal of Engineering Science and Technology Review, 8*(3)*,* 83–90.

Li, F., Jackson, T. J., Kustas, W. P., Schmugge, T., French, A., Cosh, A. M., & Bindlish, R. (2004). Deriving land surface temperature from Landsat 5 and 7 during SMEX02/SMACEX. *Remote Sensing of Environment, 92*(4)*,* 521–534.

Li, Z. L., Tang, B. H., Wu, H., Ren, H., Yan, G., Wan, Z., … & Sobrina, A. J. (2013). Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment, 131,* 14–37.

Lv, Z., & Zhou, Q. (2011). Utility of Landsat image in the study of land cover and land surface temperature change. *Procedia Environment Science, 10,* 1287–1292.

Ning, J., Gao, Z., Meng, R., Xu, F., & Gao, M. (2017). Analysis of relationships between land surface temperature and land use changes in the Yellow River delta. *Frontiers of Earth Science,* 1–13.

Rajendran, P., & Mani, K. (2015). Estimation of spatial variability of land surface temperature using Landsat 8 imagery. *International Journal of Engineering and Science, 11*(4), 19–23.

Rajeshwari, A., & Mani, N. D. (2014). Estimation of land surface temperature of dindigul district using Landsat 8 data. *International Journal of Research in Engineering and Technology, 3*(5), 122–126.

Srivanit, M., Hokao, K., & Phonekeo, V. (2012). Assessing the impact of urbanization on urban thermal environment: A case study of Bangkok metropolitan. *International Journal of Applied Science and Technology, 2*(7)*,* 243–254.

Srivastava, P., Majumdar, T. J., & Bhattacharya, A. K. (2009). Surface temperature estimation in Singhbhum shear zone of India using Landsat-7 ETM+ thermal infrared data. *Advances in Space Research, 43*(10), 1563–1574.

Suresh, S., Ajay, S. V., & Mani, K. (2016). Mountain landscape of Devikulam Taluk using Landsat 8 Data. *International Journal of Research in Engineering and Technology, 5,* 92–96.

Application of Almost Increasing Sequence for Absolute Riesz $|\bar{N}, p_n^\alpha; \delta; \gamma|_k$ Summable Factor

Sonker, Smita and Munjal, Alka*

Department of Mathematics, National Institute of Technology Kurukshetra, Kurukshetra-136119, Haryana, India

ABSTRACT

A moderated theorem was established in this study to derive an infinite series as an absolute Riesz $|\bar{N}, p_n^\alpha; \delta; \gamma|_k$ summable by applying an almost increasing sequence. Suitable conditions were applied for obtaining some well-known applications of the presented theorem. It was observed from the investigation that absolute summability could be used for Bounded Input Bounded Output stability and to bind the output of the signals as absolute summable, it was a necessary and sufficient condition for BIBO stability.

Keywords: Almost increasing sequence, absolute summability, infinite series, $|\bar{N}, p_n^\alpha; \delta; \gamma|_k$ summability

INTRODUCTION

Summability is a field that studies the non-convergent series/integrals and assigns a value (number) to it. In mathematical analysis, the summability method is an alternative formulation of convergence of a series that is divergent in the conventional sense. Throughout the 19th century, many mathematicians studied various sets of divergent series and defined numerous summability methods such as Abel summability, Cesàro summability, Euler summability, Hausdorff summability, Nörlund summability, Riesz summability etc. Riesz summability was introduced by Marcel Riesz in 1911 as an improvement to the Cesàro mean. It is a generalisation of Cesàro (C, 1) summability, which can be obtained using $p_n = 1$. The present paper is devoted to the study of the Riesz summability method and a theorem that has been established for the absolute Riesz summable factor of an infinite series.

With the help of a minimal set of sufficient conditions for a soft computing model, error can be minimised by using summability methods so that the output data are filtered according to the user's interest. The output data are given by u_n, which is a sequence-to-sequence transformation. For such computing,

$\{s_n\}$ should be the sequence of partial sums of an infinite series $\sum_{n=0}^{\infty} a_n$ and the nth sequence-to-sequence transformation (mean) of $\{s_n\}$ be given as u_n i.e.

$$u_n = \sum_{k=0}^{\infty} u_{nk} s_k. \tag{1}$$

However, before discussing $|\overline{N}, p_n^{\alpha}; \delta; \gamma|_k$ summability, it would be meaningful to consider some well-known basic summabilities that can be helpful for understanding the $|\overline{N}, p_n^{\alpha}; \delta; \gamma|_k$ summability.

Definition 1

Let $\sum_{n=0}^{\infty} a_n$ be an infinite series with a sequence of partial sums $\{s_n\}$ and said to be absolutely summable, if

$$\lim_{n \to \infty} u_n = s, \tag{2}$$

and

$$\sum_{n=1}^{\infty} |u_n - u_{n-1}| < \infty, \tag{3}$$

where, u_n represents the nth sequence-to-sequence transformation (mean) of $\{s_n\}$.

Definition 2 (Flett, 1957)

If t_n represents the $n^{th} (C,1)$ means of the sequence (na_n), then the series $\sum_{n=0}^{\infty} a_n$ is said to be $|C,1|_k$ summable for $k \geq 1$, if

$$\sum_{n=1}^{\infty} \frac{1}{n} |t_n|^k < \infty. \tag{4}$$

Definition 3

Let (p_n) be a sequence with $p_0 > 0$ and $p_n \geq 0$ for $n > 0$

$$P_n = \sum_{v=0}^{n} p_v \to \infty. \tag{5}$$

For $\alpha > -1$, define

$$\varepsilon_0^{\alpha} = 1, \varepsilon_n^{\alpha} = \frac{(\alpha+1)(\alpha+2)...(\alpha+n)}{n!}, (n=1,2,3,...), \tag{6}$$

$$p_n^{\alpha} = \sum_{v=0}^{n} \varepsilon_{n-v}^{\alpha-1} p_v, \tag{7}$$

$$P_n^\alpha = \sum_{v=0}^{n} p_v^\alpha \to \infty,\ n \to \infty, \tag{8}$$

and $P_{-n}^\alpha = p_{-n}^\alpha = 0,\ n \geq 1,$

then, the sequence-to-sequence transformation t_n defines the $(\overline{N}, p_n^\alpha)$ mean of series $\sum a_n$ and is given by

$$t_n = \frac{1}{P_n^\alpha} \sum_{k=0}^{n} p_k^\alpha s_k,\ P_n^\alpha \neq 0,\ n \in N \tag{9}$$

and $lim_{n \to \infty} t_n = s,$ and the series is said to be $(\overline{N}, p_n^\alpha)$ summable, generated by the sequence of coefficients $\{p_n^\alpha\}$.

Further, if sequence $\{t_n\}$ is of bounded variation with index $k \geq 1$ i.e.

$$\sum_{n=1}^{\infty} \left(\frac{P_n^\alpha}{p_n^\alpha}\right)^{k-1} |\Delta t_{n-1}|^k < \infty, \tag{10}$$

then the series $\sum a_n$ is said to be absolutely $(R, p_n^\alpha)_k$-summable with index k or $|\overline{N}, p_n^\alpha|_k$-summable to s.

Definition 4

The series is said to be $|\overline{N}, p_n^\alpha; \delta|_k$ summable, if

$$\sum_{n=1}^{\infty} \left(\frac{P_n^\alpha}{p_n^\alpha}\right)^{\delta k + k - 1} |\Delta t_{n-1}|^k < \infty, \tag{11}$$

and summable $|\overline{N}, p_n^\alpha; \delta; \gamma|_k$, if

$$\sum_{n=1}^{\infty} \left(\frac{P_n^\alpha}{p_n^\alpha}\right)^{\gamma(\delta k + k - 1)} |\Delta t_{n-1}|^k < \infty, \tag{12}$$

where, $k \geq 1, \delta \geq 0$ and γ is a real number and

$$\Delta t_n = -\frac{p_n^\alpha}{P_n^\alpha P_{n-1}^\alpha} \sum_{v=1}^{n} P_{v-1}^\alpha a_v,\ n \geq 1. \tag{13}$$

Bor and Seyhan (1999) determined the set of sufficient conditions for an infinite series to be absolutely Riesz $|\overline{N}, p_n^\alpha; \delta|_k$ summable using an almost increasing sequence. Bor and Özarslan (2002) redesigned the problem of Mazhar (1997) under weaker conditions using a quasi β-power increasing sequence, and Bor (2014) generalised the theorem, dealing with a general class of power increasing sequences and absolute Riesz summability factors of an infinite series.

Özgen (2013) used the concept of positive non-decreasing sequence for an infinite series to be absolutely $|A|_k (k \geq 1)$ summable. Sonker and Munjal (2016a) determined the generalised absolute Cesáro summable factor and obtained the minimal sufficient conditions for the

boundedness of an infinite series (2016b). Sonker and Munjal (2017a, 2017b, 2017c, 2017d), and Sonker, Krasniqi and Munjal (2017) also worked out various interesting theorems on the absolute summable factor. Bor (2015) applied absolute Cesáro and Nörlund summability and established two theorems using more general conditions for the infinite series.

KNOWN RESULTS

Using $|\overline{N}, p_n; \delta|_k$ summability, Bor and Seyhan (1999) proved the following theorem with a minimal set of sufficient conditions for an infinite series to be absolutely Riesz summable.

Theorem 2.1 (Bor and Seyhan, 1999)

Let p_n be a sequence of positive numbers such that

$$P_n = O(np_n) \ \text{as } n \to \infty. \tag{14}$$

Let (X_n) be an almost increasing sequence and assuming that there exists sequences (β_n) and (λ_n) such that

$$|\Delta\lambda_n| \le \beta_n, \tag{15}$$

$$\beta_n \to 0 \ \text{as } n \to \infty, \tag{16}$$

$$\sum_{n=1}^{\infty} n |\Delta\beta_n| X_n \le \infty, \tag{17}$$

$$|\lambda_n| X_n = O(1) \ \text{as } n \to \infty, \tag{18}$$

$$\sum_{n=v+1}^{\infty} \left(\frac{P_n}{p_n}\right)^{\delta k-1} \frac{1}{P_{n-1}} = O\left\{\left(\frac{P_v}{p_v}\right)^{\delta k} \frac{1}{P_v}\right\}, \tag{19}$$

$$\sum_{n=1}^{m} \left(\frac{P_n}{p_n}\right)^{\delta k-1} |t_n|^k = O(X_m) \ \text{as } m \to \infty, \tag{20}$$

where, $t_n = \dfrac{1}{n+1} \displaystyle\sum_{v=1}^{n} v a_v,$ \hfill (21)

then, the series $\sum a_n \lambda_n$ is $|\overline{N}, p_n; \delta|_k$ summable for $k \ge 1$ and $0 \le \delta \le 1/k$.

MAIN RESULTS

A sequence is of bounded variation i.e. $(\lambda_n) \in BV$, if

$$\sum_{n=1}^{\infty} |\Delta\lambda_n| = |\lambda_n - \lambda_{n+1}| < \infty.$$

A positive sequence (g_n) is an almost increasing sequence if $\exists$ is a positive increasing sequence (h_n) and there are two positive constants, M and N s.t.

$$Mh_n \leq g_n \leq Nh_n.$$

Theorem 3.1

Let (X_n) be an almost increasing sequence and the sequences (β_n) and (λ_n) be such that conditions $(15)-(18)$ of theorem 2.1 are satisfied. If the following conditions also satisfy

$$\sum_{n=v+1}^{\infty} \frac{1}{P_{n-1}^{\alpha}} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)-k} = O\left\{\frac{1}{P_v^{\alpha}} \left(\frac{P_v^{\alpha}}{p_v^{\alpha}}\right)^{1-k+\gamma(\delta k+k-1)}\right\}, \tag{22}$$

$$\sum_{n=1}^{m} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)-k} |t_n|^k = O(X_m), \tag{23}$$

$$\sum_{n=1}^{m} \frac{|\lambda_n|}{n} = O(1), \tag{24}$$

and

$$\sum_{n=1}^{m} \frac{1}{n} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{1-k+\gamma(\delta k+k-1)} |t_n|^k = O(X_m) \text{ as } m \to \infty, \tag{25}$$

then, the series $\sum a_n \lambda_n$ is $|\overline{N}, p_n^{\alpha}; \delta; \gamma|_k$ summable for $k \geq 1$, $0 \leq \delta \leq 1/k$ and γ is a real number.

Proof of the Theorem 3.1

Let T_n denote the $(\overline{N}, p_n^{\alpha})$ mean of the series $\sum a_n \lambda_n$. Then, by definition and by changing the order of summation, we have

$$T_n = \frac{1}{P_n^{\alpha}} \sum_{v=0}^{n} p_v^{\alpha} \sum_{i=0}^{v} a_i \lambda_i = \frac{1}{P_n^{\alpha}} \sum_{v=0}^{n} (P_n^{\alpha} - P_{v-1}^{\alpha}) a_v \lambda_v. \tag{26}$$

Then, for $n \geq 1$, we have

$$\Delta T_n = T_n - T_{n-1} = \frac{p_n^{\alpha}}{P_n^{\alpha} P_{n-1}^{\alpha}} \sum_{v=1}^{n} P_{v-1} a_v \lambda_v$$

$$= \frac{p_n^{\alpha}}{P_n^{\alpha} P_{n-1}^{\alpha}} \sum_{v=1}^{n} \frac{P_{v-1}\lambda_v}{v} v a_v.$$

By Abel's transformation, we have

$$\Delta T_n = \frac{n+1}{nP_n^{\alpha}} p_n^{\alpha} t_n \lambda_n - \frac{p_n^{\alpha}}{P_n^{\alpha} P_{n-1}^{\alpha}} \sum_{v=1}^{n-1} p_v^{\alpha} t_v \lambda_v \frac{v+1}{v}$$

$$+ \frac{p_n^{\alpha}}{P_n^{\alpha} P_{n-1}^{\alpha}} \sum_{v=1}^{n-1} P_v^{\alpha} t_v \Delta \lambda_v \frac{v+1}{v} + \frac{p_n^{\alpha}}{P_n^{\alpha} P_{n-1}^{\alpha}} \sum_{v=1}^{n-1} P_v^{\alpha} t_v \lambda_{v+1} \frac{1}{v}$$

$$= T_{n,1} + T_{n,2} + T_{n,3} + T_{n,4}. \tag{27}$$

In order to prove that the theorem works, it is sufficient to show that

$$\sum_{n=1}^{\infty} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}} \right)^{\gamma(\delta k + k - 1)} | \Delta T_n |^k < \infty. \tag{28}$$

Using Minkowski's inequality

$$| T_{n,1} + T_{n,2} + T_{n,3} + T_{n,4} |^k \le 4^k (| T_{n,1} |^k + | T_{n,2} |^k + | T_{n,3} |^k + | T_{n,4} |^k),$$

Equation 28 is reduced to

$$\sum_{n=1}^{\infty} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}} \right)^{\gamma(\delta k + k - 1)} | T_{n,r} |^k < \infty \ for \ r = 1, 2, 3, 4. \tag{29}$$

Now the L.H.S of Equation 29 is given as

$$\sum_{n=1}^{m} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}} \right)^{\gamma(\delta k + k - 1)} | T_{n,1} |^k$$

$$= \sum_{n=1}^{m} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}} \right)^{\gamma(\delta k + k - 1)} \left| \frac{n+1}{nP_n^{\alpha}} p_n^{\alpha} t_n \lambda_n \right|^k$$

$$= \sum_{n=1}^{m} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}} \right)^{\gamma(\delta k + k - 1) - k} | t_n |^k | \lambda_n |$$

$$= O(1) | \lambda_m | \sum_{n=1}^{m} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}} \right)^{\gamma(\delta k + k - 1) - k} | t_n |^k$$

$$+ O(1)\sum_{n=1}^{m-1}\Delta \mid \lambda_n \mid \sum_{v=1}^{n}\left(\frac{P_v^{\alpha}}{p_v^{\alpha}}\right)^{\gamma(\delta k+k-1)-k} \mid t_v \mid^k$$

$$= O(1)\mid \lambda_m \mid X_m + O(1)\sum_{n=1}^{m-1}\mid \Delta\lambda_n \mid X_n$$

$$= O(1)\mid \lambda_m \mid X_m + O(1)\sum_{n=1}^{m-1}\beta_n X_n$$

$$= O(1) \text{ as } m \rightarrow \infty, \tag{30}$$

$$\sum_{n=2}^{m+1}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)} \mid T_{n,2}\mid^k$$

$$= \sum_{n=2}^{m+1}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)}\left| \frac{p_n^{\alpha}}{P_n^{\alpha}P_{n-1}^{\alpha}}\sum_{v=1}^{n-1}p_v^{\alpha}t_v\lambda_v\frac{v+1}{v}\right|^k$$

$$= O(1)\sum_{n=2}^{m+1}\frac{1}{P_{n-1}^{\alpha}}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)-k}\sum_{v=1}^{n-1}p_v^{\alpha}\mid \lambda_v \parallel t_v \mid^k \times$$

$$\times \left(\frac{1}{P_{n-1}^{\alpha}}\sum_{v=1}^{n-1}p_v^{\alpha}\right)^{k-1}$$

$$= O(1)\sum_{v=1}^{m}p_v^{\alpha}\mid \lambda_v \parallel t_v \mid^k \sum_{n=v+1}^{m+1}\frac{1}{P_{n-1}^{\alpha}}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)-k}$$

$$= \sum_{v=1}^{m}p_v^{\alpha}\mid \lambda_v \parallel t_v \mid^k \frac{1}{P_v^{\alpha}}\left(\frac{P_v^{\alpha}}{p_v^{\alpha}}\right)^{1-k+\gamma(\delta k+k-1)}$$

$$= \sum_{v=1}^{m}\mid \lambda_v \parallel t_v \mid^k \left(\frac{P_v^{\alpha}}{p_v^{\alpha}}\right)^{\gamma(\delta k+k-1)-k}$$

$$= O(1)\mid \lambda_m \mid \sum_{n=1}^{m}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)-k} \mid t_n \mid^k$$

$$+ O(1)\sum_{n=1}^{m-1}\Delta \mid \lambda_n \mid \sum_{v=1}^{n}\left(\frac{P_v^{\alpha}}{p_v^{\alpha}}\right)^{\gamma(\delta k+k-1)-k} \mid t_v \mid^{k}$$

$$= O(1) \mid \lambda_m \mid X_m + O(1)\sum_{n=1}^{m-1} \mid \Delta\lambda_n \mid X_n$$

$$= O(1) \mid \lambda_m \mid X_m + O(1)\sum_{n=1}^{m-1}\beta_n X_n$$

$$= O(1) \ \text{as} \ m \to \infty, \tag{31}$$

$$\sum_{n=2}^{m+1}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)} \mid T_{n,3} \mid^{k}$$

$$= \sum_{n=2}^{m+1}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)}\left|\frac{p_n^{\alpha}}{P_n^{\alpha}P_{n-1}^{\alpha}}\sum_{v=1}^{n-1}P_v^{\alpha}t_v\Delta\lambda_v\frac{v+1}{v}\right|^{k}$$

$$= O(1)\sum_{n=2}^{m+1}\frac{1}{P_{n-1}^{\alpha}}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)-k}\sum_{v=1}^{n-1}P_v^{\alpha}\beta_v \mid t_v \mid^{k} \times$$

$$\times\left(\frac{1}{P_{n-1}^{\alpha}}\sum_{v=1}^{n-1}P_v^{\alpha}\beta_v\right)^{k-1}$$

$$= O(1)\sum_{v=1}^{m}P_v^{\alpha}\beta_v \mid t_v \mid^{k}\sum_{n=v+1}^{m+1}\frac{1}{P_{n-1}^{\alpha}}\left(\frac{P_n^{\alpha}}{p_n^{\alpha}}\right)^{\gamma(\delta k+k-1)-k}$$

$$= O(1)\sum_{v=1}^{m}P_v^{\alpha}\beta_v \mid t_v \mid^{k}\frac{1}{P_v^{\alpha}}\left(\frac{P_v^{\alpha}}{p_v^{\alpha}}\right)^{1-k+\gamma(\delta k+k-1)}$$

$$= O(1)\sum_{v=1}^{m}v\beta_v\frac{1}{v}\left(\frac{P_v^{\alpha}}{p_v^{\alpha}}\right)^{1-k+\gamma(\delta k+k-1)} \mid t_v \mid^{k}$$

$$= m\beta_m\sum_{v=1}^{m}\frac{1}{v}\left(\frac{P_v^{\alpha}}{p_v^{\alpha}}\right)^{1-k+\gamma(\delta k+k-1)} \mid t_v \mid^{k}$$

$$+ O(1)\sum_{v=1}^{m-1}\Delta(v\beta_v)\sum_{i=1}^{v}\frac{1}{i}\left(\frac{P_i^\alpha}{p_i^\alpha}\right)^{1-k+\gamma(\delta k+k-1)}|t_i|^k$$

$$= O(1)m\beta_m X_m + O(1)\sum_{v=1}^{m-1}|\Delta(v\beta_v)|X_v$$

$$= O(1)m\beta_m X_m + O(1)\sum_{v=1}^{m-1}X_v|\Delta\beta_v|$$

$$+ O(1)\sum_{v=1}^{m-1}\beta_{v+1}X_{v+1}$$

$$= O(1) \text{ as } m \to \infty, \tag{32}$$

$$\sum_{n=1}^{m}\left(\frac{P_n^\alpha}{p_n^\alpha}\right)^{\gamma(\delta k+k-1)}|T_{n,4}|^k$$

$$= \sum_{n=1}^{m}\left(\frac{P_n^\alpha}{p_n^\alpha}\right)^{\gamma(\delta k+k-1)}\left|\frac{p_n^\alpha}{P_n^\alpha P_{n-1}^\alpha}\sum_{v=1}^{n-1}P_v^\alpha t_v \lambda_{v+1}\frac{1}{v}\right|^k$$

$$= O(1)\sum_{n=2}^{m+1}\frac{1}{P_{n-1}^\alpha}\left(\frac{P_n^\alpha}{p_n^\alpha}\right)^{\gamma(\delta k+k-1)-k}\sum_{v=1}^{n-1}P_v^\alpha\frac{|\lambda_{v+1}|}{v}|t_v|^k \times$$

$$\times\left(\frac{1}{P_{n-1}^\alpha}\sum_{v=1}^{n-1}P_v^\alpha\frac{|\lambda_{v+1}|}{v}\right)^{k-1}$$

$$= O(1)\sum_{v=1}^{m}P_v^\alpha\frac{|\lambda_{v+1}|}{v}|t_v|^k\sum_{n=v+1}^{m+1}\frac{1}{P_{n-1}^\alpha}\left(\frac{P_n^\alpha}{p_n^\alpha}\right)^{\gamma(\delta k+k-1)-k}$$

$$= O(1)\sum_{v=1}^{m}P_v^\alpha\frac{|\lambda_{v+1}|}{v}|t_v|^k\frac{1}{P_v^\alpha}\left(\frac{P_v^\alpha}{p_v^\alpha}\right)^{1-k+\gamma(\delta k+k-1)}$$

$$= O(1)\sum_{v=1}^{m}|\lambda_{v+1}|\frac{1}{v}\left(\frac{P_v^\alpha}{p_v^\alpha}\right)^{1-k+\gamma(\delta k+k-1)}|t_v|^k$$

$$= O(1) \mid \lambda_{m+1} \mid \sum_{v=1}^{m} \frac{1}{v} \left(\frac{P_v^{\alpha}}{p_v^{\alpha}} \right)^{1-k+\gamma(\delta k+k-1)} \mid t_v \mid^k$$

$$+ O(1) \sum_{v=1}^{m-1} \Delta \mid \lambda_{v+1} \mid \sum_{i=1}^{v} \frac{1}{i} \left(\frac{P_i^{\alpha}}{p_i^{\alpha}} \right)^{1-k+\gamma(\delta k+k-1)} \mid t_i \mid^k$$

$$= O(1) \mid \lambda_{m+1} \mid X_m + O(1) \sum_{v=1}^{m-1} \Delta \mid \lambda_{v+1} \mid X_v$$

$$= O(1) \mid \lambda_{m+1} \mid X_m + \sum_{v=1}^{m-1} \beta_{v+1} X_{v+1}$$

$$= O(1) \text{ as } m \to \infty, \tag{33}$$

In merging Equations 26 to 33, we have

$$\sum_{n=1}^{\infty} \left(\frac{P_n^{\alpha}}{p_n^{\alpha}} \right)^{\gamma(\delta k+k-1)} \mid T_{n,r} \mid^k < \infty \text{ for } r = 1, 2, 3, 4. \tag{34}$$

Hence, the theorem is proven.

COROLLARIES

Corollary 5.1

Let (X_n) be an almost increasing sequence and the sequences (β_n) and (λ_n) be such that the conditions of Equations 15 to 18 and 24 are satisfied. If the following conditions also satisfy

$$\sum_{n=v+1}^{\infty} \frac{p_n^{\alpha}}{P_n^{\alpha} P_{n-1}^{\alpha}} = O(\frac{1}{P_v^{\alpha}}), \tag{35}$$

$$\sum_{n=1}^{m} \frac{p_n^{\alpha}}{P_n^{\alpha}} \mid t_n \mid^k = O(X_m) \text{ as } m \to \infty, \tag{36}$$

$$\sum_{n=1}^{m} \frac{1}{n} \mid t_n \mid^k = O(X_m) \text{ as } m \to \infty, \tag{37}$$

then, the series $\sum a_n \lambda_n$ is $\mid \overline{N}, p_n^{\alpha} \mid_k$ summable for $k \geq 1$.

The proof for the theorem is established thus: $\gamma = 1$ and $\delta = 0$ are used in Theorem 3.1 to get Equations 35, 36 and 37. We omitted the details as the proof was similar to that of Theorem 3.1, but we used Equations 35, 36 and 37 instead of Equations 22, 23 and 25.

Corollary 5.2

Let (X_n) be an almost increasing sequence and the sequences (β_n) and (λ_n) be such that the conditions of Equations 15 to 18 and 24 are satisfied. If the following conditions also satisfy

$$\sum_{n=v+1}^{\infty} \frac{p_n^\alpha}{P_n^\alpha P_{n-1}^\alpha} = O\left(\frac{1}{P_v^\alpha} \right), \tag{38}$$

$$\sum_{n=1}^{m} \frac{p_n^\alpha}{P_n^\alpha} \mid t_n \mid = O(X_m) \text{ as } m \to \infty, \tag{39}$$

$$\sum_{n=1}^{m} \frac{1}{n} \mid t_n \mid = O(X_m) \text{ as } m \to \infty, \tag{40}$$

then, the series $\sum a_n \lambda_n$ is $\mid \overline{N}, p_n^\alpha \mid$ summable.

The proof for this theorem is obtained thus: by replacing $\gamma = 1$, $\delta = 0$ and $k = 1$ in Theorem 3.1, we got Equations 38, 39 and 40. We omitted the details as the proof was similar to that of Theorem 3.1, but we used Equations 38, 39 and 40 instead of Equations 22, 23 and 25.

CONCLUSION

A necessary and sufficient condition for a system to be BIBO (Bounded Input Bounded Output) stable is that the impulse response must be absolutely summable i.e.

$$\text{BIBOstable} \Leftrightarrow \sum_{n=-\infty}^{\infty} \mid h(n) \mid < \infty.$$

Thus, the summability technique was used in this study to stabilise and bind the output of signals and to predict the behaviour of input data, the initial situation and the changes in the completed process. This study also has a number of direct applications in the rectification of signals in the FIR filter (finite impulse response filter) and the IIR filter (infinite impulse response filter). The results of this investigation are generalised and can be reduced for several well-known summabilities.

ACKNOWLEDGEMENT

REFERENCES

Bor, H. (2014). A new theorem on the absolute Riesz summability factors. *Filomat*, *28*(8), 1537–1541.

Bor, H. (2015). Some new results on infinite series and Fourier series. *Positivity*, *19*(3), 467–473.

Bor, H., & Özarslan, H. S. (2002). On the quasi power increasing sequences. *Journal of Mathematical Analysis and Applications*, *276*(2), 924–929.

Bor, H., & Seyhan, H. (1999). On almost increasing sequences and its applications. *Indian Journal Pure Applied Mathematics, 30,* 1041–1046.

Flett, T. M. (1957). On an extension of absolute summability and some theorems of Littlewood and Paley. *Proceedings of the London Mathematical Society, 3*(1), 113–141.

Mazhar, S. M. (1997). A note on absolute summability factors. *Bulletin-Institute of Mathematics Academia Sinica*, *25*, 233–242.

Özgen, H. N. (2013). A note on generalized absolute summability. *Annals of the Alexandru Ioan CUZA University - Mathematics*, *59*(1), 185–190.

Sonker, S., & Munjal, A. (2016a). Absolute summability factor $\varphi - |C, 1, \delta|_k$ of infinite series. *International Journal of Mathematical Analysis, 10*(23), 1129–1136.

Sonker, S., & Munjal, A. (2016b). Sufficient conditions for triple matrices to be bounded. *Nonlinear Studies, 23*(4), 533–542.

Sonker, S., & Munjal, A. (2017a). A note on boundness conditions of absolute summability $\varphi - |A|_k$ factors. *Proceedings of the International Conference on Advancements in Science and Technology-2017, 67,* 208–210.

Sonker, S., & Munjal, A. (2017b). Absolute summability $\varphi - |C, \alpha, \beta; \delta|_k$ of infinite series. *Journal of Inequalities and Applications, 168,* 1–7.

Sonker, S., & Munjal, A. (2017c). Absolute summability factor $|N, p_n|_k$ of improper integrals. *International Journal of Engineering and Technology, 9*(3S), 457–462.

Sonker, S., & Munjal, A. (2017d). Absolute Nörlund summability $|N, p_n|_k$ of improper integrals. *NCRAME-2017, 2*(90), 413–415.

Sonker, S., Krasniqi, X. Z., & Munjal, A. (2017). A note on absolute Cesáro $\varphi - |C, 1; \delta; l|_k$ summability factor. *International Journal of Analysis and Applications, 15*(1), 108–113.

SCIENCE & TECHNOLOGY

Effect of Tool-Pin Profile on Weld Zone and Mechanical Properties in Friction Stir Welding of Aluminium Alloy

Prashant Prakash[1]*, Sanjay Kumar Jha[2] and Shree Prakash Lal[1]

[1]*Department of Production Engineering, Birla Institute of Technology, Patna Campus, Patna, Bihar, 800014, India*

[2]*Department of Production Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, 835215, India*

ABSTRACT

This article presents the experimental analysis of tool pin-profile effects on the weld zone shape, mechanical properties and microstructure of the friction stir welding (FSW) process. To determine tool pin-profile effects on weld zone shape, mechanical properties and microstructure, four different pin-profile tools were used namely, the cylindrical pin, conical pin, cylindrical-conical pin and stepped-conical pin. The results of the experiment showed that weld-zone shape thickness of the cylindrical-conical tool near the shoulder was the highest and the conical pin-profile tool was the lowest near the bottom of the tool pin with respect to all other tool pin-profiles. Fine grain structure was produced by the stepped-conical pin-profile tool in the weld zone compared with the other three pin-profile tools. Yield strength, ultimate tensile strength, ductility and joint efficiency of the stepped-conical tool were the highest among all the tool pin-profiles. Additionally, this article also explains the effect of weld-zone shape on mechanical properties, showing that the small basin shape of the weld-zone shape produces high yield strength, ultimate tensile strength, ductility and joint efficiency.

Keywords: Friction stir welding, mechanical properties, microstructure, tool pin-profiles, weld-zone shape

INTRODUCTION

Currently, the friction stir welding (FSW) process is gaining wide acceptance for joining alloy components used in the aerospace, automobile and shipbuilding industries because the friction stir welding process has several advantages compared with fusion welding. The FSW process reduces weld defects, residual stresses and distortion and improves the dimensional stability of the welded structure (Gibson et al., 2014). It is also known as an eco-friendly welding process because the friction stir

E-mail addresses:
prashant.prakash86@yahoo.co.in (Prashant Prakash)
sanjujha@hotmail.com (Sanjay Kumar Jha)
splal@bitmesra.ac.in (Shree Prakash Lal)
*Corresponding Author

welding process does not produce smoke, fumes and arc glare (Thomas, Johnson, & Wiesner, 2003). In the friction stir welding process, the tool used has a specially designed shoulder and pin profile that rotates and plunges into the workpiece edges to be joined and moved along the weld direction. Heat is created from friction between the tool and the workpiece. At an elevated temperature, the workpiece material undergoes extreme plastic deformation and material flows around the tool (Mishra & Ma, 2005). Most of the plasticised material flows from the advancing side (AS) to the retreating side (RS) and deposited plasticised material behind the tool creates a weld zone, joining the workpiece material (Kumar & Kailas 2008; He, Gu, & Ball, 2014). Performance of the welded joint is evaluated by the weld-zone mechanical properties, which depend on various process parameters like tool welding speed, rotational speed, tool pin-profile and axial force, among others.

Several attempts have been made to identify pin-profile effect on weld-zone shape, mechanical properties and microstructure. Lorrain, Favier, Zahrouni and Lawrjaniec (2010) described the effect of the tapered cylindrical pin as producing three flat surfaces and a straight cylindrical surface in the shape of the weld zone. They observed that the weld size was larger in the straight cylindrical surface than in a TC3F pin. Su, Wu, Bachmann and Rethmeier (2015) analysed pin-profile effect on the behaviour of material flow using conical and triflat pin-profile tools. It was concluded that the weld zone caused by a conical pin-profile tool was smaller than that caused by a triflat pin-profile tool. Hasan, Bennett and Shipway (2015) predicted the effect of unworn and worn tool geometries on the shape of the weld zone. Their results showed that worn tools produced the conical shape of the weld zone.

Marzbanrad, Akbari, Asadi and Safaee (2014) analysed cylindrical and square pin-profile tool effects on the microstructure of the weld zone of AA5083. They observed that using a square pin-profile tool produced smaller recrystallised grains in the weld zone compared to when using the cylindrical pin profile. Gadakh and Kumar (2014) analysed the effect of tapering cylindrical and straight cylindrical pin-profile tools on weld joints. They observed that fine microstructure in the stir zone and higher hardness and strength of the weld joint were developed by using a tapering cylindrical pin-profile tool. Zhao, Lin, Wu and Qu (2005) explained pin-profile effects on mechanical properties and weld shape. They concluded that plastic material flow and mechanical properties of the weld joint are strongly affected by pin profile. Elangovan, Balasubramanian and Valliappan (2008) studied the various tool pin-profiles and rotation speed effects on mechanical properties of FSW AA6061 aluminum alloy. They found that pin-profile effects on mechanical properties were significantly large with respect to rotation speed. Suresha, Rajaprakash and Upadhya (2011) analysed square and conical pin-profile effects on the joint efficiency of welds of 7075–T6 aluminum alloy. They observed that joint efficiency produced by the conical tool was better than that produced by the square tool.

Previous works on this subject showed that tool pin-profile effect on the weld-zone shape, mechanical properties and microstructure are significant. However, the relationship between the weld-zone shape and mechanical properties has not been discussed extensively. The objective of this article was to investigate tool pin-profile effects on weld-zone shape, mechanical properties and microstructure and also to show the relationship among them. Experiments were conducted using four different tool pin-profiles; two of the tools are commonly used (cylindrical and conical), while the two other were combinations: cylindrical-conical and

stepped-conical pin-profile tools. These tools were used to find the effect of tool geometry on weld-zone shape along with thickness. Weld zone, mechanical properties and microstructure were determined with each tool pin-profile. This work also showed the weld-zone shape effect on the mechanical properties of the weld joint.

MATERIALS AND METHOD

To investigate tool pin-profile effects on weld-zone shape, mechanical properties and microstructure, experiments were conducted on a vertical milling machine as shown in Figure 1(a). Four different types of fabricated tool pin-profile were used namely, the cylindrical pin (T1), conical pin (T2), cylindrical-conical pin (T3) and stepped-conical pin (T4) as shown in Figure 1(b). The schematic diagram of the FSW tool is shown in Figure 1(c). The length and shoulder diameter of all the tool pins were 5.7 mm and 18 mm, respectively. The other dimensions of the tool are given in Table 1.

Table 1
Dimensions of the tool pin-profile

Tool	Tool Pin Profile	Upper Pin Diameter (mm)	Lower Pin Diameter (mm)
T1	Cylindrical pin	6	6
T2	Conical pin	6	4
T3	Cylindrical-Conical pin	6	4
T4	Stepped-Conical pin	6	3
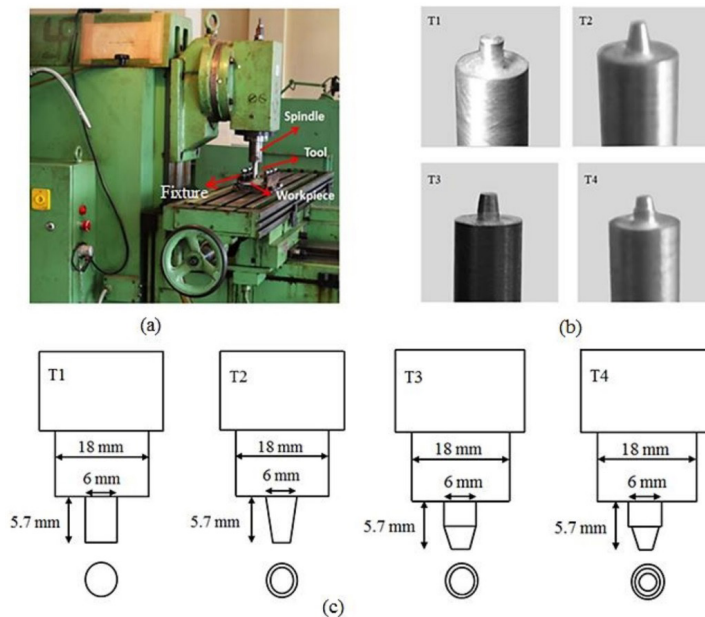
Figure 1. (a) Experimental setup; (b) Tool pin-profiles: Cylindrical pin tool (T1), conical pin tool (T2), cylindrical-conical pin tool (T3), and stepped-conical pin tool (T4); (c) Schematic diagram of the FSW tool

The workpiece material was 6061 aluminium alloy rolled plates with the dimensions 100 x 50 x 6 mm. The experiment was conducted at 1120 rpm of tool rotation and 80 mm/min of welding speed at zero tool tilt angle. For the microstructural characterisation of the weld zone, the specimens were cut in perpendicular to the weld direction by wire electro-discharge machining (WEDM). The specimens were polished using abrasive disks (600, 800, 1000 and 2000 grade), followed by a final polishing diamond on tissue disk and etched by Keller's reagent (Muthukumaran & Mukherjee, 2008). The chemical composition of Keller's reagent is nitric acid (3 ml), hydrochloric acid (2 ml), hydrofluoric acid (1 ml) and water (94 ml). The weld-zone shape and microstructure were measured along the thickness of the workpiece by an optical microscope. To determine tool pin-profile effect on weld-zone shape, four characteristic lengths were defined as shown in Figure 2. In Figure 2, L_r, L_1, L_2 and L_3 represent the four characteristic lengths at the bottommost portion, 1.5 mm away from the bottommost portion, 3 mm away from the bottommost portion and 4.5 mm away from the bottommost portion of the weld, respectively.
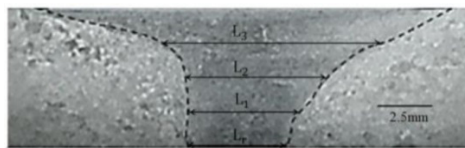
Figure 2. Weld-zone macrostructure
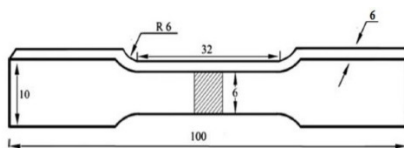
Figure 3. Workpiece dimension (mm) for testing tensile strength

Mechanical properties such as yield strength, ultimate tensile strength, ductility and joint efficiency were measured. For this purpose, three specimens were prepared per joint according to 'ASTM: E8 M' as given in Figure 3. The tensile tests were conducted at 1 mm/min cross head speed in a universal testing machine (UTM).

RESULTS AND DISCUSSION

Tool Pin-Profile Effect on Weld-Zone Shape

Tool pin-profile effect on the shape of the weld zone was determined by macrostructure of a cross section of the weld joint. In the macrostructure, the weld zone was shown by a dashed line as given in Figure 4. Weld zone includes the thermo-mechanical affected zone (TMAZ) and the nugget zone (NZ). It was observed that all pin-profile affected the basin shape of the

weld zone. A similar result was found in the literature (Mishra & Ma, 2005). The weld zone is larger near the shoulder compared to the bottom because the shoulder stirring effect was large at the top surface of the workpiece and its value decreases from top to bottom. It was also observed that the weld zone shape in the advancing and retreating side was not symmetric. The reason for the asymmetric shape was deposition of material in the retreating side from the advancing side during the welding process (Kumar & Kailas, 2008).
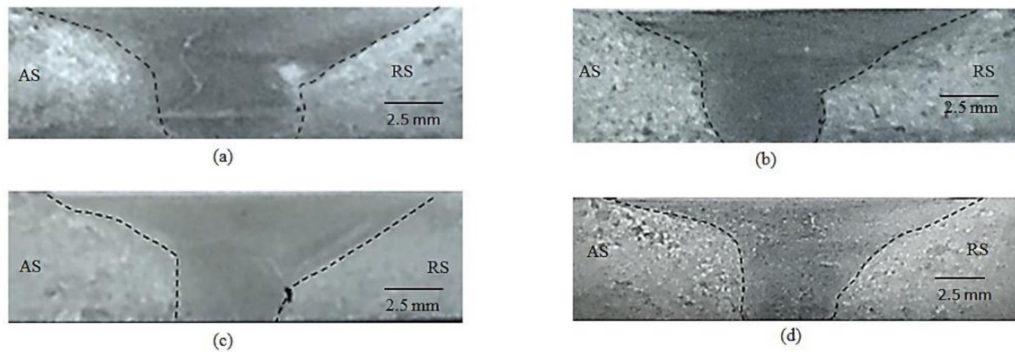
Figure 4. Experimental weld zone: (a) T1: Cylindrical pin tool; (b) T2: Conical pin tool; (c) T3: Cylindrical-Conical pin tool and; (d) T4: Stepped-Conical pin tool

However, the thicknesses of the weld zone of all the tool pin-profiles were different, as shown from top to bottom in Figure 5. It can be observed from the figure that cylindrical-conical tool weld zone thickness was large at the top of the workpiece in comparison with all the other tools.
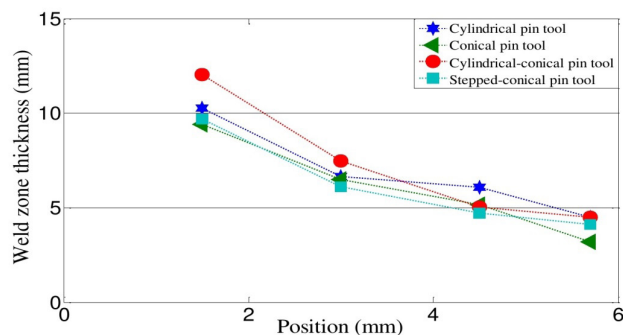
Figure 5. Experimental weld zone along with the thickness of the workpiece from top to bottom for all the tools

At the mid-point of the pin length for the conical, cylindrical-conical and stepped-conical tool weld zones, the thickness decreased gradually but for the cylindrical tool, the weld zone thickness did not decease gradually. The weld-zone thickness of the conical pin-profile tool was small near the bottom of the tool pin compared with all the other tools. The difference in

the weld-zone thickness between top and bottom of the pin length was lowest in the stepped-conical tool. This means that the small basin shape was produced by a stepped-conical tool rather than the other tools.

Tool Pin-Profile Effect on Microstructure of FSW Joint

In the friction stir welding process, the workpiece material undergoes extreme plastic deformation at elevated temperatures and grain is recrystallized; this affects the microstructure of the weld zone. The pin geometry affects the microstructure greatly. Figure 6 shows four different tool pin-profile effects on the microstructure of the weld zone.
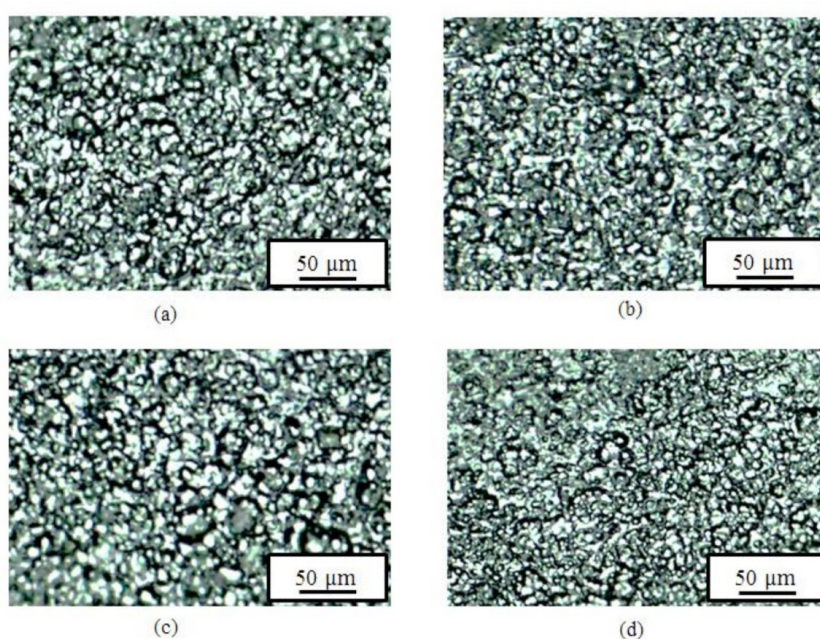
Figure 6. Microstructure in the weld nugget zone: (a) Cylindrical pin profile tool (T1); (b) Conical pin profile tool (T2); (c) Cylindrical-Conical pin profile tool (T3); (d) Stepped-Conical pin profile tool (T4)

Fine-grain structure was produced by the stepped-conical pin profile tool in the weld zone as shown in Figure 6(d). The largest grain structure was observed in the cylindrical-conical pin-profile tool as shown in Figure 6(c). Figure 6(a) and 6(b) show that grain structure produced by the conical tool was smaller than that produced by the cylindrical tool. It was found that using the stepped-conical and conical pin-profile tool weld-zone microstructure produced finer grain, which improved the mechanical property, as described in the following section.

Tool Pin-Profile Effect on Mechanical Properties

The mechanical properties of the weld joint were measured by tensile test on the universal testing machine. Table 2 shows the fracture locations of the FSW joints obtained by the different tool pin-profiles. It was observed that fractures occurred at two locations.

Table 2
Fracture location of tensile test specimens

Tool	Fracture Location	Photograph of Fractured Specimen	
		AS	RS
T1	HAZ, RS		
T2	HAZ, RS		
T3	NZ		
T4	HAZ, RS		

The first location was inside the weld zone and the second was outside the weld zone in the retreating side (RS). The fracture occurred inside the weld zone in the cylindrical-conical pin-profile tool (T3). It was also observed from Figure 4(c) that a void existed inside the weld zone that weakened the weld joint. This revealed that a low stirring effect was produced by the cylindrical-conical pin profile. Three other tool pin-profile fractures occurred outside the weld zone on the retreating side (RS). The weld zones produced by the three tools did not form a void as shown in Figure 4(a, b, d). This showed that the weld zone produced by the three tools using the stirring effect was strong. A similar result was also observed by Yuqing, Liming, Fencheng, Yuhua and Li (2017). The results were further evaluated by comparison of mechanical properties with respect to the base metal (B). Tool pin-profile effects on yield strength, ultimate tensile strength, ductility and joint efficiency are shown in Figure 7.

It can be observed from the figure that yield strength, ultimate tensile strength, ductility and joint efficiency of the weld joint produced by the stepped-conical pin profile were superior compared with those produced by the cylindrical-conical tool pin. The joints produced by the cylindrical and cylindrical-conical pin-profile tools showed equivalent yield strength and ultimate tensile strength. Ductility produced by the cylindrical-conical pin-profile tool was significantly lower compared with that produced by the cylindrical pin-profile tool.

Salari, Jahazi, Khodabandeh and Ghasemi-Nanesa (2014) explained that pin profile influences plastic deformation and material mixing during the welding process; consequently, volume and material flow vary as a function of tool geometry. It was predicted that the stepped-conical pin-profile tool would improve material mixing and enhance material movement in the weld zone, increasing the mechanical properties of the structure.
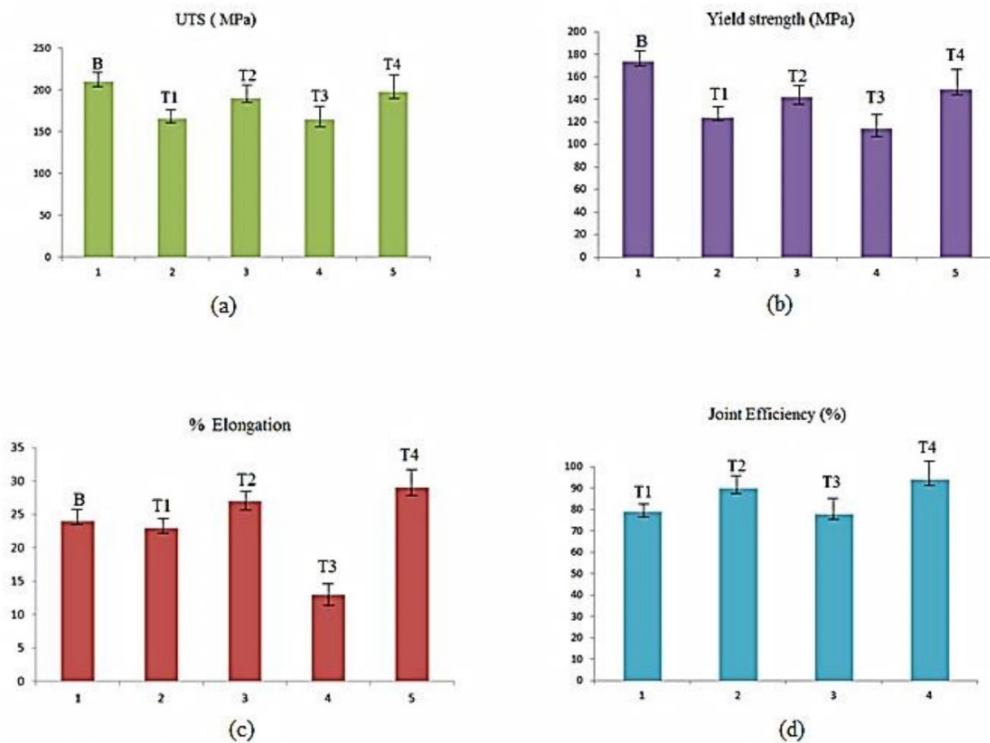
Figure 7. Tool profile effect on mechanical properties: (a) Ultimate tensile strength; (b) Yield strength; (c) Ductility (percentage of elongation); (d) Joint efficiency

It was also observed from the figure that the weld joint fabricated using the stepped-conical tool pin-profile had a high value of ductility (percentage of elongation) compared with the base metal (B). The stepped-conical tool produced maximum joint efficiency (94%), while the cylindrical-conical tool had minimum joint efficiency (78%).

This analysis also showed that size and shape of the weld zone significantly affected the mechanical properties. In the cylindrical-conical pin-profile tool, the mechanical properties were inferior and the basin shape of the weld zone was large compared with the other pin-profile tool. The mechanical properties produced by the stepped-conical pin-profile tool were superior and the basin shape of the weld zone was small compared with the other tool pin-profiles. Finally, it could be concluded that the small basin shape of the weld zone produced high yield strength, ultimate tensile strength, ductility and joint efficiency.

CONCLUSION

Tool pin-profile effects on mechanical properties and weld zone were experimentally investigated and the following conclusions were made:

1. Weld-zone thickness of the cylindrical-conical tool near the shoulder was highest but lowest for the conical pin-profile tool near the bottom of the tool pin compared with all the other tool pin-profiles.

2. The conical, cylindrical-conical and stepped-conical tool pin profiles produced weld-zone thickness that decreased gradually in the middle of the workpiece but the weld-zone thickness produced by the cylindrical tool did not decease gradually.

3. The stepped-conical and conical pin-profile tool weld-zone microstructure produced finer grain compared with that produced by the cylindrical and cylindrical-conical pin-profile tools.

4. The stepped-conical pin tool exhibited higher mechanical properties compared with the other tool pin-profiles.

5. The small basin shape of the weld zone produced higher mechanical properties.

ACKNOWLEDGEMENT

REFERENCES

Elangovan, K., Balasubramanian, V., & Valliappan, M. (2008). Effect of tool pin profile and tool rotational speed on mechanical properties of friction stir welded AA6061 aluminium alloy. *Materials and Manufacturing Processes, 23*(3), 251–260.

Gadakh, V. S., & Kumar, A. (2014). Friction stir welding window for AA6061-T6 aluminium alloy. Proceedings of the Institution of Mechanical Engineers, Part B: *Journal of Engineering Manufacture, 228*(9), 1172–1181.

Gibson, B. T., Lammlein, D. H., Prater, T. J., Longhurst, W. R., Cox, C. D., Ballun, M. C., … & Strauss, A. M. (2014). Friction stir welding: Process, automation, and control. *Journal of Manufacturing Processes, 16*(1), 56–73.

Hasan, A. F., Bennett, C. J., & Shipway, P. H. (2015). A numerical comparison of the flow behaviour in friction stir welding (FSW) using unworn and worn tool geometries. *Materials and Design, 87,* 1037–1046.

He, X., Gu, F., & Ball, A. (2014). A review of numerical analysis of friction stir welding. *Progress in Materials Science, 65,* 1–66.

Kumar, K. S., & Kailas, S. V. (2008). The role of friction stir welding tool on material flow and weld formation. *Materials Science and Engineering: A, 485*(1), 367–374.

Lorrain, O., Favier, V., Zahrouni, H., & Lawrjaniec, D. (2010). Understanding the material flow path of friction stir welding process using unthreaded tools. *Journal of Materials Processing Technology, 210*(4), 603–609.

Marzbanrad, J., Akbari, M., Asadi, P., & Safaee, S. (2014). Characterization of the influence of tool pin profile on microstructural and mechanical properties of friction stir welding. *Metallurgical and Materials Transactions B, 45*(5), 1887–1894.

Mishra, R. S., & Ma, Z. Y. (2005). Friction stir welding and processing. *Materials Science and Engineering: R: Reports, 50*(1), 1–78.

Muthukumaran, S., & Mukherjee, S. K. (2008). Multi-layered metal flow and formation of onion rings in friction stir welds. *The International Journal of Advanced Manufacturing Technology, 38*(1), 68–73.

Salari, E., Jahazi, M., Khodabandeh, A., & Ghasemi-Nanesa, H. (2014). Influence of tool geometry and rotational speed on mechanical properties and defect formation in friction stir lap welded 5456 aluminum alloy sheets. *Materials and Design, 58,* 381–389.

Su, H., Wu, C. S., Bachmann, M., & Rethmeier, M. (2015). Numerical modeling for the effect of pin profiles on thermal and material flow characteristics in friction stir welding. *Materials and Design, 77,* 114–125.

Suresha, C. N., Rajaprakash, B. M., & Upadhya, S. (2011). A study of the effect of tool pin profiles on tensile strength of welded joints produced using friction stir welding process. *Materials and Manufacturing Processes, 26*(9), 1111–1116.

Thomas, W. M., Johnson, K. I., & Wiesner, C. S. (2003). Friction stir welding – Recent developments in tool and process technologies. *Advanced Engineering Materials, 5*(7), 485–490.

Yuqing, M., Liming, K., Fencheng, L., Yuhua, C., & Li, X. (2017). Effect of tool pin-tip profiles on material flow and mechanical properties of friction stir welding thick AA7075-T6 alloy joints. *The International Journal of Advanced Manufacturing Technology, 88*(1-4), 949–960.

Zhao, Y. H., Lin, S. B., Wu, L., & Qu, F. X. (2005). The influence of pin geometry on bonding and mechanical properties in friction stir weld 2014 Al alloy. *Materials letters, 59*(23), 2948–2952.

SCIENCE & TECHNOLOGY

PERTANIKA
JOURNALS

Computation of Antifractals - Tricorns and Multicorns and Their Complex Nature

Narayan Partap[1]*, Sarika Jain[2] and Renu Chugh[3]

[1]*Northern India Engineering College*, *GGSIP University*, *New Delhi, 110053 India*

[2]*Amity Institute of Information Technology, Amity University, Uttar Pradesh*, *201313 India*

[3]*Department of Mathematics, M.D. University, Rohtak, Haryana*, *124001 India*

ABSTRACT

Since the last decade, the study of fractals and antifractals and their complex nature has been one of the significant areas of research. While many researchers are working on fractals, only a very few have focused on antifractals and their dynamic nature. This paper discovers new antifractals like tricorns for the complex antipolynomial $\bar{z}^n + c$ for $n \geq 2$ in the *GK*-orbit. The main focus of this paper is to develop a new collection of antifractals and to understand their attributes using the *GK*-orbit.

Keywords: Antifractals, antipolynomial, *GK*-orbit, multicorns, tricorns

INTRODUCTION

Anti-Julia, tricorns and multicorns are some of the striking examples of antifractals that can be generated through the dynamics of antiholomorphic complex polynomials $\bar{z}^n + c$ of the complex polynomial $z^n + c$ for $n \geq 2$. They have been named antifractals as tricorns and multicorns are generated through antiholomorphic complex polynomials $\bar{z}^n + c$ for $n \geq 2$. In mathematics, we define tricorn in a similar way as the Mandelbrot set, but we use the $z \to \bar{z}^2 + c$ map instead of the $z \to z^2 + c$ map, the map used for the Mandelbrot set; therefore, sometimes the tricorn is also called the Mandelbar set. The tricorn was initiated by Crowe, Hasson, Rippon and Strain-Clark (1989). They worked on the antiholomorphic complex polynomial $\bar{z}^n + c$ and studied the bifurcation diagram for iterates of the above map. These tricorns and multicorns possess various attributes. The three-cornered nature, the main characteristic of a tricorn, repeats with variations at different scales, reflecting the same sort of self-similarity as the Mandelbrot set.

Apart from complications and their beauty, the same may be connected or disconnected. The set of all parameters c for

E-mail addresses:
narayan.maths81@gmail.com (Narayan Partap)
ashusarika@gmail.com (Sarika Jain)
chugh.r1@gmail.com (Renu Chugh)
*Corresponding Author

which the Julia set of $\bar{z}^2 + c$ is connected, is called a multicorn. These multicorns are simply higher-order tricorns and they all are used in the commercial field.

Milnor (1999) coined the term 'tricorn' for the connectedness locus for the antiholomorphic polynomial $\bar{z}^2 + c$, which plays an intermediate role between quadratic and cubic polynomials. The tricorn has many similarities with the Mandelbrot set due to a compact subset of C.

Crowe et al. (1989) used the complex polynomial as a formal analogy with Mandelbrot sets and they called it the Mandelbar set, which has many similarities to the Mandelbrot set. They also brought their bifurcation features along arcs rather than at points. Milnor (1999) found multicorns in a real slice of the cubic connectedness locus. Winters (1990) discovered the boundary of the tricorn, showing that it contains a smooth arc. Nakane and Schleicher (2003) revealed the various properties of tricorns and multicorns along with their complexity and their beauty and stated that multicorns are generalised tricorns or tricorns of higher order. They also explored the antifractal of a complex polynomial of the form $M_c(z) = \bar{z}^n + c$ for $n \geq 2$, showing that it is either connected or disconnected. The set of parameters c for which the Julia set of $M_c(z)$ is connected is called the multicorn. Lau and Schleicher (1996) analysed the symmetries of tricorns and multicorns.

Rani and Kumar (2009) studied the dynamics of the complex polynomial $M_c(z) = \bar{z}^n + c$ for $n \geq 2$ and Rani (2010) introduced superior tricorns and superior multicorns using Mann's iterative technique. In the same year, Chauhan, Rana and Negi (2010) introduced relative superior tricorns and relative superior multicorns using the Ishikawa iterative techniques. Kumar, Chugh and Rani (2012) generated a new class of tricorns and multicorns in the Noor orbit. Recently, Kang, Rafik and Latif (2015) presented tricorns and multicorns using the S-iteration scheme.

In this paper, we explore a new class of tricorns and multicorns in the GK-orbit and analyse their mathematical characteristics. These generated anti-fractals such as anti-Julia, tricorns and multicorns have their own attraction and are different from previously generated antifractals. Tricorn prints are being used for commercial objects like tricorn coffee cups, jugs and tricorn apparels like tricorn dresses and tricorn shirts, among others.

METHOD

Before generating a new class of tricorns and multicorns for the complex polynomial $M_c(z) = \bar{z}^n + c$, we need to define the basic terminology of the research.

Julia Set

The filled-in Julia set of the function $F(z) = z^n + c$ is defined as:

$$K(F) = \{z \in \mathbb{C} : F^k(z) \text{ does not tend to } \infty \},$$

where, $\mathbb{C}$ is the complex space, $F^k(z)$ is the kth iterate of function F and $K(F)$ denotes the filled Julia set. The Julia set of the function F is defined to the boundary of $K(F)$, that is, $J(F) = \partial K(F)$, where $J(F)$ denotes the Julia set. The set of points whose orbits are bounded under $F_c(z) = z^2 + c$ is called the Julia set (Peitgen, Jurgens, & Saupe, 1994).

The Julia and Mandelbrot sets, which have great importance in fractal geometry, can be generated through fixed points using different complex maps (Wang & Shi, 2006; Nazeer, Kang, Tanveer, & Shahid, 2015).

Multicorn

The multicorn for the complex function $M_c(z) = \bar{z}^n + c$ is defined as the collection of all $c \in \mathbb{C}$, for which orbit of the point 0 is bounded, that is:

$$M_c = \{c \in \mathbb{C} : M_c^n(0) \text{ does not tend to } \infty\},$$

where, $\mathbb{C}$ is the complex space and $M_c^n(z)$ is the nth iterate of function $M_c(z)$. An equivalent formulation is the connectedness of loci for higher degree antiholomorphic polynomials $M_c(z) = \bar{z}^n + c$ that are called multicorns (Devaney, 1992).

It has been observed that for $n = 2$, multicorns reduce to tricorns. Moreover, the tricorns naturally live in the real slice $d = \bar{c}$ in the two-dimensional parameter space of map $z \rightarrow (z^2 + d)^2 + c$. They have $(n + 1)$-fold rotational symmetries. After dividing these symmetries, the resulting multicorns are called unicorns (Nakane & Schleicher, 2003).

GK-Orbit

Consider f as a self-map from a metric space X into itself such that $f : X \rightarrow X$. Then the generalised Kransnoselskii iteration scheme (Schaefer, 1957) is defined as:

$$x_{n+1} = (r_n)x_n + (1 - r_n)f(x_n), \qquad n = 0, 1, 2, 3, \ldots$$

where, $0 \leq r_n < 1$ and $< r_n >$ are a sequence of positive numbers.

For $x_0 \in X$ (real or complex numbers), construct a sequence $< x_n >$ in X in the following manner:

$$x_1 = r_1 x_0 + (1 - r_1)f(x_0),$$

$$x_2 = r_2 x_1 + (1 - r_2)f(x_1),$$

$$x_n = r_n x_{n-1} + (1 - r_n)f(x_{n-1}),$$

where, $0 \leq r_n < 1$ and $\{r_n\}$ are a sequence converging to non-zero numbers. Then the sequence $x_1, x_2, x_3 \ldots x_n$ constructed above is called the *GK*-orbit. We may also denote as *GKO* (f, x_0, r_n).

To visualize antifractals in the *GK*-orbit for $z \rightarrow \bar{z}^n + c$, we need an escape criterion with respect to the *GK*-orbit. The escape criterion for $z \rightarrow \bar{z}^n + c$ in *GK*-orbit is ma x $\{|c|, [2/(1 - r)]^{1/(n-1)}\}$ for $0 \leq r_n < 1$.

Escape Criterions for Antifractals

To prove the general escape criterion, we first prove the following result for the cubic complex polynomial.

Escape criterion for cubic polynomial. Suppose for a function $M_{a,b} = \bar{z}^3 + a\bar{z} + b$, $|\bar{z}| \geq |b| > \{|a| + 2/(1-r)\}^{1/2}$, where, $0 \leq r < 1$ and a, b are complex numbers.

Define

$$z_1 = (1-r)M_{a,b}(z) + r\bar{z}$$

$$z_2 = (1-r)M_{a,b}(z_1) + r\bar{z_1}$$

$$\vdots$$

$$z_n = (1-r)M_{a,b}(z_{n-1}) + r\bar{z}_{n-1}, \quad n = 2, 3, \dots$$

Then $|z_n| \to \infty$ as $n \to \infty$

This is proved by definition:

$$|z_1| = \left|(1-r)M_{a,b}(z) + r\bar{z}\right|$$

$$= \left|(1-r)\,(\bar{z}^3 + a\bar{z} + b)\, + r\bar{z}\right|$$

$$= \left|(1-r)\bar{z}^3 + a(1-r)\bar{z} + r\bar{z} + (1-r)b\,\right|$$

$$\geq \left|(1-r)\bar{z}^3 + a(1-r)\bar{z} + r\bar{z}\right| - |(1-r)b\,|$$

$$\geq |\bar{z}|\left|(1-r)\bar{z}^2 + a(1-r) + r\right| - |(1-r)\bar{z}\,|$$

$$\geq |\bar{z}|(|(1-r)\bar{z}^2 + a(1-r)| - r) - |(1-r)\bar{z}|$$

$$= |\bar{z}|(|(1-r)\bar{z}^2 + a(1-r)| - r - (1-r))$$

$$= |\bar{z}|\{(1-r)|\bar{z}^2 + a| - 1\}$$

$$= |\bar{z}|(1-r)\{|\bar{z}^2 + a| - 1/(1-r)\}$$

$$\geq (1-r)|\bar{z}|[|\bar{z}|^2 - |a| - 1/(1-r)]$$

As $|\bar{z}| > \{|a| + 2/(1-r)\}^{1/2}$.

Therefore,

$$|\bar{z}|^2 - |a| - 1/(1-r)\} > 1/(1-r)$$

$$(1-r)[|\bar{z}|^2 - |a| - 1/(1-r)] > 1$$

Hence, we can find a $\gamma > 1$ such that

$$|z_1| > \lambda|\bar{z}|.$$

Repeating this argument n time, we find $|z_n| > \lambda^n |\bar{z}|$.

Therefore, the GK-orbit of $\bar{z}$ under the cubic polynomial $M_{a,b}(z)$ tends to infinity, and max

$\left\{|b|, [|a| + 2/(1-r)]^{1/2}\right\}$ is the required escape criterion. This completes the proof.

General escape criterion. Assume the function of the form $M_c(z) = \bar{z}^n + c$, $n = 1, 2, 3\ldots$

where $0 \leq r < 1$ and c is in the complex plane, Define

$$z_1 = (1-r)M_c(z) + r\bar{z}$$

$$z_2 = (1-r)M_c(z_1) + r\bar{z}_1$$

$$\vdots$$

$$z_n = (1-r)M_c(z_{n-1}) + r\bar{z}_{n-1}, \quad n = 2, 3, \ldots$$

Then, the general escape criterion is max $\{|c|, [2/(1-r)]^{1/(n-1)}\}$.

We shall prove the result by induction.

For $n = 1$, $M_c(z) = \bar{z} + c$ so, the escape criterion is $|c|$, which is obvious i.e. $|z| > \max\{|c|, 0\}$.

For $n = 2$, $M_c(z) = \bar{z}^2 + c$

so, the escape criterion is $max \{|c|, 2/(1-r)\}$, which can be easily proved.

For $n = 3$, $M_c(z) = \bar{z}^3 + c$

This result follows from above theorem with $a = 0$ and $b = c$ i.e. the escape criterion for $\bar{z}^3 + c$ is $\max\{|c|, [2/(1-r)]^{1/2}\}$; thus, the theorem is true for $n = 1, 2$, and 3.

Now, suppose the theorem is true for any n, we shall prove it for $n + 1$.

Suppose $M_c(z) = \bar{z}^{n+1} + c$ and $|\bar{z}| \geq |c| > [2/(1-r)]^{1/n}$

Then, $|z_1| = |(1-r)M_c(z) + r\bar{z}|$

$$= |(1-r)(\bar{z}^{n+1} + c) + r\bar{z}|$$

$$= |\bar{z}^{n+1}(1-r) + r\bar{z} + (1-r)c|$$

$$\geq |\bar{z}^{n+1}(1-r) + r\bar{z}| - |(1-r)c|$$

$$\geq |\bar{z}|[|\bar{z}^n(1-r) + r|] - |(1-r)\bar{z}|$$

$$\geq |\bar{z}|[|(1-r)\bar{z}^n| - r] - |(1-r)\bar{z}|$$

$$= |\bar{z}|[|(1-r)\bar{z}^n| - 1]$$

As $|\bar{z}| > [2/(1-r)]^{1/n}$, we can find a $\lambda > 0$ such that $(1-r)|\bar{z}|^n - 1 > 1 + \lambda$

Therefore,

$$|z_1| > (1+\lambda)|\bar{z}|$$

Repeating this argument, we find

$$|z_n| > (1+\lambda)^n|\bar{z}|$$

Thus, the *GK*-orbit of $\bar{z}$ under the general iteration function $\bar{z}^{n+1} + c$ tends to infinity. Therefore, max $\{|c|, [2/(1-r)]^{1/n}\}$ is the required escape criterion. This completes the proof.

Corollary. Suppose that $|c| > [2/(1-r)]^{1/n-1}, 0 \leq r < 1$. Then the *GK*-orbit *GKO* $(G_c, 0, r)$ escapes to infinity i.e. orbit of 0 escapes to infinity under $M_c(z)$.

General algorithm. Assume that for some $k \geq 0$, we have $|z_k| > max\ \{|c|, [2/(1-r)]^{1/(k-1)}\}$. Then $|z_{k+1}| > (1+\gamma)|z_k|$ and $|z_n| \to \infty$ as $n \to \infty$

The above result provides a general algorithm for computing the filled *GK* Julia set for the function $M_c(z) = \bar{z}^n + c$, n=1, 2, 3...

RESULTS AND DISCUSSION

Tricorns and Multicorns

Antifractals, tricorns and multicorns in *GK*-orbit, were generated for the complex polynomial $M_c(z) = \bar{z}^n + c$ using the general escape criterion with the software Mathematica 9.0 (see Figures 1-12).

Figure 1.
Tricorn for $r = 0.001$, $n = 2$

Figure 2.
Tricorn for $r = 0.031$, $n = 2$

Figure 3.
Tricorn for $r = 0.051$, $n = 2$

Figure 4.
Tricorn for $r = 0.191$, $n = 2$

Figure 5.
Tricorn for $r = 0.391$, $n = 2$

Figure 6.
Tricorn for $r = 0.195$, $n = 2$
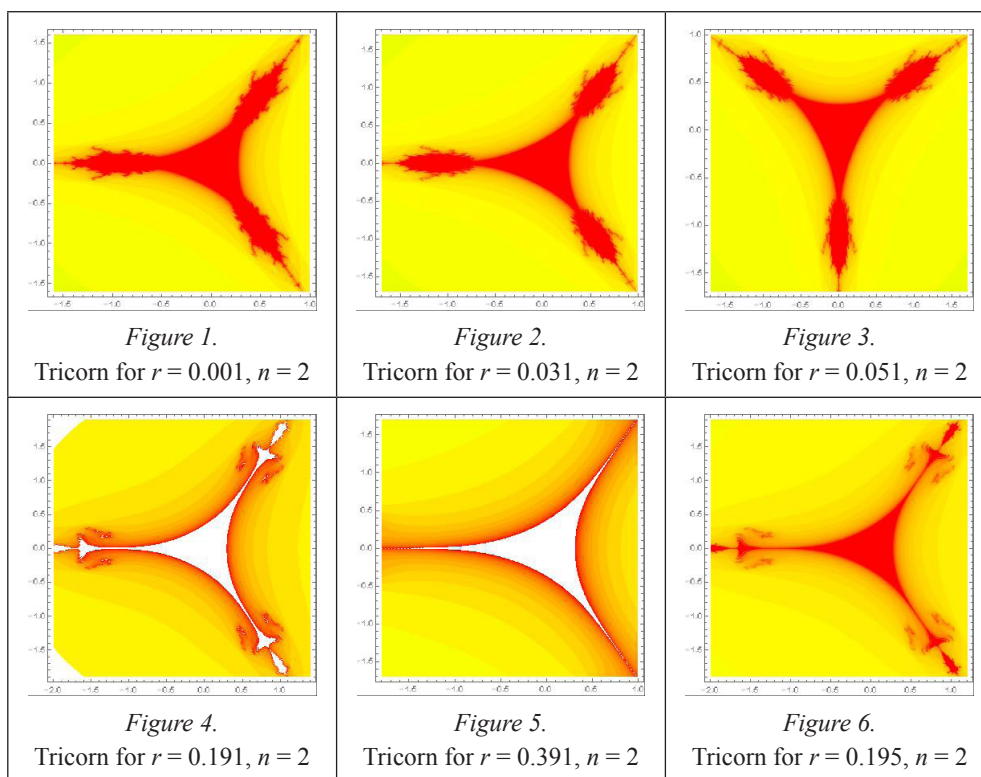
The following characteristics were noticed after the generation of tricorns and multicorns in the *GK*-orbit:

- It was observed that branches in the tricorns and multicorns were one more than the degree i.e. the number of branches was $n+1$, if n is the power of $\overline{z}$. Also, some branches had n sub-branches (Figures 1-3).
- The multicorns exhibited $(n+1)$-fold rotational symmetries.

Figure 7.
Multicorn for $r = 0.091$, $n = 3$

Figure 8.
Multicorn for $r = 0.191$, $n = 3$

Figure 9.
Multicorn for $r = 0.291$, $n = 5$

Figure 10.	*Figure 11.*	*Figure 12.*
Multicorn for $r = 0.151$, $n = 20$	Multicorn for $r = 0.151$, $n = 50$	Multicorn for $r = 0.01$, $n = 374$
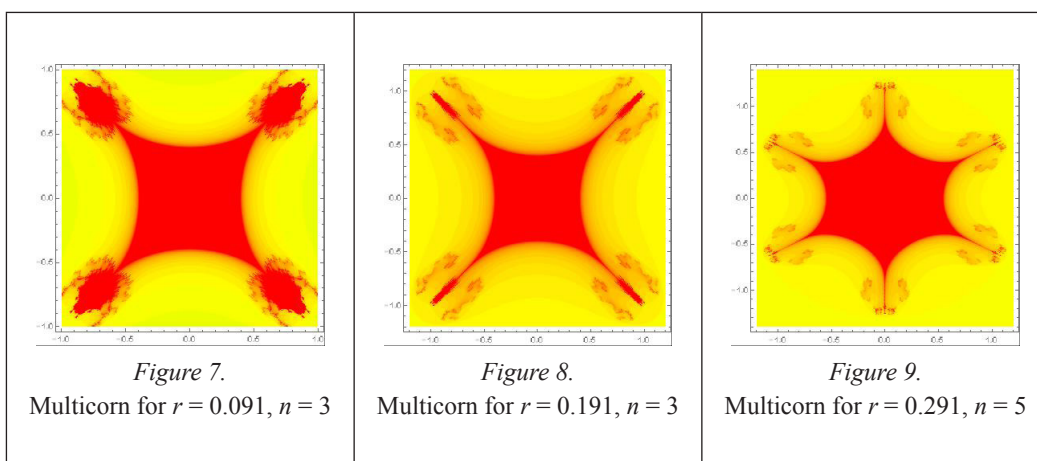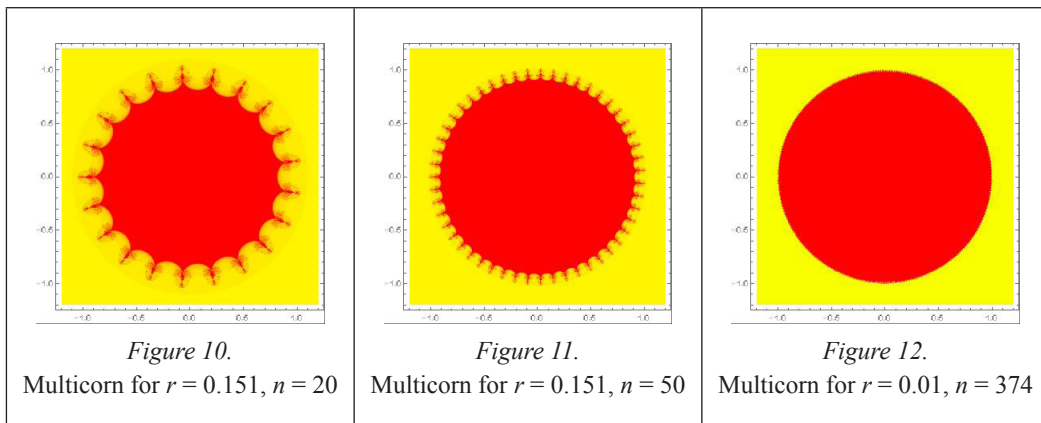
- It was also observed that if we increased the value of r, the tricorns in Figures 4-6 and the multicorns in Figures 7-9 became thinner.
- Figures 3-6 and Figures 7-8 showed that for a fixed value of n, there existed many antifractals.
- If we increased the degree of the complex polynomial, the generated multicorns became a circular saw (Figures 10-12). Kumar at al. (2012) came to a similar conclusion while generating tricorns and multicorns using different iterates. The name circular saw was first considered by Rani and Kumar (2009) for Mandelbrot sets.

Anti-Julia Sets in GK-Orbit

Anti-Julia sets in *GK*-orbit were generated with the help of a general escape criterion for the complex polynomial $M_c(z) = \bar{z}^n + c$ using the software Mathematica 9.0 (see Figures 13-21).

Figure 13.	*Figure 14.*	*Figure 15.*
Anti-Julia for $r = 0.809$, $c = -35.505-0.001I$, $n = 2$	Anti-Julia for $r = 0.809$, $c = -36.505-0.071I$, $n = 2$	Anti-Julia for $r = 0.815$, $c = -47.505-0.001I$, $n = 2$

Figure 16.
Anti-Julia for $r = 0.834$,
$c = -40.505+0.5I$, $n = 2$

Figure 17.
Anti-Julia for $r = 0.834$,
$c = -37.505+1.95I$, $n = 2$

Figure 18.
Anti-Julia for $r = 0.834$,
$c = -33.505+1.55I$, $n = 2$

Figure 19.
Anti-Julia for $r = 0.860$,
$c = -0.60-17.20I$, $n = 3$

Figure 20.
Anti-Julia for $r = 0.860$,
$c = -0.60-19.20I$,
$n = 3$

Figure 21.
(Christmas tree)
Anti-Julia for $r = 0.879$,
$c = 0.002-11.50I$, $n = 3$

The following observations were made:

- Figures 13-15 show that if we fixed the parameter r and decreased the value of m in $c = m + n\,i$, the anti-Julia sets became more gracious. Similarly, Figures 16-18 show that if we fixed the parameter r and increased the value of m in $c = m + n\,i$, the connected anti-Julia sets in these figures looked like a helix.

- The anti-Julia sets in Figures 13-18 and in Figure 21 were connected, whereas the anti-Julia sets in Figures 19-20 were disconnected.

- The anti-Julia sets in Figures 17-18 show the reflection symmetry for the real axis point, whereas the anti-Julia sets in Figures 13-15 show the reflection symmetry for both the axes.

- The anti-Julia sets in Figures 19-20 looked like the meditation posture.

- The anti-Julia set in Figure 21 looked like a Christmas tree and showed the reflection symmetry for the imaginary axis.

CONCLUSION

In this paper, a new iterative procedure was employed in the antiholomorphic complex polynomials $\bar{z}^n + c$ of the complex polynomial $z^n + c$ for $n \geq 2$ and important as well as attractive antifractals like tricorns and multicorns were generated. It was fascinating to see that some of the generated antifractals were connected, while some were disconnected, and some anti-Julia sets were shaped like a helix and one anti-Julia set looked like a Christmas tree.

REFERENCES

Chauhan, Y. S., Rana, R., & Negi, A. (2010). New tricorn and multicorn of Ishikawa iterates. *International Journal of Computer Application, 7*(13), 25–33.

Crowe, W. D., Hasson, R., Rippon, P. J., & Strain-Clark, P. E. D. (1989). On the structure of the Mandelbar set. *Nonlinearity, 2*(4), 541–553.

Devaney, R. L. (1992). *A first course in chaotic dynamical systems: theory and experiment*. New York, NY: Addison-Wesley.

Kang, S. M., Rafik, A., & Latif, A. (2015). Tricorns and multicorns of *S*-iteration scheme. *Journal of Function Spaces, 2015,* 1–7. doi: 10.1155/2015/417167

Kumar, A., Chugh R., & Rani, M. (2012). Dynamics of antifractal in Noor orbit. *International Journal of Computer Applications, 57*(4), 11–15.

Lau, E., & Schleicher, D. (1996). Symmetries of fractals revisited. *Mathematical Intelligencer, 18*(1), 45–51.

Milnor, J. (1999*)*. Dynamics in one complex variable. *Introductory lectures*. Germany: Friedrich Vieweg & Sohn, Braunschweig.

Nakane, S., & Schleicher, D. (2003). On multicorn and unicorns I: Antiholomorphic dynamics, hyperbolic components and real cubic polynomials. *International Journal of Bifurcation and Chaos, 13*(10), 2825–2844.

Nazeer, W., Kang, S. M., Tanveer, M., & Shahid ,A. A. (2015). Fixed point results in the generation of Julia and Mandelbrot sets. *Journal of Inequalities and Applications, 2015*(1), 1–16.

Peitgen, H. O., Jurgens, H., & Saupe, D. (1994). *Chaos and fractals*. New York, NY: Springer-Verlag.

Rani, M. (2010). Superior tricorns and multicorns. In N. Mastorakis, V. Mladenov, A. Zaharim, & C. A. Bulucea (Eds.), *Proceedings of the 9th WSEAS International Conference on Applications of Computer Engineering* (pp. 58–61). Penang, Malaysia: Universiti Kebangsaan Malaysia.

Rani, M., & Kumar, M. (2009, December). Circular saw Mandelbrot sets. In C. A. Bulucea, V. Mladenov, E. Pop, M. Leba, & Nikos Mastorakis (Eds.), *Proceedings of the 14th WSEAS International Conference on Applied Mathematics* (pp. 131–136). Tenerife, Canary Islands, Spain: Puerto De La Cruz.

Schaefer, H. (1957). About the method of successive approximation. *Annual Report of the German Mathematical Association, 59,* 131–140.

Wang, X. Y., & Shi, Q. J. (2006). The generalized Mandelbort-Julia sets from a class of complex exponential map. *Applied Mathematics and Computation, 181*(2)*,* 816–825.

Winters, R. (1990). *Bifurcations in families of antiholomorphic and biquadratic maps*. (Doctoral dissertation). Department of Mathematics, Boston University, United States.

SCIENCE & TECHNOLOGY

An Explicit Drain Current Model in Subthreshold Regime for Graded Channel Schottky Barrier Gate All Around MOSFET to Improve Analog/RF Performance

Suman Sharma[1]*, Rajni Shukla[1] and Malay Ranjan Tripathy[2]

[1]*Department of Physics, Murthal Deenbandhu Chottu Ram University, Sonepat, Haryana, India*

[2]*Department of Electronics and communication Engineering, Amity University, Noida, Uttar Pardesh, India*

ABSTRACT

In this paper, an explicit analytical model for a Graded Channel Gate All Around Schottky Barrier MOSFET is presented. The problem of carrier mobility degradation in a uniformly highly-doped Schottky Barrier MOSFET is assessed by graded channel architecture in which high doping is considered in the source and low doping in the drain. The Evencent Mode Analysis is carried out using the superposition technique to obtain the surface potential in the channel. The one-dimensional Poisson's and two-dimensional Laplace equations are solved to calculate the surface potential. The subthreshold current is obtained using the surface potential analysis. A TCAD simulator (ATLAS-3D device simulator) is used for numerical simulation purposes. The developed model results are in good agreement with the TCAD simulations.

Keywords: ATLAS-3D, Gaussian graded channel, graded channel, high frequency, surrounding gate MOSFET

INTRODUCTION

In the past decade the Gate All Around MOSFET has received traction as one of the best candidates among existing Multi-Gate MOSFET architecture such as Dual Gate (DG) FET and Fin-FET for ultra-scaled nano devices (Chiang, 2005; Li & Hwang, 2009; Tsormpatzoglou et al., 2009; Zhuge, Wang, Huang, Xing, & Wang, 2008). It offers minimum short channel effects by providing excellent electrostatic control over the entire channel region (Chiang, 2005; Kumar, Haldar, Gupta, & Gupta, 2014a; Kumar, Haldar, Gupta, & Gupta, 2014b). However, the surrounding gate MOSFET reduces SCEs but at the nano-scale the P-N junction depletion width cannot be neglected.

This junction depletion width is reasonably large and so, resistance at source-channel and drain-channel junctions increases (Kumar, Haldar, Gupta, & Gupta, 2013). To overcome these problems, novel architecture like Schottky Barrier Surrounding Gate MOSFET and Junctionless Surrounding Gate MOSFET are being extensively explored (Colinge et al., 2010; Lee, Zhuang, Di & Han, 2013; Trivedi et al., 2016; Wang, Lou, & Lee, 2013). In Junctionless MOSFET there is no P-N junction and homogeneously doped source-channel-drains are used to avoid junction formation. However, it requires very high metal gate work function to fully deplete a highly-doped channel. Even though the device resolves the problem of junction formation at nanoscale, due to the very high channel doping, Carrier Mobility Degradation (CMD) occurs and degrades device performance compared with when the conventional surrounding gate MOSFET is used (Sharma & Vishvakarma, 2012; Yang et al., 2007).

To overcome all these issues, the Schottky Barrier Surrounding Gate MOSFET has been investigated to discover how improvements can be made (Marchi et al., 2014). The source and drain regions are metallic; this offers reduced source/drain resistance with contact resistivity (ρ_c ~ 10^{-9} Ω·cm) for metallic S/D compared with ρ_c ~ 10^{-7} Ω·cm in doped S/D junction (Pavanello, Martino, Dessard, & Flandre, 2000). The high doping is a problem in the use of the nano-scaled MOSFET due to carrier mobility degradation (CMD). To overcome the problem of CMD, graded channel architecture is proposed and extensively investigated over the conventional surrounding gate MOSFET to improve the performance of the device (Kaur, Kabra, Haldar, & Gupta, 2008; Kumar, Haldar, Gupta, & Gupta, 2016b). In the graded channel architecture, high doping was kept at the source and low doping at the drain.

Recently, Kumar, Haldar, Gupta and Gupta (2016a) proposed a novel Graded Channel Schottky Barrier GAA MOSFET to reduce carrier mobility degradation. This work extended the development of an analytical model for the calculation of subthreshold current for the GC-SB-GAA MOSFET. In this research work, an analytical model for the Graded Channel Schottky Barrier GAA MOSFET was developed and compared for performance with a conventional uniformly-doped channel with high doping ($N_{A/H} = 5 \times 10^{18}$ cm^{-3}) Schottky Barrier GAA MOSFET. The device with the Graded Channel (GC) had a high-doping profile in its half-channel region of the source and an undoped channel of the drain. A huge improvement in device performance (in on-state) was observed for the GC-SB-GAA MOSFET structure.

MATERIALS AND METHOD

Simulation and Device Structure

Simulation environment and fabrication feasibility. A schematic view of the Graded Channel (GC) Gate All Around Schottky Barrier (SB) MOSFET is shown in Figure 1. For numerical simulation, the ATLAS 3-D, device simulator (Atlas User's Manual, 2013) was used and the electrical as well as RF characteristics of both devices i.e. SB-GAA MOSFET and GC-SB GAA MOSFET were studied.

The models invoked for the numerical simulation purpose are: SRH recombination/thermal generation model, field dependent mobility (FLDMOB) model for high electric field velocity saturation, concentration dependent mobility (CONMOB) model for concentration dependent

mobility and Boltzmann carrier statistics for electrons and holes. The Newton-Gummel method was adopted for a numerical solution. Cerdeira et al. (2005) have shown the advantages of the graded-channel and the feasibility of fabrication of the proposed device.

Device structure. Figure 1 shows the schematic view of the GC-SB-GAA MOSFET. Two different regions of doping are shown. The following formula was used: Gate length L = 30 nm, radius of Si pillar (R_{Si}) = 10 nm , high Gaussiann doping p-type substrate ($N_{A/H}$) = 5×10^{18} cm^{-3}, low doping p-type substrate ($N_{A/L}$) = 1×10^{15} cm^{-3}, SiO_2 thickness (t_{ox}) = 3 nm, work-function of the metal source/drain $\Phi_{S/D}$ = 4.1 eV, dielectric permittivity of SiO_2 is ε_{ox} = 3.9 and gate metal workfunction Φ_M = 4.8 eV.
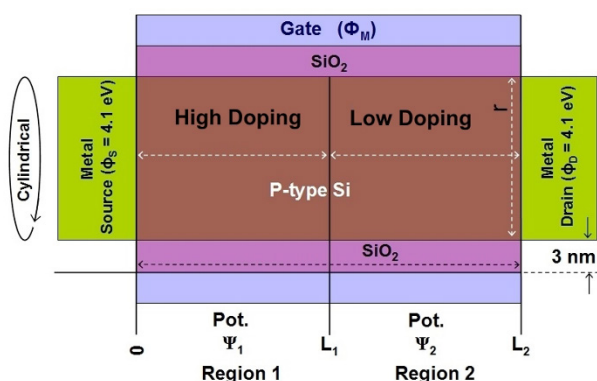
Figure 1. Schematic view of Schottky Barrier Surrounding Gate MOSFET with Gaussian Graded Channel (GGC)

The source side used high doping (5×10^{18} cm^{-3}), while the drain side was left undoped (lightly doped) with silicon (1×10^{15} cm^{-3}). Hence, the potential in these two different regions was also expected to be different. To solve the potential in the two different regions, Poisson's equation was solved for both regions.

Model derivation. The results section was divided into two parts. Part A was analog performance and part B was RF performance of the proposed GGC-SGT-SB-MOSFET.

Surface potential modelling. The silicon under the gate was assumed to be fully depleted. Channel potential including channel mobile carrier density was governed by Poisson's equation in two different regions and is given as:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\Phi_j\left(r,z\right)\right)+\frac{\partial^2}{\partial z^2}\Phi_j\left(r,z\right)=\frac{qN_{AH/AL}}{\varepsilon_{Si}} \quad (1)$$

Here, $\Phi_i(r,z)$ is the potential distribution in the silicon film, q is the electron charge, ε_{Si} is the permittivity of Silicon, i = 1 and 2 shows the undamaged (region-1) and damaged (region-2) regions, respectively. $N_{AH/AL}$ is the doping of silicon film.

Due to the different doping in the two different regions through the channel, the flat band voltage (V_{FBi}) was different in the two different regions (regions 1 and region 2) and were written as:

$$V_{FB1} = \Phi M_1 - \Phi_{Si1}; \ i = 1, \ 0 \le z \le L_1 \tag{2}$$

$$V_{FB2} = \Phi M_2 - \Phi_{Si2}; \ i = 2, \ L_1 \le z \le (L_1 + L_2) \tag{3}$$

where, Φ_{Si1} and Φ_{Si2} were the semiconductor work function in the two different regions and could be calculated using the generalised equation (Equation 5), and C_{ox} was the gate oxide capacitance per unit area of GME-SB-CGAA MOSFET and was given as:

$$C_{ox} = 2\varepsilon_{ox}\left(t_{Si} \ln\left(1 + \frac{2t_{ox}}{t_{Si}}\right)\right)^{-1} \tag{4}$$

and

$$\Phi_{Si} = \chi + (E_g / 2) - q\Phi_F \tag{5}$$

where, χ was the electron affinity of the silicon, E_g was the energy band gap and Φ_F was the Fermi potential.

By using the super position technique, the solution of electrostatic potential $\psi_i(r,z)$ could be decomposed in two components, the long channel solution $V_i(r)$ and short channel solution $U_i(r,z)$, where $V_i(r)$ and $U_i(r,z)$ were the solution of the 1D Poisson equation and 2D Laplace equation, respectively. Then, the electrostatic potential in SB-CGAA MOSFET could be written as:

$$\Psi_i(r,z) = V_i(r) + U_i(r,z) \tag{6}$$

Equation 1 was then decomposed to Equations 7 and 8 as:

$$\frac{1}{r}\frac{\partial V_i(r)}{\partial r} + \frac{\partial^2 V_i(r)}{\partial r^2} = \frac{qN_a}{\varepsilon_{si}} \tag{7}$$

$$\frac{1}{r}\frac{\partial U_i(r,z)}{\partial r} + \frac{\partial^2 U_i(r,z)}{\partial r^2} + \frac{\partial^2 U_i(r,z)}{\partial z^2} = 0 \tag{8}$$

where, $i = 1$ for the high-doped region 1 and $i = 2$ for the low-doped region. The 1D potential $V_i(r)$ could be obtained by solving Poisson's equation with the following boundary conditions:

(a) Potential at the centre of the silicon pillar was:
$$\psi_c \ \Psi_i(r = 0, z) = \Psi_c(z) \tag{9a}$$

(b) Electric flux at the centre of the silicon pillar was zero:
$$\frac{\partial}{\partial r}\Psi_i(r = 0) = 0 \tag{9b}$$

(c) Potential at the surface of the silicon pillar was:

$$\Psi_i(r,z)\Big|_{r=\frac{t_{Si}}{2}} = \Psi_s(r = \frac{t_{Si}}{2}, z) \tag{9c}$$

(d) Electric flux at the surface of the silicon pillar was continuous:

$$\frac{\partial}{\partial r}\Psi_i(r,z)\Big|_{r=\frac{t_{Si}}{2}} = \frac{C_{ox}}{\varepsilon_{Si}}\left[V_{GS} - V_{FBi} - \Psi_i(r = \frac{t_{Si}}{2}, z)\right] \tag{9d}$$

The solution of the 1D Poisson equation (Equation 7) using the boundary conditions given in Equations 9a-9d can be expressed as:

$$V_1(r) = \frac{qN_{A/H}}{4\varepsilon_{Si}}r^2 + V_{GS} - [\phi_M - \phi_{Si1}] - \frac{qN_{A/H}t_{Si}^2}{16\varepsilon_{Si}} - \frac{qN_{A/H}t_{Si}}{4\beta\varepsilon_{Si}} \tag{10a}$$

and

$$V_2(r) = \frac{qN_{A/L}}{4\varepsilon_{Si}}r^2 + V_{GS} - [\phi_M - \phi_{Si2}] - \frac{qN_{A/L}t_{Si}^2}{16\varepsilon_{Si}} - \frac{qN_{A/L}t_{Si}}{4\beta\varepsilon_{Si}} \tag{10b}$$

Here, V_{GS} was the gate-to-source voltage; t_{si} was the diameter of the silicon pillar and $\beta = C_{ox}/\varepsilon_{Si}$. The solution of the 2D Laplace equation (Equation 8) was obtained using the following boundary conditions:

(a) At the interface of two different metal gates, potential is continuous.

$$\Psi_1(r,z)\Big|_{z=L_0} = \Psi_2(r,z)\Big|_{z=L_0} \tag{11a}$$

(b) At the interface of M_1 and M_2, the electric field is continuous.

$$\frac{\partial}{\partial r}\Psi_1(r,z)\Big|_{z=L_0} = \frac{\partial}{\partial r}\Psi_2(r,z)\Big|_{z=L_0} \tag{11b}$$

(a) Potential at the source end could be taken as:
$$\Psi_i(r, z = 0) = V_{bi,S/D} - \Delta\Phi_{bS} \tag{11c}$$

(c) Potential at the drain end could be taken as:
$$\Psi_i(r, z = L_0 + L_1) = V_{bi,S/D} - \Delta\Phi_{bD} + V_{DS} \tag{11d}$$

Here, $V_{bi,S/D}$ was the built-in potential at source/drain and $\Delta\Phi_{bS}$ and $\Delta\Phi_{bD}$ were SB lowering at source and drain, respectively (Crowell & Rideout, 1969; Kumar et al., 2016b). It could be expressed as:

$$V_{bi,S/D} = \Phi_{Si} - \Phi_{S/D} \tag{12}$$

$$\Delta\Phi_{bS} = \alpha\left[\frac{\tau q}{4\pi\varepsilon_S}\right]^{\frac{1}{2}} E_S^{\frac{1}{2}} \tag{13a}$$

$$\Delta\Phi_{bD} = \alpha\left[\frac{\tau q}{4\pi\varepsilon_S}\right]^{\frac{1}{2}} E_D^{\frac{1}{2}} \tag{13b}$$

Here, χ was the electron affinity of silicon, ε_{Si} was the silicon permittivity and E_S and E_D were the electric field at the source and drain end, respectively. α and τ were treated as a fitting parameter ($\alpha =1$, $\tau = 1.1$).

The solution of Equation 8 using the above boundary conditions can be expressed as:

$$U_i\left(r,z\right) = \sum_{n=1}^{\infty} J_0\left(\lambda_n r\right)\left(A_i e^{\lambda_n z} + B_i e^{-\lambda_n z}\right) \tag{14}$$

A_1, A_2, B_1 and B_2 are the Bessel-Fourier series coefficients and are given in the Appendix. Υ_n is the Eigen value satisfying the following equation:

$$C_{ox} J_0\left(\Upsilon_n \frac{t_{Si}}{2}\right) - \varepsilon_{si} \Upsilon_n J_1\left(\Upsilon_n \frac{t_{Si}}{2}\right) = 0 \tag{15}$$

J_0 and J_1 are the Bessel function of order zero and one, respectively.

Subthreshold current modelling. The total subthreshold current was due to the thermionic current and the tunneling currents and are given as:

$$I_{Thermionic} = \pi t_{Si} \mu k T n_i \frac{\left(1 - e^{\left(\frac{-V_{DS} q}{kT}\right)}\right)}{\left[\int_0^L \left[\frac{1}{\int_0^R e^{\left[\frac{\left(\Psi\left(r,V_{DS},V_{GS},z\right)\right)}{kT}\right]} dr}\right] dz\right]} \tag{16}$$

and

$$I_{Tunneling} = \pi t_{Si} A^* T^2 e^{\left(\frac{-\phi_b}{E_{00}(\phi_b)}\right)} \left(e^{\left(\frac{V_{gs} q}{nkT}\right)} - 1\right) \tag{17}$$

Here, μ (1300 cm^2/Vs) was the mobility, n_i (1.45×10^{10} cm^{-3}) was the intrinsic carrier concentration, A^* was the Richardson constant, ϕ_b was the barrier height, and E_{00} was the constant of the material associated with the WKB expression and was given as (Balaguer, Iniguez, & Roldan, 2011):

$$A^* = \frac{4\pi q m^* k^2}{h^3} \tag{18}$$

$$E_{00} = \alpha(\Phi_b) q \frac{\hbar}{2} \sqrt{\frac{N_A}{m^* \varepsilon_{Si}}} \tag{19}$$

$$\alpha(\Phi_b) = G + H\Phi_b \tag{20a}$$

$$\hbar = h / 2\pi \tag{20b}$$

Here, m^* was the effective mass of the electron, k was the Boltzmann constant and h was the Plank constant. G and H were the fitting parameters, with the values 21.8 (1/ eV) and 36.6 (1/ eV) (Balaguer et al., 2011).

The total subthreshold current was the combination of both $I_{thermionic}$ and $I_{tunneling}$ (Balaguer et al., 2011) and was evaluated as:

$$I_{Sub\_total} = \frac{(I_{thermionic} I_{Tunneling})}{(I_{thermionic} + I_{Tunneling})} \tag{21}$$

RF performance. Cut-off frequency (fT) was the important parameter for evaluating the RF performance of the device and it was a good candidate for high-frequency switching applications.

RESULTS

The surface potential of any device is a characteristic by which we can calculate the threshold voltage of the device. The surface potential at no applied bias is shown in Figure 2. It can be observed that the surface potential for the proposed SB-GC MOSFET device shifts upwards and the minimum of the surface potential lies under the highly-doped region of the MOSFET. The upward shift of the surface potential indicated that the threshold voltage of the graded channel device would be lower compared to that of the uniformly-doped device. The analytical result presented in the figure for the proposed device was in good agreement with the simulation results obtained through the ATLAS-3D numerical simulation tool. The surface potential was also calculated for applied bias and it was observed that the minimum surface potential still lay in the highly-doped region. The analytical and simulation results were matched.

DISCUSSION

The electric field for both the devices is shown in Figures 4(a) and 4(b) at no bias applied and at applied bias, respectively. From Figure 4(a) and 4(b) it is evident that the electric field towards the drain end was lowered for the Graded Channel Schottky Barrier GAA MOSFET. This means that the reduced electric field resulted in reduced short channel effects. This was the result of lower carrier mobility degradation.

The subthreshold drain current for the Graded Channel Schottky Barrier Gate All Around MOSFET is shown in Figure 5. The analytical model results were in good agreement with the simulation results. The figure also depicts that the on current of the graded channel was drastically increased by 47%; this is desirable in scaled MOSFET devices. Due to the lower doping towards the drain side, carrier mobility was enhanced compared with that of the uniformly-doped channel SB-MOSFET. Consequently, the on-state current increased. This was also mainly due to the reduced resistance in the undoped part of the channel region.

However, one tradeoff observed in the Graded Channel SB-GAA MOSFET was that the off-state current was increased, reducing the I_{on}/I_{off} ratio of the device. This indicated that the device would be better used in analog applications and would degrade performance compared to SB-GAA MOSFET for digital applications. In the triode region of operation, the Graded Channel MOSFET had higher current than the conventional MOSFET. This was due to the slightly higher surface electron concentration as well as to higher mobility in the low-doped region. In addition, current density was concentrated near the surface of the high-doped part of the channel but spread inside the low-doped region, contributing to reducing the resistance associated with the low-doped region of the MOSFET.

The high on-current resulted in higher transconductance; thus, the high cut-off frequency for the Graded Channel SB-GAA MOSFET was expected. Figure 6 shows the simulated transconductance for both the devices and it clearly depicts that the graded channel device consisted of high transconductance. It is clear from the figure that the GC-SB-GAA MOSFET showed higher transconductance than the SB-GAA MOSFET. The improvement in transconductance was directly related to the improvement in the low-frequency open-loop gain of the device and can be expressed as $A_V=(g_m/I_{DS})V_{EA}$. The enhanced transconductance of the GC-SB-GAA MOSFET can provide very high gain and can be used in operational amplifiers. Figure 7 illustrates the variation of cut-off frequency with gate voltage for both the devices and it was clearly observed that the peak cut-off frequency of the GC-SB-GAA architecture was 260 GHz, while for SB-GAA, it was 196 GHz, respectively. This clearly shows that the graded channel architecture improved the peak cut-off frequency by 32%. It was expected that the increased transconductance would increase the cut-off frequncy, as can be seen in Figure 7.
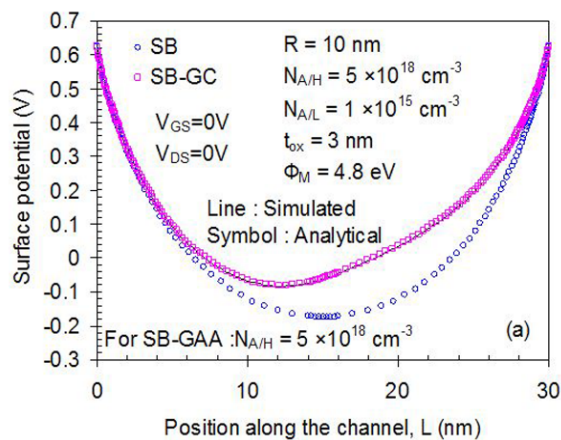
Figure 2. Variation of surface potential for both SB-GAA and GC-SB-GAA MOSFET at no applied bias
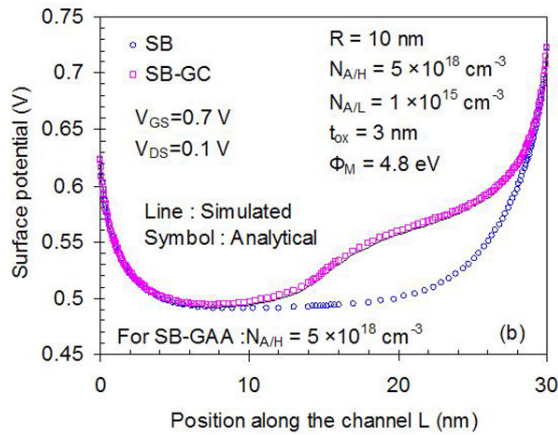
Figure 3. Variation of surface potential for both SB-GAA and GC-SB-GAA MOSFET at $V_{GS} = 0.7V$ and $V_{DS} = 0.1V$
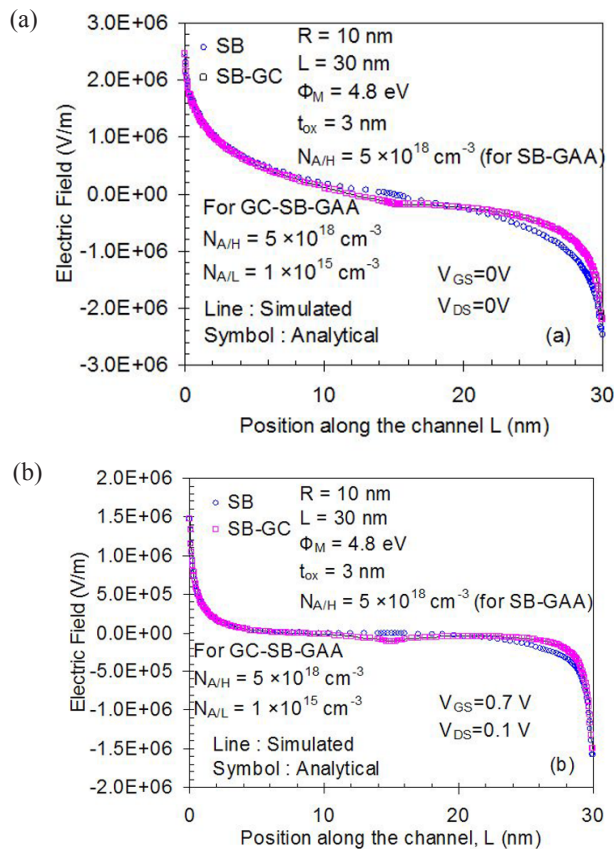
Figure 4. Variation of electric field: (a) at no applied bias; (b) at $V_{GS} = 0.7V$ and; $V_{DS} = 0.1V$ for SB-GAA and GC-SB-GAA MOSFET
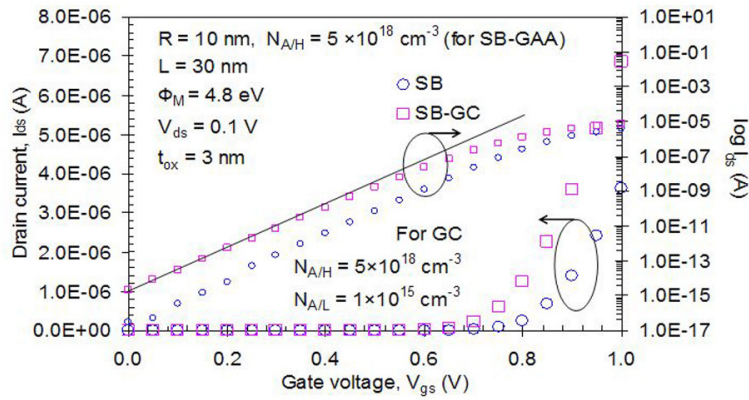
Figure 5. Variation of drain current (I_{ds}) at normal scale versus gate voltage V_{GS} at a fixed drain bias of 0.1 V
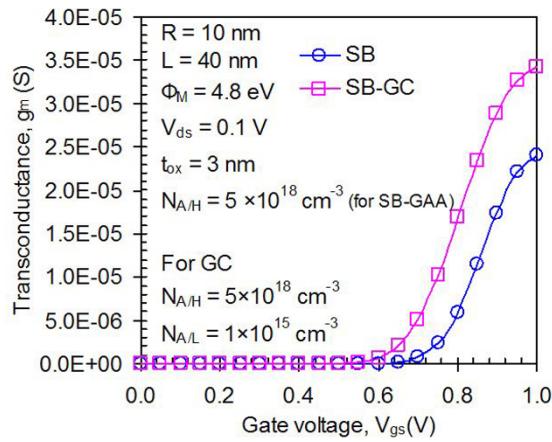
Figure 6. Variation of transconductance with applied gate voltage (V_{gs}) at a fixed V_{DS} of 0.1 V
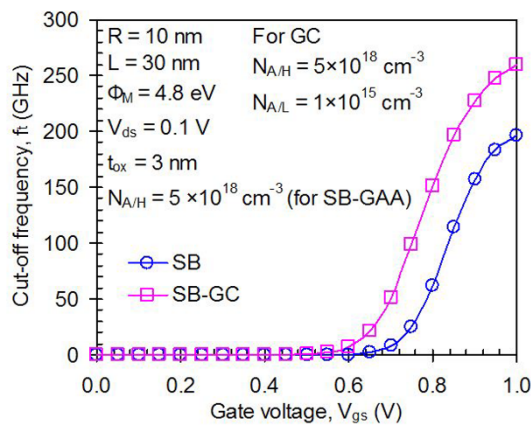
Figure 7. Variation of cut-off frequency as a function of the applied gate bias V_{gs}

CONCLUSION

The Graded Channel (GC) Schottky Barrier Gate All Around MOSFET offered better on-state performance compared to the conventional highly-doped Schottky Barrier Gate All Around MOSFET in terms of I_{on}, g_m and high cut-off frequency. The on-current was improved by 47%, resulting in higher transconductance. The cut-off frequency of the device was improved by 32%. The developed analytical model using the superposition technique showed good agreement with the ATLAS-3D simulation results. So, it was concluded that the Graded Channel (GC) Schottky Barrier Gate All Around MOSFET was a strong contender for low-power high performance devices applicable in future USLI (Ultra Large Scale Integration) applications.

ACKNOWLEDGEMENT

REFERENCES

ATLAS User's Manual. (2013). *3D Device Simulator, Silvaco Inc., Santa Clara, CA, USA*. Retrieved from the Silvaco website: https://www.silvaco.com/products/vwf/atlas/device3d/device3d_br.html

Balaguer, M., Iniguez, B., & Roldan, J. B. (2011). An analytical compact model for Schottky-barrier double gate MOSFETs. *Solid State Electronics, 64*(1)*,* 78–84. https://doi.org/10.1016/j.sse.2011.06.045

Cerdeira, A., Aleman, M. A., Pavanello, M. A., Martino, J. A., Vancaillie, L., & Flandre, D. (2005). Advantages of the graded-channel SOI FD MOSFET for application as a quasi-linear resistor. *IEEE Transactions on Electron Devices*, *52*(5), 967–972. doi: 10.1109/TED.2005.846327

Chiang, T. K. (2005). A compact analytical two-dimensional threshold voltage model for cylindrical, fully depleted, surrounding-gate (SG) MOSFETs. *Semiconductor Science and Technology, 20*(12)*,* 1173–1178. doi: 10.1109/EDSSC.2005.1635330

Colinge, J. P., Lee, C. W., Afzalian, A., Akhavan, N. D., Yan, R., Ferain, I., … & Murphy, R. (2010). Nanowire transistors without junctions. *Nature Nanotechnology, 5*(3), 225–229. doi: 10.1038/nnano.2010.15

Crowell, C. R., & Rideout, V. L. (1969). Normalized thermionic-field (TF) emission in metal-semiconductor (Schottky) barriers. *Solid-State Electronics, 12*(2)*,* 89–105. https://doi.org/10.1016/0038-1101(69)90117-8

Kaur, H., Kabra, S., Haldar, S., & Gupta, R. S. (2008). An analytical threshold voltage model for graded channel asymmetric gate stack (GCASYMGAS) surrounding gate MOSFET. *Solid State Electronics, 52*(2)*,* 305–311. doi: 10.1016/j.sse.2007.09.006

Kumar, M., Haldar, S., Gupta, M., & Gupta, R. S. (2013). A comparative assessment of Schottky-barrier source/drain GAA MOSFET with conventional and junctionless GAA MOSFETs. *IEEE Indian Conference (INDICON)* (pp. 1–6). Indian Institute of Technology, Bombay: India. doi: 10.1109/INDCON.2013.6726127

Kumar, M., Haldar, S., Gupta, M., & Gupta, R. S. (2014a). Impact of gate material engineering (GME) on analog/rf performance of nanowire Schottky-barrier gate all around (GAA) MOSFET for low power wireless applications: 3D T-CAD simulation. *Microelectronics Journal, 45*(11)*,* 1508–1514. doi: 10.1016/j.mejo.2014.07.010

Kumar, M., Haldar, S., Gupta, M., & Gupta, R. S. (2014b). A new t-shaped source/drain extension (T-SSDE) gate underlap GAA MOSFET with enhanced subthreshold analog/rf performance for low Power applications. *Solid State Electronics Journal, 101,* 13–17.

doi: 10.1016/j.mejo.2014.07.010

Kumar, M., Haldar, S., Gupta, M., & Gupta, R. S. (2016a, November 4–5). Gaussian graded channel (GGC) Schottky-barrier surrounding gate MOSFET for analog/rf performance enhancement. *ICANN 2016*. Retrieved from http://www.icannjmi.in/index.php

Kumar, M., Haldar, S., Gupta, M., & Gupta, R. S. (2016b). Physics-based analytical model for surface potential and subthreshold current of cylindrical Schottky-barrier gate all around MOSFET with high-k gate stack. *Superlattices and Microstructures, 90,* 215–226. doi: 10.1016/j.spmi.2015.12.029

Lee, C., Zhuang, Y., Di, S., & Han, R. (2013). Subthreshold behavior models for nanoscale short-channel junctionless cylindrical surrounding-gate MOSFETs. *IEEE Electron Device Letters, 60*(11), 3655–3662. doi: 10.1109/TED.2013.2281395

Li, Y., & Hwang, C. H. (2009). The effect of the geometry aspect ratio on the silicon ellipse-shaped surrounding – Gate field-effect transistor and circuit. *Semiconductor Science and Technology, 24*(9), 095018. doi:10.1088/0268-1242/24/9/095018.53

Marchi, M. D., Sacchetto, V., Zhang, J., Frache, S., Gaillardon, P. E., Leblebici, Y., & Micheli, G. D. (2014). Top-Down fabrication of gate-all-around vertically stacked silicon nanowire FETs with controllable polarity. *IEEE Transactions on Nanotechnology, 13*(6), 1029–1038. doi: 10.1109/TNANO.2014.2363386

Pavanello, M. A., Martino, J. A., Dessard, V., & Flandre, D. (2000). Analog performance and application of graded-channel fully depleted SOI MOSFETs. *Solid State Electronics, 44*(7)*,* 1219–1222. doi: 10.1016/S0038-1101(00)00034-4

Sharma, D., & Vishvakarma, S. K. (2012). Precise analytical model for short channel cylindrical gate (CylG) vgate-al-around (GAA) MOSFET. *Solid State Electronics, 86,* 68–74. doi: org/10.1016/j.sse.2012.08.006

Trivedi, N., Kumar, M., Haldar, S., Deswal, S. S., Gupta, M., & Gupta, R. S. (2016). Analytical modeling of junctionless accumulaion mode cylindrical surrounding gate MOSFET. *International Journal of Numerical Modeling Electronic Networks Devices and Fields, 29*(6), 1036–1043. doi: 10.1002/jnm.2162

Tsormpatzoglou, A., Tassis, D. H., Dimitriadis, C. A., Ghibaudo, G., Pananakakis, G., & Clerc, R. (2009). A compact drain current model of short-channel cylindrical gate-all-around MOSFETs. *Semiconductor Science and Technology, 24*(7)*,* 075017. doi: 10.1088/0268-1242/24/7/075017

Wang, T., Lou, L., & Lee, C. (2013) A junctionless gate-all-around silicon nanowire FET of high linearity and its potential applications. *IEEE Electron Devices Letter, 34*(4), 478–480. doi: 10.1109/LED.2013.2244056

Yang, W., Whang, S., Lee, S., Zhu, H., Gu, H., & Cho, B. (2007). Schottky-Barrier Si nanowire MOSFET: Effects of source/drain metals and gate dielectrics. *Materials Research Society Symposium Proceedings* (pp. 133–138). doi: 10.1557/PROC-1017-DD14-05

Zhuge, J., Wang, R., Huang, R., Xing, Z., & Wang, Y. B. (2008). Investigation of parasitic effects and design optimization in silicon nanowire MOSFETs for rf applications. *IEEE Transactions on Electron Devices*, *55*(8), 2142–2147. doi: 10.1109/TED.2008.926279

APPENDIX

The coefficients of Equations 10-15 are:

$$m = \frac{t_{Si}}{2} \ , \kappa = V_{GS} - \frac{qN_A t_{Si}^2}{16\varepsilon_{Si}} - \frac{qN_A t_{Si}}{4\varepsilon_{Si}\beta} \ ,$$

$$X = \frac{qt_{Si}^2}{16\varepsilon_{Si}}(N_{A/H} - N_{A/L}) + \frac{qt_{Si}}{4\varepsilon_{Si}\beta}(N_{A/H} - N_{A/L}) + \Phi_{Si1} - \Phi_{Si2} + \frac{qr^2}{4\varepsilon_{Si}}(N_{A/L} - N_{A/H})$$

$$P = \frac{V_{biS}mJ_1(\lambda_{n1}m)}{\lambda_n} - \frac{mJ_1(\lambda_n m)}{\lambda_n}\left\{V_{GS} - (\Phi_M - \Phi_{Si1}) - \frac{qN_{A/H}t_{Si}^2}{16\varepsilon_{Si}} - \frac{qN_{A/H}t_{Si}}{4\varepsilon_{Si}\beta} + \frac{qN_{A/H}m^2}{4\varepsilon_{Si}}\right\} + \frac{qN_{A/H}m^2 J_2(\lambda_n m)}{2\varepsilon_{Si}\lambda_n^2}$$

$$Q = \frac{(V_{biS} + V_{DS})mJ_1(\lambda_n m)}{\lambda_n} - \frac{mJ_1(\lambda_n m)}{\lambda_n}\left\{V_{GS} - (\Phi_M - \Phi_{Si2}) - \frac{qN_{A/L}t_{Si}^2}{16\varepsilon_{Si}} - \frac{qN_{A/L}t_{Si}}{4\varepsilon_{Si}\beta} + \frac{qN_{A/L}m^2}{4\varepsilon_{Si}}\right\} + \frac{qN_{A/L}m^2 J_2(\lambda_n m)}{2\varepsilon_{Si}\lambda_n^2}$$

$$B_2 = -\sinh^{-1}(\lambda(L_1 + L_2))\left[\frac{2}{m^2 J_1^2(\lambda m)}[Q - Pe^{\lambda(L_1+L_2)}] + \frac{2X\cosh(\lambda L_1)e^{\lambda(L_1+L_2)}}{\lambda m J_1(\lambda m)}\right]$$

$$B_1 = \frac{X}{e^{-\lambda L_1}m\lambda J_1(\lambda m)} - \sinh^{-1}(\lambda(L_1 + L_2))\left[\frac{2}{m^2 J_1^2(\lambda m)}[Q - Pe^{\lambda(L_1+L_2)}] + \frac{2X\cosh(\lambda L_1)e^{\lambda(L_1+L_2)}}{\lambda m J_1(\lambda m)}\right]$$

$$A_1 = \frac{2P}{m^2 J_1^2(\lambda m)} - \frac{X}{e^{-\lambda L_1}m\lambda J_1(\lambda m)} + \sinh^{-1}(\lambda(L_1 + L_2))\left[\frac{2}{m^2 J_1^2(\lambda m)}[Q - Pe^{\lambda(L_1+L_2)}] + \frac{2X\cosh(\lambda L_1)e^{\lambda(L_1+L_2)}}{\lambda m J_1(\lambda m)}\right]$$

$$A_2 = A_1 - \frac{X}{e^{\lambda L_1}m\lambda J_1(\lambda m)}$$

Case Study

A rare case of Subglottic Ectopic Thyroid Tissue Causing Upper Airway Obstruction: Detection by CECT and MRI Neck

Ngah, N. A.[1], SitiJusnaMuhammad[1], Suraini, M. S.[2], Mohd Noh, M. S. F.[2], AbHamid, S.[2], Salleh, H.[3] and Suppiah, S.[2,4]*

[1]*Department of Diagnostic Imaging, Hospital Serdang Malaysia, Jalan Puchong, 43000 Kajang, Selangor, Malaysia*

[2]*Department of Imaging, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

[3]*Department of Otorhinolaryngology, Hospital Serdang Malaysia, Jalan Puchong, 43000 Kajang, Selangor, Malaysia*

[4]*Centre for Nuclear Diagnostic Imaging, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia*

ABSTRACT

Internal obstruction of the upper airways can be due to infection, anaphylactic reaction, congenital anomaly, foreign body inhalation or mass. This case report is of a lady who presented with upper airway obstruction due to subglottic ectopic thyroid tissue. She had a history of noisy breathing and progressive neck swelling since childhood. Biochemical results showed hypothyroid features while flexible scope showed tracheal stenosis. Computed Tomography (CT) detected a goitre and a posterior laryngeal mass at subglottic region causing laryngeal stenosis. The MRI helped to better delineate the soft tissue details of this mass. while elective direct laryngoscope showed a firm mass arising from the trachea. Debulking of the mass was done and histopathological diagnosis was nodular thyroid hyperplasia. Treatment with L-thyroxine was instituted with good clinical improvement. An ectopic thyroid gland can develop if its normal migration is halted along this tract during embryogenesis. Subglottic location of ectopic thyroid is extremely rare. However, ectopic thyroid tissue in the larynx should be considered as a possible diagnosis causing upper airway obstruction, especially in a patient that is clinically hypothyroid and having a concurrent goitre.

Keywords: Ectopic thyroid, stridor, upper airway obstruction

INTRODUCTION

Upper airway obstruction can be due to internal or external compression of the trachea. Internal obstruction can be a result of an infection, anaphylactic reaction, congenital anomaly, foreign body inhalation or mass. We report a case of a young lady who presented with upper airway obstruction due to subglottic ectopic thyroid tissue.

An 18-year-old lady with no known comorbid medical history, presented with difficulty in breathing. She had a history of noisy breathing and neck swelling, since childhood, which was progressively worsening. On examination, she had stable vital signs, normal nutritional status, and presence of an anterior neck mass which moved with swallowing and occasional stridor. She was however, not in acute respiratory distress. Thyroid function test confirmed she had hypothyroid condition. Flexible laryngeal scope showed tracheal stenosis. Elective tracheostomy insertion was performed for prophylactic airway protection. Contrast-enhanced Computed Tomography (CECT) of the neck showed the presence of enlarged thyroid gland and also a posterior laryngeal mass at the subglottic region causing laryngeal stenosis. Magnetic Resonance Imaging (MRI) delineated the well-defined subglottic mass and its location accurately. The MRI also demonstrated that the mass had a similar signal intensity to the normally placed thyroid gland; i.e. hyperintense signal on T1-weighted and T2-weighted images as well as avid enhancement in post-gadolinium images.

Elective direct laryngoscope showed a firm mass arising from within the trachea, having no evidence of invasion of the wall of the larynx and trachea. Debulking of the mass was done and histopathological diagnosis confirmed the lesion to have cells of nodular thyroid hyperplasia in origin. The patient was started on L-thyroxine and showed clinical improvement with progressive resolution of her stridor.
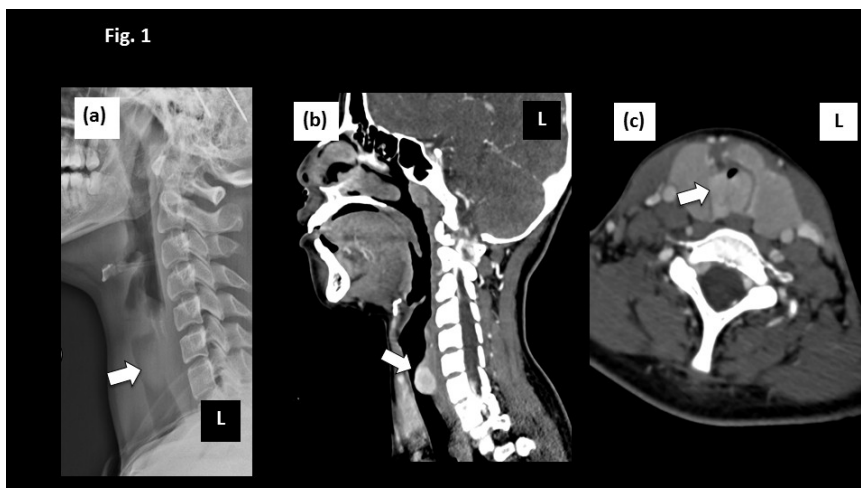
Figure 1: (a) Plain radiograph of the neck in lateral projection showed a lobulated soft tissue opacity within the larynx (white arrow). (b) CT scan of the neck in sagittal view showed a well-defined, enhanced subglottic lesion (white arrow) arising from the posterior laryngeal wall. (c) CT scan of the neck in axial view, showed a hypertrophied thyroid gland and homogenously enhancing, intra-laryngeal lesion (white arrow) causing almost total occlusion of the upper airway
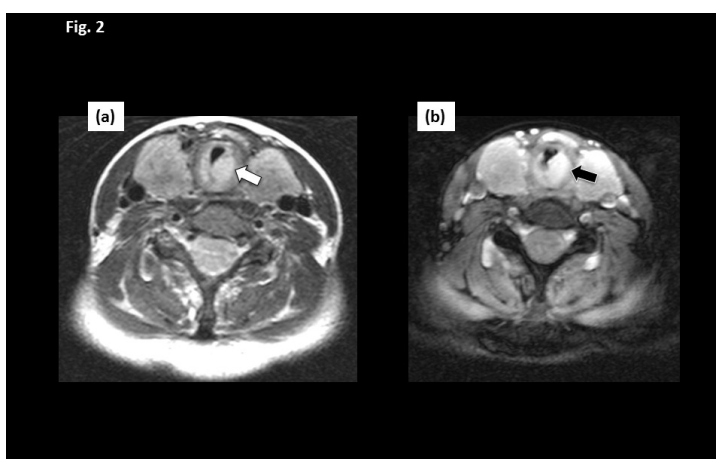
Figure 2: (a) MRI of the neck on T1-weighted images showed a well-defined intra-laryngeal hyperintense lesion (white arrow) causing occlusion of the airway. The native thyroid gland also appeared enlarged. (b) MRI of T1-weighted post gadolinium images showed avid enhancement of the intra-laryngeal lesion (black arrow) and also enhancement of the thyroid gland

DISCUSSION

The ectopic thyroid gland is an aberrant development of the gland during embryogenesis, whereby its migration is arrested somewhere along the tract of its normal migration from the floor of the primitive foregut to its final pre-tracheal position. The most frequent location of ectopic thyroid tissue is at the base of the tongue, accounting for about 90% of the reported cases (Noussios, Anagnostis, Goulis, Lappas, & Natsis, 2011). Subglottic ectopic thyroid tissue is an extremely rare abnormality that can cause airway obstruction (Sung, Lee, Han, & Cho, 2008). Only 14 well documented cases of subglottic ectopic thyroid tissue have been reported in the literature since 1966 until 2005 (Ramalingam, Ramalingam, Dhote, & Murthy, 2005).

Two theories may explain the origin of subglottic thyroid tissue. The first theory is that a foetal anomaly could occur when the thyroid is split by the developing trachea and its cartilage rings. The second is that thyroid tissue could rise into the tracheal lumen (Ammor et al., 2014). In CT scans performed without contrast material administration, the ectopic thyroid tissue demonstrates mildly increased attenuation (HU 70 ± 10); relative to adjacent muscle - a finding that is due to the intrinsic iodine content in the thyroid gland. On T1-weighted MRI images, ectopic thyroid tissue is typically isointense to mildly hyperintense compared with the musculature, and on T2-weighted MRI images, ectopic thyroid tissue is mildly hyperintense, with variable enhancement after administration of a gadolinium-based contrast agent (Zander & Smoker, 2014).

Different diagnoses for this type of mass lesion among others include metastasis from thyroid cancer, angioma, fibroma, lymphangioma, and thyroglossal duct cyst. By utilising MRI, we were able to demonstrate that the lesion was solid and not cystic in nature. The intrinsic signals of this lesion were also similar to the thyroid gland at the thyroid fossa, thus making the diagnosis of an ectopically located thyroid nodule more likely. Furthermore, MRI provided

excellent soft tissue details that aided in the planning of surgery. It also helped to exclude the presence of cervical lymphadenopathy, thus allaying fears of a more aggressive aetiology. Additional imaging techniques that could be utilised include Technetium 99m pertechnetate or Iodine-131 thyroid scintigraphy which can give functional information to differentiate an ectopic thyroid from other causes of midline neck masses. (Noussios et al, 2011). In certain instances, 18F-Fluorodeoxyglucose positron emission tomography/ computed tomography (18F-FDG PET/CT) was utilised to differentiate suspicious next masses especially involving patients with known primary cancers, although this method is non-specific for diagnosis of ectopic thyroid tissue and may mislead the clinician into making a diagnosis of a metastatic lesion as it can give abnormally elevated maximum standardised uptake value as high as 8.0 g/mL (Kim, 2013).

Our patient was clinically hypothyroid and therefore, developed a goitre as well as hyperplasia of an ectopically located thyroid tissue. This hormonal imbalance that leads to hyperplasia of the ectopic thyroid was the cause of the patient's progressive difficulty in breathing. Contrast-enhanced CT scan is a useful modality to investigate causes of difficulty in breathing as it can help to accurately detect upper airway, lung or mediastinal masses (Balakrishnan, Suppiah, Md Sidek & Noriah, 2015) as the causative factors.

A biopsy is required in order to confirm the diagnosis of ectopic thyroid; however, it can cause severe bleeding as the thyroid gland is a vascular organ (Dowling, Johnson & Collier, 1962). Therefore, non-invasive investigations such as CT scan and MRI play an essential role to determine the diagnosis. In this case, CT scan of the neck and laryngoscopic examination helped detect and characterise the mass at the subglottic region. The patient was operated upon and a reddish, firm and vascular but non-pulsatile mass, suspicious of thyroid tissue, was excised. This mass was confirmed to be of thyroid in origin by histopathological examination, which is the gold standard for diagnosis.

CONCLUSION

Subglottic ectopic thyroid tissue should be considered as a possible diagnosis causing upper airway obstruction, especially in a patient that is clinically hypothyroid and has a concurrent goitre. Congenital airway mass and malignancy could be some of the different diagnoses. The CT scan and MRI of the neck should be considered as investigations of choice to locate the mass, characterise it and delineate its extent.

ACKNOWLEDGEMENT

REFERENCES

Ammor, H., Boujarnija, H., Lamrani, Y., Boubbou, M., Maâroufi, M., Kamaoui, I., & Tizniti, S. (2014). A rare cause of tracheal obstruction: Ectopic thyroid. *EuroRad Radiological Case Database European Society of Radiology*. Retrieved from http://www.eurorad.org/eurorad/case.php?id=11839pg.1doi: http:// 10.1594/EURORAD/CASE.11839.

Balakrishnan, D., Suppiah, S., Md Sidek, S., & Noriah, O. (2015). Primary mediastinal choriocarcinoma masquerading As Lung Metastasis: A rare disease with a fatal outcome. *International Journal of Public Health and Clinical Sciences*, *2*(6), 132-40.

Dowling, E. A., Johnson, I. M., & Collier, F. C. (1962). Intratrachealgoiter: a clinicopathologic review. *Annals of Surgery, 156*(2), 258 –67.

Kim, S. Y. (2013). A case of right paratracheal ectopic thyroid, mimicking metastasis on CT and 18F-FDG PET/CT. *Open Journal of Medical Imaging, 3*(3), 82-85.

Noussios, G., Anagnostis, P., Goulis, D. G., Lappas, D., & Natsis, K. (2011). Ectopic thyroid tissue: anatomical, clinical, and surgical implications of a rare entity. *European Journal of Endocrinology, 165*(3), 375-82.

Ramalingam, K. K., Ramalingam, R., Dhote, K., & Murthy, S. (2005). Ectopic thyroid: A rare cause of tracheal obstruction. *Indian Journal of Otolaryngology and Head and Neck Surgery, 57*(3), 252-255.

Sung, Y. M., Lee, K. S., Han, J., & Cho, E. Y. (2008). Intratrachealectopic thyroid tissue with adenomatous hyperplasia in a pregnant woman. *American Journal of Roentgenology, 190*(2), 161-163.

Zander, D. A., & Smoker, W. R. K. (2014) Imaging of ectopic thyroid tissue and thyroglossal duct cysts. *RadioGraphics, 34*(1), 37-50.

Case Study

An Uncommon Vascular Cause of Spontaneous Hydropneumothorax

Balakrishnan, D.[1], Suraini, M. S.[2], Hazman, M. N.[3], Hariati, J.[4], R, Mahmud.[2] and Ezamin, A. R.[2]*

[1]*Department of Radiology, Hospital Serdang Malaysia, Jalan Puchong, 43000 Kajang, Selangor, Malaysia*

[2]*Department of Imaging, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43600 UPM, Serdang, Selangor, Malaysia*

[3]*Department of Radiology, Prince Court Medical Center, 39, Jalan Kia Peng, Kuala Lumpur, 50450 Kuala Lumpur, Malaysia*

[4]*Deputy Director Office, Hospital Putrajaya, Presint 7, Putrajaya, Malaysia*

ABSTRACT

This paper highlights a potentially life threatening and unsuspected case of multiple small pseudoaneurysms from the thyrocervical and costocervical branch of the left subclavian artery that was spontaneously ruptured. The cause was not suspected prior to thoracic CT angiography (CTA) because the initial pre-emptive diagnosis was only trivial spontaneous hydropneumothorax which became serious after a chest tube was inserted.

Keywords: Hydropneumothorax, vascular

INTRODUCTION

Conventional CT thorax is a known method to diagnose non-vascular lung pathology, while CT angiography (CTA) is a special technique in image acquisition used mainly to visualize the vascular structure. In this study CTA was used to rule out any iatrogenic injury to the vessel post chest tube insertion. This method enabled us to promptly diagnose pseudoaneurysm as the cause of spontaneous hydropneumothorax (SHP). The Conventional CT thorax would probably miss the cause of SHP and further delay diagnosis and treatment.

SHP can be due to thoracic malignancies, vascular malformations, infections, coagulation disorders or cavitating pulmonary infarction (Suppiah & Abdullah, 2015). SHP has high mortality due to rapid blood loss and aerated lung volume loss. The incidence of SHP is up to 7% of patients with spontaneous pneumothorax (Ali, Lippmann, Mundathaje & Khaleeq G, 2008, Hsu, Shih, Hsu, & Chen, 2005). SHP usually occurs with the presence of illness or injury. Anomalous vascular cause for SHP is extremely rare. The aetiology can be varied but includes lung neoplasms, blood dyscrasia, arteriovenous malformation (Osler-Weber Rendu Disease), pregnancy, aortic dissection, pulmonary emboli, connective tissue disorders such as Ehlers-Danlos syndrome (EDS) type IV, endometriosis, adhesions with pneumothorax and infection (Hsu et al, 2005). SPH has a predilection among males with 100% male dominance. Smokers are thought to have a higher incidence of SHP accounting for 76.5% of patients (Kim, Kang, Pyo, Jeon, & Lee, 2008).

RESULT

A 21-year-old male, social smoker with no underlying medical illness was presented to the Emergency Department (ED) with a sudden onset of dyspnoea associated with epigastric pain radiating to the back. Upon examination, he had severe, left-side chest pain. No history of trauma or heavy coughing was identified. A Chest X-ray (CXR) (Figure 1A) confirmed a left hydropneumothorax with tracheal deviation to the right. An intercostal chest drain was inserted and approximately 1.7 litres of haemoserous fluid was drained over three hours.

The patient became progressively hypotensive with a haemoglobin level of 6.8 g/dL and Haematocrit of 19%. An urgent non-cardiac gated CT-angiography with 128-DECT (dual energy) CT scanner (Somatom Sensation Flash, Siemens Medical Solutions) was immediately performed to exclude any iatrogenic chest drain-related lung injury. The CTA demonstrated a left hydropneumothorax with an active contrast extravasation into the left pleural space that was suggestive of active bleeding (Figure 1B). The actual origin of the haemorrhage was not well delineated on the CTA. However, a well-defined lobulated arterial density structure was seen in the upper lobe suggesting pseudoaneurysm (Figure 2 A, 2B). No active contrast extravasation was noted around the chest drain site. The patient was transferred to the Angiography suite from red zone ED. The interventional radiology team swiftly proceeded with an angiography using the right femoral puncture approach. The left subclavian artery (SCA) digital subtraction images revealed a leaking pseudoaneurysm supplied from the thyrocervical branch of the left SCA (Figure 3A, B and C). The feeder vessel of the thyrocervical pseudoaneurysm branch was embolized using Polyvinyl Alcohol (PVA) particles, 250-355 μm size and a single 2.0 mm x 20.0 mm coil. The subsequent post-embolization run of the left SCA demonstrated another, smaller leaking pseudoaneurysm from the costocervical branch of left subclavian artery (Figure 3D), which was embolized with PVA particles 250-355 μm.

The post-embolisation subsequent angiogram showed complete resolution of contrast extravasation. The patient's condition steadily improved upon receiving a total of 4 pints of pack red blood cells (RBC) for resuscitation. Serial CXR's showed improvement of the left hemopneumothorax allowing the patient to be discharged a week later.
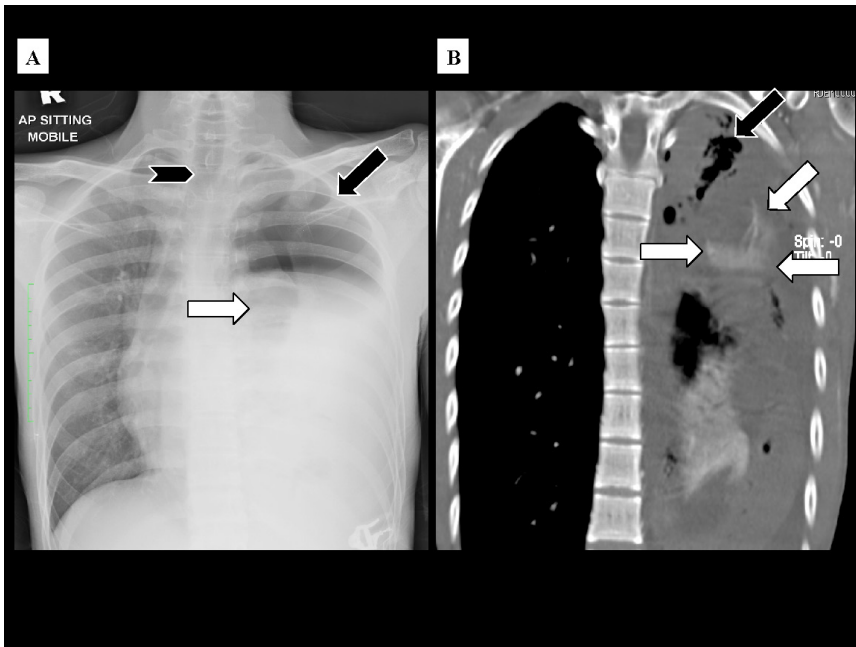
Figure 1. CXR in casualty showed left hydropneumothorax; pneumothorax (black arrow) with fluid level within (white arrow) and tracheal deviation to the right (notched arrow) (A). Coronal CT Thorax image in delayed images showing pooling of contrast in the left upper hemithorax (white arrow) and air pockets within the effusion confirming the pneumothorax component (black arrow) (B)

Figure 2. Axial CTA Thorax image in arterial phase showing the lobulated enhancing structure in apical region (black arrow) suggestive of a pseudoaneurysm (A). Coronal CT Thorax image in arterial phase showing active contrast extravasation in the left apical region (black arrow) (B)
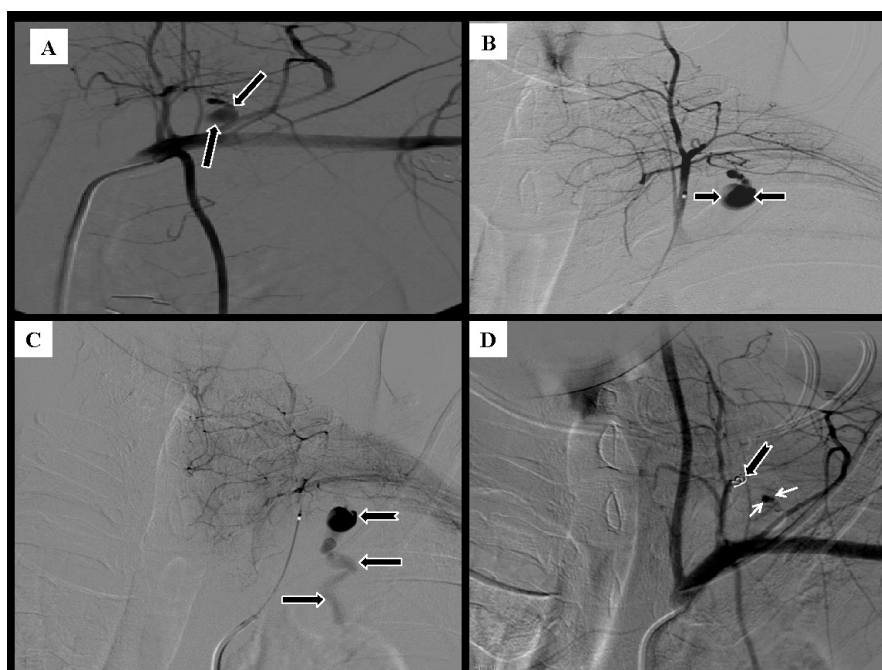
Figure 3. Left SCA angiogram showing pseudoaneurysm (black arrow) (A). Selective angiogram in left thyrocervical trunk confirms pseudoaneurysm (black arrow) (B). Pseudoaneurysm (notch arrow) with active extravasation of contrast into left hemithorax (black arrow) (C). Post-coiling run showing coil in situ (notch arrow) and left costocervical trunk pseudoaneurysm (thin white arrow) (D)

DISCUSSION

The cause of haemothorax is hypothesized due to a normal vessel having undergone chronic stress, due to abnormal vascular morphology or due to torn adhesion of the lung pleura or even rupturing of the luxury perfused bullae (Ali et al., 2008; Kanazawa, Yamazaki, Aoki, & Sakurai, 1996). Conditions such as coagulopathy, neoplasia, and endometriosis are also a few of the causes (Patrini et al., 2015). Mucoid degeneration within the aberrant vessel wall was also reported to be one of the causes of SHP (Tatebe et al., 1996). In our patient, there was no evidence the above-mentioned conditions, coagulopathy or trauma. In such circumstances, spontaneous rupture of intrathoracic blood vessels needs to be considered.

Despite the initial CXR, which demonstrated a hemopneumothorax, the bleeding source in our patient was identified during angiography arising from a pseudoaneurysm supplied by the thyrocervical and costocervical branches of the left Subclavian artery (SCA). The thyrocervical arterial branch that originates from SCA is well protected from blunt trauma due to its deep bodily location. It is most vulnerable during central venous catheterization. Our patient had no history of any previous attempts at central venous access.

There is a long list for the causes of pseudoaneurysm, such as direct trauma or indirect trauma to the vessel wall, disease vessel wall, vasculitides, dissections and even inflammation. Pseudoaneurysm happens when there is a defect in the vessel wall with an intact adventitia,

causing blood to leak through the defect, forming a sac that is only lined by the tunica adventitia. Due to its poor wall strength, the risk of rupture is significantly greater (Marx, Gardiner, & Miller, 1985).

An uncommon correlation between neurofibromatosis (NF) and arterial aneurysm due to wall fragility was also reported in patients with SHP and the incidence is approximately 3.6% (Salyer & Salyer, 1974). In regards of NF, the bleeding is probably either due to rupture of the abnormal vessels or bleeding schwannoma. Bleeding from the thyrocervical trunk in NF type 1 has been reported, although it is not common compared with bleeding from SCA and intercostal arteries which is more common in NF type 1 (Salyer & Salyer, 1974; Miura et al., 1997). Our patient had no features suggestive of NF.

Historically, surgery has been the main stay of treatment for SHP; however, the trend for less invasive methods such as the video-assisted thoracoscopic (VATS) approach has become popular. Endovascular treatment is also effective and its increasing availability in tertiary care centres all over the world has made it a good alternative. Coils embolization is the treatment of choice in the end vessels whereby the small artery feeding the pseudoaneurysm can be sacrificed. A covered stent is the best choice to control arterial bleeding in pseudoaneurysm that arises from the main parent artery. In our patient, the bleeding site was a ruptured pseudoaneurysm from the small thyrocervical branch and costocervical arteries, hence coil embolization was utilized.

CONCLUSION

SHP is potentially life threatening. It can result in massive blood loss, severe respiratory compromise and mortality if not promptly treated. A thoracotomy however is not always indicated, as has been the case in the past. A contemporary endovascular approach with coils embolization may be technically feasible being less invasive and safer than open surgery. However, had this patient continued to bleed post-embolization or if the latter failed, emergency surgical exploration would have been necessary. Similarly, surgery is more expedient for a hemodynamically unstable patient. Hence close collaboration between the interventional radiologist and cardiothoracic surgeon is vital to ensure optimal care.

ACKNOWLEDGEMENT

REFERENCES

Ali, H. A., Lippmann, M., Mundathaje, U., & Khaleeq, G. (2008). Spontaneous hemothorax: a comprehensive review. *Chest Journal*, *134*(5), 1056-65.

Hsu, N. Y., Shih, C. S., Hsu, C. P., & Chen, P. R. (2005). Spontaneous hemopneumothorax revisited: clinical approach and systemic review of the literature. *The Annals of Thoracic Surgery*, *80*(5), 1859-63.

Kim, E. S., Kang, J. Y., Pyo, C. H., Jeon, E. Y., & Lee, W. B. (2008). 12-year experience of spontaneous hemopneumothorax. *Annals of Thoracic and Cardiovascular Surgery*, *14*(3), 149-153.

Marx, M., Gardiner, G. A. Jr., & Miller, R. H. (1985). The truth about false aneurysms. *American Journal of Roentgenology*, *145*(1), 193-194.

Miura, H., Taira, O., Uchida, O., Usuda, J., Hirai, S., & Kato, H. (1997). Spontaneous haemothorax associated with von Recklinghausen's disease: review of occurrence in Japan. *Thorax*, *52*(6), 577-578.

Patrini, D., Panagiotopoulos, N., Pararajasingham, J., Gvinianidze, L., Iqbal, Y., & Lawrence D. R. (2015). Etiology and management of spontaneous haemothorax. *Journal of Thoracic Disease*, *7*(3), 520–526.

Salyer, W. R., & Salyer, D. C., (1974). The vascular lesions of neurofibromatosis. *Angiology*, *25*(8), 510-519.

Suppiah, S., & Abdullah, B. J. J. (2015). Spontaneous haemopneumothorax due to cavitating pulmonary infarction - a rare condition revisited. *International Journal of Public Health and Clinical Sciences*, *2*(2), 45 -54.

Tatebe, S., Kanazawa, H., Yamazaki, Y., Aoki, E., & Sakurai, Y. (1996). Spontaneous hemopneumothorax. *The Annals of thoracic surgery*, *62*(4), 1011-1015.

REFEREES FOR THE PERTANIKA
JOURNAL OF SCIENCE AND TECHNOLOGY

VOL. 26 (2) APR. 2018

The Editorial Board of the Journal of Science and Technology wishes to thank the following for acting as referees for manuscripts published in this issue of JST.

Pertanika Journals

Our goal is to bring high quality research to the widest possible audience

INSTRUCTIONS TO AUTHORS
(Manuscript Preparation & Submission Guide)

Revised: June 2016

MANUSCRIPT PREPARATION

Manuscript Types

Pertanika accepts submission of mainly **four** types of manuscripts for peer-review.

1. **REGULAR ARTICLE**

   Regular articles are full-length original empirical investigations, consisting of introduction, materials and methods, results and discussion, conclusions. Original work must provide references and an explanation on research findings that contain new and significant findings.

   Size: Generally, these are expected to be between 6 and 12 journal pages (excluding the abstract, references, tables and/or figures), a maximum of 80 references, and an abstract of 100–200 words.

2. **REVIEW ARTICLE**

   These report critical evaluation of materials about current research that has already been published by organizing, integrating, and evaluating previously published materials. It summarizes the status of knowledge and outline future directions of research within the journal scope. Review articles should aim to provide systemic overviews, evaluations and interpretations of research in a given field. Re-analyses as meta-analysis and systemic reviews are encouraged. The manuscript title must start with "Review Article:".

   Size: These articles do not have an expected page limit or maximum number of references, should include appropriate figures and/or tables, and an abstract of 100–200 words. Ideally, a review article should be of 7 to 8 printed pages.

3. **SHORT COMMUNICATIONS**

   They are timely, peer-reviewed and brief. These are suitable for the publication of significant technical advances and may be used to:

   (a) report new developments, significant advances and novel aspects of experimental and theoretical methods and techniques which are relevant for scientific investigations within the journal scope;

   (b) report/discuss on significant matters of policy and perspective related to the science of the journal, including 'personal' commentary;

   (c) disseminate information and data on topical events of significant scientific and/or social interest within the scope of the journal.

   The manuscript title must start with "*Brief Communication*:".

   Size: These are usually between 2 and 4 journal pages and have a maximum of three figures and/or tables, from 8 to 20 references, and an abstract length not exceeding 100 words. Information must be in short but complete form and it is not intended to publish preliminary results or to be a reduced version of Regular or Rapid Papers.

4. OTHERS

Brief reports, case studies, comments, concept papers, Letters to the Editor, and replies on previously published articles may be considered.

PLEASE NOTE: NO EXCEPTIONS WILL BE MADE FOR PAGE LENGTH.

Language Accuracy

Pertanika **emphasizes** on the linguistic accuracy of every manuscript published. Articles must be in **English** and they must be competently written and argued in clear and concise grammatical English. Contributors are strongly advised to have the manuscript checked by a colleague with ample experience in writing English manuscripts or a competent English language editor.

Author(s) **must provide a certificate** confirming that their manuscripts have been adequately edited. A proof from a recognised editing service should be submitted together with the cover letter at the time of submitting a manuscript to Pertanika**. All editing costs must be borne by the author(s)**. This step, taken by authors before submission, will greatly facilitate reviewing, and thus publication if the content is acceptable.

Linguistically hopeless manuscripts will be rejected straightaway (e.g., when the language is so poor that one cannot be sure of what the authors really mean). This process, taken by authors before submission, will greatly facilitate reviewing, and thus publication if the content is acceptable.

MANUSCRIPT FORMAT

The paper should be submitted in one column format with at least 4cm margins and 1.5 line spacing throughout. Authors are advised to use Times New Roman 12-point font and *MS Word* format.

1. *Manuscript Structure*

Manuscripts in general should be organised in the following order:

Page 1: *Running title*

This page should **only** contain the running title of your paper. The running title is an abbreviated title used as the running head on every page of the manuscript. The running title should not exceed 60 characters, counting letters and spaces.

Page 2: *Author(s) and Corresponding author information.*

This page should contain the **full title** of your paper not exceeding 25 words, with name(s) of all the authors, institutions and corresponding author's name, institution and full address (Street address, telephone number (including extension), hand phone number, and e-mail address) for editorial correspondence. First and corresponding authors must be clearly indicated.

The names of the authors may be abbreviated following the international naming convention. e.g. Salleh, A.B.[1], Tan, S.G[2]*., and Sapuan, S.M[3].

Authors' addresses. Multiple authors with different addresses must indicate their respective addresses separately by superscript numbers:

George Swan[1] and Nayan Kanwal[2]
[1]Department of Biology, Faculty of Science, Duke University, Durham, North Carolina, USA.,
[2]Office of the Deputy Vice Chancellor (R&I), Universiti Putra Malaysia, Serdang, Malaysia.

A **list** of number of **black and white / colour figures and tables** should also be indicated on this page**.** Figures submitted in color will be printed in colour. See "*5. Figures & Photographs"* for details.

Page 3: *Abstract*

This page should **repeat** the **full title** of your paper with only the **Abstract** (the abstract should be less than 250 words for a Regular Paper and up to 100 words for a Short Communication), and **Keywords**.
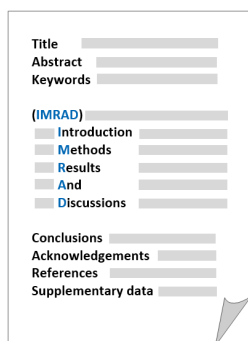
Keywords: Not more than eight keywords in alphabetical order must be provided to describe the contents of the manuscript.

Page 4: *Introduction*

This page should begin with the **Introduction** of your article and followed by the rest of your paper.

2. Text

Regular Papers should be prepared with the headings *Introduction, Materials and Methods, Results and Discussion*, *Conclusions, Acknowledgements, References, and Supplementary data* (if avavailble) in this order.

Title
Abstract
Keywords

(IMRAD)
   Introduction
   Methods
   Results
   And
   Discussions

Conclusions
Acknowledgements
References
Supplementary data

MAKE YOUR ARTICLES AS CONCISE AS POSSIBLE

Most scientific papers are prepared according to a format called IMRAD. The term represents the first letters of the words Introduction, Materials and Methods, Results, And, Discussion. It indicates a pattern or format rather than a complete list of headings or components of research papers; the missing parts of a paper are: Title, Authors, Keywords, Abstract, Conclusions, and References. Additionally, some papers include Acknowledgments and Appendices.

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

3. Equations and Formulae

These must be set up clearly and should be typed double spaced. Numbers identifying equations should be in square brackets and placed on the right margin of the text.

4. Tables

All tables should be prepared in a form consistent with recent issues of Pertanika and should be numbered consecutively with Roman numerals. Explanatory material should be given in the table legends and footnotes. Each table should be prepared on a new page, embedded in the manuscript.

When a manuscript is submitted for publication, tables must also be submitted separately as data - .doc, .rtf, Excel or PowerPoint files- because tables submitted as image data cannot be edited for publication and are usually in low-resolution.

5. Figures & Photographs

Submit an **original** figure or photograph. Line drawings must be clear, with high black and white contrast. Each figure or photograph should be prepared on a new page, embedded in the manuscript for reviewing to keep the file of the manuscript under 5 MB. These should be numbered consecutively with Roman numerals.

Figures or photographs must also be submitted separately as TIFF, JPEG, or Excel files- because figures or photographs submitted in low-resolution embedded in the manuscript cannot be accepted for publication. For electronic figures, create your figures using applications that are capable of preparing high resolution TIFF files. In general, we require **300 dpi** or higher resolution for **coloured and half-tone artwork**, and **1200 dpi or higher** for **line drawings** are required.

Failure to comply with these specifications will require new figures and delay in publication.
NOTE: Illustrations may be produced in colour at no extra cost at the discretion of the Publisher; the author could be charged Malaysian Ringgit 50 for each colour page.

6. References

References begin on their own page and are listed in alphabetical order by the first author's last name. Only references cited within the text should be included. All references should be in 12-point font and double-spaced.

NOTE: When formatting your references, please follow the **APA reference style** (6th Edition). Ensure that the references are strictly in the journal's prescribed style, failing which your article will **not be accepted for peer-review**. You may refer to the *Publication Manual of the American Psychological Association* for further details (http://www.apastyle.org/).

7. General Guidelines

Abbreviations: Define alphabetically, other than abbreviations that can be used without definition. Words or phrases that are abbreviated in the introduction and following text should be written out in full the first time that they appear in the text, with each abbreviated form in parenthesis. Include the common name or scientific name, or both, of animal and plant materials.

Acknowledgements: Individuals and entities that have provided essential support such as research grants and fellowships and other sources of funding should be acknowledged. Contributions that do not involve researching (clerical assistance or personal acknowledgements) should **not** appear in acknowledgements.

Authors' Affiliation: The primary affiliation for each author should be the institution where the majority of their work was done. If an author has subsequently moved to another institution, the current address may also be stated in the footer.

Co-Authors: The commonly accepted guideline for authorship is that one must have substantially contributed to the development of the paper and share accountability for the results. Researchers should decide who will be an author and what order they will be listed depending upon their order of importance to the study. Other contributions should be cited in the manuscript's Acknowledgements.

Copyright Permissions: Authors should seek necessary permissions for quotations, artwork, boxes or tables taken from other publications or from other freely available sources on the Internet before submission to Pertanika. Acknowledgement must be given to the original source in the illustration legend, in a table footnote, or at the end of the quotation.

Footnotes: Current addresses of authors if different from heading may be inserted here.

Page Numbering: Every page of the manuscript, including the title page, references, tables, etc. should be numbered.

Spelling: The journal uses American or British spelling and authors may follow the latest edition of the Oxford Advanced Learner's Dictionary for British spellings.

SUBMISSION OF MANUSCRIPTS

Owing to the volume of manuscripts we receive, we must insist that all submissions be made electronically using the **online submission system ScholarOne™**, a web-based portal by Thomson Reuters. For more information, go to our web page and click "**Online Submission**".

Submission Checklist

1. **MANUSCRIPT**: Ensure your MS has followed the Pertanika style particularly the first four pages as explained earlier. The article should be written in a good academic style and provide an accurate and succinct description of the contents ensuring that grammar and spelling errors have been corrected before submission. It should also not exceed the suggested length.

   COVER LETTER: All submissions must be accompanied by a cover letter detailing what you are submitting. Papers are accepted for publication in the journal on the understanding that the article is **original** and the content has **not been published** either **in English** or **any other language(s)** or **submitted for publication elsewhere**. The letter should also briefly describe the research you are reporting, why it is important, and why you think the readers of the journal would be interested in it. The cover letter must also contain an acknowledgement that all authors have contributed significantly, and that all authors have approved the paper for release and are in agreement with its content.

   The cover letter of the paper should contain (i) the title; (ii) the full names of the authors; (iii) the addresses of the institutions at which the work was carried out together with (iv) the full postal and email address, plus telephone numbers and emails of all the authors. The current address of any author, if different from that where the work was carried out, should be supplied in a footnote.

   The above must be stated in the cover letter. Submission of your manuscript will not be accepted until a cover letter has been received

2. **COPYRIGHT**: Authors publishing the Journal will be asked to sign a copyright form. In signing the form, it is assumed that authors have obtained permission to use any copyrighted or previously published material. All authors must read and agree to the conditions outlined in the form, and must sign the form or agree that the corresponding author can sign on their behalf. Articles cannot be published until a signed form *(original pen-to-paper signature)* has been received.

Please do **not** submit manuscripts to the editor-in-chief or to any other office directly. Any queries must be directed to the **Chief Executive Editor's** office via email to nayan@upm.my.

Visit our Journal's website for more details at http://www.pertanika.upm.edu.my/home.php.

HARDCOPIES OF THE JOURNALS AND OFF PRINTS

Under the Journal's open access initiative, authors can choose to download free material (via PDF link) from any of the journal issues from Pertanika's website. Under "**Browse Journals**" you will see a link, "*Current Issues*" or "*Archives*". Here you will get access to all current and back-issues from 1978 onwards.

The **corresponding author** for all articles will receive one complimentary hardcopy of the journal in which his/her articles is published. In addition, 20 off prints of the full text of their article will also be provided. Additional copies of the journals may be purchased by writing to the Chief Executive Editor.

Why should you publish in *Pertanika?*

BENEFITS TO AUTHORS

PROFILE: Our journals are circulated in large numbers all over Malaysia, and beyond in Southeast Asia. Our circulation covers other overseas countries as well. We ensure that your work reaches the widest possible audience in print and online, through our wide publicity campaigns held frequently, and through our constantly developing electronic initiatives such as Web of Science Author Connect backed by Thomson Reuters.

QUALITY: Our journals' reputation for quality is unsurpassed ensuring that the originality, authority and accuracy of your work are fully recognised. Each manuscript submitted to Pertanika undergoes a rigid originality check. Our double-blind peer refereeing procedures are fair and open, and we aim to help authors develop and improve their scientific work. Pertanika is now over 38 years old; this accumulated knowledge has resulted in our journals being indexed in SCOPUS (Elsevier), **Thomson (ISI) Web of Science™ Core Collection, Emerging Sources Citation Index (ESCI)**, Web of Knowledge [BIOSIS & CAB Abstracts], EBSCO, DOAJ, ERA, AGRICOLA, Google Scholar, ISC, TIB, Journal Guide, Citefactor, Cabell's Directories and MyCite.

AUTHOR SERVICES: We provide a rapid response service to all our authors, with dedicated support staff for each journal, and a point of contact throughout the refereeing and production processes. Our aim is to ensure that the production process is as smooth as possible, is borne out by the high number of authors who prefer to publish with us.

CODE OF ETHICS: Our Journal has adopted a Code of Ethics to ensure that its commitment to integrity is recognized and adhered to by contributors, editors and reviewers. It warns against plagiarism and self-plagiarism, and provides guidelines on authorship, copyright and submission, among others.

PRESS RELEASES: Landmark academic papers that are published in Pertanika journals are converted into press-releases as a unique strategy for increasing visibility of the journal as well as to make major findings accessible to non-specialist readers. These press releases are then featured in the university's UK and Australian based research portal, ResearchSEA, for the perusal of journalists all over the world.

LAG TIME: The elapsed time from submission to publication for the articles averages 3 to 4 months. A decision on acceptance of a manuscript is reached in 3 to 4 months (average 14 weeks).

Address your submissions to:
The Chief Executive Editor
Tel: +603 8947 1622
nayan@upm.my

Journal's Profile: www.pertanika.upm.edu.my/

Call for Papers 2017-18

now accepting submissions...

Pertanika invites you to explore frontiers from all key areas of agriculture, science and technology to social sciences and humanities.

Original research and review articles are invited from scholars, scientists, professors, post-docs, and university students who are seeking publishing opportunities for their research papers through the Journal's three titles; JTAS, JST & JSSH. Preference is given to the work on leading and innovative research approaches.

Pertanika is a fast track peer-reviewed and open-access academic journal published by Universiti Putra Malaysia. To date, Pertanika Journals have been indexed by many important databases. Authors may contribute their scientific work by publishing in UPM's hallmark SCOPUS & ISI indexed journals.

Our journals are open access - international journals. Researchers worldwide will have full access to all the articles published online and be able to download them with zero subscription fee.

Pertanika uses online article submission, review and tracking system for quality and quick review processing backed by Thomson Reuter's ScholarOne™. Journals provide rapid publication of research articles through this system.

For details on the Guide to Online Submissions, please visit http://www.pertanika.upm.edu.my/guide_online_submission.php

About the Journal

Pertanika is an international multidisciplinary peer-reviewed leading journal in Malaysia which began publication in 1978. The journal publishes in three different areas — Journal of Tropical Agricultural Science (JTAS); Journal of Science and Technology (JST); and Journal of Social Sciences and Humanities (JSSH). All journals are published in English.

JTAS is devoted to the publication of original papers that serves as a forum for practical approaches to improving quality in issues pertaining to tropical agricultural research- or related fields of study. It is published four times a year in *February, May, August* and *November.*

JST caters for science and engineering research- or related fields of study. It is published twice a year in *January* and *July.*

JSSH deals in research or theories in social sciences and humanities research. It aims to develop as a flagship journal with a focus on emerging issues pertaining to the social and behavioural sciences as well as the humanities, particularly in the Asia Pacific region. It is published four times a year in *March, June, September* and *December.*

Contents