

Jordan Journal of Mechanical and Industrial Engineering (JJMIE)

JJMIE is a high-quality scientific journal devoted to fields of Mechanical and Industrial Engineering. It is published by The Hashemite University in corporation with the Jordanian Scientific Research Support Fund.

EDITORIAL BOARD

Editor-in-Chief

Prof. **Mousa S. Mohsen**

Editorial board

Prof. **Bilal A. Akash**
Hashemite University

Prof. **Adnan Z. Al-Kilany**
University of Jordan

Prof. **Ayman A. Al-Maaaitah**
Mutah University

Prof. **Moh'd A. Al-Nimr**
Jordan University of Science and Technology

Prof. **Ali A. Badran**
University of Jordan

Prof. **Naseem M. Sawaqed**
Mutah University

Assistant Editor

Dr. **Ahmed Al-Ghandoor**
Hashemite University

THE INTERNATIONAL ADVISORY BOARD

Abu-Qudais, Mohammad
Jordan University of Science & Technology, Jordan

Abu-Mulaweh, Hosni
Purdue University at Fort Wayne, USA

Afaneh Abdul-Hafiz
Robert Bosch Corporation, USA

Afonso, Maria Dina
Institute Superior Tecnico, Portugal

Badiru, Adedji B.
The University of Tennessee, USA

Bejan, Adrian
Duke University, USA

Chalhoub, Nabil G.
Wayne State University, USA

Cho, Kyu-Kab
Pusan National University, South Korea

Dincer, Ibrahim
University of Ontario Institute of Technology,
Canada

Douglas, Roy
Queen's University, U. K

El Bassam, Nasir
International Research Center for Renewable
Energy, Germany

Haik, Yousef
United Arab Emirates University, UAE

Jaber, Jamal
Al- Balqa Applied University, Jordan

Jubran, Bassam
Ryerson University, Canada

Kakac, Sadik
University of Miami, USA

Khalil, Essam-Eddin
Cairo University, Egypt

Mutoh, Yoshiharu
Nagaoka University of Technology, Japan

Pant, Durbin
Iowa State University, USA

Riffat, Saffa
The University of Nottingham, U. K

Saghir, Ziad
Ryerson University, Canada

Sarkar, MD. Abdur Rashid
Bangladesh University of Engineering &
Technology, Bangladesh

Siginer, Dennis
Wichita State University, USA

Sopian, Kamaruzzaman
University Kebangsaan Malaysia, Malaysia

Tzou, Gow-Yi
Yung-Ta Institute of Technology and Commerce,
Taiwan

EDITORIAL BOARD SUPPORT TEAM

Language Editor

Dr. Abdullah Jaradat

Publishing Lavout

Eng. Hasan B. Al-Ba'ba'a

SUBMISSION ADDRESS:

Prof. **Mousa S. Mohsen**, Editor-in-Chief
Jordan Journal of Mechanical & Industrial Engineering,
Hashemite University,
PO Box 330127, Zarqa, 13133 , Jordan
E-mail: jjmie@hu.edu.jo



Hashemite Kingdom of Jordan



Hashemite University

Jordan Journal of
Mechanical and Industrial Engineering

JJMIIE

An International Peer-Reviewed Scientific Journal

<http://jjmie.hu.edu.jo/>

ISSN 1995-6665

Jordan Journal of Mechanical and Industrial Engineering (JJMIE)

JJMIE is a high-quality scientific journal devoted to fields of Mechanical and Industrial Engineering. It is published by The Jordanian Ministry of Higher Education and Scientific Research in corporation with the Hashemite University.

Introduction: The Editorial Board is very committed to build the Journal as one of the leading international journals in mechanical and industrial engineering sciences in the next few years. With the support of the Ministry of Higher Education and Scientific Research and Jordanian Universities, it is expected that a heavy resource to be channeled into the Journal to establish its international reputation. The Journal's reputation will be enhanced from arrangements with several organizers of international conferences in publishing selected best papers of the conference proceedings.

Aims and Scope: Jordan Journal of Mechanical and Industrial Engineering (JJMIE) is a refereed international journal to be of interest and use to all those concerned with research in various fields of, or closely related to, mechanical and industrial engineering disciplines. Jordan Journal of Mechanical and Industrial Engineering aims to provide a highly readable and valuable addition to the literature which will serve as an indispensable reference tool for years to come. The coverage of the journal includes all new theoretical and experimental findings in the fields of mechanical and industrial engineering or any closely related fields. The journal also encourages the submission of critical review articles covering advances in recent research of such fields as well as technical notes.

Guide for Authors

Manuscript Submission

High-quality submissions to this new journal are welcome now and manuscripts may be either submitted online or mail.

Online: For online submission upload one copy of the full paper including graphics and all figures at the online submission site, accessed via E-mail: jjmie@hu.edu.jo. The manuscript must be written in MS Word Format. All correspondence, including notification of the Editor's decision and requests for revision, takes place by e-mail and via the Author's homepage, removing the need for a hard-copy paper trail.

By Mail: Manuscripts (1 original and 3 copies) accompanied by a covering letter may be sent to the Editor-in-Chief. However, a copy of the original manuscript, including original figures, and the electronic files should be sent to the Editor-in-Chief. Authors should also submit electronic files on disk (one disk for text material and a separate disk for graphics), retaining a backup copy for reference and safety.

Note that contributions may be either submitted online or sent by mail. Please do NOT submit via both routes. This will cause confusion and may lead to delay in article publication. Online submission is preferred.

Submission address and contact:

Prof. **Mousa S. Mohsen**, Editor-in-Chief
Jordan Journal of Mechanical & Industrial Engineering,
Hashemite University,
PO Box 330127, Zarqa, 13115 , Jordan
E-mail: jjmie@hu.edu.jo

Types of contributions: Original research papers

Corresponding author: Clearly indicate who is responsible for correspondence at all stages of refereeing and publication, including post-publication. Ensure that telephone and fax numbers (with country and area code) are provided in addition to the e-mail address and the complete postal address. Full postal addresses must be given for all co-authors.

Original material: Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, without the written consent of the Publisher. Authors found to be deliberately contravening the submission guidelines on originality and exclusivity shall not be considered for future publication in this journal.

Supplying Final Accepted Text on Disk: If online submission is not possible: Once the paper has been accepted by the editor, an electronic version of the text should be submitted together with the final hardcopy of the manuscript. The electronic version must match the hardcopy exactly. We accept MS Word format only. Always keep a backup copy of the electronic file for reference and safety. Label the disk with your name. Electronic files can be stored on CD.

Notification: Authors will be notified of the acceptance of their paper by the editor. The Publisher will also send a notification of receipt of the paper in production.

Copyright: All authors must sign the Transfer of Copyright agreement before the article can be published. This transfer agreement enables Jordan Journal of Mechanical and Industrial Engineering to protect the copyrighted material for the authors, but does not relinquish the authors' proprietary rights. The copyright transfer covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microfilm or any other reproductions of similar nature and translations.

PDF Proofs: One set of page proofs in PDF format will be sent by e-mail to the corresponding author, to be checked for typesetting/editing. The corrections should be returned within 48 hours. No changes in, or additions to, the accepted (and subsequently edited) manuscript will be allowed at this stage. Proofreading is solely the author's responsibility. Any queries should be answered in full. Please correct factual errors only, or errors introduced by typesetting. Please note that once your paper has been proofed we publish the identical paper online as in print.

Author Benefits

Page charge: Publication in this journal is free of charge.

Free off-prints: Three journal issues of which the article appears in along with twenty-five off-prints will be supplied free of charge to the corresponding author. Corresponding authors will be given the choice to buy extra off-prints before printing of the article.

Manuscript Preparation:

General: Editors reserve the right to adjust style to certain standards of uniformity. Original manuscripts are discarded after publication unless the Publisher is asked to return original material after use. If online submission is not possible, an electronic copy of the manuscript on disk should accompany the final accepted hardcopy version. Please use MS Word for the text of your manuscript.

Structure: Follow this order when typing manuscripts: Title, Authors, Affiliations, Abstract, Keywords, Introduction, Main text, Conclusions, Acknowledgements, Appendix, References, Figure Captions, Figures and then Tables. For submission in hardcopy, do not import figures into the text - see Illustrations. For online submission, please supply figures imported into the text AND also separately as original graphics files. Collate acknowledgements in a separate section at the end of the article and do not include them on the title page, as a footnote to the title or otherwise.

Text Layout: Use double spacing and wide (3 cm) margins. Ensure that each new paragraph is clearly indicated. Present tables and figure legends on separate pages at the end of the manuscript. If possible, consult a recent issue of the journal to become familiar with layout and conventions. All footnotes (except for table and corresponding author footnotes) should be identified with superscript Arabic numbers. To conserve space, authors are requested to mark the less important parts of the paper (such as records of experimental results) for printing in smaller type. For long papers (more than 4000 words) sections which could be deleted without destroying either the sense or the continuity of the paper should be indicated as a guide for the editor. Nomenclature should conform to that most frequently used in the scientific field concerned. Number all pages consecutively; use 12 or 10 pt font size and standard fonts. If submitting in hardcopy, print the entire manuscript on one side of the paper only.

Corresponding author: Clearly indicate who is responsible for correspondence at all stages of refereeing and publication, including post-publication. The corresponding author should be identified with an asterisk and footnote. Ensure that telephone and fax numbers (with country and area code) are provided in addition to the e-mail address and the complete postal address. Full postal addresses must be given for all co-authors. Please consult a recent journal paper for style if possible.

Abstract: A self-contained abstract outlining in a single paragraph the aims, scope and conclusions of the paper must be supplied.

Keywords: Immediately after the abstract, provide a maximum of six keywords (avoid, for example, 'and', 'of'). Be sparing with abbreviations: only abbreviations firmly established in the field may be eligible.

Symbols: All Greek letters and unusual symbols should be identified by name in the margin, the first time they are used.

Units: Follow internationally accepted rules and conventions: use the international system of units (SI). If other quantities are mentioned, give their equivalent in SI.

Maths: Number consecutively any equations that have to be displayed separately from the text (if referred to explicitly in the text).

References: All publications cited in the text should be presented in a list of references following the text of the manuscript.

Text: Indicate references by number(s) in square brackets in line with the text. The actual authors can be referred to, but the reference number(s) must always be given.

List: Number the references (numbers in square brackets) in the list in the order in which they appear in the text.

Examples:

Reference to a journal publication:

- [1] M.S. Mohsen, B.A. Akash, "Evaluation of domestic solar water heating system in Jordan using analytic hierarchy process". Energy Conversion & Management, Vol. 38, No. 9, 1997, 1815-1822.

Reference to a book:

- [2] Strunk Jr W, White EB. The elements of style. 3rd ed. New York: Macmillan; 1979.

Reference to a conference proceeding:

- [3] B. Akash, S. Odeh, S. Nijmeh, "Modeling of solar-assisted double-tube evaporator heat pump system under local climate conditions". 5th Jordanian International Mechanical Engineering Conference, Amman, Jordan, 2004.

Reference to a chapter in an edited book:

- [4] Mettam GR, Adams LB. How to prepare an electronic version of your article. In: Jones BS, Smith RZ, editors. Introduction to the electronic age, New York: E-Publishing Inc; 1999, p. 281-304

Free Online Color: If, together with your accepted article, you submit usable color and black/white figures then the journal will ensure that these figures will appear in color on the journal website electronic version.

Tables: Tables should be numbered consecutively and given suitable captions and each table should begin on a new page. No vertical rules should be used. Tables should not unnecessarily duplicate results presented elsewhere in the manuscript (for example, in graphs). Footnotes to tables should be typed below the table and should be referred to by superscript lowercase letters.

PAGES	PAPERS
285 – 289	Conceptual Design of a Single DOF Human-Like Eight-Bar Leg Mechanism <i>O. Al-Araidah, W. Batayneh, T. Darabseh, S. M. BaniHani</i>
291 – 300	A Diverse Neural Network Ensemble Team for Mean Shift Detection in X-Bar and CUSUM Control Charts <i>M. A. Barghash</i>
301 – 306	Fatigue Properties under Constant Stress/Variable Stress Amplitude and Coaxing Effect of Acicular Ductile Iron and 42 CrMo4 Steel <i>A.R. I. Kheder, N. M. Jubeh, E. M. Tahah</i>
307– 314	Investigation of the Endurance Characteristics of a Compression Ignition Engine runs on Jatropha Biodiesel <i>A. Shanono, J. Enaburekhan</i>
315 – 323	An Assessment of Maintenance Practices and problems in Jordanian Industries <i>K. K. Tahboub</i>
325 – 333	Efficiency of Free Cooling Technique in Air Refrigeration Systems <i>A.Al-Salaymeh, M.R. Abdelkader</i>
335 – 343	A Comparison Study between Various Fuzzy Clustering Algorithms <i>K. M. Bataineh, M. Naji, M. Saqer</i>
345 – 351	Machining Characteristics of Multiwall-CNT Reinforced Al/Al-Si Composites using Recurrence Quantification Analysis <i>K. V. Gangadharan, K. S. Umashankar, Ravish, V. Desai</i>
353 – 357	Modeling of the MEMS Reactive Ion Etching Process Using Neural Networks <i>M. Ashhab, N. Talat</i>
359 – 368	Developing a Software to Predict Thermal Comfort of Humans at Work <i>N. J. Rabadi</i>
369 – 372	An Experiment of Chocolate Softness Measurements Using a New Design Machine <i>Bassam A. Al-Helou</i>
373 – 382	Influence of Camera's Optical Axis Non-perpendicularity on Measurement Accuracy of Two-dimensional Digital Image Correlation <i>A. Hijazi, A. Friedl and C. J. Kähler</i>

Conceptual Design of a Single DOF Human-Like Eight-Bar Leg Mechanism

O. Al-Araidah^{*a}, W. Batayneh^b, T. Darabseh^c, S. M. BaniHani^d

^aIndustrial Engineering Department, Jordan University of Science and Technology, Irbid, 22110, Jordan

^bMechanical Engineering Department, Jordan University of Science and Technology, Irbid, 22110, Jordan

^cAeronautical Engineering Department, Jordan University of Science and Technology, Irbid, 22110, Jordan

^dMechatronics Engineering Department, The Hashemite University, Zarqa 13115, Jordan

Abstract

Legs are the most important elements for accomplishing human physical work including transportation or displacement. The article presents a mechanical reproduction of the human walking apparatus. Using design rules, a final mechanism configuration is achieved such that the crank is a binary link connected to a binary ground link. The resulting linkage is a single degree-of-freedom (DOF) eight-bar mechanism. The mechanism exemplifies the shape and movement of a human leg. The mechanism is simulated and tested to verify the proposed synthesis.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: kinematic synthesis; DOF; Biped walking; Leg mechanism

1. Introduction

Legs are the primary means to satisfy relocation needs for human and many types of animals and insects. Moreover, legs help protect other parts of the body by keeping them far from ground. In the literature, Weber and Weber [1] studied the human walking apparatus and showed that the skeletal system constitutes a mechanism that can potentially be reproduced mechanically. The human leg has a cylindrical shape with the driving mechanism and controllers located at the upper part of the leg. Many researchers studied the walking apparatus and showed that walking patterns are measurable, predictable and repeatable. Hence, mechanical reproductions of walking mechanisms prove useful in the development of humanoid for applications ranging from toys to space exploration.

This article presents a closed loop one degree of freedom (DOF) eight-bar human-like leg mechanism. Like it in a human, the proposed leg design includes a hip, a knee and ankle joints, in addition to a clear femur and fibula. Proportions of the mechanism are appropriately scaled to approximate the human leg. Moreover, the resulting movements of the leg proportions approximate that of a typical person walking. The approximate mechanism precision points represent the human-type walking gait, known as the ovoid path [2]. Figure 1 illustrates an approximation of the desired path. The path is composed of a straight-line, line H₁H₂H₃, portion that represents the times when the foot contacts the ground. Points H₁ and H₃ represent the extreme points of the path where distance H₁H₃ represents the horizontal stride. The arched portion H₃H₄H₁ represents the motion of the foot

to assume a new position through which the foot has no contact with the ground. To this end, the authors followed a typical design procedure to develop, simulate and test design alternatives.

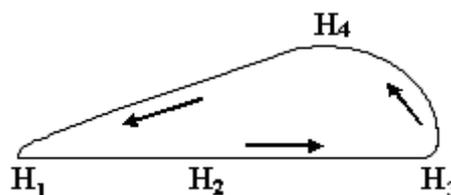


Figure 1: An approximation of the desired walking gait.

The rest of the paper is organized as follows: Section 2 presents a review of related literature. Section 3 describes the mechanism and presents kinematic synthesis of the mechanism. Section 4 provides simulation results obtained from testing the mechanism. Concluding remarks are presented in the final section.

2. Literature Review

Many researchers studied the human walking apparatus and produced motion diagrams representing the movement. A mechanical reproduction of these mechanisms and their movement were useful in various fields such as the development of human prosthetics, human mimicking robots, and advancements in research areas such as biomimetics, military combat, cinematography, toys, and terrestrial and extraterrestrial exploration [3-4].

* Corresponding author. e-mail: alarao@just.edu.jo

In the literature, researchers developed various open and closed kinematic chains for human-like walking machines. While open kinematic chains are generally more flexible and easier to design, their large number of degrees of freedom (DOF) makes them more expensive and harder to control. On the other hand, most of the existing closed chain configurations are generally bulky and do not look like a human leg. The closest approximation of human-like leg with a closed kinematic chain uses a six-bar mechanism [5]. Batayneh, et al. [5] presented a single DOF Watt I six-bar mechanism that typifies the human's motion. The proposed mechanism includes a hip, femur, knee, and fibula. Shieh et al. [6-7] proposed a two-stage optimization process of a two DOF leg mechanism. In the first stage, leg dimensions are optimized with respect to the design objectives, which include minimizing leg size and actuating forces. In the second stage, spring elements with various placement configurations are considered for further reduction of the actuating forces. Sangwan et al. [8] introduced a design for a self-excited biped walking mechanism consisting of two legs that are connected in series at the hip joint through a motor. Each leg has a thigh and a shank connected at a passive knee joint that has a knee stopper restricting hyperextension similar to the human knee. Mukherjee et al. [9] studied the stability of the previous bipedal gait and verified their work by comparing the results with human gait.

Many researchers used motor/motors as actuators while others tend to use other kinds of actuations. Verrelst et al. [10] designed a biped walking mechanism actuated using pneumatic muscles. Hosoda et al. [11] studied antagonistic pneumatic actuators that can be utilized to achieve locomotion for different cases. In [12], the authors presented a five DOF design for a three dimensional bipedal walking robot with human-like morphology and gait. Yavin [13] considered the modeling and control of a three DOF walking four-bar open chain linkage robot. Capi [14] introduced an optimal scheme for a biped robot that has two legs with two DOF each; one translational DOF achieved by a DC motor attached at the body of the robot, and the other DOF is the rotational motion at the ankle joint. Companies such as Honda Motors, Kawada Industries and Sony played a major roll in the competition towards developing legged robots [15-19]. Hurmuzlu et al. [20] studies different existing bipedal robots including "Honda's P3 robot" and cited that most of these existing robots need development in terms of walking patterns and in terms of stability. Mousavi and Bagheri [21] presented a mathematical interpolation of a seven-link robot. The authors presented a simulation study of the biped robot for both nominal and un-nominal gaits. Barai and Nonami [22] presented a two-degree-of-freedom fuzzy controller for hydraulically actuated six leg robotic mechanisms. The proposed designs generally focus on creating the desired walking gait while overlooking the shape of the leg. This article utilizes observations of previous works in designing a closed-loop 8-bar human-like leg mechanism.

3. Mechanism Description and Kinematic Synthesis of the Linkage

A planar one DOF eight-bar mechanism is shown in Figure 2a. For convenience, an alternative sketch of the mechanism is shown in Figure 2b. The proposed path generator $O_AAO_BBCDEFH$ where, B is a coupler point on O_BBC , D is a coupler point on CE , and H is a coupler

point on FE that represents the nickel joint. The rotation of the short crank O_AA provides forward/backward motion. The mechanism could be provided with a turning capability about its vertical axis using a second DOF motion. In this design, the lengths of the linkages are selected such that $CE = O_BF = H'H$ and $O_BC \geq FE$. Moreover, $CD' \leq CE/3$ measured from C , B' is located midway on segment O_BC and H' is located midway on segment EF .

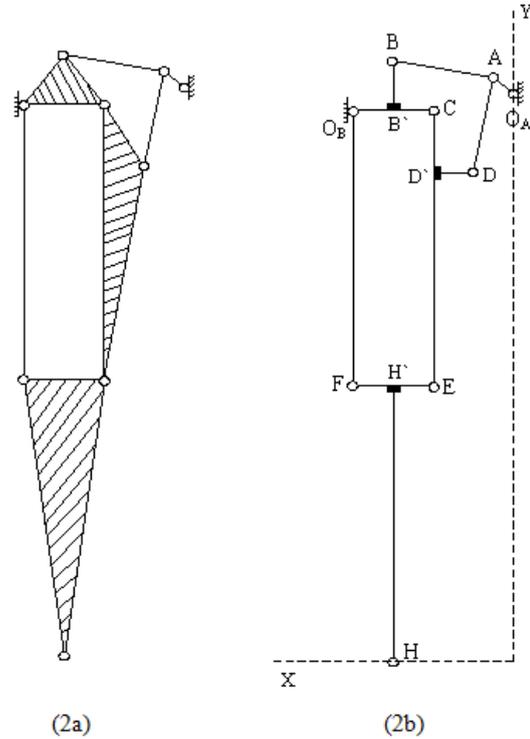


Figure 2: Configuration of the proposed leg mechanism.

Four precision points are selected to reflect the desired motion, and the complex number method is used to synthesize the mechanism. Figure 3 shows the general notated configuration in addition to the individual closed loops.

Equations (1) through (4) illustrate the standard form equations for resulting closed loops.

Loop1:

$$\begin{aligned}
 Z_1(e^{i\alpha_1} - 1) + Z_3(e^{i\theta_{31}} - 1) + Z_4(e^{i\theta_{41}} - 1) + Z_6/2(e^{i\theta_{61}} - 1) &= 0, \quad J = 2,3,4 \\
 Z_1(e^{i\alpha_2} - 1) + Z_3(e^{i\theta_{32}} - 1) + Z_4(e^{i\theta_{42}} - 1) + Z_6/2(e^{i\theta_{62}} - 1) &= 0 \\
 Z_1(e^{i\alpha_3} - 1) + Z_3(e^{i\theta_{33}} - 1) + Z_4(e^{i\theta_{43}} - 1) + Z_6/2(e^{i\theta_{63}} - 1) &= 0 \\
 Z_1(e^{i\alpha_4} - 1) + Z_3(e^{i\theta_{34}} - 1) + Z_4(e^{i\theta_{44}} - 1) + Z_6/2(e^{i\theta_{64}} - 1) &= 0
 \end{aligned}
 \tag{1}$$

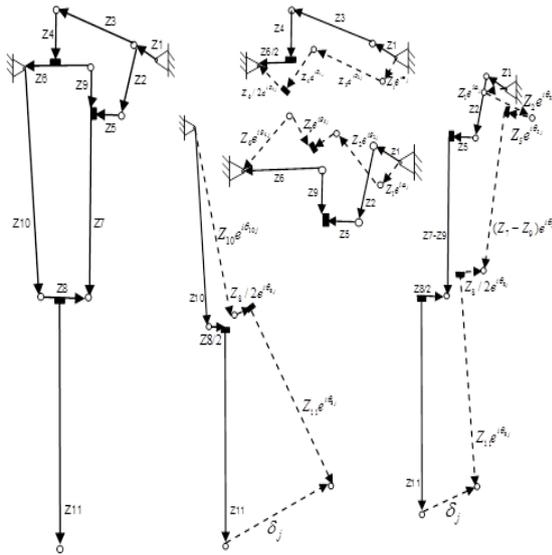


Figure 3: Notated Mechanism Configuration and Closed Loops.

Loop 2:

$$\begin{aligned}
 Z_{10}(e^{i\theta_{10j}} - 1) + Z_8/2(e^{i\theta_{8j}} - 1) + Z_{11}(e^{i\theta_{11j}} - 1) &= \delta_j, \quad J = 2,3,4 \\
 Z_{10}(e^{i\theta_{102}} - 1) + Z_8/2(e^{i\theta_{82}} - 1) + Z_{11}(e^{i\theta_{112}} - 1) &= \delta_2 \\
 Z_{10}(e^{i\theta_{103}} - 1) + Z_8/2(e^{i\theta_{83}} - 1) + Z_{11}(e^{i\theta_{113}} - 1) &= \delta_3 \\
 Z_{10}(e^{i\theta_{104}} - 1) + Z_8/2(e^{i\theta_{84}} - 1) + Z_{11}(e^{i\theta_{114}} - 1) &= \delta_4
 \end{aligned} \tag{2}$$

Loop 3:

$$\begin{aligned}
 Z_1(e^{i\alpha_j} - 1) + Z_2(e^{i\theta_{2j}} - 1) + Z_5(e^{i\theta_{5j}} - 1) - Z_9(e^{i\theta_{9j}} - 1) \\
 + Z_6(e^{i\theta_{6j}} - 1) &= 0, \quad j = 2,3,4 \\
 Z_1(e^{i\alpha_2} - 1) + Z_2(e^{i\theta_{22}} - 1) + Z_5(e^{i\theta_{52}} - 1) - Z_9(e^{i\theta_{92}} - 1) + Z_6(e^{i\theta_{62}} - 1) &= 0 \\
 Z_1(e^{i\alpha_3} - 1) + Z_2(e^{i\theta_{23}} - 1) + Z_5(e^{i\theta_{53}} - 1) - Z_9(e^{i\theta_{93}} - 1) + Z_6(e^{i\theta_{63}} - 1) &= 0 \\
 Z_1(e^{i\alpha_4} - 1) + Z_2(e^{i\theta_{24}} - 1) + Z_5(e^{i\theta_{54}} - 1) - Z_9(e^{i\theta_{94}} - 1) + Z_6(e^{i\theta_{64}} - 1) &= 0
 \end{aligned} \tag{3}$$

Loop 4:

$$\begin{aligned}
 Z_1(e^{i\alpha_j} - 1) + Z_2(e^{i\theta_{2j}} - 1) + Z_5(e^{i\theta_{5j}} - 1) + (Z_7 - Z_9)(e^{i\theta_{7j}} - 1) - Z_8/2(e^{i\theta_{8j}} - 1) \\
 + Z_{11}(e^{i\theta_{11j}} - 1) &= \delta_j, \quad J = 2,3,4 \\
 Z_1(e^{i\alpha_2} - 1) + Z_2(e^{i\theta_{22}} - 1) + Z_5(e^{i\theta_{52}} - 1) + (Z_7 - Z_9)(e^{i\theta_{72}} - 1) - Z_8/2(e^{i\theta_{82}} - 1) \\
 + Z_{11}(e^{i\theta_{112}} - 1) &= \delta_2 \\
 Z_1(e^{i\alpha_3} - 1) + Z_2(e^{i\theta_{23}} - 1) + Z_5(e^{i\theta_{53}} - 1) + (Z_7 - Z_9)(e^{i\theta_{73}} - 1) - Z_8/2(e^{i\theta_{83}} - 1) \\
 + Z_{11}(e^{i\theta_{113}} - 1) &= \delta_3 \\
 Z_1(e^{i\alpha_4} - 1) + Z_2(e^{i\theta_{24}} - 1) + Z_5(e^{i\theta_{54}} - 1) + (Z_7 - Z_9)(e^{i\theta_{74}} - 1) - \\
 Z_8/2(e^{i\theta_{84}} - 1) + Z_{11}(e^{i\theta_{114}} - 1) &= \delta_4
 \end{aligned} \tag{4}$$

The above equations are constructed such that four precision points are considered. Consequently, this yields six independent equations for each loop and a total of 24 independent nonlinear equations to solve, with 43 unknowns (namely; the 11 two dimensional vectors, which represents the links used in the mechanism namely: $Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, Z_9, Z_{10}, Z_{11}$, and the links corresponding angles: α_j (corresponding to link $O_A A$ (namely link Z_1)), θ_{2j} (corresponding to link AD (namely link Z_2)), θ_{3j} (corresponding to link AB (namely link Z_3)), θ_{4j} (corresponding to link BB'

(namely link Z_4), CB' (namely link $Z_6/2$) and CO_B (namely link Z_6)), θ_{5j} (corresponding to link DD' (namely link Z_5), CD' (namely link Z_9) and CE (namely link Z_7)), θ_{8j} (corresponding to link FE (namely link Z_8), FH' (namely link $Z_8/2$) and $H'H$ (namely link Z_{11})), θ_{10j} (corresponding to link $O_B F$ (namely link Z_{10})) for $j = 1,2,3$). All the angles are measured clockwise from the positive x -axis. $\delta_j, j = 2,3,4$, represent the precision points (for point H, which represent the motion of the foot joint), are prescribed based on desired values of the horizontal and vertical strides. Notice that increasing the number of independent equations by increasing the number of precision points will increase the number of unknowns and hence increase the complexity of the problem.

4. Modeling and Results

Matlab software package is used to solve the synthesis problem such that a horizontal (vertical) stride of no less than 15cm (5cm) is required as illustrated in the ovoid path described in Figure 1. To limit the search, the leg segments are selected proportional to that of an adult human [23] such that $O_A O_B \leq$ the 95th percentile of the thigh clearance of males ≈ 19 cm, $CE \approx 45$ cm, and $O_B C \leq 10$ cm and $O_B C \geq FE$. Moreover, $CD' \leq CE/3$. These constraints guide the selection of free choice variables and are translated into the following additional equations:

$$|Z_7 + Z_9| = |Z_{10}| = |Z_{11}| \cong 45cm \tag{5}$$

$$\angle Z_7 = \angle Z_9 \tag{6}$$

$$|Z_1 + Z_3 + Z_4 + Z_6/2| \cong 19cm \tag{7}$$

$$|Z_6| \leq 10cm \tag{8}$$

$$|Z_6| \geq |Z_8| \tag{9}$$

$$|Z_9| \leq |Z_7/3| \tag{10}$$

Figure 4 illustrates sample Matlab results at different crank angles where the fibula is longer than the femur such that $Z_1=1.4, Z_2=21.0, Z_3=10.7, Z_4= 8.4, Z_5=3.5, Z_6=8.4, Z_7=27.6, Z_8=7.6, Z_9=17.4, Z_{10}=45.0$, and $Z_{11}=76.3$. The three diagrams to the left represent positions of the leg at specific crank angles and the one to the right illustrates the performance of the leg over a discrete range of crank angles. Other sample results obtained using Matlab are simulated using Working Model software package. Figure 5 illustrates an alternative leg model that typifies the shape of the cylindrical human leg. The model is tested by allowing a trace of the path of the foot joint as illustrated in Figure 6. The figure shows clearly that the generated path is comparable to the desired

one illustrated in Figure 1. The accuracy of the path can be further enhanced by increasing the number of precision points during calculations. Figure 7 illustrates another alternative in addition to the corresponding bath. The figure illustrate that the motions of the various portions of the proposed leg typify that of the human leg. To illustrate, notice the motions of the knee and foot joints produced using small motion increments of the upper portion of the leg.

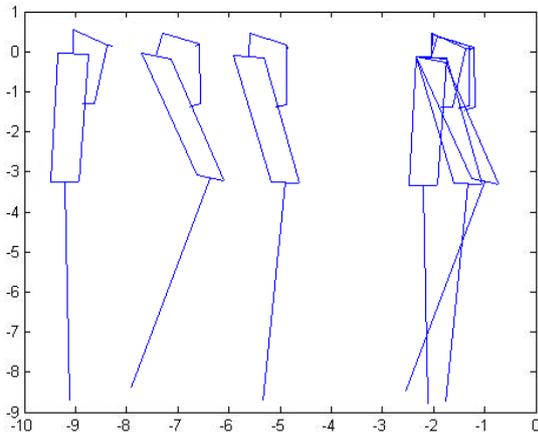


Figure 4: Sample Matlab results of the mechanism.

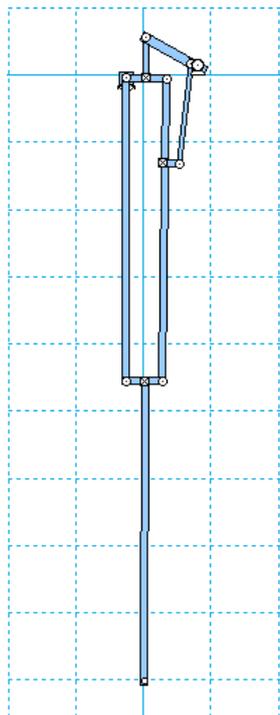


Figure 5: Simulated leg mechanism.

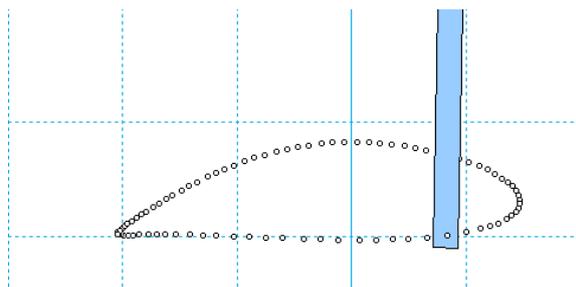


Figure 6: An illustration of the generated gait.

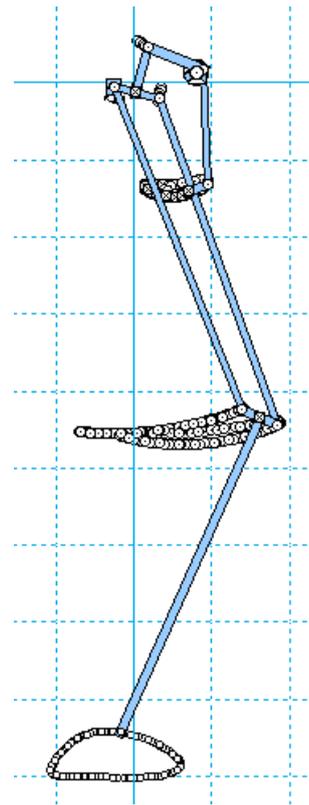


Figure 7: Alternative simulated leg mechanism.

5. Concluding Remarks

The paper presents an eight-bar single DOF path generator that typifies the shape and motion of a human leg. Considered design specifications include the slenderness of the leg and the shape of the walking gait. Moreover, the actuator of the mechanism is located in the upper portion of the linkage similar to it in a human leg. The proposed mechanism is suitable for the fabrication of legged robots. Proportions of the linkage are estimated utilizing anthropometric measures of the human leg. Matlab and Working Model software packages are used to simulate and validate the usability of the mechanism. The proposed mechanism demonstrates that a one DOF closed loop mechanical linkage can be designed to the shape and movement of the biped human walking apparatus.

References

- [1] Weber W., Weber E., Mechanics of the human walking apparatus. Springer; 1992.
- [2] Al-Araidah O., Analysis and design of two-degree of freedom pantograph type leg mechanism. Masters Thesis, Department of Mechanical Engineering, Eastern Mediterranean University, TRNC, Turkey; 1998.
- [3] M. Gracia, A. Chatterjee, A. Ruina, M. Coleman, "The simple walking model: stability, complexity, and scaling". Journal of Biomechanical Engineering, Vol. 120, 1998, 281-288.
- [4] V. Vakili, L. Shu, "Toward biomimetic concept generation". Design Theory and Methodology Conference, Pittsburgh, PA, 2001.

- [5] W. Batayneh, O. Al-Araidah, C.A. Mattson, A. Ismail-Yahaya, "Design and implementation of human-like biped walking mechanism". 3rd International Conference in Mechatronics, Kuala Lumpur, Malaysia, 2008.
- [6] W.B. Shieh, L.W. Tsai, S. Azarm, A.L. Tits, "Multi-objective optimization of a leg mechanism with various spring configurations for force reduction". NSF Technical Research Report, University of Maryland, Harvard University and Industry, 1996.
- [7] W.B. Shieh, L.W. Tsai, S. Azarm, "Design and optimization of a one-dof six-bar leg mechanism for a walking machine". Journal of Robotics Systems, Vol. 14, No. 12, 1997, 871-880.
- [8] V. Sangwan, A. Taneja, S. Mukherjee, "Design of a robust self-excited biped walking mechanism". Mechanism and Machine Theory, Vol. 39, 2004.
- [9] S. Mukherjee, V. Sangwan, A. Taneja, B. Seth, "Stability of an underactuated bipedal gait". Biosystems, Vol. 90, 2007, 582-589.
- [10] B. Verrelst, B. Vanderborcht, J. Vermeulen, R.V. Ham, J. Naudet, D. Lefeber, "Control architecture for the pneumatically actuated dynamic walking biped "Lucy"". Mechatronics, Vol. 15, 2005
- [11] K. Hosoda, T. Takuma, A. Nakamoto, S. Hayashi, "Biped robot design powered by antagonistic pneumatic actuators for multi-modal locomotion". Robotics and Autonomous Systems, Vol. 56, 2008, 46-53.
- [12] S.H. Collins, A. Ruina, "A Bipedal walking robot with efficient and human-like gait". IEEE International conference on robotics and automation, Barcelona, Spain, 2005.
- [13] Y. Yavin, "Modeling and control of a walking four link robot". Mathematical and Computer Modeling, Vol. 35, 2002.
- [14] G. Capi, S. Kaneko, K. Mitobe, L. Barolli, Y. Nasu, "Optimal trajectory generation for a prismatic joint biped robot using genetic algorithms". Robotics and Autonomous Systems, Vol. 38, 2002.
- [15] K. Hirai, M. Hirose, Y. Haikawa, T. Takenaka, "The development of Honda humanoid robot". IEEE International Conference on Robotics and Automation, Leuven, Belgium, 1998.
- [16] M. Hirose, Y. Haikawa, T. Takenaka, K. Hirai, "Development of humanoid robot ASIMO". IEEE/RSJ International Conference on Intelligent Robots and Systems, Maui, USA, 2001.
- [17] K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, M. Hirata, K. Akachi, T. Isozumi, "Humanoid robot HRP-2". IEEE International Conference on Robotics and Automation, New Orleans, USA, 2004.
- [18] Y. Kuroki, T. Ishida, J. Yamagushi, M. Fujita, T. Doi, "A small biped entertainment robot". IEEE/RAS International Conference on Humanoid Robots, Piscataway, NY, 2001.
- [19] H. Hirukawa, F. Kanehiro, K. Kaneko, S. Kajita, K. Fujiwara, Y. Kawai, F. Tomita, S. Hirai, K. Tanie, T. Isozumi, K. Akachi, T. Kawasaki, S. Ota, K. Yokoyama, H. Handa, Y. Fukase, J. Maeda, Y. Nakamura, S. Tachi, H. Inoue, "Humanoid robotics platforms developed in HRP". Robotics and Autonomous Systems, Vol. 48, No. 4, 2004.
- [20] Y. Hurmuzlu, F. Genot, B. Brogliato, "Modeling stability and control of biped robots-a general framework". Automatica, Vol. 40, 2004, 1647-1664.
- [21] P. Mousavi, A. Bagheri, "Mathematical simulation of a seven link biped robot on various surfaces and ZMP considerations". Applied Mathematical Modelling, Vol. 31, 2007, 18-37.
- [22] R. Barai, and K. Nonami, "Optimal two-degree-of-freedom fuzzy control for locomotion control of hydraulically actuated hexapod robot". Information Science, Vol. 177, 2007, 1892-1915.
- [23] Eastman Kodak Company, Ergonomic Design for People at Work, Volume1, Lifetime Learning Publications, Belmont, CA., 1983.

A Diverse Neural Network Ensemble Team for Mean Shift Detection in X-Bar and CUSUM Control Charts

M. A. Barghash ^{*,a}

^aIndustrial Engineering Department, Faculty of Engineering and Technology, University of Jordan, 11942, Amman, Jordan

Abstract

In manufacturing processes, maintaining quality is associated with proper process mean parameters and product quality metrics. The early detection of mean changes is important to reduce the number of defectives or non-conformities in the production. In this work, a diverse ensemble of Artificial Neural Networks (ANNs) with a leader network have been used to achieve this purpose, then a performance comparison was conducted on two types of control charts: X-bar and CUSUM in addition to comparing it to individual neural network performance. The traditional and individual neural network performances were obtained from published literature. It was found that, the diverse ensemble of ANNs detects small shift in process mean far earlier (Shorter Average Run Length (ARL)) than individual ANN's, traditional X-bar and CUSUM control charts. The postulated reason for this improvement is that the ensemble ANN system analyses more than one sample point, rather, it considers the points pattern and overcomes the instabilities in individual ANN's.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: Neural networks ensemble; Artificial intelligence; Pattern recognition; team of networks; diversified artificial neural networks; X-bar control charts; CUSUM control charts

1. Introduction

1.1. Quality control and supervisory process control:

Quality control charts are graphical tools commonly used to discover patterns and assignable causes in the manufacturing processes. They help manufacturers to discover the root cause of the problem and to reduce the number of defective products. X-bar charts and CUSUM control charts play an important role in maintaining quality.

Within the scope of mean shift detection, the basic idea for quality process control analysis is to continuously sample the manufacturing process, where for each new sample collected, one of two binary decisions can be made: the process mean did not shift "Normal or no-shift" or the mean had changed "process not normal or shift existing". Also, two types of mistakes can be made, particularly, no-shift decision while the process mean had changed and a shift decision with no change in the process mean existing (false alarm). If the decision is no-shift, the sampling process is continued however, if the decision is a shift then the process is stopped and an attempt is made to discover the reason for the change in the process mean.

1.2. Traditional control charts development:

Shewart developed the basic X-bar control charts in the early 1920s. Among the earliest work in X-bar control charts was devoted to heuristics and zone testing [1-2]. The developed rules enhances the ability to discover mean

shifts but also reduces the Average Run Length (ARL) when there is no-shift in the process that is, it increases the number of false alarms. The X-bar control charts perform well for large shifts, While, it is less effective for small shifts. This is due to the memory-less property of the Shewart X-bar control chart. Developed by Page in 1958 , the CUmulative SUM (CUSUM) control chart were introduced to include all previous sample points through summation. Roberts proposed the Exponentially Weighted Moving Average (EWMA) control charts [3]. It also includes a weighted sum of all previous sample points where the importance of any point decline over time by a selected factor. Both charts include a memory of previous points. But while the sum increases the power of discovering small shifts, it dilutes the effect of large shifts in the mean with previous non-shifted points and it also delays large shifts from being discovered. The CUSUM and EWMA detect small shifts earlier that Shewart X-bar control charts, while the X-bar chart outperforms the CUSUM and EWMA charts for large shifts. Lucas developed a V-mask for the CUSUM control chart to improve its abilities[4-5].

1.3. Artificial intelligence based methods:

One main Limitation for the traditional methods is that the basic X-bar and CUSUM charts (no V-mask or sensitizing rules) look only at the last point and not at the pattern. This is traditionally solved by adding sensitizing rules (i.e., 2 of 3 consecutive points outside 2-sigma limit, 4 of 5 consecutive points outside the 1-Sigma limit, 8 consecutive points on one side of the center line, etc). The developed sensitizing rules –if used- have a major

* Corresponding author. e-mail: mabargha@yahoo.com

drawback. In particularly, it increases the number of false alarms.

There is a need for a technique to analyze more than just the last point without increasing the number of false alarms. This raised the reason why many researchers turned to the application of artificial intelligence.

Several researchers have utilized Artificial Intelligence (AI) in control charts. AI methods can handle a specified number of input points at the same time. It categorizes the shift and no-shift cases into two different patterns. Training population is then prepared through computer process simulation or through process sampling. This is then used to train the neural network. That is, to estimate the different parameters (weights and biases) of the neural network. The neural network will then make the decision as shift and no-shift for runs other than those used for the training process. Although, the neural network will give perfect and accurate results for the population used in training, it might not be as successful for new population that has not been tested yet. This is due the infinite number of possible shapes of patterns (being a random process).

Another important perspective of AI technique is that it is mostly a continuous function. The output of the AI is usually a number or a score for each pattern. The usual procedure is to compare this output to a threshold, if the score is greater/less than the threshold, then the pattern is classified as a shift pattern. Threshold is usually set to achieve a low percentage of false alarms.

Early studies compared neural networks and Shewhart control charts for both large and small shifts in the mean and reported that neural networks are comparable to Shewhart X-bar and R-charts for large shifts and outperform them for small shifts [6-7]. Several artificial Intelligence tools were used including: Multi-layer perceptron [6-9], Linear Vector Quantization[10], Probabilistic neural networks[11], Modular neural networks [12], Adaptive Resonance Theory [13], Neuro Fuzzy techniques[14-17], Fuzzy technique[18-21], Support Vector Machine [22], Self-Organized Approaches[23] and Competitive Neural Networks[24].

The basic types of Artificial intelligence techniques, which are referred to "Individual" in this paper have outperformed the traditional techniques, but as will be illustrated shortly in section 5.3, the magnitude of difference was small. This might put some doubt on the abilities of the Artificial Intelligence techniques. However, one reason for this humble performance of individual type control charts is instabilities in performance of the AI techniques[25]. This can be overcome and heavily improved by group type or "ensemble type" Techniques, where multi-stage decision is made as will be illustrated in this paper.

1.4. Instabilities in neural networks and artificial intelligence:

Instability is a feedback control concept. A stable system reaches a target value, while unstable systems oscillate (entails instabilities) and do not reach their required target. Figure (1) shows a sample behavior of ANN's. The upper chart is the process sample pattern, while the lower part is the decision of the ANN. It can be noticed that, the sample ANN in Figure (1) changes its decision with time. This is a form of instability noted in individual ANN's. However, this type of behavior can be

eliminated in group type or ensemble type ANN systems as will be shown in this work.

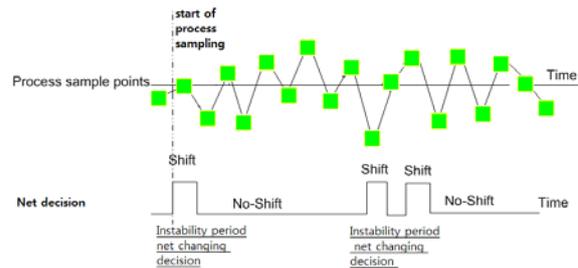


Figure 1: A sample behavior of ANN's.

Instabilities have direct effect on the number of false alarms. Usually, the threshold for classification is modified to reduce the number of false alarms and thus, reducing -as a consequence- the ability of the neural network to discover small shift.

In this work, the instabilities are handled differently. By tolerating the instabilities initially and keeping the threshold used for classification at 0.5. Thus, the instabilities will occur at the initial stage of the ANN, and then another ANN stage is used for classification. The initial stage ANN, and for the case of large shift size, the neural network will be mostly giving a shift decision, but some time a no-shift decision can occur, as for the small shift, most likely the decision will be a no-shift decision but a shift decision is likely to occur. The second stage ANN needs to see more than one decision from the initial stage to make the final decision.

1.5. Ensemble type artificial neural networks:

To reduce the effect of ANN instability and to reach to smooth, robust and stable decision, it is possible to utilize neural networks in a hierarchical manner similar to organizations charts, where a group of "Individual Neural Networks" assess process pattern and make initial decisions (Figure (2)). Later, other networks are used to obtain the final decision. It is possible - at the initial stage- to have multiple decisions where some ANNs signal for a shift and some do not. This is called a neural networks ensemble. Instead of handling the original pattern, the initial neural networks provide their analysis to the following neural networks.

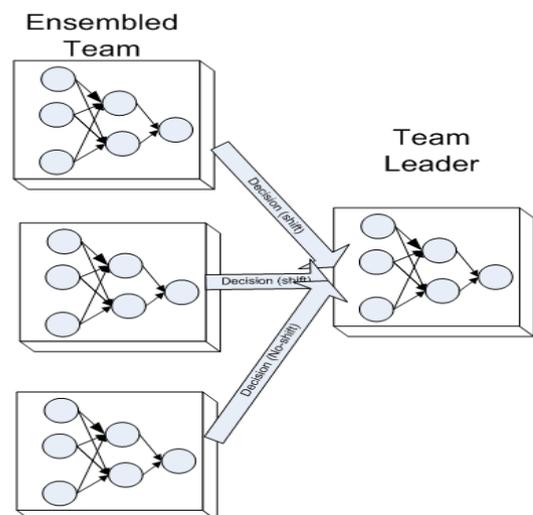


Figure 2: An ensemble team of ANN's

Ensemble based neural network schemes have been used by several researchers [26-28]. With strong emphasis on ensemble diversity [22]. The collective method can then be: plurality voting, majority voting, winner takes all for classification purposes [24]. The diversity of the neural networks is achieved through varying the topology and initial conditions or algorithm involved. The training for the individual networks can be done by sequential (Ada-Boost) or parallel training (Bagging). The sequential method allows the initial nets to learn the training set. The outputs of the initial nets are the training set for the next group of nets and so on. The Bagging method diversifies neural networks by randomly selecting the input set for each network, then making a collective opinion for these sets [23]. ANNs are extremely affected in their decision by population percentage used for the training [16]. Thus can be used for diversification of the neural networks [15]. Such diversified ANN's are trained using different pattern percentages had shown tendency to have an error in judgment titling towards the pattern available more in the training population [15]. Figure (3) shows an illustrative example of neural network diversity, the process encounter a shift pattern. The first, second and third neural network respond in a different style. The three neural networks encounter instability, in that, they don't have a stable target decision, and instead, they fluctuate. However, the key difference between the three networks is in the number of shift decisions for a specified number of sample points. This is largest for the third neural network and smallest for the first.

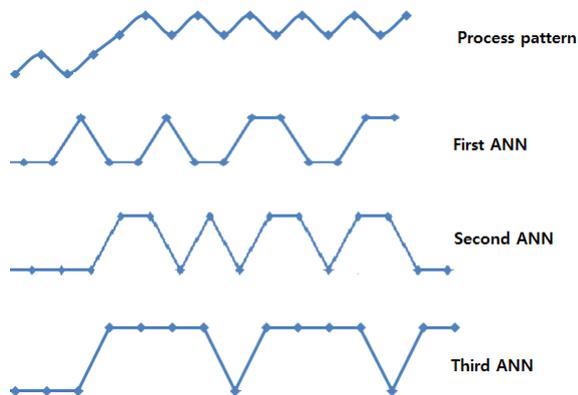


Figure 3: An example of diverse neural network response.

The process variations are the deviations in the process and product parameters. They include both random and assignable causes of product variations. Diversity, on the other hand, is a recent trend in algorithm design accompanied with population based algorithms such as genetic algorithms. A survey of population diversity measures using genetic programming (GP) is done by Burke [29]. Maintaining population diversity in genetic algorithms improves their performance.

Diversity is viewed as an enhancing addition to algorithms and in this work diversity is used to form a team of neural networks. Diversity in this sense is the diversity of the decision makers. Maintaining diversity is important to enhance the decision process. The diversity and its analysis had been proved to be critical in fields such as genetic algorithms; however, the potential of this important concept is not yet fully explored.

1.6. The focus in this paper:

In this work, a diverse team of neural networks is utilized to make an initial decision on shifts in control chart means. A leader neural network then makes the final decision using the initial decisions as inputs. Two types of charts are considered X-bar and CUSUM control charts. The CUSUM of input pattern tends to amplify the shift with the addition of samples. The performance of the ensemble is fine-tuned through a prior scaling factor and threshold based classification. This paper proves that teaming up diverse ANNs and including a leader results in a lower Average Run Length (ARL) and thus a lower Average Time-To-Signal (ATS) than traditional CUSUM control charts.

2. Methodology and Model Construction

2.1. The basic artificial neural network:

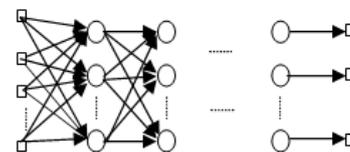
A neural network is composed of several interconnected neurons with the following activation function for each neuron [32].

$$output = \phi \left(\sum_{i=1}^p w_i x_i - \theta \right) \tag{1}$$

Where

- ϕ : is the neuron function
- w_i : is the neuron i th weight
- x_i : is the i th input to the neuron
- θ : is the constant threshold value
- p : number of inputs to the neuron.

Neural networks have high interconnectivity, where all the neurons are arranged in layers. The outputs of the first layer are the inputs for the second layer as shown in Figure (4). The first layer is the input layer and the final layer is output layer. The layers in between are called hidden layers.



Input layer Hidden layers Output layer

Figure 4: Layers of neural networks.

Training of the neural network is done using back propagation where the error value and the local gradient of the error are used to estimate the new value of the weights and biases.

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad (2)$$

Where

- W_{ji} : is the synaptic weight connecting the output of neuron i to the input of neuron j .
 η : is the learning rate
 $\delta_j(n)$: is the local gradient for the error in output neuron j at iteration n .
 $y_i(n)$: is the output of the network at neuron i at iteration n .
 n : iteration number. Several types of neuron activation functions from Mathworks Matlab 7.0 were used including the LogSig sigmoid function, TanSig and PureLin as shown below

$$\text{Logsig}(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

$$\text{Purelin}(x) = b + x \quad (4)$$

$$\text{Tansig}(x) = \text{sign}\left(\frac{2}{1 + e^{-2x}} - 1\right) \quad (5)$$

2.2. Diverse team of nets:

2.2.1. Diversity of the team:

Pattern percentages have been used to achieve ANN's diversification as shown in Table 1. Column 2 shows the percentages of the patterns for each of the neural networks given in column 1.

Table 1: Population composition and training error for the ANN's.

NET	Training Population N: Normal (no shift) S: Shift	Training Output N: Normal (no-shift) S: Shift	Average of Square Error
1	(50%, N; 50%, S)	(1, N; 0, S)	$4.01 * 10^{-5}$
2	(75%, N; 25%, S)	(1, N; 0, S)	$3.23 * 10^{-5}$
3	(90%, N; 10%, S)	(1, N; 0, S)	$4.21 * 10^{-6}$
4	(50%, S; 50%, N)	(1, S; 0, N)	$5.32 * 10^{-5}$
5	(75%, S; 25%, N)	(1, S; 0, N)	$4.25 * 10^{-5}$
6	(90%, S; 10%, N)	(1, S; 0, N)	$5.12 * 10^{-5}$

2.2.2. ANN configuration and training:

The number of samples or input size is of little significance. The number of input samples has little impact on the performance of the ANN[16]. Thus, an arbitrary input size of 20 was satisfactory for the purpose of this paper.

The construction of the ANN's was in four layers: input layer, two hidden layers and one output layer. The size of the layers (number of neurons) is not highly significant[16]. Increasing the layer size more the 15 does not improve the performance. However, if we reduce it appreciably below 15 the performance may deteriorate.

The selected size for the number of neurons is 15. The layers were composed of LogSig functions for the input layer and the second layer, TanSig for the third layer and PureLin for the fourth layer.

The training average sum of square errors for the different ANNs is shown in Table 1-column 4. The output of the neural networks is shown in column 3. Any deviation from these values will be considered as error and is included in the sum of squares error in column 4.

As is shown in column 4, the ANN's are trained well to a very small sum of square error and can perform extremely well to classify the training population. But the ANN might not be successful in classifying other populations. The average sum of square error is a measure of the training quality for the individual ANN. But it is not a measure of its decision quality.

2.3. Leader network:

The leader network is a neural network that accepts the outputs of the different neural networks and is trained to make the final decision. The leader network is trained in the same manner as the rest of the networks, however, the input to the team leader is the decisions of the other networks. The configuration of the leader ANN's consists of four layers: input layer, two hidden layers and one output layer. The LogSig function was used for the input layer and the second layer, TanSig was used for the third layer and PureLin was used for the fourth layer.

A threshold level classification for the output of the leader net is used. Patterns getting an output of less than a selected threshold value are classified as "no-shift" (normal) and those getting above the threshold are classified as "Shift" patterns. The threshold level classification is the most used classification for the output of the neural nets among the researchers. In this work the threshold of classification was chosen to establish a near-value to the traditional X-bar chart and traditional CUSUM charts for the no-shift case.

2.4. Ensemble construction:

2.4.1. X-bar type ensemble:

As discussed earlier ANN's suffer from instabilities. The ANN seems to change their decision for the same pattern as more sample points are collected. Thus, in this work, three consecutive decisions for each individual neural network are used. These are then submitted as inputs to the leader pattern (As shown in Figure (5)). The size of the input to the leader pattern is 18 (see Figure (5)). To smooth the output of the Leader-net, a 2-point moving average is used. This procedure is explained through an example later in section 4.0. The 2-point moving average is effective in smoothing the output of the ensemble.

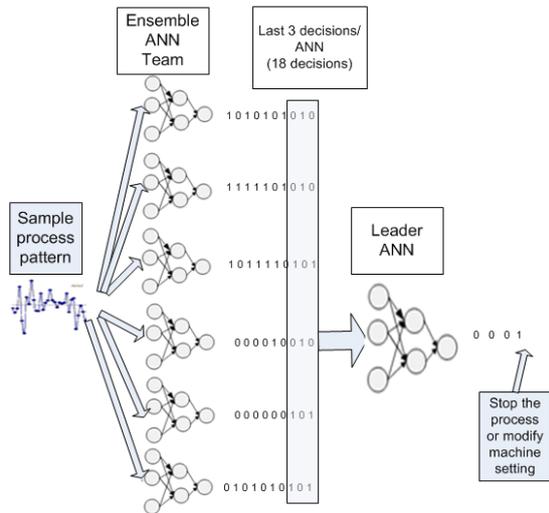


Figure 5: The framework of the ANN ensemble for mean shift detection in X-bar control charts.

2.4.2. CUSUM based ensemble:

Figure (6) shows the framework for the mean shift detection. The data from the process are standardized and summed according to the following equations:

CUSUM ($S_0 = 0$)

$$S_i = \bar{X}_i + S_{i-1} \tag{6}$$

$$\text{Normalized CUSUM} = \frac{S_i - \mu}{\sigma_{\bar{x}}} \tag{7}$$

Where

i : is the sample number

The Normalized CUSUM is divided by a factor “A” before passing them to the ensemble of nets. A is a tuning parameter for the performance of the ensemble.

$$Q = \text{Normalized CUSUM} / A \tag{8}$$

For the initial networks, the input size is 20 process sample points for each decision. Three consecutive decisions / net are assembled for the input to the leader net. Altogether 18 inputs are used for the leader net. To smooth the leader net output a “two-point moving average” is used. A threshold based classification is then used. If the smoothed output of the leader net is below the threshold process, the process is considered normal (i.e., 0), otherwise, a shift (i.e., 1) decision is made.

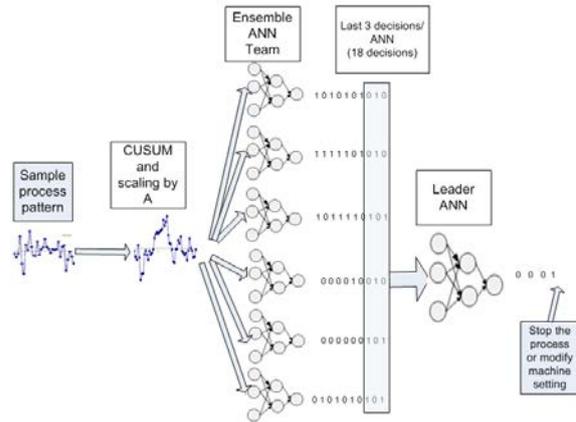


Figure 6: The framework for Ensemble ANN mean shift detection in CUSUM control charts.

3. Pattern Generation

Equation (3) is used to generate the shift and no-shift patterns:

$$\text{Generated pattern } y(i) = \mu + n(t) + d(t) \tag{9}$$

where

- μ : is the process mean
- $n(t)$: is the normal cause of variations which is $N(0, \sigma)$.
- $d(t)$: is the special cause of variations

For the case of a shift:

$$d(t) = u * s$$

Where:

- u : is the position of the shift (it can be either 0 or 1)
- s : is the value of the shift in the mean.

The following standardization was chosen for the patterns:

$$Y(i) = \frac{y(i) - \mu}{\sigma_{\bar{x}}} \tag{10}$$

The final variable Y(i) is a standardized variable with mean zero and a standard deviation of one for the no shift case. The mean shift values (s) can be expressed in a standardized form: as 4, -3, -2,-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1, 2, 3 and 4 or they can be expressed as a multiple of the standard deviation. The “s” variable is set to zero for the no-shift case.

4. An Illustration of the Suggested Model for the X-bar Chart

In this section, the basic mechanics of the suggested model (shown in Figure (5)) will be illustrated. I will use a numerical example with the following assumptions: Those were selected to eliminate the need for the standardization step. I will consider two cases: no-shift case $d(t)= 0$, and shift case with $d(t)=0.25$. The function in MATLAB generates normally distributed data with a mean of zero,

and a standard deviation of one. The chosen input size for the neural network ensemble is 20, thus 20 sample points need to be generated.

Iteration 1 samples 1-20:

Using $y = \text{randn}(20,1)$; to obtain 20 sample points:

$y(1), \dots, y(20) = \{-0.1172, 0.9572, 0.3863, 0.5589, 0.9682, -0.1059, -1.6905, -1.2510, -0.8900, 0.3454, 0.7927, 0.2004, -0.2301, -0.5378, 1.8070, -0.8698, 0.0232, -0.1273, 0.8690, -0.2178\}$.

This data acts as the input for the individual ANNs. Table 2 shows the decisions of the individual nets. These values are obtained using the function "sim(network_name, data_vector)". I cannot still submit these inputs to the leader net because two more consecutive decisions / net are required. So, I move to the next two iterations.

Table 2: Iteration 1 results.

NET	Output	Classification		Decision after classification
		>0.5 1 shift	<0.5 0 no-shift	
1	0.4821	Shift		0
2	0.7995	No-shift		1
3	0.9970	No-shift		1
4	-0.0038	No-shift		0
5	-0.2756	No-shift		0
6	0.9063	Shift		1

Iteration 2 and Iteration 3:

For iteration 2 and 3, two new sample points are generated using randn():

$y(21)$ and $y(22) = 1.1071, -0.0009$.

With each new point, the newly generated point is added at the end of the already generated pattern in iteration 1 and remove the first point. The number of inputs to the individual neural network should remain 20. The input to the networks in iteration 2 becomes: $\{0.9572, 0.3863, 0.5589, 0.9682, -0.1059, -1.6905, -1.2510, -0.8900, 0.3454, 0.7927, 0.2004, -0.2301, -0.5378, 1.8070, -0.8698, 0.0232, -0.1273, 0.8690, -0.2178, 1.1071\}$.

The modifications to iteration 2 are shown in bold font. For the third iteration it becomes: $\{0.3863, 0.5589, 0.9682, -0.1059, -1.6905, -1.2510, -0.8900, 0.3454, 0.7927, 0.2004, -0.2301, -0.5378, 1.8070, -0.8698, 0.0232, -0.1273, 0.8690, -0.2178, 1.1071, -0.0009\}$.

Table 3 shows the individual neural network decisions for iterations 2 and 3.

Table 3: Iteration 2 and 3 results.

NET	Iteration 2		Iteration 3	
	Output	Value for the decision	Output	Value for the decision
1	0.1813	0	-0.1959	0
2	1.0486	1	1.1628	1
3	0.9974	1	0.9998	1
4	0.0026	0	0.0011	0
5	-0.0486	0	-0.1628	0
6	0.8187	1	1.1959	1

At the end of iteration 3, enough data will be available to start activating the leader network. The input to the leader net is formed as row-by-row combination of the decision columns combining Tables 3 and 4, that is: $\{0,0,0,1,1,1,1,1,1,0,0,0,0,0,1,1,1\}$.

Iteration 4 (final iteration for the no-shift case)

We then use the MATLAB function `sim(leader_net,data_vector)` to obtain the output of the leader net. MATLAB result is given as 0.0142. This value is below the threshold value and is classified as no-shift pattern.

For the shift pattern case, I have randomly selected 0.25σ or 0.25 standard shift starting at the 20th sample.

Iteration 2 (following iteration 1):

The variables u, s, d and y in equation (3) are given as follows, note that the changes to iteration 1 are shown in bold:

$u = [0\ 1\ 1\ 1]$,

$s = 0.25$,

$d = [0\ 0.25\ 0.25\ 0.25]$,

$y = [-0.1172, 0.9572, 0.3863, 0.5589, 0.9682, -0.1059, -1.6905, -1.2510, -0.8900, 0.3454, 0.7927, 0.2004, -0.2301, -0.5378, 1.8070, -0.8698, 0.0232, -0.1273, 0.8690, 0.0322, 0.3571, 0.2491]$.

The same procedure shown above for the no-shift case is then used, by selecting sample points 1-20, to obtain the individual networks decision, then the sample points 2-21, and sample points 3-22. The inputs for the leader net are obtained in the same manner from the networks individual decision. Then the leader network can be used to make the final decision.

5. Results and Discussion

The results part of this work is composed of three sections: the first section analyses the output of the individual networks and the leader net. The second section analyses and optimizes the threshold and the third part compares the optimized results obtained to those of the traditional X-bar chart and individual neural networks.

5.1. Analyzing the output of the neural networks:

The intention of this section is to analyze the output of the individual networks having a threshold level equal to 0.5 for the individual nets and a level of 0.85 for the leader net. A standardized shift of 1 is chosen for the analysis because it has a relatively short ARL. The mean shift occurs at the sample number 21.

Figure (7) shows a sample run results for the suggested model. Figure (7)-a shows the sample points generated. It clearly shows that a mean shift has occurred around sample number 21. Figure (7- b, c and d) shows the response of individual Networks 1-3. These are classifying the pattern: as “1” for no-shift case and “0” for the shift case (see Table 2). Some instability is observed in their decisions. It is clearly noted in Net 2 which had changed its decision 6 times (i.e., 1 → 0 and 0 → 1). This instability reduces the benefit of individual networks for pattern classification and definitely justifies the need for an ensemble model. Figure (7- e,f and g) shows the decision of the Nets 4,5 and 6. Instabilities are noted also in the behavior of the Nets 4,5 and 6. Instabilities are noted also in the behavior of Net 5. Figure (7-h) shows the output of the leader net. The output of the leader net is certainly stable and consistent and no instabilities are noted in its behavior. The value of the output exceeds the threshold level at sample 27 and a sample run length of 6 is recorded for this case (the run started at sample 21 and is detected at sample 27).

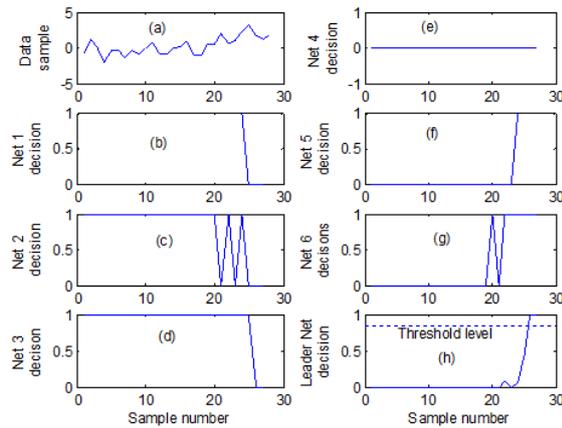


Figure 7: Output for the individual and the leader nets for a sample pattern for a standardized mean shift size of 1.

High variability is noticed at some sample points. Sample 21 -for example- has one of the highest decision variability where the Nets 1,3 and 6 clearly state it is a shift pattern, while the other Nets considers it as no-shift pattern. Sample 27 on the other hand, has the lowest variability. This conflicting decision behavior also advocates the need for an ensemble based technique rather than individual neural network.

It can be noticed that Net 2 and Net 6 have discovered the change in the process mean rather early. Their first signal for a change in the process mean comes in sample 21. The ensemble on the other hand detected the shift at sample 27. However, these two individual nets suffer from extreme instabilities.

5.2. Output analysis for suggested ensemble and threshold selection:

Table 4 shows the ARL results for several shift patterns. Column 1 is the standardized value for the shift and columns 2 to 5 are selected results at different threshold levels.

Table 4: ARL at different threshold level.

Shift	Threshold level			
	0.85	0.94	0.95	0.96
4.00	3	3.04	3	3.04
3.00	3.44	3.56	3.44	3.4
2.00	4.2	4.24	5.04	4.52
1.00	10.56	10.28	11.64	9.72
0.75	14.68	13.56	14.88	14.52
0.50	27.6	34.12	33.48	32.76
0.25	100.6	114.28	134.56	84.64
0 (no shift case)	235.52	237.32	294.28	416.28
-0.25	47.36	102.64	81.64	75.16
-0.50	22.64	20.68	20.12	37.52
-0.75	9.16	11.52	14.44	14.48
-1.00	9	9.72	9.36	8.76
-2.00	3.68	4.16	4.36	3.96
-3.00	2.44	3.2	2.88	3
-4.00	2.28	2.6	2.76	2.56

The results show that as the threshold increases, the ARL for the no shift case increases. However, there is slight change for the shift case.

5.3. Comparing the ensemble model to X-bar and individual ANN's:

Table (5) compares the ARL for traditional X-bar chart and individual ANN to the suggested ensemble model. The average run length for the best arrangement in Table 4 (column 5) is selected for comparison and is given in Table 5- column 4. It is compared to traditional techniques X-bar and individual neural network results. Column 2 and 3 are obtained for benchmarking purpose from Yi et al (2001).

Table 5: Comparison of the results of the X-bar chart, NN and the suggested ensemble.

Shift	X-bar chart 3-sigma limit	NN[30]	Selected ensemble results
4.00	1.19	1.19	3.04
3.00	1.98	1.97	3.4
2.00	6.28	6.17	4.52
1.00	44.9	43.48	9.72
0.75	83.27	79.48	14.52
0.50	157.1	151.09	32.76
0.25	280.16	272.85	84.64
0 (no shift case)	372.28	371.96	416.28
-0.25	282.48	292.57	75.16
-0.50	153.81	163.16	37.52
-0.75	43.94	45.87	14.48
-1.00	14.99	15.68	8.76
-2.00	6.31	6.51	3.96
-3.00	2.02	2.05	3
-4.00	1.18	1.19	2.56

Table (5)-column 4 shows that the suggested model is detecting small process mean shifts far earlier than both individual ANN and X-bar chart. Let's take for example the 0.25 standardized shift, here, the ARL for the suggested ensemble is 97.72 as compared to 272.85 for the individual NN and 280.16 for the X-bar 3 sigma limit. This represents a large difference and shows that the ensemble model is superior to the two other methods. It is also noticed that the ensemble model detects very large shifts such as 3-sigma or 4 sigma nearly one sample point afterward as compared to the other methods. This is due to using more than one decision point and using the moving average technique to smooth the output results.

5.4. Results for the CUSUM control chart:

Table 6 shows the various results that can be obtained from the CUSUM through varying the upper and lower control limits (CL factor).

Table 6: Comparison of tradition CUSUM and new the framework.

	Traditional CUSUM		Selected ensemble results	
	CL =4	CL=5	Threshold=0.09 A=2.8	Threshold=0.09 A=2.7
Standard mean shift size				
4	1.71	2.01	4.36	5.2
3	2.19	2.57	5.4	6
2	3.34	4.01	6.24	7.52
1	8.38	10.4	10.32	13.64
0.75	13.3	17	12.24	18.72
0.5	26.6	30	15.08	23.44
0.25	74.2	140	39.68	51.76
0	336	930	324.04	906.84
-0.25	74.2	140	33.2	38.12
-0.5	26.6	30	18.64	27.48
-0.75	13.3	17	15.2	15.72
-1	8.38	10.4	10.36	11.8
-2	3.34	4.01	7.24	8.64
-3	2.19	2.57	4.44	6.52
-4	1.71	2.01	4.6	5.36

Table (6) shows the results obtained for a traditional CUSUM for different control limits (columns 2, 3) and the suggested ensemble (columns 4, 5). It is noted that for low shift values the suggested model is highly superior to that of the traditional X-bar technique. Let us for example take the 0.25 standardized shift case and compare columns 3 and 5. The traditional CUSUM has a high ARL value of 140, while the suggested model has an ARL of 51.76. A large difference in value is noticed which shows the superiority of the suggested model.

5.5. A sample ARL distribution:

Figure (8) shows a sample ARL histogram for the CUSUM ensemble for a shift size = 0.5*sigma. ARL distribution is generally characterized with wide spread. The current ensemble ARL is not different. 500 ARL data points are used in the construction of the histogram in Figure (8). The mean ARL is nearly 23. However, an ARL more 4 time this amount is found (nearly 106). This is a general characteristic of ARL distributions and is noticed in traditional X-bar charts and in CUSUM control charts as well.

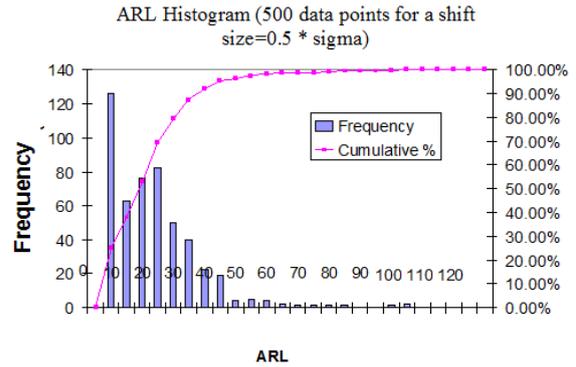


Figure 8: A sample ARL distribution for the CUSUM based ensemble with shift size=0.5*sigma.

6. Case study: Material Variations in plastics sheet

6.1. Introduction:

In this section, the developed model will be applied to plastics sheet extrusion. Plastics manufacturing is a batch type process and the batch-to-batch variations is a common disturbance in polymer manufacturing processes. If the material is highly variable then defective parts may be produced. Figure (9) shows the sheet extrusion process, The extrusion process melts the plastics pellets to a viscosity level V and the viscous form is pumped using a rotating screw extruder at head speed HS towards a heated die to temperature DT, the sheet shaped extruded material is pulled in between two rollers rotating at roller speed RS.

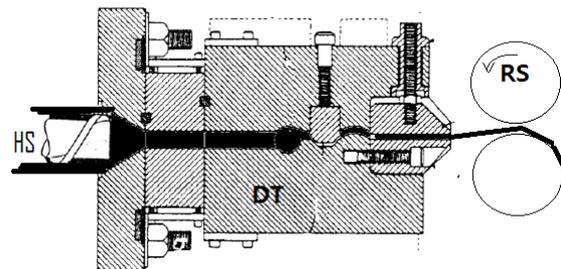


Figure 9: The sheet metal extrusion process schematic.

6.2. SIMULATOR software:

The following section uses simulation for the plastic sheet extrusion generated by the simulation package called SIMULATOR. Simulator performs Monte-Carlo simulation of products and processes for the purpose of training in designed experiments, robust design and tolerance analysis[31]. The Simulator software generates

random values of sheet thickness simulating actual process values. Although, the focus of the software is for robust process design, it can be used for control charts design.

6.3. Data Generation:

The sheet thickness is an important customer quality objective. A model relating thickness to the input parameters is found by the following expression[31].

$$TH = 64.93479 - 0.10766RS - 0.001058RS^2 + 0.48434DT - 0.001058DT^2 + 0.001084RS*DT + 5.0V + 0.0075HS + ERR \tag{5}$$

Where, RS: The Roller speed (rpm), DT: The Die temperature oF, V: The Viscosity (coef), HS: Head speed (rpm). The data is generated using the Simulator software randomly. The mean and the standard deviation of the data is estimated for normalizing purpose. The CUSUM framework is then used to analyze the data.

6.4. Viscosity variation detection:

The batch-to-batch variation is an important variation cause in sheet extrusion. It affects the product thickness and material strength. In this subsection, I have used the CUSUM based ensemble to detect the viscosity variations. Figure (10) shows the ARL for the CUSUM framework. The X-axis being the viscosity variation and the Y-axis represents the ARL required to detect the change. The threshold and scaling factor A are set to achieve a low false alarm rate.

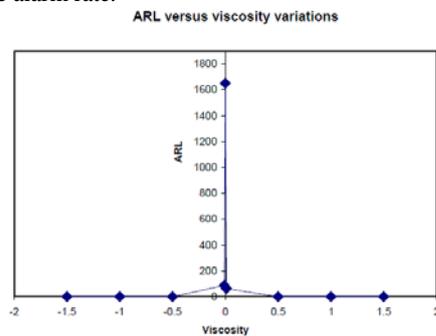


Figure 10: ARL versus viscosity variations for the sheet extrusion process.

Table 7: The tabulated results of ARL versus viscosity.

Viscosity variation	ARL
-1.5	1.5
-1.2	1
-0.5	2.9
-0.01	87
0	1650
0.01	63
0.5	3
1.2	4
1.5	1.3

From Figure (10) and Table (7), the suggested framework can be used to detect the variations as much as 0.5 in an ARL less than 3. Smaller variations as low as 0.01 can be detected in less than 87 ARL. However, if there was no viscosity variations false alarms will appear once every 1650 points which is a relatively low rate of false alarms.

7. Conclusion

It was found that the new suggested model can be tuned through added parameters such as threshold or a scaling factor. If the new model is tuned such that it gives close to the same amount of false alarms as that of the X-bar or CUSUM control charts. Then this model will be able to detect small shifts (0.25sigma) far earlier than the traditional X-bar or CUSUM control charts. However, it will detect larger shifts slightly later than the traditional control charts.

The ensemble was successfully tuned to detect batch-to-batch variation in sheet extrusion, with a false alarm ARL of 1650 and it can discover a small shift of 0.01 in viscosity with an average ARL = 87.

References

- [1] Western Electric Company. Statistical quality control handbook. Indianapolis: Western Electric; 1958.
- [2] L.S. Nelson, "The Shewart control chart-tests for special causes". Journal of quality technology, Vol. 16, 1984, 237-239.
- [3] S.W. Roberts, "Control chart tests based on geometric moving averages". Technometrics, Vol. 1, 1959, 239-250.
- [4] J.M. Lucas, "Combined Shewhart-CUSUM quality control schemes". J. Qual. Technol., Vol. 14, 1982, 51-59.
- [5] J.M. Lucas, "A modified V mask control scheme". Technometrics, Vol. 15, 1973, 833-847.
- [6] A.E. Smith, "X-bar and R control charts interpretation using neural computing". International Journal of Production Research, Vol. 32, 1994, 309-320.
- [7] G.A. Pugh, "A comparison of neural networks to SPC charts". Computers in industrial engineering, Vol. 21, 1991, 253-255.
- [8] D.T. Pham, E. Oztemel, "Control chart pattern recognition using neural networks". Journal of Systems Engineering, Vol. 2, No. 4, 1992, 256-62.
- [9] C.S. Cheng, "A multi-layer neural network model for detecting changes in the process mean". Computers and Industrial Engineering, Vol. 28, No.1, 1995, 55- 61.
- [10] D.T. Pham, E. Oztemel, "Control chart pattern recognition using learning vector quantization networks". International Journal of Production research, Vol. 32, No. 3, 1994, 721-9.
- [11] J. Plummer, "Tighter process control with neural networks". AI Expert, Vol. 8, 1993, 49-55.
- [12] C S. Cheng, "A neural network approach for the analysis of control charts". International Journal of Production Research, Vol. 35, No. 3, 1997, 667-97.
- [13] HB. Hwring, WC. Chong, "Detecting process non-randomness through a fast and cumulative learning ART-based pattern recognizer". International Journal of Production Research, Vol. 33, 1995,1817-1833.
- [14] Chang, CA. Aw, "A neural fuzzy control chart for detecting and classifying process mean shifts". International Journal of production research, Vol. 34, 1996, 2265- 2278.
- [15] N. Santarisi , M A Barghash, "Neuro-fuzzy based model for pattern recognition in control charts". Proceeding of the 37th international conference on computers and Industrial Engineering 2007, 1881-1897.
- [16] M. Barghash, N. Santarisi, "Analyzing the Effect of Training Parameters on the Performance of the ANN Using Fractional

- Factorial Experimental Design". Journal of intelligent manufacturing Vol. 15, 2004, 635-644.
- [17] DT Pham, MA Wani, "Feature based control charts pattern recognition". International Journal of production research, Vol. 35 No.7, 1997, 1875-1890.
- [18] M. Gülbaya, C. Kahramana, "Development of fuzzy process control charts and fuzzy unnatural pattern analyses". Computational Statistics & Data Analysis, Vol. 51, 2006, 434-451.
- [19] S.K. Guuri, S. Chakraborty, "Recognition of control charts using improved selection of features". Computers in Industrial Engineering, 2008,
- [20] S.K. Guuri, S. Chakraborty, "Feature-based recognition of control chart patterns". Computers in Industrial Engineering, Vol. 51, 2006, 726-742.
- [21] K. Assaleh, Y. Al-Assaf, "Features Extraction and Analysis for Classifying Causable Patterns in Control Charts". Computers and Industrial Engineering, Vol. 49, 2005, 168-181.
- [22] C.S. Cheng, H.P. Cheng, "Identifying the Source of Variance Shifts in the Multivariate Process Using Neural Networks and Support Vector Machines". Expert Systems with Applications, Vol. 35, 2008, 198-206.
- [23] R.S. Guh, Y.R. Shiue, "On-Line Identification of Control Chart Patterns Using Self-Organizing Approaches". International Journal of Production Research, Vol. 43, 2005, 1225-1254.
- [24] C.H. Wang, W. Kuo, H. Qi, "An Integrated Approach for Process Monitoring Using Wavelet Analysis and Competitive Neural Network". International Journal of Production Research, Vol. 45, 2007, 227-244.
- [25] M. Barghash, N. Santarisi, "A study on the response and stability of Artificial Neural Networks for pattern recognition in control charts". 12th International Conference on Machine Design and Production, Kusadasi, Turkey 2006.
- [26] S. Geman, E. Bienenstock, R. Doursat, "Neural networks and the bias/variance dilemma". Neural computations, Vol. 4, 1992, 1-58.
- [27] J. Yu, L. Xi, "A neural network ensemble-based model for on-line monitoring and diagnosis of out of control signal in multivariate manufacturing processes". Expert systems with applications, Vol. 36, 2009, 909-921.
- [28] Z.H. Zhou, J.X. Wu, W. Tang, "Ensembling neural networks: many could be better than all". Artificial intelligence, Vol. 137, 2002, 239-263
- [29] Edmund Burke, Steven Gustafson, Graham Kendall, "A survey and analysis of diversity measures in genetic programming". In GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference, pages 716-723. Morgan Kaufmann Publishers, 2002.
- [30] J. Yi, V R. Prybutok, HR. Clayton, "ARL Comparisons Between Neural Network Models and \bar{x} -Control Charts for Quality Characteristics that are Nonnormally Distributed". Economic Quality Control, Vol. 16 No.1, 2001, 5-15.
- [31] SIMULATOR software, 2009, <http://www.variation.com/sim/>
- [32] S. Haykin. Neural Networks a Comprehensive Foundation. New Jersey: Macmillan; 1994.

Fatigue Properties under Constant Stress/Variable Stress Amplitude and Coaxing Effect of Acicular Ductile Iron and 42 CrMo4 Steel

A.R. I. Kheder^a, N. M. Jubeh^{*b}, E. M. Tahah

^aAl-Balqa' Applied University, Faculty of Engineering, Department of Materials and Metallurgical Engineering, Al-Salt, Jordan

^bAl-Balqa' Applied University, Faculty of Engineering Technology, Department of Mechanical Engineering, Amman, Jordan

Abstract

Fatigue characteristics of as-cast acicular ductile iron (ASGI) produced by alloying and controlled cooling and as-rolled 42CrMo4 steel were studied. Rotating bending fatigue tests at a stress ratio of $R = -1$ was employed. Tests included constant stress amplitude loading, variable stress amplitude high-low, low-high loading and coaxing. $S - N$ curves were established and fatigue strengths were determined, (450 N/mm² for ASGI, 460 N/mm² for steel). The variable stress amplitude loading, high-low, showed high N_T/N_f for the materials, while variable stress amplitude loading, low-high showed a decrease in N_T/N_f for steel and only small improvement for ASGI (N_T = total stress cycles in variable stress amplitude, N_f = stress cycles in constant stress amplitude to failure). Both materials showed coaxing effect, pronounced more in ASGI. The crack propagation rate determined using Paris equation at constant stress amplitude for ASGI was $da/dN = 2.0 * 10^{-5} (\Delta K)^{2.6902}$ and for steel it was $da/dN = 1.0 * 10^{-6} (\Delta K)^{3.5205}$.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: fatigue strength; as-cast ASGI; 42CrMo4 steel; constant stress; amplitude; variable stress amplitude; coaxing effect

1. Introduction

Spheroidal graphite cast irons (SGI) are generally used in as-cast condition or after relatively simple annealing or normalizing heat treatment [1]. Remarkable engineering properties and the full potential of the material can be achieved when the SGI is heat treated and converted to austempered ductile iron [2], but the cost of adding elaborate heat treatment facilities is high. An alternative method is to employ a controlled cooling schedule closely approximating the thermal schedule for austempering with suitable alloying elements [3-4]. The heat treatment conditions can be selected, either to produce a structure of upper bainite which combines high strength, ductility and toughness or lower bainite which is characterized by good wear resistance and high strength.

The austempered ductile iron has excellent toughness and wear-resistance and has been shown to have superior fatigue properties as well. The fatigue strength tests of gears showed that the bainitic ductile iron grades were superior to the pearlitic and quenched and tempered grade [5].

The fatigue strength of the materials is related to their tensile strength. The endurance ratio (ratio of unnotched fatigue strength to tensile strength) is approximately 0.45 for nodular graphite irons and compacted graphite irons [6]. Both nodular and compacted graphite irons do not exhibit the lack of notch sensitivity shown by the flake graphite cast irons because they do not contain the sharp-edge graphite flakes which act as internal notches in flake

graphite irons. Both types of irons are notch sensitive in fatigue. With a relatively sharp notch the fatigue limit is reduced to about a quarter of the tensile strength [7], various kinds of cast irons containing different graphite shapes were investigated and the influence of graphite shape on fatigue strength of cast iron was determined from stand points of the external effect, the notch effect of graphite and the effect of reduction of effective sectional area caused by graphite [8].

The matrix of cast iron is comparable to that of steel and thus can be in ferrite, pearlite and other phases possible in steel. The morphology of graphite in cast iron determines the type of cast iron and it is the controlling factor of the mechanical and physical properties of the cast iron.

This paper describes the fatigue strength and crack growth behavior of as-cast acicular spheroidal graphite cast iron. The results are compared with those of as-rolled steel, and discussed from several viewpoints and compared with the data available in the literature.

2. Experimental Procedure

In this study, specimens of acicular spheroidal graphite cast iron (ASGI) and 42CrMo4 steels were subjected to rotating bending fatigue tests. The ASGI specimens for fatigue tests and other mechanical properties were prepared from ingots produced by controlled cooling and alloying method [4]. They were tested in as-cast

* Corresponding author. e-mail: jubeh@wanadoo.jo

condition. The steel specimens for tests were prepared from steel rods of 40 mm diameter.

The preparation of the fatigue test specimens from ingots is shown in Fig. 1, CNC wire cutting machine was used for this purpose and then the square specimens were further machined to obtain the fatigue test specimens as shown in Fig. 2. The fatigue specimens from rolled steel were machined directly to the required dimensions, Fig. 2. The final machining of the fatigue specimens were carried out on CNC lathe machine. To obtain smooth surface, the fatigue test specimens were ground using emery papers 320, 500, 800, and 1000 grades and then diamond paste of $\frac{2}{3}$ μm particle size was utilized for polishing the surface. The average surface roughness of $R_{\text{max}} = 1.89 \mu\text{m}$ was obtained and taken into consideration. The chemical composition and mechanical properties of the materials employed are given in Tables 1 and 2 respectively.

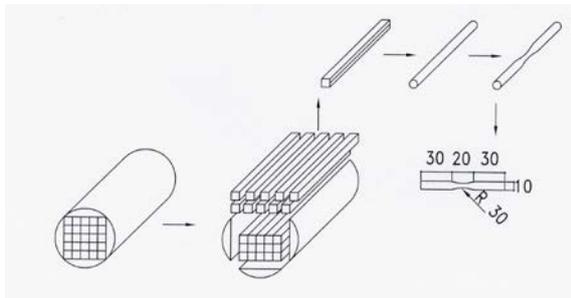


Figure 1: Preparation of the fatigue specimen of the ASGI.

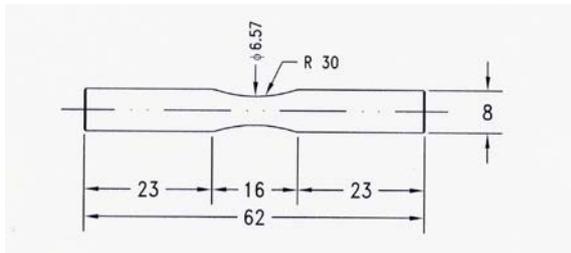


Figure 2: Fatigue specimen.

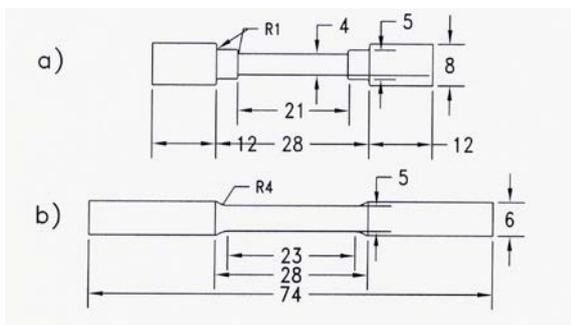


Figure 3: Tensile test specimen:
 a) ASGI according to ISO 1083.
 b) Steel according to DIN 50125.

Table 1: Chemical composition of the materials.

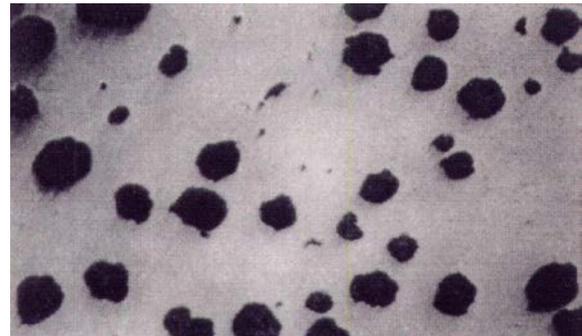
Elements	C	Si	Mn	Cr	Ni	Mo	S	P	Mg
ASGI	3.3	2.8	0.2	0	2.8	0.2	0.01	0	0.033
Steel 42CrMo4	0.4	0.2	0.5	0.9	0.2	0.1	0.03	0	-

Table 2: Mechanical properties of the materials

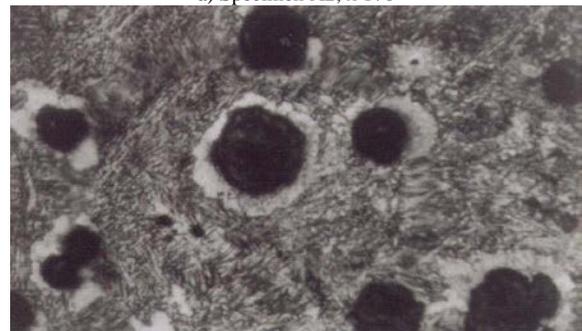
Mechanical properties	Yield strength N/mm ²	Tensile strength N/mm ²	Hardness HBN	Elongation %
ASGI	780	1050	370	5.5
Steel 42CrMo4	885	1096	217	10

To determine the tensile properties of the materials, specimens were machined to the dimensions given in Fig. 3 according to ISO 1083 for ASGI and DIN 50125 for steel.

The ASGI specimens were subjected to microstructural examinations to determine the number of graphite nodules/mm² and the matrix microstructure. The matrix microstructure consisted of about 95% bainite and about 5% retained austenite. These percentages were estimated by applying the direct comparison method [9]. The constituents of the microstructure are shown in Fig. 4 for ASGI in unetched and etched conditions. The details of the graphite nodules across the cross section of the ingot, such as the number of nodules/mm², the nodule size and the area fraction of the nodules, as well as the hardness variation of the structure with distance, are shown in Figs. 5 and 6. The structure of the steel was mainly pearlite.



a) Specimen A2, x 170



b) Specimen A4, x 270

Figure 4: a) Graphite nodules in as-cast ASGI, unetched, x 170.
 b) Graphite nodules and the matrix structure in as-cast ASGI, etched, x 270.

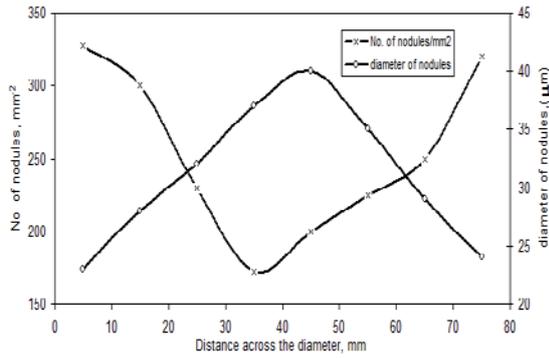


Figure 5: Change of the number of graphite nodules/mm² and their diameter with the distance.

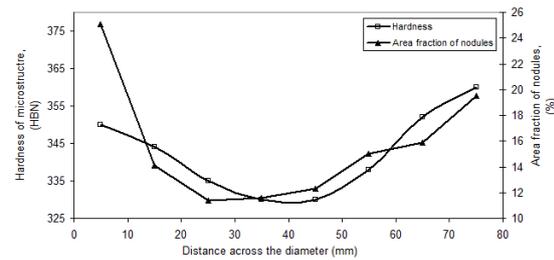


Figure 6: Change on the area fraction of graphite and hardness with the distance.

Fatigue tests were carried out using the rotating bending fatigue machine with a stress ratio of R= -1. The tests were divided into three groups:

- Constant stress amplitude fatigue tests on specimens of as-cast ASGI, as-rolled steel.
- Variable stress amplitude fatigue tests including two-step, low-high and high-low loading tests on specimens of as-cast ASGI and as-rolled steel.
- Coaxing effect on specimens of as-cast ASGI and as-rolled steel.

The fatigue crack propagation was followed up by using plastic replica technique. Cellulose acetate foils of 35µm thickness were employed for this purpose. The foils were cut into strips suitable for sticking on the surface of the specimens after wetting by acetone and left for 3-4 minutes for drying, then the replicas were lifted from the surface carefully and fixed between two glass slides for observations and crack length measurements using optical microscope.

To study the coaxing effect, some of the fatigue specimens were tested at stresses 3 MPa below the fatigue limits of the concerned material for duration of 7x10⁶ cycles, then loading the specimens to stresses more than the fatigue limits and measuring the number of cycles to failure. The results obtained in this way were compared with the number of cycles to failure under constant stress amplitude at the stresses above the fatigue limits as shown in Table 3.

Table 3: Data of the coaxing effect.

	σ_1 (MPa)	σ_2 (MPa)	N_1 under σ_1	N_2 under σ_2	N_f under σ_2 Constant stress	N_2/N_f
ASGI	447	520	7 000 000	1 050 000	93 000	11.29
42 CrMo4	457	520	7 000 000	880 000	200 000	4.4

3. Results and Analysis

3.1. Results of constant amplitude loading:

From the test results under constant stress amplitude loading tests, the stress-stress cycle (S-N) curves were drawn for the specimens of as-cast ASGI and as-rolled 42CrMo4 steel. These results are shown in Fig. 7. These curves can be expressed mathematically as follows:

For as-cast ASGI:

$$\Delta\sigma = 957.56 N_f^{-0.049} \tag{1}$$

For as-rolled 42CrMo4 steel

$$\Delta\sigma = 1387.9 N_f^{-0.073} \tag{2}$$

Where

$$\Delta\sigma = \sigma_{\max} - \sigma_{\min}$$

The fatigue limits (σ_w) for the materials used were obtained from the results given in Table 4 and Fig. 7, and they are: 450 N/mm² for as-cast ASGI and 460N/mm² for as-rolled steel. There are more up-to-date and elaborate models of expressing the fatigue data in the literature [10, 11, &12]. These models can lead to a better fit and provide more appropriate information when compared with the traditional model applied in this study. Increasing the number of specimens will lead to a better fit and more accurate results. The σ_w for as-cast ASGI and for as-rolled steel are quite reasonable when compared with $\sigma_w = 380$ MPa of S 45C quenched and tempered steel used by [13] reflecting the strength of both materials, particularly the ASGI specimens of bainitic structure.

Table 4: Data of stress ($\bar{\sigma}$) – number of stress cycles (N), for ASGI and 42 CrMo4 steel.

Specimen Number	ASGI		42 CrMo4 steel	
	$\bar{\sigma}$ (N/mm ²)	N	$\bar{\sigma}$ (N/mm ²)	N
1	625	44000	700	44200
2	600	46000	650	61000
3	550	73000	625	110000
4	520	93000	600	150000
5	500	160000	550	185000
6	500	220000	520	200000
7	485	2760000	500	400000
8	475	2800000	500	370000
9	460	6800000	470	7200000
10	450	10000000 No failiure	460	10000000 No failiure
11	400	10000000 No failiure	450	10000000 No failiure

The endurance ratios (σ_w/σ_{uts}) of the materials concerned are: 0.428 and 0.419 for as-cast ASGI and as-rolled steel respectively, where σ_{uts} = ultimate tensile stress. The values for endurance ratio of ASGI and rolled steel are in quite reasonable agreement with published data in [6,8, &14]. The ratios for ASGI and rolled steel are very close indicating that the graphite nodules effect as internal notches in ASGI is negligible emphasizing the conclusion drawn by the authors in [8], where they state that the influence of the notch effect of graphite is far smaller than what has been considered before. This influence can be recognized slightly in case of flake graphite cast iron, but hardly in case of spheroidal graphite cast iron. Many short cracks were formed and were stopped from propagation, at graphite nodules in ASGI after duration of 7×10^6 cycles at stress levels just below σ_w .

The data given in Figs. 5 and 6 are of important value in castings of parts with different thicknesses or diameters which graphite nodule number, size, and volume fraction, as well as the hardness of the structure vary, hence other mechanical properties will vary and should be considered. The data available in this area is very rare, and the results obtained in this study were encouraging.

3.2. Results of variable amplitude loading:

Specimens of as-cast ASGI and as-rolled 42CrMo4 steel were subjected to variable stress amplitude fatigue tests, namely two-steps, low-high (L-H) and high-low (H-L) at initial loading of about $N_1/N_f = 0.25$ where N_1 = number of the stress cycles at initial loading, and N_f = number of the stress cycles at failure. The results are included in Table 4. It is clear from the results that the variable amplitude loading mode H-L leads to higher fatigue life in both materials, while L-H amplitude loading mode results in lower fatigue life in case of as-rolled steel, and in as-cast ASGI there was an increase but not as high as in loading mode H-L, the increase was relatively small compared with the results of constant loading mode at the same stress level.

The increase in fatigue life in case of H-L loading can be explained in terms of formation of large plastic zone at the tip of the crack which delays the crack growth rate at the lower stress level and increases the fatigue life. A reason for the increase of the fatigue life of ASGI specimens in loading mode L-H rather than in steel is the high closure rate of the crack under the higher stress of 600 MPa after changing from 500 MPa which is larger than the closure rate of the crack in tests at constant stress amplitude test. This can be also due to the higher values of the stress intensity factor K than the threshold level which leads to a decrease in the force (stress) responsible of the crack growth at 600 MPa and results in a ratio of $NT/N_f > 1$. Therefore the value of the NT/N_f for ASGI is larger than the value for 42CrMo4 steel in variable loading mode L- H, (NT is the sum of stress cycles at both stress levels, $NT = N_1 + N_2$, and N_f is the number of stress cycles to failure at constant stress amplitude).

3.3. Results of coaxing effect:

Coaxing effect was studied on specimens of as-cast ASGI and as-rolled 42CrMo4 steel under the loading mode, two steps, L-H. The results for ASGI are shown in Fig. 8. First the fatigue tests were carried out at stresses 3 MPa below σ_w of both materials for a period enough to

create short cracks to appear which are not able to propagate, and then the specimens were subjected to stresses higher than σ_w of both materials. The coaxing effect means an increase of the fatigue strength or fatigue life after testing the specimens at stresses below σ_w and then above σ_w of the concerned material.

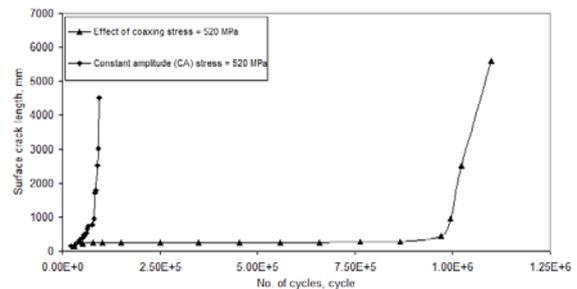


Figure 8: Coaxing effect in as-cast ASGI.

Both materials showed this phenomenon after subjecting the fatigue specimens to initial stress of 447 MPa and 457 MPa for as-cast ASGI and as-rolled 42CrMo4 steel respectively for a period of 7×10^6 cycle. From the results given in Fig. 8 and Table 4, the as-cast ASGI specimens failed after 9.3×10^4 cycles at constant stress amplitude of 520 MPa, while in L-H loading cycle 447-520 MPa the number of cycles to failure was 1.05×10^6 which gives a ratio of $N_2/N_f = 1.05 \times 10^6 / 9.3 \times 10^4 = 11.29$. Similar calculation for as-rolled steel cycled within the stress range of 457-520 MPa and tested at constant stress amplitude of 520 MPa gives a value of $N_2/N_f = 8.8 \times 10^5 / 2 \times 10^5 = 4.4$. In a recent study of the coaxing effect in spheroidal graphite cast iron (SGI) and compacted graphite cast iron (CGI), the following values were obtained: $N_2/N_f = 8.96$ for SGI and 7.34 for CGI, which are of the similar order reported in this work indicating a phenomenon of the same cause [15]. This effect is clearly shown in Fig. 8 for ASGI. The more positive response of the as-cast ASGI to coaxing effect may be related to the higher closure rate of the cracks and the graphite particles acting as obstacles against the crack propagation as shown in Fig. 9, besides the strain aging effect in ferrous materials [13, 16]. The SEM micrographs show clearly the path of the crack propagation which is between the graphite nodules after it has started at the surface of the fatigue specimen. Some branching can be seen, but the main crack is obvious.

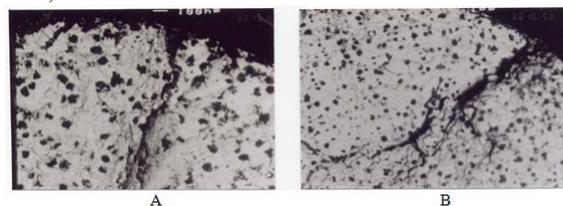


Figure 9: SEM micrographs showing the Crack propagation between the graphite nodules in the acicular ASGI.

A. 600 MPa, $\times 80$

B. 520 MPa, $\times 33$

The crack propagation rate for both materials was calculated according to Paris equation [17], for constant stress amplitude:

$$da/dN = C * (\Delta K)^n \quad (3)$$

Where C and n are constants, ΔK is the average stress intensity factor.

The constants were estimated by applying the least square method and the equations obtained are:

For as-cast ASGI

$$da/dN = 2.0 * 10^{-5} * (\Delta K)^{2.6902} \quad (4)$$

For as-rolled steel

$$da/dN = 1.0 * 10^{-6} * (\Delta K)^{3.5205} \quad (5)$$

The values for constant C for both materials are relatively larger when compared with the C values reported for spheroidal graphite iron, but the values for the constant n are close (n=3.3 to 5) [18, 19].

This study shows the different behavior of the crack propagation in as-cast ASGI and as-rolled steel specimens, because the crack propagation paths are different. In ASGI the crack propagates mainly between the graphite nodules, as shown in Fig. 9. Apparently the graphite nodules provide obstacles against the crack propagation besides other obstacles such as the various phases present in the microstructure, like bainite, martensite, retained austenite and the grain boundaries. In steel the microstructural barriers are the different phases, phase and grain boundaries, carbide particles and dislocations. The presence of graphite nodules increases the threshold stress determined by crack tip blunting during the crack transfer from the matrix to the graphite particles while crack transfer from graphite into matrix occurs if the $\Delta K \geq \Delta K_{th}$ for the matrix [20].

It was noted that there are major cracks and secondary cracks. At initial stages of the fatigue tests it is difficult to distinguish between them but as the test proceeds a chief crack forms and its rate of propagation can be followed and measured while those minor cracks remain inactive until the specimen fails.

4. Conclusion

- Acicular spheroidal graphite cast iron produced by alloying and controlled cooling constitute an important type of ductile iron and has fatigue properties in as-cast condition similar to that of as-rolled 42CrMo4 steel. Their endurance ratios are, 0.428 and 0.419 respectively.
- The number of graphite nodules/mm², the area fraction of the graphite nodules (%) and the hardness of the microstructure decrease with the distance from the outer surface towards the center of the ingot, while the graphite nodule size increases in the same direction.
- The variable stress amplitude high-low mode leads to higher fatigue life for as-cast ASGI and as-rolled steel, while low-high amplitude loading mode results in lower fatigue life in case of steel and an increase in that of ASGI but not as high as in high- low loading mode.
- The as-cast ASGI and as-rolled 42CrMo4 steel showed coaxing effect; it was more pronounced in the former material.

5. Acknowledgment

The second author "N. M. Jubeh" would like to acknowledge the support of Al-Balqa Applied University during his sabbatical leave, which he spends at the Hashemite University; this research was carried out during this period.

References

- [1] I.C. Hugues, A.G. Fuller, R.A. Harding, "Austempered ductile irons – their significance and present applications". Foundry Trade Journal, 10, 1985, 277-286.
- [2] H.L. Morgan, "Introduction to foundry production and control of austempered ductile irons". British Foundryman, 80, part 2, 1987, 98-108.
- [3] J.F. Janowak, R.B. Grunlach, "Approaching austempered ductile iron properties by controlled cooling in the foundry". Int. Conf. on austempered ductile iron, Chicago, Illinois, USA, ASM, 63-69, 2-4 April 1984.
- [4] A.R.I. Kheder, G. Al-Marahleh, "Effect of controlled cooling on the mechanical properties and microstructure of alloyed acicular ductile cast iron". Sci. Bull. Fac. Eng., Ain Shams Univ., Vol. 39, No. 2, 2004, 529-540.
- [5] M. Johansson, "Austenitic-Bainitic Ductile Irons". AFS transaction, 85, 1977, 117-122.
- [6] G.F. Sergeant, E.R. Evans, "The production and properties of compacted graphite irons". British Foundryman, 71, May 1978, 115-124.
- [7] J. Powell, "A review of some recent work on compacted graphite irons". British Foundryman, 77, part 9, Nov 1984, 472-483.
- [8] T. Shiota, S. Komatsu, "Influence of notch and effective sectional area on fatigue strength of cast irons with various graphite shapes". Jap. J. of Materials, 27(294), 1978, 291-297.
- [9] Cullity B.D, Stock S.R. Elements of X-ray diffraction. 3rd Ed. Prentice Hall; 2001.
- [10] E. Castillo, A.S. Hadi, (1995), "Modeling lifetime data with application to fatigue models". Journal of the American Statistical Association, 90, 1995, 1041-1054
- [11] F.G. Pascual, W.Q. Meeker, "Estimating fatigue curves with the random fatigue limit model". Technometrics, 41, 1999, 277-302.
- [12] E. Castillo, A. Fernandez-Canteli, "A general regression model for lifetime evaluation and prediction". International Journal of Fracture, 107, 2001, 117-137.
- [13] S. Ishihara, A.J. McEvily, "A coaxing effect in the small fatigue crack growth regime". Scripta Materialia, Vol. 40, No. 5, 1999, 617-622.
- [14] Haruyoshi Sumimoto, Kokichi Nakamura, Yoshinori Umeda, "The relation between graphite structure and fatigue limit ratio, fatigue limit in CV, spheroidal graphite cast iron". Jap. J. of Foundry Soc. 105 Conf., Kobe City, May, 19, 1984, 279-285.
- [15] A.R.I. Kheder, G. Marahleh, "Effect of coaxing on the fatigue strength of SGI and CGP", to be published in J. Applied Science.
- [16] Silverleaf D.J. Fatigue design of machine components. 6th ed. Pergamon press;1971.
- [17] J.B. Chang, E. Kleven, "A review and assessment of fatigue crack growth rate relationships for metallic air frame

- materials". Fracture and failure analysis mechanics and applications, ASM, 1981, 35-53.
- [18] Shen-Chill Lee and Yin-Bean Chang., "Fracture toughness and crack growth rate of ferritic and pearlitic compacted graphite cast iron at 25°C and 150°C". Metallurgical transaction A, 22A, Nov-1991, 2645-2653.
- [19] A.R.I. Kheder, G. Marahleh, S. Al-Goussous, "Fatigue crack propagation in SGI and CGI". J. Applied Science, Vol. 5, No. 5, 2005, 862-867.
- [20] O.N. Romaniv, A.N. Tkach, "Experimental study and modeling of near threshold crack growth in nodular cast iron". Fatigue, 90, 1990, 2155-2160.

Investigation of the Endurance Characteristics of a Compression Ignition Engine runs on Jatropha Biodiesel

A. Shanono^{*a}, J. Enaburekhan^b

^aCommander Shanono is currently serving with the Nigerian Navy, and undertook the Masters Degree dissertation under the supervision of Dr. Enaburekhan, 13141, Kano, Nigeria

^bDepartment of Mechanical Engineering, Bayero University, Kano, Nigeria

Abstract

The need to ensure diversity in energy supply and its long-term sustainability necessitated the quest for renewable energy sources. Jatropha biodiesel is a renewable energy source that can be used to run a diesel engine without any modification to the engine. This work investigated the endurance characteristics of a compression ignition engine run on Jatropha biodiesel in comparison to fossil diesel. Results of the investigation indicate that an engine run on Jatropha biodiesel blends could endure longer hours of operation due to lower engine operating temperatures, than when it is run on diesel fuel. The maximum engine temperature recorded when the engine was run on diesel fuel was 161 oC. Whereas, the maximum engine temperature recorded when it was run on the Jatropha biodiesel blends was 147 oC. However, an engine run on Jatropha biodiesel blends would result to more frequent choking of oil filters and earlier lubrication oil change.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: Renewable Energy Source; Sustainability; Biodiesel; CI-engine; Endurance; Poverty Alleviation

1. Introduction

The quest for renewable energy sources emanated from the need to diversify the supply of energy and reduce the dependence on fossil fuels as major source of energy. This need was to ensure sustainable and secure energy supply in order to satisfy the growing demand for energy. In addition, there was growing paranoia on the various forecasts made on fossil fuels depletion. Asif and Muneer [1] made noteworthy analysis on fossil fuels depletion. They averred that the world's ultimate conventional oil reserve was estimated at 2000 billion barrels. This was the amount of production (sic) that would have been produced when production eventually ceases. They went further to state that the expert consensus was that the world's midpoint of reserve depletion would be reached when 1000 billion barrels of oil have been produced. That is to say, half the ultimate reserve of 2000 billion barrels. It was estimated that around 1000 billion barrels have already been consumed and 1000 billion barrels of proven reserve are left in the world. This paranoia made biomass a major source of diversifying the supply of energy.

Biomass is material derived from recently living organism. This includes plants, animals and their by products for example, manure, garden waste and crop residues. It is a renewable energy source based on the carbon cycle, unlike other natural resources such as fossil fuels derived from biological material. Humans have used biomass fuels in the form of solid bio fuels for heating and cooking since the discovery of fire, liquid bio fuels have been used since the early days of the automobile industry.

Nikolaus August Otto, the German inventor of the Internal Combustion Engine (ICE) conceived his invention to run on ethanol. Rudolf Diesel, the German inventor of the diesel engine, designed it to run on peanut oil, and Henry Ford originally designed the Ford Model T, a car produced from 1903 to 1926, to run completely on hemp derived bio fuel. [2]

Venkateswara et al [3] noted that Pryde [4] reviewed the reported successes and shortcomings for alternative fuel research. Long-term engine test results showed that durability problems were encountered with vegetable oils because of deposit formation, carbon build up and lubricating oil contamination. It was concluded that vegetable oils must either be chemically altered or blended with diesel fuel to prevent premature engine failure. They further noted that blending, cracking/pyrolysis, emulsification or transesterification of vegetable oils could overcome these problems. Heating of vegetable oils may reduce the viscosity and improve volatility of the oils but its molecular structure remains unchanged. Hence, polyunsaturated character remains. Blending of vegetable oils with diesel, however, reduces the viscosity drastically and the fuel handling system of the engine can handle vegetable oil-diesel blends without any problem. Based on experimental investigations, it was found that converting vegetable oils into simple esters is an effective way to overcome all the problems associated with the vegetable oils. Fangrui and Hanna [5], Pryde [4] and Srivastava and Prasad [6], all noted that most of the conventional production methods for biodiesel use basic or acidic catalyst. A reaction time of 45min to 1hr and reaction temperature of 55–65°C are required for completion of reaction and formation of respective esters.

* Corresponding author. e-mail: aashanono@yahoo.com

Biodiesel consists of methyl esters of fatty acids produced by the transesterification of vegetable oils. The use of biodiesel in diesel engines requires no hardware modification. In addition, biodiesel is a superior fuel than diesel because of lower sulphur content, higher flash point and lower aromatic content. Biodiesel fuelled engine emits fewer pollutants. Biodiesel can be used in its pure form or as a blend of diesel. It can also be used as a diesel fuel additive to improve its properties. [3] Agarwal and Das [7] observed significant improvement in engine performance and emission characteristics for the biodiesel-fuelled engine compared to diesel-fuelled engine. Thermal efficiency of the engine improved, brake specific fuel consumption reduced and a considerable reduction in the exhaust smoke opacity was observed.

Venkateswara et al [3] blended *Jatropha* biodiesel with fossil diesel to several blends, with which they ran diesel engine tests and investigated the performance and emission characteristics of the engine. The blends were B5, B10, B20, B25, B30, B40, B60, B80 and B100. 'B' denotes a blend of biodiesel and fossil diesel, while the figure indicates the percentage of biodiesel in the blend. The results of their study indicated that the kinematic viscosities (at room temperature of 35°C) of the different blends were higher than the viscosity of diesel. However, up to B20 the viscosities of the blends were very close to the viscosity of diesel. Their densities were also close to the density of diesel, so the blends of B5, B10, B15 and B20 can be used without any heating arrangement and are alternative fuels for diesel. While the flash points of different blends increased with increase in biodiesel percentage, the efficiency dropped, smoke density reduced and carbon monoxide reduced compared to diesel. However, for higher blends (B60–B100), smoke density and carbon monoxide increased due to insufficient combustion.

Generally, after design and fabrication are completed, an engine or component is tested for reliability and endurance. Long-term engine or component operations on a continuous basis are conducted to verify their durability and reliability. Specially equipped test cells provide a comprehensive range of test capabilities including engine health monitoring, low/high temperature simulation, diffusive/premixed combustion simulation, ignition delay/advance simulation, and low-load running simulation. In addition, blow-by monitoring, combustion-heat release analysis, continuous and real-time oil consumption measurement, exhaust emission evaluation, and failure analysis can be carried out. Moreover, fuel analysis, in-cylinder combustion pressure measurement, torsional (sic) vibration testing, and used oil analysis are conducted. Furthermore, the study of combustion processes in the cylinder using laser diagnosis can be carried out. [8]

Kaufman and Ziejewski [9] carried out an endurance test on a 4-cylinder Allis Chalmers diesel engine run on sunflower methyl ester. The test results indicated that some difficulties were experienced with engine starting when run on the methyl ester. At a test cell temperature of 170C, one minute of cranking time was necessary to start the engine. For an additional minute, the engine speed remained at 400 to 500rpm irrespective of throttle position. Thus, there could be problems associated with the use of transesterified vegetable oil at low ambient temperature. They reported that McCutchen [10] durability tested a turbocharged Caterpillar 3306, direct injection diesel engine using methyl ester of rapeseed oil for 150hrs and

found no performance deterioration. In addition, Bacon and Co-workers [11] found much less nozzle coking with methyl stearate than with ethyl oleate or ethyl ester of sunflower oil, and that injector coking was less severe for all of these ester fuels than with unmodified vegetable oils. Furthermore, Fort and Co-workers [12] noted in endurance tests with transesterified cottonseed oil a ring-groove deposit effect, but in general, the appearance was about the same as when diesel fuel was used.

Hawkins [13] however, reported no problems in completing 1500hrs on a Massey Ferguson tractor with a Perkins 4.236 direct injection engine running on an ethyl ester of sunflower oil. German et al [14] also carried out field endurance test of diesel engine fuelled with sunflower oil/diesel fuel blends. Engine inspection after the tests showed no difference from what could be expected from such a test carried out with diesel fuel. Quick and Co-workers [15] observed a 10% reduction in power after just 7hrs of operation using filtered raw linseed oil in a Lombardini single cylinder, air-cooled, direct-injected engine. However, with a methyl ester of linseed oil, the engine operated for 1000hrs. All these tests confirm that transesterified vegetable oils are good fuel for direct injected engines. The greatly reduced viscosity, increased volatility, and improved cetane number of the ester compared to the neat vegetable oil leads to more efficient injection and combustion, even though the iodine numbers, as measures of unsaturation, remain essentially the same for the esters as for the neat vegetable oil. These earlier endurance tests utilising the vegetable oil and methyl esters of various seeds served as guides for this study on *Jatropha* biodiesel.

The most important parameters for determining engine endurance are engine operating temperature and lubrication oil. High engine temperature leads to overheating, burning of cylinder head gaskets and cylinder liner ferrules, reduction in lubrication oil viscosity and engine seizure. While low lubrication oil viscosity leads to friction, overheating and engine seizure, high lubrication oil viscosity leads to choking of filters, oil starvation and, engine seizure. On the other hand, low engine temperature leads to high fuel consumption and incomplete combustion. The endurance test for this study involved running a diesel engine with nine test fuels and analysing the engine's endurance characteristics. The test fuels used were; fossil diesel, B20, B30, B40, B50, B50, B60, B70, B80 and B100, and the endurance characteristics analysed were; engine operating temperatures and, lubrication oil and blow-by.

2. Test Fuels and Lubrication Oil Properties

The reference diesel fuel was obtained from MOBIL fuel filling station, which satisfied the ASTM D975 2-D diesel standard. On the other hand, National Research Institute for Chemical Technology (NARICT), Zaria (NIGERIA) extracted the *Jatropha* oil from the *Jatropha* seeds. The fatty acid profile of the oil are; 47.3% linoleic, 12.5% oleic, 11.3% palmitic, 17.0% stearic and 4.7% arachidic. NARICT also transesterified the oil into the methyl ester – *Jatropha* biodiesel – and analysed both the test fuels, using American Society of Testing and Materials standard procedures. Table 1 show several important properties of the fuels that influenced the engine performance and eventually, its endurance. Compared to the reference diesel fuel, the density, flash point, cetane

number, water content and kinematic viscosity of the methyl ester were significantly higher, though it has lower calorific value.

Table 1: Test Fuels Properties.

Serial	Fuel property	Unit	ASTM D975 2-D Diesel	NARICT Produced Jatropa Biodiesel	ASTM Limits for Biodiesel
1.	Density at 15°C	kg/m ³	850	879	860 - 900
2.	Flash Point	°C	80	191	130 min.
3.	Viscosity at 40°C	mm ² /s	2.37	4.84	1.9 – 6.0
4.	Calorific value	MJ/kg	42.7	38	-
5.	Cetane number	-	50.1	51	47 min.
6.	Water Content	Vol.%	0.02	0.16	0.05 max.

Source: National Research Institute for Chemical Technology. [16]

The properties of the biodiesel test fuel satisfied the ASTM limits for biodiesel except the water content, which was significantly high. The biodiesel has a water content of 0.16 as against the ASTM limit of 0.05. SAE 40 OLEUM heavy-duty motor oil was used as the engine lubrication oil throughout the engine tests. It has a viscosity of 137.5 cSt at 40°C and 17.8 cSt at 100°C. It has ash content of 0.85 and flash point of 230°C.

3. Engine Set-up and Procedure

A single cylinder, 4-stroke Petter diesel engine was used for the endurance test. The engine specifications are given in Table 2.

Table 2: Test Engine Specifications.

Serial	Description	Specification
1.	Engine Manufacturer/Model	Petter AA1
2.	Type	Single Cylinder, Air Cooled, Naturally Aspirated 4-Stroke Diesel Engine
3.	Bore/Stroke	70mm/57mm
4.	Compression Ratio	17:1
5.	Maximum Torque	8.2 Nm at 2700 rpm
6.	Maximum Brake Power	2.6 kW at 3600 rpm
7.	Fuel Injection Timing	24 to 33° BTDC
8.	Fuel Injection Pressure	180psi

Source: Cussons Educational Technology. [17]

The engine, which was mounted on a test bench and then connected to a hydraulic dynamometer and control panel, has accessories for monitoring speed, torque, temperature and fuel consumption. The engine temperature, exhaust temperature and fuel consumption were measured at various speeds and loads. Fig. 1 shows the engine set up and control panel.



Figure 1: Engine Set Up With Control Panel.

The test Cycle adopted for this study consists of five test runs with the engine speed and torque manually simulated as required. In Test run 1, the engine was operated as an automobile engine for duration of 60 minutes, with the load determining the speed. The engine speed was initially set to 2800rpm with minimal load; this was followed with a step increase of the load on the engine. Test run 2 was a hybrid engine operation for 3 hours duration carried out to initiate durability problems. The engine was simulated to operate as a ship engine/generator in test runs 3 and 4 (3 hrs each), while test run 5 was low-load running operation to determine the engine's endurance on low-load, with the remedial action taken after the low-load operation (1hr 10mins). The differences in the simulated modes of operation of the engine depend on its governing action. It should be borne in mind that the purpose of operating the engine as an automobile engine, a ship engine or a generator was to fully exploit it and see its response to various load settings (and thus its performance), since that is the essence of a durability/endurance test.

The test engine was first run on the reference diesel fuel and then on the different blends of Jatropa methyl ester. It underwent a Test Cycle consisting of five tests runs on each test fuel. The total duration of each test cycle thus was 11 hrs 10 minutes. After each test cycle, the engine oil was drained and replaced with fresh oil, while the fuel system was flushed with the next test fuel. Fuel consumption in all the test runs was measured on volume basis. The test cycle duration was sufficient to cause a change in the condition of the oil and thus, gave a predictive idea of the condition of the oil after 100 hours of operation. This is especially with the hybrid engine operation and the simulation of low-load running operation condition. By the end of the engine tests runs, 10 lubrication oil samples were obtained (the reference lubrication oil sample, diesel run oil sample, and a sample each from the B20, B30, B40, B50, B60, B70, B80 and B100 engine test cycles). These samples were analysed by AMMASCO International Limited, Club Road, Kano-Nigeria.

4. Result and Discussions

4.1. Performance analysis:

Figs. 2 and 3 shows the performances of a diesel engine run on diesel fuel and Jatropa biodiesel blends when operated as an automobile engine.

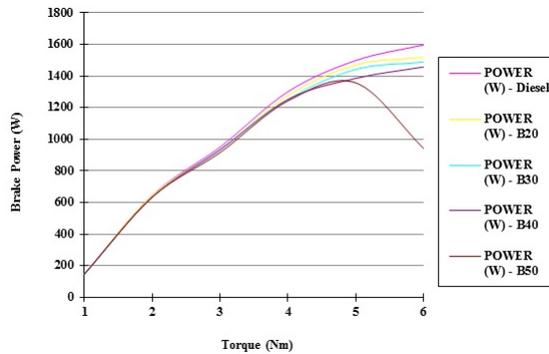


Figure 2: Performances of a Diesel Engine Run on Diesel Fuel, Jatropa Biodiesel Blends B20, B30, B40 and B50, when Operated as an Automobile Engine.

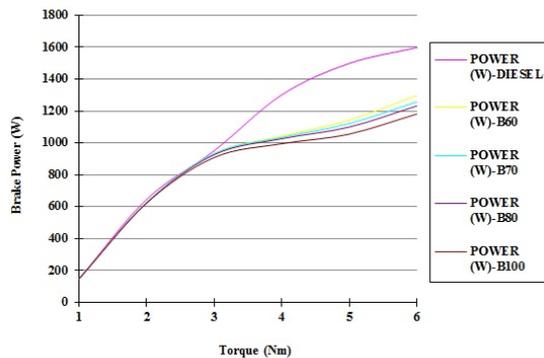


Figure 3: Performances of a Diesel Engine run on Diesel Fuel, Jatropa Biodiesel Blends B60, B70, B80 and B100, when operated as an Automobile Engine.

From the figures, the brake power from the diesel run engine was higher than that recorded from the blends. The power of the engine reduced with increase in the blend percentage. It was observed that the power output from the B20, B30 and B40 runs were closer to that of the diesel run. In addition, B50 exhibited the performance of a real engine at much lower loading than B20, B30 and B40. However, from B60 to B100, there was sharp drop in the power output of the engine at higher engine loading as compared to diesel fuel.

The performances of a diesel engine run on diesel fuel and Jatropa biodiesel blends when operated as ship engine or generator, are shown in Figs. 4 and 5.

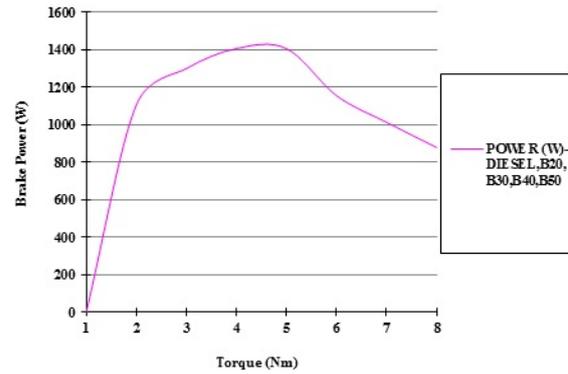


Figure 4: Performances of a Diesel Engine Run with Diesel Fuel, Jatropa Biodiesel Blends B20, B30, B40 and B50, when Operated as Ship Engine or Generator.

As a ship or generator, the engine produced the same power when run on diesel fuel, B20, B30, B40 and B50. However, the power output reduced considerably with increase in the Jatropa biodiesel blend percentage. It should be noted that in Fig. 5, with B100 the engine produced the same power output as when it was run with B80 (the B100 brake power curve is superimposed on the B80 curve).

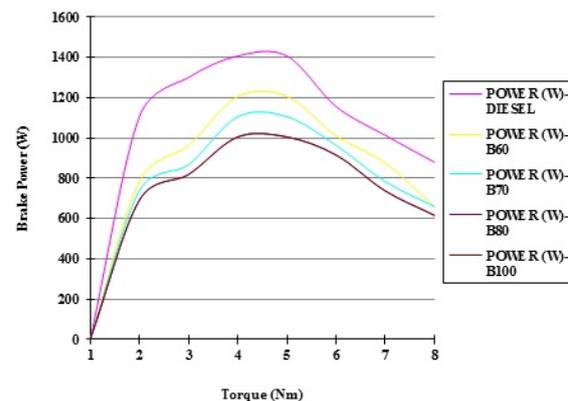


Figure 5: Performances of a Diesel Engine Run with Diesel Fuel, Jatropa Biodiesel Blends B60, B70, B80 and B100, when Operated as Ship Engine or Generator.

Generally, knocking sound of the engine was not noticed throughout the engine test runs, and low-load running of the engine did not result to overheating of the engine. Thus, the possibility of the formation of hot spots in the engine, and crankcase explosion when running an engine with Jatropa biodiesel blends are very unlikely. However, hard starting of the engine was experienced when it was run on B80 and B100. This corresponds to the findings of Kaufman and Ziejewski [9], when they carried out a diesel engine endurance test using sunflower methyl ester.

4.2. Endurance characteristics:

4.2.1. Engine operating temperature:

Figs. 6 and 7 shows comparisons of the variation of engine temperature with load, of an engine run with diesel fuel and that run with Jatropa biodiesel blends, when operated as an automobile engine.

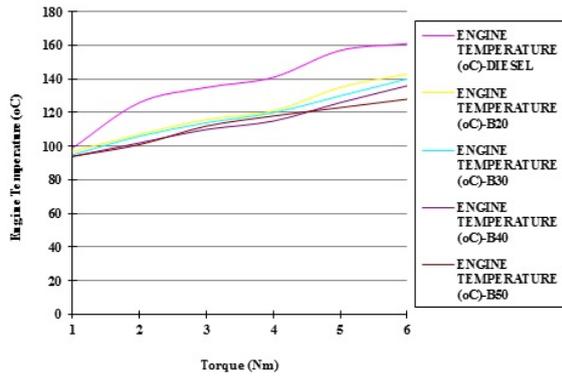


Figure 6: Comparison of the Variation of Engine Temperature with Load, of an Engine Run with Diesel Fuel and that Run with Jatropa Biodiesel Blends B20, B30, B40 and B50, when Operated as an Automobile Engine.

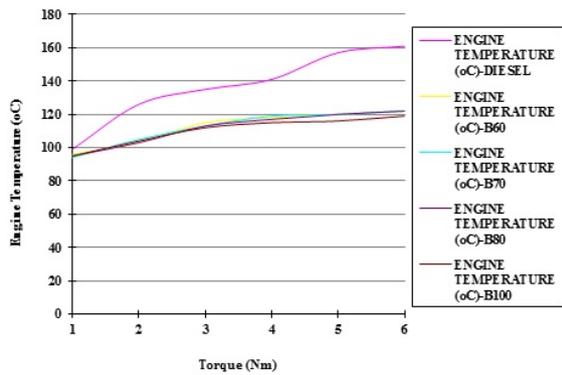


Figure 7: Comparison of the Variation of Engine Temperature with Load, of an Engine Run with Diesel Fuel and that Run with Jatropa Biodiesel Blends B60, B70, B80 and B100, when Operated as an Automobile Engine.

The figures indicate that the maximum engine temperature recorded on test runs with diesel fuel was 161°C. However, the maximum engine temperature recorded with the biodiesel blends was 143°C; this was with the B20 blend. The engine temperature reduced further with increase in the biodiesel blend percentage. The lowest engine temperature recorded at high loading was 119°C, which was with the B100 blend. The high water content in the Jatropa biodiesel could be the factor responsible for the low temperature recorded. Thus, a diesel engine operated as an automobile engine could endure longer hours of operation when run on Jatropa biodiesel blends, than when it is run on diesel fuel. The blend B20 could be an alternative to diesel fuel for use in an automobile engine.

Comparisons of the variation of engine temperature with load, of an engine run with diesel fuel and that run with Jatropa biodiesel blends, when operated as a ship engine or generator are shown in Figs. 8 and 9.

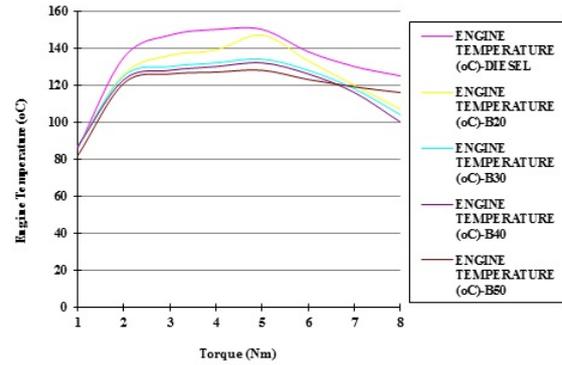


Figure 8: Comparison of the Variation of Engine Temperature with Load, of an Engine Run with Diesel Fuel and that Run with Jatropa Biodiesel Blends B20, B30, B40 and B50, when Operated as Ship Engine or Generator.

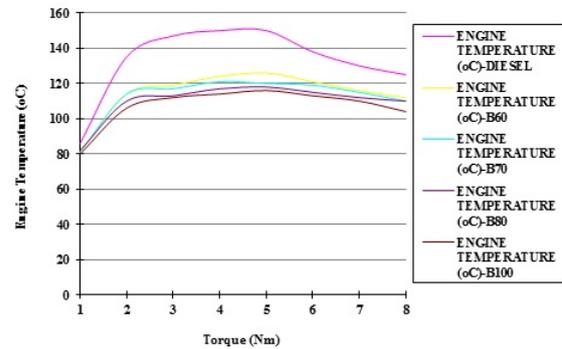


Figure 9: Comparison of the Variation of Engine Temperature with Load, of an Engine Run with Diesel Fuel and that Run with Jatropa Biodiesel Blends B60, B70, B80 and B100, when Operated as Ship Engine or Generator.

From the figures, the maximum temperature recorded with diesel fuel was 152°C, and that recorded with the biodiesel blends was 147°C; this was with the B20 blend. This temperature was too close to that of diesel fuel. The next maximum temperature recorded with the biodiesel blends was 135°C; this was with the B30 blend. It should be noted from Fig. 5 that the power output from an engine run on B50 is the same as that run on diesel fuel, when operated as ship engine or generator. The highest temperature recorded at higher loading from this engine was 128°C. The engine temperature reduced further with increase in the biodiesel blend percentage. The lowest engine temperature recorded at high loading was 114°C, which was with the B100 blend. The high water content in the Jatropa biodiesel could be the factor responsible for the low temperature recorded. Thus, a diesel engine operated as ship engine or a generator could endure longer hours of operation when run on Jatropa biodiesel blends, than when it is run on diesel fuel. The B50 blend could be an alternative to diesel fuel for use in ship engine or generator.

4.2.2. Lubrication oil and blow-by:

The comparison of kinematic viscosities at 40°C of the reference oil, and used oil samples of diesel fuel run and Jatropa biodiesel blends' runs is shown in Fig. 10.

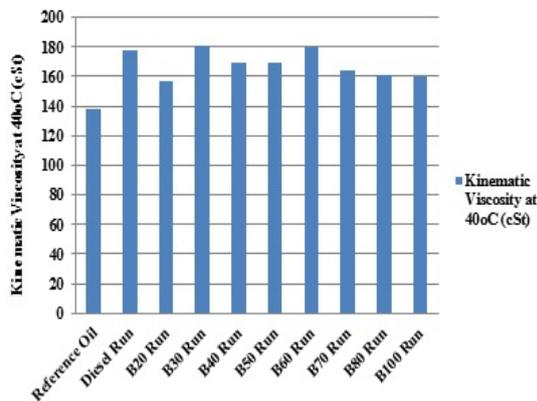


Figure 10: Comparison of the Kinematic Viscosities at 40°C of the Reference Oil and Used Engine Lubrication Oil Samples.

From the figure, the kinematic viscosities at 40°C of the used oils from the engine runs on diesel fuel, B30 and B60 were above the tolerated limits ($\pm 25\%$ of the viscosity of reference/new oil), and are almost of the same value. This implies that the endurance of engines run on B30 and B60 are closer to the endurance of engines run on diesel fuel. The figure also indicates that the kinematic viscosities of the used oil from the engines run on B20, B80 and B100 were closer to that of the reference oil.

Now, consider the kinematic viscosities at 100°C of the used oils for B20, B80 and B100 engine runs in Fig. 11.

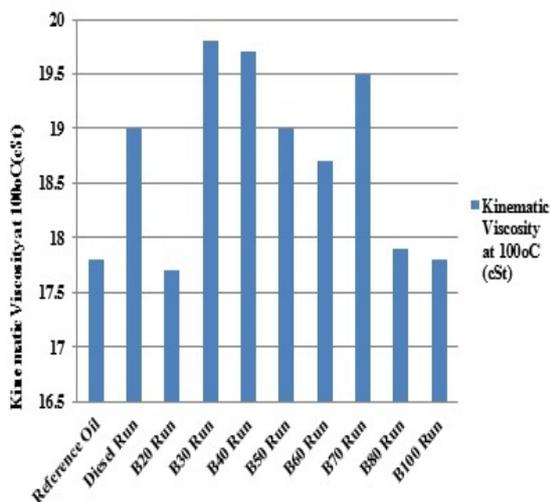


Figure 11: Comparison of the Kinematic Viscosities at 100°C of the Reference Oil and Used Engine Lubrication Oil Samples.

The figure indicates that they were within tolerated limits, and were closer to that of the reference oil in comparison to that of the diesel run. Thus, engines run on Jatropha biodiesel blends B20, B80 and B100 have superior endurance characteristics based on the viscosities of the lubrication oil, than the engines run with diesel fuel.

Fig. 12 shows a comparison of the flash points of the reference oil, and used engine oil samples of diesel fuel run and Jatropha biodiesel blends' runs.

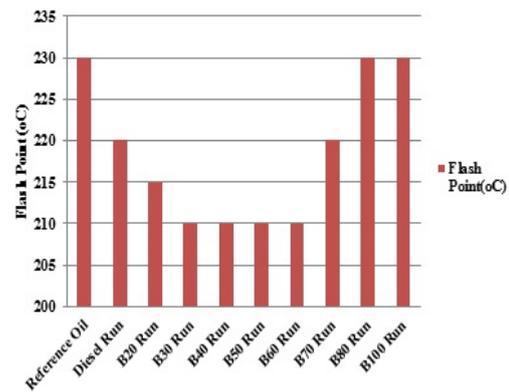


Figure 12: Comparison of the Flash Points of the Reference Oil and Used Engine Lubrication Oil Samples.

The lower flash points of the used oils for B20 to B60 runs as shown in the figure may suggest symptom of fuel dilution during the test runs. This was not the case as the viscosities of the used oils at 40°C indicated there was no fuel dilution during these test runs (the viscosities were not below the tolerated limits). In addition, the flash points of the used oils for all the test cycles were within/closer to the standard limit ($\geq 220^\circ\text{C}$).

Fig. 13 shows the comparison of ash content of the reference oil, and the used engine oil samples.

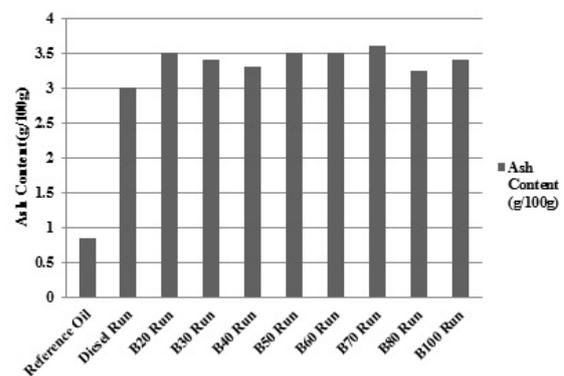


Figure 13: Comparison of the Ash Content of the Reference Oil and Used Engine Lubrication Oil Samples.

From the figure, the ash content in the used oils from test runs with the Jatropha biodiesel blends were higher than that of the diesel fuel test run. Their values were also higher than the standard specification (max. 3%). The lower calorific value of the blends and thus, incomplete combustion could be the reason for this. Therefore, engines run on Jatropha biodiesel blends would result to more frequent choking of filters and earlier lubrication oil changes, than the engines run with diesel fuel.

It should be noted, that the increase in the lubrication oil viscosities at 40°C recorded in test runs with diesel fuel and the Jatropha biodiesel blends, was evidence of the build up of total insoluble/ash content caused by blow-by. However, this did not result to the formation of hot spots in the engine during the test runs and as such, there was no danger of crankcase explosion throughout. In addition, just as with the diesel fuel run, low-load operation of the engine when it was run on the biodiesel blends did not result to dilution of the lubrication oil with fuel. Thus, engines run on Jatropha biodiesel blends would endure low-load running operation.

5. Conclusion

The experimental results were obtained after operating single cylinder 4-stroke cycle diesel engine that was fuelled with *Jatropha* biodiesel blends B20, B30, B40, B50, B60, B70, B80 and B100. The endurance characteristics of the engine were thus analysed and compared to that when the engine was run with diesel fuel. Following are the conclusions based on this investigation:

- A diesel engine operated either as an automobile engine, ship engine or generator could endure longer hours of operation based on the operating temperatures, when it is run on *Jatropha* biodiesel blends (B20, B30, B40, B50, B60, B70, B80 or B100), than when it is run on diesel fuel.
- A diesel engine run on *Jatropha* biodiesel blends B20, B80 and B100 have superior endurance characteristics based on the viscosities of the engine lubrication oil, than the engine run on diesel fuel.
- An engine run on *Jatropha* biodiesel blends (B20, B30, B40, B50, B60, B70, B80 or B100) will result to more frequent choking of oil filters and earlier lubrication oil change, than the engine run on diesel fuel.
- An engine run on *Jatropha* biodiesel blends (B20, B30, B40, B50, B60, B70, B80 or B100) will endure low-load operation in the same way as that run on diesel fuel.
- Hard starting of the test diesel engine was encountered when it was run on *Jatropha* biodiesel blends B80 and B100. This corresponds to the findings of Kaufman and Ziejewski [9] when they carried out a diesel engine endurance test using sunflower methyl ester.
- The *Jatropha* biodiesel blend B20 could be an alternative to diesel fuel for use in an automobile engine.
- The *Jatropha* biodiesel blend B50 could be an alternative to diesel fuel for use in ship engine or generator.

References

- [1] M. Asif, T. Muneer, "Energy Supply, its Demand and Security Issues for Developed and Emerging Economies". *Renewal and Sustainable Energy Reviews*. Elsevier; 19 December 2005, p.3.
- [2] A.T. Marshal, "Bio energy from Waste: A Growing Source of Power". *Waste Management World Magazine*. New York; April 2007, p. 34-37.
- [3] T. Venkateswara Rao, G. Prabhakar Rao, K. Hema Chandra Reddy, "Experimental Investigation of Pongamia, *Jatropha* and Neem Methyl Esters as Biodiesel on Compression Ignition Engine". *Jordan Journal of Mechanical and Industrial Engineering*, Vol. 2, No. 2, 2008, 117-122.
- [4] E. H. Pryde, "Vegetable Oil Fuels Standards". *American Society of Agricultural Engineers' International Conference on Plant and Vegetable oil Fuels*. 1982.
- [5] M.A. Fangrui, M.A. Hanna, "Biodiesel Production: A Review", *Bio Source Technology*, Vol.70, 1999, 1-15.
- [6] A. Srivastava, R. Prasad, "Triglycerides-Based Diesel Fuels". *Renewable Energy Reviews*, Vol. 24, 2004, 111-133.
- [7] A.K. Agarwal, L.M. Das, "Biodiesel Development and Characterization for use as a Fuel in Compression Ignition Engine". *Journal of Engineering, Gas Turbine and Power (ASME)*, Vol. 123, 2000, 440-447.
- [8] M. Megel, "Engine Durability and Reliability Testing". San Antonio. Southwest Research Institute, 2000, p. 3.
- [9] K.R. Kaufman, M. Ziejewski, "Sunflower Methyl Esters for Direct Injected Diesel Engines". *Transactions of the American Society of Agricultural Engineers*, Vol. 27, No. 6, 1984, 1626-1633.
- [10] McCutchen, (1981). In: K.R. Kaufman, M. Ziejewski, "Sunflower Methyl Esters for Direct Injected Diesel Engines". *Transactions of the American Society of Agricultural Engineers*, Vol. 27, No. 6, 1984, 1626-1633.
- [11] Bacon, Co-workers, (1981). In: K.R. Kaufman, M. Ziejewski, "Sunflower Methyl Esters for Direct Injected Diesel Engines". *Transactions of the American Society of Agricultural Engineers*, Vol. 27, No. 6, 1984, 1626-1633.
- [12] Fort, Co-workers, (1982). In: K.R. Kaufman, M. Ziejewski, "Sunflower Methyl Esters for Direct Injected Diesel Engines". *Transactions of the American Society of Agricultural Engineers*, Vol. 27, No. 6, 1984, 1626-1633.
- [13] Hawkins, (1983). In: K.R. Kaufman, M. Ziejewski, "Sunflower Methyl Esters for Direct Injected Diesel Engines". *Transactions of the American Society of Agricultural Engineers*, Vol. 27, No. 6, 1984, 1626-1633.
- [14] T.J. German, K.R. Kaufman, G.L. Pratt, J. Derry, "Field Endurance Test of Diesel Engines Fuelled with Sunflower Oil/Diesel Fuel Blends". *Society of Automotive Engineers*, Paper 850239, 1985.
- [15] Quick, Co-workers, (1983). In: K.R. Kaufman, M. Ziejewski, "Sunflower Methyl Esters for Direct Injected Diesel Engines". *Transactions of the American Society of Agricultural Engineers*, Vol. 27, No. 6, 1984, 1626-1633.
- [16] National Research Institute for Chemical Technology. *Fuels Standards and Specifications*. Zaria. Federal Ministry of Science and Technology. 2009.
- [17] Cussons, G. *Cussons Educational Technology*. Manchester. G Cussons Limited. 1985.
- [18] B. K. Barnwal, M.P. Sharma, "Prospects of Biodiesel Production from Vegetable Oils in India". *Renewable and Sustainable Energy Reviews*, Vol. 9, 2005, 363-378.
- [19] P. De-Filippis, C. Giavarini, M. Scarsella, M. Sorrentino, "Transesterification Processes for Vegetable Oils: A Simple Control Method of Methyl Ester Content". *Journal of the American Oil Chemists' Society*, Vol. 72, No. 11, 1995, 1399-1404.
- [20] M.R. Goyal, "Performance, Emissions, and Durability Testing of Diesel Engine with Bio fuels". *Bio Oils Symposium*. Saskatoon, Saskatchewan, Canada, 1994.
- [21] C. Guangyi, Q. Daren, "A Brief Discussion on the Alternative Fuel for Internal Combustion with Plants Oils". *Trans Chin Soc Agric Eng*, 1987, 90-97.
- [22] Y. Ishii, R. Takeuchi, "Transesterified Curcas Oil Blends for Farm Diesel Engines". *Transactions of the American Society of Agricultural Engineers*, Vol. 30, No. 3, 1987, 605-609.
- [23] F. Karaosmanoplu, "Vegetable Oil Fuels: A Review". *Energy Sources*, Vol. 70, 1999, 221-231.
- [24] M. S. Kumar, A. Ramesh, "An Experimental Comparison of Methods to Use Methanol and Pongamia Oil in a

- Compression Ignition Engine". *Biomass and Bio Energy*, Vol. 25, 2003, 309-318.
- [25] M. S. Kumar, A. Ramesh, B. Nagalingam, "Experimental Investigation of Jatropha Oil Methanol, Dual Fuel Engine". *SAE Journal*, 2001.
- [26] A.S. Ramadhas, S. Jayaraj, C. Muralidheeren, "Use of Vegetable Oils as I.C engine Fuels: A Review". *Renewable Energy*, Vol. 29, 2003, 727-742.
- [27] A.S. Ramadhas, S. Jayaraj, C. Muralidheeren, "Characterization and Effect of using Rubber Seed Oil as Fuel in C.I Engine". *Renewable Energy*, Vol. 30, 2004, 795-803.
- [28] R. Sarin, M. Sharma, "Jatropha Palm Biodiesel Blends: An Optimum Mix for Asia". *FUEL*, Vol. 86, 2007, 1365-1371.

An Assessment of Maintenance Practices and problems in Jordanian Industries

K. K. Tahboub^{*,a}

^aDepartment of Industrial Engineering, Faculty of Engineering and Technology, University of Jordan, Amman, Jordan

Abstract

In the move towards world-class manufacturing, many firms are realizing a need for the use of proper maintenance of production facilities and systems. Industrial plants, machines and equipment are becoming technologically more advanced and at the same time more complex and difficult to control. Therefore, the importance of the maintenance function has been greater than before, due to its role in maintaining and improving availability, performance efficiency, on-time deliveries, safety requirements and overall plant productivity.

In this context, Jordanian industries are more and more realizing the importance of the maintenance function. This paper aims at assessing the maintenance requirements in local industries. To achieve this goal, a questionnaire was designed in order to investigate maintenance needs and problems, namely, maintenance planning, spare parts, equipment calibration and maintenance staff training. The questionnaire was distributed on a selected sample of Jordanian industries and was filled, mostly, through personal interviews.

A hundred industries from all different existing sectors and from all sizes responded to the questionnaire. The collected data were analyzed using SPSS. It was found that most of the industrial firms in Jordan have maintenance departments, but still not all of them allocate budgets to these departments. About half of the firms apply preventive maintenance along side with corrective maintenance. Nearly 15% of the respondent firms were found not well aware of the different types of maintenance systems. Half of the firms calibrate their equipment locally. In this study, the main problems in machines' maintenance were highlighted. It was found that the major problem is the unavailability of spare parts in the local market. About 44% of the respondent firms consult foreign experts. The needs of industries related to specialties in the field of maintenance were also investigated. Finally, conclusions regarding the current situation of maintenance in Jordanian industries were derived.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

1. Introduction

According to [1], maintenance is the total of activities serving the purpose of retaining the production units or retaining them to the state considered necessary for fulfillment of their production function. Traditionally, maintenance was an activity that was put into action to solve production problems. Its objective was to keep the process running. Little time was spent on planning maintenance activities.

Nowadays, many firms are realizing a need for the use of proper maintenance of production facilities and systems. Industrial plants, machines and equipment are becoming technologically more advanced and at the same time more complex and difficult to control [2]. Therefore, the importance of the maintenance function has been greater than before, due to its role in maintaining and improving availability, performance efficiency, on-time deliveries, safety requirements and overall plant productivity [1, 3, 4]. As a matter of fact, the total losses due to maintenance omission or ineffectiveness has been highlighted by practitioners and researchers [5, 6, 7]. In order to assess the maintenance practices in their countries, several researchers have conducted surveys and studies. For example, a study was conducted in Hong Kong to assess

maintenance practices there. In [9, 10], a study on maintenance practice and challenges was conducted for Nigerian industries.

Jordanian industries are no exception. They are more and more realizing the importance of the maintenance function [11, 12]. This paper aims at assessing the maintenance requirements in local industries. To achieve this goal, a questionnaire was designed in order to investigate maintenance needs and problems, namely, maintenance planning, spare parts, equipment calibration and maintenance staff training. The questionnaire was distributed on a selected sample of Jordanian industries and was filled, mostly, through personal interviews.

2. Research Methodology

As mentioned earlier, the objective of this study is to assess the status of maintenance practices in Jordanian industries. In order to achieve this objective, a questionnaire was prepared and was filled during a direct meeting with the related people. The data collected in the questionnaire covered the following aspects: Maintenance department and expenditure, Problems regarding machines and calibration, Expenditure on spare parts, Maintenance staff qualifications, and Maintenance staff training.

* Corresponding author. e-mail: ktahboub@ju.edu.jo

2.1. Questionnaire design

The questionnaire was divided into four main parts as follows:

Part 1: General information

This part includes questions that are related to the classification that is followed in this study:

- **Classification by capital:** Capital is divided into five categories according to the classification followed by [13].
- **Classification by sector:** Companies have been classified into twelve major sectors of industry, following the classification of Amman Chamber of Industry (ACI) as shown in Figure-1.
- **Classification by size:** is achieved by combining the number of employees in an industrial firm and its capital according to the ACI classification (Table 1). The year of foundation of an industrial firm was also included in this part in order to see how it affects maintenance practices in general.

Table 1: Classification by size.

Class	No. employees (N)		Capital Investment (CI) (million JD)
Micro	N ≤ 50	Or	CI ≤ 0.25
Small	50 ≤ N < 100	Or	0.25 ≤ CI < 1.0
Medium	100 ≤ N < 500	Or	1.0 ≤ CI < 5.0
Large	500 ≤ N < 1000	Or	5.0 ≤ CI < 15.0
Giant	N ≥ 1000	Or	CI ≥ 15.0

Part 2: Maintenance department

This part includes 16 different questions that are related to the organization of the maintenance department. The first three questions measure the awareness of local industries of the importance of having a maintenance department, allocating a budget for the maintenance department, and the size of this budget with respect to the total operational cost.

Question 4 is concerned with the type of maintenance system applied. Questions 5 and 6 ask about scheduled maintenance and having records for machines breakdowns. Question 7 gives an idea about the effectiveness of maintenance planning. Questions 8 and 9 ask about machine downtimes (dead-times). Question 10 is concerned with the readiness of the maintenance department. Specifically, it asks about having enough tools to perform maintenance operations. Questions 11 and 12 are mainly to assess firms' ability of repairing their own machines, and to explore the major reasons for not repairing some machines. In question 13, companies are asked to summarize problems of maintenance in Jordanian industries in order to find out the major ones (vital few problems). Questions 14 to 16 aim at calibration problems in Jordan.

Part 3: Spare parts

Part 4: Maintenance staff

Questions 1 to 5 assess the technical qualifications of maintenance staff and their educational background. Questions 6 to 10 explore the shortages of any needed specialty in maintenance. Finally, questions 11 and 12 ask for industrial firms' opinions regarding the performance of both foreign and local experts.

2.2. Sample selection:

According to the ACI index, around 8730 industries are registered in the Amman Chamber of Industry. The ACI index classifies those facilities with four employees or less as workshops. About 80-85% of the 8730 facilities are considered to be small workshops. Hence, the remaining 20% (1746 firms) is the population that the survey is concerned with. This portion is selected because the objective of this survey is to study the needs of those industries that contribute considerably to the economy of the country. Accordingly, a sample of 10% of the population was selected and a total of 100 responses (in the form of a filled questionnaire) were achieved.

The requirements of a random, homogenous, and representative sample were taken into consideration. There are twelve different sectors of industry, as mentioned earlier. The number of firms that is chosen in each sector is proportional to the total number of registered industries in that sector (e.g. food companies represent about 7 % of all industries and the percentage of questionnaires that are filled by food companies is around this portion). Moreover, industries in each sector were chosen randomly. The survey covered different geographical areas, like Wadi elsir, Na'our, Zarqa, Fuhais, and most concentrated in the Amman Industrial City (AIC). As shown in Table 2 and Figure 1, the industries who responded to the questionnaire represent all sectors with, nearly, the desired ratio (the designed ratio is proportional to the actual total number of firms in each sector). The respondent firms covered the five ranges of capital investment as shown in Table 3.

2.3. Responses to the questionnaire:

A total of 150 industries were approached in the survey. Only 100 of these industries responded positively and provided answers to the questionnaire, which represented a 67% response rate.

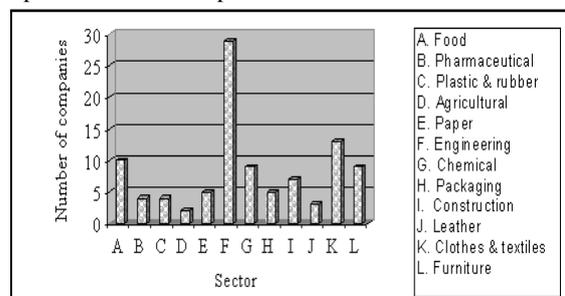


Figure 1: Distribution of the sample by sector.

Table 2: Distribution of respondent industries by sector.

Sector	Number of Firms	Sector %	Respondents %
Leather Products	433	5%	3%
Medical Supplies	81	1%	4%
Plastics and Rubber Products	208	2%	4%
Chemical Products	278	3%	9%
Engineering Industries	2789	32%	29%
Wood and Metal Furniture	1724	20%	9%

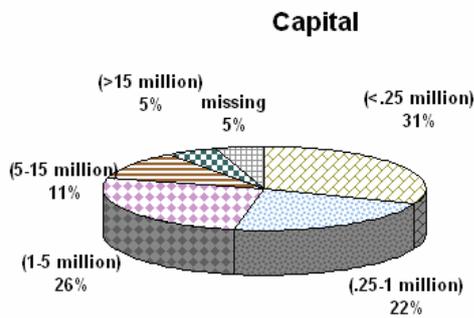


Figure 2: Distribution of the sample by capital.

Table 3: Classification of respondent industries by size.

Class	No. industries
Micro	31
Small	24
Medium	26
Large	12
Giant	5

3. Analysis of the Questionnaire

The questionnaire was analyzed in four different ways to come up with conclusions regarding maintenance practices in Jordanian industries. First, a simple description of the results is presented. Next, the effect of sector, size and year of foundation with respect to some questions in the questionnaire is investigated. Also, confidence intervals were constructed on the results for some questions. Finally, Pareto charts are developed in order to highlight the most important problems (vital few) that industries in Jordan face.

3.1. Description of the results:

3.1.1. Responses related to maintenance department:

Table 4 summarizes the responses of the surveyed industries to questions related to having a maintenance department.

Table 4: Responses related to maintenance department.

Question	%
Having maintenance department	86.0%
Allocating budget for maintenance	53.0%
Maintenance budget relative to total operating cost:	
• < 5%:	34.0%
• 5% - 10%:	27.6%
• > 10%:	24.3%
Adopted maintenance system:	
• Corrective:	26.0%
• Corrective & Preventive:	44.8%
• Corrective, Preventive & Predictive:	13.5%
Having periodic maintenance schedules	79.0%
Keeping records of machine breakdowns	77.0%
Ratio of sudden to planned maintenance:	
• < 20%:	68.8%
• 20% - 30%:	22.6%
• > 30%:	8.6%
Short time between a breakdown and repair	84.0%
Having adequate equipment for maintenance	89.0%
Having machines that failed and could not be repaired	20.0%
Calibration of machines (local or abroad):	
• Local:	52.0%
• Abroad:	3.0%
• Local & abroad:	12.0%
• Do not calibrate:	31.0%
Problems in calibration locally	31.0%

3.1.2. Responses related to spare parts:

Table 5 summarizes the responses of the surveyed industries to questions related to maintaining spare parts inventories.

Table 5: Responses related to spare parts.

Question	%
Keeping inventory of spare parts	83.0%
Annual expenditure on spare parts (JD):	
• < 20000	62.1%
• 20000 – 50000	17.9%
• 50000 – 100000	5.3%
• > 100000	14.7%

3.1.3. Responses related to maintenance staff:

Table 6 summarizes the responses of the surveyed industries to questions related to maintenance staff qualifications.

Table 6: Responses related to maintenance staff.

Question	%
Technicians facing problems understanding instructions in machine catalogues	45.0%
Technicians specializations satisfying maintenance needs	68.0%
Providing maintenance staff training	48.0%
Consulting foreign experts	44.0%
Dependability of local expert in solving technical problems:	
• Dependable to a large degree:	42.0%
• Dependable to a limited degree:	45.0%
• Not dependable	4.0%

3.2. Effect of sector, size and year of foundation:

The effect of industry sector, size and year of foundation of industries was investigated in this study in two ways. First, by constructing cross tables between these variables for some selected questions. Some percentages were calculated to derive conclusions. Secondly, analysis of variance (ANOVA) tables were constructed.

3.2.1. Effect of sector, size and year of foundation:

Tables 7(a), 7(b), 8(a), 8(b), 9(a), and 9(b) show how the answers to the question of having a maintenance department are distributed with respect to sector, size, and year of foundation.

With respect to sector, Table 7(a) shows that the sectors having least percentages of maintenance departments are plastics (50%), chemical (66.6%) and construction industries (71.4%). It is also noted that plastics and agricultural industries do not allocate budgets to their maintenance departments. Also, almost all pharmaceutical, food, leather and agricultural industries maintain periodic maintenance schedules. On the other hand, in the furniture sector, about 44% of firms do not have any periodic maintenance schedules. It is clear that food industries pay great attention to maintenance whether by specifying a budget for the department of maintenance or planning for its works. The same thing is true for the pharmaceutical industries. It is noticed that, in general, all sectors are interested in having machine breakdown tables, except for the plastic and furniture industries.

Table 7(a): Responses by sector related to maintenance department.

Sector	Having maintenance department	Budget for maintenance	Periodic maintenance schedules	Machine Breakdown records	Adequate maintenance equipment
Food	100.0%	90.0%	100.0%	100.0%	100.0%
Pharmac.	100.0%	75.0%	100.0%	100.0%	100.0%
Plastics	50.0%	0.0%	75.0%	50.0%	100.0%
Agricultural	100.0%	0.0%	100.0%	100.0%	100.0%
Paper	100.0%	20.0%	80.0%	100.0%	80.0%
Engineering	86.2%	60.0%	72.0%	79.0%	79.0%
Chemical	66.6%	83.3%	78.0%	67.0%	78.0%
Packaging	100.0%	60.0%	80.0%	100.0%	100.0%
Construction	71.4%	100.0%	86.0%	71.0%	86.0%
Leather	100.0%	100.0%	100.0%	100.0%	100.0%
Textiles	84.6%	81.8%	77.0%	69.0%	92.0%
Furniture	88.8%	14.3%	56.0%	33.0%	100.0%

Table 7(b): Responses by sector related to maintenance department.

Sector	Maintenance budget to total operating cost			Maintenance system		
	< 20 %	20% -30%	> 30%	corrective	Corrective & preventive	Corrective, preventive & predictive
Food	50.0%	50.0%	0.0%	10.0%	50.0%	30.0%
Pharmac.	75.0%	0.0%	25.0%	25.0%	75.0%	0.0%
Plastics	25.0%	75.0%	0.0%	25.0%	25.0%	0.0%
Agricultural	50.0%	50.0%	0.0%	50.0%	0.0%	0.0%
Paper	75.0%	0.0%	25.0%	0.0%	20.0%	20.0%
Engineering	72.0%	24.0%	4.0%	7.7%	57.7%	7.7%
Chemical	100.0%	0.0%	0.0%	25.0%	12.5%	25.0%
Packaging	80.0%	20.0%	0.0%	0.0%	80.0%	0.0%
Construction	57.1%	14.3%	28.6%	0.0%	42.9%	28.6%
Leather	100.0%	0.0%	0.0%	0.0%	33.3%	33.3%
Textiles	63.6%	18.2%	18.2%	15.4%	23.1%	15.4%
Furniture	75.0%	25.0%	0.0%	22.2%	66.7%	0.0%

From Table 8(a), it is noted that large and giant industries are interested in allocating a budget for maintenance department more than micro and small industries. 72% of medium sized industries having maintenance departments do allocate budget for maintenance; while 80.0% of the giant industries and 90.9% of the large industries allocate budgets for

maintenance. In general, companies (from medium to giant) are greatly interested in having machine-breakdown tables. Again, it is a result of their attention in following up machines' estate in order to avoid repeated breakdown. It is well known that for large-size companies, stopping of machines for several minutes costs them a lot.

Table 8(a): Responses by size related to maintenance department.

Size	Having maintenance department	Budget for maintenance	Periodic maintenance schedules	Machine Breakdown records	Adequate maintenance equipment
Micro	70.9%	40.9%	55.0%	45.0%	77.0%
Small	79.2%	57.8%	88.0%	83.0%	88.0%
Medium	96.2%	72.0%	92.0%	92.0%	88.0%
Large	91.7%	90.9%	83.0%	91.0%	92.0%
Giant	100.0%	80.0%	80.0%	100.0%	100.0%

Table 8(b): Responses by size related to maintenance department.

Size	Maintenance budget to total operating cost			Maintenance system		
	< 20 %	20% -30%	> 30%	corrective	Corrective & preventive	Corrective, preventive & predictive
Micro	60.0%	28.0%	12.0%	17.9%	32.1%	3.6%
Small	90.9%	9.1%	0.0%	14.3%	61.9%	4.8%
Medium	53.8%	34.6%	11.6%	7.7%	42.3%	23.1%
Large	81.8%	0.0%	18.2%	9.1%	63.6%	9.1%
Giant	80.0%	20.0%	0.0%	20.0%	20.0%	60.0%

Table 7(b) and Table 8(b) investigate the ratio of maintenance budget to the total operational cost versus sector and size. Table 7(b) shows how maintenance budget percentage is affected by sector. It is noticed that the highest percentages are in the food industries (55.6%), construction (80%) and textile industries (88%). Table 8(b) shows how the ratio of maintenance budget to the total operating cost is affected by size. Giant industries (as should be expected) have the largest percentage (75%).

It is noticed from Table 9(a) that industries that have been established recently are not so interested in having maintenance departments compared to older ones. The reason for this may be that they are new in the field and they have no genuine problems related to machines' maintenance. Industries which have been established in the 1950's to 1970's pay more attention to planning for maintenance works than recent industries.

Table 9(a): Responses by year of foundation related to maintenance department.

Years old	Having maintenance department	Periodic maintenance schedules	Machine Breakdown records	Adequate maintenance equipment
> 40	100.0%	83.0%	100.0%	83.3%
30 - 40	100.0%	88.0%	88.0%	100.0%
20 - 30	100.0%	100.0%	100.0%	100.0%
10 - 20	83.7%	78.0%	73.0%	83.8%
< 10	80.4%	73.0%	71.0%	90.2%

Table 9(b): Responses by year of foundation related to maintenance department.

Years old	Maintenance budget to total operating cost		
	< 20 %	20% -30%	> 30%
> 40	83.3%	16.7%	0.0%
30 - 40	87.5%	12.5%	0.0%
20 - 30	57.1%	28.6%	14.3%
10 - 20	68.8%	18.7%	12.5%
< 10	66.7%	25.6%	7.7%

According to tables 7(a) and 8(a), most of the responded companies from all sectors and sizes have enough tools and equipment for regular maintenance works.

Analysis of variance has been conducted to see whether the year of foundation has an effect on having a maintenance department in a company. That is, the year of foundation has no effect on having a maintenance department. Further, analysis of variance has been conducted to investigate the effect of sector. It was concluded that sector has no effect on the type of the applied maintenance system. Also, ANOVA table was constructed to study the effect of the size of an industry to having periodic maintenance schedules. The test showed that size has an effect on having periodic maintenance schedules.

Table 9(a) shows that older companies are more interested in breakdowns tables than recent ones. As it is said before, recent companies do not suffer from genuine problems in maintenance. The effect of the size of the company is studied regarding this question; the ANOVA test confirmed that the size of the company has an effect on having machine-breakdown tables.

3.2.2. Spare parts:

Table 10 presents the responses of industries by sector to keeping spare parts inventories. Most interested companies in keeping enough inventory of spare parts are from the food, pharmaceutical, packaging and leather industries. Table 11 presents the responses to keeping spare parts by industry size. It also presents industry responses to annual expenditure on spare parts. The table shows that most of the responded companies (especially for micro, small and medium sizes) spend yearly less than JD 50,000 on spare parts.

Table 10: Responses by sector related to spare parts.

Sector	keeping inventory of spare parts
Food	100.0%
Pharmaceutical	100.0%
Plastics	75.0%
Agricultural	50.0%
Paper	80.0%
Engineering	72.0%
Chemical	66.0%
Packaging	100.0%
Construction	85.7%
Leather	100.0%
Textiles	92.3%
Furniture	88.9%

Table 11: Responses by size related to spare parts.

Size	keeping inventory of spare parts	Annual expenditure on spare parts			
		< 20000 JDs	20000 - 50000 JDs	50000 - 100000 JDs	> 100000 JDs
Micro	74.0%	90.0%	10.0%	0.0%	0.0%
Small	67.0%	81.8%	13.6%	4.6%	0.0%
Medium	88.0%	44.0%	28.0%	8.0%	20.0%
Large	92.0%	10.0%	40.0%	0.0%	50.0%
Giant	100.0%	0.0%	0.0%	20.0%	80.0%

3.2.3. Maintenance staff:

Table 12(a) and Table 12(b) show the qualifications of maintenance staff according to the sector of industry.

Table 12(a): Qualification of maintenance staff percentage according to sector.

Sector	mechanical engineer	Electrical engineer	Chemical engineer	Industrial engineer
Food	11.0%	10.0%	1.0%	8.0%
Pharmaceutical	10.5%	10.5%	1.8%	31.6%
Plastics	25.0%	8.3%	0.0%	8.3%
Agricultural	5.6%	5.6%	16.7%	0.0%
Paper	30.0%	17.5%	0.0%	17.5%
Engineering	18.1%	10.6%	0.5%	13.9%
Chemical	4.8%	19.0%	4.8%	0.0%
Packaging	25.0%	8.3%	0.0%	8.3%
Construction	18.2%	9.0%	13.6%	11.4%
Leather	25.0%	25.0%	0.0%	0.0%
Textiles	26.1%	8.7%	0.0%	8.7%
Furniture	13.3%	13.3%	0.0%	20.0%

Table 12(b): Qualification of maintenance staff percentage according to sector.

Sector	Polytechnic	vocational	Community college
Food	12.0%	19.3%	9.2%
Pharmaceutical	35.1%	0.0%	12.3%
Plastics	0.0%	0.0%	0.0%
Agricultural	16.7%	0.0%	0.0%
Paper	17.5%	10.0%	7.5%
Engineering	23.6%	15.3%	7.4%
Chemical	47.6%	0.0%	4.8%
Packaging	0.0%	0.0%	0.0%
Construction	15.9%	2.3%	11.4%
Leather	0.0%	0.0%	0.0%
Textiles	15.9%	5.8%	5.8%
Furniture	13.3%	20.0%	6.7%

Table 13 shows the answers to the rest of the questions related to maintenance staff by sector of industry. It is obvious from Most industries, except leather industries, do not face real problems in dealing with machines catalogues. Leather industries have the highest percentage of companies that are not satisfied with their technicians' specialties. Plastic and agricultural industries are mostly not interested in providing their employees with training in maintenance field.

Table 14: Confidence intervals for some industry responses.

Question	Answer: Yes	Sample Size	P (%)	Confidence Interval
<i>Maintenance Department:</i>				
Have maintenance department	86	100	86.0%	80.3 – 91.7
Have maintenance budget	53	87	60.9%	52.3 – 69.5
Budget percent (<5 %)	21	52	40.4%	29.2 – 51.6
Have periodic maintenance schedule	79	100	79.0%	72.3 – 85.7
Have machine breakdown records	77	100	77.0%	70.1 – 83.9
Sudden to planned maintenance (< 20%)	64	92	69.6%	61.7 – 77.5
Long breakdown duration	16	96	16.7%	0.0 – 23.0
Have adequate maintenance equipment	88	99	88.9%	83.7 – 94.0
<i>Spare Parts:</i>				
Keeping enough spare parts inventory	81	98	82.7%	76.4 – 89.0
Annual spending on spare parts (< 20000 JD)	59	94	69.8%	62.0 – 77.6
<i>Maintenance staff:</i>				
Have problems with maintenance catalogues	44	94	46.8%	38.3 – 55.3
Satisfactory maintenance staff specializations	68	91	74.7%	67.2 – 82.2
Providing maintenance staff with training	46	91	50.5%	41.9 – 59.1

Table 13: Responses by sector related to maintenance staff.

Sector	Problems following catalogues	Satisfactory staff specializations	Providing staff training
Food	60.0%	80.0%	60.0%
Pharmaceutical	75.0%	100.0%	75.0%
Plastics	75.0%	50.0%	0.0%
Agricultural	50.0%	100.0%	0.0%
Paper	60.0%	100.0%	40.0%
Engineering	55.2%	58.6%	45.0%
Chemical	66.7%	55.6%	56.0%
Packaging	60.0%	80.0%	60.0%
Construction	47.1%	85.7%	43.0%
Leather	0.0%	0.0%	67.0%
Textiles	30.8%	69.2%	54.0%
Furniture	66.7%	66.7%	44.0%

3.3. Confidence intervals:

Confidence intervals have been computed for some selected questions. Point estimates are first calculated for these questions (proportion estimates), then α interval estimates are found. For example, companies who said that they have maintenance department are 86 companies from 100 companies, then

$$C.I. = p \pm Z_{\alpha/2} (p(1-p)/n)^{1/2}$$

Where;

$$p \text{ (the proportion estimator)} = 86/100 = 0.86$$

$Z_{\alpha/2}$: upper $\alpha/2$ percentage point of the standard normal distribution [13].

Taking a significance level of 90 %, then:

$$C.I. = 0.86 \pm 1.645 (0.86(1-0.86)/100)^{1/2}$$

$$C.I. = [80.3 - 91.7]$$

This means that with 90 % confidence the proportion of companies who have maintenance department in Jordan lies within a [80.3% - 91.7%] interval. Table 14 shows all of the computed intervals.

3.4. Maintenance problems in Jordan:

In the questionnaire, question 13 asked companies to summarize maintenance problems in Jordan based on their own experiences. Figure 3 shows the frequency of each problem. Shortage of spare parts and their raw materials in the local market occupies the highest frequency (58 answers). The next main problem is the unavailability of local agent for machines and parts. The third problem is the lack of some specialties in the field of machines' maintenance in the local market. Some sectors of industry face other main problems as a consequence of their type of activity or production they are engaged in.

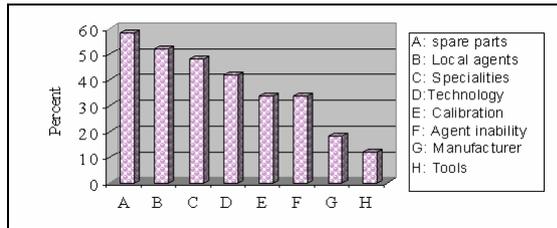


Figure 3: Problems of maintenance in Jordan.

For the food industries, in addition to the main three problems, about 40% of companies face problems in calibration, as shown in Figure 4.

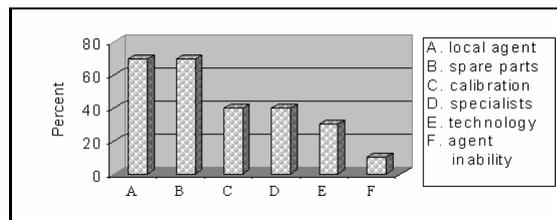


Figure 4: Maintenance problems in food industries.

For the paper industries, about 60% of companies have main problems regarding the lack of knowledge about the basics of technologies used in machines. About 40% of companies face problems in calibration. Figure 5 shows the problems faced in this sector.

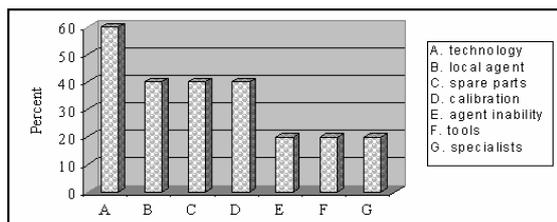


Figure 5: Maintenance problems in paper industries.

For the furniture industries, about 30% of the companies in this sector face problems related to calibration (see Figure 6).

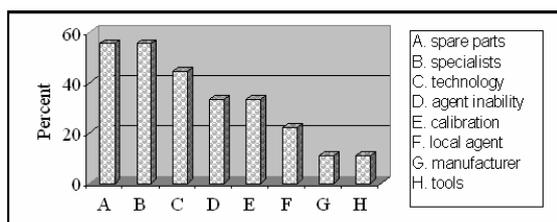


Figure 6: Maintenance problems in furniture industries.

For clothes and textiles industries, nearly 58% of companies suffer from the lack of knowledge about basic technology used in machines, as shown in Figure 7.

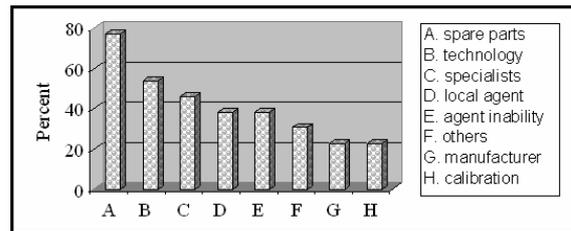


Figure 7: Maintenance problems in clothes and textiles industries.

For the engineering industries, there are 10 companies and so some percentages can be generalized over this sector. About 31% of companies have problems related to calibration. 66% of the companies have problems regarding the unavailability of local agent for machines and tools. The lack of spare parts in the local market is a main problem for 62% of companies in this sector of industry. Nearly 41% of companies say that the local agent, in case of availability, is not capable to solve technical problems. The companies that face problems resulted from the lack of some specialties in the field of machines' maintenance are 41% of the 29 companies. Figure 8 shows problems mentioned above and other problems.

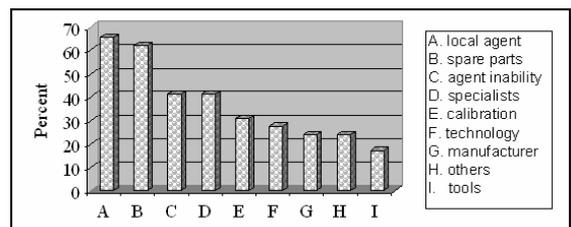


Figure 8: Maintenance problems in engineering industries.

For the packaging industries, about 60% of the companies in this sector face problems in calibration. This may be due to their type of work that needs high accuracy in the fabrication of packages. Figure 9 shows problems faced in this sector.

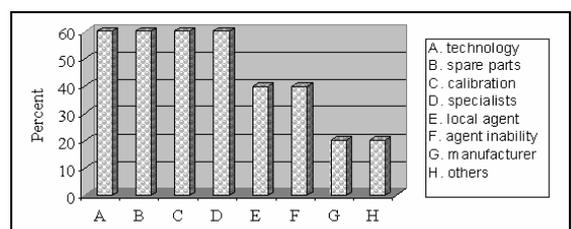


Figure 9: Maintenance problems in packaging industries.

For the plastic industries, it can be said that all companies in this sector of industry suffer from the lack of some technical specialties in the field of maintenance. About 75% of companies have problems related to the knowledge of the basics of technologies used in machines. Figure 10 displays the main problems of this sector.

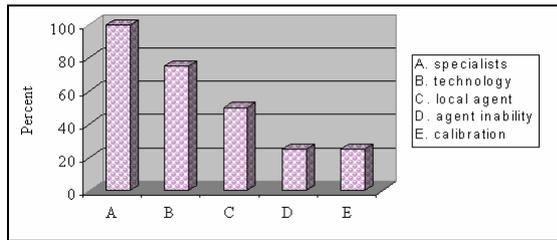


Figure 10: Maintenance problems in plastic industries.

3.5. Companies' ability of repairing their own machines:

The answers to question 11 (in maintenance department part) are almost satisfactory. Only 20 companies have machines that are out of order and have not been repaired yet. 19 companies of those twenty gave reasons for not repairing machines in question 12. Figure 11 shows the frequencies for each reason. About 47% of companies that have a problem say that this is due to the unavailability of spare parts for old machines. 42% of them find it more feasible to buy a new machine comparing with the cost of repairing the existing one. About 37% of companies refer the reason of not repairing a machine to the frequent breakdowns and the useless repair of it.

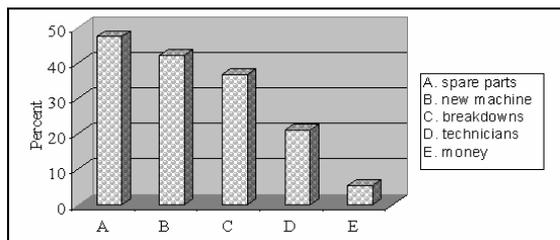


Figure 11: Reasons of not repairing a machine.

3.6. Calibration problems in Jordan:

Twenty three companies answered question 15 that they face problems in calibrating some sets of measurement and control. 22 companies answered question 16 regarding the type of problems they face. Figure 12 displays those problems and their frequencies. The main problems upon which 86 % of responded companies agree is the lack of local capabilities of calibration and the lack of experts in this field.

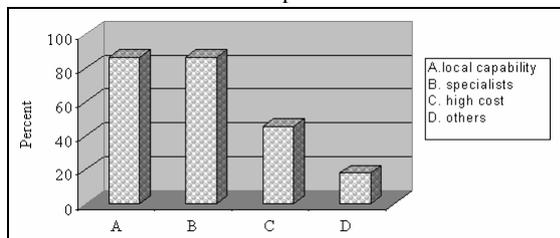


Figure 12: Calibration problems in Jordan.

3.7. Foreign experts specialties needed in maintenance works:

In question 10 (in maintenance crew part) companies that need experts from abroad are asked to specify the specialties they need. Figure 13 shows the frequencies of specialties needed. 44 companies answered this question. About 59% of the companies need experts for control

systems. 48 % of companies require experts specialized in printed circuits. The third required specialty, which is needed by 43% of the companies, is the calibration works. In general, companies facing problems locally in calibration are those who request foreign experts for it.

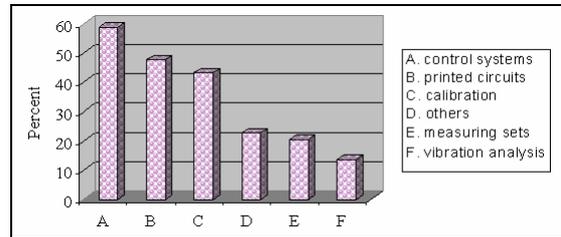


Figure 13: Foreign experts specialties needed.

4. Summary of Major Findings

Through the previous discussion of the questionnaire, it has been found that 86 % of the respondent companies have maintenance departments. Meanwhile, plastic, chemical and construction industries are the sectors that have the lowest percentages of companies that have maintenance departments. Companies that activate the maintenance department role by allocating a budget for it constituted 61.6% of those who said they have departments. The most interested sectors in allocating a budget are food, leather and construction industries. Large companies showed more concern than small ones regarding the size of maintenance budget relative to the total operating cost.

It seems that local industries are in, a rather, good situation since 44.8 % of the respondent companies apply both corrective and preventive maintenance systems. 26 % of the companies apply only corrective maintenance. These companies should start thinking of applying preventive maintenance in order to improve their machines' efficiency. 12.5 % of the companies apply only preventive maintenance, which is not logical, because it is necessary to do corrective maintenance for out-of-order machines. 1 % of the companies apply predictive maintenance only; this is again unreasonable since predictive maintenance requires applying preventive maintenance. 1 % of the respondent companies apply preventive and predictive maintenance and 1 % of them apply corrective and predictive maintenance systems. It is clear from companies answers to this question (the applied maintenance system) that some companies are not well aware of the types of maintenance systems (nearly 15.5%).

Seventy seven percent of the respondent companies have machine-breakdown tables. Companies, except micro ones, are interested in keeping both periodic maintenance tables and machine-breakdown records (above 80 % of the companies in each size).

Eighty four percent of the companies do not have relatively long dead-times. This may be referred to the fact that 89% of the respondent companies have the required tools for regular maintenance works. In addition, an out-of-order machine may lead to stopping of a complete production line, which in return affects the company's estate negatively.

More than half of the respondent companies do their machines' calibration locally (nearly 52% of companies). The main problems that companies face in local calibration are related to the inability of performing calibration works

and the lack of specialties in this field. Thirty one percent of the companies do not calibrate their machines. This may lead to change in the specifications of their products or affect their quality.

Regarding maintenance problems in Jordan, the main problems on which most of the companies agreed are:

- The unavailability of spare parts or their raw materials in the local market (58 %).
- The unavailability of local agents for parts and machines (52 %).
- The lack of some specialties in maintenance field (48 %).
- The lack of knowledge about basic technology used in machines (42 %).
- Problems related to calibration (34 %).

Related to spare parts, 83% of the companies said that they have enough spare parts. Companies (medium, large and giant) are more concerned than small ones with having sufficient inventory of spare parts. The annual expenditure volume on spare parts shows that 62% of the companies spend less than JD 20,000 annually. As can be concluded from the ANOVA test, the expenditure is affected by the size of the company.

About 70% of the respondent companies are satisfied with their technicians specialties. Some sectors such as the plastic and leather industries have large percentages of companies who do not have satisfactory specialties, which cover their needs. 44 % of the respondent companies compensate for this lack of specialties by consulting foreign experts. Foreign experts are mainly needed in the field of control systems, printed circuits and calibration.

Related to the training sessions, it was found that 48 % of the respondent companies said that they do provide their employees with sessions related to maintenance. Companies with micro (32%), small (54%) and medium (46%) sizes were found less concerned with technicians' training.

The plastic industry is one of the sectors that face real problems related to maintenance. Its main problems are technical ones; all of the respondent companies in this sector agreed that there is a lack of knowledge regarding the basics of technology used in machines. In spite of their need to specialists, plastic industries have the least percentage of companies that provide their technicians with training sessions concerning maintenance works and machines technology. Only 50% of plastic companies have maintenance departments, and non of them specifies a budget for its department.

To the contrary of plastic industries, the responses of food industries revealed that these companies are in good situation. All of the respondent food companies have maintenance departments and 90% of them specify separate budgets for their departments. All of these companies are interested in keeping periodic maintenance tables and machine-breakdown records, which reveals their efficiency in planning. They pay attention for training since 60 % of the companies provide their technicians with sessions.

Related to experts, 52% of the respondent companies see that foreign experts have an ordinary performance. 42% of companies stated that they can depend on the local experts to a great degree.

References

- [1] Ahlmann, H., *The Economic Significance of Maintenance in Industrial Enterprises*, Lund Institute of Technology, Lund University, Sweden, 1998
- [2] Stephen, T. D., Positioning Maintenance as a Competitive Advantage, *Plant Engineering*, 54(5), pp66-69, 2000.
- [3] Al-Najjar, B. (1996), Total Quality Maintenance: An Approach for Continuous Reduction in Cost of Quality Products, *Journal of Quality in Maintenance Engineering*, 2(3), pp2-20, 1996.
- [4] Cooke, F. L., Implementing TPM in Plant Maintenance: Some Organizational Barriers, *International Journal of Quality and Reliability Management*, 17:9, pp1003-1016, 2000.
- [5] Madu, C., Competing Through Maintenance Strategies, *International Journal of Quality and Reliability Management*, 17:9, pp937-948, 2000.
- [6] Mirghani, M. A., A Framework for Costing Planned Maintenance, *Journal of Quality in Maintenance Engineering*, 7(3), pp170-182, 2001.
- [7] Mitchell, E., Robson, A., and Prabhu, V. B., The Impact of Maintenance Practices on Operational and Business Performance, *Managerial Auditing Journal*, 17(5), pp234-240, 2002.
- [8] Peter W. Tse, *Maintenance practices in Hong Kong and the use of the intelligent scheduler*, *Journal of Quality in Maintenance Engineering*, Vol. 8 No. 4, pp 369 – 380, 2002.
- [9] Eti, M.C., Ogaji, S.O.T., and Probert, S.D., *Development and implementation of preventive-maintenance practices in Nigerian industries*, *Applied Energy* 83, pp 1163-1179, 2006.
- [10] Eti, M.C., Ogaji, S.O.T., and Probert, S.D., *Impact of corporate culture on plant maintenance in the Nigerian electric-power industry*, *Applied Energy* 83, pp 299-310, 2006.
- [11] Al-Zoubi, M., *A Maintenance System for Belt Conveyors at Jordan Phosphate Mining Company*, Master's Thesis, University of Jordan, Amman, Jordan, 2003.
- [12] Al-Awaysheh, I., *Maintenance Activity-Based Costing: Analysis and its Role in Maintenance Management*, Master's Thesis, University of Jordan, Amman, Jordan, 2003.
- [13] Rayyan, L.A. and Tahboub, K.K., *An Assessment of the Perception of Jordanian Industries Towards TQM and ISO9000*, *Dirasat, Natural and Engineering Sciences*, Volume , 25, No. 2, 1998.
- [14] Montgomery, D. and Runger J., *Applied Statistics and Probability for Engineers*, John Wiley & Sons, Inc, 1994.

Efficiency of Free Cooling Technique in Air Refrigeration Systems

A.Al-Salaymeh^{*a}, M.R. Abdelkader^a

^aMechanical Engineering Department, Faculty of Engineering and Technology, University of Jordan, Amman 11942, Jordan

Abstract

Free cooling techniques can be used to substantially reduce energy costs. During cold weather, the outside ambient temperature can help in saving energy in refrigeration systems. The low temperature of the cooling ambient air supply enables free cooling technique to store fresh fruits and vegetables. This energy-efficiency measure can save enough compressor electric power to pay for modulating damper installation costs in approximately one year. Free cooling has a motorized damper that conducts the two flows of internal and external air. When the damper is open it takes the air necessary for cooling directly from the exterior, excluding compressor operation. It starts the evaporator fan that takes external air if $T_{\text{external}} < T_{\text{internal}}$. A case study has been carried out for 17 Ton cooling load in a storage room and the COP can be reached to 24 where the only energy consumption is from the use of evaporative fans.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: Refrigeration; Cooling; Efficiency; Energy Saving

1. Introduction

Free cooling techniques are used to substantially reduce energy costs. This system effectively provides nearly free cooling using the cooling air during cold weather without the requirement to run a compressor. The compressor is essentially shut off during this period, there by saving energy and also allowing scheduled preventative maintenance to take place. The low temperature of the cooling ambient air supply enables free cooling of cold stores for fruits and vegetables, e.g. Al-Salaymeh et al. [1]. The cooling and air refrigeration systems are designed to cool cold stores to keep the fruits and vegetables for long time without damage. They are designed for internal or external installation. Their use allows eliminating problems caused by high temperatures, dirt and humidity which are present in the environment. The unit, which is basically made up of a sealed cooling circuit where the coolant circulates, is divided into two suction, hermetically separated from each other, where the air in the environment and the air in the cabinet are treated without coming into the contact with each other. The air refrigeration system that uses free cooling techniques has a motorized damper that conducts the two flows of internal and external air. When the damper is open it takes the air necessary for cooling directly from the exterior, excluding compressor operation, e.g. Al-Salaymeh et al. [1]. If the external temperature is less than the internal temperature, the damper will be opened and the evaporator fan will start to take external air. If the external temperature is higher than the internal temperature, the damper remains closed and air is recycled.

Refrigeration in the simplest terms is cooling by removing heat. It could also be said that refrigeration is the transfer of heat from one place where it is not wanted to

another less objectionable place. The normal strategy in mechanical refrigeration is to get the heat to the refrigerant. Then transfer the refrigerant to a place where the heat can be removed from it. The refrigerant makes the heat transfer possible. Most of the heat transfer occurs because the refrigerant changes state. The liquid refrigerant in the evaporator absorbs its latent heat of vaporization, and in the process changes from a liquid to a vapor. The gas refrigerant within the condenser rejects its latent heat of vaporization, thus changing from a gas to a liquid. It is this cycle change that moves the removed heat from one place to another.

Al-Salaymeh et al. [1] mentioned that free cooling takes place when the external ambient air enthalpy is less than the indoor air enthalpy. They showed that free cooling may be used with mixed outside air and recirculation systems by the use of modulating dampers. Dampers are provided on the outside air intake ductwork, exhaust air ductwork and the recirculation ductwork. In the event of cool outside air the quantity of outside air is increased and the quantity of recirculated air is reduced to provide the required supply air temperature. In this way cooling by means of refrigeration equipment is avoided altogether at certain times of year and often at night times. This system of free cooling uses thermostats to determine when the outside temperature is lower than the room temperature and the proportion of opening the outside air damper. More accurately the proportion of outside air should be increased when the outside air temperature is lower than the room temperature. When the outside air temperature is higher than the cold store temperature, the dampers will modulate to the minimum outside air position to keep the load on the refrigeration equipment to a minimum.

The coefficient of performance, COP, is simply a ratio of the effect we want (a heat transfer) to the quantity that we must buy (work) in order to cause the desired effect. In the refrigeration system, the COP is the ratio between the

* Corresponding author. e-mail: salaymeh@ju.edu.jo

heat removed from the low-temperature reservoir in the case of refrigeration to the compressor and fan work. However, if the free cooling is used, the compressor is switched off and the input work is equal to the fan work. Song et al. [2] studied the energy performance of a cooling plant system using the inverter chiller for industrial building. They clarified the energy performance of the cooling plant system in the industrial building using actual measured operating data and numerical simulation analysis. One aspect of industrial buildings is that they have large energy consumption for manufacturing and air conditioning compared with office and commercial buildings. The maximum COP of the inverter chiller reaches about 18 under certain conditions and integrated cooling towers make lower temperature cooling water as the whole capacity is large. The actual operating data in their study indicates satisfied values for chiller and system COP during the running period and the simulation results show that the cooling plant system can cut down annual electric power consumption by about 48% compared with conventional cooling system. Mina et al. [3] investigated a generalized coefficient of performance for conditioning moist air. In air-conditioning systems, the general goal is to effect a change of state of air from one condition to another one that is more desirable. They examined expressions for determining the limiting coefficient of performance for conditioning air in a manner that combines the effects of sensible and latent processes. They developed an expression that accounts for the change in energy and exergy due to change in air temperature and water vapor content of the air, i.e. humidity ratio.

Medved and Arkar [4] made a correlation between the local climate and the free-cooling potential of latent heat storage. They presented studies of the free-cooling potential for different climatic locations. They selected six cities from around Europe with a wide range of climatic conditions for their study. In the majority of the presented free-cooling systems the emphasis is placed mainly on modeling the latent heat thermal energy storage's (LHTES) and less emphasis was placed on the free-cooling efficiency, where various approaches have been tried. Zalba et al. [5] determined the thermal response of a plate latent heat thermal energy storage's for a step change of the inlet temperature. When designing a real free-cooling system a constant ambient air temperature during the night-time and a constant indoor air temperature in the building during the daytime were assumed. Arkar et al. [6] used a project day with an assumed sinusoidal temperature variation for optimizing the geometrical and performance parameters of the LHTES. Arkar et al. [6] studied the efficiency of free cooling using latent heat storage integrated into the ventilation system of a low energy building. They presented the results of an investigation into the free cooling efficiency in a heavyweight and lightweight low energy building using a mechanical ventilation system with two LHTESs, one for cooling the fresh supply air and the other for cooling the re-circulated indoor air. They found that the free cooling technique enables a reduction in the size of the mechanical ventilation system, provides more favourable temperatures and therefore enables better thermal comfort conditions and also fresh air for the occupants.

Arkar and Medved [7] presented a study of the free cooling of a low-energy building using a LHTES device integrated into a mechanical ventilation system. They developed a numerical model of the LHTES to identify the parameters that have an influence on the LHTES's thermal

response, to determine the optimum phase-change temperature and to form the LHTES's temperature-response function. Their analyses of the temperatures in a low-energy building showed that free cooling with an LHTES is an effective cooling technique. Moeseke et al. [8] focused in their paper on the impact of management strategies for external mobile shadings and cooling by natural ventilation. For natural ventilation, strategies limiting the flow rate when external temperature drops are found to be efficient to save energy.

Zalba et al. [5] and Zhang et al. [9] defined the free-cooling as a means to store outdoors coolness during the night, to supply indoors cooling during the day. Zhang et al. [9] showed that free-cooling can make the indoor air temperature in the comfortable region all the year if the thermophysical properties of building envelope material are in the desired range. Those properties are obviously related to the outdoor climate condition, internal heat source intensity, building configuration, ventilation mode etc. For a given region and a given building, the critical values of those ideal thermal physical properties can be determined through modeling and simulation. Zalba et al. [5] studied the application of phase change materials (PCM) in free-cooling systems. The use of PCMs is suitable because of the small temperature difference between day indoors and night outdoors. They designed and constructed an installation that allows testing the performance of PCMs in such systems. They performed experiments following the design of experiments strategy and they developed an empirical model in which a real free-cooling system was designed and economically evaluated.

Ghiaus and Allard [10] pointed out that natural ventilation is one of the most effective techniques for cooling. They showed that its potential for cooling may be assessed by using a method based on the indoor-outdoor temperature difference of the free-running building, the adaptive comfort criteria and the outdoor temperature. They demonstrated that the free-running temperature may be used instead of the balance temperature in energy estimation methods. The indoor-outdoor temperature difference of the free-running building becomes a characteristic of the thermal behavior of the building which is decoupled from comfort range and outdoor temperature. They found that a measure related to the energy saved and the applicability of free-cooling is given by the probabilistic distribution of the degree-hours as a function of the outdoor temperature and time.

In this work, self-developed computer software for calculating the total cooling load for the cold stores that contain fruits or vegetables room storage at each hours of the year has been carried out. The cooling load for refrigeration unit (compressor and fan) or evaporator fans only before and after using free cooling techniques has been made. Then, a monthly electricity cost comparison before and after using free cooling is done and the percentage of the energy saving for each types of product is calculated. Finally, the payback period is calculated to check if the reduction in the electricity cost can cover the cost of the components of the new system in a short time and therefore to know if the free cooling system has a good application in cold stores for fruits and vegetables in Jordan.

2. Economizer Damper

There are two dampers on a typical economizer that modulate when the outdoor conditions are right. An outside damper opens to draw air into the room and a return air damper closes, cutting off the return air flow from the room. When this outdoor damper opens, the extra added air will pressurize inside the room, just like blowing up a balloon. This can cause outside doors to blow open or doors that open to the inside to stick closed. A small amount of pressure can have a big effect. A pressure relief damper is used upstream of the return air damper to relieve this pressure in the room. This may be a very simple metal square that will swing open under a small amount of pressure or a large mechanical systems will use a mechanical damper that opens in conjunction with the intake damper and a fan to expel the extra air. The free cooling version has a motorized damper that conducts the two flows of internal and external air. When the damper is open it takes the air necessary for cooling directly from the exterior, excluding compressor operation as shown in Figure (1).

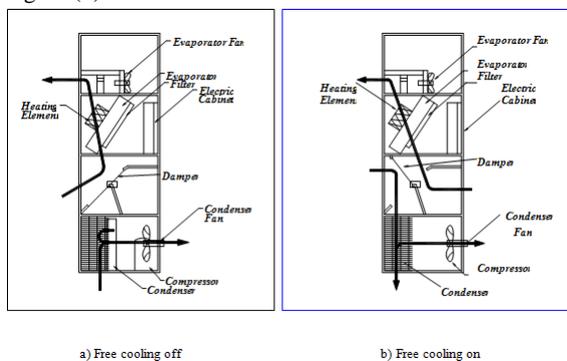


Figure 1: The motorized damper for free cooling purposes in air refrigeration unit which conducts the two flows of internal and external air.

The dampers are driven by an electric or pneumatic actuator, which is controlled from an outdoor temperature or enthalpy control. The control switches the system from mechanical cooling to economizer cooling when the outdoor conditions are right. Preferably an enthalpy control is used to sense both humidity and temperature. Some units have 'temperature only' controls that do not take in to consideration the effects of humidity, which can be considerable. If the enthalpy of the outdoor air is less than the enthalpy of the indoor air by volume, then the outdoor air can be used for cooling the building.

Some economizers use just a temperature control mounted on the intake hood with the minimum damper position controlled by a potentiometer on the actuator. Whenever the fan is running, the damper should drive to the minimum fresh air position. This may only amount to a slight crack in the damper opening but is enough to provide a standard 10% of the total fan volume of fresh air. When ever the fan is off this damper should drive completely closed by a return spring. Turn power off at the roof top unit while the fan is running and you should see the damper move closed.

When the outdoor conditions are below the control set point and the room thermostat is calling for cooling, the supply air fan will start and the economizer damper will modulate without starting the compressor. If the thermostat fan switch is set to "On" position the fan runs continuously and the damper will already be at the minimum fresh air

position. A second controls modulates the dampers positions to maintain a constant supply air temperature.

The sensor should be located downstream of the evaporator, this will prevent the air from over cooling if the compressor is running. If the economizer is integrated with the compressor control, it is very important for the sensing bulb to be downstream of the evaporator coil. This will close off the damper when the compressor comes on and stops it from drawing in air that is too cold for the evaporator and prevents the evaporator coil from freezing up. If the outdoor conditions rise above the set point, the dampers will modulate closed to the Minimum fresh air position and the mechanical cooling will start. If the enthalpy control is set too low, the economizer will not open and take advantage of the cool outdoor temperature. If it is set too high, then the building may get too warm, because the economizer cannot provide enough cooling. It may take a few adjustments to determine what set point work best for your system. Systems with integrated controls allow you to set the enthalpy control higher without comprising comfort. An integrated control allows the compressor and the economizer to run at the same time by using a two-stage thermostat. When the first stage calls for cooling, the economizer opens. If the conditions do not improve in the space and the temperature in the cold stores continues to climb, the thermostat will then call for second stage cooling. This brings on the compressor to provide additional cooling required. If the outdoor temperature is too cold to run the compressor another control called a "Low Ambient Lock Out" will prevent the compressor from starting. By keeping the mechanical compressor off a little longer we can maximize the savings with economizer.

3. Case Study

The main objective of air refrigeration system is to maintain the environment in the enclosed spaces at conditions that reached the storage temperature of the product in the spaces. The case study in the present work includes a design for cold stores to storage any types for fruits and vegetables. The calculations of cooling load for the products have been carried out per hour and therefore a self developed software program for calculations has been built. There are many factors that affect the free cooling calculations. When we design the storage room, the heat of respiration must be considered. Also, the period of storage is very important factor and should be taken into account. Sensible cooling load depends on differences of temperatures between inside and outside room storage, and value of chilling factor and chilling time which it varies with the type of product, that chilling time and factor and number of entering products each month. Also, we mustn't neglect the cooling load for boxes which have the same chilling factor and time and depends on the types of it. In addition to the previous calculation of cooling load for the products, another component that have relation to the total cooling load for the room such as walls, ceiling, ground, door, lights, and natural ventilation should be calculated.

After calculating the total cooling load for all hours in the year (8760 hours) for each type of stored products (Potato, Lemon and Tomato), then the maximum cooling load is selected for design purposes. The components of refrigeration cycle such as condensing unit and evaporator unit should be selected based on the maximum cooling load. Also, the total electricity bill which resulted from

operating the compressor and fans in the air refrigeration system has been calculated for each hour of the year. Then an economical study has been conducted based on the payback period for the cost of the new components of the free cooling technique such as motorized damper, controller system and sensors which results from the reduction in electricity cost.

In calculating the cooling load for cold stores, the self developed software has been carried out to calculate the cooling load for each hour in the year taken into account all structure components (walls, ceiling door, ground and door), the type of the products, and the type of boxes. Then the total cooling load for the stored products including the natural ventilation, person, and lights are calculated. The selection of the refrigeration cycle components and the new added components for free cooling techniques are based on the calculated cooling load. Then, the cooling load which needs fan work only after using free cooling techniques has been calculated and the reduction in cost by using this technique has been estimated. Finally, the payback period at which the new system covers the cost of the new components is found.

The dimensions of the storage room are 6 m*8 m*6 m and it has a door in the west direction and its dimension is 2.4 m x 2.0 m. Table (1) shows the overall heat transfer coefficients for the room of the cold store. The storage room has 3 lamp lights and the power of each lamp is equal to 250 W. It is assumed that the number of persons who enters the storage room is equal to 3. The period of opening the door of the storage room is assumed to be 3 hours for each time of inserting the products in the room or taking the products from it. The number of air change per hour in the room storage is taken to be equal to 4.84. Three types of products which are potato, Lemon and tomato have been stored in the storage room. The type of boxes is taken as wood which has a specific heat of $C_{p,boxes} = 1.38 \text{ kJ/kg.}^\circ\text{C}$.

Table 1: Overall heat transfer coefficients for the room of the cold store.

Type of walls	Construction	Thickness (mm)	U value (W/m ² .K)
Wall	L.W Concrete + Insulation	304.8	$U_{walls} = 0.642$
Ceiling	L.W Concrete + Insulation	203.4	$U_c = 0.715$
Ground	100 mm Fallen Board + 125 mm Tiles + 75 mm Reinforced Concrete		$U_g = 0.375$
Door	Type of door is Metal – Steel		$U_d = 5.8$

The total cooling load for the storage room before and after using free cooling has been calculated. Also, the economical study has been carried out and the payback period for the damper system is computed. The values of the outside temperatures for each hour of the year are taken from meteorological department and royal scientific society at Jordan. The heat gain from walls, ceiling, ground and door is calculated and the cooling load temperatures differences (CLTD) correction method is used. The total cooling load for products is divided into

sensible cooling load and the heat of respiration. Also, the heat from natural ventilation which depends on the number of change of air in the room for fruits and vegetables is taken into account. After summation all components of the cooling load, the total cooling load for the storage room at each hour around the year is found. The second step was the selection of the components of refrigeration cycles and then determines compressor and fan work and the mass flow rate of air through the fans.

For free cooling techniques, all hours around the year at which the outside temperature is less than the inside design temperature are determined. Then, the mass flow required at each hour is calculated. The cost of the electricity for compressor and fan is calculated in the presence of free cooling and in the absence of the free cooling and finally a comparison between the two costs is made. The cooling load which is required from compressor and/or fans before and after using free cooling is calculated. The energy saving for each type of product is estimated.

As it was mentioned above, the selection of the components of refrigeration cycle should be based on the maximum cooling load. In the present work, the selected air cooled condensing unit with semi-Hermetic Reciprocating compressor has a cooling capacity of 67.1 kW and a compressor work of 21.2 kW and it has been selected from Bitzer company. The required evaporator which is selected from Guntner Company has a mass flow rate of 12.48 kg/s and two fans; the work for each fan is 1.4 kW. The motorized damper has a size of 4 x 1.25 m².

4. Theory

The cooling load due to heat transfer from outside to the room of the cold stores through walls and ceiling in the storage room is:

$$Q_{w,c} = U_{w,c} A_{w,c} \Delta T \quad (1)$$

where U_w is the overall heat transfer coefficient for all construction for the walls, U_c is the overall heat transfer coefficient for all construction for the ceiling, A_w is the area for each wall, A_c is the area for the ceiling and ΔT is the cooling load temperatures differences CLTD and it is taken after correction.

$$(\text{CLTD})_{\text{corr.}} = (\text{CLTD} + \text{LM})k + (25.5 - T_i) + (T_{o,m} - 29.4)f \quad (2)$$

where CLTD is the cooling load temperatures differences before correction and it depends on the type of material for the walls and ceiling, LM is the latitude correction factor, T_i is the inside design temperature which equals the final storage temperature, $T_{o,m}$ is the outdoor mean temperature, k is the color adjustment factor, and f is the attic or room fan factor. The outdoor mean temperature is equal to:

$$T_{o,m} = \frac{T_{\text{max}} + T_{\text{min}}}{2} \quad (3)$$

where T_{max} is the average maximum daily temperature, and T_{min} is the average minimum daily temperature. The cooling load due to ground is

$$Q_g = U_g A_g (T_g - T_o) \quad (4)$$

where U_g is the overall heat transfer coefficient for all construction of the ground, A_g is the area for the ground and T_g is the temperature of the ground. The cooling load due to the door is

$$Q_d = U_d A_d (T_o - T_i) \quad (5)$$

where U_d is the overall heat transfer coefficient for the door, A_d is the area of the door, T_o is the outside temperature and T_i is the inside temperature. The cooling load due to light, persons and ventilation is calculated only when the door of the storage room is opened. The cooling load for light, e.g. Dossat and Horan [11], is:

$$Q_{Ll} = \frac{n * P * 3.6}{3600} \quad (6)$$

where P is the power for each lamp light, and n is the number of light lamps. The cooling load for heat gain due to persons, e.g. Dossat and Horan [11], is

$$Q_{person} = \frac{m * 753.62}{3600} \quad (7)$$

where m is the number of persons. The cooling load for natural ventilation is

$$Q_{vent} = \frac{V * c}{\rho} * C_p * (T_o - T_i) \quad (8)$$

where V is the volume of storage room, c is the number of change air per hour, and C_p is the specific heat for air. The cooling load for product consists of two components. The first component is the sensible cooling load from output design temperature to the inside design temperature:

$$Q_{product} = \frac{m_p C_{p, before} (T_o - T_i)}{t * F} \quad (9)$$

where m_p is the mass of the product, $C_{p, before}$ is the specific heat before freezing, t is the chilling time and F is the chilling factor. The second component for the cooling load of the product is the heat of respiration which is equal to:

$$Q_{breathing} = m_p R \quad (10)$$

where R is the reaction factor. The cooling load for boxes is:

$$Q_{boxes} = \frac{m_b C_{p, Boxes} (T_o - T_i)}{t * F} \quad (11)$$

where m_b is the mass of boxes, and $C_{p, boxes}$ is the specific heat before freezing. The mass flow rate of the cooling air that is needed to cover the final total cooling load for the room when applying the free cooling techniques is:

$$\dot{m}_{fan} = \frac{Q_{tot}}{C_p (T_o - T_i)} \quad (12)$$

The properties of the three stored products (Potato, Lemon and Tomato) which were used in the present investigation are summarized in Table (2), e.g. Dossat and Horan [11]. As it can be shown in Table (2), the three types of the products have approximately similar properties, but the maximum period for storage is high for potato and low for tomato.

Table (2): Stored Products Properties.

	Potato	Limon	Tomato
Storage Temperature (°C)	15	15	15
Max. Period For Storage	(4-6) Months	(1-4) Months	(3-5) Weeks
Chilling Factor	1	1	1
Chilling Time	20	20	34
C_p (before freezing) (kJ/kg°C)	3.6	3.76	3.85
Freezing Point (°C)	-1.8	-2.2	-0.8
Heat of Reaction (kJ/kg. hr)	0.162	0.144	0.302

The electricity cost per hour for fan and compressor work can be calculated from the following:

- If the total cooling load is less than or equal zero, then there is no need to use compressor and fans and therefore the electricity cost is zero.
- If the compressor is switched off and the cooling load is resulted from the free cooling, then the fan is operated and the electricity cost equals:

$$\text{Electricity cost}_{fan} = W_{fan} * PF * \text{Electricity cost per kWh} \quad (13)$$

where W_{fan} is the fans work and PF is the product factor.

- If the air refrigeration system is working, then the cooling load is resulted from the fan and compressor work and the electricity cost equals:

$$\text{Electricity cost}_{compressor+fan} = (W_{compressor} + W_{fan}) * PF * \text{Electricity cost per kWh} \quad (14)$$

where $W_{compressor}$ is the compressor work and PF is the product factor.

The coefficient of performance (COP) can be reached a very high value in the case of free cooling. As it is known, the COP is defined as the ratio between the output cooling load and input electrical power. The output cooling load is the same in the two cases which are the presence of free cooling air refrigeration unit and in the absence of free cooling. But, the electrical input power is reduced in the case of free cooling and the input power is only the power needed for evaporator fans. The value of the coefficient of performance (COP) is:

$$COP = \frac{Q_{Evaporator}}{W_{Total}} \quad (15)$$

As it can be seen from equation (15), the cooling load from evaporator which is required to keep the storage product at a certain temperature remains the same before and after using the free cooling technique. However, the COP depends on the total work. For the case study in the present work, the fan power in the air refrigeration system is 2.8 kW and the compressor power is 21.2 kW. The required cooling load has been calculated after applying the above equations and it is found to be equal to 67.1 kW. Therefore, the COP before using free cooling techniques is:

$$COP = \frac{Q_{Evaporator}}{W_{Total}} = \frac{Q_{Evaporator}}{W_{Compressor} + W_{Fan}} = \frac{67.1}{21.2 + 2.8} = 2.80$$

On the other hand, the COP after using free cooling techniques becomes:

$$COP = \frac{Q_{Evaporator}}{W_{Total}} = \frac{Q_{Evaporator}}{W_{Fan}} = \frac{67.1}{2.8} = 23.96$$

These results show that the COP increases to a very high value after using the free cooling techniques. The ratio between the COP before and after using free cooling technique can be calculated as:

$$\frac{COP_{with\ free\ cooling}}{COP_{without\ free\ cooling}} = \frac{23.96}{2.8} = 8.56$$

Finally, the payback period for the total cost of the new component such as control system and damper for free-cooling technique can be calculated as a ratio between the capital cost of the new component and the total saved money which is resulted from energy saving after the free cooling technique is applied.

$$\text{Payback Period} = \frac{\text{Capital Cost for Free Cooling Components}}{\text{Money Saved After Using Free Cooling Technique}} \quad (16)$$

In the case study of the present investigation, it has been found that the payback period for each type of stored product is approximately one year.

5. Result

Self-developed computer software was designed and built to calculate the total cooling load per each hour in the year by using different types of storage products with and without free cooling. Utilization of free cooling techniques depends on the if-clause sentences in using mass flow rate of evaporator and with the total cooling load needed. The results of the cooling load which needs compressor work before and after using free cooling are plotted. Then, a comparison between the monthly electricity cost before and after using free cooling for each types of product is carried out as well as the percentage of the electricity cost for fan and compressor to the total cost in each month. The results are drawn for potato to show the performance of free cooling technique.

The free cooling technique is used when the outside temperature is below the inside storage temperature and in this case the evaporator fan is operated without compressor to reduced the electricity cost. Figure (2) shows the hourly ambient temperature for January and July months as a function of month days for Amman city where the cold stores in the present study are found. As we see from this

figure, there are many days in the year where the ambient temperature is below the inside desired temperature and therefore free cooling technique can be used beside the air refrigeration device to keep the cold stores at a constant temperature. As we know, the cold stores contain some vegetables or fruit products which have respiration heat and so it needs to be cooled all time around the year. By using free cooling technique, the number of operating hours for the compressor of air refrigeration system can be reduced and therefore the electricity cost can be minimized.

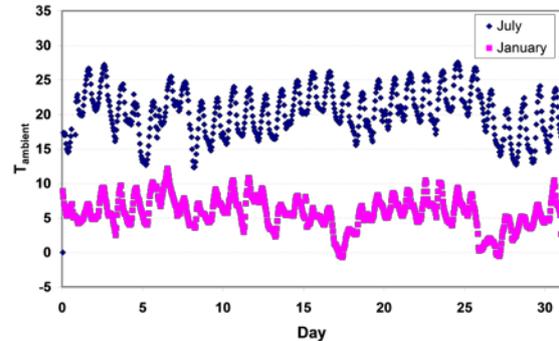


Figure 2: A comparison between the hourly average ambient temperature for January and July months as a function of month days for the location of the case study in Amman.

The cooling load at different months which needs compressor work before and after using free cooling technique for potato are presented in Figures (3)-(7) which show that the best month for using free cooling technique is January because the outside temperature is very low. Figure (3) shows that there is no need to switch on the compressor during January month and therefore the compressor work is equal zero. Also, in January month, the fan is switched off in many hours because the heat loss from walls, ceiling and doors is more than the cooling load which needed from the product. Also, it has been found that the total cost of electricity is decreased 83% for potato in this month with respect to the value before using free cooling.

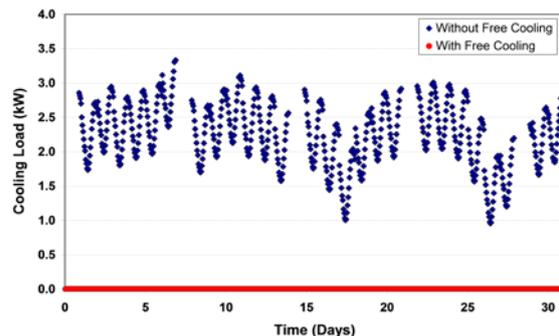


Figure 3: Cooling load in January month which needs compressor work before and after using free cooling technique for potato.

In March month, there is no need to switch on the compressor in some days to keep the potato product in the cold store as shown in Figure (4). Also, the evaporator fan is switched off for short time and this is due to the heat loss from walls, ceiling and doors is higher than the cooling load which is needed from the product. If a new product is inserted in the cold stores, then the air refrigeration unit should be operated for short time to decrease the temperature of the product to the required value. In this month, it has been found that the percentage

of reduction in the total cost of electricity is about 66.5 % with respect to the value before using free cooling. In April month, the effectiveness of the free cooling technique is less than cold months and the percentage of energy saving after using free cooling technique in this month can reach 41%. Figure (5) shows the cooling load in April month that needs compressor work before and after using free cooling technique for potato product. At the end of the month, using of free cooling is very low compared with compressor work and this is due to the fact that a new product was inserted in the cold stores and its cooling load is high.

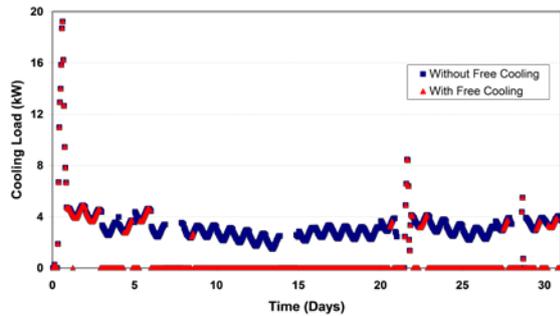


Figure 4: Cooling load in March month which needs compressor work before and after using free cooling technique for potato.

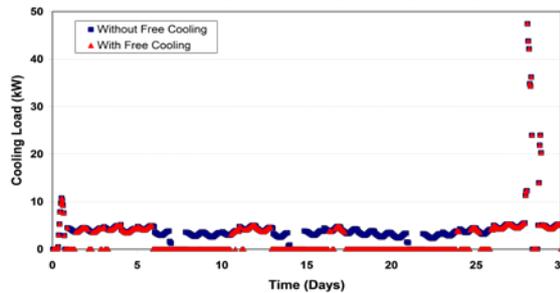


Figure 5: Cooling load in April month which needs compressor work before and after using free cooling technique for potato

The summer months such as July and August can be considered as the worst months for using free cooling techniques. The percentage of energy saving in these months is very small and negligible. For example, in August month the percentage of energy saving can not reach more than 5% for all periods of the month as shown in Figure (6). In November month, the ambient temperature is dropped again and therefore the effect of free cooling starts to increase again. Figure (7) presents the cooling load in November month before and after using free cooling technique. It is clear from Figure (7) that the percentage of energy saving is high and the compressor is switched off in the most days of November month.

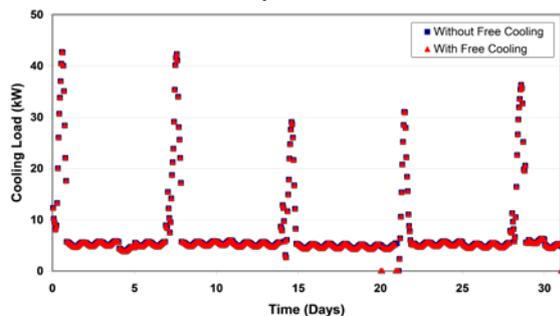


Figure 6: Cooling load in August month which needs compressor work before and after using free cooling technique for potato.

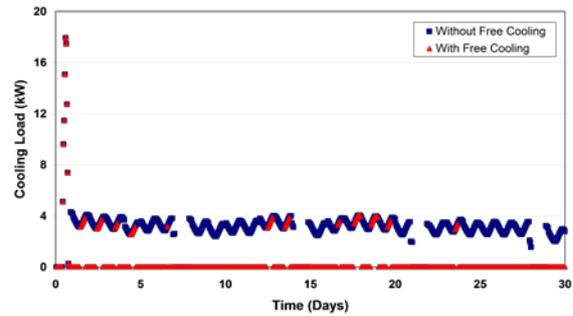


Figure 7: Cooling load in November month which needs compressor work before and after using free cooling technique for potato.

A comparison of the saved money before and after utilizing of free cooling technique has been carried out and the total money saved for the three studied stored product in each month is presented Table (3). The total cost of operating air refrigeration unit in each month that operates 24 hours per day is the summation of the electricity cost for compressor and evaporator fans and it is the electricity cost for evaporator fans in the presence of free cooling technique. When the compressor switched off, the evaporator fan of the air refrigeration unit is continuously operated and the cost of operating the evaporator fan has been calculated. Usually, a payback period can be taken as guide for justifying the economical aspect of utilizing the free cooling technique in air refrigeration systems. Assuming that the cost of 1.0 kWh electricity in Jordan is about 0.063 JD/kWh, then we can calculate the cost of operating the air refrigeration unit which has 21.2 kW of an electrical input power for the compressor and 2.8 kW of an electrical input power for the two evaporator fans. The cost of electrical energy consumption is calculated by multiplying the compressor and fans power with the number of operating hours for air refrigeration unit. The total cost for the components of free cooling technique (control system and motorized damper) for the refrigeration unit in the present investigation is approximately equal to 3000 JD. The data presented in Table (3) shows that the payback period for each type of product is approximately one year. The payback period of free cooling components is reasonable and therefore it is recommended to use free cooling technique in all air-conditioning and air refrigeration units.

Table 3: Money saved in (JD) from using free cooling technique in each month and for three different types of product.

Months	Money Saved (JD)		
	Potato	Lemon	Tomato
January	482.5	482.5	490
February	443.4	440.2	446.6
March	406.7	405.6	416.3
April	243	243	250.4
May	158.1	158.1	154.9
June	103.6	103.6	102.6
July	32.5	33.4	32.5
August	76.9	76.9	69.4
September	75.9	75.9	69.4
October	158.1	209.4	219
November	456.7	440.2	452.4
December	479.3	479.3	490
Total	3116.7	3148.2	3193.6

A comparison between the monthly saved money for air refrigeration system before and after utilizing of free cooling technique has been carried out and the results are presented in Figure (8). Figure (8) illustrates that there is a saving in electricity and therefore saving money if free cooling technique is used in the air refrigeration unit and as it is expected, more energy or money can be saved if the free cooling technique is implemented in winter months. Figure (8) presents also the electricity cost for compressor with and without free cooling in addition to the fan at different months that is needed to keep the temperature inside the storage room where the study has been carried out at a constant temperature around 15°C. This figure shows also the electricity cost which resulted from number of daily operating hours for evaporator fan at different months for desired location in Amman city in the case of utilization of free cooling. It is clear from this figure that the electricity cost and therefore number of operating hours for free cooling technique which is represented by evaporator fan decreases in summer where the operating hours for compressor increases. The number of operating hours for evaporator fan is higher than the number of operating hours for compressor at the most time in year except the summer season. The air refrigeration unit in the storage room is permanently operated and as long as there is vegetables or fruits inside the room and therefore the evaporator fan will be worked in the case at which the compressor is switched off. It is clear from this figure that the number of operating hours for evaporator fan decreases as the number of operating hours for compressor in summer increases. Since our selected air refrigeration unit has a total input electrical power of 24 kW (21.2 kW for compressor and 2.8 kW for fan), therefore the total cost of operating such a unit during each month at the desired location has been calculated. As it is expected, the maximum electricity cost occurs during summer time especially at July and August months.

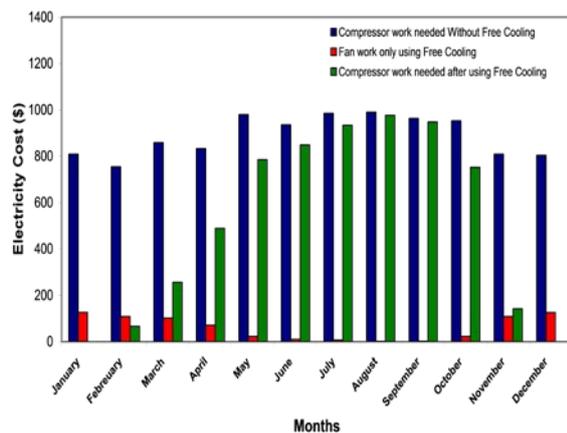


Figure 8: A comparison between the total monthly electricity cost in JD for compressor work with and without using free cooling technique in addition to the evaporator fan in the presence of free cooling for Potato.

Figure (8) summarizes the results for the total electricity cost of the selected air refrigeration unit in addition to the cost of the compressor and evaporator fan during all months in the year. There are some of months in the year such as January month where the compressor is not needed to be operated. The electricity cost for operating air refrigeration unit in the presence of free cooling is lower than the electricity cost in the absence of

free cooling. Figure (8) shows that the maximum electricity cost for evaporator fans occurs during winter season. The ratio of the total electricity cost that is resulted from operating the compressor and/or the evaporator fans in the air refrigeration unit with and without free cooling is shown in Figure (9). The results presented in Figure (9) show that 83% from the total cost of electricity can be saved in January month and this value is decreased to be less than 5% in summer months such as July. The capital cost of the free cooling components such as the control system and the motorized dampers can be recovered after one year from operation. A comparison between the total electricity cost for operating the air refrigeration unit in the case of presence or absence of free cooling is shown in Figure (9). The amount of money saving per each month in the case of utilizing free cooling technique is clearly presented in this figure.

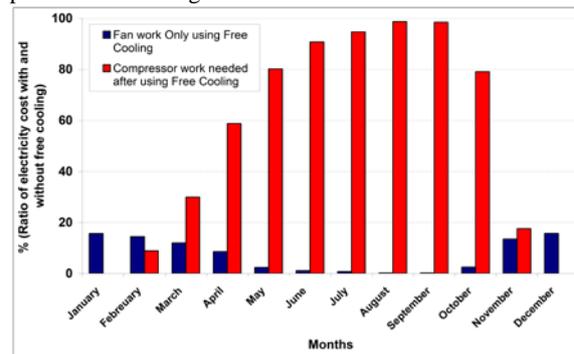


Figure 9: Percentage of the electricity cost for fan and compressor to the final cost in each month for Potato.

6. Conclusions

If the outside ambient wet-bulb temperature drops below the required cooling air set-point, then free cooling can be used to save energy whenever. Free cooling technique is environmental friendly and its usage will have a positive impact on reduction the CO₂ emission and on the global warming phenomena. The main purpose of this study is to focus some sheds on the utilizing of free cooling in the air refrigeration system especially for cold stores which contain some vegetable or fruits that required to be cooled all days during the storage period. If the outside temperature is less than the set point of the cold stores, then a cooling load is required due to the heat of respiration of the products. The air refrigeration system in small cold stores uses free cooling techniques to substantially reduce energy costs. The typical value for the coefficient of performance is located between 3 and 5. However there are periods of the year where cooling ambient temperature can be utilized to provide nearly free cooling and the COP can reach to 24 where the only energy consumption is from the use of evaporative fans. Therefore, the coefficient of performance after using free cooling techniques reaches approximately 8 times its value before using this technique.

During free cooling the compressor of the air refrigeration unit is essentially shut off, thereby saving energy and also allowing scheduled preventative maintenance to take place. It has been found that this energy-efficiency measure is economically attractive and can save enough compressor electric work to pay for the installation cost of the damper for free cooling purposes in less than one year in Jordan.

Free cooling system is a good technique to store fresh products such as potato, lemon and tomato in Jordan because the outside temperature is less than the storage temperature of some products in cold months. The utilization of free cooling techniques in cold stores depends on the location of the stores and a high energy saving can be received if the ambient temperature in the location of the cold stores is low. Free cooling technique is very efficient to be implemented in the air conditioning units and air refrigeration systems during cold or winter

months such as January, February, November and December and the percentage in energy reduction can reach more than 80%, because the compressor is switched off most of time in these months. However, in other months such as March and April, the energy saving in these months is good and it can reach 50% after free cooling technique is used. On the other hand, in summer semester the percentage of energy saving that resulted from free cooling technique is less than 10% because the ambient temperature is high.

References

- [1] A. Al-Salaymeh, M. Al-Salaymeh, M. Rabah, M. abdelkader, "Enhancement of the coefficient of performance in Air Conditioning systems by Utilizing Free Cooling", Proceedings of the 2nd International Conference on Thermal Engineering: Theory and Application, Al-Ain, UAE, Jan. 3-6, 2006.
- [2] Y. Song, Y. Akashi, J. Yee, "Energy performance of a cooling plant system using the inverter chiller for industrial building", Energy and Buildings, Vol. 39, 2007, p. 289-297.
- [3] E.M. Mina, T.A. Newell, A.M. Jacobi, "A generalized coefficient of performance for conditioning moist air", International Journal of Refrigeration, Vol. 28, 2005, p. 784-790.
- [4] S. Medved, C. Arkar, "Correlation between the local climate and the free-cooling potential of latent heat storage", Energy and Buildings, Vol. 40, 2008, p. 429-437.
- [5] Zalba, J.M. Marin, L.F. Cabeza, H. Mehling, "Free-cooling of buildings with phase change materials", International Journal of Refrigeration, Vol. 27, No. 8, 2004, p. 839-849.
- [6] Arkar, B. Vidrih, S. Medved, "Efficiency of free cooling using latent heat storage integrated into the ventilation system of a low energy building", International Journal of Refrigeration, Vol. 30, No. 1, 2007, p. 134-143.
- [7] Arkar, S. Medved, "Free cooling of a building using PCM heat storage integrated into the ventilation system", Solar Energy, Vol. 81, 2007, p. 1078-1087.
- [8] G. Moeseke, I. Bruyere, A. Herde, "Impact of control rules on the efficiency of shading devices and free cooling for office buildings", Building and Environment, Vol. 42, 2007, p. 784-793.
- [9] Y. Zhang, K. Lin, Q. Zhang, H. Di, "Ideal thermophysical properties for free-cooling (or heating) buildings with constant thermal physical property material", Energy and Buildings, Vol. 38, 2006, p. 1164-1170.
- [10] C. Ghiaus, F. Allard, "Potential for free-cooling by ventilation", Solar Energy, Vol. 80, 2006, p. 402-413.
- [11] R.J. Dossat, T.J. Horan, "Principles of Refrigeration", 5th Ed. Prentice Hall; 2002.

A Comparison Study between Various Fuzzy Clustering Algorithms

K. M. Bataineh^{*a}, M. Najj^a, M. Sager^a

^aDepartment of Mechanical engineering, Jordan University of Science and Technology, Irbid, Jordan

Abstract

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset shares some common features. This paper reviews and compares between the two most famous clustering techniques: Fuzzy C-mean (FCM) algorithm and Subtractive clustering algorithm. The comparison is based on validity measurement of their clustering results. Highly non-linear functions are modeled and a comparison is made between the two algorithms according to their capabilities of modeling. Also the performance of the two algorithms is tested against experimental data. The number of clusters is changed for the fuzzy c-mean algorithm. The validity results are calculated for several cases. As for subtractive clustering, the radii parameter is changed to obtain different number of clusters. Generally, increasing the number of generated cluster yields an improvement in the validity index value. The optimal modelling results are obtained when the validity indices are on their optimal values. Also, the models generated from subtractive clustering usually are more accurate than those generated using FCM algorithm. A training algorithm is needed to accurately generate models using FCM. However, subtractive clustering does not need training algorithm. FCM has inconsistency problem where different runs of the FCM yields different results. On the other hand, subtractive algorithm produces consistent results.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: data clustering; fuzzy c-means; subtractive clustering; system modeling

1. Introduction

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset shares some common features, often proximity according to some defined distance measure. Clustering plays an important role in our life, since we live in a world full of data where we encounter a large amount of information. One of the vital means in dealing with these data is to classify or group them into a set of categories or clusters. Clustering finds application in many fields. For example, data clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Also, clustering is used to discover relevance knowledge in data.

Clustering finds application in system modelling. Modelling of system behaviour has been a challenging problem in various disciplines. Obtaining a mathematical model for a complex system may not always be possible. Besides, solving a mathematical model of a complex system is difficult. Fortunately, clustering and fuzzy logic together provide simple powerful techniques to model complex systems. Clustering is considered powerful tool for model construction. It identifies the natural groupings in data from a large data set, which allows concise representation of relationships hidden in the data. Fuzzy logic is efficient theory to handle imprecision. It can take imprecise observations for inputs and yet arrive to precise

values for outputs. The Fuzzy Inference System (FIS) is a simple way to construct systems models without using complex analytical equations.

Several clustering techniques have been developed. Hierarchical clustering produces a graphic representation of data [1]. This method is often computationally inefficient, with the possible exception of patterns with binary variables. Partitional clustering is considered the second general category of clustering. It concerns with building partitions (clusters) of data sets according to the relative proximity of the points in the data sets to each other. Generally, algorithms can be categorized according to their way in unveiling the patterns inside the raw data sets. These classifications are the objective function and the mountain function based clustering algorithms.

In Sugeno type models, the consequent of a rule can be expressed as a polynomial function of the inputs and the order of the polynomial also determines the order of the model. The optimal consequent parameters (coefficients of the polynomial function) for a given set of clusters are obtained by the Least Square Estimation (LSE) method.

This paper reviews and compares between the two famous clustering techniques: Fuzzy C-mean algorithm and Subtractive clustering algorithm. These methods are implemented and their performances are tested against highly nonlinear functions and experimental data from Ref. [2]. Three validity measures are used to assess the performance of the algorithms: Dave's, Bezdek and Xie and Beni indices. Also a comparison between the FCM algorithm, Subtractive algorithm and algorithms presented in Alta et al. [3] and Moaqt [4] is made. The effects of different parameters on the performance of the algorithms are investigated. For the FCM, the number of clusters is

* Corresponding author. e-mail: k.bataineh@just.edu.jo

changed and the validity results are calculated for each case. For subtractive clustering, the radii parameter is changed to obtain different number of clusters. The calculated validity results are listed and compared for each case. Models are built based on these clustering results using Sugeno type models. Results are plotted against the original function and against the model. The modeling error is calculated for each case.

2. Data Clustering Techniques

In this section a detailed discussion of FCM and Subtractive clustering is presented.

2.1. Fuzzy c-Means clustering:

Fuzzy c-means algorithm (FCM), also known as Fuzzy ISODATA, is by far the most popular objective function based fuzzy clustering algorithm. It is first developed by Dunn [5] and improved by Bezdek [6]. The objective (cost) function used in FCM presented in [6] is:

$$J_m(U, V, X) = \sum_{i=1}^c \sum_{j=1}^N u_{ik}^m d_{ik}^2 = \sum_{i=1}^c \sum_{j=1}^N u_{ik}^m \|x_j - v_i\|^2 \tag{1}$$

Where, $U = [u_{ik}] \in M_{fc}$ is a fuzzy partition matrix of Z . $V = [V_1, V_2, V_3, \dots, V_c], V_i \in R^n$ is a vector of cluster prototypes (centers), which has to be determined. $D_{ikA} = \|x_k - V_i\|^2_A = (x_k - V_i)^T A (x_k - V_i)$ is a squared inner-product distance norm, and $m \in [1, \infty)$ is a parameter which determines the fuzziness of the resulting clusters. The value of the cost function can be seen as a measure of the total variance of x_k from v_i . The necessary conditions for minimizing equation (1) are:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}}}, 1 \leq i \leq c, 1 \leq k \leq n \tag{2}$$

And

$$v_k = \frac{\sum_{i=1}^c u_{ik}^m x_k}{\sum_{i=1}^c \left(\frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}}}, 1 \leq i \leq c \tag{3}$$

Several parameters must be specified in order to carry FCM algorithm; the number of clusters c , the fuzziness exponent m , the termination tolerance ϵ , and the norm-inducing matrix A . Finally, the fuzzy partition matrix U must be initialized. The number of clusters c is the most important parameter. When clustering real data without any priori information about the structures in the data, one usually has to make assumptions about the number of underlying clusters. The chosen clustering algorithm then searches for c clusters, regardless of whether they are really present in the data or not. The validity measure approach and iterative merging or insertion of clusters approach are the main approaches used to determine the appropriate number of clusters in data. Figure 1 shows the steps of carrying FCM algorithm.

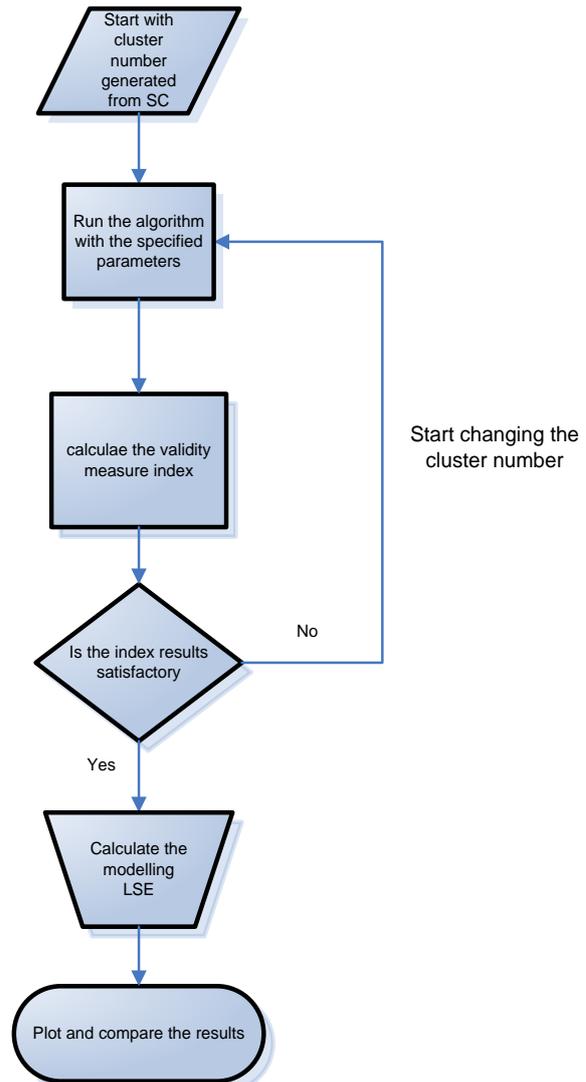


Figure 1: FCM clustering algorithm procedure.

2.2. Subtractive clustering algorithm:

Clustering algorithms typically require the user to pre-specify the number of cluster centers and their initial locations. The Fuzzy C-Means algorithm [6] is well-known examples of such clustering algorithms. The quality of the solution depends strongly on the choice of initial values (i.e., the number of cluster centres and their initial locations). Yager and Filev [7] proposed a simple and effective algorithm, called the mountain method, for estimating the number and initial location of cluster centers. Their method is based on gridding the data space and computing a potential value for each grid point based on its distances to the actual data points. A grid point with many data points nearby will have a high potential value. The grid point with the highest potential value is chosen as the first cluster center. The key idea in their method is that once the first cluster center is chosen, the potential of all grid points is reduced according to their distance from the cluster center. Grid points near the first cluster center will have greatly reduced potential. The next cluster center is then placed at the grid point with the highest remaining potential value. This procedure of acquiring new cluster center and reducing the potential of surrounding grid points repeats until the potential of all grid points falls

below a threshold. Although this method is simple and effective, the computation grows exponentially with the dimension of the problem because the mountain function has to be evaluated at each grid point.

Chiu [8] proposed an extension of Yager and Filev's mountain method, called subtractive clustering.

This method solves the computational problem associated with mountain method. It uses data points as the candidates for cluster centers, instead of grid points as in mountain clustering. The computation for this technique is now proportional to the problem size instead of the problem dimension. The problem with this method is that sometimes the actual cluster centres are not necessarily located at one of the data points. However, this method provides a good approximation, especially with the reduced computation that this method offers. It also eliminates the need to specify a grid resolution, in which tradeoffs between accuracy and computational complexity must be considered. The subtractive clustering method also extends the mountain method's criterion for accepting and rejecting cluster centres.

The parameters of the subtractive clustering are χ_i is the normalized data vector of both input and output dimensions defined as: $\chi_i^i = \frac{x_i^i - \min\{x^i\}}{\max\{x^i\} - \min\{x^i\}}$, n is

total number of data vectors, r_a is hyper sphere cluster radius in data space, r_b is the hyper sphere penalty radius in data space, P_i is the potential value of data vector i , η is the squash factor $= \frac{r_a}{r_b}$, $\bar{\epsilon}$ is the accept ratio, and $\underline{\epsilon}$ is

the reject ratio. The subtractive clustering method works as follows. Consider a collection of n data points $\{x_1, x_2, x_3, \dots, x_n\}$ in an M dimensional space. Without loss of generality, the data points are assumed to have been normalized in each dimension so that they are bounded by a unit hypercube. Each data point is considered as a potential cluster center. The potential of data point x_i is defined as:

$$P_i = \sum_{j=1}^n e^{-\frac{4\|x_i - x_j\|^2}{r_a^2}} \quad (4)$$

Where, the symbol $\| \cdot \|$ denotes the Euclidean distance, and r_a is a positive constant. Thus, the measure of the potential for a data point is a function of its distances to all other data points. A data point with many neighboring data points will have a high potential value. The constant r_a is effectively the radius defining a neighborhood; data points outside this radius have little influence on the potential. After the potential of every data point has been computed, we select the data point with the highest potential as the first cluster center. Let x_1^* be the location of the first cluster center and P_1^* be its potential value. We then revise the potential of each data point x_i by the formula:

$$p_i = p_i - p_1^* e^{-\frac{4\|x_i - x_1\|^2}{r_b^2}} \quad (5)$$

where r_b is a positive constant. Thus, we subtract an amount of potential from each data point as a function of its distance from the first cluster center. The data points

near the first cluster center will have greatly reduced potential, and therefore will unlikely be selected as the next cluster center. The constant r_b is effectively the radius defining the neighborhood which will have measurable reductions in potential. When the potential of all data points has been revised, we select the data point with the highest remaining potential as the second cluster center. This process continues until a sufficient number of clusters are obtained. In addition to these criteria for ending the clustering process are criteria for accepting and rejecting cluster centers that help avoid marginal cluster centers. Figure 2 demonstrates the procedure followed in determining the best clustering results output from the subtractive clustering algorithm.

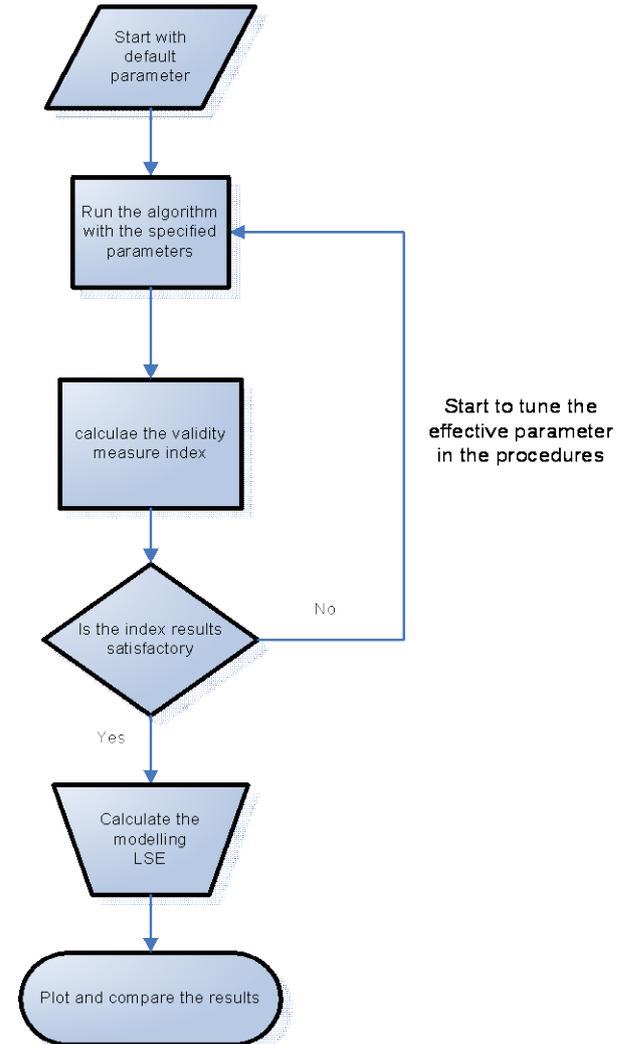


Figure 2: Subtractive clustering algorithm procedure.

3. Validity Measurement

This section presents the widely accepted three validity indices: Bezdek, Dave's, and Xie-Beni. These indices are used to evaluate the performance of each algorithm. Validity measures are scalar indices that assess the goodness of the obtained partition. Clustering algorithms generally aim to locate well separated and compact clusters. When the number of clusters is chosen equal to the number of groups that actually exist in the data, it can be expected that the clustering algorithm will identify them

correctly. When this is not the case, misclassifications appear, and the clusters are not likely to be well separated and compact. Hence, most cluster validity measures are designed to quantify the separation and the compactness of the clusters. However, as Bezdek [6] pointed out that the concept of cluster validity is open to interpretation and can be formulated in different ways. Consequently, many validity measures have been introduced in the literature, Bezdek [6], Dave's [9], Gath and Geva [10], Pal and Bezdek [11].

3.1. Bezdek index :

Bezdek [6] proposed an index that is sum of the internal products for all the member ship values assigned to each point inside the U output matrix. Its value is between $[1/c, 1]$ and the higher the value the more accurate the clustering results is. The index was defined as:

$$V_{PC} = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n u_{ij}^2 \quad (6)$$

In general, an optimal cluster number c^* that produces the best clustering performance for the data is found by solving:

$$\max_{2 \leq c \leq n-1} V_{PC} \quad (7)$$

3.2. Dave's Validity Index:

The Dave's validity index presented in [9] has been successfully used by many researchers. It has an excellent sense to hidden structures inside the data set. Dave's [9]'s defined the validity measure as:

$$V_{MPC} = 1 - \frac{c}{c-1} (1 - V_{PC}) \quad (8)$$

Where V_{PC} is defined in equation (6)

This is a modified version of validity measure V_{PC} proposed by Bezdek [6]. Dave's index usually ranges between 0 and 1. The higher the value of the index is, the better the results are. The study by Wang et al [12] reported that V_{MPC} has successfully discovered the optimal number of clusters in most of the testing benchmarks.

3.3. Xie and Beni Index:

Xie and Beni validity index [13] is defined as:

$$\chi(Z; U, V) = \frac{\sum_{i=1}^c \sum_{k=1}^N u_{ik}^m \|z_k - v_i\|^2}{c \min_{i,j} (\|v_i - v_j\|^2)} \quad (9)$$

This index has been found to perform well in practice. It can be interpreted as the ratio of the total within-group variance and the separation of the cluster centres. The optimal c^* that produces the best clustering performance for the data is found by solving

$$\min_{2 \leq c \leq n-1} \chi(Z, U, V)$$

4. Results and Discussion

The Fuzzy C-mean algorithm and Subtractive clustering algorithm are implemented to find the number & the position of clusters for a set of highly non-linear data. The clustering results obtained are tested using validity measurement indices that measure the overall goodness of the clustering results. The best clustering results obtained are used to build input/output fuzzy models. These fuzzy models are used to model the original data entered to the algorithm. The results obtained from the subtractive clustering algorithm are used directly to build the system model, whereas the FCM output entered to a Sugeno-type training routine. The obtained least square modeling error results are compared against similar recent researches in this field. When possibility allowed, the same functions and settings have been used for each case.

This section starts with presenting some simple example to illustrate the behavior of the algorithms studied in this paper. Next two highly nonlinear functions are modeled using both algorithms. Finally experimental data are modeled.

4.1. Illustrative cases:

In order to clarify the basic behavior of the two clustering algorithms; three cases are discussed:

4.1.1. Sine wave data modeling

Figure 3 compares between FCM and Subtractive clustering in modeling sine wave data. For this data, Subtractive clustering outperforms the FCM as the eventual goal of the problem is to create a model for the input system. The FCM prototypes are likely to be in the middle area especially in the rounded area as in the peak and the bottom of the sinusoidal wave. This is because the FCM would always estimate the cluster center over a circular data area. When remodeling these prototypes to obtain the original system model, the prototypes generated by the FCM will be shifted outward. As a result, the entire model will be shifted away from the original model. The FCM resulting cluster centers are highlighted by a circle. We can see that subtractive clustering was successful in defining the cluster center because the subtractive clustering assumes one of the original data points to act as the prototype of the cluster center. The resulting subtractive clustering centers are pointed by black box.

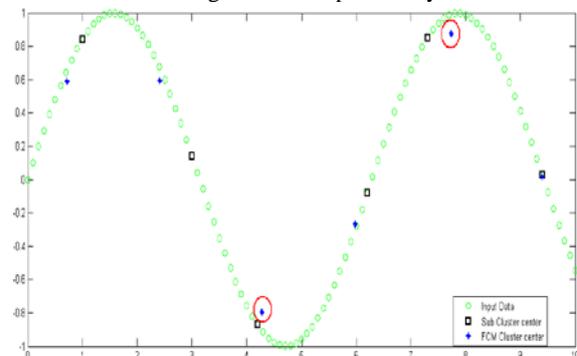


Figure 3: Sine wave data clustering, FCM vs Subtractive.

4.1.2. Scattered data clustering

The behavior of the two algorithms is tested against scattered discontinuous data. The subtractive still deploy the same criteria to find out the cluster center from the system available points. Thus, it might end to choose relatively isolated points as a cluster center. Figure 4 compares between the two algorithms with scattered data. The suspected points for Subtractive algorithms are circled. However, the FCM still deploy the weighted center. It can reach more reasonable and logical distribution over the data. The performance of the FCM is more reliable for this case.

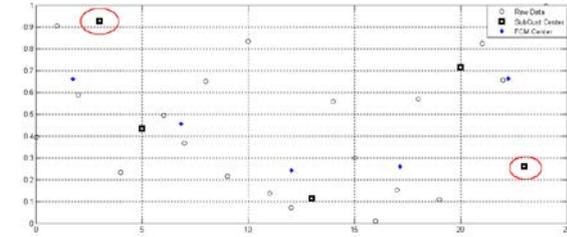


Figure 4: Scattered points clustering FCM versus Subtractive.

4.1.3. Well separated clusters

In this section, comparison between the two algorithms for clustering well separated clusters is presented. A sample data set provided by Mathwork [50] that contains 600 points distributed in three well defined clusters, are clustered using FCM and Subtractive algorithms. The effect of the number of clusters parameter in the FCM performance is investigated. Also the effect of radii parameters in the Subtractive algorithm is studied. Dave's, Bezdek and Xi-Beni validity indices variations with number of clusters are calculated and listed.

Table 1 lists the results by using FCM with varying number of clusters. It can be easily seen from table 1 that there are three well-defined clusters.

Table 1: FCM clustering algorithms VS number of clusters.

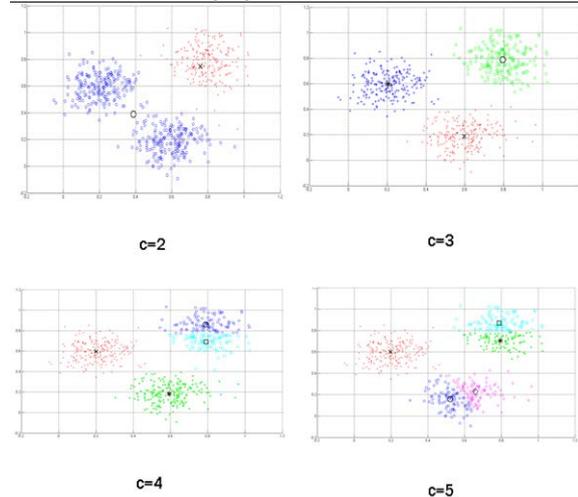


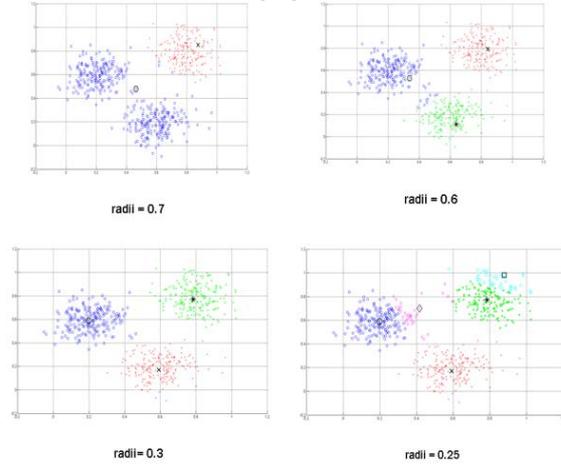
Table 2 list of Dave's validity index V_{MPC} and Bezdek index V_{PC} as a function of number of clusters assigned for FCM clustering. The parameters like fuzziness $m=2$, number of iteration =100, and minimum amount of improvement $1*10^{-5}$ are used. It can be seen from table 2 that the optimal value for the validity indices obtained when the number of chosen clusters is three, which reflects the natural grouping of the structure inside this data sample.

Table 2: Dave's validity index V_{MPC} and Bezdek index V_{PC} versus number of clusters c .

Number of clusters c	2	3	4	5
V_{MPC}	0.5029	0.7496	0.6437	0.5656
V_{PC}	0.7514	0.8331	0.7328	0.6500

As discussed earlier, Subtractive algorithm has many parameters to tune. In this study, only the effect of radii parameter has been investigated. Since this parameter has the highest effect in changing the resulting clusters number. Table 3 lists the results by using Subtractive algorithm with varying the radii.

Table 3: Subtractive clustering algorithms Vs radii.



Dave's validity and Bezdek indices are computed for several values of radii parameter for subtractive clustering. Parameters such as squash factor = 1.25, accept ratio = 0.5 and reject ratio = 0.15 are used. Table 4 lists the radii parameter with the generated number of clusters versus Dave's validity index V_{MPC} and Bezdek index V_{PC} . It is apparent from the validity measures that the optimal clustering results are obtained with 3 clusters which generated from using radii = 0.3 units. Even though the radii=0.6 and radii =0.3 generates three clusters, the V_{MPC} for radii =0.3 is better than that associated with 0.6. Similar behavior is observed for Bezdek validity index.

Table 4: Dave's validity index V_{MPC} and Bezdek index V_{PC} versus radii for Subtractive algorithm.

Radii	0.7	0.6	0.3	0.25
Number of clusters c	2	3	3	5
V_{MPC}	0.5071	0.6515	0.7478	0.5715
V_{PC}	0.7536	0.7677	0.8318	0.6572

Figures 5 and 6 show the Bezdek and the Dave's validity index variations with number of cluster for both FCM and Subtractive algorithms, simultaneously. Both indices agree that the optimal values for both algorithms are obtained when three clusters are used.

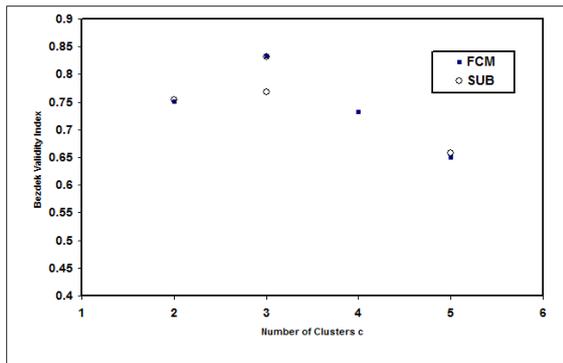


Figure 5: Bezdek validity index Vs number of clusters c for FCM and Sub Algorithms.

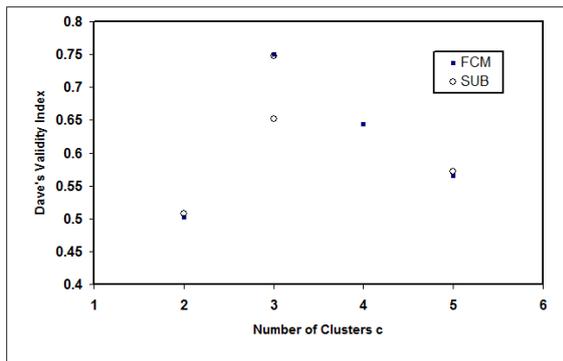


Figure 6: Dave's validity index Vs number of clusters c for FCM and Sub Algorithms.

Figure 7 shows Xi-Beni validity indices variations with number of clusters c for FCM and Subtractive algorithms. Same parameters mentioned previously are used. Although for Subtractive algorithm the same number of cluster are generated for the radii 0.6 and 0.3, the validity measure goes to its optimal value for radii= 0.3.

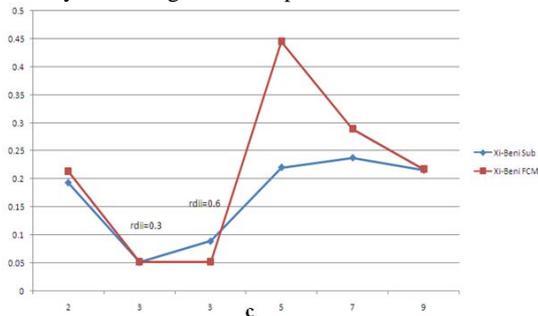


Figure 7: Xi-Beni validity indices variations with number of clusters c for FCM and Subtractive algorithms.

4.2. Non-linear system modeling:

In this section, several examples are presented in deploying the clustering algorithms to obtain an accurate model of highly nonlinear functions. The models are generated from the two clustering methods. For subtractive clustering the radii parameters are tuned. This automatically generates the number of clusters. The number of clusters is input to the FCM algorithm as starting point only. The generated models is trained using hybrid learning algorithm to identify the membership function parameters of single-output, (Sugeno type fuzzy inference systems FIS). A combination of least- squares and back propagation gradient descent methods are used

for training FIS membership function parameters to model the given model. The training routine is used as a tool to shape out the resulting model and to optimize the rule.

The results obtained from using FCM and Subtractive clustering for example 1 are compared against two recent studied conducted by Alta et al. [3] and Moaqt [4]. Alta et al.[3] used GA to optimize the clustering parameters. Moaqt [4] used 2 nd orders Sugeno consequent system to optimize the resulting model from subtractive clustering algorithm.

Example 1: Modeling the nonlinear function $y = \frac{\sin x}{x}$

Figure 8 shows a plot of the nonlinear function $\frac{\sin x}{x}$ over two cycles and the optimal resulting generated

model from Subtractive and FCM clustering. Generated models are obtained by using 100 clusters for FCM and 0.01 is set for radii parameter for Subtractive algorithm. It can be seen from Figure 8 that both algorithms accurately model the original functions. Table 5 lists the validity measures for subtractive algorithm. Changing the Radii parameter has effectively affects the performance of the algorithm. Generally, lowering the radii parameter has increased the performance measures indices, except for the case of 5 clusters. This is because of the nature of the data considered in this example. Table 6 lists validity measures for the FCM model. Similar behavior is observed for FCM method, that is, as the number of clusters increases, the performance improves. The ** superscript indicates that the index values are undefined but it is approaching 1:

$$\lim_{c \rightarrow n} \text{Dave's index} = 1$$

and

$$\lim_{c \rightarrow n} \text{Xi-Beni index} = 1$$

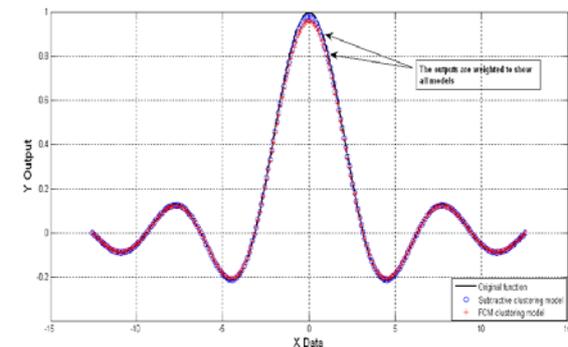


Figure 8: FCM Vs Subtractive for modeling $y=\sin(x)/x$, 100 clusters.

Table 5: Validity measure for subtractive clustering.

Radii value	Generated number of clusters	Dave's index	Bezdek index	Dispersion index(XB)
1.1	2	0.6737	0.8369	0.0601
0.9	3	0.6684	0.7790	0.1083
0.5	5	0.6463	0.7171	0.1850
0.02	100	0.8662	0.8676	0.1054
0.01	126	1	1	$1.6077 * 10^{-29}$

Table 6: Validity measure for FCM clustering.

Number of assigned clusters	Dave's index	Bezdek index	Dispersion index(XB)
2	0.7021	0.8511	0.0639
3	0.7063	0.8042	0.0571
5	0.6912	0.7530	0.0596
100	0.8570	0.8487	0.0741
126	NA ~ 1**	NA ~ 1**	NaN~0**

The Modeling errors are measured through LSE. Table 7 summaries the modelling error from each method and those obtained by Moaqt and Ramini. The LSE obtained by FCM followed by FIS training is 4.8286×10^{-9} . On the other hand, the same function is modeled using the subtractive clustering algorithm and the outcome error is 2.2694×10^{-25} . The default parameters with very minimal changes to the radii are used for the subtractive clustering algorithm. The above example was solved by Alata et al. [3] and the best value through deploying Genetic algorithm optimization for both the subtractive parameter and for the FCM fuzzifier is 1.28×10^{-10} . The same problem was solved by H. Moaqt [4]. Moaqt [4] used subtractive clustering to define the basic model and then a 2nd order ANFIS to optimize the rule. The obtained modeling error is 11.3486×10^{-5} .

Table 7: Modeling results for Subtractive Vs FCM.

Used Algorithm	Number of generated clusters	Output error With training	Output error Without training	Alta (2 nd order Sugeno to optimize the rule)	Ramini (Using GA optimization)
FCM	100(user input)	4.8286×10^{-9}	27.0883	----	1.28×10^{-10}
Subtractive	100	----	2.2694×10^{-25}	11.349×10^{-5}	---

Example 2: Single-input single-output function

Another nonlinear function $y = \frac{\sin x}{x^3}$ is modeled by

using subtractive and FCM clustering. Figure 9 shows the original function and output models generated from FCM and Subtractive clustering using 126 clusters. Table 8 lists validity indices (Dave's index, Bezdek index, and Dispersion index) versus radii parameter for subtractive clustering. Based on the validity measure, table 8 shows that using radii parameter equals to 1.1 that generates two clusters is better than that associated with using 3, 5, or 60 clusters. Optimal representation is obtained when setting radii parameter equals to 0.01 (generated 126 clusters).

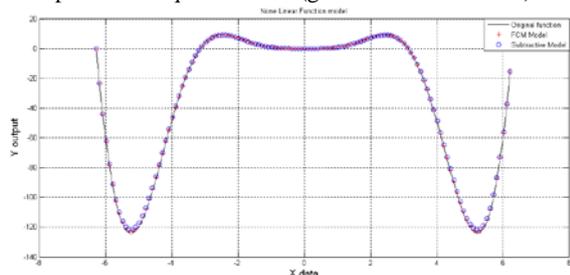


Figure 9: Single input function models, FCM Vs Subtractive, 126 clusters.

Table 8: Validity measure Vs cluster parameter for subtractive clustering.

Radii parameter	Generated number of clusters	Dave's index	Bezdek index	Dispersion index(XB)
1.1	2	0.8436	0.9218	0.0231
0.9	3	0.7100	0.8067	1.1157
0.4	5	0.7010	0.7608	0.6105
0.05	60	0.7266	0.7311	1.1509
0.01	126	1	1	2.1658×10^{-5}

The same function is modeled using FCM clustering. Table 9 lists the validity measures (Dave's, Bezdek and Dispersion index) versus number of clusters. Similar behavior to Subtractive clustering is observed when using FCM, such that optimal values associated with using 126 clusters. Table 10 lists the resulting modeling error for both algorithms.

Table 9: Validity measure Vs cluster parameter for FCM clustering.

Number of assigned clusters	Dave's index	Bezdek index	Dispersion index(XB)
2	0.8503	0.9251	0.0258
3	0.8363	0.8908	0.0300
5	0.7883	0.8307	0.0513
60	0.7555	0.7416	2.7595
110	0.9389	0.9271	0.1953
126	Na~1	Na~1	Na~0

Table 10: Modeling results for subtractive Vs FCM.

Used Algorithm	Number of generated clusters	Output LSE error With training	Output LSE error Without training
FCM	110(user input)	5.6119×10^{-5}	2.82×10^{-25}
Subtractive	110	----	1.4155×10^{-25}

4.3. Modeling experimental data:

A set of data that represent a real system with an elliptic shaped model obtained from real data experiments [14] is modeled. Figure 10 shows the real data points and the generated models obtained from FCM and subtractive clustering using 38 clusters.

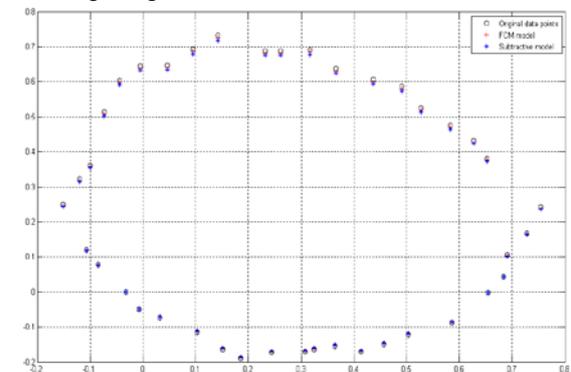


Figure 10: Elliptic shaped function models, FCM Vs Subtractive.

Table 11 summaries the validity measure variations with radii parameter for subtractive clustering. For this set of data, optimal value of validity measure obtained when radii parameter is 0.02. Generally, the smaller the radii parameter is the better the modeling results are. From the table 11, the model generated by using radii =1.1 that generated 2 clusters performed better than that obtained from using three clusters. The reason behind this is that the data is distributed into two semi-spheres (one cluster for each semi-sphere). However, further decrease in the value of Radii improves the resulting model.

Table 11: Validity measure Vs cluster parameter for subtractive clustering.

Radii parameter	Generated number of clusters	Dave's index	Bezdek index	Dispersion index(XB)
1.1	2	0.5725	0.7862	0.1059
0.9	3	0.5602	0.7068	0.1295
0.5	6	0.6229	0.6858	0.0806
0.05	39	0.9898	0.9895	0.0077
0.02	40	1	1	4.9851*10 ⁻²⁵

Table 12 summaries the validity measure variation with the number of clusters using FCM algorithm. Similar behavior to the subtractive clustering is observed.

Table 12: Validity measure Vs cluster parameter for FCM clustering.

Number of assigned clusters	Dave's index	Bezdek index	Dispersion index(XB)
2	0.5165	0.7582	0.1715
3	0.5486	0.6991	0.1151
20	0.7264	0.7287	0.0895
30	0.8431	0.8395	0.1017
39	0.9920	0.9813	0.0791
40	Na~1	Na~1	Na~0

Table 13 compares the LSE error with and without training between FCM and subtractive clustering. It can be seen that FCM algorithm without training routine does not model the data. However, using FCM with training algorithm yields acceptable results. On the other hand, the LSE error resulted from using subtractive algorithm is on the order of 10⁻²⁹. Double superscript * indicates that the value listed here is an average of 20 different runs.

Table 13: Modeling results for subtractive Vs FCM.

Used Algorithm	Number of generated clusters	Output LSE error With training	Output LSE error Without training
FCM	39 (user input)	1.9906*10 ^{-3**}	4.2625
Subtractive	39	--	2.1244*10 ⁻²⁹

5. Conclusion

In this work a comparison study has been performed between FCM clustering algorithm and subtractive clustering algorithm according to their capabilities to model a set of non-linear systems and experimental data. A concise literature review is provided. The basic parameters that control each algorithm are presented as well. The clustering results from each algorithm are assessed using Dave's, Bezdek, and Xi-Beni validity measurement indices.

For the majority of the system discussed earlier; increasing the number of generated cluster yields an improvement in the validity index value. The optimal modelling results are obtained when the validity indices are on their optimal values. Also, the models generated from subtractive clustering usually are more accurate than those generated using FCM algorithm. A training algorithm is needed to accurately generate models using FCM. However, subtractive clustering does not need training algorithm. FCM has inconsistency problem such that, different running of the FCM yield different results as the algorithm will choose an arbitrary v matrix each time. On the other hand, subtractive algorithm produce consist results.

The best modelling results obtained for example 1 when setting the cluster number equals to 100 for the FCM and the radii to 0.01. Using these parameter values yields an optimal validity results. The LSE was 4.8286*10⁻⁹ for the FCM and 2.2694*10⁻²⁵ using the subtractive clustering.

For the data in example 2 optimal modelling obtained when using 126 clusters for the FCM and 0.01 radii input for the Subtractive. The LSE are 5.6119*10⁻⁵ and 7.0789*10⁻²¹ for the FCM and subtractive respectively. As for the data in example 3 going to 110 clusters for the FCM and 0.03 radii yields the best clustering results that is, 5.6119*10⁻⁵ for the FCM and 1.4155*10⁻²⁵ for the subtractive clustering. As for experimental data which for an elliptic shape, optimal models obtained when using 39 clusters for FCM. For subtractive clustering, optimal values obtained by setting radii=0.02. The LSE is 1.9906*10⁻³ and 2.4399*10⁻²⁸ for the FCM and subtractive respectively.

Optimising the resulting model from FCM using a Sugeno training routine has produced a significant improvement in the system modelling error. For data that lacks of natural condensation, the optimal number of clusters tends to be just below the same number of data points in the raw data set. Tuning the radii parameter for the subtractive clustering was an efficient way to control the generated cluster number and consequently the validity index value and finally modelling LSE.

References

- [1] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2nd edition, John Wiley, New York; 2001.
- [2] Hussein Al-Wedyan, *Control of whirling vibrations in BTA deep hole boring process using fuzzy logic modeling and active suppression technique*, 2004.
- [3] M. Alata, M. Molhim, A. Ramini, "Optimizing of fuzzy C-means clustering algorithm using GA". *Proceeding of world academy of science, Engineering and technology*, Vol. 29, 2008, 1307-6884.
- [4] H. Moaqt, "Adaptive neuro-fuzzy inference system with second order Sugeno consequent". *Jordan University of science and technology, Master Theses*, 2006.
- [5] J. C. Dunn, "A fuzzy relative of the Isodata process and its use in detecting compact well-separated clusters". *J. Cybern.*, 1973, 32-57.
- [6] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function". New York, Plenum Press, NewYork, 1981.
- [7] R. Yager, D. Filev, "Generation of Fuzzy Rules by Mountain Clustering". *Journal of Intelligent & Fuzzy Systems*, Vol. 2(3), 1994, 209-219.
- [8] S.L. Chiu, "Fuzzy model identification based on cluster estimation". *J. Intell. Fuzzy Systems*, Vol. 2, 1994, 267-278.
- [9] R.N. Dave, "Validating fuzzy partition obtained through c-shells clustering". *Pattern Recognition Lett.*, Vol. 17, 1996, 613-623.
- [10] Gath, A. Geva, "Unsupervised optimal fuzzy clustering". *IEEE Transactions on Pattern Analysis and, Machine Intelligence*, Vol. 7, 1989, 773-781.
- [11] N.R. Pal, and J.C. Bezdek, "On cluster validity for the fuzzy C-means model". *IEEE Trans. on Fuzzy Systems*, Vol. 3, 1995, 370-379.
- [12] W. Wang, Y. Zhang, "On fuzzy cluster validity indices". *Fuzzy Sets and Systems*, Vol. 158, 2007, 2095-2117.
- [13] Xie X., G. Beni, "Validity measure for fuzzy clustering". *IEEE Trans. PAMI*, Vol. 3(8), 1991, 841-846.

Machining Characteristics of Multiwall-CNT Reinforced Al/Al-Si Composites using Recurrence Quantification Analysis

K. V. Gangadharan^{*a}, K. S. Umashankar^a, Ravish^a, V. Desai^a

^aDepartment of Mechanical Engineering National Institute of Technology Karnataka, Surathkal, 576102, India

Abstract

Aluminium (Al)/Aluminium alloy composites are emerging as very promising materials, especially in the fields of aerospace and automotive for their various attractive and technically demanding properties. Discontinuously reinforced aluminium metal matrix composites with reinforcements as nanoparticles of ceramics in general and carbon nanotubes in particular have emerged as the forerunner for a variety of general and special engineering and structural applications. In many of the fields where these materials find applications, machining is invariably required for getting correct geometries, dimensions and surface finish of the components. Hence, establishing the machining characteristics of these materials in terms of the deterministic nature of dynamic signals such as cutting force signals and vibration signals is very important and sought after. Machining process has been understood to be nonlinear and chaotic in nature. In this paper a relatively new technique called Recurrence Plots (RP) and Recurrence Quantification Analysis (RQA), a tool to analyse nonlinear and chaotic systems, is used to study the machining characteristics of cast and powder metallurgy Al and Al-Si alloys (LM6 and LM25), CNT reinforced Al/Al-Si composites produced by powder metallurgy route. Cutting force signals were sensed, acquired and analysed using RQA technique. Determinism (DET), which is one of the variables of RQA, indicates the determinism present in a signal. The values of DET were used to compare the machining characteristics. For all the three materials the deterministic nature of the cutting force signal was highest when reinforced with 0.5 weight percentage CNT, followed by respective base alloys produced by powder metallurgy method and casting route.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: Aluminium; Carbon Nanotube; Determinism; Recurrence Plot; Recurrence Quantification Analysis

1. Introduction

Most of the popular engineering materials are highlighted because of their versatile structural properties. Among the metal matrix composites (MMC) synthesized, aluminium metal matrix composites in general and discontinuously reinforced aluminium metal matrix composites in particular, have emerged as the forerunner for a variety of general and special engineering and structural applications. This trend has been attributed to their superior specific strength and specific stiffness, high temperature capability, lower coefficient of thermal expansion, better wear resistance, improved dimensional stability, and amenability to conventional metal forming techniques [1-4].

Stiffer materials with high damping property are actively sought for dynamic mechanical systems such as in spacecrafts, semiconductor equipments and robotics [5]. Al-Si alloys provide good mechanical properties as well as high damping capacity, particularly for powder metallurgy parts. Discovery of Carbon Nanotubes (CNT) has provided further wings for such applications.

Carbon Nanotube is a nanosize anisotropic material belonging to Fullerene family identified in 1991 by Iijima and has been proved to be very attractive in terms of properties like weight, strength, modulus values and

dynamic behaviour. Carbon Nanotubes have been given a great deal of attention because of their unique properties which are leading to many promising applications [6-8]. The mechanical properties reported on carbon nanotubes can be used for developing an entire new class of composite materials. Although most of the research works are focused on the development of nanotube based polymer composites [9] and metal matrix composites using ceramic as reinforcement material [10, 11], attempts have also been made to produce composites using metals such as aluminium [6], lead [12], silver [13], copper [14], magnesium [15] as matrix materials with nanotubes as reinforcement.

Components manufactured by different processes invariably require machining, at least to some extent. Establishing machinability of these newer materials helps the manufacturer to choose machining conditions and component performance. Present work deals with the machining characteristic studies on cast and powder metallurgy Al and Al-Si alloys, CNT reinforced Al/Al-Si composites produced by powder metallurgy route. An attempt has been made to use a relatively new technique called Recurrence Plots (RP) and Recurrence Quantification Analysis (RQA) to study the machining characteristics of these materials.

* Corresponding author. e-mail: : kvganga@nitk.ac.in

2. Recurrence Plots

Recurrence Plots (RP) were first described by Eckmann et al. in 1987 [16]. RP is a technique by which we can qualitatively assess a time series signal embedded in phase space. A recurrence plot can be represented as:

$$R_{i,j} = \theta(\epsilon_i - \|x_i - x_j\|) \quad i, j = 1, 2, \dots, N \quad (2.1)$$

where x_i stands for the point in phase space at which the system is situated at time i , and ϵ is a predefined threshold for whose selection there are plenty of criteria. $\|\cdot\|$ is the norm used to calculate the distances between points in phase space. θ is the heaviside function. The matrix corresponding to $R_{i,j}$ consists of values of only 1 and 0. RP will ultimately be a black and white plot with time on both the axes. A black point in a RP means that the system returns to an ϵ -neighbourhood of the corresponding point in phase space [17, 18]. This recurrence gives the name to the method. Figures 1 to 5 demonstrate obtaining RP from time series data.

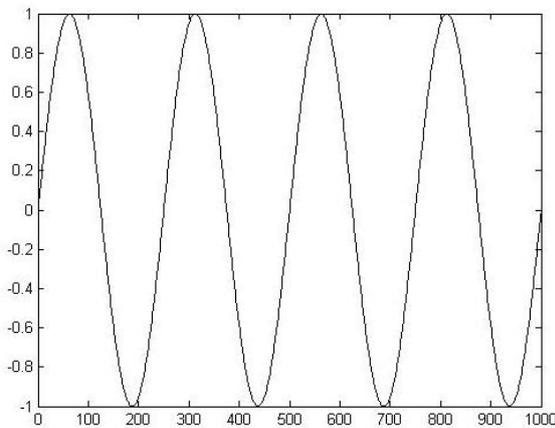


Figure 1: Time domain plot of sine wave (frequency 4 Hz).

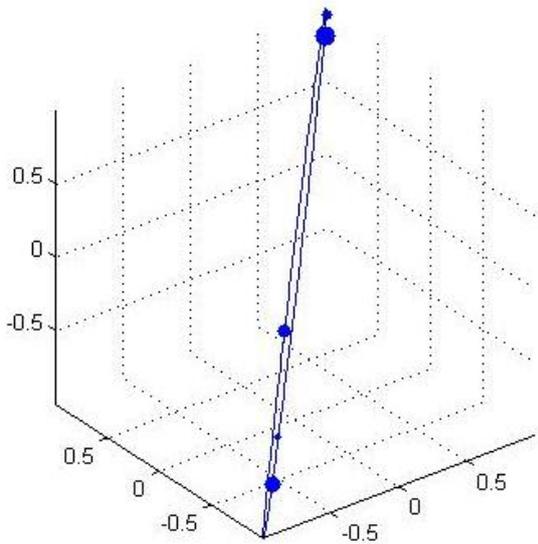


Figure 2: Phase Space plot of sine wave

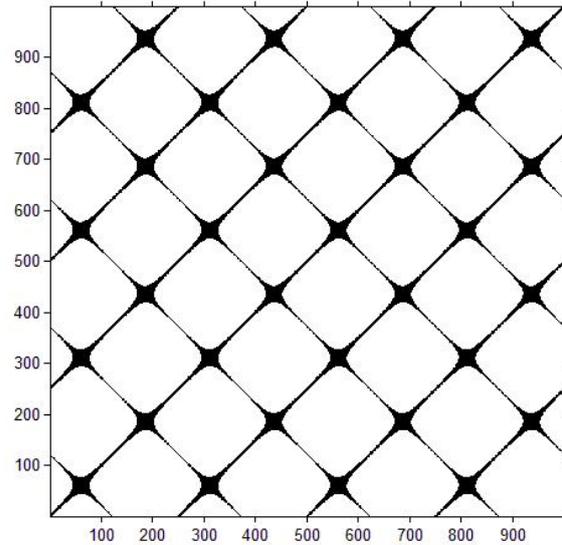


Figure 3: Recurrence Plot of sine wave. $m = 2, \tau = 2, \epsilon = 0.091$ (10 % of mean phase space diameter), Maximum norm.

While figures 1 to 3 have sine wave as the underlying time series, random noise with a standard deviation of 1 is the underlying time series for figures 4 and 5. One can notice easily the difference in RPs of the two cases. While it is the characteristic of deterministic signals to show diagonal lines in RP, homogeneous RP with scattered points will be exhibited by random signals [17, 18, 20]. These are just only two typical type of plots to pick from a large pool of Recurrence Plots. RPs require quite a few criteria and input parameters to be set carefully. The time delay (τ) and embedding dimension (m) required for state space embedding are obtained correspondingly by Mutual Information method [17] and False Nearest Neighbour algorithm [17, 21]. Threshold for the RP in figure 3 was set to be 10 % of the mean phase space diameter [17] whereas that for the RP in figure 5 was 10 % of the maximum phase space diameter, only to get more number of points in the RP. There are numerous criteria to select a proper threshold for a given application [17, 18, 22, 23] even as there lies no fixed single method to select an appropriate threshold.

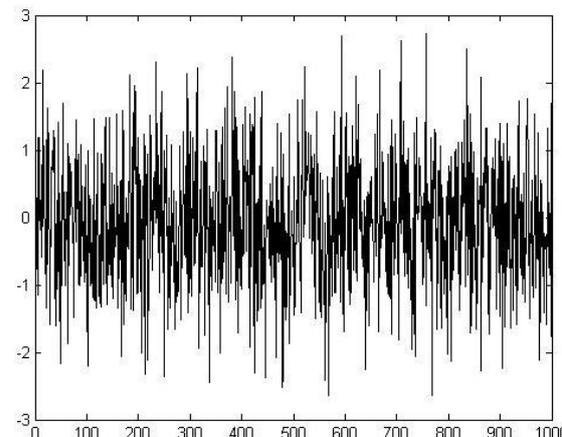


Figure 4: Random noise signal (standard deviation of 1).

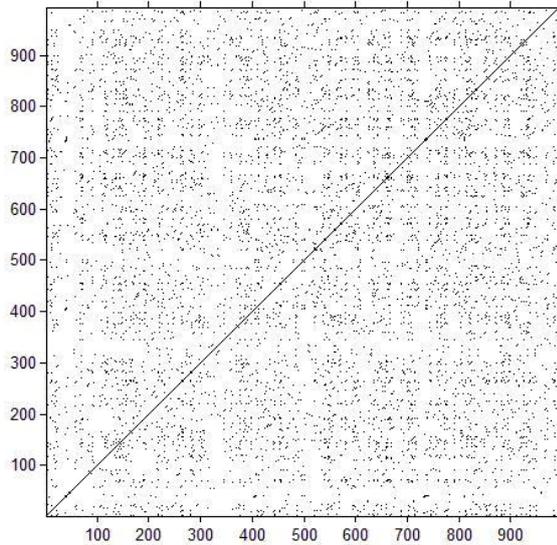


Figure 5: RP of random noise. $m = 3$, $\tau = 3$, $\varepsilon = 0.63$ (10 % of maximum phase space diameter), Maximum norm.

It is possible to look at RPs and qualitatively assess the dynamics of the underlying system. However, it demands practice and is not easy always. There are some general guidelines present to interpret the behaviour of a system by observing its corresponding Recurrence Plot [17, 19, 23]. In general, deterministic and regular signals tend to form diagonal line structures whereas random signals form scattered points distributed throughout the RP. Laminar systems (states are changing slowly with time or they are stationary) show vertical line structures in their corresponding RPs.

3. Recurrence Quantification Analysis

It is always difficult to judge the status of a system just by observing the corresponding RP. Some means of quantification of RPs would make understanding the behaviour of the system easy. Charles L. Webber et al. came up with a technique called Recurrence Quantification Analysis in 1992 which was based on quantifying the diagonal line structures present in RPs. In 2002, Norbert Marwan et al. successfully added the quantifications based on vertical line structures. Some of the important variables in RQA are listed below [17].

- Recurrence Rate (RR): Percentage of recurrence points in RP.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j} \quad (3.1)$$

- Determinism (DET): Percentage of recurrence points which form diagonal lines.

$$DET = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{i,j} R_{i,j}} \quad (3.2)$$

- Averaged diagonal line length (L): Average length of diagonal lines.

$$L = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_{min}}^N P(l)} \quad (3.3)$$

- Entropy (ENTR): Shannon entropy of the probability distribution of diagonal line lengths.

$$ENTR = -\sum_{l=l_{min}}^N p(l) \ln p(l) \quad (3.4)$$

- Laminarity (LAM): Percentage of recurrence points which form vertical lines.

$$LAM = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=1}^N vP(v)} \quad (3.5)$$

- Trapping Time (TT): Average length of vertical lines.

$$TT = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=v_{min}}^N P(v)} \quad (3.6)$$

Table 1 shows the above discussed RQA variables for a sine wave with a frequency of 4 Hz and for a random noise signal. The threshold was set to be equal to 10 % of the mean phase space diameter for both the signals and Maximum norm was used for the calculations. A DET value of 98.66 % for sine wave shows that the signal is deterministic. Since deterministic signals form diagonal lines, all the variables quantifying the diagonal structures (DET, L and ENTR) show higher values for sine wave. Since sine wave is not very agile in behaviour, the variables LAM and TT which quantify vertical structures attain higher values.

Table 1: RQA variables' values for sine wave and random noise (Maximum norm)

Variable/parameter	Sine wave (4 Hz)	Random noise
τ	2	3
m	2	3
ε	0.09	0.14
RR (%)	8.39	0.04
DET (%)	98.66	1.05
L	12.34	2
ENTR	2.56	0
LAM (%)	99.98	0
TT	12.01	0

RP of the noise signal is very scattered as can be seen from figure 5, thus contrasting in nature the RP of a deterministic signal such as sine series which is shown in figure 3 where one gets structured diagonal lines. An interesting comparison can be made between the RQA

variables of the sine series and that of the random noise, drawing conclusions about the system dynamics. Very low value of DET for random noise (1.05 % in table 1) confirms that the system has low determinism. This in turn will imply that the plot has no diagonal lines of considerable length. Hence, one gets a lower value of L also in this case. As ENTR is dependent on the probability distribution of diagonal line lengths, it will be lower for noisy signals. In case of random noise, the system is very agile. Hence, the system will hardly be laminar. LAM shows very low values because of this very reason. For the very same reason, TT will also be low for random noise. These variables give a feel of systems' dynamic behaviour. All these behaviours are exactly opposite to that of a deterministic system such as sine wave.

Visual Recurrence Analysis (VRA), CRP Toolbox for Matlab, Dataplore, TISEAN and Bios Analyzer are few of the codes and softwares available for Recurrence Plots and Recurrence Quantification Analysis. For all the RPs and RQA in this work, CRP Toolbox, developed as part of the dissertation work of and by Dr. Norbert Marwan, University of Potsdam, Germany, was used.

4. Work piece Preparation

Commonly available industrial grade cast Al, LM6 and LM25 were procured from the suppliers for preparing work specimen for machining. The powder metallurgy workpieces of Al, LM6 and LM25 and CNT reinforced variants of these materials were prepared in the laboratory using conventional powder metallurgy technique.

4.1. Preparation of Al, LM6 and LM25 by powder metallurgy:

LM6 is a eutectic mixture of Aluminium and Silicon with 11.8 % Si. LM25 is a hypoeutectic alloy with 7 % Si and 0.3 % Mg. Al, LM6 and LM25 powders of 75 micron size were procured from the suppliers and the powders were compacted according to ASTM B 925 03 to form solid billets. These billets were sintered in vacuum furnace at appropriate temperature and time suggested in literatures. The compaction and sintering procedure was verified by measuring the hardness and comparing it with standard values for powder metallurgy materials. The sintered billets were hot extruded at a temperature of 350°C to get the workpieces required for machining.

4.2. Preparation of 0.5 % (weight fraction) CNT reinforced workpieces:

Multiwalled CNT was procured from M/s Sigma Aldrich, Bangalore. Impurities such as graphitic particle, amorphous carbons or any other present in the raw CNT powder were removed by immersing them in concentrated Nitric acid, then filtered and washed with de-ionized water and dried at 120°. The Al/Al alloy powders were mixed with CNTs in ethanol solution. These mixed powders were dispersed with mechanical stirring assisted with ultrasonic shaker for 30 minutes. Finally, the mixed powders were dried at 120° in vacuum (less than 10⁻² Pa) and ball milled for 10 minutes in Retsch PM100 high energy ball mill. The mixed powders were compacted and sintered into billets and finally extruded into rods at 350°. The specification of the CNT is listed in Table 2.

Table 2: Specifications of the CNT used.

Parameter	Values
Outer diameter	1 to 15 nm
Length	0.1 to 10 μm
Density	2100 kg/m ³

5. Experimental Setup

Figure 6 shows the schematic of experimental setup for the present work. Machining was performed on the workpieces on a Panther 1530/1650 lathe. The machine has a highest attainable spindle speed of 1250 rpm. As indicated in the past research, tungsten carbide tools with better tool life than HSS tools in the machining of eutectic (LM6) and hypoeutectic (LM25) Al-Si alloys [24] has been used. Kennametal DNMG 150608 tool insert was chosen with standard tool holder PCLNR 2020 K12. Cutting forces were sensed using a Kistler 9257 B dynamometer on which the tool post was mounted. The dynamometer in turn was connected to Kisteler 9257 B charge amplifier. The force signals in three directions from the amplifier were acquired by National Instruments PCI-4472 eight channel data acquisition card at a sampling rate of 1000 samples per second with LabVIEW 8.0 being the software interface. Cutting force signals in the cutting direction (F_z) were then taken as the inputs for analysis with CRP toolbox. All the workpieces were of the same size of 8 mm diameter and 80 mm length. The following nine specimens were machined.

- Cast and P/M Al
- 0.5 wt. % CNT reinforced Al composite
- Cast and P/M LM6
- 0.5 wt. % CNT reinforced LM6 composite
- Cast and P/M LM25
- 0.5 wt. % CNT reinforced LM25 composite

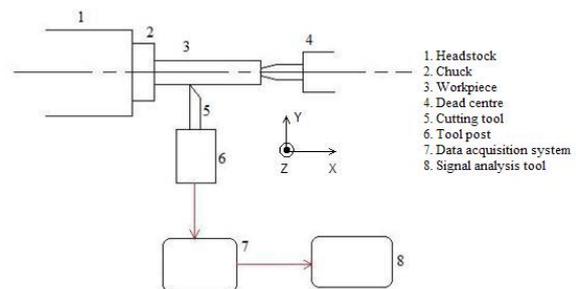


Figure 6: Scheme of experimental setup.

Cutting force signals during machining were sensed and acquired at a sampling rate of 1000 samples per second. For the present work, cutting force signals in the cutting direction (F_z) was considered for analysis. The force signals were the inputs for CRP Toolbox used with Matlab. The criterion to select the threshold was to keep RR at 1 %.

6. Results and Discussions

Turning operation for the materials listed in section 5.0 under the machining conditions of 1250 rpm speed , 0.0375 mm/rev feed and 0.5 mm DOC were carried out and cutting force signal was sensed and acquired to process by RQA technique. The combination of machining parameters was chosen for good surface finish at low MRR. From the RQA, the DET was calculated and used to assess machining characteristics of the materials in view of randomness and the expected value of the variable. The time domain signal and the corresponding RP for LM6 and its variants are given in figures 7, 8 and 9. In all the three plots, recurrence rate (RR) was kept at 10 % so that more recurrence points will be visible in the RPs and hence the line structures are clearly visible.

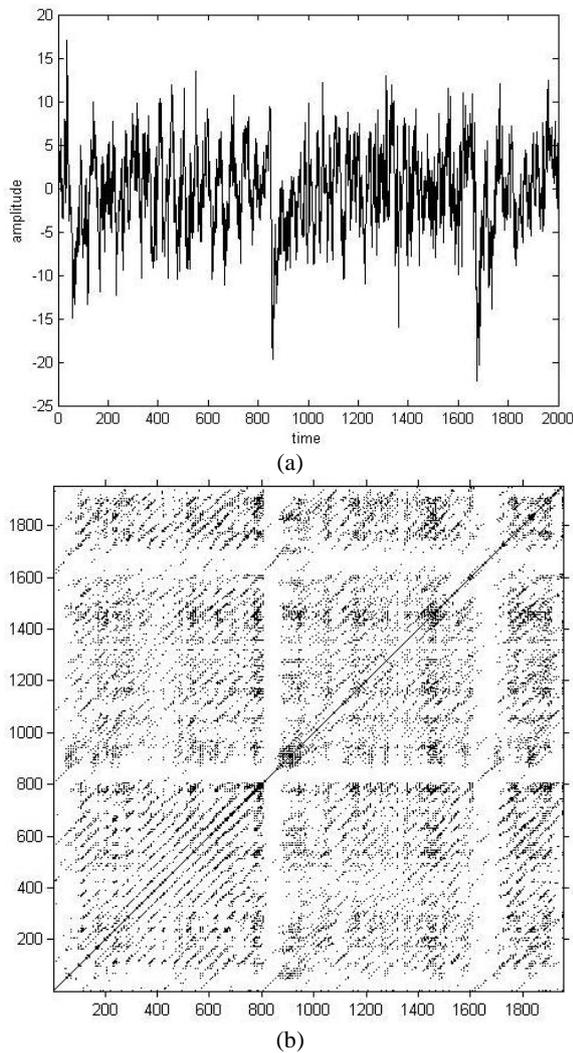


Figure 8: (a and b): Time domain cutting force signal and corresponding RP for P/M LM6.

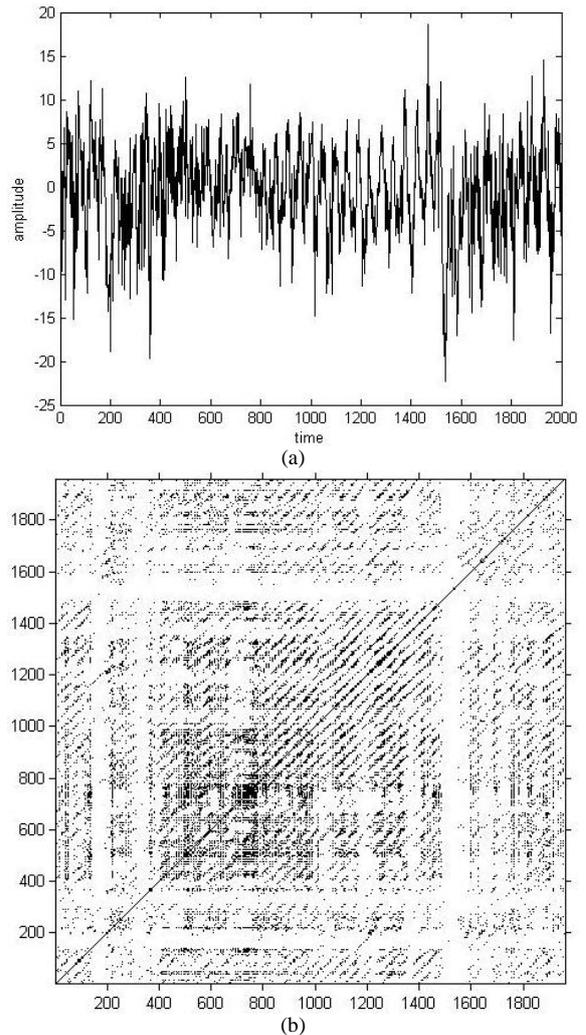


Figure 9: (a and b): Time domain cutting force signal and corresponding RP for 0.5 % CNT reinforced LM6.

From the RPs for LM6, the amount of diagonal structures are more in case of LM6 produced by powder metallurgy route compared to that produced by casting route. The diagonal structures further improves with reinforcement of 0.5 wt. % CNT in LM6. It indicates that the force signal is more deterministic in case of 0.5 wt. % CNT reinforced LM6 composite. This is a good indication to understand that machining characteristics of powder metallurgy material are better than the cast material. This may be attributed to controlled particle size and uniform distribution of Si as particles which are presumed to act as chip breakers. In addition to this the porosity factor in case of powder metallurgy materials may also aid in breaking the chips. The addition of CNT to LM6 primarily to improve its damping behaviour further improves the deterministic nature of the signal also. This is a good indication that during sintering no reaction has taken place in forming the hard phase Silicon Carbide at the interface and also the CNTs are uniformly dispersed. The CNTs also contribute in improving the deterministic nature of the signal by way of increasing the anti-friction property at the tool-material interface. To quantify the RP, the DET values were calculated and based on the fact that in RQA, dynamic systems giving higher values of DET are considered to be more regular in nature. The DET values for LM6 and its variants are given in Table 3.

Table 3: DET values for LM6 and its variants.

Material	DET (in per cent)
LM6- cast	6.34
LM6- powder metallurgy	17.44
LM6 with 0.5 % CNT	18.67

From Table 3, the value of DET is highest for 0.5 wt. % CNT reinforced LM6 composite produced by powder metallurgy route. This indicates that the machining characteristics of LM6 evaluated in terms of force signal (dynamic in nature) regularity is better for 0.5 wt. % CNT reinforced LM6 composite followed by LM6 produced by powder metallurgy route and LM6 produced by casting route. Similar trends were also observed for Al and LM25. The DET values for Al and its variants and LM25 and its variants are shown in Tables 4 and 5 respectively.

Table 4: DET values for Al and its variants.

Material	DET (in per cent)
Al- cast	11.56
Al- powder metallurgy	21.32
Al with 0.5 % CNT	22.87

Table 5: DET values for LM25 and its variants.

Material	DET (in per cent)
LM25- cast	6.21
LM25- powder metallurgy	11.73
LM25 with 0.5 % CNT	12.61

7. Conclusions

Machining experiments were conducted on Al, LM6 and LM25 produced by casting and powder metallurgy routes. Experiments were also conducted on 0.5 % CNT reinforced composites of the above materials through powder metallurgy route. The following conclusions are drawn from the experimental results. Machining operation is dynamic in nature and RQA technique can be used to assess the machining by determining the regular or random nature of the cutting force signal. For all the three materials, the DET value is highest for the 0.5 wt % CNT reinforced composites produced by powder metallurgy route. The uniformity of particle size, the uniform distribution of Si particles and the anti-friction property of CNT are the factors that makes the cutting force signal more deterministic. To an extent a favourable presence of porosity in terms of shape and size of the pores in powder metallurgy processing contribute towards forming smaller and discontinuous chips. This makes the cutting force signal more deterministic. LM6 having good mechanical properties can be further improvised in terms of its vibration damping ability and stability during machining by reinforcing it with 0.5 wt. % CNT of smaller aspect ratio and processing through P/M route. This material offers excellent scope for manufacturing components for aeronautical, automotive and structural applications.

8. Acknowledgments

Free license for CRP Toolbox was given by Dr. Norbert Marwan, University of Potsdam, Germany under GNU. Dr. Marco Thiel, Nonlinear Dynamics Group, University of Potsdam, Germany shared his knowledge, research works and publications with us. The experimental facilities were provided by Technical Education Quality Improvement Programme (TEQIP), National Institute of Technology Karnataka, Surathkal.

References

- [1] K.G. Satyanarayana, R.M. Pillai, and B.C. Pai, Aluminium Cast Metal Matrix Composites, Handbook of Ceramics and Composites, Vol 1, Synthesis and Properties, N.P. Cheremisinoff, Ed., Marcel Dekker Inc., 1990, p 555-599.
- [2] P.K. Rohatgi, Future Directions in Solidification of Metal Matrix Composites, Key Engineering Materials, G.M. Newaz et al., Ed., Trans. Tech., Switzerland, Vol 104-107, 1995, p 293-312.
- [3] D.J. Lloyd, Particle Reinforced Aluminium and Magnesium Matrix Composites, Int. Mater. Rev., Vol 39, 1994, p 1-23.
- [4] J.E. Allison and G.S. Cole, Metal Matrix Composites in the Automotive Industries, J. Met., Vol 45 (No. 4), 1993, p 10-15.
- [5] B.J. Lazan, Damping of Materials and Members in Structural Mechanics, Pergamon Press, New York, 1968.
- [6] Erik T. Thostenson, Zhifeng Ren, Tsu-Wei Chou, Advances in the science and technology of carbon nanotubes and their composites: a review, Composites Science and Technology 61 (2001) 1899–1912.
- [7] Rouff R S, Qian D, Lie W K, Mechanical properties of carbon nanotubes: theoretical predictions and experimental measurements, Physics 2003;4:993-1008.
- [8] Kin Tak Lau a, Mei Lu, David Hui Coiled carbon nanotubes: Synthesis and their potential applications in advanced composite structures Composites: Part B 37 (2006) 437–448.
- [9] Kin-tak Lau a, Chong Gu, David Hui, A critical review on nanotube and nanotube/nanoclay related polymer composite materials, Composites: Part B 37 (2006) 425–436.
- [10] William A Curtin, Brian W Sheldon, CNT-reinforced ceramics & metals, Materials today, (2004) 44-49.
- [11] Seung I. Cha, Kyung T. Kim, Kyong H. Lee, Chan B. Mo, Soon H. Hong, Strengthening and toughening of carbon nanotube reinforced alumina nanocomposite fabricated by molecular level mixing process, Scripta Materialia 53 (2005) 793–797.
- [12] S.M.L. Nai, J. Wei, M. Gupta, Improving the performance of lead-free solder reinforced with multi-walled carbon nanotubes, Materials Science and Engineering: A Volume 423, Issues 1-2, 2006, Pages 166-169.
- [13] Yi Feng, Hai Long Yuan and Min Zhang, Fabrication and properties of silver-matrix composites reinforced by carbon nanotubes, Materials Characterization, Volume 55, Issue 3, September 2005, Pages 211-218.
- [14] Kyung Tae Kim, Seung Il Cha, Seong Hyeon Hong and Soon Hyung Hong, Microstructures and tensile behaviour of carbon nanotube reinforced Cu matrix nanocomposites, Materials Science and Engineering: A Volume 430, Issues 1-2, 25 August 2006, Pages 27-33.
- [15] A. Peigney, S. Rul, F. Lefèvre-Schlick and C. Laurent, Densification during hot-pressing of carbon nanotube-metal-magnesium aluminate spinel nanocomposites, Journal of the European Ceramic Society Volume 27, Issue 5, 2007, Pages 2183-2193
- [16] Eckmann, J. P., Kamphorst, S. O., Ruelle, D., Recurrence plots of Dynamical Systems. Europhysics Letters 5, 1987, 973-977.
- [17] Norbert Marwan, Encounters With Neighbours; Current developments of concepts based on recurrence plots and

- their applications. Ph.D. thesis, University of Potsdam, Institute for Physics, 2003.
- [18] Norbert Marwan, M. Carmen Romano, Marco Thiel, Jürgen Kurths, Recurrence plots for the analysis of complex systems. *Physics Reports* 438 (2007) 237 – 329.
- [19] <http://www.recurrence-plot.tk>
- [20] Fabretti, A., Ausloos, M., Recurrence plot and recurrence quantification analysis techniques for detecting a critical regime. Examples from financial market indices.
- [21] Carl Rhodes, Manfred Morari, The false nearest neighbors algorithm: an overview. *Computers chem Engng*, Vol. 21, 1997.
- [22] Marco Thiel, M. Carmen Romano, Jürgen Kurths, Riccardo Meucci, Enrico Allaria, F. Tito Arcucci, Influence of observational noise on the recurrence quantification analysis. *Physica D* 171 (2002) 138–152.
- [23] Norbert Marwan, Jürgen Kurths, Line Structures in Recurrence Plots. Nonlinear Dynamics Group, Institute of Physics, University of Potsdam, Potsdam 14415, Germany.
- [24] P. Narahari, B.C. Pai, and R.M. Pillai, Some Aspects of Machining Cast Al-SiCp Composites with Conventional High Speed Steel and Tungsten Carbide Tools, *Journal of Materials Engineering and Performance*, Volume 8(5), October 1999 538-542.

Modeling of the MEMS Reactive Ion Etching Process Using Neural Networks

M. Ashhab^{*a}, N. Talat^b

^aDepartment of Mechanical Engineering, The Hashemite University, Zarqa 13115, Jordan

^bMechanical Engineering Department, University of Jordan, Amman 11942, Jordan

Abstract

Reactive ion etch (RIE) is commonly used in microelectromechanical systems (MEMS) fabrication as plasma etching method, where ions react with wafer surface substrate in plasma environment. Due to the importance of RIE in the MEMS field, two prediction models are established to predict the wafer status in reactive ion etching process: back-propagation neural network (BPNN) and principle component analysis BPNN (PCABPNN). These models have the potential to reduce the overall cost of ownership of MEMS equipment by increasing the wafer yield, and not depend upon monitoring wafers or expensive metrology rather it will enable inexpensive real-time wafer-to-wafer control applications in RIE. The artificial neural net (ANN) is trained with historical available input-output process data. Once trained, the ANN forecasts the process output rapidly if given the input values.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: MEMS; Reactive ion etching; Modelling; Neural networks

1. Introduction

Micro-Electro-Mechanical Systems (MEMS) is an emerging technology, which uses the tools and techniques that were developed to build microscopic machines, sensors, actuators, and electronics. MEMS enable expanding the space of possible designs and applications of nearly every product category on a common silicon substrate through microfabrication technology. MEMS devices are manufactured using batch fabrication techniques similar to those used for integrated circuits. The fabrication of modern semiconductor products requires thousands of processing steps. A key element in achieving high yields during semiconductor fabrication is to minimize the amount of defected wafers. Therefore, detecting the defected wafers is a very important issue.

Neural networks have seen an explosion of interest over the last few decades. They have been utilized to study the possibility of designing control techniques to significantly improve the performance of the RIE process [1]. Optical emission spectroscopy data were used to construct neural network models of plasma etch process in [2]. A virtual metrology system for MEMS is proposed in [3] that fulfills real-time quality measurement of each wafer and detects the performance degradation of the corresponding machines from the information of manufacturing processes. Neural networks are being successfully applied in microfabrication inspection systems [4-8]. The major reason for adopting neural networks is because neural networks have potential capability in modelling and control of non-linear systems.

Moreover neural networks have the ability of learning arbitrary nonlinear mappings between noisy sets of input and output data. Back-propagation neural network (BPNN) is currently the most popular learning rule used in supervised learning, which is also known as feed forward neural network and multilayer perceptron (MLP).

Back propagation is a very powerful tool with application to solve the problems of prediction, optimization, control, and diagnosis in the MEMS manufacturing processes [9-12]. Most of the literature adopt BPNN because it has the advantages of an easier-comprehended theory, faster recalling speed and higher learning accuracy. However, the determination of the structure architecture and the parameters under this network is difficult.

Many researchers have studied pattern classification by using BPNN for the automatic inspection system in the MEMS industry [13-15]. Zoroofi et al. [13] used curve recognition to detect the contamination on a wafer surface during semiconductor production. Three conventional classification models: a back-propagation technique, a minimum distance algorithm and a maximum likelihood classifier, were used and the performance of these three models was compared. The results showed that the back-propagation classifier has a better classification performance. Su et al. [14] proposed a neural-network approach for semiconductor wafer post-sawing inspection. BPNN, radial basis function network (RBFN), and learning vector quantization (LVQ) were employed in the inspection models. The inspection results showed that both BPNN and LVQ have excellent prediction results with 100% accuracy. Chen et al. [15] used BPNN in the etch semiconductor process to identify and classify endpoint curves. By real-time monitoring of changes in the endpoint

* Corresponding author. e-mail: sami@hu.edu.jo

curve, the abnormalities of products can be detected immediately. The system can reduce the uncertainty in the process curve classification and provide machine shut-down suggestion immediately when necessary. In this respect, back propagation neural network is utilized to identify the wafer status during reactive ion etching, RIE, which is an important MEMS fabrication process.

2. Reactive Ion Etching

Reactive ion etching (RIE) is a very difficult process to control, since the physical mechanism of this process is not well understood. Consequently wafer defect occurs in RIE when there is a sudden change in the etching behavior. This change can happen due to operator errors or machine errors, such as gas leak, power fault, and pressure fault. The main defect in oxide RIE is un-open etch. This defect costs 10%-20% yield loss in a factory. Un-open etch signify the inadequate etching space in the wafer surface as shown in Figure 1.

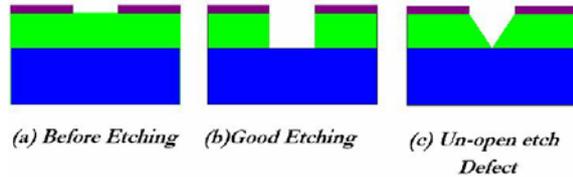


Figure 1: Wafer surface shape before and after etching process.

A silicon dioxide film etching without any errors is a complex task. One of the significant detractive defect in this process is un-opened etch. Predicting wafer status in RIE is important to enhance yield, quality, and efficiency. Towards this end, thorough analysis has been done to get improved prediction models with the purpose of providing valuable benefits. One of the attractive prediction models is artificial neural networks. Data preparation for neural networks has been discussed and applied in different ways.

3. Case Study: Oxide RIE

The advanced reactive ion etching equipment used in the factory is 2300 Exelan Flex made up of two chambers as shown in Fig. 2. The main etching chamber is configured inside the vacuum RIE chamber for optimal efficiency, where these two chambers are separated by quartz confinement rings. The RIE chamber consists of two parallel plates, RF power supply and pumping system.

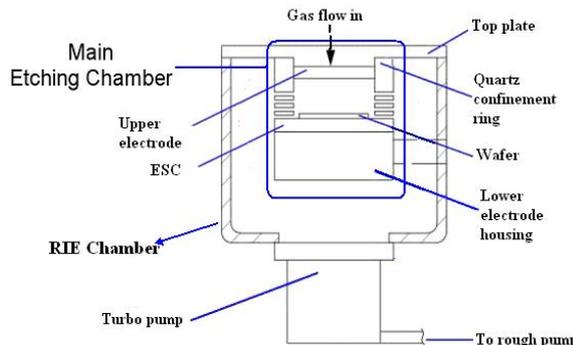


Figure 2: Sketch of the main etching chamber inside the RIE chamber.

Forty different signals were collected from 2300 Exelan Flex equipment for the reactive ion etching process. Twenty two signals are carefully chosen from all the signals. The chosen signals directly impact the wafer status. Table 1 provides more elaborations about RIE factors.

Table 1: RIE factors and their clarification.

#	Factor	Factor explanation
1	Bias Voltage	Direct-current voltage at electrode. The value of the bias voltage depends on the size of the electrode and on the gas pressure.
2	ESC Clamp Voltage	Electrostatic chuck (a mechanism for holding wafers using electrostatic attraction.)
3	ESC Current1	The Current pass in the first selected point on ESC
4	ESC Current2	The Current pass in the second selected point on ESC
5	ESC Temperature	Electrostatic chuck. Temperature (a mechanism for holding wafers using electrostatic attraction.)
6	Fluorine Pressure	The mixed Gases pressure inside the pipe
7	Forward Power 27 MHz	Applied 2 MHz for dissociation
8	Forward Power 2MHz	Applied 2 MHz for bombard
9	Gas 1	C ₄ F ₈ mass flow rate
10	Gas 10	Xe mass flow rate
11	Gas 11	C ₄ F ₆ mass flow rate
12	Gas 4	Ar mass flow rate
13	Gas 7	O ₂ mass flow rate
14	He Flow Inner	Mass flow rate for Helium gas inside the Inner tube of cooling system
15	He Flow Outer	Mass flow rate for Helium gas inside the outer tube of cooling system
16	He Pressure Inner	Helium gas Pressure inside the Inner tube of cooling system
17	He Pressure Outer	Helium gas Pressure inside the outer tube of cooling system
18	Pressure	The pressure value inside the RIE chamber
19	Process Time	The time from the initial to the end process
20	Reflect Power 27 MHz	The reflected 27 MHz which has not been used at the process
21	Reflect Power 2 MHz	The reflected 2 MHz which has not been used at the
22	Top Plate Temperature	The temperature on the upper plate

4. Data Preparation

This section explains the details of the data preparation performed in this study. Fig. 3 presents the percentages of the training and test wafers where the total number of wafers is one hundred twenty. The wafers were collected from the 2300 Exelan Flex machine. The ratio of the number of training wafers to the number of test wafers is three. Fourteen wafers (12%) from the ninety training wafers (75%) stand for unopened etch defected wafers, and five wafers (4%) from the thirty test wafers (25%) stand for unopened etch defected wafers.

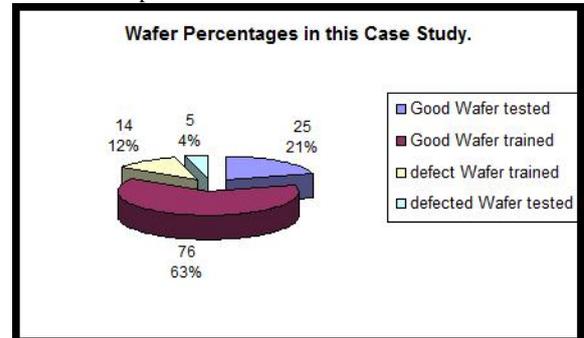


Figure 3: Percentages of the training and test wafers.

Data preparation techniques are used to obtain good prediction results. Three different data preparation techniques are suggested: raw data, sampling data and statistical summary data.

- Raw data preparation: one hundred eighty four data points is the minimum number of data points from the collected data. Thus the first one hundred eighty data

points are suggested as raw data inputs for offline prediction models and twenty data points are suggested as raw data inputs for online prediction models.

- Sampling: non-symmetric sampling is the second suggested preparation technique, which has ability to cover all etching steps, at the same time focusing on the main three etching steps (step 4, step 5 and step 10). Table 2 shows the number of captured samples in each step, where two samples are captured from stabilization steps (steps 1, 2, 3, 7, 8, and 9) and two from plasma ramp down steps (steps 6 and 11). More than half of the samples are captured from the main etching steps (steps 4, 5, and 10). As a result thirty-four captured samples cover all the etching steps. These thirty-four captured samples are used as inputs for offline prediction models.

Table 2: Number of suggested sampling for each step in the sampling technique.

Step Number	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10	Step 11
Number of sample	2	2	2	6	6	2	2	2	2	6	2

Figure 4 illustrates the position of captured samples, where the first data point of each step is captured, and the suggested sampling rate for the main etching steps is five whereas it is equal to three for other steps. These samples include the first data point of each stabilization step.

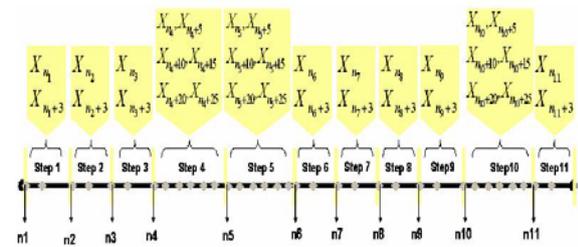


Figure 4: The position of captured samples.

- Statistical summary preparation: the last suggested preparation technique depends on mean and standard deviation values. Since many samples have no data for the sixth step it became out of interest. This means ten steps will be statistically summarized and applied in prediction models.

5. Principal Component Analysis

Principal component analysis (PCA) is an important analysis technique in multivariate statistics. It was first suggested in 1901 by Pearson [16], and formally developed by Hotelling [17]. The main idea of principal component analysis (PCA) is to represent a number of correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component (PC) accounts for the variability in the data as much as possible, the second PC is the linear combination with the second largest variance and orthogonal to the first PC, and so on. There are as many PCs as the number of the original variables. For many datasets, the first several PCs explain most of the variance, so that the rest can be disregarded with minimal loss of information. The objectives of using PCA are to reduce the dimensionality of a data set and to identify new underlying variables that are now orthogonal.

To enhance performance of the prediction model in this study, PCA is suggested to represent the RIE factors,

since simple neural networks with few nodes and connections tend to have better generalization capability. In this section, PCA technique automatically extracts three principle components (PCs) from all RIE factors (twenty two factors). It is important to treat each etching step separately in PCA, because each step has different inherent physical/chemical characteristics. Considering the overall process characteristics and the objective of model simplicity, it was decided that utilizing one PCA for each of the eleven steps would yield a better solution than utilizing a single PCA for the entire process. In this paper, principal component analysis was utilized for 90 training wafers. The principle components are found by computing the sample covariance matrix and selecting its eigenvectors (loading vectors) for the largest eigenvalues.

6. Architecture of Prediction Models

As stated before, this study combines back propagation neural network (BPNN) and principle component analysis (PCA) to construct the two prediction models. Prediction models are concerned with all etching process steps to predict the wafer status at the end of the etching process.

The BPNN in this study consists of three layers of neurons: the input layer, hidden layer, and output layer. The input layer receives external information such as RIE processing factors or principle components. From the output layer, predictions are produced with binary values to represent the wafer status. Since the network output is between zero to one, the zone that is smaller than a minimum threshold value is set to zero and the zone that is greater than a maximum threshold value is set to one. If the network output value is equal to one that means the wafer status is good, and otherwise it is defected. When the network output value is between the minimum and maximum values, then the network fails to predict the wafer status.

The BPNN also incorporates hidden layers of neurons, which do not interact with the outside world, but assist in performing nonlinear feature extraction on the data provided by the input and output layers. The number of hidden layers was set to one in this application. Training matters have to be settled with respect to the description of the BPNN network structure.

6.1. Training:

During training, the network is trained to associate outputs with input patterns. This principle is referred to as supervised learning. The training is continued until the training reached the maximum number of epochs or training neural network has MSE (mean square error) less than 10⁻⁶. The maximum number of epochs used during training the networks is set to 10000.

After training, the prediction performance of the models is evaluated with two test sets. In the first test set computation of the prediction error for new data points is performed. Data of twenty-five good wafers and five-defected wafers stand for the test data set. Two types of errors are obtained in the first test: type I prediction error occurred when good wafers are predicted as defected wafers, and type II prediction error occurred when defected wafers are predicted as good wafers.

The second test set depends on the recognition / rejection rate. The recognition rate is the percentage of test

samples recognized correctly and the output value is located outside the range between the minimum and maximum threshold values. The rejection rate is the percentage of input samples that could not be assigned to any particular class; because the output value is located somewhere between the minimum and maximum threshold values. The minimum and maximum values are determined for every prediction model after testing the training wafers, where the minimum value is the highest output value for a defected wafer, and the maximum value indicates the lowest value for a good wafer.

7. Evaluation of Prediction Models

The major aspect in this section is to evaluate the performance of the prediction models, and decide the best prediction model with respect to a constraint. The same training and test samples are used for all prediction models. The experimental data examined were collected from reactive ion etching of silicon dioxide thin film.

7.1. BPNN prediction model:

Figure 5 illustrates the offline BPNN prediction model, where the five significant factors (bias voltage, He inner flow, He outer flow, pressure, and reflect power 2MHZ) are the model inputs. The number of input neurons of the BPNN is different for each of the three data preparation techniques described above. When the raw data preparation technique is applied in the offline BPNN the number of input neurons is nine hundred, whereas it is one hundred seventy when captured sampling is applied, and one hundred for the statistical summary preparation technique.

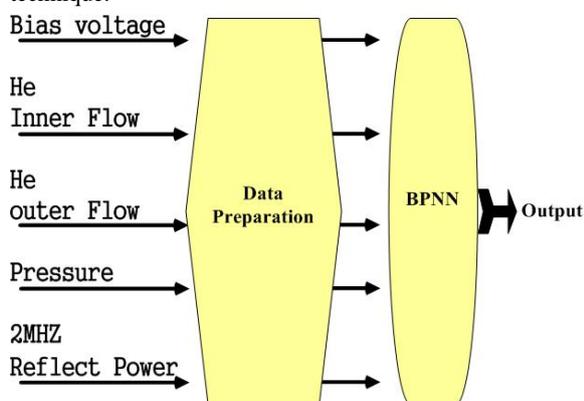


Figure 5: BPNN prediction model.

The raw data preparation technique does not prepare data as well as other data preparation techniques, even if it has good performance for predicting the wafer status by using two factors (He outer flow and bias voltage). In addition, the raw data preparation technique has no ability to cover all etching steps. Table 3 illustrates the offline prediction model performance by using sampling and statistical summary preparation techniques. Both data preparation techniques assist the offline BPNN prediction model to achieve zero error and 100% recognition rate. Recognition rate represents the percentage of test samples recognized correctly and the corresponding output values are located outside the range between the minimum and maximum threshold values, namely, 0.1 and 0.9, respectively.

Table 3: The performance of offline BPNN prediction model.

	BPNN Characteristic	
	Sampling	Statistical summary
Recognition rate	100%	100%
Training MSE	1.94311E-07	1.47594E-06
Testing MSE	5.12016E-09	7.46886E-07
# of input neurons	170	100
Error prediction	0%	0%

7.2. PCABPNN prediction model:

To construct offline PCABPNN prediction model the principle component analysis (PCA) and back propagation neural network (BPNN) are combined together. First, the PCA is adopted to extract valuable information from the twenty two RIE factors for each step. Then the extracted principle components for all the steps are combined together and prepared by the three different data preparation techniques described earlier in this paper. The neural network is trained afterwards by the prepared PC data of 90 training wafers and tested for prediction accuracy by the prepared PC data of 30 test wafers.

Figure 6 illustrates the offline PCABPNN prediction model. Principle component analysis is applied in this model to extract the input parameters for the neural network. Moreover the required time to find the principle components is much less than the required time to find the significant parameters.

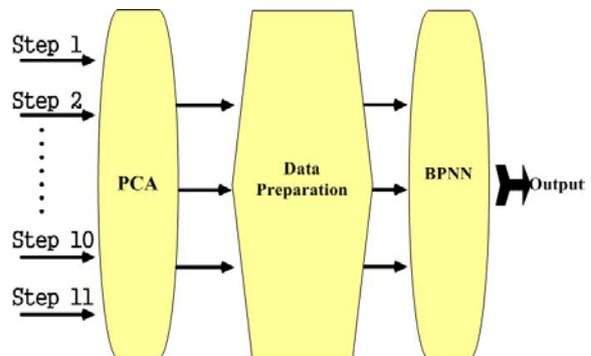


Figure 6: Offline PCABPNN prediction model.

One hundred eighty data points for each principle component parameter are prepared by the raw data preparation technique. Totally, there are five hundred and forty neuron inputs for the three PCs. Since the raw data preparation technique covers the first one hundred eighty data points of the PCs, the rest of data points are ignored. The ignored points of PCs may contain important information. Three defected wafers are incorrectly predicted as good wafers by using the raw data preparation technique.

In order to reduce the input data size one hundred and two captured samples from the three principle components are trained and tested using the BPNN (see Table 4). Four samples out of thirty tested samples were clamped between 0.1 and 0.9 (the minimum and maximum threshold values) by applying the captured sampling in the offline PCABPNN prediction model. Applying the statistical summary data preparation technique yields no samples in the failure zone, due to high accuracy (100% recognition

rate) of the prediction model. In general, the statistical summary data preparation technique has the best ability to enhance the accuracy of the offline PCABPNN prediction model.

Table 4: The performance of the offline PCABPNN prediction model.

	PCABPNN Characteristics		
	Raw PC	Sampling	Statistical summary
Recognition rate	96.67%	86.67%	100%
Training MSE	1.29021E-06	2.4213E-05	4.52214E-06
Testing MSE	0.101489096	0.041548198	4.05693E-05
# of input neurons	540	102	60
Error prediction	10.00%	3.33%	0%

8. Conclusion

Two prediction models for etching of silicon dioxide thin film in the MEMS fabrication process reactive ion etching were developed to predict the wafer status correctly by using statistical summary preparation techniques. Back propagation neural network (BPNN) is the first prediction model and is the backbone for the second prediction model, namely, the principle component analysis BPNN (PCABPNN). The two models achieve the objective of achieving fast and robust predictions. This paper describes the potential of these prediction models to reduce the overall cost of MEMS equipment, to achieve high yields and throughput during MEMS fabrication. The two prediction models do not depend upon monitoring wafers or expensive metrology rather they will enable inexpensive real-time wafer-to-wafer inspection application. The results from the evaluation of the prediction models indicate that robust, accurate and stable predictors have been constructed. Furthermore, a greater accurate performance of prediction models has been achieved by online BPNN prediction model.

References

- [1] N. Tudoroia , R.V. Patelb, K. Khorasania, "Neural Network Based Control Strategies for Improving Plasma Characteristics in Reactive Ion Etching". *Neurocomputing*, Vol. 69, 2006, 786-802.
- [2] B. Kim, M. Kwona, "Prediction of Plasma Etch Process by Using Actinometry-based Optical Emission Spectroscopy Data and Neural Network". *Journal of Materials Processing Technology*, Vol. 209, 2009, 2620-2626.
- [3] P. H. Choua, M. J. Wua, K. K. Chenb, "Integrating Support Vector Machine and Genetic Algorithm to Implement Dynamic Wafer Quality Prediction System". *Expert Systems with Applications*, Vol. 37, 2010, 4413-4424.
- [4] S. R. Bhatikar, R. L. Mahajan, "Artificial Neural-Network-Based Diagnosis of CVD Barrel Reactor". *IEEE Transactions on Semiconductor Manufacturing*, Vol. 15, 2002, 71-78.
- [5] F. L. Chen, S. F. Liu, "A Neural-Network Approach to Recognize Defect Spatial Pattern in Semiconductor Fabrication". *IEEE Transactions on Semiconductor Manufacturing*, Vol. 13, 2000, 366-373.
- [6] J. H. Lee, S. J. You, S. C. Park, "A New Intelligent SOFM-based Sampling Plan for Advanced Process Control". *Expert Systems with Applications*, Vol. 20, 2001, 133-151.
- [7] F. L. Chen, S. F. Liu, K. Y. Y. Doong, K. L. Young, "LOGIC Product Yield Analysis by Wafer Bin Map Pattern Recognition Supervised Neural Network". *IEEE International Symposium on Semiconductor Manufacturing*, Vol. 1, 2003, 501-504.
- [8] F. Di Palma, G. D. Nicolao, G. Miraglia, E. Pasquinetti, F. Piccinini, "Unsupervised Spatial Pattern Classification of Electrical-Wafer-Sorting Maps in Semiconductor Manufacturing". *Pattern Recognition Letters*, Vol. 26, 2005, 1857-1865.
- [9] S.S. Han, M. Ceiler, S. A. Bidstrup, P. Kohl, G. May, "Modeling the Properties of PECVD Silicon Dioxide Films Using Optimized Back Propagation Neural Networks". *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, Vol. 17, 1994, 174-182.
- [10] S. S. Han, G. S. May, "Using Neural Network Process Models to Perform PECVD Silicon Dioxide Recipe via Genetic Algorithms". *IEEE Transactions on semiconductor manufacturing*, Vol. 10, 1997, 279-287.
- [11] B. Kim, D. W. Kim, G. T. Park, "Prediction of Plasma Etching Using a Polynomial Neural Network". *IEEE Transactions on Plasma Science*, Vol. 31, 2003, 1330-1336.
- [12] B. Kim, W. S. Hong, "Use of Neural Network to Characterize a Low Pressure Temperature Effect on Refractive Property of Silicon Nitride Film Deposited by PECVD". *IEEE Transactions on Plasma Science*, Vol. 32, 2004, 84-89.
- [13] R. A. Zoroofi, H. Taketani, S. Tamura, Y. Sato, K. Sekiya, "Automated Inspection of IC Wafer Contamination". *Pattern Recognition*, Vol. 34, 2001, 1307-1317.
- [14] C. T. Su, T. Yang, C. M. Ke, "A Neural-Network Approach for Semiconductor Wafer Post-Sawing Inspection". *IEEE Transactions on Semiconductor Manufacturing*, Vol. 15, 2002, 260-266.
- [15] W. C. Chen, C. T. Chen, T. H. Ho, J. H. Chen, L. J. Sheu, "Use of Neural Network in Pattern Recognition of Semiconductor Etching Process". *The Proceedings of the 11th International Conference on Industrial Engineering and Engineering Management*, Vol. 1, 2005, 719-725.
- [16] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space". *Philosophical Magazine*, Vol. 2, 1901, 559-572.
- [17] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components". *Journal of Educational Psychology*, Vol. 24, 1933, 417-441.

Developing a Software to Predict Thermal Comfort of Humans at Work

N. J. Rabadi^{*,a}

^aDepartment of Mechanical Engineering, Faculty of Engineering and Technology, University of Jordan, Amman, Jordan

Abstract

The goal of this research work is to develop a user-friendly software that is capable of estimating the thermal comfort of a human working in the open air. Several numerical approaches were considered in the literature. The computational time and the complexity of those computational methods make it usable by professionals only. The developed computer software under this work can be used by low ranking field managers, who may not be computer experts, operating in the field where the workers are performing their duties. An important parameter crucial for the calculation of the energy transfer around the human is the mean radiant temperature, MRT. A method has been developed here to evaluate the MRT. It is suitable for indoor as well as outdoor applications. Weather and solar data are expected to be available for the outdoor applications. It can either be actively measured at the actual site under consideration, or calculated from the available data documented for numerous sites around the world.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: thermal comfort; human comfort

1. Introduction

The goal of this research work is to develop a user-friendly software that is capable of estimating the thermal comfort of a human working in the open air. Several approaches were considered in the literature. They vary from very sophisticated codes that use well-known commercially available CFD codes (ANSYS[1], FLUENT[2], PHEONIX[3]) that segment the human body into large number of segments each represent an independent source/sink of heat. The computational time and the complexity of the codes make it usable by professionals only. The goal of this work is to develop computer software that can be used by low ranking field managers, who may not be computer experts, operating in the field where the workers are performing their duties.

The approach followed here is similar to that recommended by ISO standard 7933[4] second edition 2004-8-15 entitled "Ergonomics of the thermal environment — Analytical determination and interpretation of heat stress using calculation of the predicted heat strain." An important parameter crucial for the calculation of the energy transfer around the human is the mean radiant temperature, MRT. The methods recommended by ASHRAE or ISO Standards to determine the MRT are not suitable for the applications under consideration. The reason is that this method is designed to calculate thermal comfort indices in black enclosures where there is no direct solar radiation; the case under consideration.

A method (called thereafter as "Modified ASHRAE method") has been developed here to evaluate the MRT. It

is suitable for indoor as well as outdoor applications. Weather and solar data are expected to be available for the outdoor applications. It can either be actively measured at the actual site under consideration, or calculated from the available data documented for numerous sites around the world.

The clothing insulation properties of the subject under consideration must be supplied as input data. Thermal insulation properties for normal clothing are available from literature ([5-9]).

The computer program developed under this work is capable of handling the following scenarios:

- a. Operator defines all environmental conditions in the vicinity of the subject:

This scenario is applied when the operator wishes to know how a human, wearing specified clothes and doing specified work, feels thermally if subjected to any practical, field type conditions. Air temperature, air velocity, humidity, solar radiation, space dimensions, geographical data, and radiative properties must be well defined within the vicinity of the worker.

- b. Operator defines the subject's geographical location; country and city name:

The program reads weather and geographical information on over a thousand cities worldwide from data available on the web (data has been compiled and linked to the program.) The compiled data includes all of the above-mentioned parameters. This scenario is applied only for outdoor operating conditions. Certainly the computed

* Corresponding author. e-mail: rabadi@ju.edu.jo

results depend largely on the accuracy of the available weather data.

- c. Operator defines the subject's geographical location through latitude and longitude:

Empirical correlations are used to estimate the solar radiation ([10-11]). The mean radiant temperature is then calculated. Compilation of those correlations is provided through this work. The correlations cover Jordan, however it can be extended to cover most regions of the world. Information on these correlations' uncertainty is not available yet. A subroutine has been written to accommodate this scenario. Extending the correlations to cover most regions of the world in the third scenario is a future extension for this research work.

2. Factors Effecting Thermal Comfort

Aims of comfort research are to find the factors for an individual or group that make them feel comfortable. Six major variables determine how warm or cold (i.e., how comfortable) a person feels[12]. Four of them classified as environmental; Air temperature, Air speed, Humidity, and Mean radiant temperature, and two are classified as personal; Activity, and Clothing insulation. Other factors that could be of importance in some cases are the individual differences, and recent thermal history.

3. Thermal Comfort Models

3.1. ASHRAE STANDARD 55 – 2004[13]

In this standard, thermal comfort is defined as the condition of mind which expresses satisfaction with the thermal environment and is assessed by subjective evaluation. The standard presents a method for predicting the thermal sensation and the degree of discomfort (thermal dissatisfaction) of healthy people exposed to moderate thermal environments. It also specifies acceptable environmental conditions for comfort of adults at atmospheric pressures equivalent to altitudes up to 3000 m. It applies to indoor environments where the aim is to attain thermal comfort or indoor environments where moderate deviations from comfort occur.

The standard determines an index that predicts the mean value of the votes (PMV) of a large group of people on a 7-point thermal sensation scale;

	3	Hot
	2	Slightly Warm
	1	Warm
PMV =	0	Neutral
	-1	Slightly Cool
	-2	Cool
	-3	Cold

and the PPD (predicted percentage of dissatisfied). The PMV/PPD model takes into account the main factors related to the steady-state thermal balance of the body;

- Environmental parameters:
 - Air temperature,

- Mean radiant temperature,
- Air velocity (< 0.2 m/s).
- Partial water vapor pressure (< 1.910 kPa).
- Activity level, affecting metabolic rate (1.0 met < metabolic rate < 2.0).
- Clothing level, affecting thermal resistance (< 1.5 clo).

Equation for PMV: The heat balance equation can be written as follows;

$$(M-W) = 3.96 \times 10^{-8} f_{cl} [(t_{cl} + 273)^4 - (t_r + 273)^4] + f_{cl} h_c (t_{cl} - t_a) + 3.05 [5.73 - 0.007 (M-W) - p_a] + 0.42 [(M-W) - 58.1] + 0.0173 M (5.87 - p_a) + 0.0014 M (34 - t_a) + L \quad (1)$$

Where:

$$t_{cl} = 35.7 - 0.0275(M-W) - I_{cl} \{ (M-W) - 3.05 [5.73 - 0.007 (M-W) - p_a] - 0.42 [(M-W) - 58.1] - 0.0173 M (5.87 - p_a) - 0.0014 M (34 - t_a) \} \quad (2)$$

and L equals the imbalance between heat generated and heat dissipated. The values of h_c and f_{cl} can be estimated from tables and equations given in the Engineering Data and Measurements section, Table 6, Chapter 8 ASHREA Fundamental Vol., 2001.

$$h_c = \begin{cases} 2.38(t_{cl} - t_a)^{0.25} & \text{for } 2.38(t_{cl} - t_a)^{0.25} > 1.21\sqrt{v_a} \\ 12.1\sqrt{v_a} & \text{for } 2.38(t_{cl} - t_a)^{0.25} < 1.21\sqrt{v_a} \end{cases}$$

$$f_{cl} = \begin{cases} 1.00 + 2.0 I_{cl} & I_{cl} < 0.5 clo \\ 1.05 + 0.1 I_{cl} & I_{cl} > 0.5 clo \end{cases} \quad (3)$$

The PMV empirical equation is expressed as follows;

$$PMV = [0.303 e^{(-0.036M)} + 0.028] L \quad (4)$$

Where L is the imbalance between net heat generated within the body and the actual heat transferred to the environment;

$$L = (M-W) - \{ 3.96 \times 10^{-8} f_{cl} [(t_{cl} + 273)^4 - (t_r + 273)^4] + f_{cl} h_c (t_{cl} - t_a) + 3.05 [5.73 - 0.007 (M-W) - p_a] + 0.42 [(M-W) - 58.1] + 0.0173 M (5.87 - p_a) + 0.0014 M (34 - t_a) \} \quad (5)$$

Substituting for (M-W) in the t_{cl} equation to get

$$t_{cl} = 35.7 - 0.0275(M-W) - I_{cl} \{ 39.6 \times 10^{-9} f_{cl} [(t_{cl} + 273)^4 - (t_r + 273)^4] + f_{cl} h_c (t_{cl} - t_a) \} \quad (6)$$

t_{cl} should be found by iterative methods.

PMV	is the Predicted Mean Vote.
M	is the metabolic rate, in W/m ² of body surface area.
W	is the external work, in W/m ² , equal to zero for most activities.
I _{cl}	is the thermal resistance of clothing, in (m K / W.)
f _{cl}	is the ratio of a person's surface area while clothed, to the surface area while nude.
v _a	the average wind velocity
t _a	is the air temperature, in °C.
t _r	is the mean radiant temperature, in °C.
p _a	is the partial water vapor pressure, in Pa.
h _c	is the convective heat transfer coefficient, in W/m ² K.
t _{cl}	is the surface temperature of clothing, in °C.

PMV is derived for steady state conditions but can be applied during minor fluctuations of one or more variables (time-weighted averages during previous 1- hour period should be used). It is calculated for conditions when the human body is in thermal equilibrium. PMV index should only be used for PMV between -2 and 2 and it is recommended for use only for the following range of conditions

$M = 58$ to 116 W/m^2 (1.0 to 2 met)
 $I_{cl} = 0$ to $0.310 \text{ m}^2\text{k/W}$ (0 to 2 clo)
 $t_a = 10$ to $30 \text{ }^\circ\text{C}$; $t_r = 10$ to $40 \text{ }^\circ\text{C}$
 $v_a = 0$ to 0.2 m/s (or less if draughts are important)
 $p_a = 0$ to 1910 Pa . (also relative humidity should be between 30 and 60%)

3.2. Predicted percentage dissatisfied:

PPD is a measure of the number of people likely to feel uncomfortably warm or cool in a given environment, i.e., the number of people voting -3, -2, +2 or +3 within the PMV scale.

$$\text{PPD} = 100 - 95.0 \exp[-(0.03353 \text{ PMV}^4 + 0.2179 \text{ PMV}^2)] \quad (7)$$

The PPD index predicts the number of thermally dissatisfied people within a large group. The rest of the group would vote -1, 0 or +1 on the PMV scale. The minimum PPD is 5%.

3.3. The MEMI model[14-15]:

The predicted mean vote, PMV, model explained earlier is a steady-state theoretical model that is widely recognized and used worldwide. It can only apply to humans exposed for a long period to constant conditions at a constant metabolic rate. The major limitation of the PMV model is the explicit constraint of skin temperature and evaporative heat loss to values for comfort and "neutral" sensation at a given activity level. The PMV model has a physical and also thermophysiological background and considers all relevant meteorological parameters. However it can only be used to determine a comfort index. It cannot be used to quantify real values of heat fluxes or body temperatures for a given environment. The MEMI (Munich Energy Balance Model for Individuals) takes care of this issue by introducing an assumption that the sensible heat lost from the skin is equal to the heat carried by the blood flow from the core to the skin. MEMI uses the following heat balance equation:

$$H + C + R + E_D + E_{sr} + E_{lr} + E_{sw} + E_f = S \quad (8)$$

H = Heat production;
 C = Convective heat flux;
 R = Radiative heat flux
 ED = Water vapor diffusion
 E_{sr} = Sensible heat loss by respiration
 E_{lr} = Latent heat lost by respiration
 E_{sw} = Heat lost by sweat evaporation
 E_f = Heat added or lost from food or drink
 S = Net heat stored in the core.

The output of this method is: S, the net heat stored in the body, t_c , the core temperature and t_{sk} , the skin

temperature. The details of this method can be found in [14-15].

3.4. The 2- Node model (ET*- DISC)

The 2-Node theoretical Model also uses a heat balance method to predict thermal comfort, but the model is transient with time rather than being steady state like PMV. ET* stands for New Effective Temperature where "effective temperature" is an temperature index that accounts for radiative and latent heat transfers. ET* can be calculated using the '2-Node' model. The 2-node model determines the heat flow between the environment, skin and core body areas on a minute-by-minute basis. Starting from an initial condition at time=0, the model iterates until equilibrium has been reached (60 minutes is a typical time). The final mean skin temperature and skin wettedness are then associated with an effective temperature. DISC predicts thermal discomfort using skin temperature and skin wettedness. Details of this method can be found in [16].

3.5. PD model[16]

PD or "predicted percent dissatisfied due to unwanted local cooling (draft)," is an empirical model. It is a fit to data of persons expressing thermal discomfort due to drafts. The inputs to PD are air temperature, air velocity, and turbulence intensity. The PD equation is:

$$\text{PD} = 3.413 (34 - T_a) (v - 0.05)^{0.622} + 0.369 v T_u (34 - T_a)(v - 0.05)^{0.622} \quad (8)$$

T_u is the turbulence intensity expressed as a percent. **0** represents laminar flow and **100** means that the standard deviation of the air velocity over a certain period is of the same order of magnitude as the mean air velocity. v is the air velocity (m/s) and T_a is the air temperature in $^\circ\text{C}$.

The PD equation arises from two studies in which 100 people were exposed to various combinations of air temperature, air velocity, and turbulence intensity. For each combination of conditions, the people were asked if they felt a draft. PD represents the percent of subjects who voted that they felt a draft for the selected conditions.

3.6. PS model [16]

PS is a fit to data of comfortable persons choosing air velocity levels. The PS equation predicts the air velocity that will be chosen by a person exposed to a certain air temperature when the person has control of the air velocity source. The PS equation is

$$\text{PS} = 1.13 (T_{op})^{0.5} - 0.24 T_{op} + 2.7 (v)^{0.5} - 0.99 v \quad (9)$$

T_{op} is operative temperature (in degrees Celsius) and v is the air velocity in m/s. The PS equation arises from a study in which **50** people were asked to adjust an air velocity source as they pleased when exposed to a specific air temperature. PS represents the cumulative percent of people choosing a particular air velocity at the specific temperatures tested in this experiment.

3.7. TS model [16]

The TS model is a fit to data of thermal sensation as a linear function of air temperature and partial vapor pressure. TS is an equation that predicts thermal sensation vote using a linear function of air temperature T_a , and partial vapor pressure p , in kPa. The TS equation is:

$$TS = 0.245T_a + 0.248p - 6.475 \quad (10)$$

4. Input Data

4.1. Surface Area of Body

Du Bois area [17]: The surface area of skin of an "average" adult is 1.8 m^2 . The total heat production of an "average" person at rest per hour is $58.2 \times 1.8 = 104.76 = 105$ watts.

The Du Bois area normally varies between 1.3 m^2 and 2.2 m^2 and in any setting the heat produced by sedentary adults will vary between about 75.66 watts for 1.3 m^2 and 128 watts for 2.2 m^2 .

4.2. Activity Level in Mets

The level of activity of the person must be determined to calculate his thermal comfort level. ASHRAE standards 55-2004 list various activities with the metabolic rate expected from a healthy body. Table 1 is an extract from ASHRAE standards 55.

Table 1: Metabolic Rates for Typical Tasks.*

Activity	Metabolic rate in mets
Reclining	0.8
Seated, quietly	1.0
Sedentary activity (office, dwelling, lab, school)	1.2
Standing, relaxed	1.2
Light activity, standing (shopping, lab, light industry)	1.6
Medium activity, standing (shop assistant, domestic work, machine work)	2.0
High activity (heavy machine work, garage work)	3.0

*See ASHREA for more data

4.3. Clothing:

Detailed data on the clothing of the person under consideration must be available. ASHRAE standards 55-2004 list clothing insulation data for typical ensembles as well as specific garments. Insulation values for special ensembles made of combination of garments can be calculated by adding insulation values of the individual garments. Below is useful data for clothing insulation;

- Clothing insulation is measured in clo units (Icl)
- $1 \text{ clo} = 0.155 \text{ m}^2 \text{ }^\circ\text{C}/\text{W}$
- Lowest clo value is 0 (naked body)
- Highest practical clo value = 4 clo (Eskimo clothing, fur pants, coat, hood, gloves etc.)

- Summer clothing ~ 0.6 clo
- Winter clothing ~ 1 clo
- $I_{cl} \sim 0.15 \times \text{weight of clothes in lbs.}$
- 10 lbs of clothing ~ 1.5 clo
- 1 clo maintains sedentary man (1 met) indefinitely comfortable at 21°C , 50% RH, 0.01 m/sec air speed.
- Wind speed affects clothing insulation
- Porosity - water vapor transfer through clothing affects its insulation value
- Permeation efficiency factor (Fpcl) ranges from 0 = completely impermeable fabric 1 = absence of clothing
- Thickness - effects insulation value
- Tog - European unit of thermal insulation 1 tog = 0.645 clo
- Tested insulation values for 7 chairs ranged between 0.1 – 0.3 clo for chairs with solid seats and backs.

4.4. Thermal Conditions:

Atmospheric influences on the thermal sensations of a person depend on four thermal environmental variables listed below.

- Air temperature (t_a): is the average temperature of the air surrounding the body. The average is with respect to location and time.
- Mean radiant temperature (MRT): is the average temperature of the surfaces in a space. Mean radiant temperature may be higher or lower than the air temperature in a space. Mean Radiant Temperature (t_r) is the uniform temperature of the surface of an imaginary black enclosure where the radiant exchange of heat between this enclosure and a man would be equal to the radiant exchanges in the actual environment.
- Air speed v_a : Is the average speed of the air to which the body is exposed. The average is with respect to location and time.
- Humidity: Humidity refers the mass of water vapor present in a unit volume of air (moisture content). It may be expressed in terms of several thermodynamic variables, including vapor pressure, dew point temperature, humidity ratio, and relative humidity.

4.5. Solar and Site Geographical Data:

Site data including altitude, elevation, absorptivity or albedo, slope azimuth, solar azimuth, slope angle, zenith angle, declination angle, hour angle, date and time of the day are all necessary to calculate the ground temperature needed for calculating the mean radiant temperature [18-23].

5. Calculation of Mean Radiant Temperature

5.1. ASHRAE Standard for Calculation of MRT (t_r):

The classical method to calculate t_r uses the values of the surrounding surface (i.e. wall, window, sofa) temperatures (ASHRAE, 2001). Each temperature is weighted according to its position relative to the person. The method assumes the surface materials have a high enough emittance, ϵ , to be considered radiatively black or

ideal. This assumption is reasonably valid for most enclosures, but its effect should be considered when analyzing the results. If the surfaces of the analyzed enclosure do not have a high emittance, then the results are not reliable. This assumption imposes a critical limitation on this method. In addition, this method does not take into account low-E glass or other types of advanced glazing systems. The published emissivity of low-E glass is less than 0.1 in the infrared wavelength range. Since the glass is opaque in that range, the rest of the radiant energy is reflected back into the room. The classical t_r method does not have the capability to handle this situation. Another case where the ASHRAE method would fail is solar radiation, which is short-wavelength radiation, shining through a window. This method fails to consider any window transmission and only considers the wall surface temperatures as boundary conditions.

Each of the surfaces is considered to be isothermal or has a uniform temperature, t_i . If this assumption is not valid for a single large surface, the surface is sub-divided until the assumption is valid. The view factors, F_{p-i} between the point P to be analyzed and all the surfaces i are calculated by some method. The t_r is then calculated as (ASHRAE, 2001).

$$t_r^4 = t_1^4 F_{p-1} + t_2^4 F_{p-2} + t_3^4 F_{p-3} + \dots + t_n^4 F_{p-n} \tag{11}$$

The temperatures for the calculation are in Kelvin and the view factors are dimensionless. For rectangular surfaces, ASHRAE (2001) provide view factor charts for the human body. View factors for standard geometric shapes can be found in a standard heat transfer text such as Incropera and DeWitt (1990) [24] or Siegel and Howell (1981) [25].

5.2. Radiant Intensity Method [26-27]:

The second method uses the fundamental t_r definition to calculate t_r in terms of the radiant intensity balance at a particular point in the room. This definition states that t_r is the uniform surface temperature of an imaginary black enclosure in which the radiation from the occupant equals the radiant heat transfer in the actual non-uniform enclosure. In a room where all the surfaces and the air are at the same temperature, the mean t_r and the air temperature are equal. As the difference between the surface temperatures and the air temperature increases, the difference between the t_r and the air temperature increases. This approach determines the radiant intensity field within the room and then uses that intensity field to calculate the actual radiant heat transfer from the occupant. The two primary advantages of this method are that: (1) determination of the view factors is unnecessary; and (2) the intensity field includes the effect of wall surface properties and any other intensity boundary condition, such as solar insolation.

The basic formulation begins with writing the definition of the t_r in terms of mathematical quantities;

$$Q_{p,b} (t_r) = Q_{p,act} \tag{12}$$

In this equation, the term $Q_{p,b}$ represents the radiant heat transfer from the occupant in a room with radiatively black walls and a uniform surface temperature. According to this definition, this term is determined by

$$Q_{p,b} = A_{eff} \sigma t_r^4 \tag{13}$$

The effective area is calculated by (ASHRAE, 1996)

$$A_{eff} = f_{eff} A_D \tag{14}$$

The effective radiation area of a person, f_{eff} equals 0.73 for a standing person (ASHRAE, 1996) and the DuBois area, A_D is estimated from a person's height and mass. For an average person, A_D equals 1.821 m². The term on the right side of Equation (6.2) represents the actual radiation heat transfer experienced by an occupant. This term is directly related to the intensity field by

$$Q_{p,act} = \int I_\lambda (\Omega) A_p (\Omega) d\Omega \tag{15}$$

This equation is an integral over all the directions represented by the solid angle Ω [28]. The intensity and projected area in the direction Ω are represented by $I_\lambda (\Omega)$ and $A_p (\Omega)$, respectively.

Figure 2 shows a sample two-dimensional intensity distribution for a point. At this point in the development, some method must be employed to determine the intensity field. The development is discussed next.

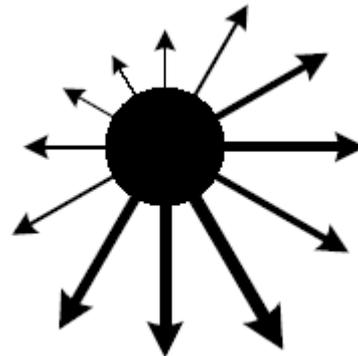


Figure 1: Sample two-dimensional radiant intensity field.

The net radiation at a particular point in the room as represented by the intensity arrows in Figure 1 is calculated using an approximation of the continuous form around the point [Equation (15) above]. This approximation was first employed in the early 1960's as a technique to determine neutron transport. Since then, the approximation has been extensively validated and used in radiation heat transfer studies. The approximation is given by

$$\dot{Q}_p = \sum I^j A^j w^j \tag{16}$$

where, the variable I^j is the intensity coming from a given discrete direction j , w^j is the quadrature weighting function for that direction, and A_p^j is the projected area in the given direction. The projected area from direction j is provided in the *HVAC Systems and Equipment Handbook* (ASHRAE, 1996). The general equation for the projected area is

$$A_p^j = F_p^j f_{eff} A_D \tag{17}$$

where f_p^j is the projected area factor in direction j . Factor charts for sitting and standing people are given by Fanger (1967) [28] and ASHRAE (1996).

Combining Equations (12) through (17) results in the final form of a very generalized method for determining t_r .

$$t_r \cong \left[\frac{\sum I^j A_p^j w^j}{f_{eff} A_D \sigma} \right]^{0.25} \quad (18)$$

This equation provides a more generalized approach to calculate t_r than using the surrounding surface temperatures given in the classical method. This approach, using the localized radiant intensity field, was extensively validated lately since it accurately incorporates the various emissivities and non-uniform surface temperatures of the surrounding environment. When enough information about I^j is available, Equation 18 can be executed and an approximate value of t_r is calculated.

5.3. Modified ASHRAE method:

To calculate t_r , the relevant properties and dimensions of the radiating surfaces and the sky view factor as well as the posture of the human body (e.g. seated or standing) must be known. The entire surroundings of the human body are divided into n thermal surfaces with the temperatures t_i ($i = 1, n$) and emission coefficients \mathcal{E}_i , to which the solid angle proportions ("angle factors") F_i are to be allocated as weighting factors. Long-wave radiation [$E_i = \mathcal{E}_i * \sigma * t_{si}^4$, with σ the Stefan-Boltzmann constant ($=5.67 * 10^{-8} \text{ W/m}^2\text{K}^4$) and t_{si} the temperature of the i^{th} surface,] diffuse short-wave radiation D_i are emitted from each of the n surfaces of the surroundings. This results in a value for t_r (K) as:

$$t_r = \left[\frac{1}{\sigma} \sum_{i=1}^n \left(E_i + \alpha_k \frac{D_i}{\mathcal{E}_p} \right) F_i \right]^{0.25} \quad (19)$$

\mathcal{E}_p is the emission coefficient of the human body (standard value 0.97.) D_i is the total of diffuse solar radiation and diffusely reflected global radiation. α_k is the absorption coefficient of the irradiated body surface for short-wave radiation (standard value 0.7.) The mean radiant temperature, t_r is replaced by t_r^* , if there is also direct solar radiation:

$$t_r^* = \left[t_r^4 + f_p a_k \frac{I}{\sigma \mathcal{E}_p} \right]^{0.25} \quad (20)$$

I is the radiation intensity of the sun on a surface perpendicular to the incident radiation direction. The surface projection factor f_p is a function of the incident radiation direction and the body posture. For practical application in human-biometeorology, it is generally sufficient to determine f_p for a rotationally symmetric person standing up or walking. f_p ranges from 0.308 for 0° of the angle of the solar altitude and 0.082 for 90° .

Therefore;

$$t_r^* = \left[\left(\frac{1}{\sigma} \sum_{i=1}^n (E_i + 0.722 D_i) F \right) + 0.722 f_p \frac{I}{\sigma} \right]^{0.25} \quad (21)$$

The problems associated with determining the angle factors F_i are discussed in detail in ASHRAE Fundamentals 2001. In the case of large flat surfaces without any restrictions of the horizon, the problem of determining F_i is reduced to an upper and a lower hemisphere with angle factors of 0.5 for both situations.

For cases where the object is in the outdoor conditions, solar irradiation can be calculated if solar data is available for outdoor conditions for that location. Solar data is available on the web or can be purchased and stored for faster recall. Note that the equation for calculating mean radiant temperature would then be;

$$t_r^* = \left[\frac{\sigma \mathcal{E}_g F_g t_g^4}{\sigma} + \frac{0.722 * F_{uh} D_{uh}}{\sigma} + 0.722 f_p \frac{I}{\sigma} \right]^{0.25} \quad (22)$$

Where the subscript uh and g stand for upper hemisphere and ground respectively, $F_g = F_{uh} = 0.5$.

Therefore;

$$t_r^* = 100 \left[0.45 * 10^{-8} t_g^4 + 0.06 D + 0.127 f_p I \right]^{0.25} \quad (23)$$

Where the first term between square brackets accounts for the long wave radiation from the ground, the second term accounts for the short wave diffuse radiation from the sky on a horizontal surface, and the third term accounts for the short wave direct beam radiation on a surface normal to the beam direction. The diffuse irradiation intensity D is different than (D_i) and I is the beam solar radiation in that outdoor location, $\mathcal{E}_g \cong 0.9$ is the emissivity of the ground and t_g is the temperature of the ground. The ground temperature is calculated using the procedure described in [29].

6. Conclusion and Recommendations

A computer code is developed to estimate the thermal comfort indices PMV/PPD for a human carrying some well-defined activity and wearing well-defined clothes and exposed to well-defined environmental conditions. A method recommended by ASHRAE and modified by the author is used to calculate the comfort indices and in the construction of the computer code. The method is simple and it is based on a steady state analysis. The modification included the effect of solar direct and diffused radiation and the radiation reflected and emitted from the ground surrounding the worker on the mean radiant temperature. It also included the effect on the mean radiant temperature of any radiating surfaces close to the subject. More data is needed regarding environmental conditions and insulation properties of special operations clothing.

Future work may include transient conditions whereby the code should deal with the worker intervention in a harsh environmental condition gradually. That is to say, a worker can be connected to a computer by remote device that sends information about the conditions in his

immediate vicinity. The computer code will then calculate the expected comfort indices and warn the worker ahead of time before reaching the critical limit of comfort.

7. Acknowledgment

This work is completed during the sabbatical leave from the Department of Mechanical Engineering at The University of Jordan, Amman, Jordan. This work was started at the Center of Computer Aided Design at the University of Iowa at Iowa City, Iowa. The author appreciates the help provided to him by both institutes to make this piece of research a reality.

References

- [1] <http://www.ansys.com/>
- [2] Fluent – Commercially available CFD software package based on the Finite Volume method. A product of Fluent Inc., Center Resource Park, 10 Cavendish Court, Lebanon, NH, USA.
- [3] <http://www.cham.co.uk/>
- [4] INTERNATIONAL STANDARD ISO 7933, Second edition, "Ergonomics of the thermal environment — Analytical determination and interpretation of heat stress using calculation of the predicted heat strain 2004-08-15."
- [5] HOLMER, H. NILSSON, G. HAVENITH and K. PARSONS, "Clothing Convective Heat Exchange; Proposal for Improved Prediction in Standards and Models. Hyg., Vol. 43, No. 5, pp. 329-337, 1999.
- [6] G. HAVENITH, I. HOLMER, E. A. DEN HARTOG, and K. C. PARSONS, "Clothing Evaporative Heat Resistance; Proposal for Improved Representation in Standards and Models." Ann. Occup. Hyg., Vol. 43, No. 5, pp. 339-346, 1999.
- [7] K. C. PARSONS, G. HAVENITH, I. Holmer, H. Nilson, and J. Malchair, "The Effects of Wind and Human Movement on the Heat and Vapor Transfer Properties of Clothing," Ann. Occup. Hyg., Vol. 43, No. 5, pp. 347-352, 1999.
- [8] K. Kuklane and I. Holmer, "Ergonomics of Protective Clothing." Proceedings of the 1st European Conference on Protective Clothing held in Stockholm, Sweden, May 7–10, 2000, isbn 91-7045-559-7.
- [9] T.M. McLellan, J.I. Pope, J.B. Cain, S.S. Cheung, "Effects of metabolic rate and ambient vapor pressure on heat strain in protective clothing," Eur J Appl Physiol (1996) 74: 518-527.
- [10] M.S. Audi and M.A. Alsaad, "Simple hourly global solar radiation prediction models," J. of Renewable Energy, Volume 1, Issues 3-4, 1991, Pages 473-478
- [11] Al-Salaymeh, "Modelling of Globally daily Solar Radiation on Horizontal Surfaces for Amman City," Emirates J. for Eng. Research, 11(1), 49-56 (2006).
- [12] Kenneth C. Parsons, "Human thermal environments: the effects of hot, moderate, and cold environments on human health, comfort, and performance," publisher; Taylor & Francis, 2003.
- [13] ASHRAE Standard 55-2004 -- Thermal Environmental Conditions for Human Occupancy (ANSI Approved)
- American Society of Heating, Refrigerating and Air-Conditioning Engineers / 2004
- [14] H. Mayer and P. Höppe, "Thermal comfort of man in different urban environments," Theoretical and Applied Climatology Volume 38, Number 1, 43-49, DOI: 10.1007/BF00866252.
- [15] Jianhua Huang, "Prediction of air temperature for thermal comfort of people in outdoor environment," Intl. J. of Biometeorology, vol. 51, no. 5, pp. 375-382, 2007.
- [16] de Dear, Richard and Brager, G. S., "Developing an Adaptive Model of Thermal Comfort and Preference." University of California, Berkeley, Publication Date: 01-01-1998.
- [17] <http://www.halls.md/body-surface-area/refs.htm>.
- [18] Jiménez, J. I., Alados-Arboledas, L., Castro-Diez, Y. and Ballester, G., "On the estimation of long-wave radiation flux from clear skies." (1987) Theoretical. Appl. Climatology., 38: 37-42.
- [19] Johnson G. T. and Watson I. D., "The determination of view-factor in an urban canyon." (1984) Jour. of Climate and Appl. Meteor., 23: 329-335.
- [20] Kyle, H. L., Ardanuy, P. E. and Hurley, E. J., "The status of the Nimbus-7 earth-radiation budget data set." (1985) Bull. Am. Meteor. Soc., 66: 1378-1387.
- [21] Oke, T. R., "Urban climatology and the tropical city: an introduction." Proc. of the technical conference of urban climatology and its applications with special regard to tropical areas, (1986) WMO No.652, pp1-25.
- [22] Oke, T. R., "Boundary layer climates." (1987) 2nd ed., Methuen, N.Y.
- [23] Terjung, W. H. and O'Rourke, P. A., "Energy exchanges in urban landscapes: selected climatic models." (1980) Publ. in Clim., Vol. XXXIII, C. W. Thornthwaite associates.
- [24] Incropera, DeWitt: Fundamentals of Heat and Mass Transfer, 5th Edition.
- [25] Siegel, R. and Howell, J.R, Thermal radiation heat transfer, Hemisphere Pub. Corp
- [26] Nasri Rabadi & Anith Mathai, "Thermal comfort of Humans in action," Report submitted to CCAD, University of Iowa at Iowa City, 2004.
- [27] Nasri Rabadi and Nicole Grosland, "Developing a Software to Predict Thermal comfort of a Virtual Soldier in action," Report submitted to CCAD, University of Iowa at Iowa City, 2005.
- [28] Charles, K.E., "Fanger's Thermal Comfort and Draught Models," IRC-RR-162 October 10, 2003.
- [29] S. Khoury, A. Abdel-Bagi, & Y. haddadin, "Thermal Comfort for Human in Action," Final year project Submitted to the Mechanical Eng. Dept. Univ. of Jordan, 2006. Supervisor: Prof. Nasri Rabadi.

Appendices

A. The Computer Code:

The computer program handles the following scenarios:

- The operator defines all environmental conditions. This scenario is applied when the operator wishes to know how a human, wearing specific clothes and doing specific work, feels thermally if subjected to

any practical, field type conditions. Air temperature, air velocity, humidity, solar radiation, and space dimensions and radiative properties must be well defined. This scenario applies for both indoor and outdoor conditions.

- The location is determined through country and city name. The program reads weather and geographical information from data incorporated within the code. Data includes all of the above-mentioned parameters. This scenario is applied only for outdoor operating conditions. Accuracy of computed results depends largely on the uncertainty of the weather data available.
- The location is determined through elevation, longitude and latitude. Empirical correlation to estimate the solar radiation is used. The mean radiant temperature is then calculated. A subroutine is written to accommodate such needs, however the available correlations cover the city of Amman, Jordan. This of course can be extended to cover other regions of the world as needed.

The work done by many Jordanian researchers provides a useful resource of modeling of the solar energy data in Jordan. The empirical correlations reported by A. Al-Salaymeh [11] are considered here to feed the necessary information to the computer code. The global solar radiation $G(\text{kW}\cdot\text{hr}/\text{m}^2\cdot\text{day})$ for Amman, Jordan is expressed using the following 4th order polynomial correlation

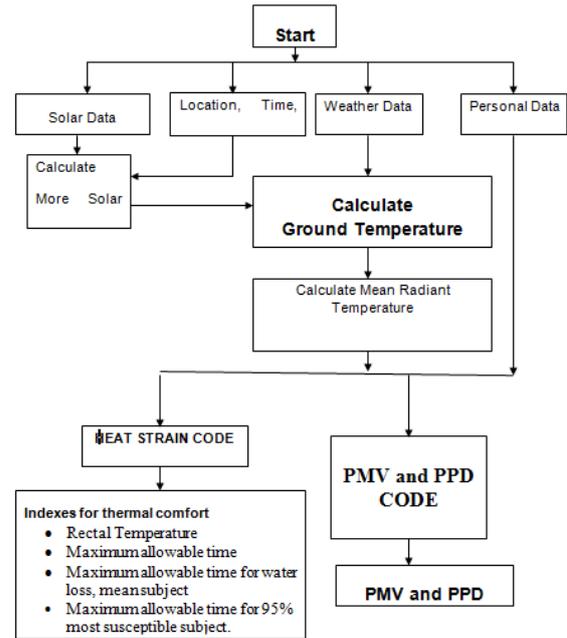
$$G = 2706.8 - 15.5x + 0.84x^2 - 0.000431x^3 + 5.9 \cdot 10^{-6}x^4 \quad (24)$$

Where x is the day number in the year.

This correlation is used because its regression coefficient is high (≈ 0.96) which means excellent fit for the measured data of the global solar radiation of Amman-Jordan.

B. The Code Flowchart

The flow chart explaining the code calculation steps is shown below. The boxes shown on the flow chart are detailed in Appendix c. The code is written in Visual Basic and is very much user friendly. Examples on code execution results are also shown in Appendix D.



Flow chart showing the general procedure in estimating the thermal comfort indexes.

C. Data boxes

Location, time, and date BOX

- Country (optional)
- City (optional)
- Latitude: $\Phi = -90^\circ \rightarrow +90^\circ$
- Longitude $L = -180 \rightarrow +180$
- Altitude; $z = -400\text{m} \rightarrow 5000\text{m}$
- Date; m-d-y
- Time: 24 hr base
- Slope angle $\phi = 0 \rightarrow 90^\circ$
- Slope azimuth angle $\Omega_{sl} = 0 \rightarrow 360$

Weather Data BOX

Air temperature $t_a = -15 \rightarrow 55^\circ\text{C}$
 Air pressure $P_a = 99 \text{ kPa} \rightarrow 102 \text{ kPa}$
 Vapor pressure $P_v = 0.1 \text{ kPa} \rightarrow 15.68 \text{ kPa}$
 Humidity ratio $H_u = 30\% \rightarrow 100\%$
 Wind velocity $v = 0 \rightarrow 0.3$
 Bowen ratio $\delta = 0.1 \rightarrow 12$
 Thermal conductivity of substrate $k = 0.25 \rightarrow 2.7$
 Ground surface emissivity $\epsilon_g = 0.8 \rightarrow 0.9$

Solar Data BOX

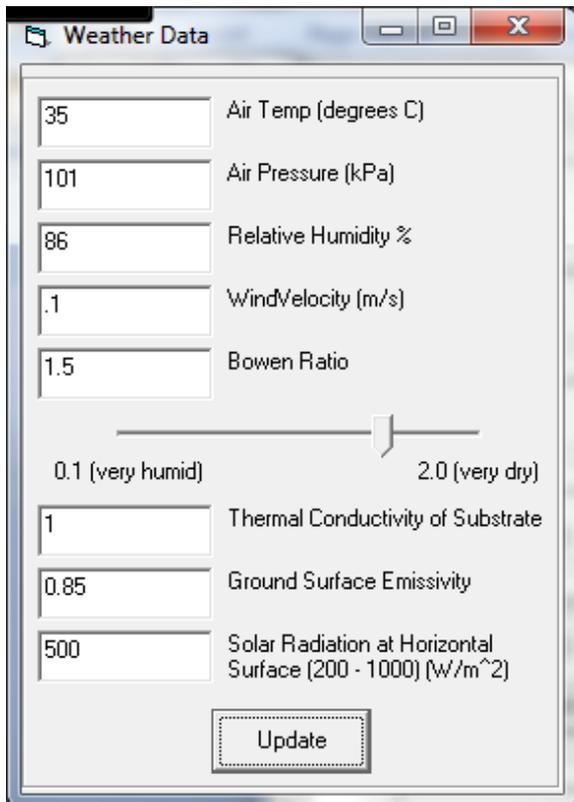
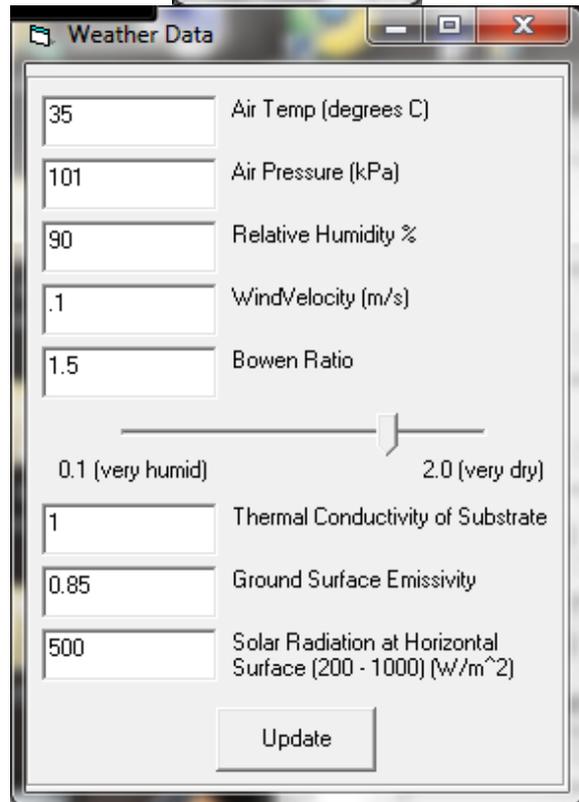
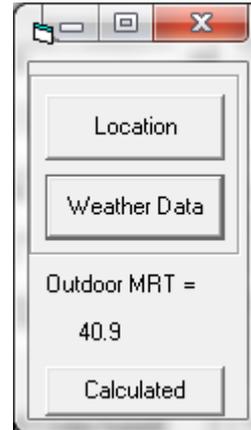
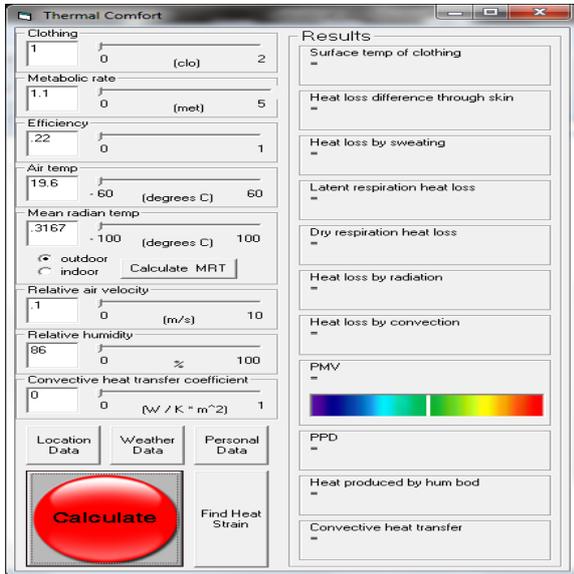
INPUT:
 Solar radiation at top of atmosphere $I_{et} = 1370 \text{ W/m}^2$
 Stefan Boltzmann constant, $\sigma = 5.67 \cdot 10^{-8} \text{ W/m}^2/\text{K}^2$
 Solar radiation at horizontal surfaces, $I_o = 200 \rightarrow 1000 \text{ W/m}^2$

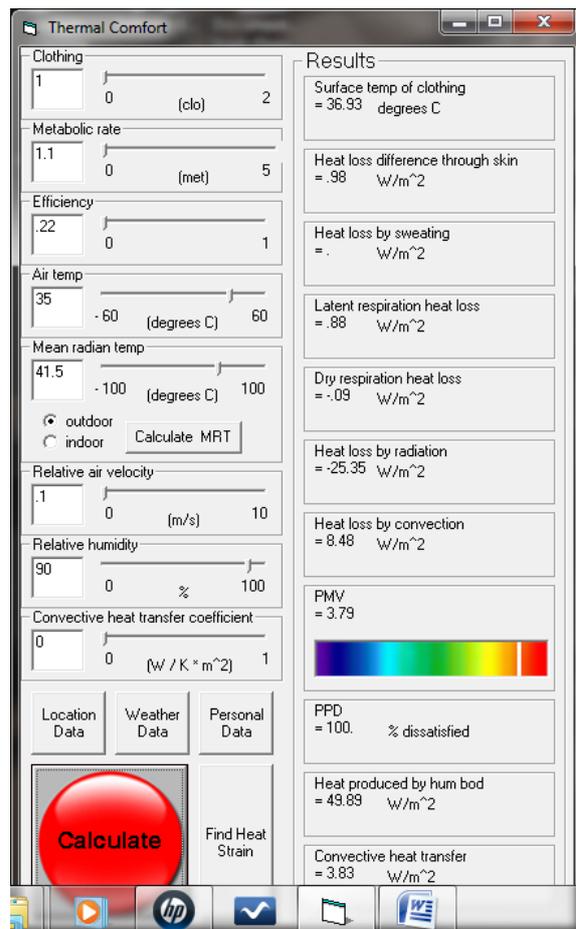
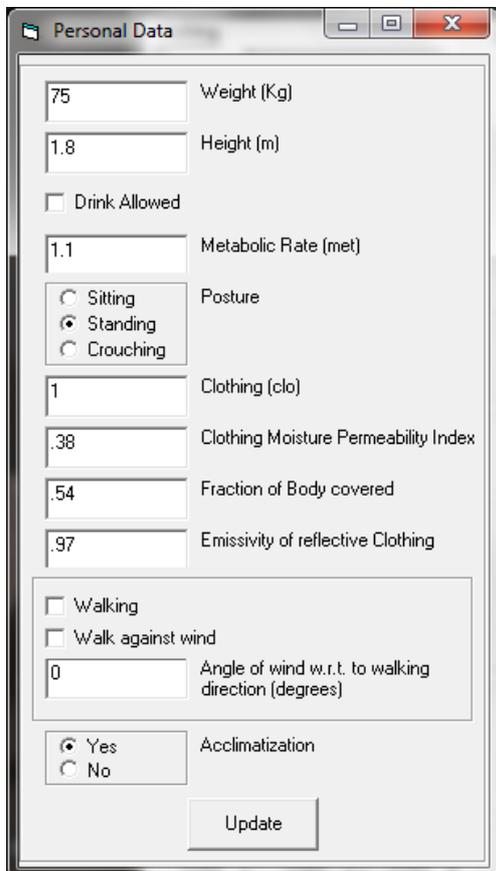
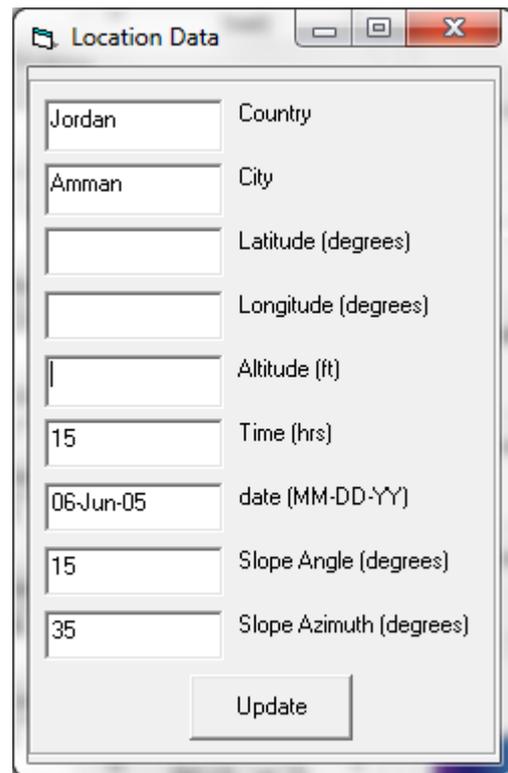
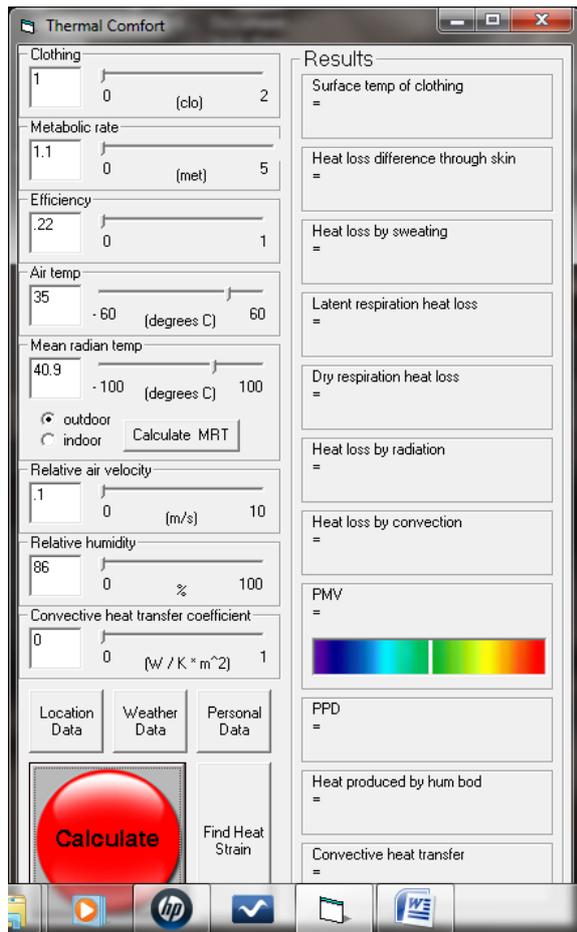
Subject Personal Data Box

Height, Weight

Kind of clothing he is wearing
 % of his body surface not covered
 Type of activity and metabolic rate
 Motion velocity
 Posture

D. Results from actual Runs of the code.





An Experiment of Chocolate Softness Measurements Using a New Design Machine

B. A. Al-Helou*^a

^aFaculty of Engineering Technology- Zarqa University, P.O. Box 32222 Zarqa 13132, Jordan

Abstract

In principle, the study is based on the objective of upgrading chocolate softness and homogeneity by designing a new machine that would be alternative to the costly current one. The eventual target is to elevate the present level of chocolate delicacy to match with that in the developed countries.

The findings of the study are satisfactory; the mixing period in the factory machines has been reduced from 24 to 12 hours in the newly designed machine, while chocolate softness is lowered from 55 to 25 micron. Hence the chocolate conformity and cohesion.

Coarse sugar, which was used in chocolate mix in this experiment, is now substituted by powdered sugar. Eventual results demonstrate ample saving in energy consumption and reduce the cost production due to the reduced mixing time, in addition to improving the quality product.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: Chocolate softness; Primary mixer; Weiner; Grinding; Marble; Mixing time; Viscosity

1. Introduction

Chocolate is, in fact, a matrix of discrete solid particles (sugar, cocoa and milk) set in a continuous fat medium. This fat makes up about a third of the weight of the chocolate, but because it has a lower density, it occupies almost half the volume. When the fat is solid, the chocolate is harder to bite through, although at higher temperatures it becomes softer and eventually turns into a liquid. This is what happens when chocolate is eaten, as the melting temperature of the fat is below that of the mouth [1,2].

The fat in milk chocolate is a mixture of cocoa butter (derived from cocoa beans) and milk fat. The geographical origin of cocoa beans affects the hardness of the chocolate produced and the rate at which it melts, with the fat from beans grown nearer the equator being harder and slower to melt than those grown further from the equator [3-5].

Chocolate is highly affected by temperature, a factor that has to be reckoned with before and during manufacturing, and in stockpiling and distribution as well.

Chocolate manufacturing greatly depends on how well its components of cocoa mass, cocoa butter, sugar and milk are mixed [6]. At the Philadelphia factories, this process takes about 24 hours daily, and carried out in two 500kg mixers (primary mixers) (Fig.1). The temperature generated by this process should not exceed 45°C. Otherwise, the control unit would stop off the mixer and thus save the chocolate mass.

The chocolate would then be carried by double jacket pipes, at 45°C, to two storage tanks, the capacity of each is 1000 kg. These latter two are also double jacket tanks with

40-45°C though preferably to be at 42°C. After that, chocolate would be carried to the Tempering Unit, where its temperature drops to 29°C, then to the Depositing Unit, which is connected with the control unit, with a view to determining the quantities required for poured in the molds. By that time, those molds would have been exposed to preheating. Indeed, the temperature of the mold should be very close to that of the chocolate so that chocolate breaking would be avoided when being poured in molds. Then the molds would be carried mechanically to the Vibration Unit, where chocolate is conformed inside the molds. After that, it would be carried to the three-phase Refrigeration unit, then for packing, and finally to the refrigerated stores.

2. Advanced Mixers (Five-Roll Refiner)

The five-roll refiner (Figure2) has five barrel-shaped rollers, which are normally between 800 and 2500 mm wide and are approximately 400 mm in diameter. These rolls become parallel under the pressure of operation. The particles are broken by the shearing action between the two counter-rotating rolls. The gap between them becomes narrower until the top gap is the same size as the largest particles within the chocolate, i.e. less than 30µm (micron).

*Corresponding Author. Email: heloub@hotmail.com

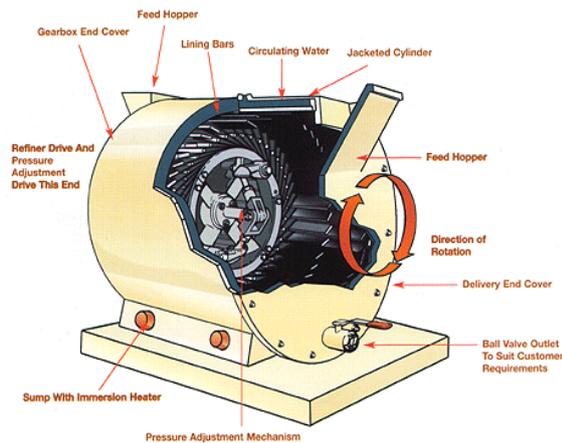


Figure 1: Shows the Primary Mixer (Pre-mixer) [7].

The material, with a maximum particle size of around $150\mu\text{m}$, is fed in between the two lower rollers and is transferred up the 'stack' due to the increased speed of the next roller. The film of material becomes thinner by a factor related to the speed difference. The ratio of roll speeds in fact is one of the ways of controlling the particle size of the machine. This is normally between 1.5 and 1.7 with the roll speeds varying from less than 50 r/min to greater than 400 r/min. In some five-roll refiners the rolls have set speeds and in this case the maximum particle size is manipulated by adjusting the feed gap between the first two rolls. The greater this feed gap, the larger will be the particles between the top two rolls. The pressure between the rolls does not control the particle size but is just used to ensure a uniform coating of the rollers by the chocolate. The Bühler Company [8] in Switzerland has produced an automated feedback system, in which a thickness-measuring device on the final roll gives a measurement which adjusts the relative speeds of the first two rollers. The viscosity of the chocolate is critical, however, and, if there is insufficient fat to bind the solid particles into a film, the material may be thrown from the rollers. On the other hand, if the feed material is very thin, because too much fat is present, some of the sugar and milk particles will segregate out in the hopper, resulting in uneven grinding. Temperature has a very large effect upon the flow properties of this fat and so accurate cooling and heating of the rolls is required to obtain an optimum particle size distribution within the chocolate [1].

As shown above, mixing chocolate components is the important section of the production processes as a whole. Chocolate quality depends a lot on the mixing process which takes quite a long time.

Good chocolate is distinguished by taste. Like cream, it should melt in the mouth [9]. Hence to get a better chocolate taste with an elevated delicacy, the mixing process should be developed. In order to obtain the correct flavor and texture, solid particles must be ground to a size smaller than about $30\mu\text{m}$, which use a highly specialized and costly machine for that purpose [9,10]. However, the mass production of these machines is not fit to the needs of the local market, to say nothing about their costly maintenance that is only compatible to developed countries. In these machines, chocolate components are mixed by a group of moving rolls with very high level of softness. So the developing process we seek to achieve is to design an alternative machine at a low cost.



Figure 2: Shows an Advance Machine, Five-Roll Refine [1].

3. The Newly Designed Machine (Wiener):

The study is based on milling the chocolate components while being passed and revolved on chrome marbles. These components would be softened by the movement of the weighty marbles which are operated by an electric motor that is tied to a gearbox. The design incorporates the manufacture of a double jacket tin made up of a stainless steel material with an inner wall thickness no less than 6mm and 5mm for the outer wall, whereas the space between the two walls will be about 8cm, and the capacity of the machine will be about 250kg of chocolate. Argon gas will be used in the welding so that chocolate would not be affected by any oxidation.

In the middle of the tinplate, an internally moving bar of stainless steel will be fixed, with 24 stainless brushes to do the mixing. The bar and brushes will be operated by the above mentioned electric motor and the gearbox, which will be fixed on a special-made basis above the upper lid of the tinplate.

At the bottom of the tinplate, chrome marbles of different or same diameter (pending availability in the marketplace) weighing to help in the mixing and milling process. However, the marbles weight should gradually be increased to reach the proper level of chocolate mixing and softness.

As the chocolate mix is well prepared, and its conformity achieved in the primary mixers of 500kg weight, a special pump at the lower part of the mixer will pump it to the upper part of the wiener (Fig.3). Meanwhile, the chocolate mix at the lower part of the wiener will also be drawn by a special pump and pumped to the upper part of the primary mixer. Thus a continued revolving of the mix will run between the primary mixer and the wiener to realize the chocolate's required delicacy.

A special control plate will run the chocolate stirring and pumping, and the cooling of the manufactured system will be done by a special cooling tower.



Figure 3: Shows the Newly Designed Machine (Wiener).

4. Means of Measurement

When chocolate conformity is realized (which usually takes one hour in this factory), we start measuring its softness once every half hour by a gauge device (Figure 4). A chocolate sample will be taken by a small-spoon. The grading surface device will be varnished by this sample. The grading device involves a straight line apertures, the highest grade of which is 185 micron and the lowest is 1 micron. So if the sample is immersed in an aperture, then it is the indication that the sample's grading is identical to that of the aperture. But if the sample is larger than 185 micron, then the chocolate softness can by no means be read. In other words, this softness is out of the grading reach.

As the factory used the coarse sugar in making chocolate, a comparison has been made on the primary mixer using two separate batches with coarse sugar and powder sugar. Figure (5) explains the curves of softening the chocolate by the coarse sugar and the powder sugar. From the curve of the coarse sugar, it is noted that the mix began gradation after 5.5 hours, and reached the highest delicacy level after 24 hours of mixing, while it has recorded 55 micron only, whereas the powder sugar curve showed that the mix began gradation after 2.5 hours, and the maximum chocolate delicacy was realized after 13 hour mixing. This shows a time saving of 11 mixing hours, with a 40 micron delicacy after 22 hour mixing; an apparently better delicacy level achieved by the powder sugar than that by the coarse sugar.

That is a primary indication for dispensing with coarse sugar. To make sure of that, an experiment was carried on the wiener with powder sugar after having added 210 kg of chrome marbles with a view to facilitating the mixing and milling process.

(Figure 6), which includes the curve mentioned in (Figure5), i.e. the powder sugar curve used in the primary mixer and the other curve pertaining to the wiener. Hence, it can be noted that the wiener performance by using powder sugar was better than of the primary mixer when using the same sugar. In fact, the required softness, i.e. 25

micron has been achieved after 24 hours of mixing 250 kg of raw chocolate material by the wiener. Given that, and to obtain better results, the mixing method in the factory has been changed. Coarse sugar has been replaced by powder sugar on the one hand, and the primary mixer and the wiener have been connected together so that the softening process will be realized between them by circulation on the other.

(Figure 7) shows the softness that has been achieved by this experiment of circulation, where the chrome marble maintained its weight, i.e. 210 kg, and the way of mixing was the only thing that has been changed. The figure also shows that the blend started to come into the grading scale only after 3-hours of mixing, and that the required softness, i.e. 25 micron has been achieved after 16.5 mixing hours. This means that the wiener works positively, and that the blend works positively, too.

Given all that, the marbles weight in the wiener by 10 kg each time up to 260kg as shown in Figure (8), where weights are 220, 230, 240, 250, 260 kg respectively. In all cases, gradation started on hour after the mixing process, excluding the case of 260kg when gradation began immediately.



Figure 4: Shows the Chocolate's Gage Device

It is noted that the curves of 240, 250, and 260 kg weights began coming closer to each other after 11 hour mixing. They registered 35, 34, and 30 micron respectively, and that the required softness of the chocolate started to be achieved by the 12 hour of mixing time, with 28, 27, and 25 micron respectively. The curve also shows that 25 micron required delicacy took 16 hours for a 220kg weight, 14.5 hours for 230kg, 14 hours for 250 kg, and 12 hours for 260 kg weights. Eventually, it can be safely said that the required chocolate softness has been achieved in a 12 hour mixing period for 500 kg of chocolate revolved between the primary mixer and the wiener.

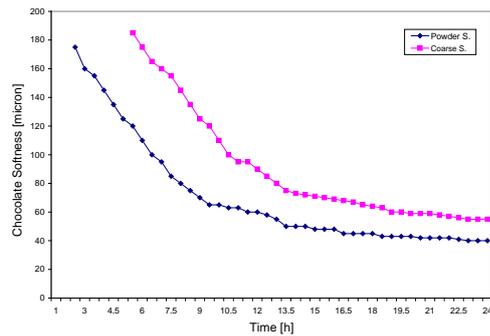


Figure 5: Shows the Chocolate's Softness by the Primary Mixer Using Coarse and Powder Sugar.

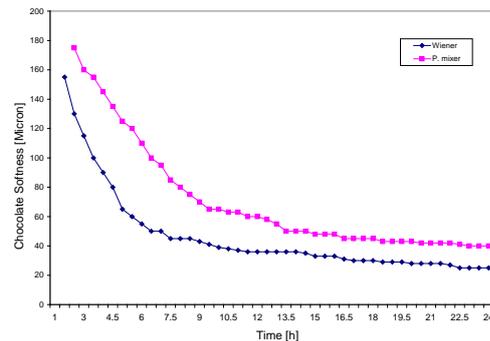


Figure 6: Shows the Chocolate's Softness by the Primary Mixer and the Wiener Using Powder Sugar.

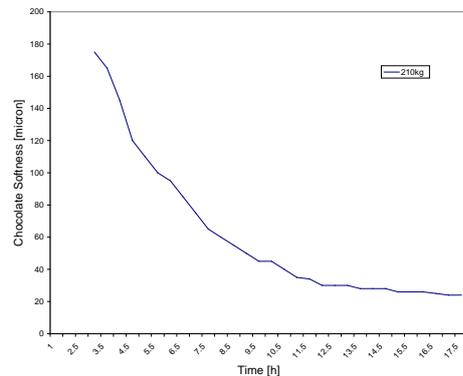


Figure 7: Shows the Chocolate's Softness by the Circulation Method between Primary Mixer and Wiener Using 210kg of Chromic Marbles.

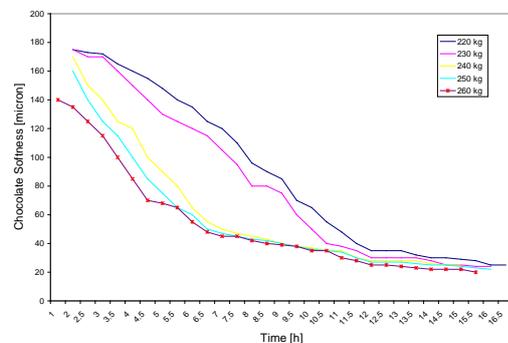


Figure 8: Shows the Chocolate's Softness by the Circulation Method Between Primary Mixer and Wiener Using (220-260) kg of Chromic Marbles.

5. Conclusions

The obtained results are summarized as follows:

- The results of the design and manufacture of the wiener machine, and the use of circulation method to have led the satisfactory results by reducing the chocolate softness from 55 to 25 microns thus, improving the quality of the product.
- The curves show that 25 micron required delicacy took 16 hours for a 220kg weight, 14.5 hours for 230kg, 14 hours for 250 kg, and 12 hours for 260 kg weights.
- The undertaken measurements show that the wiener has reduced the mixing time from 24 to 12 hours for 500kg chocolate (i.e. 50 percent of the time consumed by the primary mixer).
- The designing and manufacture of the wiener has lent coarse sugar dispensable and should be replaced by powdered sugar.
- The best softness that was realized by the primary mixer by using powder sugar was 40 micron after 22 hours of blending, whereas it was achieved after 9.5 hours by the wiener.
- The improved chocolate softness, has simultaneously improved the viscosity and homogeneity of the mixture, and accordingly the chocolate taste much better than before.

6. Acknowledgment

The author would like to thank the Higher Council of Science and Technology, The Fund of Supporting Scientific Research and Industry Development, for the financial support in conducting this study.

References

- [1] S. T. Beckett, "Is the taste of British milk chocolate different". International Journal of Dairy Technology, Vol. 56. No.3 August 2003.
- [2] K.E Kayligian, "Milk fat fractions in chocolate". Manufacturing Confectioner, 77, (1997), 79–84.
- [3] S. T. Beckett, "The Science of Chocolate. Cambridge". Royal Society of Chemistry, (2000).
- [4] S. T. Beckett, "Industrial Chocolate Manufacture and Use". 3rd edition, (Blackwell Science, Oxford), 1999.
- [5] Cocoa and Chocolate Manual, (Lobas, Ede, The Netherlands) 1997.
- [6] M. Bitter, "Chocolate plant 'refines' efficiency of grinding". Candy Industry, Vol.158 No.9, Sep 1993, 50-54.
- [7] http://www.macintyre.co.uk/a3_500kg_chocolate_refiner.htm.
- [8] Manufacturer's Catalogue, Bühler, AG, CH-9240 Uzwill, Switzerland. 6 Sweet Facts Nestle, (Nestle UK Limited, York) 1998.
- [9] S.T. Beckett, "Milling, mixing and tempering—an engineering view of chocolate". Institution of Mechanical Engineers, Journal of Process Mechanical Engineering, Vol. 215 Part E, 2001, 1-8,
- [10] S.O. Hansen, P.S. Hansen, "Spray dried whole milk powder for the manufacture of milk chocolate". Scandinavian, Dairy Information, 2, 1990, 79–82,

Influence of Camera's Optical Axis Non-perpendicularity on Measurement Accuracy of Two-dimensional Digital Image Correlation

A. Hijazi^{a,*}, A. Friedl^b and C. J. Kähler^b

^aDepartment of Mechanical Engineering, The Hashemite University, Zarqa, Jordan.

^bInstitute of Fluid Mechanics and Aerodynamics, Universität der Bundeswehr München, Neubiberg, Germany.

Abstract

The two-dimensional digital image correlation technique is very commonly used in a wide variety of solid mechanics applications for measuring in-plane deformations of planar surfaces. The perpendicularity of the camera's optical axis to the surface being observed is one of the basic conditions for the validity of the measurement. Small magnitudes of camera misalignment angles, up to two or three degrees, can go easily unnoticed during the initial setting of the experimental setup especially when the stand-off-distance between the camera and the surface is not small. In the work presented in this paper we investigate the errors in strain measurements caused by the non-perpendicularity of the camera's optical axis, with respect to the surface being observed, both theoretically and experimentally. In-plane rigid-body-translations in the directions perpendicular and parallel to the camera tilt axis are used for estimating the resulting strain errors. Results show that the non-perpendicularity of the camera causes errors in both the normal and shear strains. Misalignment angles as small as 2° are found to cause strain error greater than 10³ μ-strains. The magnitude of strain error is found to increase linearly with both the misalignment angle and the magnitude of in-plane translation while it is inversely related to the stand-off-distance. The analyses show that simple in-plane rigid-body-translations experiments can be used to estimate the resulting strain errors and more importantly to detect and hence correct any existing non-perpendicularity between the camera's optical axis and the target surface. Experiments show that misalignment angles smaller than 1° can be detected using in-plane rigid-body-translations.

© 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved

Keywords: Digital image correlation; strain error; error analysis; camera non-perpendicularity; camera misalignment; in-plane translation

1. Introduction

Nowadays, digital image correlation (DIC), also referred to as the digital speckle correlation method (DSCM), has become one of the most widely used full-field optical methods for motion and deformation measurements. The DIC method was first introduced by Sutton et al. [1] in the early 1980s and during the past three decades it underwent continuous modifications and significant improvements [2, 3]. In its simpler version, using a single camera, DIC is used for two-dimensional in-plane measurements (2D-DIC). Also, photogrammetric three-dimensional measurements (3D-DIC) [4] can be made using two cameras in stereo configuration. Besides the good measurement accuracy of the DIC method, it also offers other attractive features which include; relatively simple experimental setup, simple or no specimen preparation and low requirements for the measurement environment. All of that have made the DIC method extremely popular among the experimental mechanics community, and both 2D-DIC and 3D-DIC are being increasingly used in a very wide range of applications ranging from material science to manufacturing, mechanical, biomedical and structural engineering [5, 6].

The basic idea of DIC is to compare two digital images acquired at different states (e. g., one before deformation and the other one after), of a surface having a random speckle pattern, by dividing the image into subsets of several pixels then mathematically matching those subsets between the two images (based on intensity levels) to determine the new location of each subset in the second image. From that, the full-field deformation map can be obtained and the strain map can then be easily determined. In principle, 2D-DIC can be used for deformation measurements under three conditions; the specimen has a planar surface, it undergoes in-plane deformations and the camera's optical axis is perpendicular to the specimen surface. If any of these three conditions is not reasonably satisfied, the accuracy of the measurements will be compromised. In one of the key papers addressing the applications of DIC in experimental mechanics, Chu et al. [7] used in-plane rigid-body-translations to demonstrate the viability of the DIC method for actual measurements. When a body undergoes a rigid-body-translation, the measured strains should theoretically be zero and thus any obtained strain readings reflect the magnitude of error in the measurement. This simple experiment remains to be one of the most widely used experiments for estimating the magnitude of background strain error expected in DIC strain measurements. The measurement accuracy of 2D-DIC depends on several factors which include; a) the

*Corresponding author: hijazi@hu.edu.jo

speckle pattern, b) quality and perfection of the imaging system (distortions, noise, resolution, etc.) and c) the selection of the correlation algorithm and parameters (subset and step size, correlation and shape functions, sub-pixel algorithm, etc.) [3, 5]. Numerous studies have investigated the different sources of errors and tried to estimate the resulting errors and to suggest remedies in some cases. Listings of the different studies can be found in references [3, 5, 8, 9].

As mentioned earlier, the in-plane deformations and the perpendicularity of the viewing axis (i.e., the parallelism of the camera focal-plane-array with the specimen surface) are two of the basic assumptions for 2D-DIC. The occurrence of any out-of-plane translation or misalignment between the camera focal-plane-array and the specimen surface will lead to errors that are embedded in the measured displacements and strains. Out-of-plane translations may occur during the loading of the specimen, while the non-perpendicularity of the camera, on the other hand, may occur during loading (due to out-of-plane rotation of the specimen) or may exist before loading. Some recent studies [9-12] have investigated the errors caused by out-of-plane translations and/or non-perpendicular camera alignment. Haddadi et al. [9] did an experimental investigation where they used rigid-body-translations to investigate the different sources of error in 2D-DIC measurements and to estimate these errors. In that study, they estimated the strain errors induced by out-of-plane translations where they noted that; a) the error is linearly proportional to the amount out-of-plane displacement and b) for a fixed amount of out-of-plane displacement, the error decreases as the stand-off-distance between the camera and surface increases. Sutton et al. [10] studied the effects of out-of-plane displacements and rotations (that may occur during the loading) both theoretically and experimentally. Their results show that the strain error is proportional to the ratio $\Delta Z/Z$ (where ΔZ is the out-of-plane translation and Z is the stand-off-distance). They also showed that the use of telecentric lenses minimizes the error to a manageable level. Meng et al. [11] theoretically investigated the displacement errors caused by the camera non-perpendicularity and reported that the errors are linearly related to the displacement and non-perpendicularity when these two are small. Based on numerical analysis, they reported that measuring sensitivity of 0.01 pixels can be attained under misalignment angles up to 5 degrees. Lava et al. [12] studied the strain error caused by non-perpendicular camera alignment using numerically rotated images having an imposed finite element displacement field. They compared the strain error for a sample subjected to large plastic strain using 2D-DIC and 3D-DIC and proposed a method for rectifying the distorted image, resulting from camera non-perpendicularity, for 2D-DIC measurements. The proposed image rectification method helps in eliminating the error resulting from camera non-perpendicularity (e.g., when physical constraints prevent the camera from being oriented perpendicular to the surface) if the misalignment angle is known.

In this paper, we investigate the errors introduced in the measured strains resulting from camera non-perpendicularity in 2D-DIC measurements both theoretically and experimentally. We also demonstrate the effect of the misalignment angle, in-plane translation and camera stand-off-distance on strain error. We show that simple in-plane rigid-body-translation experiments can be used to estimate the resulting strain errors and more

importantly to detect, and hence correct, any existing non-perpendicularity between the camera's optical axis and the target surface. The development of the theoretical equations for estimating the strain errors (due to camera non-perpendicularity) associated with in-plane rigid-body-translations is presented in section 2. Section 3 presents the details of the experimental setup and the procedure used for measuring the DIC strain error at different camera orientation angles. The results of the theoretical and experimental investigation are presented in section 4 and the results are discussed in section 5. Finally, section 6 presents the concluding remarks.

2. Effect of Camera Non-perpendicularity on Strain Measurements

The perpendicularity of the camera's optical axis to the specimen surface is one of the three conditions, mentioned previously, for the validity of 2D-DIC measurements. When the camera is non-perpendicular to the specimen surface, in-plane translations of the specimen surface will not be faithfully replicated at the image plane. Consequently, DIC will show that the surface was deformed (strains not equal to zero) even if the surface underwent a pure rigid-body-translation. The resulting strain errors depend on the direction of in-plane translation relative to the camera tilt axis. If the translation is in the direction perpendicular to the tilt axis, it will change the actual stand-off-distance between the camera and all points on that surface and that in turn will change the magnification in both the x and y directions. Thus, DIC will show non-zero normal strain readings in both the x and y directions. On the other hand, if the translation is in the direction parallel to the tilt axis, points at different distances from the tilt axis will show different amounts of translation since they have different stand-off-distance from the camera. Therefore, DIC will show non-zero shear strain. In order to estimate the amount of strain error expected in DIC measurements, simple pinhole camera model can be used. According to pinhole imaging formulation, the imaged length, l^i , of an object of length l that is parallel to the image plane can be found as:

$$l^i = \frac{l}{D} f \quad (1)$$

where D is the distance between the camera and the object (i.e., the stand-off-distance) and f is the distance between the pinhole plane and the image plane (i.e., the focal distance). In geometric optics, f usually refers to the focal length of the lens which is slightly smaller than the distance to the image plane. However, it is reasonable to assume that f used in the pinhole camera formulation represents the focal length of a real lens.

Though DIC software packages calculate the strains based on the large strains theory (Green-Lagrange strain), for simplicity, small strains theory is used here for deriving the theoretical strain expressions. The obtained strain expressions are numerically evaluated using parameters matching those of the experimental setup. Since the expected strains are fairly small, the theoretically obtained strain error values are expected to be in good agreement with the experimental values obtained using DIC software.

2.1. In-plane Translation Perpendicular to the Tilt Axis

Figure 1 shows a schematic for a pinhole camera imaging a surface where the surface is tilted about the y axis by an angle θ . The surface rigidly translates along the x direction by a distance Δx and for clarity the positions of the surface before and after the translation are shown in two separate sketches as seen in the figure. To determine the amount of the induced “apparent” strain (i.e., strain error) when the surface rigidly moves from position (1) to position (2), two points on the surface (**A** and **B**) are considered. When the surface is at position (1), the x coordinates “at the image plane” for points **A** and **B** are found as:

$$(x_A^i)_1 = \frac{x_A \cos \theta}{S + x_A \sin \theta} f \tag{2}$$

$$(x_B^i)_1 = \frac{x_B \cos \theta}{S + x_B \sin \theta} f \tag{3}$$

where S is the nominal stand-off-distance between the camera and the surface and x_A & x_B are the coordinates of the two points “at position (1)”. The length of line **AB** at the image plane is simply found as:

$$(l_{AB}^i)_1 = (x_A^i)_1 - (x_B^i)_1 \tag{4}$$

When the surface rigidly translates a distance Δx to position (2), the x coordinates of the two points at the image plane become:

$$(x_A^i)_2 = \frac{(x_A - \Delta x) \cos \theta}{S + (x_A - \Delta x) \sin \theta} f \tag{5}$$

$$(x_B^i)_2 = \frac{(x_B - \Delta x) \cos \theta}{S + (x_B - \Delta x) \sin \theta} f \tag{6}$$

and the length of line **AB** at the image plane becomes:

$$(l_{AB}^i)_2 = (x_A^i)_2 - (x_B^i)_2 \tag{7}$$

Based on small strains theory, the average Cauchy strain in the x direction can be found as:

$$\varepsilon_{xx} = \frac{(l_{AB}^i)_2 - (l_{AB}^i)_1}{(l_{AB}^i)_1} \tag{8}$$

Similarly, an “apparent” strain is also expected in the y direction. To determine the strain in the y direction, a line (oriented along the y direction) of length l_y that is located at an arbitrary distance x is considered. When the surface is at position (1), the length of that line “at the image plane” can be found as:

$$(l_y^i)_1 = \frac{l_y}{S + x \sin \theta} f \tag{9}$$

When the surface rigidly translates a distance Δx to position (2), the length of that line at the image plane becomes:

$$(l_y^i)_2 = \frac{l_y}{S + (x - \Delta x) \sin \theta} f \tag{10}$$

and the strain in the y direction can simply be found as:

$$\varepsilon_{yy} = \frac{(l_y^i)_2 - (l_y^i)_1}{(l_y^i)_1} \tag{11}$$

2.2. In-plane Translation Parallel to the Tilt Axis:

Figure 2 shows a schematic for a pinhole camera imaging a surface where the surface is tilted about the y axis by an angle θ (similar to what is shown in figure 1 but the view is rotated 90° about the camera axis). The surface rigidly translates along the y direction by a distance Δy and two points on the surface (**A** and **B**) are considered. For simplicity, points **A** and **B** are defined along the x axis at arbitrary distance from the origin. The positions of points **A** and **B** before and after the translation are shown in the figure. Before the translation, the y coordinates of the two points are equal to zero. After the surface is rigidly translated by distance Δy , the y coordinates “at the image plane” for points **A** and **B** become:

$$y_A^i = \frac{\Delta y}{S + x_A \sin \theta} f \tag{12}$$

$$y_B^i = \frac{\Delta y}{S + x_B \sin \theta} f \tag{13}$$

Therefore, the shear strain can simply be found as:

$$\varepsilon_{xy} = \frac{1}{2} \left(\frac{y_A^i - y_B^i}{x_A^i - x_B^i} \right) \tag{14}$$

where x_A^i and x_B^i (i.e., the x coordinates at the image plane for points **A** and **B**) are found using equations 2 and 3 given previously.

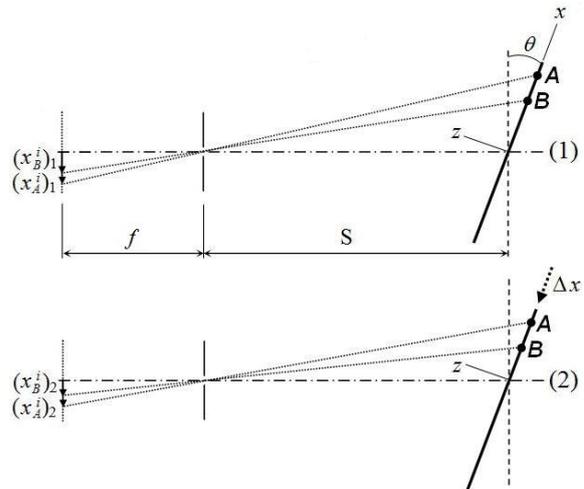


Figure 1: Non-perpendicular pinhole camera schematic; surface translating perpendicular to tilt axis.

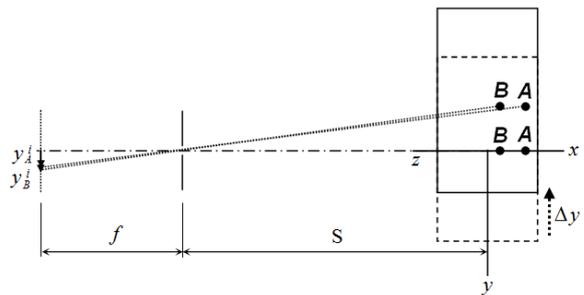


Figure 2: Non-perpendicular pinhole camera schematic (view rotated 90° about the camera axis); surface translating parallel to tilt axis.

3. Experiments

In order to experimentally quantify the DIC strain error resulting from camera non-perpendicularity, different groups of experiments were conducted where in the first group the camera was perfectly perpendicular to the surface and in each of the other groups the camera was tilted with a different angle (from 1° to 5° in 1° steps). A schematic view of the experimental setup is shown in figure 3. As can be seen from the figure, the camera was tilted around its own axis which is parallel to the y axis. In each group of experiments, while the camera is at a fixed angle with respect to the surface, the surface was translated in two steps along the x direction (i.e., perpendicular to the tilt axis) then in two steps along the y direction (i.e., parallel to the tilt axis). As a result, images were captured while the surface is at five different positions; a reference position, two translation steps in the x direction and two translation steps in the y direction.

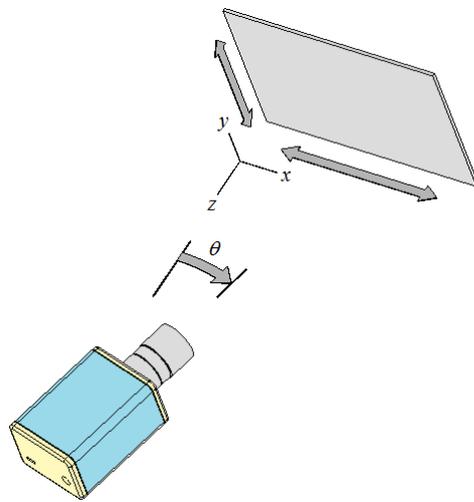


Figure 3: Schematic view for the setup used in the experiments.

The vision system used in this study consists of; a) DALSA Genie-M1410 camera having a 2/3 inch monochrome interline CCD chip with 1360×1024 pixel resolution, b) Pentax C5028-M machine vision lens having 50mm focal length, c) a 2mm extension ring between the lens and the camera to reduce the minimum working distance of the lens. The aperture of the lens was set to $f/8$ to obtain a reasonable depth of field in order to accommodate the depth gradient resulting from camera non-perpendicularity. The camera was mounted on a rotating-stage to allow orienting the camera at the desired tilt angles. A printed black and white random speckle pattern was attached to the target surface which is mounted on a multi-axis translating stage. The distance between the front end of the lens and the target surface was set to about 680mm and at that working distance the field-of-view observed by the camera was about a 113mm wide (12 pixels/mm scale factor). At the magnification level provided by the vision system, the average diameter for the black dots of the speckle pattern was about 6 pixels and the average dot center-to-center spacing was about 13 pixels. Each of the two translation steps, in both the x and y directions, was equal to 8mm (about 7% of the field-of-view width). In the first group of experiments (i.e., camera perpendicular to the surface), the camera was first placed

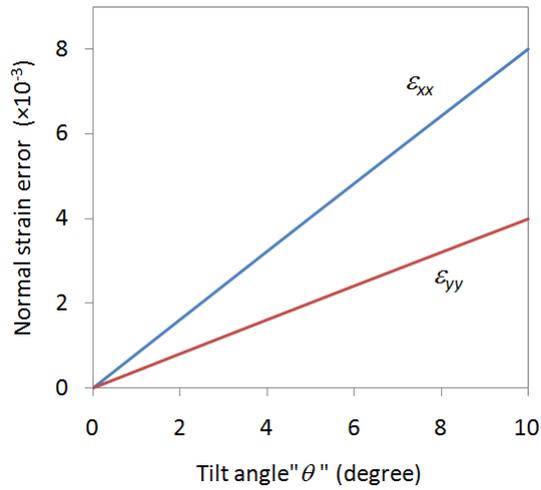
close to the surface and a right-angle triangle was used to verify its perpendicularity then it was translated backwards with the aid of an optical rail to the required working distance. In each of the preceding groups of experiments, the camera was tilted by 1° using the rotating stage and the same experiments were repeated.

For each group of experiments corresponding to a different camera orientation angle, the reference position image was correlated with the images corresponding each of the two translation steps in each of the two directions. Commercial 2D-DIC software called “VIC-2D” [13] was used for the analyses and the same analyses were also repeated (with identical parameters) using other software called “MatchID-2D” [14] for further verification of the results. A subset size of 41×41 pixels with a step size of 20 pixels was used for the correlation. The Green-Lagrange strains were calculated using a strain window size of 7×7 points. The entire area of the reference image was used for the correlation, rather than choosing a partial region of interest, such that a more reliable averages of the resulting strains can be obtained.

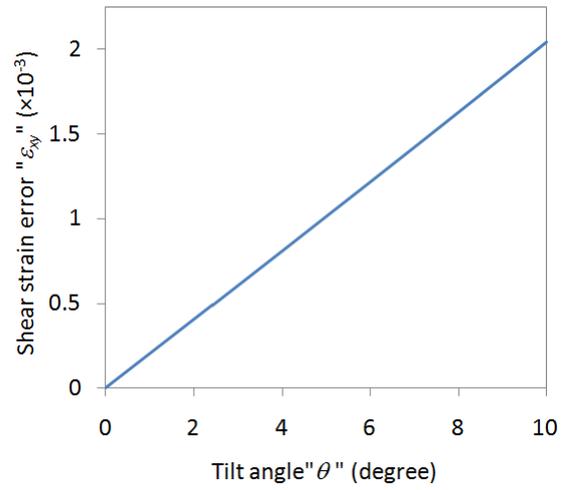
4. Results

4.1. Theoretical Results:

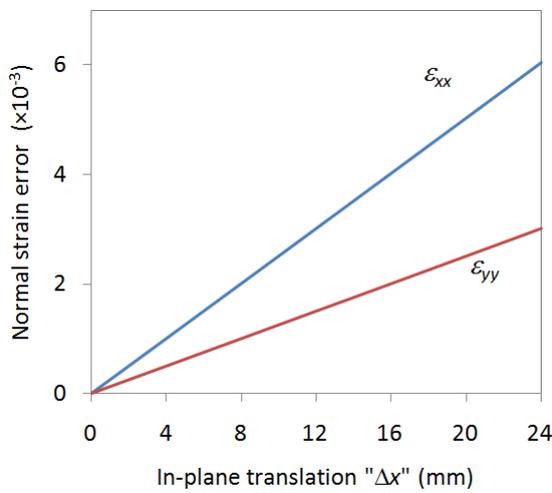
The normal strains (ϵ_{xx} and ϵ_{yy}) error resulting from in-plane translation perpendicular to the tilt axis was obtained by solving equations 8 and 11 numerically using parameters similar to those of the experimental setup. It should be noted that some of the parameters involved in equations 8 and 11 has very minor or no effect on the obtained strain values. For instance, by substituting equations 2 through 7 into equation 8, the focal length f will cancel out from the equation and the same will happen in equation 11. Also, other parameters were found to have very insignificant influence on the results. However, values have to be assigned to these parameters for the numerical evaluation. The values of these parameters used in the evaluation are; $x_A = 1$ mm, $x_B = 0$ mm, $l_y = 1$ mm and $x = 1$ mm. On the contrary, the tilt angle θ , the magnitude of in-plane-translation Δx and the stand-off-distance S have a significant influence on the resulting normal strains. The influence of these three parameters was investigated by considering an appropriate range for each and the results are shown in Figure 4. As can be seen from the figure, the strain error in the x direction (i.e., the direction of translation) is larger than that in the y direction where its value is twice of that in the y direction. From figure 4 (a) and (b) it can be seen that the strain error is linearly related to both the tilt angle and in-plane translation, while the strain error has an inverse relation with the stand-off-distance as seen in figure 4 (c). The inverse relation between stand-off-distance and the strain error is consistent with the findings reported previously in literature for the case of out-of-plane translation [9, 10]. This similarity is rather expected because, as mentioned earlier, in-plane translations occurring while the camera is non-perpendicular to the surface change the actual stand-off-distance between the camera and all points on that surface. It should be mentioned that in the theoretical analyses, the stand-off-distance S was set to be 690mm which is 10mm more than the actual distance between the front end of the lens and the target surface. The reason for adding this 10mm was to account for the distance from the front end of the lens to its optical center.



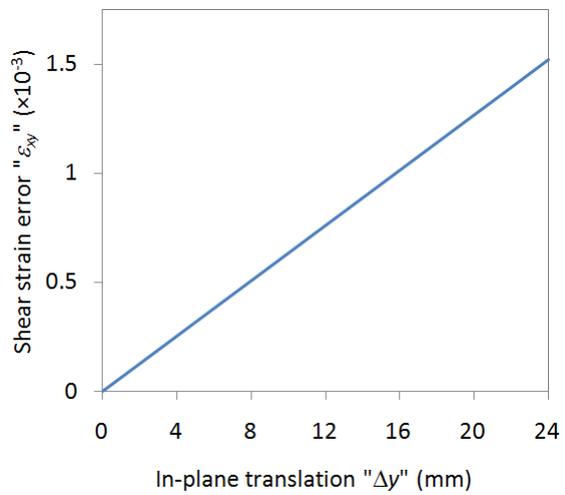
(a)



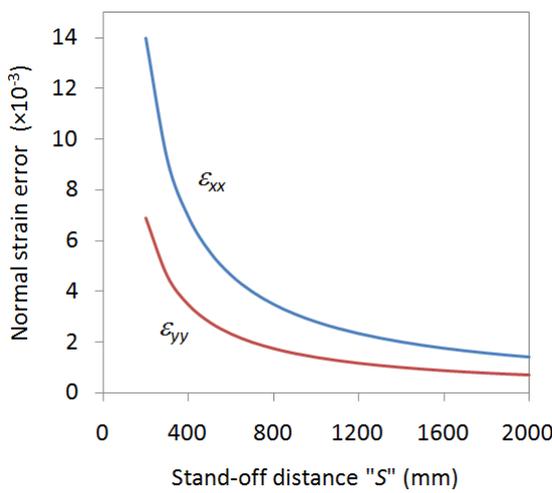
(a)



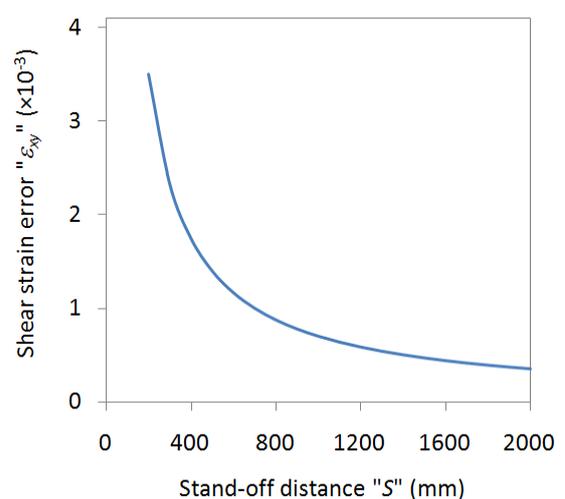
(b)



(b)



(c)



(c)

Figure 4: Normal strain error resulting from rigid-body in-plane translation in the direction perpendicular to the tilt axis; **a)** influence of tilt angle ($S = 690 \text{ mm}$ & $\Delta x = 16 \text{ mm}$), **b)** influence of the magnitude of translation ($S = 690 \text{ mm}$ & $\theta = 5^\circ$), **c)** influence of stand-off-distance ($\Delta x = 16 \text{ mm}$ & $\theta = 5^\circ$).

Figure 5: Shear strain error resulting from rigid-body in-plane translation in the direction parallel to the tilt axis; **a)** influence of tilt angle ($S = 690 \text{ mm}$ & $\Delta y = 16 \text{ mm}$), **b)** influence of the magnitude of translation ($S = 690 \text{ mm}$ & $\theta = 5^\circ$), **c)** influence of stand-off-distance ($\Delta y = 16 \text{ mm}$ & $\theta = 5^\circ$).

Similarly, the shear strain error resulting from in-plane translation parallel to the tilt axis was obtained by solving equation 14 numerically using parameters similar to those of the experimental setup. Again, by substituting in the equation, the focal length f cancels out and the assigned values for the other parameters are; $x_A = 1$ mm and $x_B = 0$ mm. The influence of the tilt angle θ , the magnitude of in-plane-translation Δy and the stand-off-distance S on shear strain error are shown in figure 5. From the figure it can be seen that the influence of these three parameters on the shear strain error is similar to that on the normal strains error seen previously in figure 4. By closely inspecting the values of shear strain error, it can be seen that its magnitude is one fourth of that of the normal strain ϵ_{xx} error resulting from in-plane translation perpendicular to the tilt axis.

It should be mentioned here that though the focal length f cancels out from the equations used for calculating the strain (eqns. 8, 11 and 14), the strain error depends of the focal length of the lens being used. This is simply due to the fact that in any experiment, certain magnification factor is required according to the size of specimen being observed. Therefore, to increase the stand-off-distance while maintaining the magnification, a higher focal length lens needs to be used. Practically, this means that the focal length have the same influence on strain error as the stand-off-distance.

4.2. DIC Results:

As mentioned previously, at each value of the tilt angle four correlations were carried out. In each correlation, the same reference image was correlated with one of the four translated images corresponding to in-plane translations in the x or y directions with a translation value of 8mm or 16 mm. Figure 6 shows an exemplary DIC output for 16mm translation in the x direction (i.e., the direction perpendicular to the tilt axis) where the camera was tilted by 5 degrees. From 6 (a) it can be seen that the results shows a clear gradient along the x direction of more than four pixels in the value of the horizontal translation (U) though the surface was rigidly translated. A similar gradient but of smaller magnitude and different direction is also observed in the vertical translation (V) values, as can be seen in 6 (b), though the surface was not translated in the vertical direction. It can be seen from the figure that these gradients are slightly rotated where that is most likely due to a very small amount of in-plane rotation that happened during the translation. These gradients in the translation values give indication to the magnitude of the error in measured displacements resulting from camera non-perpendicularity where, for instance, the four pixels difference in the horizontal displacement value simply represents the magnitude of displacement error (over the entire field-of-view) in the measurement. The presence of such gradients in the translation values indicates that the

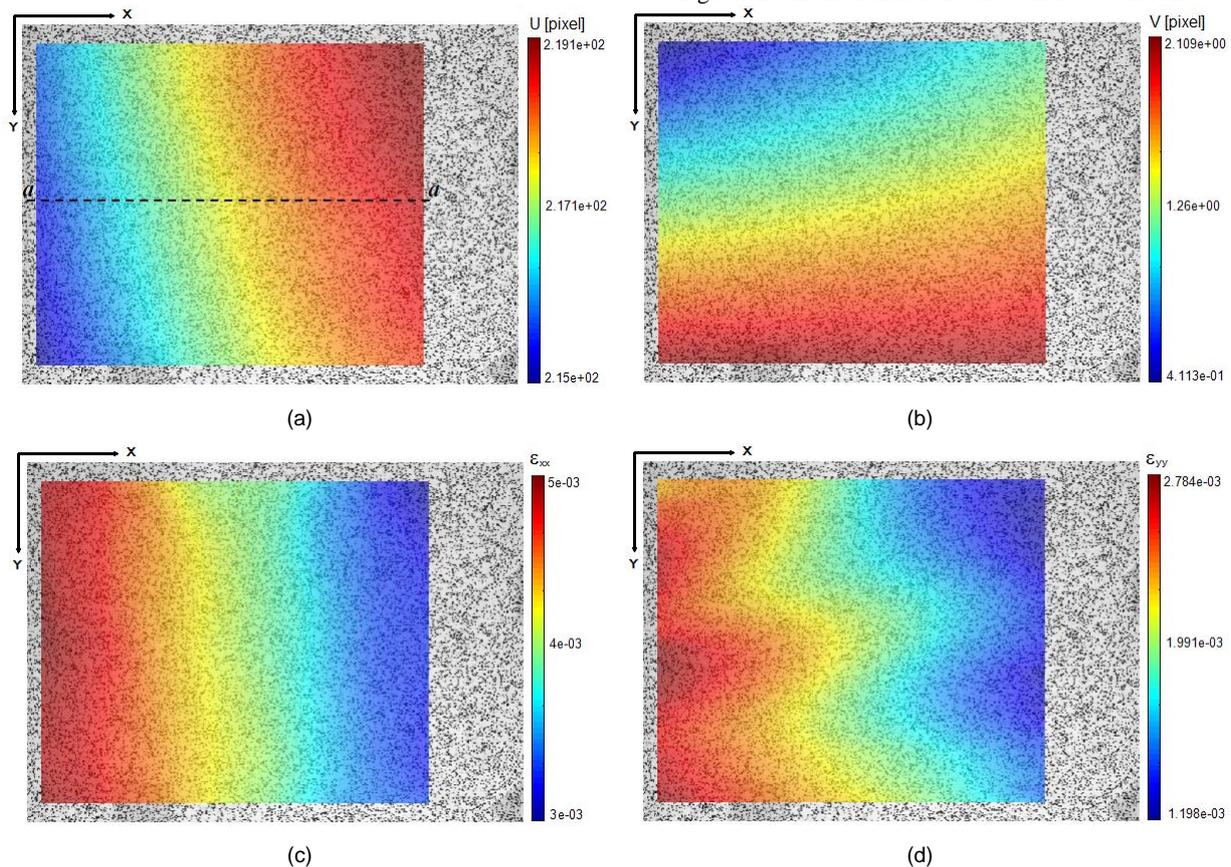


Figure 6: Typical DIC results for in-plane rigid-body-translation in the direction perpendicular to the tilt axis ($\theta = 5^\circ$, $\Delta x = 16$ mm); **a**) measured horizontal displacement, **b**) measured vertical displacement, **c**) measured normal strain in x direction, **d**) measured normal strain in y direction.

Strains in both directions have non-zero values (the values of these strains represent the strain error in the measurement). The resulting normal strain errors in the x and y directions can be seen in 6 (c) and (d). From the figure it can be seen that the strain in the x direction is larger than that in the y direction where that is consistent with the theoretical results presented earlier which shows that ε_{xx} is twice of ε_{yy} . The figure also shows a large gradient along the x direction in the values of both ε_{xx} and ε_{yy} . This gradient in the strain value, ε_{xx} for instance, indicates that the horizontal translation (U) is not linearly related to the x position as will be discussed in the next section.

Similarly, figure 7 shows DIC results for 16mm translation in the y direction (i.e., the direction parallel to the tilt axis) with 5 degrees camera tilt angle. From 7 (a) it can be seen that the horizontal translation (U) is fairly uniform (the slight non uniformity seen in the figure is most likely due to the lens optical distortions) and has a very small value. However, on the contrary, the vertical displacement (V) shows a clear gradient along the x direction, as can be seen in 7 (b), which indicates a non-zero shear strain. The resulting shear strain error is shown in 7 (c) where it can be seen that it has a large gradient along the x direction as well.

5. Discussion

The experimental results presented in the previous section confirms the theoretical conclusion that in-plane translations perpendicular to the tilt axis cause normal strain error while translations parallel to the tilt axis cause shear strain error. However, as mentioned earlier, equations 8, 11 and 14 which are used for calculating the strain error values are dependent on three parameters only (the tilt angle, the magnitude of in-plane-translation and the stand-off-distance). This indicates that the strain error values calculated by these equations are uniform over the entire surface. While, on the other hand, the DIC results show an obvious gradient in the strain values along the x direction. In order to understand the reason behind this difference, the horizontal translation at the image plane when the surface translates a distance Δx along the x direction is calculated as:

$$U^i = (x^i)_2 - (x^i)_1 \quad (15)$$

where $(x^i)_1$ and $(x^i)_2$ are the x coordinates “at the image plane” of a point on the surface before and after the surface is rigidly translated. The values of $(x^i)_1$ and $(x^i)_2$ can be calculated using equations 2 and 5, respectively. By assigning an appropriate range of values for x_A and substituting the correct values for the other parameters, the horizontal displacement as a function of the x position “at the image plane” can be obtained. Figure 8 shows the experimental relation between the horizontal translation and the x position which is obtained along the horizontal line “a-a” taken in the middle of the image as seen in figure 6 (a). Figure 8 also shows the theoretical relation obtained from equation 15, using the same parameters of that experiment, after being scaled to be plotted in “pixel” units. From the figure it can be seen that the theoretical analysis gives a linear relation while the DIC yields a nonlinear relation, while both relations still give the same net value change of about four pixels over the entire field-of-view width. This difference between the theoretical and

experimental relations indicates that the simple pinhole camera model used in developing the theoretical equations does not accurately replicate the behavior of real lenses when depth gradients are involved. It can also be seen from the figure that the theoretical normal strain ε_{xx} will have a constant value that equals the derivative of the U with respect to x which is simply the slope of the straight line. On the other hand, when the derivative of the experimental curve is taken, that will give a normal strain ε_{xx} value that varies along the x direction as seen previously in figure 6 (c).

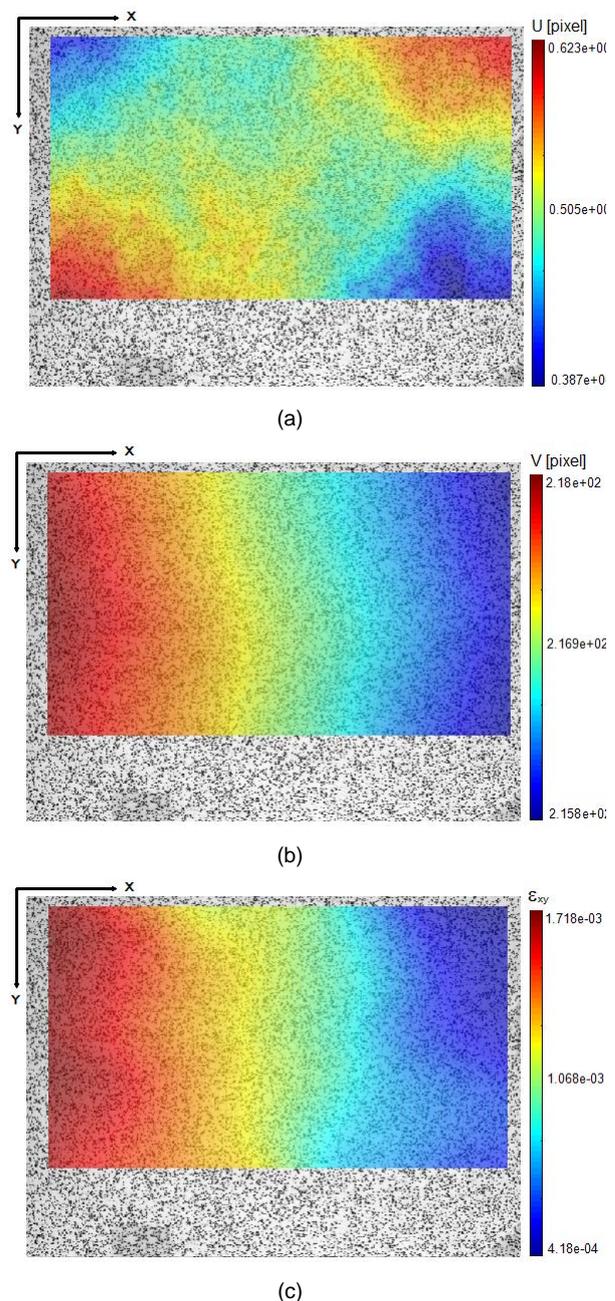


Figure 7: Typical DIC results for in-plane rigid-body-translation in the direction parallel to the tilt axis ($\theta = 5^\circ$, $\Delta y = 16$ mm) ; **a**) measured horizontal displacement, **b**) measured vertical displacement, **c**) measured shear strain.

In order to compare the experimentally measured strain error values with the theoretical estimates, the “mean”

strain value ($\frac{1}{N} \sum \varepsilon$) for each strain component (ε_{xx} , ε_{yy} and ε_{xy}) was calculated over the entire area of the image for each experiment. Figure 9 shows a comparison of the experimental and theoretical normal strain error values resulting from in-plane translation in the direction perpendicular to the tilt axis. Figure 10 shows the same comparison for shear strain error resulting from in-plane translation in the direction parallel to the tilt axis. From both figures it can be seen that the experimental results are in very good agreement with the theoretical estimates. The figures also show that the biggest difference between the experimental results and the theoretical values is observed when the camera is perfectly perpendicular to the surface where the theoretical analyses predict that all strain components are equal to zero while the experiments show small magnitudes of strain error. The experimental error observed there is mainly due to the optical distortions caused by the imaging system. The results show that, for the experimental setup used here, an in-plane translation of 16mm (i.e., about 14% of the field-of-view width) can cause a normal strain error as high as 4×10^{-3} at 5° tilt angle and the error will be even higher for shorter camera stand-off-distance. While such value of strain error might not seem that noteworthy if large plastic strains are being measured, but in fact such error value is significant knowing that it exceeds the elastic strain limit for many metallic alloys. It should also be noted that the magnitude of the strain error is not uniform over the entire surface area (see figures 6 and 7) and more variation in the error values will be present when making measurements of a surface undergoing deformation. These results clearly show that in-plane rigid-body-translations of the surface being measured using 2D-DIC can be used for detecting any misalignment between the surface and the camera focal plane array. The magnitude of the misalignment angle, if present, can be calculated using simple theoretical equations such as those presented in this paper. The calculation of the misalignment angle can be useful in cases where physical constraints prevent the camera from being oriented perpendicular to the surface. In such case the images can be rectified using an approach similar to that proposed by Lava et al. [12].

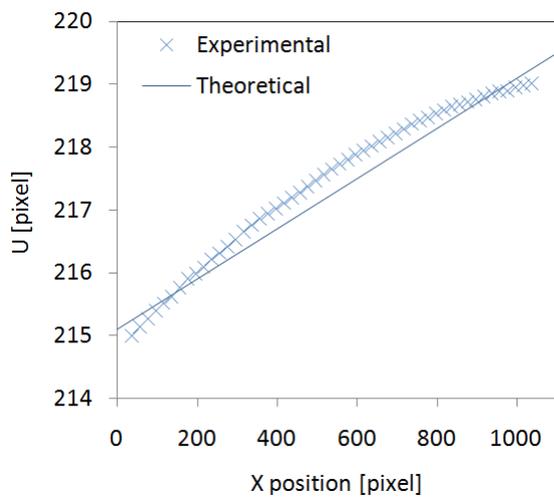
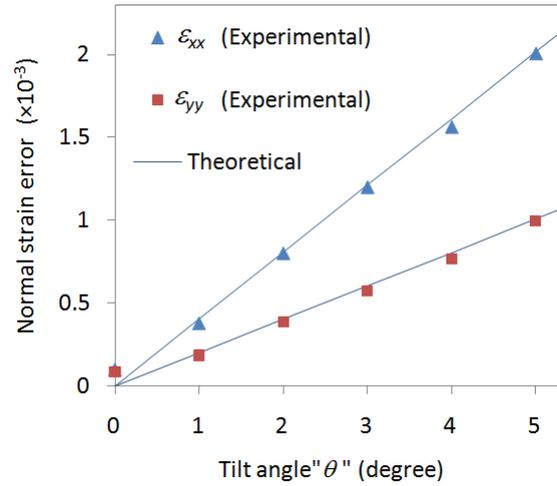
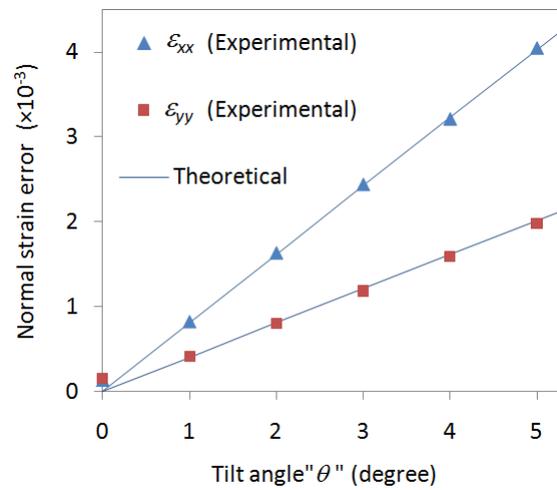


Figure 8: Comparison of the theoretically predicted and experimentally obtained relation between the horizontal translation and the position along the image x axis ($\theta = 5^\circ$, $\Delta x = 16$ mm).



(a)



(b)

Figure 9: Comparison of experimentally measured “mean” normal strains error with theoretical values for in-plane translation in the direction perpendicular to the tilt axis; a) $\Delta x = 8$ mm, b) $\Delta x = 16$ mm.

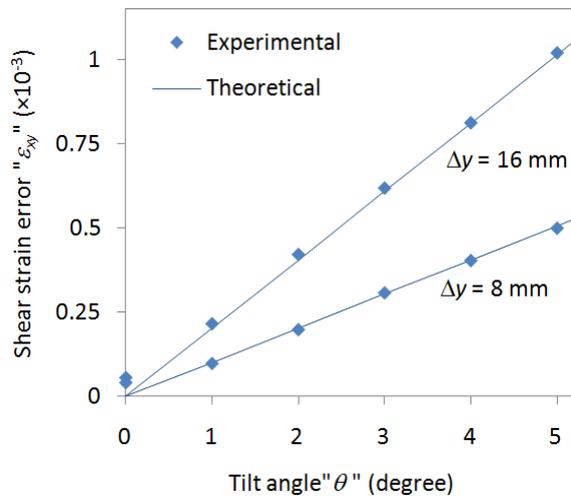


Figure 10: Comparison of experimentally measured “mean” shear strain error with theoretical values for in-plane translation in the direction parallel to the tilt axis.

In most realistic scenarios, the initial adjustment of the camera orientation with respect to the target surface is done manually without the use of any tools for angle measurement. In such cases, misalignment angles of up to two or three degrees are very well likely to go unnoticed and it is usually assumed that the DIC measurement error resulting from such small angular misalignments is negligible. The results presented here show that a misalignment angle as small as 2° can result in a strain error of more than $10^3 \mu$ -strains for an in-plane displacement of 14% of the field-of-view width. Such magnitude of error is enormous when the expected strains are not large, such as the measurement of elastic properties of materials. Though it is possible to measure the exact value of the misalignment angle using the simple equations presented here, in most cases, it is not very practical to do so. In principle, it is possible to use in-plane rigid-body-translations to identify the presence of any misalignment simply by inspecting the magnitudes of strain errors resulting from these translations. However, the fact that the magnitude of strain error depends on several factors (e.g., the optical system, the speckle pattern, the correlation parameters, etc.) makes it difficult to make such judgment. Figure 11 (a) shows the distribution of normal strain error (a histogram but shown using lines for ease of view) resulting from 16mm translation in the x direction at three different values of camera tilt angle. At zero degrees (i.e., camera is perpendicular), the error is completely random and its mean value is very close to zero. As mentioned earlier, the error in this case is mostly due to optical distortions caused by the imaging system. As the camera tilt angle increases, the error distribution is shifted (to the positive or negative direction according to the direction of tilt angle and direction of in-plane translation) due to the additional error caused by the camera non-perpendicularity. By calculating the “mean” value of the error ($\frac{1}{N} \sum \epsilon$) and the “mean of the absolute values” ($\frac{1}{N} \sum |\epsilon|$) and showing them versus the tilt angle, figure 11 (b), an interesting observation can be made. As can be seen from the figure, when the camera is perpendicular, there is a big difference between the magnitudes of the mean and the mean of absolute values of strain error. As the tilt angle increases, the mean starts approaching the mean of absolute values and at 1° their magnitudes become almost identical (for angles larger than one degree, they will have the exact same magnitude). This clearly indicates that it is possible to detect camera non-perpendicularity simply by calculating the “mean” and the “mean of absolute values” of strain errors and comparing their magnitudes. If their magnitudes are identical, or very close to each other, this means that the camera is tilted, and thus attempts can be made to correct its orientation. From figure 11 (b) it can be seen that for the experimental setup used here, misalignment angles smaller than 1° can be detected. Furthermore, the results presented in this paper indicate that a translation in one direction, x direction for instance, can indicate whether the camera is tilted around the x axis or the y axis (or both) since the normal strains error indicates that the camera is tilted around the y axis (i.e., the axis perpendicular to the translation direction) and the shear strain error indicates that it is tilted around the x axis (i.e., the axis parallel to the translation direction).

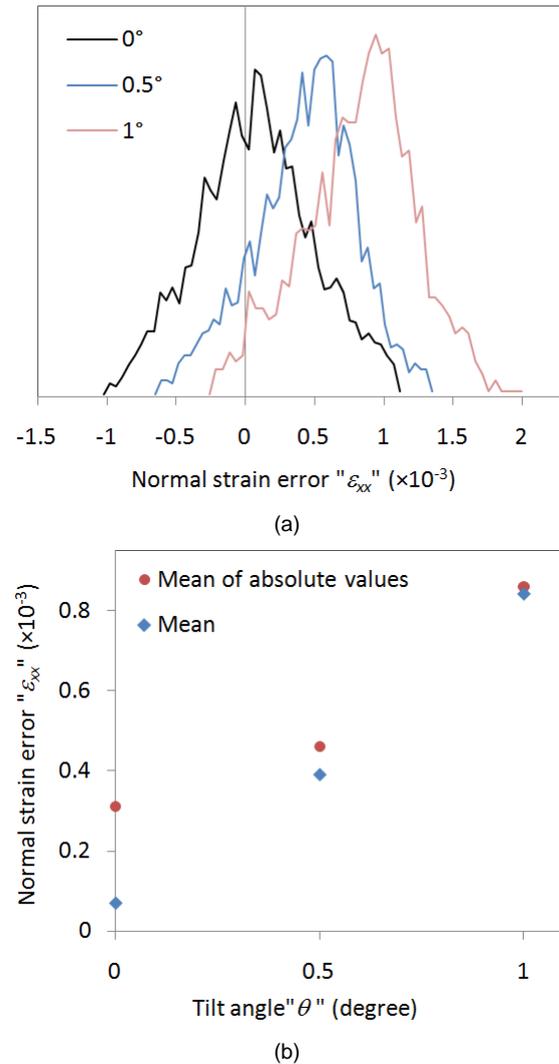


Figure 11: Experimentally measured normal strain error resulting from in-plane translation ($\Delta x = 16$ mm) at different tilt angles; **a**) error distribution histogram, **b**) the “mean” and the “mean of absolute values” of error.

6. Concluding Remarks

In this work we investigated the influence of the camera’s optical axis non-perpendicularity to the surface being observed on strain measurements errors both theoretically and experimentally. Theoretical models were developed based on simple pinhole camera model to estimate the errors in the measured normal and shear strains associated with in-plane translations in the directions perpendicular and parallel to the camera tilt axis. The conclusions of this investigation can be summarized in the following points.

- When the camera’s optical axis is not perpendicular to the surface being observed, any in-plane translation in the direction perpendicular or parallel to the camera tilt axis will cause errors in strain measurements.
- In-plane translations in the direction perpendicular to the camera tilt axis cause errors in both normal strain components.
- In-plane translations in the direction parallel to the camera tilt axis cause error in shear strain.

- In-plane rigid-body-translations can be used for estimating the resulting normal and shear strain errors.
- Theoretical equations based on simple pinhole camera model cannot predict the actual gradient of the strain errors resulting from camera non-perpendicularity; however, they can predict the mean values of strain errors.
- The magnitude of error in all strain components increases linearly with the tilt angle (for the range of angles investigated here; up to 5°) and the magnitude of in-plane translation.
- The strain error is inversely related to the stand-off-distance and, practically, the same can be said about the focal length of the lens when a constant magnification factor is desired.
- Misalignment angles as small as 2°, which could be easily overlooked during the initial setting of the experimental setup, are found to cause strain error greater than $10^3 \mu$ -strains.
- The theoretical strain error analyses are in very good agreement with the mean values of the strain errors measured experimentally. This indicates that the simple theoretical models presented here can be used to estimate the camera tilt angle and hence correct any existing non-perpendicularity between the camera's optical axis and the target surface.
- Simple in-plane rigid-body-translation experiments can be used to detect any existing non-perpendicularity between the camera's optical axis and the target surface by comparing the magnitudes of the "mean" and the "mean of absolute values" of strain errors. Experiments show that misalignment angles smaller than 1° can be detected.
- An in-plane translation in one direction (x or y) can indicate whether the camera is tilted around the x axis or the y axis (or both) since the normal strains errors are associated with translations perpendicular to the tilt axis and the shear strain error is associated with translations perpendicular to the tilt axis.

7. Acknowledgment

The first author is pleased to acknowledge the financial support provided by the German Research Foundation (DFG) for funding his research visit to the Institute of Fluid Mechanics and Aerodynamics at Universität der Bundeswehr München. He also acknowledges the support provided by the Deanship of Scientific Research at the Hashemite University for covering his travel expenses

References

- [1] Sutton, M., Wolters, W., Peters, W., Ranson, W. and McNeill, S., "Determination of displacement using an improved digital correlation method", *Image and Vision Computing*, Vol. 1, 1983, 133-139
- [2] Sutton, M., McNeill, S., Helm, J. and Chao, Y., "Advances in two-dimensional and three-dimensional computer vision", *Topics in Applied Physics*, Vol. 77, ed. Rastogi P (Berlin: Springer), 2000, p. 323-372
- [3] Sutton, M., Orteu, J. and Schreier, H., *Image Correlation for Shape, Motion and Deformation Measurements*, New York: Springer, 2009
- [4] Luo, P., Chao, Y., Sutton, M. and Peters, W., "Accurate measurement of three-dimensional deformations in deformable and rigid bodies using computer vision", *Experimental Mechanics*, Vol. 33, 1993, 123-132
- [5] Pan, B., Qian, K., Xie, H. and Asundi, A., "Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review", *Measurement Science and Technology*, Vol. 20, 062001, 2009, (17pp)
- [6] Orteu, J., "3-D computer vision in experimental mechanics", *Optics and Lasers in Engineering*, Vol. 47, 2009, 282-291
- [7] Chu, T., Ranson, W., Sutton, M. and Peters, W., "Applications of digital-image-correlation techniques to experimental mechanics", *Experimental Mechanics*, Vol. 25, 1984, 232-244
- [8] Bornert, M., Brémand, F., Doumalin, P., Dupré, J., Fazzini, M., Grédiac, M., Hild, F., Mistou, S., Molimard, J., Orteu, J., Robert, L., Surrel, Y., Vacher, P. and Wattrisse, B., "Assessment of digital image correlation measurement errors: methodology and results", *Experimental Mechanics*, Vol. 49, 2009, 353-370
- [9] Haddadi, H. and Belhabib, S., "Use of rigid-body motion for the investigation and estimation of the measurement errors related to digital image correlation technique", *Optics and Lasers in Engineering*, Vol. 46, 2008, 185-196
- [10] Sutton, M., Yan, J., Tiwari, V., Schreier, H. and Orteu, J., "The effect of out-of-plane motion on 2D and 3D digital image correlation measurements", *Optics and Lasers in Engineering*, Vol. 46, 2008, 746-757
- [11] Meng, L., Jin, G. and Yao, X., "Errors caused by misalignment of the optical camera axis and the object surface in the DSCM", *Journal of Tsinghua University (Science and Technology)*, Vol. 46, 2006, 1930-1932
- [12] Lava, P., Coppieters, S., Wang, Y., Van-Houtte, P. and Debruyne, D., "Error estimation in measuring strain fields with DIC on planar sheet metal specimens with a non-perpendicular camera alignment", *Optics and Lasers in Engineering*, Vol. 49, 2011, 57-65
- [13] VIC-2D manual, < www.CorrelatedSolutions.com >
- [14] MatchID-2D manual, < www.MatchID.org >



الجامعة الهاشمية



المملكة الأردنية الهاشمية

المجلة الأردنية
للهندسة الميكانيكية والصناعية

JJMIE

مجلة علمية عالمية محكمة

<http://jjmie.hu.edu.jo/>

ISSN 1995-6665

المجلة الأردنية للهندسة الميكانيكية والصناعية

مجلة علمية عالمية محكمة

المجلة الأردنية للهندسة الميكانيكية والصناعية: مجلة علمية عالمية محكمة تصدر عن عمادة البحث العلمي و الدراسات العليا في الجامعة الهاشمية بالتعاون مع صندوق دعم البحث العلمي في الأردن.

هيئة التحرير

رئيس التحرير

الأستاذ الدكتور موسى محسن

قسم الهندسة الميكانيكية، الجامعة الهاشمية، الزرقاء، الأردن.

الأعضاء

الأستاذ الدكتور أيمن المعاينة

الأستاذ الدكتور بلال العكش

الأستاذ الدكتور نسيم سواقد

الأستاذ الدكتور عدنان الكيلاني

الأستاذ الدكتور محمد النمر

الأستاذ الدكتور علي بدران

مساعد رئيس هيئة التحرير

الدكتور أحمد الغندور

فريق الدعم

تنفيذ وإخراج

المهندس حسن البعباع

المحرر اللغوي

الدكتور عبدالله جردات

ترسل البحوث إلى العنوان التالي:

رئيس تحرير المجلة الأردنية للهندسة الميكانيكية والصناعية

عمادة البحث العلمي والدراسات العليا

الجامعة الهاشمية

الزرقاء - الأردن

هاتف : 00962 5 3903333 فرعي 4147

Email: jjmie@hu.edu.jo

Website: www.jjmie.hu.edu.jo