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 Hashemite University in cooperation with the Jordanian Scientific Research and Innovation Support Fund, Ministry of Higher Education and Scientific Research.

EDITORIAL BOARD

THE INTERNATIONAL ADVISORY BOARD

EDITORIAL BOARD SUPPORT TEAM

SUBMISSION ADDRESS:

Prof. Moh'd Sami Ashhab, 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

JJMIE

An International Peer-Reviewed Scientific Journal

Financed by Scientific Research Support Fund

Volume 14, Number 2, June 2020

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 Hashemite University in cooperation with the Jordanian Scientific Research and Innovation Support Fund, Ministry of Higher Education and Scientific Research.

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 (Materials, Manufacturing, Management, Design, Thermal, Fluid, Energy, Control, Mechatronics, and Biomedical). 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 email.

Online: For online and email submission upload one copy of the full paper including graphics and all figures at the online, submission site, accessed via http://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

Submission address and contact :
Prof. Moh'd Sami Ashhab
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 and Technical reports

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 itstan abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, tha 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.

Withdrawing: If the author chooses to withdraw his article after it has been assessed, he shall reimburse JJMIE with the cost of reviewing the paper.

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. Please use MS Word for the text of your manuscript

Structure: Follow this order when typing manuscripts: Title, Authors, Authors title, Affiliations, Abstract, Keywords, Introduction, Main text, Conclusions, Acknowledgements, Appendix, References, Figure Captions, Figures and then Tables. 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 1.5 line 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.

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 quantities 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 (1997) No. 9, 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.

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.

Proof Reading: One set of page proofs in MS Word 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:

No page charges: Publication in this journal is free of charge.

Free offprints: One journal issues of which the article appears will be supplied free of charge to the corresponding author and additional offprint for each co-author. Corresponding authors will be given the choice to buy extra offprints before printing of the article.

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665

Jordan Journal of Mechanical and Industrial Engineering

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665
Pages 183 - 193

Jordan Journal of Mechanical and Industrial Engineering

Simulation and Performance Evaluation of CO_2 Booster System Integrated with Modified Evaporative Cooling for Supermarket Application in India

Manju Lata, Rishabh Yadav, Dileep Kumar Gupta*

Institute of Infrastructure Technology Research And Management (IITRAM) Ahmedabad, Gujarat, India

Received December 7 2019 *Accepted April 25 2020*

Abstract

CO_2 is one of the rediscovered, ecologically safe refrigerants with very low global warming potential which has favorable thermo-physical properties. The CO_2 booster refrigeration system has already been identified as a suitable choice for the supermarket application to replace the conventional R404A (high GWP) system. However, the performance of the CO_2 booster system is still comparatively lower than the conventional R404A system, especially when operated at high ambient temperature, which compels to improve the performance of the CO_2 system with suitable modification. In the present work, an attempt has been made to examine the year-round performance of the basic booster system and basic booster system with the integrated modified evaporative cooling system for Ahmedabad city weather conditions (Hot and Dry climate region). The experimentally investigated and validated data are used for the modified evaporative cooling system with real-time weather data taken from the weather station installed in the institute. Subsequently, the performance of the BBS, BBS-MEC, and R404A systems have been compared in terms of COP, power consumption, and seasonal energy efficiency ratio. The results show that for BBS-MEC, SEER enhances by 28.66% and annual power consumption decreases by 22.89% as compared with BBS. In addition to that, the total environment warming impact is also found significantly lower in the case of the BBS-MEC system.

Keywords: CO_2 refrigeration, booster system, evaporative cooling, R744, natural refrigerator, supermarket;

1. Introduction

Supermarket installations are continuously increasing worldwide, which lead to an increase in the use of R404A as a refrigerant for the supermarket. The refrigerant leakage rate predicted using this refrigerant ranges from 3% to 35% [1], which contributes to direct global warming due to high GWP. Subsequently, the power consumption of the system contributes to increasing global warming indirectly. Several alternative low GWP refrigerants, viz. CO_2, R600a, R32, NH4, R290, and R1234yf, etc. have been proposed to reduce the direct contribution of global warming. However, except CO_2, afore said refrigerants are either toxic and/or flammable. As a natural refrigerant with excellent thermo-physical and heat transfer characteristics, CO_2 has been recognized as an encouraging substitute [2]. Further, safety features i.e. non-flammability and non-toxicity in nature make CO_2 an attractive replacement of the conventional refrigerants. CO_2 has already been effectively commercialized in low-temperature climate regions, however, using in high-temperature climate regions, is still a challenge due to poor system performance at high ambient temperature [3].The lower critical temperature of CO_2i.e. 31.1oC makes the CO_2 cycle trans-critical, consequently, the CO_2 system performs vulnerably at high-temperature climates. Several studies have been carried out on the use of CO_2 as a refrigerant for supermarket applications in different ambient conditions, some of the recent developments are summarized below, The water spray technique was used for reducing gas cooler outlet temperature when operated at high ambient temperatures [4]. The yearly energy savings were limited to 3-5% and stated that saving is significant due to the high cost of electricity in peak summer periods. A theoretical investigation was done on two main trans-critical CO_2 systems centralized with an accumulation tank at the medium temperature and parallel with two separate circuits for low and medium temperature [5]. The performance of the centralized system with two-stage compression was found better in ambient temperature range 10-40ºC.Theoretical analysis is done of the trans-critical booster system for supermarket applications [6].The possible parameters were identified which affect the performance of the system at high ambient temperature. It was concluded that high side pressure, is highly dependent on compressor efficiency, ambient temperature, and suction line heat exchanger.

Comparison is done for different configurations of the CO_2 system with the baseline R404A direct expansion system using bin analyses in the eight climate zones of the

* Corresponding author e-mail: dileep.vnit@gmail.com.

United States for supermarket applications [1]. It was concluded that the trans-critical booster system with bypass compressor, performance is comparable to the R404A direct expansion system. Later, the oretical analysis was done for the integrated CO_2 trans-critical system with parallel compression, heat recovery, and air conditioning [7]. It was identified that in summer conditions, the performance of parallel compression is better than flash gas by-pass. The CO_2 system as compared HFC system is more efficient in lower ambient temperature (range 20-25°C) and less efficient in high ambient above 25°C. A theoretical investigation was done on the trans-critical CO_2 refrigeration system with an enhanced booster and parallel compression for supermarket applications in warm climatic conditions (35-50°C)[9]. The results indicated that the proposed cycle configuration has an advantage over the other modified cycles in warm climatic conditions. Further, a comparative analysis of five different CO_2 booster systems for supermarket application of four prominent cities based on annual hourly average temperature presented [10]. It was concluded that the performance of the CO_2 booster system with parallel compression with flooded LT evaporator and work recovery expander is better with maximum annual energy savings of 22.16% for New Delhi, India.A discussion was done on the control strategies of the CO_2 refrigeration system for supermarkets application [11]. It was concluded that integrated refrigeration and heating by CO_2 system for the supermarket saves up to 13% of primary energy compared to the conventional heating method. The conclusion derived from the study is that CO_2 is the only refrigerant which can be used as a refrigerant even at high ambient condition [12].A study on the integrated CO_2 trans-critical booster systems showed results indicating the two-stage heat recovery, flooded evaporation, parallel compression, and integration of air conditioning are the most promising features of the state-of-the-art integrated CO_2 system [13]. Subsequently, the booster system with gas cooler evaporative cooling was also investigated and concluded that annual energy saving was 1% and 3% for Stockholm and Barcelona. Results show that using evaporative cooled gas cooler does not contribute much in energy-saving at a moderate and humid place, but performance is better when operated for hot and dry climatic conditions.

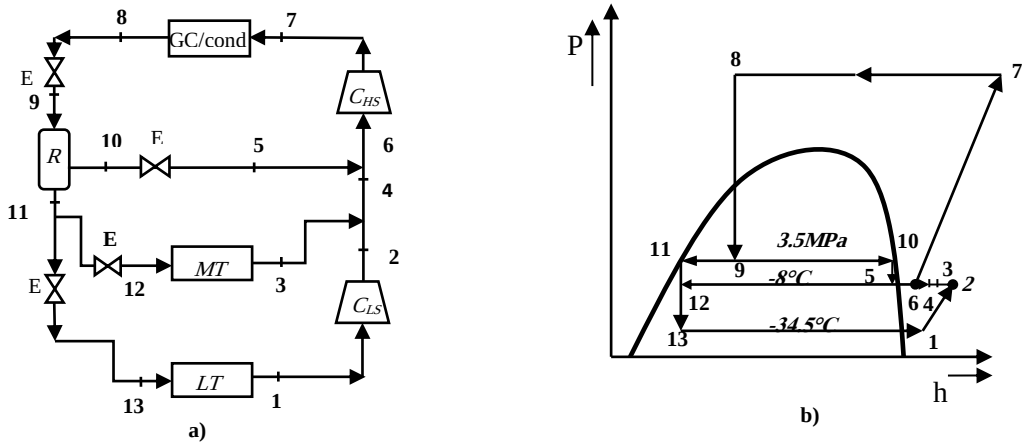
Recently, a comparison was done on the different CO_2 refrigeration systems with an internal heat exchanger, a parallel compression, two-stage compression, and mechanical subcooling [14]. It was observed that the mean COP of a two-stage compression and mechanical subcooling was enhanced by 49.8% and 75.8% respectively. Theoretical analysis is done on the CO_2 booster systems for supermarket applications with dedicated and integrated mechanical subcooling with a thermal load of 41 kW and 140 kW for low and medium temperature respectively [15]. The variation among both investigated systems in terms of annual energy reductions is 1.5-2.9% and 1.4-2.9% for tempered regions, 2.9-3.4%, and 2.9-3.4% for warm, 3.0-5.1%, and 1.3-2.4% for hot regions respectively. Investigation of ten different CO_2 refrigeration systems and its comparison with a conventional R404A system is done [16]. It was concluded that parallel compression, over fed MT and LT evaporators, intercooler, and mechanical subcooling reduced energy consumption by 8.53% yearly compared to the conventional R404A system. It was also concluded that the simple booster CO2 system is found to be less efficient than the conventional R404A system. A study showed R744 refrigeration as an alternative to the supermarket sector in Mauritius and concluded that the R744/R134a cascade system is a feasible solution both from energy efficiency and environmental perspective [17].

It has been observed from the above literature that there are limited studies available in the evaporative cooled CO_2 refrigeration system, some of the literature [4, 18] indicate that the evaporative cooling system is preferable for hot and dry climates. In this paper, an attempt has been made to investigate the yearly performance of the BBS and BBS-MEC and, compared with the conventional R404A system for supermarket application. The year-round performance has been analyzed using hourly ambient air condition (DBT and RH) real-time data, which is taken from the weather station (installed at institute lab) Ahmedabad city India. Subsequently, the reduced temperature is calculated using the experimentally investigated and validated MEC. These hourly data are used to generate temperature-bin hour's profile for the performance analysis of the aforesaid systems.

2. Description of the systems

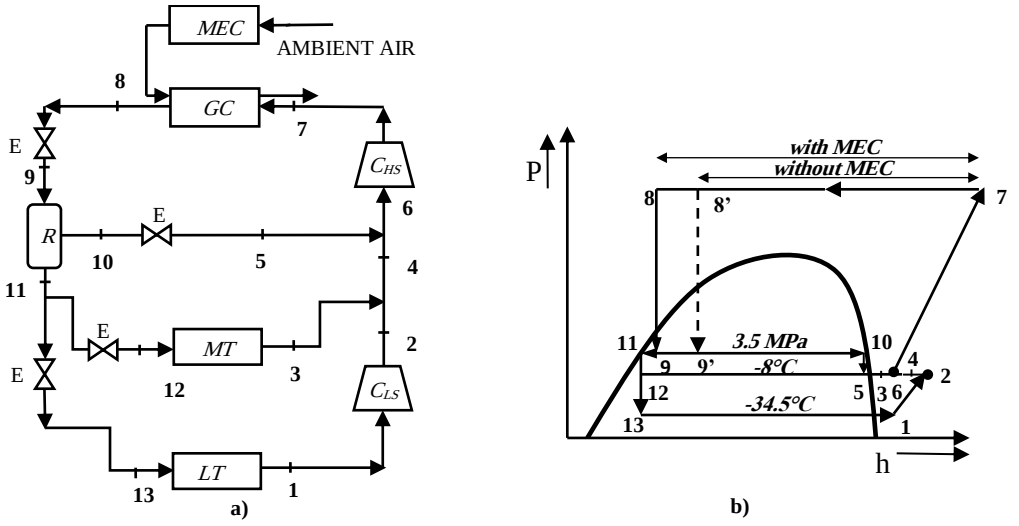
The refrigeration systems selected for the study are BBS, BBS-MEC, and conventional R404A system. The BBS system includes LT and MT loads, and the system operates in three pressures i.e. high, intermediate, and low pressure. The refrigerant from the HS compressor at high pressure enters into the gas cooler to reject heat and then passes through an expansion device to the R at the intermediate pressure. The refrigerant flows in two streams from the receiver i.e. in liquid and gaseous forms. The gaseous refrigerant is bypassed to the suction line of the HS compressor and the liquid form enters into the MT and LT evaporators through the expansion devices. The gaseous refrigerant from the MT evaporator mix with LS compressed gaseous refrigerant from the LT evaporator (low pressure). Subsequently, it mixes with the bypass refrigerant (gaseous form) from the receiver and enters into the HS compressor at the end of the cycle. The schematic and p-h diagrams of the BBS system are shown in Fig. 1.

The BBS-MEC system works similarly as the BBS system with an additional system (MEC) to the gas cooler. The MEC reduces the temperature of ambient air and supplies it to the gas cooler. Subsequently, the refrigerant temperature at the gas cooler outlet gets further reduced, which enhances the performance of the booster system. This can be seen clearly in Fig. 2(b) by point 8 (exit of the gas cooler), while point 8' describes the exit of the gas cooler when MEC was not used. Process 8'-9' represents the expansion process without using MEC and process 8-9 represents the expansion process using MEC. The schematic and p-h diagrams of the BBS-MEC system are shown in Fig. 2. (BBS-MEC) a) Schematic b) p-h diagram.



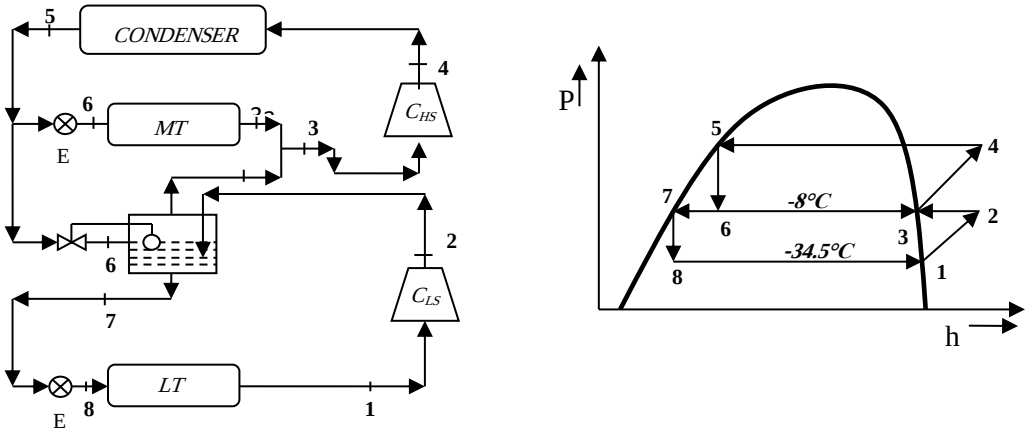
a) Schematic Diagram b) p-h diagram

Figure 1. CO_2 Basic Booster System (BBS)



a) Schematic Diagram b) p-h diagram

Figure 2. CO_2Basic Booster System with the integrated Modified Evaporative Cooling system (BBS-MEC)



b) p-h diagram

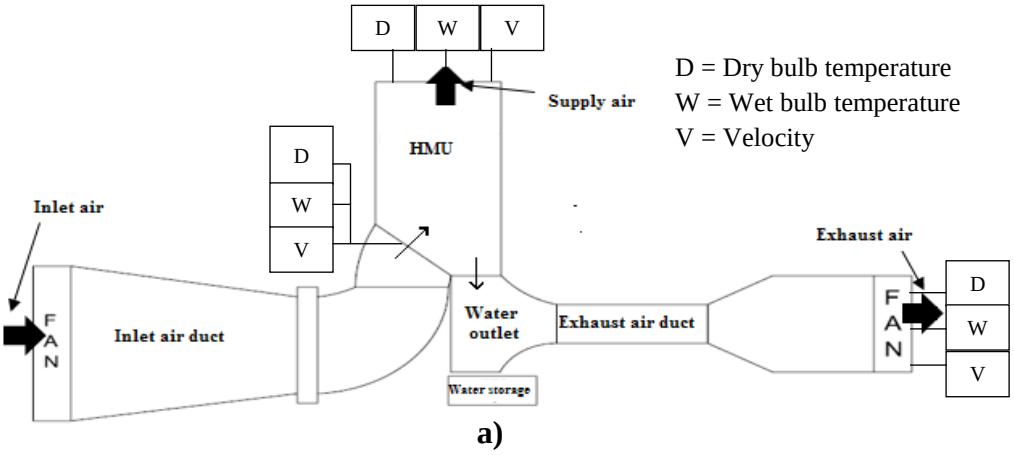
a) Schematic Diagram

Figure 3. Conventional R404A system

In the present scenario, the conventional R404A system is mostly used for supermarket refrigeration applications. Fig. 3 shows the schematic and p-h diagrams of the conventional R404A refrigeration system.

3. MEC System

There is an additional component which has been integrated into the BBS system to analyze the BBS-MEC system. The schematic and actual experimental setup fabricated in the laboratory is shown in Fig. 4 (a and b), and Thermo-Anemometer (Model 8912) in Fig. 3 (c), the measuring instrument used for measuring DBT, WBT, and velocity at the inlet, supply, and exhaust section of the MEC system. Experimental and theoretical analysis of the MEC system has been carried out on the laboratory scale. Initially experimental set up is tested to measure the RAT, subsequently, a steady-state mathematical model has been developed in MATLAB® and the results are validated using the experimental data. The results show good agreement with a maximum error of 3.66% for RAT and 8.89% for the TD. Further, the validated model has been used to obtain the RAT, throughout the year, which is given as input to the gas cooler.



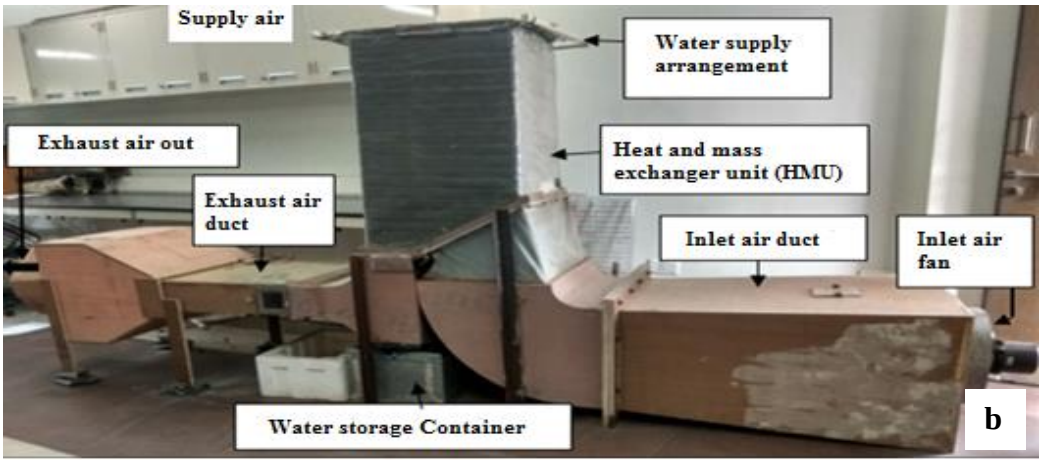
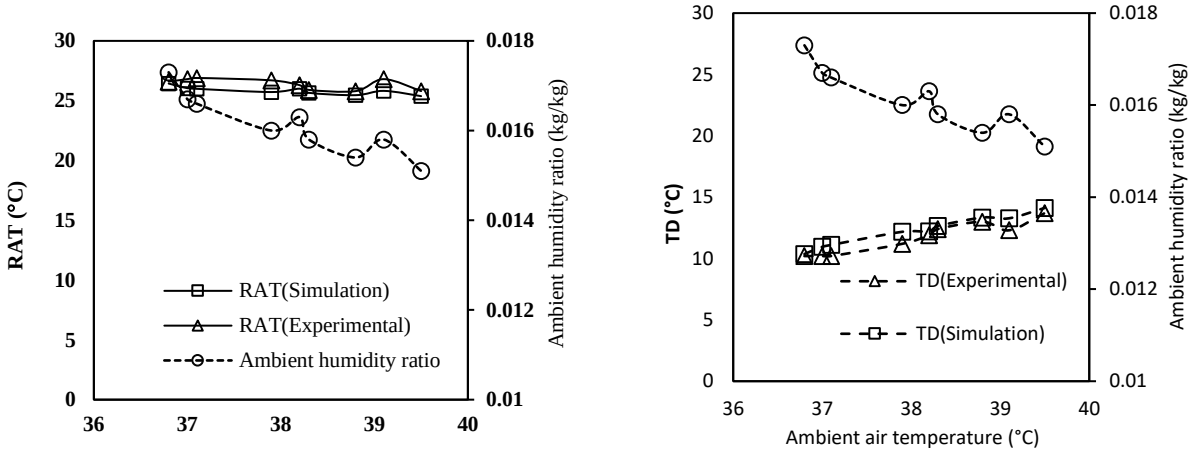


Figure 4. Experimental Facility (MEC) in Laboratory a) Schematic Diagram b) Actual Experimental Setup c) Measuring Instrument Vane Thermo-Anemometer (Model 8912)

Fig. 5 (a and b) shows the variation of RAT (exit or supply air temperature from MEC, which is reduced ambient temperature), and TD (temperature drop of supply air after passing through MEC) measured after the process of cooling in MEC with actual environment conditions i.e. DBT and Ambient humidity ratio as input variables. The results included the experimental as well as the simulation results at the respective operating conditions. It has been observed that there is a very good agreement between the experimental and simulated results. Moreover, the model has been used as a subroutine for the integration with the BBS for modeling the BBS-MEC systems.

The yearly data of Ahmedabad city is taken from the weather station data installed at the institute lab in IITRAM Ahmedabad, India. Fig. 6 shows the variation in bin hours of the ambient air conditions (DBT and RH) throughout the year (2017-18). The ambient air temperature is represented on X-axis. Y-axis (right) indicates ambient relative humidity and Y-axis (left) represents bin hours. The annually, DBT and RH range vary between 8°C-44°C and 23%-100 % respectively.



(a) Variation of RAT with ambient conditions

(b) Variation of TD with ambient conditions

Figure 5. Comparison of experimental and simulation results at different ambient air conditions

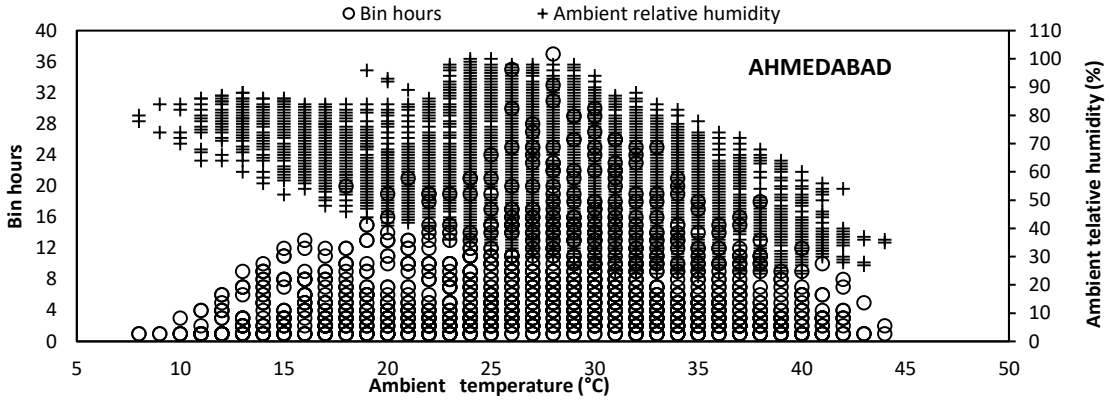


Figure 6. Bin hours at ambient air conditions

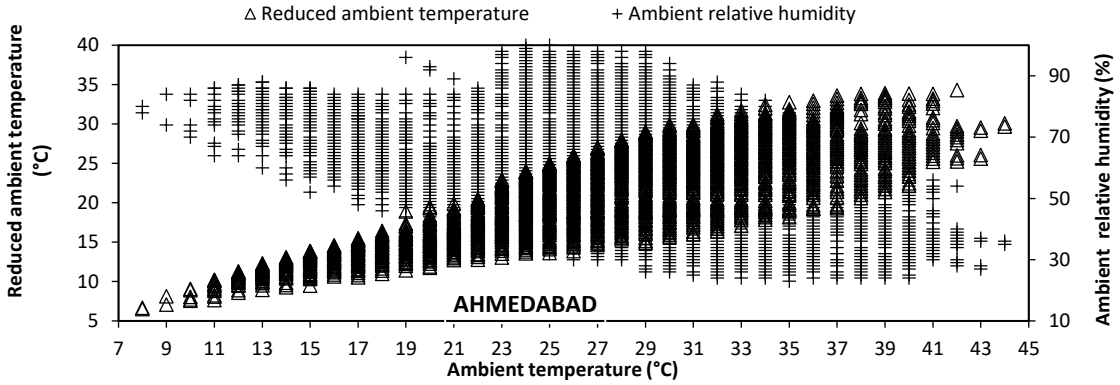


Figure 7. The reduced ambient temperature at ambient air conditions

Fig. 7 depicts the variation of RAT after the cooling process in MEC at different ambient air conditions (DBT and RH) throughout the year evaluated using the validated model of the MEC. The RAT has been evaluated using the mathematical equations of MEC, published elsewhere in Lata and Gupta (2020) [8].The inputs are taken from the ambient conditions as shown in Fig. 6. It has been observed that minimum and maximum reduced air temperature obtained after MEC is 6.52°C at the ambient condition of 8°C DBT and 78% RH, and 34.31°C, at the ambient condition of 42°C DBT and 54% RH, respectively. This RAT has been used as input for the analysis of the BBS-MEC system.

4. Thermodynamic Modeling

The steady-state thermodynamic model of the BBS, BBS-MEC, R404A systems has been developed by considering the following assumptions i.e. throttling process as expansion process, pressure loss, and heat transfer in components and piping are negligible. Table 1, summarizes the mass balance and energy balance equations for the different systems, and these equations have been solved in MATLAB® to evaluate the performances of the systems. The rmophysical properties of the refrigerants, air, and water are taken from REFPROP 9.0 by using the subroutine. Isentropic efficiencies for both the compressor are calculated by using correlations given in the literature [9]. Subsequently, the implementation of the mathematical model of the three systems i.e. R404A, BBS, and BBS-MEC is described in Fig. 8.

5. Operating Conditions and Parameters

The system performance has been analyzed at the ambient temperature of Ahmedabad weather conditions ranging from 8ºC to 44ºC with a gas cooler pressure range from 4.6MPa to 10.6MPa. The correlations of temperature and pressure and the operating conditions for the investigated configurations are taken from the literature [10], as shown in Table 2 and Table 3.

Table 1. Thermodynamic equations for the simulation of the investigated systems

R404A	**BBS**
$\dot{W}_{LS} = \dot{m}_{LT} \times (h_2 - h_1)$	$\dot{W}_{LS} = \dot{m}_{LT} \times (h_2 - h_1)$
$\dot{W}_{HS} = \dot{m}_{MT} \times (h_4 - h_3)$	$\dot{W}_{HS} = (\dot{m}_{LT} + \dot{m}_{MT} + \dot{m}_{F1}) \times (h_7 - h_6)$
$\dot{Q}_{cond.} = \dot{m}_{HS} \times (h_5 - h_4)$	$\dot{Q}_{cond./gc} = (\dot{m}_{LT} + \dot{m}_{MT} + \dot{m}_{F1}) \times (h_7 - h_8)$
$\dot{m}_F \times (h_3 - h_6) = \dot{m}_{LS} \times (h_3 - h_2)$	$(\dot{m}_{LT} + \dot{m}_{MT} + \dot{m}_{F1}) \times h_9 = (\dot{m}_{F1} \times h_{10}) + (\dot{m}_{LT} + \dot{m}_{MT}) \times h_{11}$
$\dot{m}_{HS} - \dot{m}_{MT}) \times (1 - x_6) = \dot{m}_{LS}$	$\dot{Q}_{MT} = \dot{m}_{MT} \times (h_3 - h_{12})$
$\dot{m}_{HS} = \dot{m}_{MT} + \dot{m}_{Flash} + \dot{m}_{LS} \times \left(\frac{x_6}{1 - x_6}\right)$	$\dot{Q}_{LT} = \dot{m}_{LT} \times (h_1 - h_{13})$
$\dot{Q}_{MT} = \dot{m}_{MT} \times (h_3 - h_6)$	
$\dot{Q}_{LT} = \dot{m}_{LT} \times (h_1 - h_8)$	
BBS-MEC	
$\dot{W}_{LS} = \dot{m}_{LT} \times (h_2 - h_1)$	
$\dot{W}_{HS} = (\dot{m}_{LT} + \dot{m}_{MT} + \dot{m}_{F1}) \times (h_7 - h_6)$	
$\dot{Q}_{cond./gc} = (\dot{m}_{LT} + \dot{m}_{MT} + \dot{m}_{F1}) \times (h_7 - h_8)$	
$(\dot{m}_{LT} + \dot{m}_{MT} + \dot{m}_{F1}) \times h_9 = (\dot{m}_{F1} \times h_{10}) + (\dot{m}_{LT} + \dot{m}_{MT}) \times h_{11}$	
$\dot{Q}_{MT} = \dot{m}_{MT} \times (h_3 - h_{12})$	
$\dot{Q}_{LT} = \dot{m}_{LT} \times (h_1 - h_{13})$	
$m_p C_p(t_{p\_in} - t_{p\_out}) = h_p dA_s(t_{p\_a} - t_{l\_a})$	
$\dot{m}_w(W_{w\_in} - W_{w\_out}) = h_m dA_s(\rho_{wf\_a} - \rho_{w\_a})$	
$\dot{m}_{wf\_in} - \dot{m}_{wf\_out} = \dot{m}_w(W_{w\_in} - W_{w\_out})$	
$\dot{m}_w(H_{w\_out} - H_{w\_in}) = h_w dA_s(t_{wf_a} - t_{w_a}) + H_{wv} h_m dA_s(\rho_{wf\_a} - \rho_{w\_a})$	
$m_p C_p(t_{p\_in} - t_{p\_out}) + \dot{m}_w(H_{w\_out} - H_{w\_in}) = m_{wf} C_{wf}(t_{wf\_in} - t_{wf\_out}) - \dot{m}_w(W_{w\_in} - W_{w\_out}) C_{wf} t_{wf}$	

Table 2. Correlations for the investigated CO_2 booster configurations[10].

Sub-critical		**Transition**	**Trans-critical**
T_{amb}< 4°C	**4°C <T_{amb}< 17°C**	**17°C <T_{amb}< 28°C**	**T_{amb}> 28°C**
T_{cond}= 9°C	$T_{cond} = T_{amb} + T_{approach}$	$T_{gc/cond} = 0.9 * T_{amb} + 4.7$	$T_{out,gc} = T_{amb} + T_{approach}$
$T_{out,cond}$= 7°C	$T_{out,cond} = T_{cond} - T_{sub-cooling}$	$P_{gc/cond} = (166.33 * T_{gc/cond} + 2676.3)$ kPa	$T_{approach}$ = 2°C
$T_{sub-cooling}$ = 2°C	$T_{sub-cooling}$= 2°C		P_{gc} = optimized
	$T_{approach}$ = 3°C		

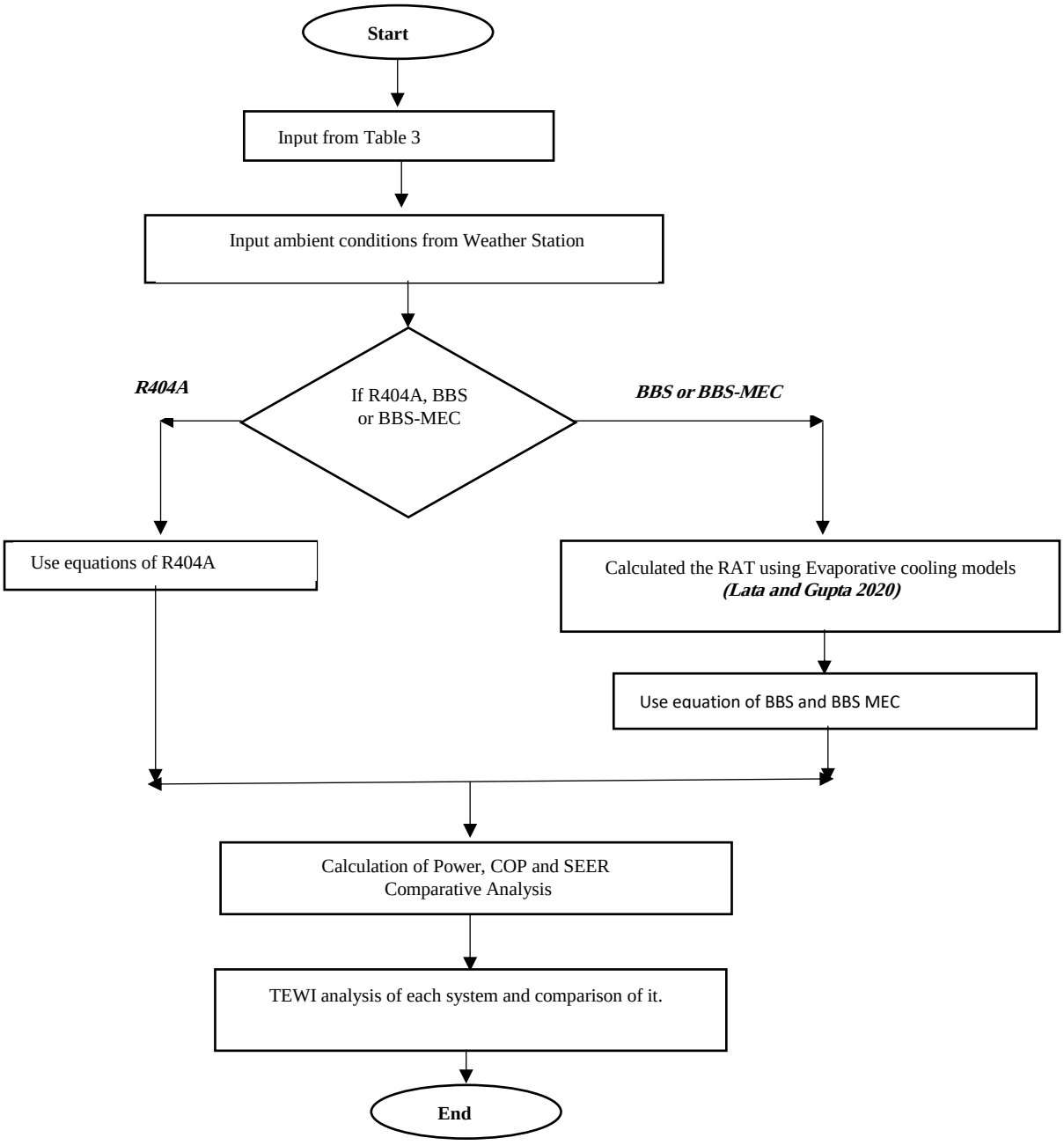


Figure 8. Model implementation of R404A, BBS, and BBS-MEC systems

Table 3. Operating parameters of the investigated configurations [10].

Operating Parameters	Values
MT load	120 kW
LT load	65 kW
Approach temperature for R404A system	10 °C
Approach temperature for gas cooler	2 °C
MT evaporating temperature	−8 °C
LT evaporating temperature	-34.5 °C
Minimum condensing temperature for CO_2/R404A system	9/25 °C
Superheating	5 °C
Circulation ratio(CR) for LT flooded evaporator	2.5
Intermediate vessel (R1) pressure	3.5 MPa
Maximum gas cooler pressure	10.6 MPa

6. Model Validation

The model developed for BBS is validated against published data [10] using the same operating conditions. Fig. 9 and Fig. 10 depict the validation of COP and optimum gas cooler pressure, respectively for the BBS system. The results show good agreement with the published data, the maximum deviation found in COP and optimum pressure, is 10.54 % and 1.6% respectively. There is no data available in the literature for BBS-MEC as it has not been explored much in the past. Hence, the analysis of such a system adds the novelty to this work.

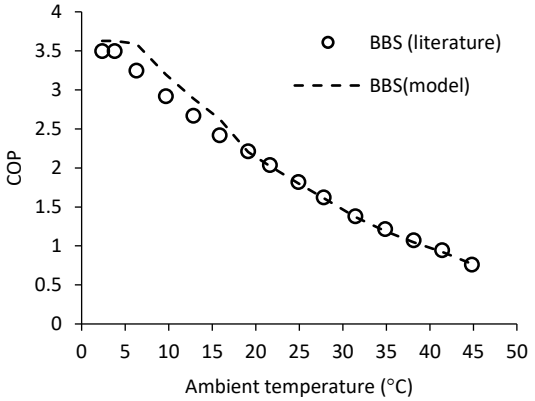


Figure 9. COP comparison of the developed model with published data

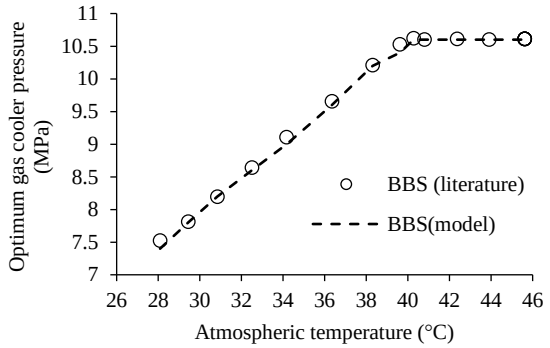


Figure 10. Optimum gas cooler pressure comparison of the developed model with published data

7. Results and discussion

The detailed analysis of all three configurations i.e. BBS, BBS-MEC, and R404A has been carried out using the thermodynamic models. The bin hours with different ambient conditions through a year have been calculated using Fig. 9 and Fig. 10 for the BBS and BBS-MEC system respectively as shown in Fig. 11. It is observed that for the BBS-MEC system, maximum bin hours obtained are 846 at 26°C DBT, whereas for BBS and R404A system, the maximum bin hours is 656 at 28°C DBT. It is a clear indication that the BBS-MEC system operates more at a lower temperature which leads to better performance.

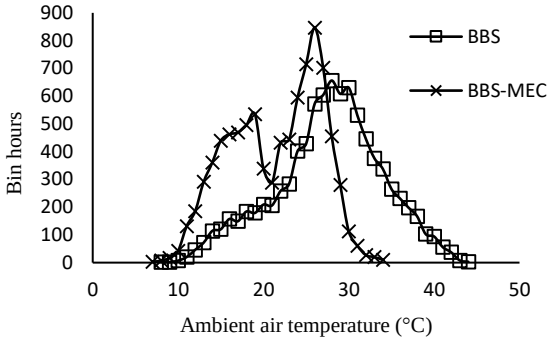


Figure 11. Bin hours at different ambient air temperature

Fig. 12, indicates the variation of COP with the bin hours throughout the year for conventional R404A system, BBS, and BBS-MEC systems. It has observed that maximum bin hour for the conventional R404A system is 656 at COP 2.03, followed by 656 for BBS at COP 1.60 and 846 at COP 1.73 for BBS-MEC system. Similarly, Fig. 13, indicates the variation of power consumption in kW with the bin hours throughout the year for conventional R404A, BBS, and BBS-MEC systems. It is observed that maximum bin hour for the conventional R404A system is 670 with a power consumption of 93.78 kW, followed by 670 for BBS, and 835 for BBS-MEC with the power consumption of 120.83 kW and 106.997 kW respectively. The variations of these bin hours mainly depend on the local ambient temperature and relative humidity at a specific time.

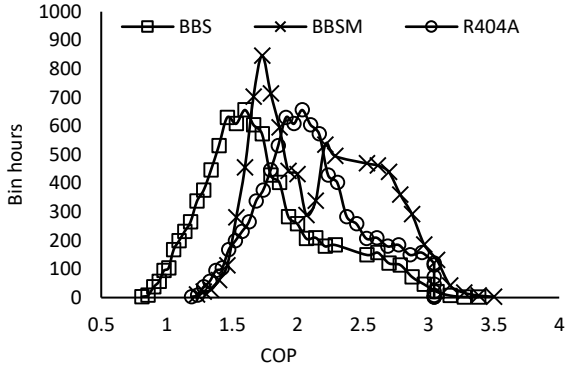


Figure 12. COP with Bin hours at different ambient air temperature

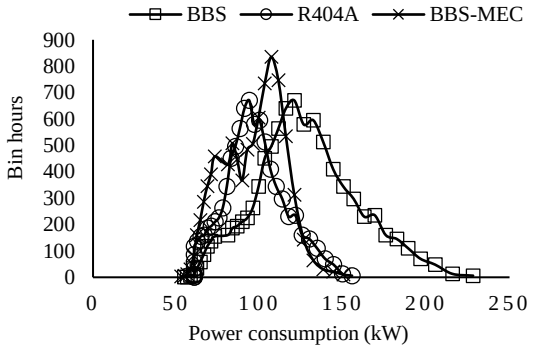


Figure 13. Power Consumptions with Bin hours at different ambient air temperature

Fig. 14 depicts the optimum gas cooler operating pressure with different ambient air temperatures. It has been observed that the maximum gas cooler pressure found is 10.6 MPa for BBS, which reduced to 9 MPa for BBS-MEC

systems at the same ambient conditions. This is because optimum gas cooler pressure decreases with a decrease in ambient air temperature, using the MEC system.

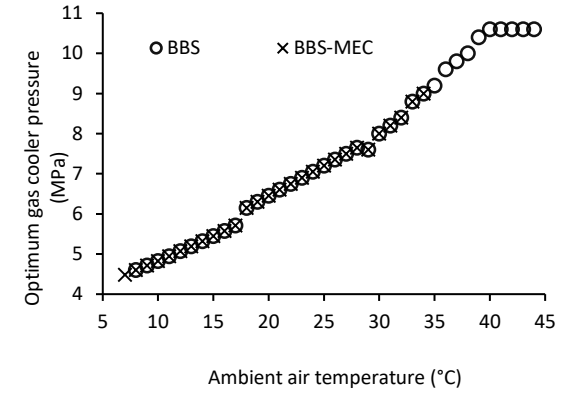


Figure 14. Optimum gas cooler pressure at various ambient temperatures

Further, the overall performance of all three systems has been evaluated in terms of SEER, and annual power consumption, throughout the year. Fig. 15 shows the percentage change in SEER and annual power consumption for BBS-MEC and conventional R404A, with reference to BBS.

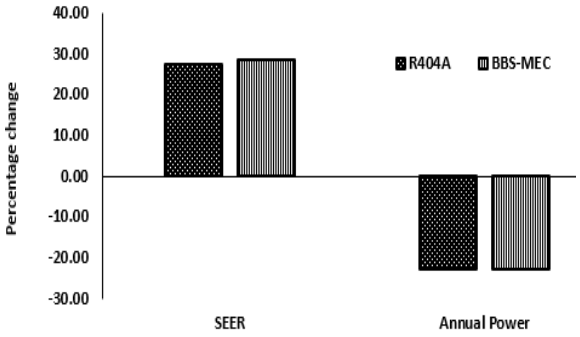


Figure 15. Percentage Change in SEER and Annual Power Consumption

It is observed that as compared to the BBS system, the BBS-MEC and conventional R404A systems, have 28.66% and 27.44% higher SEER respectively. Whereas, annual power consumption decreases for BBS-MEC and conventional R404A systems by 22.89% and 22.72% respectively. However, there is no significant difference in the performance of BBS-MEC and R404A systems. Moreover, the CO_2 BBS-MEC system can be considered as a suitable alternative of conventional R404A systems with comparable performances.

8. TEWI Analysis

The Total Equivalent Warming Impact (TEWI) has been evaluated to analyze the overall impact of the system on the environment. The emission of carbon equivalents (in the form of weight) into the atmosphere of the three systems has been carried out. Subsequently, the comparative analysis of TEWI for the BBS and BBS-MEC systems comparing to the conventional R404A system has also been carried out. The following correlations are used for TEWI evaluations for the corresponding systems, taken from the literature[19].

$$TEWI = TEWI_{direct} + TEWI_{indirect}$$

$$TEWI_{direct} = GWP * L * n + GWP * m * (1 - a_{recovery})$$

$$TEWI_{indirect} = E_{annual} * \beta * n$$

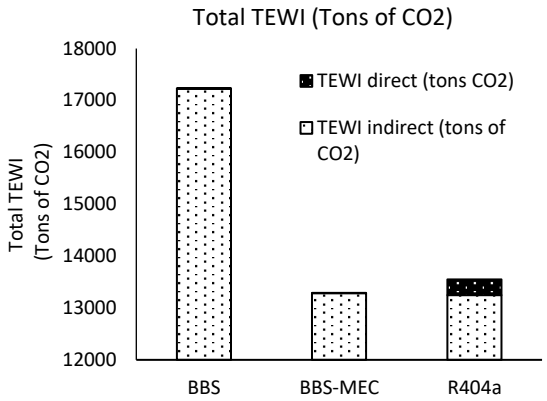


Figure 16. Total TEWI in Tons of CO_2 Equivalent of different systems

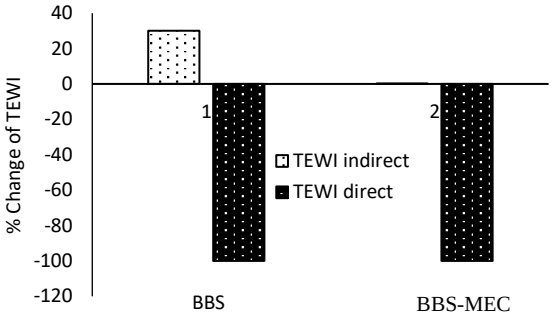


Figure 17. Percentage Change TEWI with reference to R404A system

The TEWI analysis, including direct and indirect contribution, the total TEWI in tons of CO_2 equivalent, are shown in Fig. 16. It is observed that the BBS system has the highest (17224.95 Tons of CO_2) value of TEWI followed by the R404A system (13550.86Tons of CO_2), whereas the value of TEWI is lowest in the case of BBS-MEC system (13285.85Tons of CO_2). Subsequently, the percentage change of TEWI emission with reference to the R404A system, as shown in Fig. 17. It has been found that there is a significant reduction in TEWI direct emissions (99.97%) in both BBS and BBS-MEC systems, moreover, the BBS-MEC system can be considered as an alternative of, conventional and higher GWP R404A system for the supermarket applications in warm climatic conditions.

9. Conclusions

A detailed thermodynamic analysis of the BBS, BBS-MEC, and conventional R404A system has been carried out using the real-time ambient data of Ahmedabad city (hot and dry climate) India. An additional system i.e. MEC has been integrated into the BBS to reduce the gas cooler/condenser outlet temperature, which enhanced the performance of the system. It is concluded that using BBS-MEC system annual SEER has been improved by 28.66% and power consumption has been reduced by 22.89% compared to the BBS. It has also been found that there is a

significant reduction in TEWI emissions in the BBS-MEC systems as compared to the conventional R404A and BBS systems for supermarket Applications. These results will help to provide design guidelines for the designer to get better performance using CO_2 as an eco-friendly refrigerant and alternative of the high GWP refrigerant R404A for supermarket applications at different climatic conditions.

Acknowledgment

The authors would like to thank the management department of the Institute of Infrastructure Technology Research And Management (IITRAM) Ahmedabad, Gujarat, India (the host Institute), for providing financial assistance to fabricate the MEC setup and other required support for the experimentation.

Nomenclature

a	Recovery/recycling factor
A_s	surface area (m/s)
C	Compressor
C	specific heat capacity at constant pressure (J/kgK)
CO_2	carbon dioxide
E	expansion valve
E	Energy consumption (kWh/year)
G	passage gap
h	heat transfer coefficient (W/m^{2}K)
h_m	mass transfer coefficient (m/s)
H	enthalpy (kJ/kg)
H_l	latent heat of vaporization of water at 0 °C (J/kg)
H_{wv}	enthalpy of the water vapor at water film temperature (J/kg)
L	Annual leakage rate (kg/year)
m	Refrigerant charge (kg)
$\dot{m}$	mass flow rate (kg/s)
n	System operating years
P	Pressure
P	Pump
$\dot{Q}$	refrigeration load (kW)
R	Receiver
t	temperature (K)
T_{supply}	temperature of supply air (°C)
$\dot{W}$	compressor work (kW)
W	humidity ratio (kg/kg of dry air)
W_{evap}	Water evaporation rate (kg/s)
x	dryness fraction
ρ	density (kg/m^3)

Subscript

A	Average
amb	Ambient
cond	Condenser
F	flashed at receiver
Gc	gas cooler
HS	high stage
In	Inlet
L	Wall
LS	low stage
LT	low temperature
MT	medium temperature
Out	Outlet
P	Primary air
S	Supply air
W	working air
Wb	wet bulb
w_d	dry working air
Wf	Water film
Wv	Water vapor
Wb	wet bulb
w_d	dry working air
Wf	Water film
Wv	Water vapor

Abbreviations

BBS	Basic Booster System
BBS-MEC	Basic Booster System with Modified Evaporative Cooler
COP	coefficient of performance
DBT	dry bulb temperature
GWP	Global Warming Potential
HMU	Heat and Mass Exchanger Unit
MEC	Modified Evaporative Cooler
RAT	Reduced Ambient Temperature
RH	relative humidity
SEER	Seasonal Energy Efficiency Ratio
TD	Temperature Drop
TEWI	Total Environment Warming Impact

References

[1] S.Vishaldeep, F. Brian, B. Pradeep,"Comparative analysis of various CO_2 configurations in supermarket refrigeration systems". International Journal of Refrigeration, 46, 2014, 86 -99.

[2] K. Man-Hoe, P. Jostein, B.W. Clark,"Fundamental process and system design issues in CO_2 vapor compression systems". Progress in Energy and Combustion Science, Vol. 30, No. 2, 2004,119–174..

[3] N. Petter, W. T. Harald, H. Armin,"CO_2 -a refrigerant from the past with prospects of being one of the main refrigerants in the future". In Proceedings of 9th IIR Gustav Lorentzen Conference, Sydney, Australia 2010.

[4] L. Giovanni, S. Filippini, F. Zoggia,"Using "water-spray" techniques for CO_2 gascoolers". Technological Innovations in Air Conditioning and Refrigeration Industry, Milano, Italy, June 8-9, 2017.

[5] S. Samer, "Theoretical evaluation of trans-critical CO_2 systems in supermarket refrigeration. Part I: Modeling, simulation and optimization of two system solutions".International Journal of Refrigeration, 31, 2008, 516 – 524..

[6] Y.T. Ge, S.A. Tassou, "Thermodynamic analysis of trans-critical CO_2 booster refrigeration systems in supermarket". Energy Conversion and Management, Vol. 52,No. 4, 2011, 1868–1875.

[7] K. Mazyar, S. Samer, "Theoretical analysis of CO_2 trans-critical system with parallel compression for heat recovery and air conditioning in supermarkets". 24th IIR Refrigeration Congress of Refrigeration. IIF/IIR,Yokohama, Japan, vol. 304. 2015.

[8] M. Lata, and D. K. Gupta, Experimental Investigation and Simulation of Modified Evaporative Cooling System; International conference on Advances of mechanical engineering, held at VNIT Nagpur Jan 2020,(paper Id 117) accepted for Scopus book chapter springer publication.

[9] P. Nilesh, G. K. Dileep, D.S. Mani,"Thermodynamic analysis of CO_2 trans-critical booster system for supermarket refrigeration in warm climatic conditions". Proceedings of the 4th IIR Conference on Sustainability and the Cold Chain, Auckland, New Zealand, 2016.

[10] P. Nilesh, G. K. Dileep, D.S. Mani, "Energetic and economic analysis of trans-critical CO_2 booster system for refrigeration in warm climatic condition". International Journal of Refrigeration, 80, 2017, 182–196.

[11] S. Liang, C.I. Ferreira, G. Jan, K. Hendrine, "Control strategies of CO_2 refrigeration/heat pump system for supermarkets". Proceedings of the 12th IEA Heat Pump Conference, 15th–18th May. 2017.

[12] G. Paride, H. Armin, B. Krzysztof, "Trans-critical R744 refrigeration systems for supermarket applications: Current status and future perspectives". International Journal of Refrigeration, 93, 2018, 269-310.

[13] K. Mazyar, S. Samer, "State-of-the-art integrated CO_2 refrigeration system for supermarkets: A comparative analysis". International journal of refrigeration, 86, 2018, 239-257.

[14] B. Evangelos, T. Christos, "A comparative study of CO_2 refrigeration systems. Energy Conversion and Management", 2019, https://doi.org/10.1016/j.ecmx.2018.100002.

[15] C. Jesus, L. Rodrigo, S. Daniel, A.N. Laura, C. Ramon, "Energyanalysis of dedicated and integrated mechanical subcooled CO_2 boosters forsupermarket applications". International Journal of Refrigeration, 101, 2019, 11-23.

[16] M. Georgios, S. Evangelos, T. Dimitrios, B. Evangelos, T. Christos, A. Stavros, "Annual performance of a supermarket refrigeration system using different configurations with CO_2 refrigerant". Energy Conversion and Management, 2019, https://doi.org/10.1016/j.ecmx.2019.100006.

[17] S. Darmanaden, P.Nilesh, M. Romeela, M. Francis, D.S. Mani, "R744 refrigeration as an alternative for the supermarket sector in Small Tropical Island Developing States: The Case of Mauritius". International Journal of Refrigeration, 103, 2019, 264-273. https://doi.org/10.1016/j.ijrefrig.2019.03.034.

[18] B. Riangvilaikul, S. Kumar, "Numerical study of a novel dew point evaporative cooling system". Energy and Buildings, Vol.42, No.11,2010, 2241–2250.

[19] Z. Mylona, M. Kolokotroni, K.M. Tsamos, S.A. Tassou, "Comparative analysis of the energy use and environmental impact of different refrigeration systems for frozen food supermarket applications", Energy Procedia, 123, 2017, pp.121-130.

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665
Pages 195 - 204

Jordan Journal of Mechanical and Industrial Engineering

Experimental Investigation of the Performance of a Vortex Tube with Conical Control Valve

Ahmad Al-Qaisia[a], Jamil Al Asfar [a], Nabeel Abu Shaban[b], Areej Eniezat [a,c]

[a] Mechanical Engineering Department, The University of Jordan, Amman, 11942, Jordan
[b] Mechanical Engineering Department, Al-Zaytoonah University of Jordan, Amman, 11733, Jordan
[c] Graduate Student, Mechanical Engineering Department, The University of Jordan, Amman, 11942, Jordan

Received February 16 2020 *Accepted May 31 2020*

Abstract

The Vortex tube (Ranque-Hilsch type/ RHVT) is a simple device used to obtain both cold and hot gas streams simultaneously from a compressed gas. The obtained cold gas is widely used in many low temperature commercial applications. In this work, the performance of the vortex tube was examined experimentally by studying the effects of inlet pressure variations, conical valve opening percentage, tube diameter and hot gas tube length on the thermal performance of the tube. Four tubes with diameters 14, 21, 25 and 32 mm with four hot exhaust gas lengths; 25, 50, 75 and 100 cm were tested experimentally and the obtained data were analyzed. It was observed that the 75 cm length of the hot side length, gives the maximum coefficient of performance for RHVT, as well as the maximum cold temperature reduction, either with or without insulation. Furthermore, the thermal performance of the RHVT was optimum for the two inner diameters; 14 and 21mm. This indicates that, when the ratio of hot side length to inner diameter (Lh/D) lies between 36 and 50, then the thermal performanmce of the RHVT is optimum and is highly recommended, which agrees with the results published in previous studies and invesitigations.

Keywords: Vortex tube, Conical control valve, experimental, performance;

1. Introduction

The Ranque-Hilsch vortex tube (RHVT) is a simple mechanical device operating as a refrigerating and heating apparatus that generates both hot and cold gas streams simultaneously from compressed air without utilizing any moving parts. The cold gas stream will be used in many industrial and commercial applications. Historically, the vortex tube was observed many decades ago, by Ranque [1] in 1933, and Hilsch [2] in 1947. Its principle of operation is based on the separation of a single compressed air stream with a uniform temperature into two streams (hot and cold), as it can be seen from Figure 1. Such a phenomenon is referred to as the temperature (or energy) separation effect, Figure 2. These figures show the operating principle and flow directions of the RHVT.

In Figure 3, a schematic diagram, is shown, to indicate the geometrical parameters of a given RHVT, i.e.; the inner diameter and length for cold and hot sides, orifice and nozzle diameters.

The vortex tube consists mainly of inlet tangential nozzles, vortex chamber or (hot side tube), cold orifice plate and control valve as shown. Basically, the vortex tube consists of one inlet in which the pressurized air is tangentially injected through a nozzle in the vortex chamber and exhausts on the right and left ends, as shown in Figures 1 and 2. The fraction of the air that leaves the vortex tube in the two directions is controlled by the conical valve. Due to the tangential injection, the velocity of air has significantly high rotational component in the chamber and splits into two streams; hot stream which exhausts at the right side periphery, and cold stream that exhausts at the left part of the tube. The vortex tube is a mechanical device and it operates as a refrigeration unit without any moving parts, no electricity or chemicals parts are needed, low cost, maintenance free, small and lightweight, adjustable temperature range, and no response time to reach the cold temperature, provided that compressed air is available. The geometrical parameters are: the vortex tube inlet diameter, cold orifice diameter, inlet nozzles diameter and number, conical control valve angle, cold tube length and hot tube length. When high pressure air enters into the vortex chamber through one or more tangential nozzles, a strong vortex flow is created and is split into two regions, the first one is the high temperature air near the boundary of the tube which leaves circumferentially through the conical valve, while the other one is the low temperature air that leaves through the cold orifice.

The principle of operation of the vortex tube, was first discussed by Ranque and Hilsch in 1933 and 1947,

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

respectively. Ranque [1] noticed that the air entering tangentially into a pipe exits from one outlet at a lower temperature and from the other outlet at a higher temperature than the inlet flow temperature. Hilsch [2] studied on the Ranque's findings and performed experimental and theoretical studies on the vortex tube gas flow to improve its efficiency. He worked on spiral flow using smoke visualization inside a vortex tube in which inner wall is covered by oil. Some researchers have been able to explain the principle of operation of the vortex tube numerically, due to the complexity of modeling the Rangue-Hilsh vortex tube mathematically. Due to its practical importance, the performance of the vortex tube, has been the subject of numerous numerical and experimental investigations over the years [3- 26]. A review of the relevant literature can be found in e.g. [28, 29]. The present work is intended to report the results obtained experimentally to study the performance of a vortex tube by measuring the temperatures and the mass flow rates, for 10 samples of vortex tubes with different lengths and diameters at different values of inlet pressure, which ranges from 1-4 bars [27]. The goal of this work is to present the results obtained experimentally, and to study the effect of the physical parameters of the vortex tube: hot tube diameter and length, percentage of conical valve opening, inlet pressure and insulation, on the thermal performance such as, cold temperature separation and coefficient of performance of the vortex tube. Based on the results and investigations presented in [27], it was concluded that the inner diameter of the vortex tube is one of the most important factors that affect its performance, since any increase in vortex tube inner diameter, will result in a decrease in its coefficient of performance and cold temperature reduction. In general, the results indicate that 20% of conical valve opening has produced maximum cold temperature reduction at all inlet pressures regardless of the geometrical parameters of hot tube, and 10% of conical valve opening has caused the vortex tube to operate at the maximum coefficient of performance. It is very necessary to mention that results also reflect the fact that, increasing inlet pressure causes the cold temperature reduction to increase regardless of the percentage of conical valve opening.

As mentioned before, the RHVT, is a very simple mechanical device and can be used as a refrigerating system, unlike the traditional air conditioning systems [30]

In light of the above review, the objective of this work is to study experimentally the performance of a vortex tube with conical control valve. The interest here, will be on studying and analyzing the thermal performance of the vortex tube, i.e.; coefficient of performance, reduction in ambient temperature and effect of insulation. Moreover, the performance will be thouroghly analyzed for different values of : inner diameters, hot and cold sides lenghths, conical control valve opening, and inlet pressures.

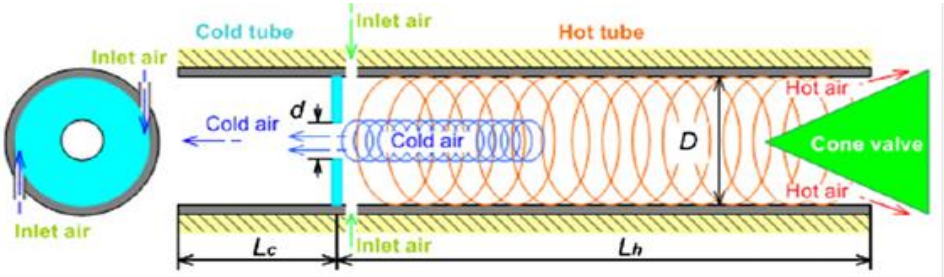


Figure 1. Air flow directions inside counter-flow type tube (Eiamsa-ard et al, 2010)

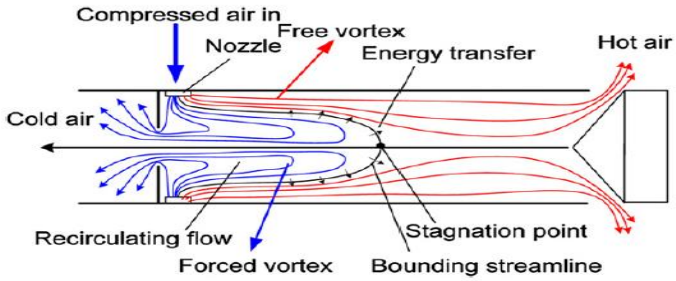


Figure 2. Stagnation point on vortex tube (Im and Yu, 2012)

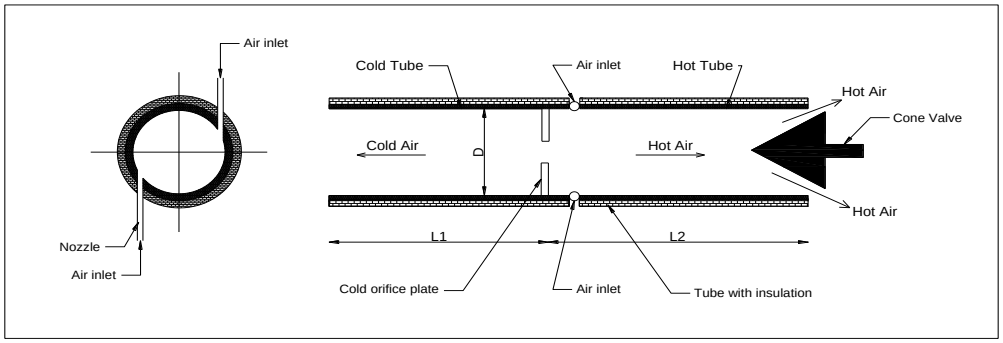


Figure 3. Schematic diagram of a general vortex tube (front and side views)

2. Experimental set-up description

The experimental setup of the vortex tube under consideration is shown in Figure 4. The test rig consists of a frame to hold the tube and a sliding type conical control valve

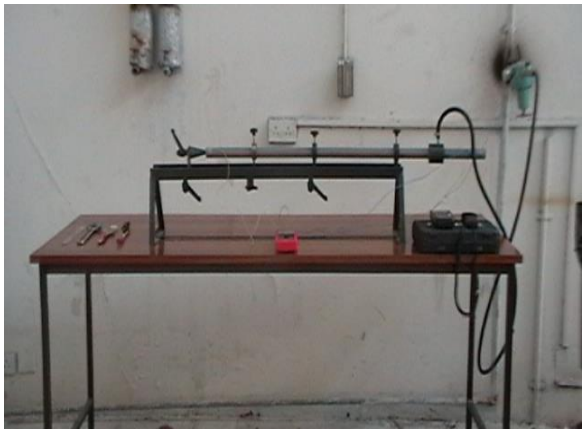


Figure 4. The Test rig used in the study

In this work, an Anemometer and three thermocouples were used to measure, respectively, the mass flow rates and temperatures at the inlet and two exits (hot and cold). The thermocouples of K-type located directly downstream at the inlet, hot and cold flows to measure temperatures by connecting them with a digital multi-meter type (TC-100).

The installed pressure gage on the main supply line (compressed air) was used to measure the pressure of the inlet air stream, i.e. the compressed air that is tangentially injected via the nozzle to the vortex chamber, and the target pressure was set by adjusting the regulation valve of the main supply line. In addition, on the vortex tube, a conical flow control valve was fixed at the end of the hot tube to control the mass flow rate of the hot stream (mass fraction), which in turn regulate also the cold stream mass fraction. The conical valve is fabricated in a very simple way and fixed on the mainframe of the test rig. It moves horizontally and a linear scale is used to calculate the percentage of the opening at the hot airside. Also, clamps are mounted on the steel frame to fix the vortex tubes during the experiments. As can be seen from the figure, the steel frame can be used to hold any vortex tubes individually during the experiment, to prevent any movements and vibrations resulted from the high pressures at the inlet. This steel frame has been designed to match all vortex tubes used regardless of their lengths and diameters. Ten vortex tubes made from Chlorinated polyvinyl chloride (CPVC) with different diameters were fabricated and used. The dimensions and geometrical parameters of these tubes are given in Table (1).

3. Experimental results

During the experiments, for a given vortex tube, the input air pressure was regulated through the supply valve and varied from 1 to 4 bar with an increment of 0.5 bar. In order to minimize the errors encountered in the measurements, the following procedure has been followed in all tests and for all vortex tubes. For a given inlet pressure, the steady-state condition was obtained first by obtaining a steady record for the temperature on both sides (hot and cold).

1. The repeatability of all results was examined, re-measuring the temperatures at hot and cold sides, for some inlet pressure values.
2. The ambient temperature was recorded before each test and run.
3. All steps mentioned above were repeated in all tests conducted on all 10 vortex tubes, regardless of the diameter and length of the tube under consideration.

In the following table (2), some temperature measurements at hot and cold sides, for a given vortex tube (14 mm inner diameter, 100 cm length and 10% conical valve opening) are presented.

The cold and hot tempartures are measured for three values of inlet pressure, form 1.5 bar to 2.5 bar. For each set of data, the average value is calculated, and the absolute errors, were estimated with respect to the average value of each temperature measurement.

The temperature records shown in the table, indicate that, the percentage of errors varies from 1 % to 6%. Other results but not shown here for the sake of brevity, have indicated that the maximum absolute errors for inlet pressure and flow rates were; 5 % and 6.5 %, respectively. It could be concluded that the variations and the percentages of errors are very little small and can be neglected, which is acceptable in any experimental work. As mentioned before, for a given vortex tube, i.e. one of the tubes listed in Table (1), first the inlet pressure is set to 1 bar, which is the minimum value of pressure. While the cold and hot air streams temperatures are measured as well as the mass flow rate at the inlet and at the hot and cold sides, at a given value of the conical valve opening. For each vortex tube and at a given percentage of conical valve opening, the procedure is repeated for different values of inlet pressure, i.e. by increasing the inlet pressure by 0.5 bar. As an example, in Table (3) below, results of the vortex tube No. 4 from table (2) are shown for 10% conical valve opening.

Table 1. Dimensions of geometrical parameters employed in the study

Sample Number	Hot Tube		Cold Tube		Orifice Diameter (mm)	Nozzles Diameter (mm)
	Inner Diameter (mm)	Length (cm)	Inner Diameter (mm)	Length (cm)		
1	14	25	14	20	7.0	3
2		50				
3		75				
4		100				
5	21	25	21		10.5	
6		50				
7		75				
8		100				
9	25	100	25		12.5	
10	32	100	32		16	

4. Discussion of the Results

Since the conical valve is mounted on the hot side exhaust tube, it was noticed that when the conical valve is fully closed, all inlet air will escape from the cold side in the form of unsteady air currents at different temperatures, which is an expected behavior. On the other hand, when the conical valve is completely opened, all the compressed air will exhaust from the hot side, and the ambient air also goes inside from the cold side due to pressure drop formed by the tangential movement "circulation " of the injected compressed air into the vortex tube. Thus, the effect of the percentage of conical valve opening on the vortex tube operation and performance was investigated in this work. The opening was gradually increased by 10% in each step. For a given percentage of opening and inlet pressure, the temperatures and anemometer reading were recorded, and the corresponding mass flow rates were calculated.

4.1. Effects of conical valve opening on temperature difference

In Fig. 5, the temperature difference between inlet and cold side air stream (Ti-Tc) versus inlet pressure, for different values of conical valve opening (10%, 20%, 30% and 40%) is presented. The tested vortex tube was of 14 mm inner diameter and 100 cm hot tube length. While Fig. 6 presents the effect of tube opening on the coefficient of performance of the vortex tube expressed as percentage ratio to maximum possible coefficient of performance that may be obtained by Carnot ideal cycle (or second law efficiency). It was found that the coefficient of performance ratio decreases with the increase of conical valve opening for all inlet pressures, which means that coefficient of performance is maximum at 10% opening, corresponding to cold volume flow rate of (0.27 - 0.34) m^3/min. While the maximum cold air temperature difference for same vortex tube occurred at 20% of conical valve opening, regardless of inlet pressure value, which corresponds to cold volume flow rate of (0.11 - 0.15) m^3/min. For this vortex tube, the maximum cold temperature difference recorded was 17.1 °C at 4.0 bar inlet pressure and 20% conical valve opening, with coefficient of performance of 2.1%. The maximum second law coefficient of performance was only 5.3% with 7.9 °C cold temperature difference at inlet pressure of 1.5 bar and 10% conical valve opening.

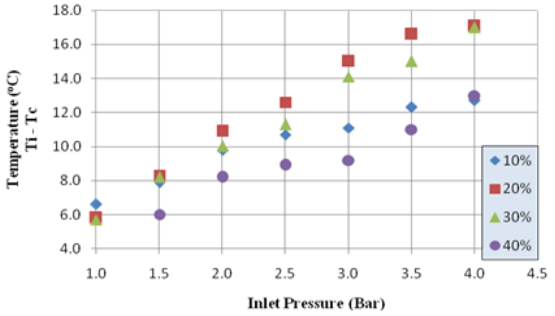


Figure 5. Temperature differences versus inlet pressures for a vortex tube of 14 mm inner diameter and 100 cm hot tube length for different values of conical valve opening

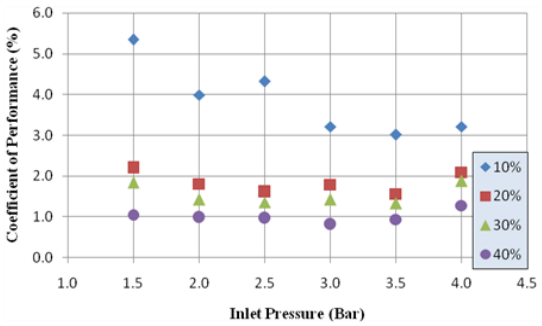


Figure 6. Coefficient of performance versus inlet pressures for a vortex tube of 14 mm inner diameter and 100 cm hot tube length for different conical valve opening

Figure 7 shows the results obtained by operating a vortex tube of 21 mm inner diameter and 100 cm hot tube length, for this vortex tube the cold temperature difference is largest at 20% of conical valve opening only for inlet pressures larger than 2.0 bar or in other words; for cold mass fraction between (0.07 - 0.10) m^3/min.. Note that this vortex tube needs more than 4.0 bar inlet pressure to operate at more than 30% of conical valve opening.

Figure 8 below gives a good view about the effect of conical valve opening on the coefficients of performance. Although coefficient of performance is higher at 10% of conical valve opening with cold volume flow rate fraction between (0.16 - 0.18) m^3/min., it drops down with increasing inlet pressure. On the contrary, the coefficient of performance increases with increasing inlet pressure at 20% of conical valve opening, and it is almost negligible for 30% of conical valve opening.

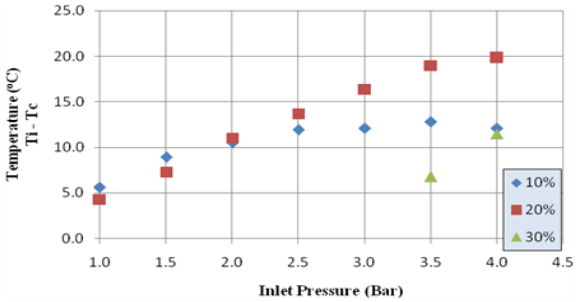


Figure 7. Cold temperature differences versus inlet pressures for a vortex tube of 21 mm inner diameter and 100 cm hot tube length for different conical valve opening.

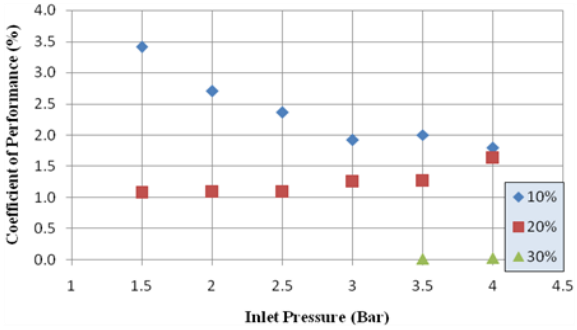


Figure 8. Coefficient of Performance versus inlet pressures for a vortex tube of 21 mm inner diameter and 100 cm hot tube length for different conical valve opening.

The maximum temperature drop in this vortex tube equals to 19.8 ºC which occurred at 4.0 bar inlet pressure and 20% of conical valve opening, the coefficient of performance in this case is 1.6%. The maximum coefficient of performance is 3.4% with cold temperature difference equals to 8.9 ºC which was recorded at 1.5 bar and 10% of conical valve opening.

The third tested vortex tube was of 25 mm inner diameter and 100 cm hot tube length. The maximum cold temperature difference was 16.3 ºC with 0.5 % coefficient of performance ratio at 4.0 bar at 20% of conical valve opening. On the other hand, the maximum coefficient of performance was 0.7% with 4.6 ºC cold temperature difference at 1.5 bar and 10% of conical valve openings. Those results are shown in Figures 9-10.

The last tested Vortex tube was of 32 mm inner diameter and 100 cm hot tube length. It operated properly under inlet pressures below 4.0 bar when the percentage of conical valve opening was below 10%. For larger openings, pressures more than 4.0 bar were needed to operate it. So that, it produced a maximum cold temperature difference of 10.6 ºC with maximum coefficient of performance of 0.5% at 4.0 bar inlet pressure and 10% of conical valve opening.

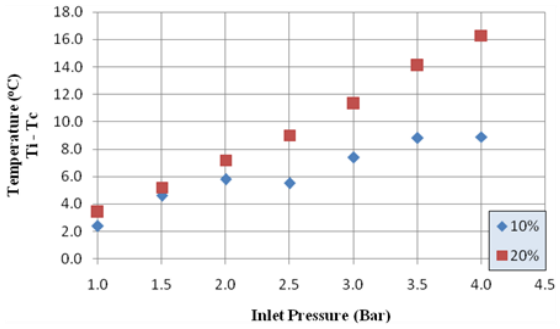


Figure 9. Cold temperature difference versus inlet pressures for a vortex tube of 25 mm inner diameter and 100 cm hot tube length for different conical valve opening.

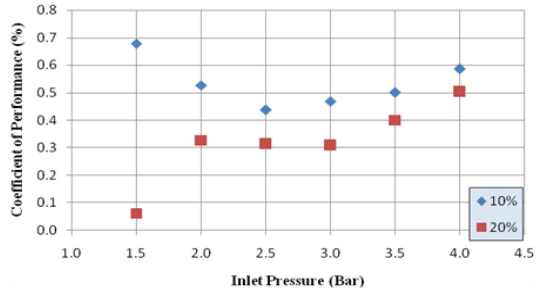


Figure 10. Coefficient of Performance versus inlet pressures for a vortex tube of 25 mm inner diameter and 100 cm hot tube length for different conical valve opening.

Based on above, it may be stated that the coefficient of performance for any vortex tube is maximum at conical valve opening of 10%, while the cold temperature deference is maximum at 20% valve opening for all inlet pressures as shown in table 4.

4.2. Effect of inlet pressure

The inlet flow pressure has been investigated. Figures 11-13 show the variations of cold air temperature difference versus the conical valve opening for different inlet pressures. It may be concluded that increasing the inlet pressure causes the cold air temperature difference to increase, as a result of increasing inlet angular momentum or centrifugal force, i.e. mixing rate. It is important to note that there is critical percentage of conical valve opening for each vortex tube which depends on tube's diameter. At this percentage, the cold temperature differences starting to decrease regardless of inlet pressures. For further details see table 5.

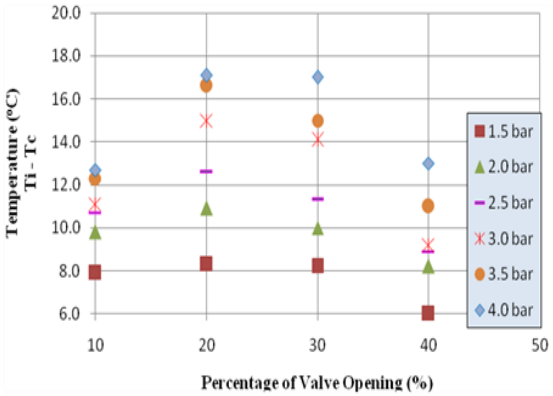


Figure 11. Cold temperature differences versus percentages of conical valve opening for different inlet pressures at a vortex tube of 14 mm inner diameter and 100 cm hot tube length

Table 2. Absolute Errors in temperature measurements (14 mm inner diameter, 100 cm length and 10% conical valve opening)

Trial Number	Cold temperature (ºC)	Error (%)	Hot temperature (ºC)	Error (%)
Inlet Pressure equals to 1.5 bar:				
1	9.8	2.97	23.5	1.29
2	10.4	2.97	23.4	0.86
3	10.3	1.98	23.0	0.86
4	9.9	1.98	22.9	1.29
Average	10.10	2.475	23.2	1.075
Inlet Pressure equals to 2.0 bar:				
1	8.1	5.19	23.6	1.25
2	7.3	5.19	23.4	2.0
3	7.4	3.89	24.2	1.25
4	8.0	3.89	24.4	2.09
Average	7.7	4.540	23.9	1.647
Inlet Pressure equals to 2.5 bar:				
1	6.0	4.76	23.6	2.88
2	6.5	3.17	24.0	1.23
3	6.6	4.76	24.7	1.64
4	5.9	6.35	24.8	2.05
Average	6.3	4.76	24.3	1.950

Table 3. Results for vortex tube, with a hot tube of 14 mm inner diameter and 100 cm length at 10% of conical valve opening and ambient temperature 18.5 (°C).

Inlet Flow			Cold Flow		Hot Flow	
Pressure (Bar)	Temperature (°C)	Flowrate (m^3/min.)	Temperature (°C)	Flowrate (m^3/min.)	Temperature (°C)	Flowrate (m^3/min.)
1.0	18.2	0.600	11.6	0.170	20.1	0.430
1.5	18.0	0.800	10.1	0.230	23.2	0.570
2.0	17.5	1.000	7.7	0.270	23.9	0.730
2.5	17.0	1.200	6.3	0.410	24.3	0.790
3.0	15.9	1.600	4.8	0.460	24.4	1.140
3.5	15.2	1.900	2.9	0.520	24.6	1.380
4.0	14.7	2.000	2.0	0.620	25.5	1.380

Table 4. Summarize of the critical values resulted from analyzing the effects of conical valve opening.

Vortex tube diameter (mm)	Maximum cold temperature difference ΔT_c (°c)					Maximum cold Coefficient of performance ratio COP_c (%)				
	Max. ΔT_c	Opening percent (%)	Cold mass fraction	Inlet pressure (bar)	COP_c	Max. COP_c	ΔT_c	Opening percent (%)	Cold mass fraction	Inlet pressure (bar)
14	17.1	20	0.150	4.0	2.1	5.3	7.9	10	0.288	1.5
21	19.8	20	0.100	4.0	1.6	3.4	8.9	10	0.163	1.5
25	16.3	20	0.038	4.0	0.5	0.7	4.6	10	0.063	1.5
32	10.6	10	0.061	4.0	0.5	0.5	10.6	10	0.061	4.0

Table 5 . Critical percentages of conical valve opening for different vortex tube diameters

Vortex tube diameter (mm)	Critical percent of opening conical valve (%)
14	21 – 24
21	16 - 19
25	13 - 15

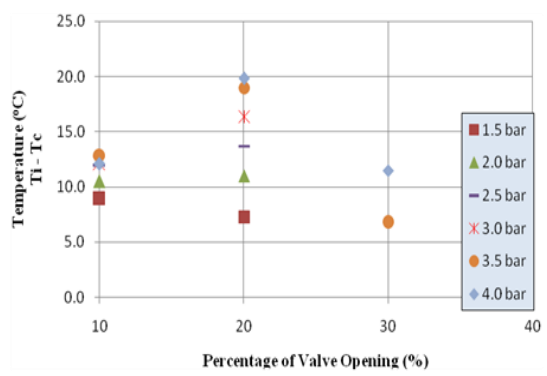


Figure 12. Cold temperature differences versus percentages of conical valve opening for different inlet pressures at a vortex tube of 21 mm inner diameter and 100 cm hot tube length.

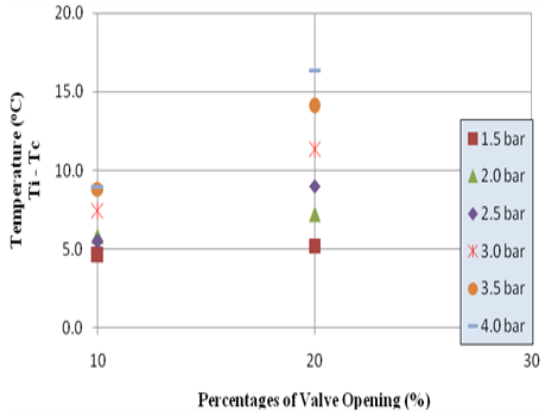


Figure 13. Cold temperature differences versus percentages of conical valve opening for different inlet pressures at a vortex tube of 25 mm inner diameter and 100 cm hot tube length.

On the contrary, the coefficient of performance is increasing with decreasing inlet pressure up to critical percentages of conical valve opening. This is clearly shown in Figures 14 - 16 below.

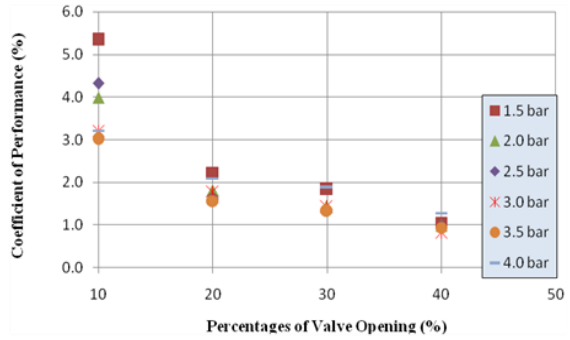


Figure 14. Coefficients of Performance versus percentages of conical valve opening for different inlet pressures at a vortex tube of 14 mm inner diameter and 100 cm hot tube length.

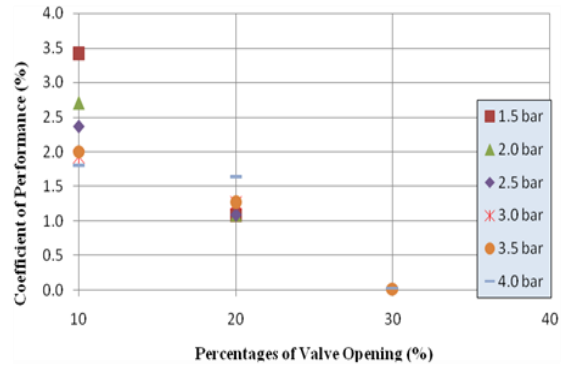


Figure 15. Coefficients of Performance versus percentages of conical valve opening for different inlet pressures at a vortex tube of 21 mm inner diameter and 100 cm hot tube length.

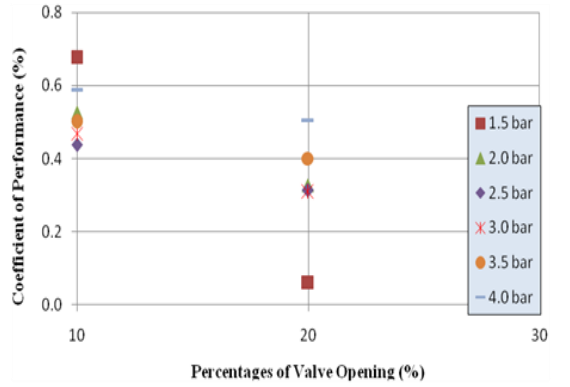


Figure 16. Coefficients of Performance versus percentages of conical valve opening for different inlet pressures at a vortex tube of 25 mm inner diameter and 100 cm hot tube length.

4.3. Effect of vortex tube diameter

Previous studies have indicated that a very small diameter of the vortex tube would offer considerably higher back pressures. Therefore, the tangential velocities between the periphery and the core would not differ substantially due to the lower specific volume of air (still high density), while the axial velocities at the core region are high. This would lead to low diffusion of kinetic energy which means low temperature separation. On the other hand, a very large tube diameter would result in lower overall tangential velocities both at the core and at the peripheral region, which would produce low diffusion of mean kinetic energy and also low temperature separation. In order to be able to study the practical effect of vortex tube diameter on its operation, we have taken 4 vortex tubes with different inlet diameters (14, 21, 25, and 32 mm). The results are shown in Figures 17-20

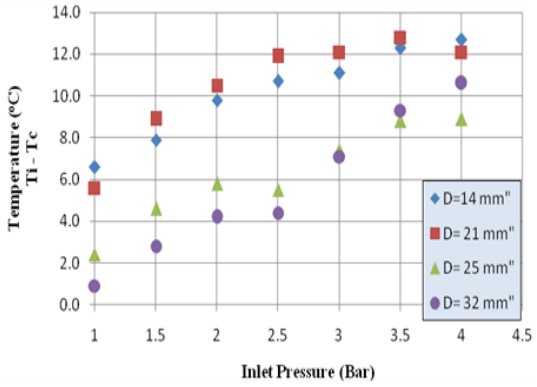


Figure 17. Cold temperature differences versus inlet pressures at 10% of conical valve opening for deferent vortex tube inner diameters.

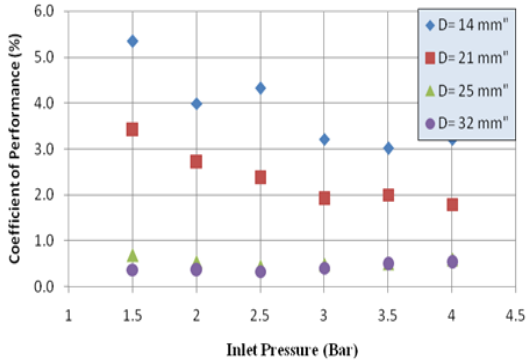


Figure 18. Coefficients of Performance versus inlet pressures at 10% of conical valve opening for deferent vortex tube inner diameters.

It is generally shown in Figures 17-18 that any increase in inner diameter of vortex tube will decrease the cold temperature deference and the coefficient of performance simultaneously. As counter to this rule, vortex tube with 21 mm inner diameter has introduced the best cold temperature differences during our experiments.

4.4. Effect of Hot tube length

In order to study the effect of length on the performance of the vortex tube, the hot tube lengths for two samples have been changed. Therefor vortex tubes with inner diameters equal 14 and 21 mm were used and their hot tube lengths were also changed to 25, 50, 75 and 100 cm.

Figures 19 - 20 below show the results of changing the hot tube length in a vortex tube of 14 mm inner diameter under 40% of conical valve opening. It is clearly seen that lengths 50 cm and 75 cm have produced the best performance regardless of the inlet pressures.

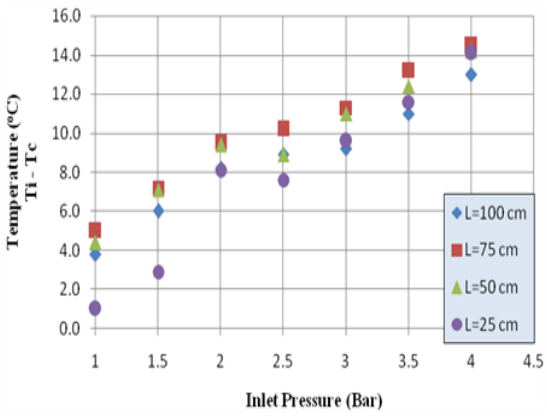


Figure 19. Cold temperature differences versus inlet pressures at 14 mm inner diameter vortex tube and 40% of conical valve opening for deferent hot tube lengths.

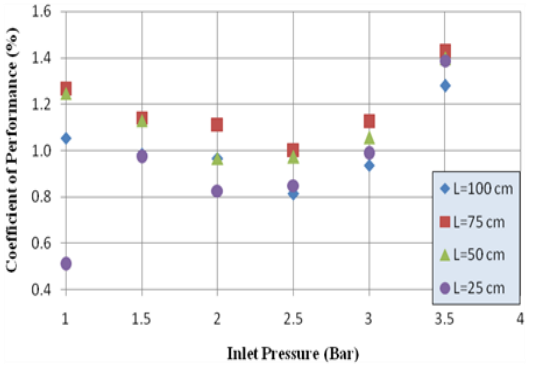


Figure 20. Coefficients of Performance versus inlet pressures at 14 mm inner diameter vortex tube and 40% of conical valve opening for deferent hot tube lengths.

Figures 21 - 23 below show the results for another vortex tube of 21 mm inner diameter operating at 20% of conical valve opening to prove the conclusions. It is clear here that hot tube length between 75 cm and 100 cm is too suitable for this vortex tube diameter.

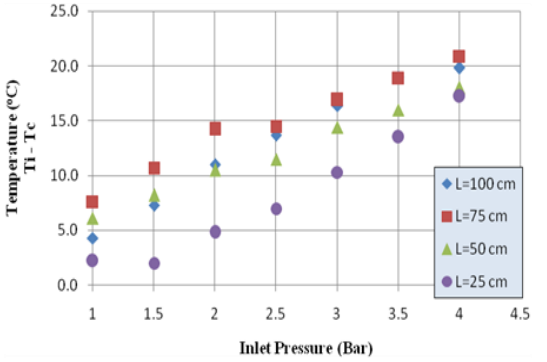


Figure 21. Cold temperature differences versus inlet pressures for a vortex tube of 21 mm inner diameter operated at 20% of conical valve opening for deferent hot tube lengths.

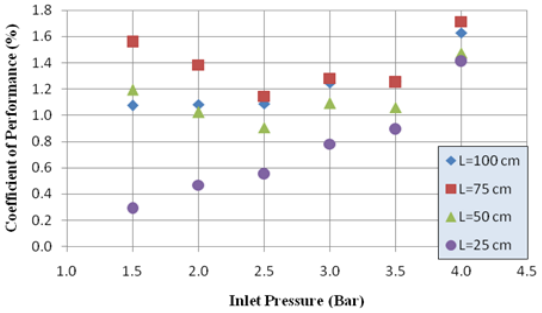


Figure 22. Coefficients of Performance versus inlet pressures for a vortex tube of 21 mm inner diameter operated at 20% of conical valve opening for deferent hot tube lengths.

As a result, using vortex tube with (L_h/D) ratio between 36 and 50 has been recommended, but using ratio equal to 36 is cost-saving - with maintaining the same level of performance - specially for vortex tubes operating at inlet pressures of more than two bars.

4.5. Effect of insulation

Figures 23-28 show the relationship between inlet pressures, temperature differences at the two exits and coefficients of performance for vortex tubes of inner diameters 14 and 21 mm operating with and without insulation. The non-insulated vortex tube provided a higher cold temperature reduction compared to the insulated one by about 2 degrees. This is due to the transition heat that has been prevented from escaping to the surroundings by the insulation being acquired by cold and hot streams.

This result has been proved in Figures 25-26 by showing the increase in hot temperature of the hot stream using insulated tube. The coefficient of performance decreases slightly with using insulation, this is clearly seen in Figures 27- 28 below.

In addition to what mentioned above, the insulation effect becomes more obvious with the increase of hot tube diameter and length due to the increase in the heat exchanging area insulated.

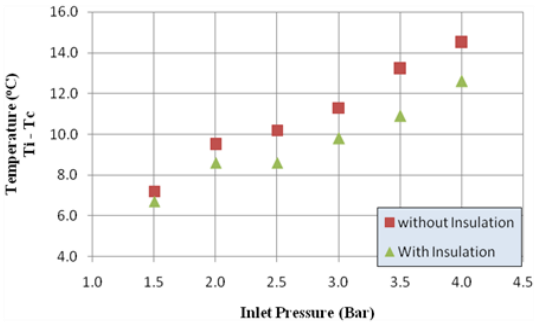


Figure 23. Cold temperature differences versus inlet pressures for a vortex tube of 14 mm inner diameter and 75 cm hot tube length operated at 40% of conical valve opening with and without insulation.

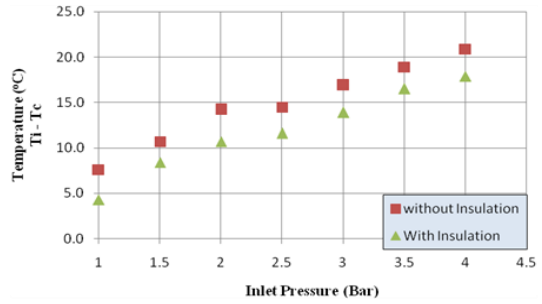


Figure 24. Cold temperature differences versus inlet pressures for a vortex tube of 21 mm inner diameter and 75 cm hot tube length operated at 20% of conical valve opening with and without insulation.

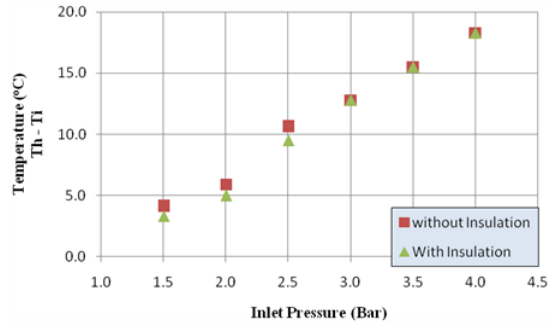


Figure 25. Hot temperature differences versus inlet pressures for a vortex tube of 14 mm inner diameter and 75 cm hot tube length operated at 40% of conical valve opening with and without insulation.

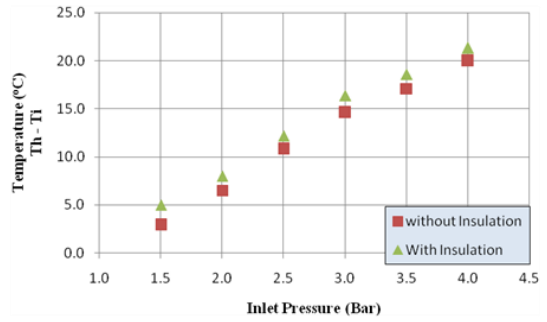


Figure 26. Hot temperature differences versus inlet pressures for a vortex tube of 21 mm inner diameter and 75 cm hot tube length perated at 20% of conical valve opening with and without insulation.

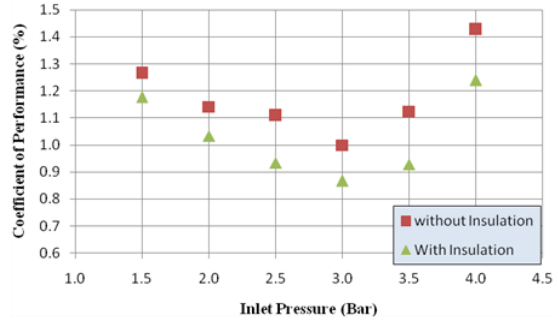


Figure 27. Coefficients of Performance versus inlet pressures for a vortex tube of 14 mm inner diameter and 75 cm hot tube length operated at 40% of conical valve opening with and without insulation.

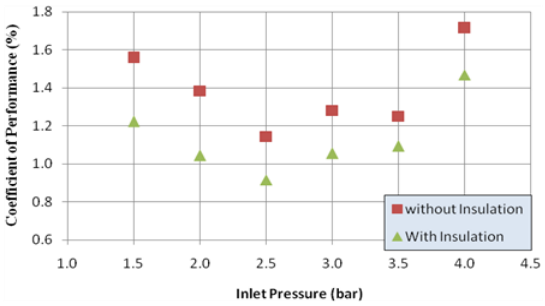


Figure 28. Coefficients of Performance versus inlet pressures for a vortex tube of 21 mm inner diameter and 75 cm hot tube length operated at 20% of conical valve opening with and without insulation.

5. Conclusions

The thermal performance of a vortex tube with conical control valve is studied experimentaly. Results were obtained for different values of: conical control valve opening, inner diameter of the vortex tube, hot and cold side lengths, inlet pressures and with and without insulation.

As a result of the presented study, it can be said that the inner diameter of vortex tube is one of the most important parameters that affect the tube performance; since any increase in vortex tube inner diameter produces a decrease in its coefficient of performance and cold temperature reduction. It was observed that the 75 cm length of the hot side length, gives the maximum coeffecint of performance for RHVT, as well as the maximum cold temperature reduction, either with or with out insulation. Moreover, the thermal performance of the RHVT, was optimum for the two inner diameters; 14 and 21mm.

This would indicate that, when the ratio of hot side length to inner diameter (Lh/D) lies between 36 and 50, then the thermal performanmce of the RHVT is optimum and it is highly recommended, which agrees with the results of published previous studies.

In general, it was found that 20% of conical valve opening produced maximum cold temperature reduction at all inlet pressures, regardless of geometrical parameters of hot tube, while 10% of conical valve opening has caused the vortex tube to operate at the maximum coefficient of performance. It is very necessary to mention that results also reflect the fact that, increasing inlet pressure causes the cold temperature reduction to increase regardless of conical valve opening.

It was found that conical valve opening is characterized by the following:

- It is inversely proportional to the inner diameter of the vortex tube.
- The vortex tube operates at maximum cold temperature reduction regardless of the inlet pressure at this range.
- At this range, the coefficients of performance converge to closed values.
- Increasing inlet pressure will decrease coefficients of performance up to these critical ranges and then the relation will be reversed, so that increasing inlet pressure will increase coefficients of performance.

The insulated vortex tube gave less energy loss to the surroundings than the non-insulated one, causing a higher hot temperature difference within the hot tube and lower

cold temperature difference in the cold tube, giving a lower coefficient of performance.

References

[1] Ranque MG. Experiences sur la de´tente giratoire avec productions simultanees dun echappement dair chaud et dun echappement dair froid. J Phys Rad 1933;7(4):112.

[2] Hilsch R. The use of the expansion of gases in a centrifugal field as cooling process. Rev Sci Instrum 1947;18(2):108.

[3] B. Ahlborny, J. Camirey, J. Kellerz, "Low-pressure vortex tubes". J. Phys. D: Appl. Phys, Vol. 29, 1996, 1469-1472.

[4] U. Behera, P. Paul, K. Dinesh, S. Jacob "Numerical investigations on flow behaviour and energy separation in Ranque–Hilsch vortex tube". International Journal of Heat and Mass Transfer, Vol. 51, 2008, 6077-6089.

[5] U. Behera, P. Paul, S. Kasthurirengan, R. Karunanithi, S. Ram, K. Dinesh, S. Jacob, "CFD analysis and experimental investigations towards optimizing the parameters of Ranque–Hilsch vortex tube". International Journal of Heat and Mass Transfer, Vol. 48, 2005, 1961-1973.

[6] K. Dincera, S. Baskaya, B. Uysa, I. Ucgu, "Experimental investigation of the performance of a Ranque–Hilsch vortex tube with regard to a plug located at the hot outlet". International Journal of Refrigeration, Vol. 32, 2009, 87-94.

[7] S. Eiamsa-ard, K. Wongcharee, P. Wongcharee, "Experimental investigation on energy separation in a counter-flow Ranque–Hilsch vortex tube: Effect of cooling a hot tube". International Communications in Heat and Mass Transfer, Vol. 37, 2010, 156–162.

[8] W. Fröhlingsdorf, H. Unger, "Numerical investigations of the compressible flow and the energy separation in the Ranque–Hilsch vortex tube". International Journal of Heat and Mass Transfer, Vol. 42, 1999, 415-422.

[9] M. Gao, K. Bosschaart, J. Zeegers, A. de Waele, "Experimental study on a simple Ranque–Hilsch vortex tube". Cryogenics, Vol. 45, 2005, 173-183.

[10] S. Im, S. Yu, "Effects of geometric parameters on the separated air flow temperature of a vortex tube for design optimization". Energy, Vol. 37, 2012, 154-160.

[11] M. Kargaran, M. Farzaneh, "Experimental Investigation the Efect of Orifice Diameter and Tube Length on a Vortex Tube Performance". International Journal of Recent advances in Mechanical Engineering, Vol. 2, 2013, 1213-1225.

[12] M. Kargaran, A. Arabkoohsar, S. Javad, V. Farzaneh, M. Farzaneh, "Second Low Analysis of Natural Gas Behavior within a Vortex Tube". Thermal Science, Vol. 17, 2013, 1079-1092.

[13] C. Linderstrøm-Lang, "Gas separation in the Ranque-Hilsch vortex tube". International Journal of Heat and Mass Transfer, Vol. 7, 1964, 1195-1206.

[14] B. Markal, O. Aydın, M. Avc, "An Experimental Study on the Effect of the Valve Angle of Counter-Flow Ranque–Hilsch Vortex Tubes on Thermal Energy Separation". Experimental Thermal and Fluid Science, 34, 2010, 966-971.

[15] J. Marshall, "Effect of operating conditions, physical size and fluid characteristics on the gas separation performance of a Linderstrom-Lang vortex tube". International Journal of Heat and Mass Transfer, Vol. 20, 1977, 227-231.

[16] R. Maurya, K. Bhavsar, "Energy and Flow Separation in the Vortex Tube: A Numerical Investigation". International Journal on Theoretical and Applied Research in Mechanical Engineering, Vol. 2, 2013, 2319-3182.

[17] H. Pouraria, W. Park, "Numerical Investigation on Cooling Performance of RANQUE-HILSCH Vortex Tube". Unpublished Doctoral Dissertation, Pusan National University, Busan, Korea, 2013, 609-735.

[18] N. Pourmahmoud, A. Bramo, "The Effect of L/D Ratio on the Temperature Separation in the Counter-flow Vortex tube". IJRRAS, Vol. 6, No. 1, 2011.

[19] J. Prabakaran, S. Vaidyanathan, D. Kanagarajan, "Establishing Empirical Relation to Predict Temperature Differences of Vortex Tube Using Response Surface Methodology". Journal of Engineering Science and Technology, Vol. 7, 2012, 722-731.

[20] M. Saidi, M. Valipour, "Experimental modeling of vortex tube refrigerator". Applied Thermal Engineering, Vol. 23, 2003, 1971-1980.

[21] M. Saidi, M. Allaf Yazdi, "Exergy model of a vortex tube system with experimental results". Original Research Article Energy, Vol. 24, 1999, 625-632.

[22] A. Secchiaroli, R. Ricci, S. Montelpare, V. D'Alessandro, "Numerical simulation of turbulent flow in a Ranque–Hilsch vortex tube". International Journal of Heat and Mass Transfer, Vol. 52, 2009, 5496-5511.

[23] K. Stephan, S. Lin, M. Durst, F. Huang, D. Seher, "An investigation of energy separation in a vortex tube". International Journal of Heat and Mass Transfer, Vol. 26, 1983, 341-348.

[24] R. Shamsoddini, A. Khorasani, "A New Approach To Study And Optimize Cooling Performance Of A Ranque-Hilsch Vortex Tube". International Journal of Refregaeration, Vol. 10, No. 1016, 2012.

[25] A. Eneizat, " Numerical and Experemental Study of a Vortex Tube with Conical Control Valve ", Master Thesis, Mechanical Engineering Dept., The University of Jordan, 2014.

[26] K. Devade and P. Ashok Pise, "Parametric Review of Ranque-Hilsch Vortex Tube", American Journal of Heat and Mass Transfer, Vol. 4, No. 3, 2017.

[27] S. Eiamsa-ard and P. Promvonge, "Review of Ranque–Hilsch effects in vortex tubes", Renewable and Sustainable Energy Reviews, Vol. 12, No1, 2008

[28] A. Nayak, P. Satapathy, S. Sahoo and I. Mahapatra, "Fluid flow and performance analysis of vortex tube: a computational approach", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, DOI:10.1080/15567036.2019.1624880, 2019

[29] R. UdayaKumar and P. Solanki, "Efficiency of Vortex Tubes in spot cooling", Journal of Physics: Conference Series on Recent Advances in Fluid and Thermal Sciences, Doi:10.1088/1742-6596/1276/1/012014, 2019

[30] A. Al-Salaymeh and M. R. Abdelkader, "Efficiency of Free Cooling Technique in Air Refrigeration Systems", Jordan Journal of Mechanical and Industrial Engineering, Vol. 5, No. 4, 2011

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665
Pages 205 - 213

Jordan Journal of Mechanical and Industrial Engineering

Numerical Analysis of Combustion Characteristics and Emission of Dual and Tri-Fuel Diesel Engine under Two Engine Speeds

Radhwan Ali [a,b*], Saddam H. Raheemah[a,c], Nabeel N. Al-Mayyahi[b]

[a] Department of Mechanical Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
[b] Department of Mechanical Engineering, Universiti of Misan, Iraq
[c] Kut Technical Institute, Middle Technical University, Baghdad, Iraq

Received February 25 2020 *Accepted May 23 2020*

Abstract

The numerical simulations were performed on a single-cylinder diesel engine that operates using the direct injection technique. In this study, a two-dimensional CFD code was used in order to evaluate the emissions and combustion characteristics of a dual-fuel operation (diesel-H2, diesel-NG), tri-fuel operation (diesel-NG-H2), and normal operation of a diesel engine under different engine speeds. The percentage of diesel fuel was 100% and 50% with the remaining fraction of different mixtures of NG-H2 (100%–0%, 50%–50%, and 0%–100%). The results showed an increase in peak temperature and pressure when gaseous fuels were added and influenced directly by H2 percentage. With diesel-H2, peak in-cylinder temperature and pressure are found. The higher temperature of combustion as a result of a rising fraction of H2 in the fuel blend proves the formation of NO, whereas increasing the fraction of diesel fuel limits the increase of NO emission, and rising percentage of H2 linearly increases NO. CO emission is mostly effected by NG fraction, but the rising fraction of H2 decreases CO closer to normal diesel operation. The mixture of 50% NG and 50% H2 produces optimum stability between combustion characteristics and emissions. However, high diesel fraction content is preferable for sustaining low combustion temperature, high thermal efficiency by avoiding excessive heat loss, reduces ignition delay, and peak in-cylinder pressure.

Keywords: Diesel engine; Dual fuel engine; NG; Hydrogen;

Abbreviations

2D	2-Dimensional	CH_4	Methane
AFR_{st}	Stoichiometric Air to Fuel Ratio	LPG	Liquefied Petroleum Gas
BDC	Bottom Dead Center	CO_2	Carbon Dioxide
CFD	Computational Fluid Dynamics	CNG	Compressed Natural Gas
CAD	Crank Angle Degree	CO	Carbon Monoxide
EVC	Exhaust Valve Close	PM	Particulate Matter
HC	Hydrocarbons	H_2	Hydrogen
NO	Nitric Oxide	$\dot{m}$	Mass Flow Rate
TDC	Top Dead Center	X	Mass Fraction of Fuels
N0x	Nitrogen Oxides	NG	Natural Gas
IVC	Intake Valve Close	IVO	Intake Valve Open

1. Introduction

Pollution is greatly felt in large urban cities where heavy-duty engines and cars largely contribute to total pollutant emissions. Most of the energy consumed globally is spent on automobiles, usually supplied by diesel and gasoline because of crude oil, a source of energy with limited resources. Rigorous oil consumption has increased its risk of rapid depletion. Oil prices become unsteady year by year [1]. Thus, scientists and researchers have shifted their focus to finding alternatives to the petroleum fuels that we normally utilize. Most of these research efforts aim to achieve high energy efficiency and to lower hazardous emissions [2]. The use of gaseous fuels for internal combustion engines has long been suggested to potentially maintain engine efficiency and performance whilst reducing emissions [3]. Current studies on various alternative fuels for diesel engines have been conducted to reduce diesel fuel consumption as well as particulate and nitrogen oxide NO emissions [4]. Natural gas (NG) has been suggested as one of the most suitable alternative fuels that work not only as a petroleum fuel replacement, but also as an alternative that can lower smoke, NO, and particulate matter emissions. Moreover, high resistance to knock permits engines to work at a high compression ratio, producing high thermal efficiency. Modification of any diesel or gasoline engine to dual fuel using NG operation involves slight mechanical changes as shown in the steady increase of vehicles that use NG resulting from alteration of NG engines [5, 6]. Revisions on the features operational of diesel engine dual-fuel (diesel-NG) single-cylinder revealed the dual-fuel mode (diesel-NG) consequences in longer ignition delay compared with normal diesel engines [7, 8]. In-cylinder pressure is enhanced at high load, but they are lowered at part loads as the content of NG increases. Hence lean mixture, long ignition delay, and low flame speed at part loads. The use of dual fuel (diesel-NG) engine revealed a

* Corresponding author e-mail: Radhwan.ali@uomisan.edu.iq.

significant increase of PM, HC, and CO emissions as a result of incomplete combustion, poor mixture, and increasing the zone of flame extinction at part load [9]. Conversely, increased quantity of diesel fuel resulting in a slight reduction of NO production and can reduce CO emission via the intake of pre-heated air [10]. When a (diesel-NG) dual-fuel engine works at several engine speeds and compression ratio, it decreases the emission of CO, and thermal efficiency increases at high engine speed and compression ratio, nevertheless, poor thermal efficiency appeared at low engine load [11]. Additionally, H_2 was suggested as another good alternative fuel for internal combustion engines hence, H_2 has a high-octane number, and high auto-ignition temperature exhibition that H_2 is a more appropriate fuel for Otto engines than diesel engines. H_2 can improve engine efficiency whilst reducing emissions. It is odorless, non-toxic and can be combusted completely. It is considered a perfect medium for energy storage that can be sourced from fossil and non-fossil sources [12]. However, mixing H_2 with other fuels appears useful to achieve its benefits. An examination of the combustion behavior of a diesel-H_2 engine showed optimum peak in-cylinder pressure improvement at a 70% load that is essential for durable and safe engine performance [13]. Combustion efficiency is noticeably low with high H_2 fraction. H_2 combustion efficiency depends on engine load. Hence, when operating under high load states, H_2 must be added to achieve high conversion efficiencies for H_2 and diesel fuels. As well as, increase the stability of the engine and decreased unburned hydrogen. conversely, hydrogen's rapid burning rate led to the reduced ignition energy and increased diffusivity that made the combustion unstable at part load [14]. When H_2 fraction is increased, HC/CO/CO_2/PM emissions decrease almost linearly. This behavior indicates that decreases in particle and carbon-based gaseous emissions are affected by the amount of H_2 being added [12,15]. However, under low to middle load states, NO emission decreases. At high load state, NO emission increases as a result of H_2's fast-burning level that results in high combustion temperatures and better NO formation. Besides, engine efficiency is decided by its speed, load, and the H_2 fraction [16]. Increasing H_2 fraction increases the consumption of fuel whilst decreasing thermal efficiency [17]. The processes of ignition of the spark-ignition engine with a mixture of NG-H_2 as fuel are widely examined, and sure limitations related to the efficiency and removed emissions [18]. Ignition delay decrease has been attained with alteration of partial O_2 pressure in the air [19], which influences the combustion of normal diesel. H_2 is used to improve the rate of combustion and expand the limit of NG lean [20]. An increasing fraction of H_2 linearly in an NG-H_2 engine increases the mixture's laminar flame speed [21]. The benefit of H_2 mixing was found to reduce ignition delay as well [22].

The works mentioned above only conducted studies on the use of either NG or H_2 as a secondary fuel. Given its lean operational capacity, H_2 can help maintain efficient and stable combustion. It also produces low amounts of hydrocarbon and greenhouse gas [23]. Thus, an engine can perform better if H_2 is added to NG to create a secondary fuel. Lata et al. [24,25] recently conducted several experimental and theoretical investigations to evaluate the performance of a dual-fuel engine that used a mixture of LPG-H_2 as the main fuel and diesel fuel as pilot fuel. These studies revealed that efficiency could be improved at low load condition states in a dual-fuel operation when H_2 and LPG are mixed to serve as the secondary fuel. Thus, NG and H_2 complement each other in flame stability, emissions, and heating value. NG mixed with H2 may be useful in reducing CO2/HC emissions and extending the flammability range. The addition of H_2 to NG could result in a stable flame and improve the fuel's volumetric burning velocity. Therefore, NG and H_2 mixture could achieve high energy and low emission levels [26]. Nevertheless, the presence of NG decreases the combustion temperature of H_2 and suppresses the emission of NO. Hence, NG has low flame propagation speed, and narrow flammability can make combustion of H_2 smoother and steadier and can thus help prevent insufficient combustion [25]. This study will compare the emission features and combustion characteristics of normal diesel fuel, diesel-NG, diesel-H_2 dual fuel, and diesel-NG-H_2 tri-fuel engine under different gas substations and different engine speeds.

2. Studied Engine

The engine specs [27] and gaseous fuel properties are presented in Tables 1 and 2, respectively. In the current study, the engine was operated under diesel engine, dual and tri-fuel modes under two engine speeds 1500, and 2000 rpm. Additionally, at atmospheric pressure, intake temperature (298 K) and torque (20,18 Nm) were normally taken. Natural gas was presumed to be 100% methane according to the literature [28-29]. Diesel-NG and diesel-H2, under the dual-fuel mode, and diesel-NG-H2 under tri-fuel mode were selected in this study. Thus, the diesel quantity was 50% of the total fuel and the other 50% was substituted by hydrogen and methane. The input power is utilized to be similar for diesel, dual, and tri-operations since the engine we are working under the same processes. Adding the gaseous fuel into a diesel engine demands different techniques in order to identify the dual or tri-fuel operation. These techniques include the carburetion, direct injection, continuous manifold induction, and timing-controlled manifold/port injection. In this study, the diesel was directly injected into the cylinder, and the gaseous fuel was noticeably aspirated into the intake port in combination with air. Hence, the gaseous fuels were recognized as a mass fraction of air in FLUENT. The fresh air in the cylinder consisted of N2 and O2 with a mass fraction of 76.8% and 23.2% respectively, then to present the mass fraction of air, methane, and hydrogen, the subsequent expression is utilized:

$$X_{h2=}\, \dot{m}_{h2}/\dot{m}_{h2} + \dot{m}_{NG} + \dot{m}_{air} \quad (1)$$

Where, $\dot{m}$ was the gaseous mass flow rate and $\dot{m}_{air}$ can calculate by:

$$\dot{m_{air}} = \lambda * AFR_{Dst} * \dot{m}_D + \lambda * AFR_{st.gaseous} * \dot{m}_{gaseous} \quad (2)$$

Where λ was the exceeding air and AFR_{Dst} matching to the stoichiometric air/fuel ratio for diesel (by mass), and $AFR_{st.gaseous}$ matching to the stoichiometric air/fuel ratio (by mass) for gaseous fuels which can be achieved from Table 2.

Table 1. Engine specifications [37]

Mode	Lister–Petter TS1
General details	Single cylinder, 4-stroke, air-cooled, direct injection, compression Ignition
Bore _ stroke	95.3 mm * 88.9 mm
Connecting rod length	165.3 mm
Compression ratio	18:1
Fuel injection timing	13º BTDC
Fuel injection pressure	250 bar
Rated power output	4.5 kW at1500 rpm
Orifices _ diameter	4 * 0.25 mm
Piston type	Cylindrical bowl (diameter: 45 mm and depth: 15 mm)
IVO	36º CA before TDC
IVC	69º CA after BDC
EVO	76º CA before BDC
EVC	32º CA after TDC

Table 2. Properties of various gaseous fuels (20 -C, 1 bar)

	Diesel-NG	NG50-$H_2$50	Diesel-H_2
Density/(kg/m3)	0.65	0.37	0.0837
LHV/(MJ/kg)	50.02	57.79	119.93
Stoichiometric AFR/(kg air/kg fuel)	17.25	25.82	34.39
Flammability limits/(Vol.% in air)	5.0-15.0	---	4.0-75.0
Flammability limits (AFR)	11.2-37.7		5.1-369.9
Thermal conductivity (10^{-2}W/mk)	2.42		4.97
Laminar burning velocity/(m/s)	0.27	---	28
Auto ignition temperature/(K)	813	---	858

3. Computational Models and Numerical Setup

CFD simulations were implemented on a diesel engine, single-cylinder, and direct injection. The GAMBIT software is utilized to generate the entire computational domain of the engine which includes only part of expansion, strokes and compression were also considered. The computational domain consists of the geometry of the combustion chamber without any ports or valves. Commercial fluid dynamics software FLUENT model 15 is used to manage governing equations and post-processing findings. The model of RNG k-ε was implemented for physical modeling according to [30,31]. The program depends on the method of pressure correction and the PISO algorithm is used. Furthermore, applied for energy, turbulence, and momentum equations, the second-order upwind differencing scheme based on Abdullah et al. [29] then in the same order for the rest of the equations are selected since it yields better results, especially for complex flows and tri/tet grids [32] temporary discretization. Initially, the implicit time integration equation can be commonly solved at an all-time level and thus has the benefit of being stable concerning the time size step. Besides, the implicit time equation is most appropriate in the density-based explicit solver situation [43]. In agreement with earlier studies, this study used the implicit method and the pressure-based solver [29,33]. Based on Jafarmadar [34] the wave method was chosen for the primary breakup method (liquid atomization) and the secondary breakup method (drop breakup), for the spray model. The emissions of NO are modeled through using extended Zeldovich mechanism. There are two acknowledged mechanisms that Jayashankara and Ganesan use here [35]. That usage nitric oxide through the combustion process: Thermal NO and Prompt NO.

3.1. Chemical kinetic mechanism

The eddy breaks up is used as the foundation for the modeling of combustion. The combustion of diesel includes a large number of reaction steps and chemical species, and for which mechanism was developed, it is based on the decreased alkanes combustion mechanism of Jones and Lindstedt [36]. Diesel reaction ($C_{12}H_{23}$) includes four global sequences of reactions defined by Eqs. (3-6).

$$C_{12}H_{23} + 6O_2 \longrightarrow 12CO + 11.5H_2 \quad (3)$$

$$C12H23\ 12 + H2 \longrightarrow 12CO + 23.5\ H2 \quad (4)$$

$$H2\ 0.5 + O2 \longleftrightarrow H2O \quad (5)$$

$$CO + H2O \longleftrightarrow CO2 + H2 \quad (6)$$

The mechanism reaction and Methane involve the three steps of global reaction

$$H2 + 0.5\ O2 \longrightarrow H2O \quad (7)$$

$$CH4\ 0.5 + O2 \longrightarrow CO + 2\ H2 \quad (8)$$

$$CO + 0.5\ O2 \longrightarrow CO2 \quad (9)$$

One reaction step involves hydrogen and reaction mechanisms [37].

$$H2 + 0.5\ O2 \longrightarrow H2O \quad (10)$$

The mixtures are supposed to obey the law of ideal gas. The computation of the specific heat, thermal conductivity, and viscosity of the mixture as temperature functions are all depending on the characteristics of each species. It is also suggested that the combustion products of mass fractions followed both the instantaneous and the local values of thermodynamic equilibrium. The equivalence ratio, temperature, and pressure are strongly affected by the cylinder's equilibrium composition [26]. The chemical species proposed for the calculation of CFD are CO_2, O_2, N_2, H_2O, H_2, NO, and CO, in which the natural gas is 100% methane (CH_4).

3.2. Grid generation

Firstly, the independence mesh examination was prepared for the Omega combustion chamber (OCC) bowl. Two measurement meshes had been created, specifically medium and fine. At an engine speed of 2000 rpm, the simulation was done for each mesh to illustration independence. Figure 2 indicates the expected in-cylinder pressure against the crank angle graph. The results were apparent that there is no marked difference between the medium and fine meshes. Indeed, the refined mesh raises the CPU costs by three times and meanwhile was not given beneficially for the CFD simulation. Hence, considering the computational time, the medium mesh was selected as the most suitable for this research.

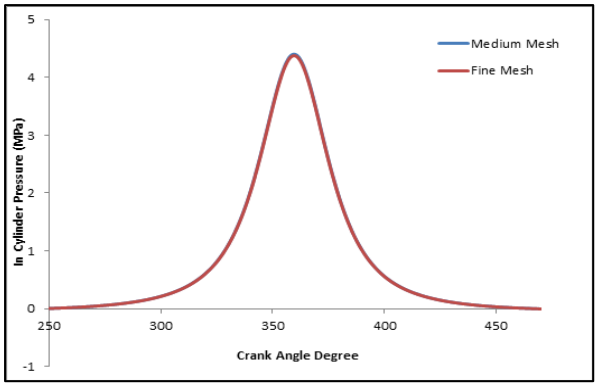


Figure 1. Calculated in-cylinder pressure results using mesh fine and medium.

3.3. Validation model

For validation, an engine model single cylinder was constructed and simulated. Engine description and further engine information utilized from the engine specifications of [27] were imported under specific conditions the speed engine (1500 rpm), temperature (330 k), and 100% diesel fuel. A 2D engine model was built and validated by comparing it with a set of experimental data for diesel operation. Moreover, to validate the simulation, the in-cylinder pressure rate was compared with the experimental results. The in-cylinder pressure validation in this study is presented in Figure 2. Generally, at most points, simulation data appear to achieve a good agreement with experimental data.

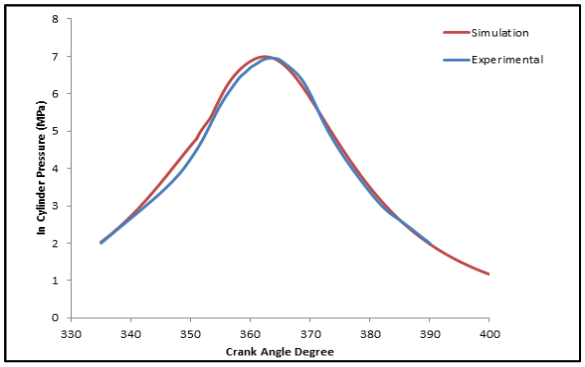


Figure 2. Validation of 2D simulation at 1500-rpm engine operation mode for In-cylinder pressure

4. Results and discussions

This section shows the numerical results of diesel, dual-fuel, and tri-fuel modes. Fuels were mixed with 100% NG and 100% H_2 for dual-fuel mode, and these percentages were alternated with 50% NG and 50% H2 for the tri-fuel mode under 1,500 and 2,000-rpm engine speeds. Using gaseous fuel will negatively affect the volumetric efficiency of the studied engine, hence, produced power reduced. In the study in hand, the effect of volumetric efficiency when using gaseous fuels with diesel ignored since the main aim is to investigate the influence on exhaust emissions.

4.1. In-cylinder pressure analysis

Figure 3 presents the effect of two selected engine speeds (1500 rpm-2000 rpm) on the peak in-cylinder pressure under pure diesel fuel, (diesel-NG, diesel-H_2) dual-fuel operations, and (diesel-NG-H_2) tri-fuel operations. The data in Figure 3 show there is no significant difference in peak in-cylinder pressure from comparing the two selected engine speeds at different types of fuel that used. However, the data also show an improvement in the peak pressure with gaseous addition to diesel fuel at two engine speeds. Furthermore, for 1500 rpm engine speed, peak pressure was improved from 7.12 MPa to 11.34 MPa (59% increase), 12.27 (72% increase) and 12.75 MPa (79% increase) for diesel fuel mode and adding NG in the following amounts, diesel-NG, diesel-NG-H_2, and diesel-H_2, respectively. Therefore, an increasing at fraction of H2 in the blend of fuel led to an increase in-cylinder pressure and advanced the occurrence of peak pressure due to the high content of specific energy and considerably faster flame as stated by Ref [37]. While the existence of H_2 increased in the blend, delays in ignition were shortened to reduced values than those of pure diesel. By contrast, the peak in-cylinder pressure decreased as the diesel fuel fraction increased due to the low flammability of the diesel fuel according to Ref [38]. The addition of pure NG to diesel enhanced ignition due to greater NG ignition power compared with diesel fuel.

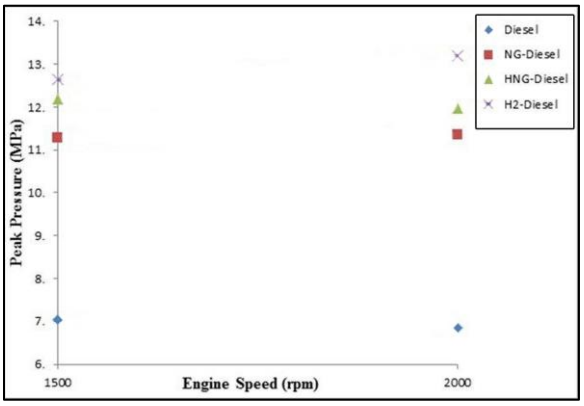


Figure 3. Effect of different ratio of gaseous addition on peak in-cylinder pressure under deferent engine speed.

Figure 4 summarizes the effects of altered mixtures of gasses fuel on peak in-cylinder pressure under two engine speeds. In the existence of H_2, the peak in-cylinder pressure was enhanced as previously demonstrated as a result of the high content of specific energy and increasingly fast flame. The combustion of H_2 increases the burning rate of diesel when combined [39]. The extensive premixed combustion phase indicates that the combustion of H_2 occurred through the premixed combustion phase of diesel fuel. Consequently, a long-premixed combustion period and a brief combustion stage of diffusion were accomplished. A fast decrease in the ignition delay was noted for diesel-H_2 combustion, which is in line with the preceding study Szwaja and Grab-Rogalinski [40]. In the current research, the shortened ignition delay is explained through two potential causes. Firstly, H_2 auto-ignition via the hot spot surface may happen under extreme circumstances or H_2 pre-ignition [41]. Secondly, a significant quantity of H_2 addition is likely to modify the chemistry of diesel oxidation

and ignition delay, thus increasing at high temperatures as earlier described by [42]. As noted in the combustion of diesel-NG, the increase of peak pressure was much smaller than that of the diesel-H_2 blend. In this case, the source of ignition was imparted by diesel, and the subsequent propagation of the flame was based on the combustion of the pre-mixed gaseous fuel-air blend. The additional significant reason for this finding is the alteration between H_2 and NG in flame propagation speeds. NG with low-speed flame propagation reduced the impact on cylinder pressure and subsequently heat release rate. Long ignition delay of NG influenced the initial phase of pressure increase in all cases.

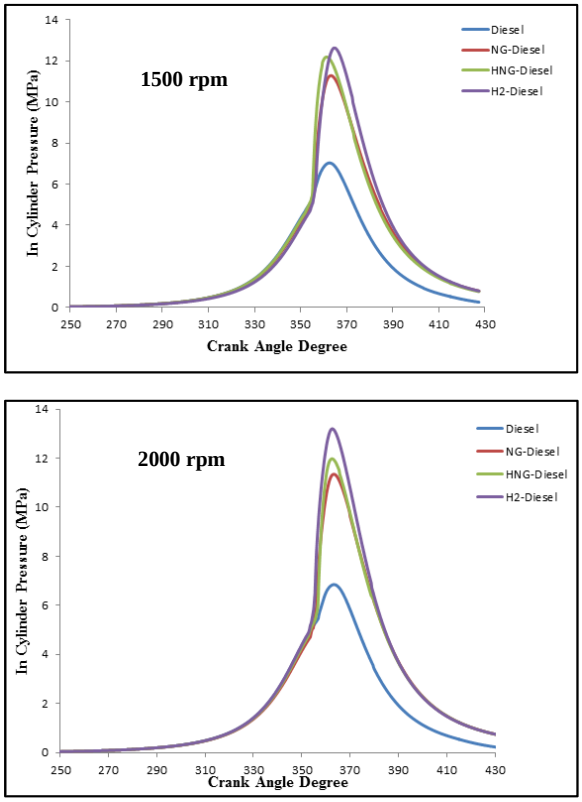


Figure 4. In-cylinder pressure curves under the different ratio of gaseous addition and engine speed

4.2. In-cylinder Temperature analysis

Figure 5 illustrates the peak in-cylinder temperature for different gases fuel substations and the different values of selected engine speeds (1500, and 2000 rpm). The data indicate there is no significant difference in peak temperature from comparing the two engine speeds at different types of fuel used. For 1500 rpm engine speed, the peak in-cylinder temperature was increased with the existence of each fuel gaseous and was obvious with an elevated H_2 percentage because hydrogen's fast-burning level led to the high combustion temperature as stated previously. The peak in-cylinder temperature enhanced from 1490 K to 2365 K (58.7% increase), 2591 K (73.9% increase), and 2787.7 K (87.1 increase) for diesel fuel mode and adding NG in the following amounts, diesel-NG, diesel-NG-H_2, and diesel-H_2, respectively. On the other hand, the peak in-cylinder temperature at diesel fuel mode decreased because of the low flammability of diesel combustion as previously confirmed.

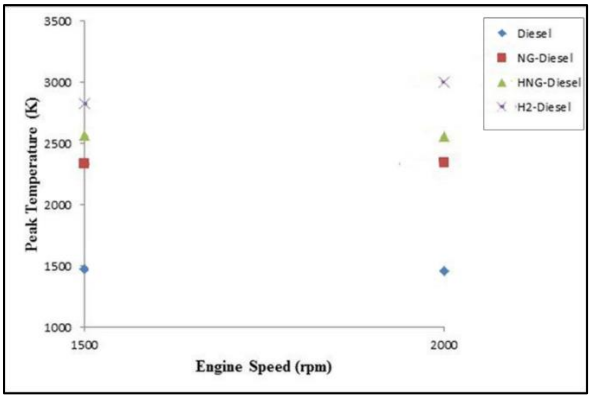


Figure 5. Effect of different ratio of gaseous addition on peak in-cylinder temperature under deferent engine speed.

Figure 6 shows the effect of altered blends mixture on the in-cylinder temperature. The variation in diesel-H_2 and diesel-NG combustion under the similar substitution ratio of energy was largely due to distinct H_2 and NG flaming propagation speeds. The NG flame propagation speed was considerably lower than that of H_2, leading to its slighter impact on temperature rate and in-cylinder pressure [34]. Furthermore, a significant ignition delay for diesel-NG was noted in all engine operating conditions. The NG presence advantage is the enhancement combustion of H_2 by preventing uncontrolled combustion, for example, a significant increase of peak pressure and temperature as illustrated in the literature.

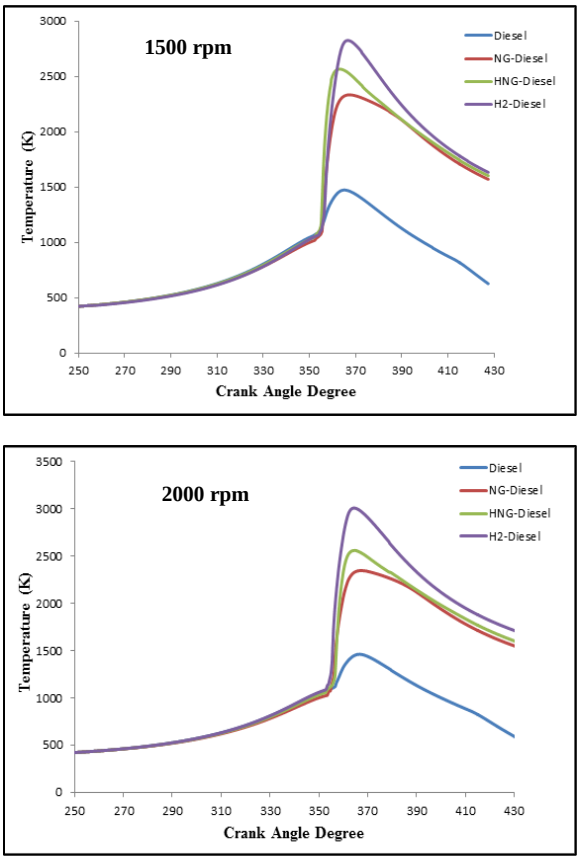


Figure 6. Temperature curves under the different ratio of gaseous addition and engine speed

Figure 7 displays the temperature contours of altered mixtures under an engine speed of 2,000 rpm. High H_2 content resulted in a large fraction of the high concentration red zone of temperature, which expanded further to the cylinder squish area. Similarly, but with decreased magnitude, diesel-NG combustion reached greater temperatures relative to the diesel fuel combustion engine.

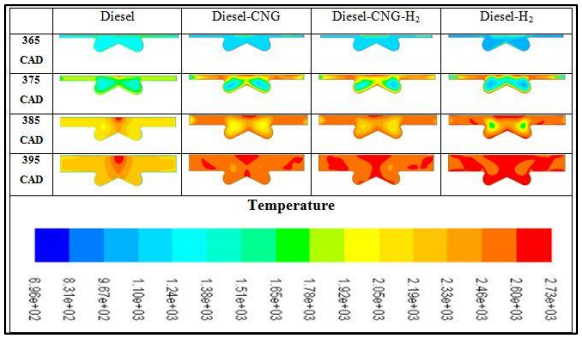


Figure 7. The development of average temperature under the different ratio of gaseous addition and 2000 rpm

5. Emissions

5.1. NO emissions

Fig. 8 demonstrates the variation of NO emissions with a different value of engine speed and gases fuel substations. The data shows there is no significant difference in NO emissions at two selected engine speeds at altered types of fuel used. For diesel-NG operation, the emission of NO slightly increased compared with the operation of diesel fuel for two engine speeds as a result of the low temperature of diesel fuel combustion according to Ref [38]. The addition of H_2 to NG increased the NO mass fraction compared with diesel and NG, with values rising from 0.00000516 to 0.00122, 0.00298, and 0.00425 when using diesel, diesel-NG, diesel-NG-H_2, and diesel-H_2, respectively under 1500 rpm engine speed. Hence, the NO formation was enhanced, and the mixture's burning rate increased as the H_2 fraction in NG increased. According to Choi et al. [43], the rise in NO emissions when H_2 is added is dependent on H_2 fuel's flame temperature, which is higher than that of the NG fuel and diesel fuel. Moreover, the emission of NO was increased in diesel-H_2 and diesel-NG-H_2 compared with pure diesel and diesel-NG under two engine speeds. In contrast, NO emission decrease is a consequence of NG fraction increasing in the blend, as discovered by Ref [37]. The mixture has the maximum amount, average, and the lowest proportion of NG in H_2 shown the best result towards the reduction of NO emission. Consequently, NG has low flame propagation speed, and narrow flammability can make combustion of H_2 smoother and steadier and can thus help prevent insufficient combustion. Besides, NO emissions in diesel-NG-H_2 operations were smaller than in diesel-H_2 operations.

Figure 9 summarises the influences of different gaseous fuel blend and diesel on the in-cylinder of NO development. Increasing diesel fraction has eliminated the formation of NO due to low temperature of diesel combustion, whereas the best for the decrease in NO emission was studied using the highest NG or the lowest H_2 values as illustrated in the previous paragraph.

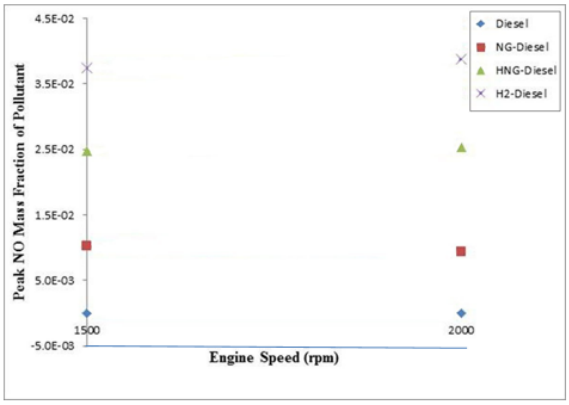


Figure 8. Effect of different ratio of gaseous addition on peak NO emissions under deferent engine speed.

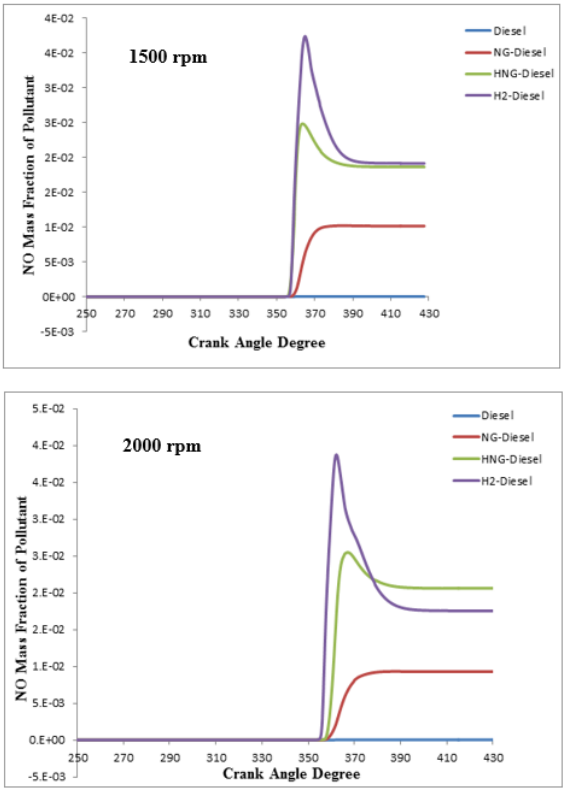


Figure 9. NO emissions curve under the different ratio of gaseous addition and engine speed

Figure 10 illustrates the contours formation of NO mass fraction at 2000-rpm engine speed. The operation of diesel-H_2 dual fuel also showed a high concentration red zone of NO distribution. The amount of formation of NO in the mode of diesel-NG dual fuel exceeded that of the normal operation of diesel fuel due to the diesel-NG combustion attained higher temperature than the normal diesel fuel. The combustion of diesel fuel engine extended to a cylinder's squish area, and the high distribution developed of NO formation at squish and swirl zones for 385° CA. However, rising diesel fraction reduced the NO formation. Another significant contributor to NO development is the equivalence ratio, where NO is maximum at stoichiometric value. However, the results indicate that while the maximum equivalence ratios were found with 50% NG and 50% H_2, temperature affects seemed to be the dominating factor.

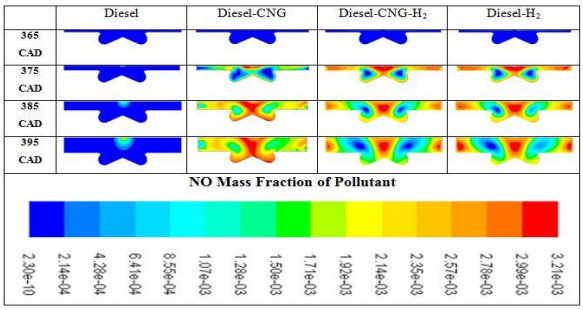


Figure 10. The development of NO mass fraction of pollutant under the different ratio of gaseous addition and 2000 rpm

5.2. CO Emissions

Incomplete combustion is the main cause of most CO emissions. Incomplete combustion occurs because of low gas temperatures, and the lack of oxidants and is overseen by the air-fuel equivalence ratio [37]. Figure 11 demonstrates the CO emission variations for operations of the diesel, dual, and tri-fuel engines under 1500 rpm, and 2000 rpm engine speeds. The figure shows there is no significant difference in CO emissions at two selected engine speeds at altered types of fuel that used. For diesel-NG operation, CO is one of the by-products of incomplete combustion of NG. CO emission increases significantly when fuels gaseous present with NG content but reductions with H2 fraction, especially compared with the addition of diesel fuel and H_2 to NG. The values rise from 0.004651 to 0.006533, 0.0325, and 0.04837 for operations of diesel fuel and with the addition of diesel-H_2, diesel-NG-H_2, and diesel-NG, respectively under 1500-rpm engine speed. Additionally, the high carbon ratio of NG could have led to an increase in CO emission. These outcomes are closer to the experimental study by Gatts et al. [44], which studied the NG combustion completion of the dual-fuel (diesel-NG) operation. Unburned CO emissions were used in their research to determine NG's combustion efficiency, which is determined by NG content and engine speed.

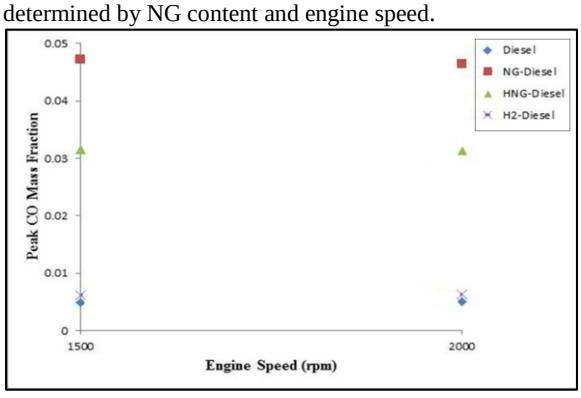


Figure 11. Effect of different ratio of gaseous addition on peak CO emissions under deferent engine speed.

Figure 12 shows the impacts of various fuel mixtures on the CO peak level with varying engine speeds. No considerable difference in CO emission was observed when the speed engine was increased. That high NG presence increases the CO peak rate, which can be related to its significantly slower flammability and the high carbon ratio of NG resulting in incomplete combustion [38]. The addition of H_2 in NG enhances NG's flammability as stated by Ref [45]. An enhancing concentration of O, H, and OH radicals was detected, rising key reaction rates to enhance NG-lean combustion efficiency with decreased CO emissions. Additionally, H_2 has a high flame velocity resulting in complete combustion that causes to reduce CO emissions. The reduction will be further decreased with the addition of H_2 in diesel. However, reducing CO emissions by increasing the diesel fraction may indicate that raised H_2 content causes CO_2 dissociation to CO. On the other hand, CO is additionally decreased at a high fraction of diesel.

Figure 13 showed the region formation of carbon monoxide at 2000 rpm under different ratios of addition gaseous. These contours illustration at 375, 385, and 395 CAD for addition H_2, as compared with diesel and NG. The formation region of CO reductions when the H_2 mass fraction addition to the gaseous and this seems obvious at diesel-H_2 dual-fuel operation and 50NG-50H_2 tri-fuel operation tables. This is due to the high flame velocity of hydrogen and resulting in complete combustion. In contrast, the formation region of CO emissions was increased as illustrated at 375, 385, and 395 CAD for diesel-NG tables. Thus, the NG has a higher carbon ratio and helps to produce CO emission.

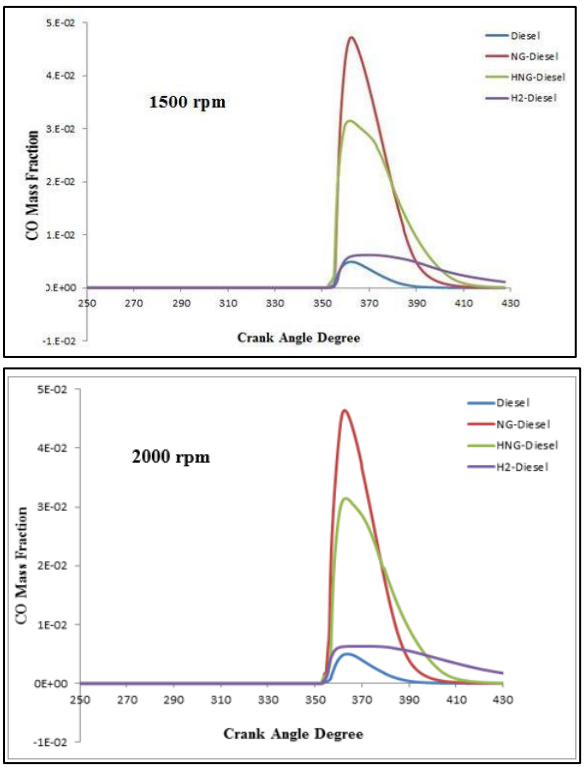


Figure 12. CO emissions curve under the different ratio of gaseous addition and engine speed

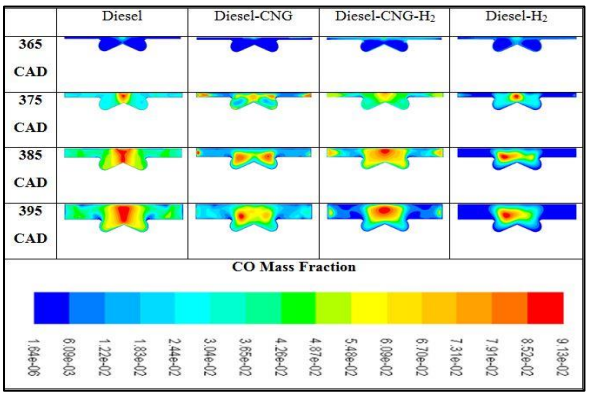


Figure 13. The development of CO mass fraction under the different ratio of gaseous addition and 2000 rpm

6. Conclusion

The present study used a direct injection method with a single cylinder to investigate the effect of using diesel with NG and H_2 at various mixture ratios under two engine speeds on a diesel engine by using a Computational Fluid Dynamics. The following conclusions are drawn from the simulation results.

- For all gaseous fuel additions under two engine speeds, in-cylinder pressure and temperature are increased as compared to normal diesel fuel operation. The improvements are related to the increase of H_2 fraction.
- Combustion is affected significantly by H_2 addition and leads to the high temperature of combustion, which can decrease the efficiency of the dual-fuel engine.
- Dual-fuel operations have a considerable influence on engine emissions. Diesel-NG mode, however, raises CO emissions, and diesel-H_2 operations raise NO emissions compared to normal diesel fuel operations.
- Tri-fuel operation involving diesel-NG-H_2 decreased CO emissions comparative to diesel-NG. It also lowered the emission of NO for all engine speed when compared to the diesel-H_2 operation. The high fraction of H_2 in NG cause to lower CO emissions. In contrast, Lower concentration of H2 in NG minimizes unwanted H_2 combustion and restricts NO emission rises.

References

[1] K. V. K. Kumar, A.S., Maheswar, D., & Reddy, "Comparision of diesel engine performance and emissions from neat and transesterified cotton seed oil," *Jordan J. Mech. Ind. Eng.*, vol. 3, no. 3, pp. 190–197, 2009.

[2] K. Srinivasa Rao, K. Bhaskara Mutyalu, and A. Ramakrishna, "Effect of exhaust gas recirculation on performance and emission characteristics of diesel engine fueled with waste cooking oil methyl ester," *Jordan J. Mech. Ind. Eng.*, vol. 10, no. 11, pp. 4799–4804, 2015.

[3] Xu J, Zhang X, Liu J, Fan L. Experimental study of a single-cylinder engine fueled with natural gas–hydrogen mixtures. Int J Hydrogen Energy 2010;35:2909–14. doi:10.1016/j.ijhydene.2009.05.039.

[4] P. Vijayabalan and G. Nagarajan, "Performance , Emission and Combustion of LPG Diesel Dual Fuel," *Jordan J. Mech. Ind. Eng.*, vol. 3, no. 2, pp. 105–110, 2009.

[5] Keskinen K, Kaario O, Nuutinen M, Vuorinen V, Künsch Z, Liavåg LO, et al. Mixture formation in a direct injection gas engine: numerical study on nozzle type, injection pressure and injection timing effects. Energy 2016;94:542–56.

[6] Kalghatgi GT. Developments in internal combustion engines and implications for combustion science and future transport fuels. Proc Combust Inst 2015;35:101–15.

[7] Papagiannakis RG, Hountalas DT. Combustion and exhaust emission characteristics of a dual fuel compression ignition engine operated with pilot Diesel fuel and natural gas. Energy Convers Manage 2004;45:2971–87.

[8] Papagiannakis RG, Hountalas DT, Rakopoulos CD. Theoretical study of the effects of pilot fuel quantity and its injection timing on the performance and emissions of a dual fuel diesel engine. Energy Convers Manage 2007;48:2951–61.

[9] Karagöz Y, Sandalcı T, Koylu UO, Dalkılıç AS, Wongwises S. Effect of the use of natural gas–diesel fuel mixture on performance, emissions, and combustion characteristics of a compression ignition engine. Adv. Mech. Eng. 2016;8.

[10] Papagiannakis RG, Rakopoulos CD, Hountalas DT, Rakopoulos DC. Emission characteristics of high speed, dual fuel, compression ignition engine operating in a wide range of natural gas/diesel fuel proportions. Fuel 2010;89:1397–406.

[11] Poompipatpong C, Cheenkachorn K. A modified diesel engine for natural gas operation: performance and emission tests. Energy 2011;36:6862–6.

[12] Saravanan N, Nagarajan G, Sanjay G, Dhanasekaran C, Kalaiselvan KM. Combustion analysis on a DI diesel engine with hydrogen in dual fuel mode. Fuel 2008;87:3591–9. doi:10.1016/j.fuel.2008.07.011.

[13] Liew C, Li H, Nuszkowski J, Liu S, Gatts T, Atkinson R, et al. An experimental investigation of the combustion process of a heavy-duty diesel engine enriched with H2. Int J Hydrogen Energy 2010;35:11357–65.

[14] Gatts T, Li H, Liew C, Liu S, Spencer T, Wayne S, et al. Heavy-duty diesel engine supplemented with H2. Int J Hydrogen Energy 2004;35(2010):11349–56.

[15] Lilik GK, Zhang H, Herreros JM, Haworth DC, Boehman AL. Hydrogen assisted diesel combustion. Int J Hydrogen Energy 2010;35:4382–98.

[16] Miyamoto T, Hasegawa H, Mikami M, Kojima N, Kabashima H, Urata Y. Effect of hydrogen addition to intake gas on combustion and exhaust emission characteristics of a diesel engine. Int J Hydrogen Energy 2011;36:13138–49.

[17] Sandalcı T, Karagöz Y. Experimental investigation of the combustion characteristics, emissions and performance of hydrogen port fuel injection in a diesel engine. Int J Hydrogen Energy 2014;39:18480–9.

[18] Korakianitis T, Namasivayam AM, Crookes RJ. Natural-gas fueled sparkignition (SI) and compression-ignition (CI) engine performance and emissions. Prog Energy Combust Sci 2011;37:89–112.

[19] Saito H, Sakurai T, Sakonji T, Hirashima T, Kanno K. Study on lean burn gas engine using pilot oil as the ignition source, SAE Technical Papers 2001-01- 0143; 2001.

[20] Tunestål P, Christensen M, Einewall P, Andersson T, Johansson B, Jönsson O. Hydrogen addition for improved lean burn capability of slow and fast burning natural gas combustion chambers, SAE Technical Papers 2002-01-2686; 2002.

[21] Zhou JH, Cheung CS, Leung CW. Combustion, performance, regulated and unregulated emissions of a diesel engine with hydrogen addition. Appl Energy 2014;126:1–12.

[22] Donohoe N, Heufer A, Metcalfe WK, Curran HJ, Davis ML, Mathieu O, et al. Ignition delay times, laminar flame speeds, and mechanism validation for natural gas/hydrogen blends at elevated pressures. Combust Flame 2014;161:1432–43.

[23] D'Andrea T, Henshaw P., Ting DS-K. The addition of hydrogen to a gasoline-fuelled SI engine. Int J Hydrogen Energy 2004;29:1541–52. doi:10.1016/j.ijhydene.2004.02.002.

[24] Lata DB, Misra A, Medhekar S. Effect of hydrogen and LPG addition on the efficiency and emissions of a dual fuel diesel engine. Int J Hydrogen Energy 2012;37:6084–96. doi:10.1016/j.ijhydene.2012.01.014.

[25] Lata DB, Misra A. Analysis of ignition delay period of a dual fuel diesel engine with hydrogen and LPG as secondary fuels. Int J Hydrogen Energy 2011;36:3746–56. doi:10.1016/j.ijhydene.2010.12.075.

[26] Miao J, Leung CW, Huang Z, Cheung CS, Yu H, Xie Y. Laminar burning velocities, Markstein lengths, and flame thickness of liquefied petroleum gas with hydrogen enrichment. Int J Hydrogen Energy 2014;39:13020–30. doi:10.1016/j.ijhydene.2014.06.087.

[27] Tarabet et al. (2014), Experimental investigation of DI diesel engine operating with eucalyptus biodiesel/natural gas under dual fuel mode. Fuel 133 (2014) 129–138.

[28] Abdullah S, Kurniawan WH, Shamsudeen A. Numerical analysis of the combustion process in a compressed natural gas direct injection engine. J Appl Fluid Mech 2008;1:65–86.

[29] Kurniawan WH, Abdullah S. Numerical analysis of the combustion process in a four-stroke compressed natural gas engine with direct injection system. J Mech Sci Technol 2008;22:1937–44.

[30] Ng HK, Gan S, Ng J-H, Pang KM. Simulation of biodiesel combustion in a light-duty diesel engine using integrated compact biodiesel–diesel reaction mechanism. Appl Energy 2013;102:1275–87. doi:10.1016/j.apenergy.2012.06.059.

[31] Abagnale C, Cameretti MC, De Simio L, Gambino M, Iannaccone S, Tuccillo R. Numerical simulation and experimental test of dual fuel operated diesel engines. Appl Therm Eng 2014;65:403–17. doi:10.1016/j.applthermaleng.2014.01.040.

[32] Didio L. Fluent - User's guide. Comput Secur 1992;11:727. doi:10.1016/0167-4048(92)90125-B.

[33] Shojaeefard MH, Noorpoor AR. Flow Simulation in Engine Cylinder with Spring Mesh. Am J Appl Sci 2008;5:1336–43. doi:10.3844/ajassp.2008.1336.1343.

[34] Jafarmadar S. Exergy analysis of hydrogen/diesel combustion in a dual fuel engine using three-dimensional model. Int J Hydrogen Energy 2014;39:9505–14. doi:10.1016/j.ijhydene.2014.03.152.

[35] Jayashankara B, Ganesan V. Effect of fuel injection timing and intake pressure on the performance of a di diesel engine - A parametric study using CFD. Energy Convers Manag 2010;51:1835–48. doi:10.1016/j.enconman.2009.11.006.

[36] W. P. Jones and R. P. Lindstedt, "Global reaction schemes for hydrocarbon combustion," Combust. Flame, vol. 73, no. 3, pp. 233–249, Sep. 1988.

[37] Masood M, Ishrat MM, Reddy AS. Computational combustion and emission analysis of hydrogen–diesel blends with experimental verification. Int J Hydrogen Energy 2007;32:2539–47.

[38] Mansor MRA, Abbood MM, Mohamad TI. The influence of varying hydrogen-methane-diesel mixture ratio on the combustion characteristics and emissions of a direct injection diesel engine. Fuel 2017;190:281–91.

[39] Zhou JH, Cheung CS, Zhao WZ, Leung CW. Diesel-hydrogen dual-fuel combustion and its impact on unregulated gaseous emissions and particulate emissions under different engine loads and engine speeds. Energy 2016;94:110–23.

[40] Szwaja S, Grab-Rogalinski K. Hydrogen combustion in a compression ignition diesel engine. Int J Hydrogen Energy 2009;34:4413–21.

[41] Verhelst S, Wallner T. Hydrogen-fueled internal combustion engines. Prog Energy Combust Sci 2009;35:490–527.

[42] Agarwal A, Assanis D. Multi-dimensional modeling of ignition, combustion and nitric oxide formation in direct injection natural gas engines, SAE Technical Papers 2000-01-1839; 2000.

[43] Ho Choi G, Jong Chung Y, Han S Bin. Performance and emissions characteristics of a hydrogen enriched LPG internal combustion engine at 1400rpm. Int J Hydrogen Energy 2005;30:77–82. doi:10.1016/j.ijhydene.2004.06.009.

[44] Gatts T, Liu S, Liew C, Ralston B, Bell C, Li H. An experimental investigation of incomplete combustion of gaseous fuels of a heavy-duty diesel engine supplemented with hydrogen and natural gas. Int J Hydrogen Energy 2012;37:7848e59.

[45] Akansu SO, Dulger Z, Kahraman N, Veziroğlu TN. Internal combustion engines fueled by natural gasdhydrogen mixtures. Int J Hydrogen Energy 2004;29:1527e39..

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665
Pages 215 - 221

Jordan Journal of Mechanical and Industrial Engineering

Finite Element Formulation of Internally Balanced Blatz – Ko Material Model

Ashraf Hadoush*

Department of Mechanical Engineering, Faculty of Engineering and Technology, Palestine Technical University – Kadoorie, Tulkarm, Palestine.

Received May 13 2020 *Accepted June 11 2020*

Abstract

Material constitutive models often include internal variables in order to capture realistic mechanical effects such as viscosity. Recent work for compressible hyperelastic material is developed based on applying the argument of calculus variation to two-factor multiplicative decomposition of the deformation gradient. The finite element formulation for this new treatment is developed, however, the implementation sheds light on a special form of constitutive model. In particular, the material model is a function of the first and third invariants of new quantities derived from the counterparts of the multiplicative decomposition. These new quantities are defined in analogy to the right Cauchy Green tensor. This work demonstrates the required treatment for a special material model that is formulated using the second and third principal invariants of these new derived quantities. Mainly, the treatment simplifies the internal balance equation that emerges from the variational treatment. This facilitates the linearization procedure of this new formulation for internally balanced compressible hyperelastic material. The present work permits the future use of more complicated internally balanced hyperelastic models.

Keywords: Hyperelasticity; internal balance; finite element;

1. Introduction

Large deformation constitutive models can be expressed in terms of deformation gradient $\boldsymbol{F}$ multiplicative decompositions [1, 2, 3] such as

$$\boldsymbol{F} = \widehat{\boldsymbol{F}}\check{\boldsymbol{F}}. \tag{1}$$

The usual treatment is that $\widehat{\boldsymbol{F}}$ models elastic response and it is associated to the rules of variational calculus. The $\check{\boldsymbol{F}}$ portion then models inelastic response usually by means of a time dependent evolution law. This multiplicative decomposition serves to pertain particular portions of $\boldsymbol{F}$ to specific parts of the material response. It has been widely used in plasticity [4, 5, 6, 7], viscoelasticity [8, 9], growth and remodeling of biological tissues [10, 11, 12, 13]. The implementation procedure of these material models in the frame of nonlinear finite element is well established [14, 15, 16, 17].

A new scheme of viewing incompressible hyperelastic material response is introduced in [18, 19]. In fact, the arguments of variation are applied to both portions of the deformation gradient decomposition. The decomposition itself is determined on the basis of an additional internal balance equation that emerges naturally from the variational treatment. Further theoretical treatment for compressible hyperelastic model is presented in [20]. The total Lagrange formulation of this new treatment is derived by linearizing the achieved weak form with respect to both portions of multiplicative decomposition [21]. Modeling material response using similar procedure presented in this paper can be found in [22, 23, 24, 25, 26].

The demonstrated implementation in [21] for compressible hyperelastic material focused on a special form of internally balance Blatz – Ko material model $W(\hat{I}_1, \hat{I}_3, \check{I}_1, \check{I}_3)$. Here $\hat{I}_1$, $\hat{I}_2$ and $\hat{I}_3$ are the principal invariants of $\widehat{\boldsymbol{C}} = \widehat{\boldsymbol{F}}^T\widehat{\boldsymbol{F}}$ (and $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}^T$) while $\check{I}_1$, $\check{I}_2$ and $\check{I}_3$ are the principal invariants of $\check{\boldsymbol{C}} = \check{\boldsymbol{F}}^T\check{\boldsymbol{F}}$ (and $\check{\boldsymbol{B}} = \check{\boldsymbol{F}}\check{\boldsymbol{F}}^T$). These new quantities $\widehat{\boldsymbol{C}}$, $\check{\boldsymbol{C}}$, $\widehat{\boldsymbol{B}}$ and $\check{\boldsymbol{B}}$ are second order symmetric tensors and they are defined in analogy with $\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F}$ and $\boldsymbol{B} = \boldsymbol{F}\boldsymbol{F}^T$. This work sheds light on the mathematical treatment that is required to implement a material model that has the form of $W(\hat{I}_2, \hat{I}_3, \check{I}_2, \check{I}_3)$ in total Lagrange and update Lagrange formulations.

2. Continuum Mechanics

Strain energy function for isotropic hyperelastic material can be expressed in terms of the principal invariants such as $W(I_1, I_2, I_3)$.To implement this type of material model in total Lagrange formulation, it is required to derive two important quantities that are Second Piola – Kirchhoff stress tensor and material elasticity tensor. The second Piola – Kirchhoff $\boldsymbol{S}$ stress can be written as

$$\boldsymbol{S} = 2\left(\frac{\partial W}{\partial I_1}\frac{\partial I_1}{\partial \boldsymbol{C}} + \frac{\partial W}{\partial I_2}\frac{\partial I_2}{\partial \boldsymbol{C}} + \frac{\partial W}{\partial I_3}\frac{\partial I_3}{\partial \boldsymbol{C}}\right). \tag{2}$$

The fourth order material elasticity tensor $\mathbb{C}$ is obtained by

* Corresponding author e-mail: a.hadoush@ptuk.edu.ps.

$$\mathbb{C} = 2\,\frac{\partial \boldsymbol{S}}{\partial \boldsymbol{C}} = 4\,\frac{\partial^2 W}{\partial \boldsymbol{C}\partial \boldsymbol{C}}, \tag{3}$$

a detailed definition for $\mathbb{C}$ in terms of principal invariants can be found in [27].

For Update Lagrange formulation, it is required to implement Cauchy stress and spatial elasticity tensors. Cauchy stress tensor $\boldsymbol{\sigma}$ is obtained using a Piola transformation of the second Piola Kirchhoff stress tensor such as

$$\boldsymbol{\sigma} = J^{-1}\boldsymbol{F}\boldsymbol{S}\boldsymbol{F}^T, \tag{4}$$

where $J = \det \boldsymbol{F}$. The Piola transformation [28] is a push forward operation that is scaled by J^{-1}. The spatial elasticity tensor $\boldsymbol{D}$ is defined in via Piola transformation of $\mathbb{C}$ [29, 27]. It is written in component form as

$$D_{abcd} = J^{-1}F_{aI}F_{bJ}F_{cK}F_{dL}\mathbb{C}_{IJKL}. \tag{5}$$

3. Internal Balance

A new scheme of viewing incompressible hyperelastic material response is introduced in [18]. The arguments of variation are applied to both portions of the deformation gradient decomposition. The decomposition itself is determined on the basis of an additional internal balance equation that emerges naturally from the variational treatment. A review of the compressible hyperelastic scheme [20] is summarized here. It has been shown that the second Piola – Kirchhoff $\boldsymbol{S}$ stress is obtained as consequence of applying the argument of variation with respect to $\boldsymbol{C}$. It can be written as

$$\boldsymbol{S}(\boldsymbol{C}, \breve{\boldsymbol{C}}) = 2\left(\frac{\partial W}{\partial \hat{I}_1}\breve{\boldsymbol{C}}^{-1} + \frac{\partial W}{\partial \hat{I}_2}\left(\hat{I}_1\breve{\boldsymbol{C}}^{-1} - \boldsymbol{M}\right) + \frac{\partial W}{\partial \hat{I}_3}\hat{I}_3\boldsymbol{C}^{-1}\right), \tag{6}$$

where $\boldsymbol{M} = \breve{\boldsymbol{C}}^{-1}\boldsymbol{C}\breve{\boldsymbol{C}}^{-1}$. Notice that $\boldsymbol{S}$ is a function of $\boldsymbol{C}$ and $\breve{\boldsymbol{C}}$. The decomposition of the deformation gradient is found by solving an internal balance equation that arises from the variation with respect to $\breve{\boldsymbol{C}}$. The internal balance equation is

$$\boldsymbol{\Psi} = \boldsymbol{0}, \tag{7}$$

where $\boldsymbol{\Psi}$ is an internal balance tensor

$$\boldsymbol{\Psi} = \widehat{\boldsymbol{\Psi}}_1 + \widehat{\boldsymbol{\Psi}}_2 + \widehat{\boldsymbol{\Psi}}_3 + \breve{\boldsymbol{\Psi}}_1 + \breve{\boldsymbol{\Psi}}_2 + \breve{\boldsymbol{\Psi}}_3 \tag{8}$$

with individual parts

$$\widehat{\boldsymbol{\Psi}}_1 = 2\frac{\partial W}{\partial \hat{I}_1}\frac{\partial \hat{I}_1}{\partial \breve{\boldsymbol{C}}} = -2\frac{\partial W}{\partial \hat{I}_1}\boldsymbol{M}, \tag{9}$$

$$\widehat{\boldsymbol{\Psi}}_2 = 2\frac{\partial W}{\partial \hat{I}_2}\frac{\partial \hat{I}_2}{\partial \breve{\boldsymbol{C}}} = 2\frac{\partial W}{\partial \hat{I}_2}\left(\boldsymbol{N} - \hat{I}_1\boldsymbol{M}\right), \tag{10}$$

$$\widehat{\boldsymbol{\Psi}}_3 = 2\frac{\partial W}{\partial \hat{I}_3}\frac{\partial \hat{I}_3}{\partial \breve{\boldsymbol{C}}} = -2\frac{\partial W}{\partial \hat{I}_3}\hat{I}_3\breve{\boldsymbol{C}}^{-1}, \tag{11}$$

$$\breve{\boldsymbol{\Psi}}_1 = 2\frac{\partial W}{\partial \check{I}_1}\frac{\partial \check{I}_1}{\partial \breve{\boldsymbol{C}}} = 2\frac{\partial W}{\partial \check{I}_1}\boldsymbol{I}, \tag{12}$$

$$\breve{\boldsymbol{\Psi}}_2 = 2\frac{\partial W}{\partial \check{I}_2}\frac{\partial \check{I}_2}{\partial \breve{\boldsymbol{C}}} = 2\frac{\partial W}{\partial \check{I}_2}\left(\check{I}_1\boldsymbol{I} - \breve{\boldsymbol{C}}\right), \tag{13}$$

$$\breve{\boldsymbol{\Psi}}_3 = 2\frac{\partial W}{\partial \check{I}_3}\frac{\partial \check{I}_3}{\partial \breve{\boldsymbol{C}}} = 2\frac{\partial W}{\partial \check{I}_3}\check{I}_3\breve{\boldsymbol{C}}^{-1}, \tag{14}$$

where $\boldsymbol{N} = \breve{\boldsymbol{C}}^{-1}\boldsymbol{C}\breve{\boldsymbol{C}}^{-1}\boldsymbol{C}\breve{\boldsymbol{C}}^{-1}$ and $\boldsymbol{I}$ is the second order identity tensor. Notice that $\widehat{\boldsymbol{\Psi}}_2$ is relatively complicated compared to other individual parts of $\boldsymbol{\Psi}$. It is tedious task to differentiate $\widehat{\boldsymbol{\Psi}}_2$ with respect to $\boldsymbol{C}$ and $\breve{\boldsymbol{C}}$ during linearization procedure. This leads to avoid the use of strain energy function based on $\hat{I}_2$. To overcome this difficulties, the following procedure is applied to simplify the expression of $\widehat{\boldsymbol{\Psi}}_2$. First of all, the push forward operation is performed

$$\boldsymbol{F}\frac{\partial \hat{I}_2}{\partial \breve{\boldsymbol{C}}}\boldsymbol{F}^T = \boldsymbol{F}\left(\boldsymbol{N} - \hat{I}_1\boldsymbol{M}\right)\boldsymbol{F}^T = \widehat{\boldsymbol{B}}^3 - \hat{I}_1\widehat{\boldsymbol{B}}^2, \tag{15}$$

then Cayley – Hamilton equation is applied to simplify (15) to

$$\boldsymbol{F}\frac{\partial \hat{I}_2}{\partial \breve{\boldsymbol{C}}}\boldsymbol{F}^T = \hat{I}_3\boldsymbol{I} - \hat{I}_2\widehat{\boldsymbol{B}}, \tag{16}$$

and finally (16) is pulled backward to get

$$\frac{\partial \hat{I}_2}{\partial \breve{\boldsymbol{C}}} = \boldsymbol{F}^{-1}\left(\hat{I}_3\boldsymbol{I} - \hat{I}_2\widehat{\boldsymbol{B}}\right)\boldsymbol{F}^{-T} = \hat{I}_3\boldsymbol{C}^{-1} - \hat{I}_2\breve{\boldsymbol{C}}^{-1}, \tag{17}$$

now $\widehat{\boldsymbol{\Psi}}_2$ has more practical expression by virtue of (17) such as

$$\widehat{\boldsymbol{\Psi}}_2 = 2\frac{\partial W}{\partial \hat{I}_2}\frac{\partial \hat{I}_2}{\partial \breve{\boldsymbol{C}}} = 2\frac{\partial W}{\partial \hat{I}_2}\left(\hat{I}_3\boldsymbol{C}^{-1} - \hat{I}_2\breve{\boldsymbol{C}}^{-1}\right). \tag{18}$$

For updated Lagrange, the Cauchy stress for internally balanced scheme is defined by push forward of (6) as

$$\boldsymbol{\sigma}(\boldsymbol{B}, \widehat{\boldsymbol{B}}) = \frac{2}{J}\left(\frac{\partial W}{\partial \hat{I}_1}\widehat{\boldsymbol{B}} + \frac{\partial W}{\partial \hat{I}_2}\left(\hat{I}_1\widehat{\boldsymbol{B}} - \widehat{\boldsymbol{B}}^2\right) + \frac{\partial W}{\partial \hat{I}_3}\hat{I}_3\boldsymbol{I}\right). \tag{19}$$

Similarly, the internal balance equation (7) becomes

$$\boldsymbol{\Xi}(\boldsymbol{B}, \widehat{\boldsymbol{B}}) = \boldsymbol{F}\boldsymbol{\Psi}\boldsymbol{F}^T = \boldsymbol{0}, \tag{20}$$

where $\boldsymbol{\Xi}$ is the internal balance tensor with the individual parts

$$\boldsymbol{\Xi} = 2\left(\widehat{\boldsymbol{\Xi}}_1 + \widehat{\boldsymbol{\Xi}}_2 + \widehat{\boldsymbol{\Xi}}_3 + \breve{\boldsymbol{\Xi}}_1 + \breve{\boldsymbol{\Xi}}_2 + \breve{\boldsymbol{\Xi}}_3\right), \tag{21}$$

these individual parts of $\boldsymbol{\Xi}$ are listed in the appendix.

4. Blatz – Ko Model

The generalized Blatz – Ko [30, 31, 32] can be written as

$$W(I_1, I_2, I_3) = \frac{\mu(1-f)}{2}\left(\frac{I_2}{I_3} + \frac{1}{\alpha}I_3^{\alpha} - \zeta\right) + \frac{\mu f}{2}\left(\frac{1}{\alpha}I_3^{-\alpha} + I_1 - \zeta\right), \tag{22}$$

where $\alpha = \nu/(1-2\nu)$ and $\zeta = 1 + 1/\nu$. Blatz – Ko model has three material parameters namely Poisson's ratio ν, shear modulus μ and volume fraction of voids in foam rubber material f. In this work a special case of (22) is achieved [30, 31, 33] by applying $\nu = 1/4$ and $f = 0$ such as

$$W_{BK}(I_2, I_3) = \frac{\mu}{2}\left(\frac{I_2}{I_3} + 2I_3^{1/2} - 5\right). \tag{23}$$

Substituting (23) into (2) to get Second Piola – Kirchhoff stress

$$\boldsymbol{S}_{BK} = \mu\left(\frac{1}{I_3}(I_1\boldsymbol{I} - \boldsymbol{C}) + \mathcal{L}\boldsymbol{C}^{-1}\right) = \mu(J\boldsymbol{C}^{-1} - \boldsymbol{C}^{-2}), \tag{24}$$

where $\mathcal{L} = -I_2/I_3 + I_3^{1/2}$. The elasticity tensor is obtained by substituting (23) into (3)

$$\mathbb{C}_{BK} = \mu J(\boldsymbol{C}^{-1}\otimes\boldsymbol{C}^{-1} - 2\boldsymbol{C}^{-1}\odot\boldsymbol{C}^{-1}) + 2\mu(\boldsymbol{C}^{-2}\odot\boldsymbol{C}^{-1} + \boldsymbol{C}^{-1}\odot\boldsymbol{C}^{-2}), \tag{25}$$

where $\otimes$ is the dyadic operator and $\odot$ operator has the same definition given in [27], further details can be found in the appendix. Now substitute (24) into (4) to obtain Cauchy stress

$$\boldsymbol{\sigma}_{BK} = \mu(\boldsymbol{I} - J^{-1}\boldsymbol{B}^{-1}). \tag{26}$$

The spatial tensor of elasticity is obtained by the virtue of (5) as

$$\boldsymbol{D}_{BK} = \mu(\boldsymbol{I}\otimes\boldsymbol{I} - 2\boldsymbol{I}\odot\boldsymbol{I}) + 2\mu J^{-1}(\boldsymbol{B}^{-1}\odot\boldsymbol{I} + \boldsymbol{I}\odot\boldsymbol{B}^{-1}). \tag{27}$$

5. Blatz – Ko Internal Balance Model

The form of internally balanced material model is motivated by special case of Blatz – Ko model

$$W_{IB} = \frac{\mu(1+\beta)}{2}\left(\frac{\hat{I}_2}{\hat{I}_3} + 2\hat{I}_3^{1/2} - 5\right) + \frac{\mu(1+\beta)}{2\beta}\left(\frac{\check{I}_2}{\check{I}_3} + 2\check{I}_3^{1/2} - 5\right), \quad (28)$$

where β is a positive material parameter that quantify the contribution of two – factor multiplicative decomposition (1). The limits $\beta \to 0$ and $\beta \to \infty$ retrieve the hyperelastic behavior in full nonlinear strain range [20]. An equivalent form of (28) is used in [34] to investigate uniaxial loading by solving nonlinear boundary value problem. In this work, the finite element formulation of (28) is presented that permits general loading scenarios.

The Piola – Kirchhoff stress of internally balanced material (28) is obtained by (6)

$$\boldsymbol{S}_{IB} = \bar{\mu}\left(\hat{\mathcal{L}}\boldsymbol{C}^{-1} + \frac{1}{\hat{I}_3}\left(\hat{I}_1\check{\boldsymbol{C}}^{-1} - \boldsymbol{M}\right)\right), \quad (29)$$

where $\bar{\mu} = \mu(1+\beta)$ and $\hat{\mathcal{L}} = -\hat{I}_2/\hat{I}_3 + \hat{I}_3^{1/2}$. The stress tensor $\boldsymbol{S}_{IB}$ is coupled to internal balance equation $\boldsymbol{\Psi} = \boldsymbol{0}$ that is achieved by the virtue of equations (9),(11) – (14) and (18)

$$\boldsymbol{C}^{-1} - \hat{J}\check{\boldsymbol{C}}^{-1} + \frac{1}{\beta}\left(\check{J}\check{\boldsymbol{C}}^{-1} - \check{\boldsymbol{C}}^{-2}\right) = \boldsymbol{0}, \quad (30)$$

where $\hat{J} = \hat{I}_3^{1/2}$ and $\check{J} = \check{I}_3^{1/2}$. This quiet form of internal balance equation is obtained by using the simplified form of $\bar{\boldsymbol{\Psi}}_2$ in (18) instead of (10) and further manipulation using Cayley – Hamilton equation. The push forward of (29) and the use of Cayley – Hamilton equation give a simplified form of Cauchy stress as

$$\boldsymbol{\sigma}_{IB} = \frac{\bar{\mu}}{J}\left(\hat{J}\boldsymbol{I} - \hat{\boldsymbol{B}}^{-1}\right) \quad (31)$$

and it is coupled to $\boldsymbol{\Xi} = \boldsymbol{0}$ that is

$$\boldsymbol{I} - \hat{J}\hat{\boldsymbol{B}} + \frac{1}{\beta}\left(\check{J}\hat{\boldsymbol{B}} - \hat{\boldsymbol{B}}\boldsymbol{B}^{-1}\hat{\boldsymbol{B}}\right) = \boldsymbol{0}. \quad (32)$$

It has been noticed that Piola – Kirchhoff stress (29) has still complicated form, a simplified form of (29) is required to avoid lengthy linearization procedure. To achieve that (32) is multiplied by $\hat{\boldsymbol{B}}^{-1}$, then it is rearranged as

$$\hat{J}\boldsymbol{I} - \hat{\boldsymbol{B}}^{-1} = \frac{1}{\beta}\left(\check{J}\mathbf{I} - \hat{\boldsymbol{B}}\boldsymbol{B}^{-1}\right), \quad (33)$$

then substitute (33) in (31) to obtain another and equivalent form of Cauchy stress such as

$$\boldsymbol{\sigma}_{IB} = \frac{\bar{\mu}}{J}\left(\hat{J}\boldsymbol{I} - \hat{\boldsymbol{B}}^{-1}\right) = \frac{\bar{\mu}}{\beta J}\left(\check{J}\mathbf{I} - \hat{\boldsymbol{B}}\boldsymbol{B}^{-1}\right), \quad (34)$$

finally pull back the new form of Cauchy stress to get

$$\boldsymbol{S}_{IB} = \frac{\mu(1+\beta)}{\beta}\left(\check{J}\boldsymbol{C}^{-1} - \boldsymbol{C}^{-1}\check{\boldsymbol{C}}^{-1}\right). \quad (35)$$

Notice that Piola – Kirchhoff stress (35) has relatively simple form compared to (29).

6. Linearization

The finite element formulation of internally balanced compressible hyperelastic material is demonstrated in [21]. It is based on calculating a condensed fourth order elasticity tensor $\mathbb{C}_{con}$ that is defined as

$$\mathbb{C}_{con} = \mathbb{C}_{\boldsymbol{C}} - \mathbb{C}_{\check{\boldsymbol{C}}}\colon \boldsymbol{\Psi}_{\check{\boldsymbol{C}}}^{-1} \colon \boldsymbol{\Psi}_{\boldsymbol{C}}. \quad (36)$$

The tensors $\mathbb{C}_{\boldsymbol{C}}$ and $\mathbb{C}_{\check{\boldsymbol{C}}}$ are obtained by differentiating second Piola – Kirchhoff stress (6) with respect to $\boldsymbol{C}$ and $\check{\boldsymbol{C}}$, respectively. The terms $\boldsymbol{\Psi}_{\boldsymbol{C}}$ and $\boldsymbol{\Psi}_{\check{\boldsymbol{C}}}$ are obtained by differentiating internal balance tensor (8) with respect to $\boldsymbol{C}$ and $\check{\boldsymbol{C}}$, respectively.

Concerning the material model of interest in this work (28), these individual parts of (36) become

$$\mathbb{C}_{\boldsymbol{C}} = \frac{\bar{\mu}}{\beta}\left(\boldsymbol{G}\odot\boldsymbol{C}^{-1} + \boldsymbol{C}^{-1}\odot\boldsymbol{G} - 2\check{J}\boldsymbol{C}^{-1}\odot\boldsymbol{C}^{-1}\right), \quad (37)$$

$$\mathbb{C}_{\check{\boldsymbol{C}}} = \frac{\bar{\mu}}{\beta}\left(\check{J}\boldsymbol{C}^{-1}\otimes\check{\boldsymbol{C}}^{-1} + \check{\boldsymbol{C}}^{-1}\odot\boldsymbol{G} + \boldsymbol{G}\odot\check{\boldsymbol{C}}^{-1}\right), \quad (38)$$

$$\boldsymbol{\Psi}_{\boldsymbol{C}} = -2\boldsymbol{C}^{-1}\odot\boldsymbol{C}^{-1} - \hat{J}\check{\boldsymbol{C}}^{-1}\otimes\boldsymbol{C}^{-1}, \quad (39)$$

$$\boldsymbol{\Psi}_{\check{\boldsymbol{C}}} = \bar{J}_1\,\check{\boldsymbol{C}}^{-1}\otimes\check{\boldsymbol{C}}^{-1} + 2\bar{J}_2\,\check{\boldsymbol{C}}^{-1}\odot\check{\boldsymbol{C}}^{-1} + \frac{2}{\beta}\left(\check{\boldsymbol{C}}^{-2}\odot\check{\boldsymbol{C}}^{-1} + \check{\boldsymbol{C}}^{-1}\odot\check{\boldsymbol{C}}^{-2}\right), \quad (40)$$

where $\boldsymbol{G} = \check{\boldsymbol{C}}^{-1}\boldsymbol{C}^{-1}$, $\bar{J}_1 = \hat{J} + \check{J}/\beta$ and $\bar{J}_2 = \hat{J} - \check{J}/\beta$. The spatial condensed elasticity tenor $\boldsymbol{D}_{con}$ for the updated Lagrange formulation can be obtained either by pushing forward of $\mathbb{C}_{con}$ or by performing push forward operation on each individual parts (37) – (40) then apply condensation procedure.

7. Solving Internal Balance Equation

It is an essential step to solve the internal balance equation in order to calculate the stress and elasticity tensor. This is carried out by solving (30) for $\check{\boldsymbol{C}}$ given the value of $\boldsymbol{C}$ or solving (32) for $\hat{\boldsymbol{B}}$ given the value of $\boldsymbol{B}$. Here it is chosen to solve (32). The multiplication of (32) by $\boldsymbol{B}$, then rearranging it, gives a simplified form of the internal balance equation

$$\hat{\boldsymbol{B}}^2 + \left(\beta\hat{J} - \frac{\check{J}}{\hat{J}}\right)\hat{\boldsymbol{B}}\boldsymbol{B} - \beta\boldsymbol{B} = \boldsymbol{0}. \quad (41)$$

It can be shown by further manipulation of (41) that the tensor $\boldsymbol{B}$ and $\hat{\boldsymbol{B}}$ have the same orthogonal eigenvectors. Then, internal balance equation (41) can be expressed in principal frame using eigenvalues of $\boldsymbol{B}$ as $\lambda_1^2, \lambda_2^2, \lambda_3^2$ and eigenvalues of $\hat{\boldsymbol{B}}$ as $\hat{\lambda}_1^2, \hat{\lambda}_2^2, \hat{\lambda}_3^2$ [35]. The internal balance equation in principal frame are expressed as a set nonlinear equations

$$\hat{\lambda}_1^4 + \beta\hat{\lambda}_1^3\hat{\lambda}_2\hat{\lambda}_3\lambda_1^2 - \frac{\lambda_1^2\check{J}\hat{\lambda}_1}{\hat{\lambda}_2\hat{\lambda}_3} - \beta\lambda_1^2 = 0, \quad (42)$$

$$\hat{\lambda}_2^4 + \beta\hat{\lambda}_1\hat{\lambda}_2^3\hat{\lambda}_3\lambda_2^2 - \frac{\lambda_2^2\check{J}\hat{\lambda}_2}{\hat{\lambda}_1\hat{\lambda}_3} - \beta\lambda_2^2 = 0, \quad (43)$$

$$\hat{\lambda}_3^4 + \beta\hat{\lambda}_1\hat{\lambda}_2\hat{\lambda}_3^3\lambda_3^2 - \frac{\lambda_3^2\check{J}\hat{\lambda}_3}{\hat{\lambda}_1\hat{\lambda}_2} - \beta\lambda_3^2 = 0, \quad (44)$$

Newton – Raphson iterative procedure is used to solve (42) – (44).

8. Results and Discussion

The conventional hyperelastic Baltz – Ko model (23) and internally balanced Blatz – Ko model (28) are implemented in $FEAPpv^{©}$. It is in – house finite element package for updated Lagrange formulation written by Prof. R.L. Taylor, University of California at Berkeley. A cube of a unit length is meshed by 8 nodes brick element with three degrees of freedom per node. The cube is discretized by eight elements in total $(2 \times 2 \times 2)$. The homogeneous deformations namely uniaxial and simple shear are performed to investigate the response of the material models.

The uniaxial loading is given as $\boldsymbol{F} = \lambda_1\boldsymbol{e}_1\otimes\boldsymbol{e}_1 + \lambda_2\boldsymbol{e}_2\otimes\boldsymbol{e}_2 + \lambda_3\boldsymbol{e}_3\otimes\boldsymbol{e}_3$ with corresponding stress $\boldsymbol{\sigma} = \sigma_{11}\,\boldsymbol{e}_1\otimes\boldsymbol{e}_1$ where $\lambda_1, \lambda_2, \lambda_3$ are principal stretches. The material is stretched in principal direction one and it is free to contract in the other principal directions. The achieved uniaxial stress σ_{11} for different values of β is shown in figure 1. It is verified that the achieved Cauchy stress value for given β coincides with achieved Cauchy stress value

for $1/\beta$. Therefore, the values of $0 < \beta \leq 1$ are used in following discussion.

The achieved Cauchy stress curves for different values of β have a specific pattern up to particular value of stretch λ_1 that is stiffer than hyperelastic achieved Cauchy stress ($\beta = 0$). However, this pattern is not maintained for large value of stretch λ_1 see figure 2 (top). A better pattern is observed when Cauchy stress is normalized by $\mu(1+\beta)$ as shown in figure 2 (bottom). Now, the curves are ordered such as the stiffer response is achieved when $\beta = 0$ (hyperelastic) and the softest response is achieved for $\beta = 1$. The normalized Cauchy stress $\bar{\sigma}_{11} = \sigma_{11}/\mu(1+\beta)$ decreases as β increases from 0 to 1. The achieved Cauchy stress for $\beta = 0$ reaches an asymptotic value of one for large stretch $\lambda_1 \to \infty$; this is the retrieved value by (26). For internally balanced achieved results $\beta > 0$, the normalized Cauchy stress $\bar{\sigma}_{11}$ increases with stretching λ_1 up to a critical maximum value then it decreases. For the special case $\beta = 1$, this maximum value of normalized stress is found to be $\bar{\sigma}_{11} = 0.5824$ at $\lambda_1 = 6^{0.8} = 4.193$ this is shown in figure 2 (bottom).

The achieved finite element results are verified analytically by examining two special cases. The first case is for $\beta = 0$ this results in $\widehat{\boldsymbol{B}} \to \boldsymbol{B}$. This means that the hyperelastic Cauchy stress expressed in (26) is retrieved by internal balance Cauchy stress (31) at $\beta = 0$, the material parameter $\mu(\beta+1)$ becomes simply μ. The hyperelastic Cauchy stress is also retrieved at $\beta \to \infty$. Now, the normalized uniaxial stress has analytical expression such as $\bar{\sigma}_{11} = 1 - \lambda_1^{-5/2}$. The second special case is for $\beta = 1$. It can be shown that $\widehat{\boldsymbol{B}} = \boldsymbol{B}^{1/2} \to \hat{J} = J^{1/2}$ is a solution for (41). Then, corresponding normalized uniaxial stress becomes $\bar{\sigma}_{11} = \lambda_1^{-1/4} - \lambda_1^{-3/2}$. The achieved finite element results for both special cases have very good agreement with the achieved results by the analytical expressions, see figure 3.

Simple shear deformation has the form of $\boldsymbol{F} = \boldsymbol{I} + \gamma\, \boldsymbol{e}_1 \otimes \boldsymbol{e}_2$ where γ is the amount of shear that is related to angle of shear θ by $\gamma = \tan\theta$. The Cauchy stress for hyperelastic material (26) becomes $(1/\mu)\,\boldsymbol{\sigma} = \sigma_{22}\, \boldsymbol{e}_2 \otimes \boldsymbol{e}_2 + \sigma_{12}\,(\boldsymbol{e}_1 \otimes \boldsymbol{e}_2 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_1)$ where $\sigma_{22} = -\gamma^2$ and $\sigma_{12} = \gamma$. In general for hyperelastic material, the normal stress components do not vanish but the Cauchy stress components satisfy the universal relation $\sigma_{11} - \sigma_{22} = \gamma\sigma_{12}$, this is known as Poynting effect [36]. The hyperelastic Cauchy stress $\sigma_{12} = \gamma$ is increasing monotonically with amount of shear as plotted in figure 4, this theoretically means that the cube can sustain infinite shear deformation. The internally balance shear stress component shows softer response for $0 < \beta \leq 1$. The achieved curves of normalized shear stress is ordered such as hyperelastic ($\beta = 0$) is the most stiff, it becomes softer with increasing β, and the result of $\beta = 1$ is the most soft response. The internally balanced curves tend to reach an asymptotic value at large deformation. This agrees with previous findings in [19, 35].

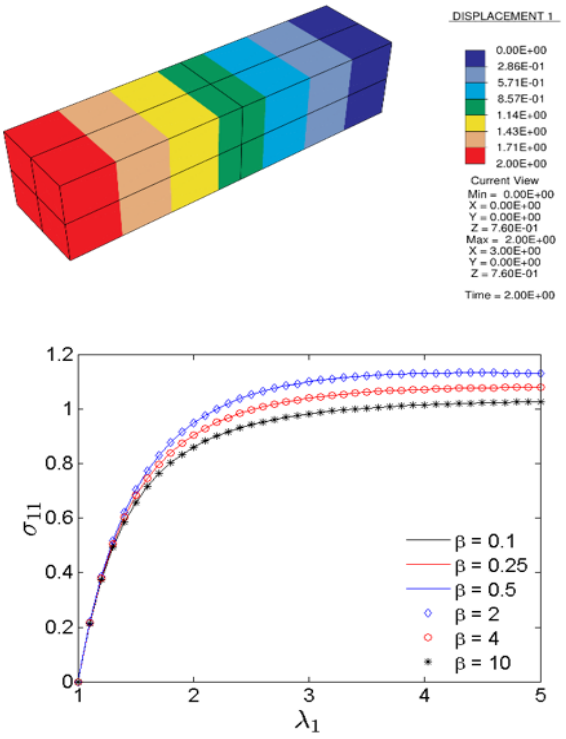


Figure 1. Uniaxial homogeneous deformation (top). Uniaxial Cauchy stress for different values of β (bottom). Achieved Cauchy stress value for given β coincides with achieved Cauchy stress value for $1/\beta$.

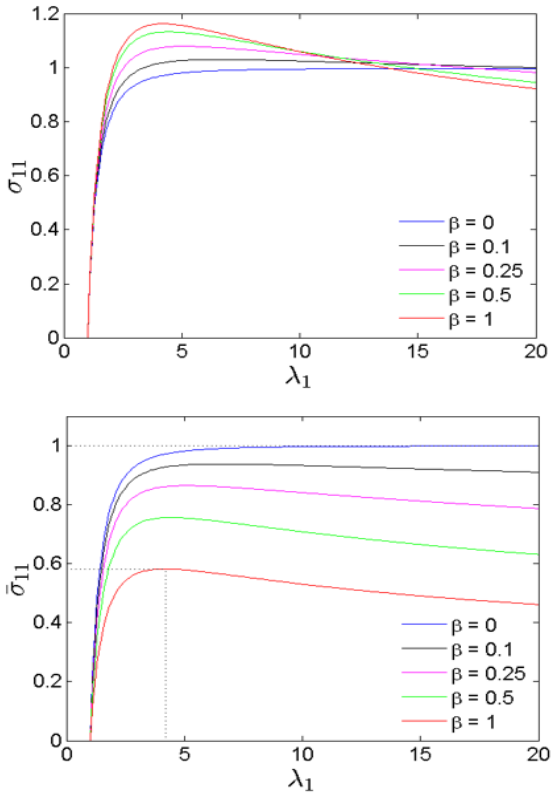


Figure 2. Uniaxial Cauchy stress for different values of β (top) and normalized Cauchy stress by $\mu(1+\beta)$ (bottm). The hyperelastic case is retrieved by $\beta = 0$.

9. Conclusion

In this work a finite element treatment is demonstrated for a material model based on new theory that applies argument of variation to both counterparts of deformation gradient multiplicative decomposition [21]. The use of Cayley – Hamilton equation facilitates significantly the implementation of internally balanced material model that has the form of $W(\hat{I}_2, \hat{I}_3, \check{I}_2, \check{I}_3)$. The response of the material model is examined in uniaxial loading and simple shear. The internally balanced theory retrieves the conventional hyperelastic theory in the special limiting case $\beta = 0$. The uniaxial stress for internally balanced material has a stiffer response compared with hyperelastic uniaxial stress up to significant value of stretching when it reaches maximum value then it shows softening behavior. For simple shear, the internally balanced shear stress shows softer response and reaches an asymptotic value in contrast with unbounded increases of hyperelastic shear stress. The presented treatment complements previous formulation demonstrated in [21] and it allows the use of complicated material models.

Acknowledgement

The author would like to acknowledges Palestine Technical University – Kadoorie for supporting this research.

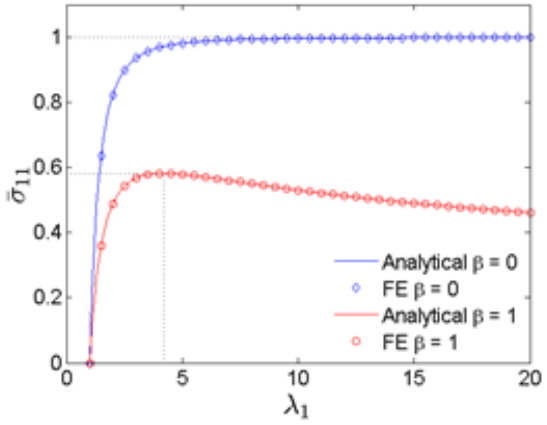


Figure 3. The verification of finite element results by analytical solution for two special cases $\beta = 0$ and $\beta = 1$.

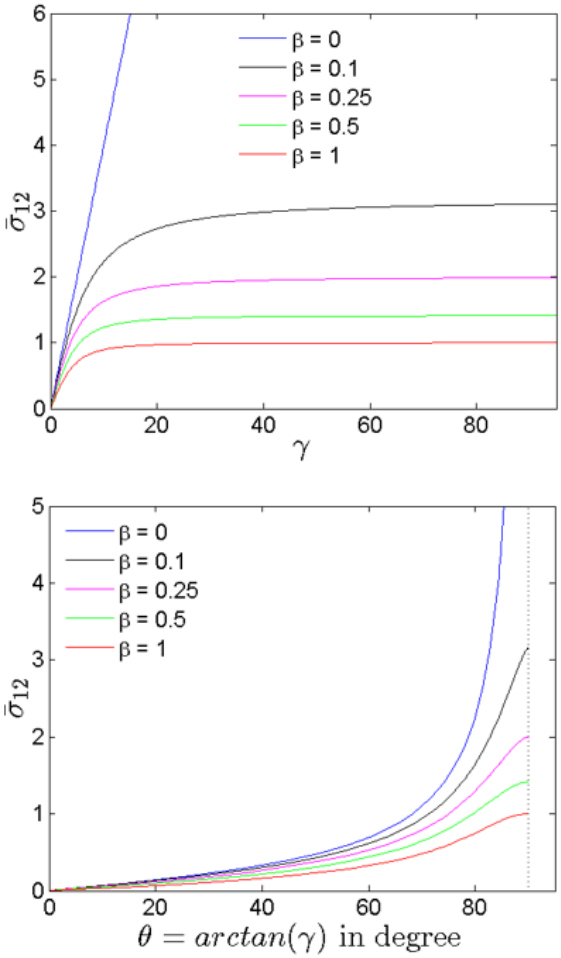


Figure 4. Normalized shear stress component $\bar{\sigma}_{12}$ for different values of β.

Appendix

Let $\boldsymbol{A}$ denote a second order symmetric tensor. The principal invariants of $\boldsymbol{A}$ are

$$I_1 = tr(\boldsymbol{A}), I_2 = \frac{1}{2}\left(\left(tr(\boldsymbol{A})\right)^2 - tr\left(\boldsymbol{A^2}\right)\right), I_3 = \det \boldsymbol{A} \quad (45)$$

and their partial derivatives with respect to $\boldsymbol{A}$

$$\frac{\partial I_1}{\partial \boldsymbol{A}} = \boldsymbol{I}, \frac{\partial I_2}{\partial \boldsymbol{A}} = I_1\boldsymbol{I} - \boldsymbol{A}, \frac{\partial I_3}{\partial \boldsymbol{A}} = I_3\boldsymbol{A^{-1}}. \quad (46)$$

Cayley – Hamilton equation can be written as

$$\boldsymbol{A}^3 - I_1\boldsymbol{A}^2 + I_2\boldsymbol{A} - I_3\boldsymbol{I} = \boldsymbol{0}, \quad (47)$$

successive multiplication of (47) by $\boldsymbol{A^{-1}}$ provides additional useful formulas such as

$$\boldsymbol{A}^2 - I_1\boldsymbol{A} + I_2\boldsymbol{I} - I_3\boldsymbol{A^{-1}} = \boldsymbol{0}, \quad (48a)$$

$$\boldsymbol{A} - I_1\boldsymbol{I} + I_2\boldsymbol{A^{-1}} - I_3\boldsymbol{A^{-2}} = \boldsymbol{0}. \quad (48b)$$

Derivative of $\boldsymbol{A^{-1}}$ and $\boldsymbol{A^{-2}}$ with respect to $\boldsymbol{A}$ are

$$-\frac{\partial \boldsymbol{A^{-1}}}{\partial \boldsymbol{A}} = \boldsymbol{A^{-1}} \odot \boldsymbol{A^{-1}}, \quad (49)$$

$$-\frac{\partial \boldsymbol{A^{-2}}}{\partial \boldsymbol{A}} = \boldsymbol{A^{-2}} \odot \boldsymbol{A^{-1}} + \boldsymbol{A^{-1}} \odot \boldsymbol{A^{-2}}, \quad (50)$$

where the $\odot$ operator is defined as

$$2\boldsymbol{A^{-1}} \odot \boldsymbol{A^{-1}}_{|IJKL} = A^{-1}_{IK}A^{-1}_{JL} + A^{-1}_{IL}A^{-1}_{JK}. \quad (51)$$

The push forward of the internal balance quantities is defined as

$$\boldsymbol{F\breve{C}^{-2}F^T} = \boldsymbol{\hat{B}B^{-1}\hat{B}}, \quad (52a)$$

$$\boldsymbol{F\breve{C}^{-1}F^T} = \boldsymbol{\hat{B}}, \quad (52b)$$

$$\boldsymbol{F\breve{C}^{-1}C\breve{C}^{-1}F^T} = \boldsymbol{\hat{B}}^2, \quad (52c)$$

$$\boldsymbol{F\breve{C}^{-1}C^{-1}F^T} = \boldsymbol{\hat{B}B^{-1}}. \quad (52d)$$

The individual parts of (21) are

$$\hat{\boldsymbol{\Xi}}_1 = \boldsymbol{F}\frac{\partial W}{\partial \hat{I}_1}\frac{\partial \hat{I}_1}{\partial \breve{\boldsymbol{C}}}\boldsymbol{F^T} = -\frac{\partial W}{\partial \hat{I}_1}\,\hat{\boldsymbol{B}}^2, \quad (53)$$

$$\hat{\boldsymbol{\Xi}}_2 = \boldsymbol{F}\frac{\partial W}{\partial \hat{I}_2}\frac{\partial \hat{I}_2}{\partial \breve{\boldsymbol{C}}}\boldsymbol{F^T} = \frac{\partial W}{\partial \hat{I}_2}\left(\hat{\boldsymbol{B}}^3 - \hat{I}_1\hat{\boldsymbol{B}}^2\right) = \frac{\partial W}{\partial \hat{I}_2}\left(-\hat{I}_2\hat{\boldsymbol{B}} + \hat{I}_3\boldsymbol{I}\right), \quad (54)$$

$$\hat{\boldsymbol{\Xi}}_3 = \boldsymbol{F}\frac{\partial W}{\partial \hat{I}_3}\frac{\partial \hat{I}_3}{\partial \breve{\boldsymbol{C}}}\boldsymbol{F^T} = -\frac{\partial W}{\partial \hat{I}_3}\hat{I}_3\hat{\boldsymbol{B}}, \quad (55)$$

$$\breve{\boldsymbol{\Xi}}_1 = \boldsymbol{F}\frac{\partial W}{\partial \breve{I}_1}\frac{\partial \breve{I}_1}{\partial \breve{\boldsymbol{C}}}\boldsymbol{F^T} = \frac{\partial W}{\partial \breve{I}_1}\boldsymbol{B}, \quad (56)$$

$$\breve{\boldsymbol{\Xi}}_2 = \boldsymbol{F}\frac{\partial W}{\partial \breve{I}_2}\frac{\partial \breve{I}_2}{\partial \breve{\boldsymbol{C}}}\boldsymbol{F^T} = \frac{\partial W}{\partial \breve{I}_2}\left(\breve{I}_1\boldsymbol{B} - \boldsymbol{B}\hat{\boldsymbol{B}}^{-1}\boldsymbol{B}\right), \quad (57)$$

$$\breve{\boldsymbol{\Xi}}_3 = \boldsymbol{F}\frac{\partial W}{\partial \breve{I}_3}\frac{\partial \breve{I}_3}{\partial \breve{\boldsymbol{C}}}\boldsymbol{F^T} = \frac{\partial W}{\partial \breve{I}_3}\breve{I}_3\hat{\boldsymbol{B}}. \quad (58)$$

References

[1] B.A. Bilby, L.R.T. Gardner, A.N. Stroh, "Continuous distributions of dislocations and theory of plasticity". Actes du Xième Congrès International de Mécanique Appliquée, vol. VIII, 1957, 35-44.

[2] E. Kröner, "Allgemeine kontinuumstheorie der versetzungen und eigenspannungen". Archive for Rational Mechanics and Analysis, vol. 4, 1959, 273-334.

[3] E. H. Lee, "Elastic plastic deformation at _nite strain". ASME Journal of Applied Mechanics, vol. 36, 1969, 1-6.

[4] J.C. Simo, "A framework for finite strain elastoplasticity based on maximum plastic dissipation and the multiplicative decomposition. Part I: Continuum formulation". Computer Methods in Applied Mechanics and Engineering, vol. 66, 1988, 199-219.

[5] B. Nedjar, "Frameworks for finite strain viscoelastic-plasticity based on multiplicative decompositions. Part I: Continuum formulations". Computer Methods in Applied Mechanics and Engineering, vol. 191, 2002, 1541-1562.

[6] V.A. Lubarda, "Constitutive theories based on multipliative decomposition of the deformation gradient: thermoelasticity, elastoplasticity, and biomechanics". Appl. Mech. Rev., vol. 57, 2004, 95-108.

[7] L. Anand, M.E. Gurtin, "The decomposition $F = F^eF^p$, material symmetry, and plastic irrotationality for solids that are isotropic-viscoplastic or amorphous". International Journal of Plasticity, vol. 21, 2005, 1686-1719.

[8] M. Latorre, F.J. Montans, "Anisotropic finite strain viscoelasticity based on the sidoroff multiplicative decomposition and logarithmic strains". Computational Mechanics, vol. 56, 2015, 503-531.

[9] A.V. Shutov, R. Landgraf, J. Ihlemann, "An explicit solution for implicit time stepping in multiplicative finite strain viscoelasticity". Computer Methods in Applied Mechanics and Engineering, vol. 265, 2013, 213-225.

[10] E. K. Rodriguez, A. Hoger, A.D. McCulloch, cStress-dependent finite growth in soft elastic tissues". Journal of Biomechanics, vol. 27, 1994, 455-467.

[11] Y. Chen, A. Hoger, "Constitutive functions of elastic materials in finite growth and deformation". Journal of Elasticity, vol. 59, 2000, 175-193.

[12] A. DiCarlo, S. Quiligotti, "Growth and balance". Mechanics Research Communications, vol. 29, 2002, 449-456.

[13] A. Goriely, M.B. Amar, "On the de_nition and modeling of incremental, cumulative, and continuous growth laws in morphoelasticity". Biomechanics and Modeling in Mechanobiology, vol. 6, 2007, 289-296.

[14] T. Belytschko, W.K. Liu, B. Moran, Nonlinear finite elements for continua and
structures. England: John Wiley and Sons; 2007.

[15] O.C. Zienkiewicz, R.L. Taylor, D.D. Fox. The finite element method for solid and Structural mechanics. seventh edition. Butterworth-Heinemann; 2013.

[16] J.C. Simo, "A framework for finite strain elastoplasticity based on maximum plastic dissipation and the multiplicative decomposition. Part II: Computational aspects". Computer Methods in Applied Mechanics and Engineering, vol. 68, 1988, 1-31.

[17] B. Nedjar, "Frameworks for finite strain viscoelastic-plasticity based on multiplicative decompositions. Part II: Computational aspects". Computer Methods in Applied Mechanics and Engineering, vol. 191, 2002, 1563-1593.

[18] H. Demirkoparan, T.J. Pence, H. Tsai, "Hyperelastic internal balance by multiplicative decomposition of the deformation gradient". Arch. Rat. Mech. Anal., vol. 214, 2014, 923-970.

[19] H. Demirkoparan, T.J. Pence, "Finite stretching and shearing of an internally balanced elastic solid". Journal of Elasticity, vol. 121, 2015, 1-23.

[20] A. Hadoush, H. Demirkoparan, T.J. Pence, "A constitutive model for an internally balanced compressible hyperelastic material". Mathematics and Mechanics of solids, vol. 22, 2015, 372-400.

[21] A. Hadoush, H. Demirkoparan, T.J. Pence, "Finite element analysis of internally balanced elastic materials". Computer Methods in Applied Mechanics and Engineering, vol. 322, 2017, 373-395.

[22] S. Reese, S. Govindjee, "A theory of finite viscoelasticity and numerical aspects". Int. J. Solids Structures, vol. 35, 1998, 3455-3482.

[23] L. Deseri, D.R. Owen, "Toward a field theory for elastic bodies undergoing disarrangements". Journal of Elasticity, vol. 70, 2003, 197-236.

[24] L. S. Weissman, J.L. Sackman, "Elastic-plastic multiplicative decomposition with a stressed intermediate configuration". Computer Methods in Applied Mechanics and Engineering, vol. 200, 2011, 1607-1618.

[25] C. Miehe, "Variational gradient plasticity at finite strains. Part I: Mixed potentials for the evolution and update problems of gradient-extended dissipative solids". Computer Methods in Applied Mechanics and Engineering, vol. 268, 2014, 677-703.

[26] J. Engelbrecht, A. Berezovski, "Internal structures and internal variables in solids". Journal of Mechanics and Materials and Structures, vol. 7, 2014, 983-996.

[27] G.A. Holzapfel. Nonlinear solid mechanics. England: John Wiley & Sons LTD; 2005.

[28] J.E. Marsden, T.J.R. Hughes. Mathematical Foundations of Elasticity. NewYork: Dover Publications; 1994.

[29] J. Bonet, R.D. Wood. Nonlinear continuum mechanics for finite element analysis. second edition. Cambridge: Cambridge University Press; 2008.

[30] P.J. Blatz, W.L. Ko, "Application of finite elasticity to the deformation of rubbery materials". Trans. Soc. Rheol., vol. 6, 1962, 223-251.

[31] P.J. Blatz, "On the thermostatic behavior of elastomers". Polymer Networks, Structure and Mechanical Properties, Plenum Press, NewYork, USA, 1971.

[32] M.F. Beatty, "Topics in finite elasticity: Hyperelasticity of rubber, elastomers and biological tissues-with examples". Appl. Mech. Reviews, vol. 40, 1987, 1699-1734.

[33] C.O. Horgan, "Remarks on ellipticity for the generalized Blatz – Ko constitutive model for a compressible nonlinear elastic solid". Journal of Elasticity, vol. 42, 1996, 165-176.

[34] A. Hadoush, H. Demirkoparan, T.J. Pence, "Modeling of soft materials via multiplicative decomposition of deformation gradient". USNCTAM, USA, 2014.

[35] A. Hadoush, H. Demirkoparan, T.J. Pence, "Simple shearing and azimuthal shearing of an internally balanced compressible elastic material". International Journal of nonlinear mechanics, vol. 79, 2016, 99-114.

[36] R.W. Ogden. Nonlinear Elastic Deformations. New York: Dover; 1997.

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665
Pages 223 - 236

Jordan Journal of Mechanical and Industrial Engineering

Hasty Fault Diagnosis of a Rotating Machinery Hinge on Stalwart Trippy Classifier with Robust Harmonized Swan Machine

Sumit Kumar Sar *, Ramesh Kumar

Bhilai institute of Technology, Bhilai House, Durg, Chhattisgarh 491001.

Received July 13 2019 *Accepted March 28 2020*

Abstract

Monitoring with fault diagnosis of machineries is critically important for production efficiency and plant safety in modern enterprises. The process of fault diagnosis along with extracting representative features from the vibrational signal with the existing Harmonised Swan Machine feature extraction technique resulted in high noise sensitivity, mixing mode problem and data loss. Moreover, measuring the correlation between the features with the existing fault diagnosis researches suffers from learning time limitations and memory constraints. Thus to commensurate a perfect diagnosis, in this research a "Robust Harmonised Swan Machine (RHSM) with Stalwart Trippy classifier" is formulated in which the iterative threshold VMD (Variational Mode Decomposition) estimation of each mode satisfying a self-consistency nature in decomposition method of RHSM is performed which in turn resolves the missing sample problem eminently independent of the signal type. Moreover, reinforcement learning uses a greedy layer wise approach empowering quick and dynamic sorts without repetition and accuracy thus measures the correlation between the features to classify the faulty features extremely thereby it takes only less memory constraint with less learning time.

Keywords Robust Harmonized Swan Machine, Stalwart Trippy Classifier, RMS, Crest factor, Kurtosis, feature Extraction;

1. Introduction

Machinery rotating mass are often with very demanding performance standards, some of which are complex used in the industry today [1]. Machine failure is the major challenge because a failure of a reliable lead-time is not possible to be predicted without effective evaluation which is consequently resulting in costly downtime that can be devastating. Therefore, effective and efficient condition monitoring and fault diagnosis is essential for the industry [2, 3]. However, the diagnosis of faults in rotating machinery is often a labour-intensive and time-consuming. Effective and efficient fault diagnosis is always a challenging task for the technicians and plant diagnostics [4]. Faults in rotating machinery mainly include bearing defects, stator faults, rotor faults or eccentricity. According to statistics, nearly 50% of the faults of rotating machinery are related to bearings [5].

In order to ensure the high reliability of bearings and reduce the downtime of rotating machinery, it is extremely important to detect and identify bearing faults quickly and accurately [6]. The rolling bearing fault diagnosis method has always been a research hotspot. The vibration signals of rolling bearings often contain important information about the running state. When the bearing fails, the impact caused by the fault will occur in the vibration signal [7, 8]. Therefore, the most common application of the bearing fault diagnosis method is to use the pattern recognition method to identify the fault by extracting the fault features of the bearing vibration signal [9].

Continuous condition monitoring and real-time fault diagnosis play an invaluable function, which not only contributes to early detection and diagnosis of fault information, but also allows fault prognosis to provide support for critical maintenance decisions. On this basis, a wide range of practical applications for condition-based failure diagnosis have been made in the literature, including bearing pumps, power transmission systems and power supply [10]. Previous studies have shown that the accuracy of the diagnostic results depends, to a large extent, on the extracted fault features normally acquired by time-frequency representations, which allow for the identification of impacts due to different types of damage, and on the different visual information and measured parameters between the impacts to be indexed by the classification methods to be judged. Researchers also explored the use of prognosis and health management for revolving devices, and addressed traditional methods for feature extraction, fault diagnosis, performance evaluation, and deterioration prediction. However, due to higher rotary machinery system complexity and heterogeneity of sensory data, the effective diagnosis of multiple health state classifications based on sensory data with strong ambient noise and fluctuations in working conditions remains a problem and a major challenge for the application of the proposed methodologies in complex engineering systems due to possible information loss and external influences.

However, the rolling bearing vibration signal is nonlinear and non-stationary. It is easily affected by the background noise and other moving parts during the transmission process, which makes it difficult to extract the

fault features from the original vibration signal, and the accuracy of the fault diagnosis is seriously affected. Traditional time-frequency analysis methods have been used in bearing fault diagnosis and have achieved corresponding results, such as short-time Fourier transform and wavelet transform. However, all these methods have defects in the lack of adaptive ability for bearing vibration signal decomposition [11]. For complex fault vibration signals, relying only on subjectively setting parameters to decompose signals may cause the omission of fault feature information and seriously affect the performance of fault diagnosis.

After the feature extraction, the classifier should be utilized to realize automatic fault diagnosis in accordance with the machine learning algorithms, including pattern recognition and neural networks, require a large number of high-quality sample data [12]. In this regard, studies based on machine learning techniques and statistical inference techniques have been conducted from multiple aspects to improve the effectiveness of health state classifications, resulting in a number of classic and typical classification methods, such as random forest (RF), filters and the auto associative neural network (AANN), as well as some state-of - the-art updated techniques for the further implementation of fault detection and multiple classification tasks. Consequently, the most important task in such studies is to effectively learn basic feature knowledge from diverse and heterogeneous signals as indicators and to accurately identify various health states based on the indexes learned. However, a problem arises when faced with complex non-linear relationships, the capacities of diagnostic algorithms with simple architectures, e.g. one hidden layer of neural network, will meet limitations. Nevertheless, there is a problem that the capabilities of diagnostic algorithms with simple architectures, such as a hidden layer of neural network, would encounter limitations when dealing with complex non-linear relationships in problems of fault diagnosis [13,14].

In fact, bearing fault identification is controlled by the application environment. In reality, a large number of fault samples cannot be obtained. Therefore, it is crucial that the classifier can handle small samples and have good generalization ability. A support vector machine (SVM), proposed by Vapnik [15], is a machine learning method based on statistical learning theory and the structural risk minimization principle. Since the 1990s, it has been successfully applied to automatic machine fault diagnosis, significantly improving the accuracy of fault detection and recognition. Compared with artificial neural networks, SVM is very suitable for dealing with small sample problems, and has a good generalization ability. SVM provides a feasible tool to deal with nonlinear problems that is very flexible and practical for complex nonlinear dynamic systems. Besides, the combination of fuzzy control and a metaheuristic algorithm is widely used in the control of nonlinear dynamic systems [16].

Bououden et al. proposed a method for designing an adaptive fuzzy model predictive control (AFMPC) based on the ant colony optimization (ACO) [17,18] and particle swarm optimization (PSO) algorithm, and verified the effectiveness in the nonlinear process. The Takagi–Sugeno (T–S) fuzzy dynamic model has been recognized as a powerful tool to describe the global behaviour of nonlinear systems. Li et al. [19] deals with a real-time-weighted, observer-based, fault-detection (FD) scheme for T–S fuzzy systems. Based on the unknown inputs proportional–integral observer for T–S fuzzy models, Youssef et al. [20] proposed a time-varying actuator and sensor fault estimation. Model-based fault diagnosis methods can obtain high accuracy, but the establishment of a complex and effective system model is the first prerequisite.

Yuan Xie and Tao Zhang [21] proposed the rotating machinery fault diagnosis with feature extraction algorithm based on empirical mode decomposition (EMD) and convolutional neural network (CNN) techniques. The fundamental purpose of our newly proposed approach is to extract distinguishing features. Frequency spectrum of the signal obtained through fast Fourier transform process is trained in a designed CNN structure to extract compressed features with spatial information. To solve the non-stationary characteristic, we also apply EMD technique to the original vibration signals. EMD energy entropy is calculated using the first few intrinsic mode functions (IMFs) which contain more energy. With features extracted from both methods combined, classification models are trained for diagnosis.

Yong Lv *et al.* [22] proposed an approach to health degradation monitoring and early fault diagnosis of rolling bearings based on a complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)and improved multivariate multi-scale sample entropy (MMSE). The smoothed coarse graining process was proposed to improve the conventional MMSE. Numerical simulation results indicate that CEEMDAN can alleviate the mode mixing problem and enable accurate intrinsic mode functions (IMFs), and improved MMSE can reflect intrinsic dynamic characteristics of the rolling bearing more accurately. During application studies, rolling bearing signals are decomposed by CEEMDAN to obtain IMFs. Then improved MMSE values of effective IMFs are computed to accomplish health degradation monitoring of rolling bearings, aiming at identifying the early weak fault phase. Afterwards, CEEMDAN is performed to extract the fault characteristic frequency during the early weak fault phase.

Liang et al., [23] presented a rolling bearing fault diagnosis method based on ensemble local characteristic-scale decomposition (ELCD) and extreme learning machine (ELM)is proposed. Vibration signals were decomposed using ELCD, and numerous intrinsic scale components (ISCs) were obtained. Next, time-domain index, energy, and relative entropy of intrinsic scale components were calculated. According to the distance-based evaluation approach, sensitivity features can be extracted. Finally, sensitivity features were input to extreme learning machine to identify rolling bearing fault types.

Pang et.al., [24] presented a bearing fault diagnosis method, namely an improved Hilbert time–time (IHTT) transform, by combining a Hilbert time–time (HTT) transform with principal component analysis (PCA). Firstly, the HTT transform was performed on vibration signals to derive a HTT transform matrix. Then, PCA was employed to de-noise the HTT transform matrix in order to improve the robustness of the HTT transform. Finally, the diagonal time series of the de-noised HTT transform matrix was

extracted as the enhanced impulsive fault feature signal and the contained fault characteristic information was identified through further analyses of amplitude and envelope spectrums.

Li et.al. [25] Presents a bearing fault diagnosis method based on fully-connected winner-take-all auto encoder. The model explicitly imposes lifetime Sparsity on the encoded features by keeping only k% largest activations of each neuron across all samples in a mini-batch. A soft voting method is implemented to aggregate prediction results of signal segments sliced by a sliding window to increase accuracy and stability.

Lu et.al. [26] explores an effective and reliable deep learning approach known as Stacked Denoising Autoencoder (SDA), which has been shown to be suitable for certain health status identifications for signals including ambient noise and working condition fluctuations. SDA has become a popular approach to achieving the promised advantages of deep architecture-based robust feature representations.

Shao et.al.[27] developed A novel deep auto-coder feature learning approach to diagnose rotating system malfunction. First, the maximum correntropy is used to design a new deep auto-encoder loss feature to improve the ability to learn from measured vibration signals. Second, the artificial fish swarm algorithm is used to optimize the key parameters of the deep auto-coder to adapt to the signal features. The proposed method is used to diagnose the fault of the gearbox and the electrical locomotive roller bearing.

From the above discussion the extraction of representative feature during fault diagnosis is still challenging because of the presence of faulty signals. The problem of empirical mode decomposition where the mode mixing leads to data loss with the existing fault detection techniques is also addressed. Furthermore the learning time and memory constraints are reduced by the proposed approach.

2. MOTIVATION AND CONTRIBUTION

Motivated by the above discussion, the Condition observing and fault determination of Machinery are basically essential for creative productivity and plant security in current undertakings. At the point when a moving bearing shortcoming is feeble at a beginning time or at a low shaft speed, frail blame highlights are frequently implanted in foundation commotion. So it's anything but a simple undertaking to remove the agent features from the original signal. Numerous researchers have evaluated the issue of blame analysis taken from frail flags and have gained ground at the same time, the extraction of complex vibration signals is extremely troublesome. Many ways to deal with enhanced extraction have been proposed, by blend with the analytic devices: impartial systems, concealed Markov models, or solitary esteem disintegration. At times, great outcomes were acquired, yet these methodologies require the setting of a large number of parameters, but when the flag is very feeble or the commotion solid, these methodologies give poor outcomes. Henceforth we have to inspire the proposed system with an appropriate component extraction technique with a reasonable analytic instrument for making the eminent fault diagnosis as per the vibration signal analysis.

3. ROBUST HARMONIZED SWAN MACHINE FOR FEATURE EXTRACTION AND STALWART TRIPPY ALGORITHM FOR FAULT DETECTION

Condition monitoring and fault diagnosis of Machinery is critically important for production efficiency and plant safety in modern enterprises. When a rolling bearing fault is weak at an early stage or a low shaft speed, weak fault features are often embedded in background noise. So it is not an easy task to extract the representative features from the original signal. Many scholars have reviewed the problem of fault diagnosis taken from weak signals and have made progress but, the extraction of complex vibration signals is very difficult Many approaches to improved extraction have been proposed, by combination with the diagnostic tools: neural networks, hidden Markov models, or singular value decomposition. In some cases, good results were obtained, but these approaches require the setting of a great many parameters, and yet when the signal is quite weak or the noise strong, these approaches give poor results.

Hence in this research, we are proposing “Hasty Fault Diagnosis of a Rotating Machinery based on Stalwart Trippy Classifier with Robust Harmonised Swan Machine (RHSM)” for making the perfect fault diagnosis in vibration signal. The analysis of the vibration signal is a preferred idea for fault diagnosis that is a classification task. For making that pattern recognition effectively, there is a need to have proper faulty feature extraction. According to existing researches, the Harmonised Swan Machine (HSM) is an outstanding technique among Signal processing techniques. However, the main weakness of this machine is its high sensitivity to noise, and it also runs into the problem of mixing modes in which mode mixing causes the mixing of higher-order components with the lower order components due to the Empirical Mode Decomposition (EMD) of this machine. Because of this mode mixing problem, some wanted data will also be loosed. So removing some of the IMFs (Intrinsic Mode Function) can corrupt the data. To overcome this missing sample problem, we propose a Robust Harmonised Swan Machine (RHSM) with a new decomposition method, termed Iterative Threshold VMD (Variational mode decomposition). Moreover, there is a difficulty to measure the correlation between the features with the existing fault diagnosis tool named “Trippy classifier “and also it considers more learning time and more memory constraints which makes the learned concept difficult to understand for classifying the faulty features prominently. To attain the perfect diagnosis, the eminent Neuro-Hybrid system is to be adapted in “Trippy classifier”. For that, we propose the Stalwart Trippy Classifier with the Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS). By this reinforcement learning it exactly measures the correlation between the features to classify the faulty features eminently thereby it takes only less memory constraint with less learning time.

Fault detection is a fundamental procedure in the scrutiny of vibrational signals yielded from turning hardware. Vibration signals must be investigated top to

bottom for that there is an expanded need to separate the features, for example, RMS, Kurtosis and Crest factor from each signal unit through acquiring features from every individual unit is gloomier. To examine every niche and corner, here we mounded a Robust Harmonized Swan Machine (RHSM), a direct nature roused system dependent on the character of Swan bird which has the stochastic nature to precisely extricate the destitute things dependent on the characteristic for feature extraction which thinks about each signal unit.

Formally, the useful features are precisely gotten by the guide of Hilbert Huang Transform which uses Iterative Threshold VMD (Variational mode decomposition) for successful IMF (Intrinsic Mode Function) extraction. The features are found out exclusively dependent on the faulty features utilizing DBN (Deep Belief Network) which at end yields the learned faulty features. For an effective fault diagnosis process, there is an expanded need to precisely characterize faulty signals in lessened time which is finished with the Stalwart Trippy classifier motivated from the prey getting conduct of Trippy fish. Stalwart Trippy fish evaluates all plausibility of prey habitation by its learning-based seeking conduct. Essentially this classifier uses the random forest algorithm which makes a rundown of accessible yields yet it experiences precise prediction. The end goal to handle the exact expectation is offered by the DENFIS which uses the knowledge-based rule for prediction. In this way, the precise order of faulty occurrence is finished utilizing Stalwart trippy classifier. The procedures utilized for every one of these steps is clarified and the general proposed architecture portrayed beneath. The block diagram is depicted below in figure 1,

This proposed architecture shows Robust Harmonized Swan Machine with stalwart TRIPPY Algorithm for feature extraction and Fault Diagnosis which aims at improving automatic identification of faults accurately. At first, the framework starts by gathering machine created crude data from proper areas by the data acquisition technique. The gathered crude vibrated data can be extremely loud, and be exposed to different natural contaminants. To detect features effectively notifying and eradicating noise, Robust Harmonized Swan Machine was developed. This feature extraction method consequently pre-prepared the crude data with the assistance of rapid preparation and uses the benefit of DBN accordingly focusing on the extraction of just required RKC (Root mean square, kurtosis, Crest factor) features. RHSM decides the features of the original signal for each instantaneous frequency independent of signal type by the greedy layer-wise learning empowering quick and dynamic sorts without repetition and clamor. Each signal's RKC features are resolved and learned which holds the faulty signal information. After this procedure with the end goal to recognize the fault diagnosis naturally, a scholarly example arrangement technique named Stalwart trippy classifier is presented. At long last fault and normal feature are classified which depends on the pattern matching idea and the predicted output is at long last recovered. The process is as shown in figure:2 and the detailed explanations are given in following sections

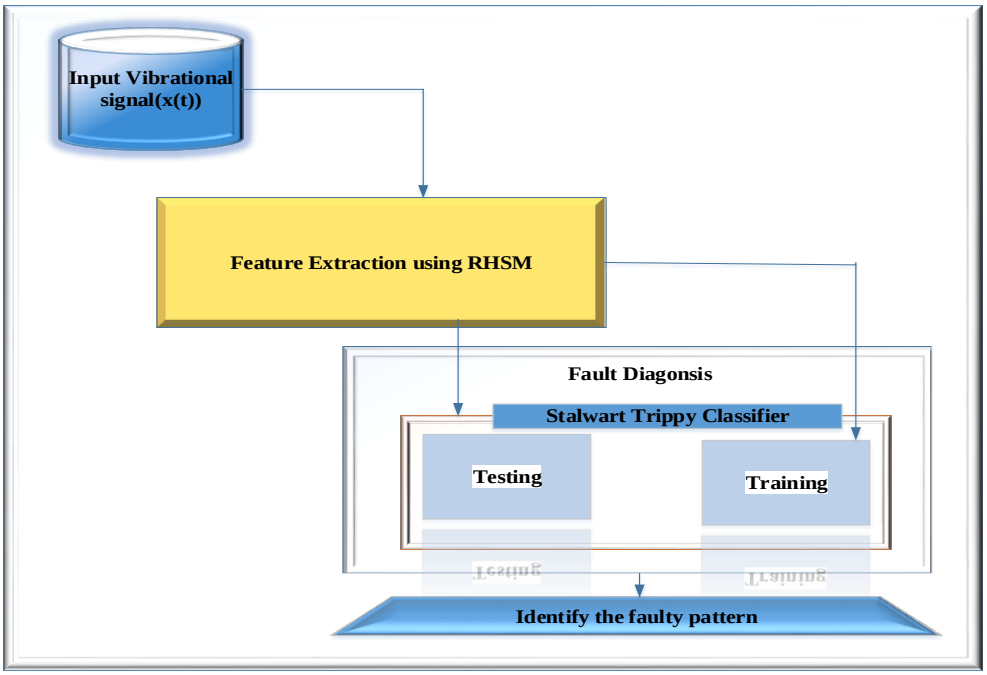


Figure 1: Basic Block Diagram of Proposed Methodology

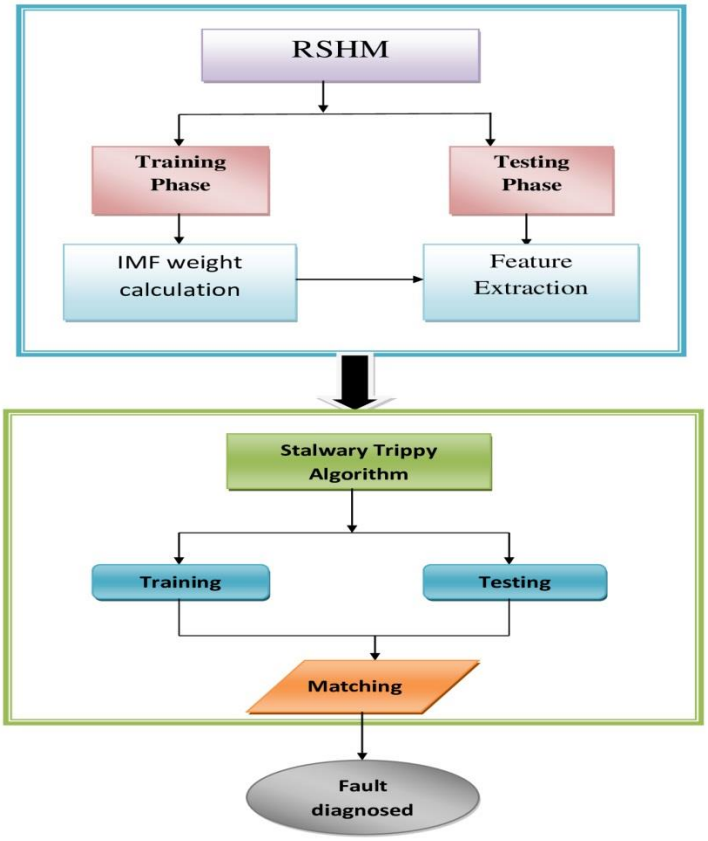


Figure 2. Process flow diagram

3.1. Data Acquisition

Data acquisition is the underlying advance of fault diagnosis, where machine singularities are estimated for further investigations and this paper is centered on vibrated signal-based fault diagnosis as described in equation (1).

$$B_i(t) = \{B_1(t), B_2(t).....B_m(t)\} \quad (1)$$

Let B_i *(t)* is the arrangement of data of vibrated unique signal from other mechanical segments of the machine. In turning apparatus, any element can be ascertained on the crude vibration data and the data must be pre-processed early, however in this work the crude data is consequently pre-processed with the assistance of our proposed feature extraction procedure. After data acquisition the vibrated signals are passed to the feature selection process.

3.2. Robust Harmonized Swan Machine (RBSH) for Feature Selection

At first vibrated data are sending to the feature extraction stage for the expulsion of vibrated commotion and exact extraction of proposed features. With the end goal to remove the precise features we have created most slam against strategy RHSM which is encircled out dependent on the motivation of the exact specific detachment of needy components by a swan behavior.Here the needy components to be the strong highlights RKC and the particular procedure is finished by using DBN thusly for exact choice HHT is used as the stochastic social of swan flying creature. The RHSM pre-preparing strategy treats each consecutive pair of layers in the learning procedure, whose joint likelihood is characterized as,

$$P_{h,u}(h,v) = \frac{1}{j(\omega)_{h,v}} . e^{(B_i(t)wh + v^T b + a^T h)} \quad (2)$$

The above equation (2) portrays the RBM (Restricted Boltzmann Machine) that has deep learning nature which comprised in each layer ‘h’ and shapes the underlying DBN structure, and after that, the regressive adjusting is connected to consistent variable ‘v’. DBN utilizes the marked information train the conditional probability which has an indistinguishable frame from that in the DBN layer, and the error spread starts to finish to tweaking the system. Contrasting the yield estimation of the deep system yield unit with the real esteem, figure the error esteem caused by the weight w. The error esteem is transmitted back to get the error caused by each layer, and afterward, the effective weight estimation procedure to extricate the fault component is computed. Presently it is important to calculate weight with appropriate inclination b to alleviate the preparation error which is given in

$$P_{h,u}(h,v) = \frac{1}{j(\omega)_{h,u}} . e^{(B_i(t)wh + (v-b)^T (v-b) + a^T h)} \quad (3)$$

The RBM parameters can be efficiently trained in an unsupervised fashion by maximizing the likelihood of the joint probability in equation (2)

$$K = \Pi_t \sum_h P_{h,u}(h, B(t)) \quad (4)$$

This K is overtraining samples of vibrated signal *B(t)*. With the end goal to separate the IMFs of a perplexing signal in feature extraction, the training phase of RHSM

calculation receives the iterative threshold VMD technique to decide the mono part of the first vibrated crude signal. Regardless of whether a signal contains missing qualities, this iterative threshold VMD breaks down a flag into meaningful IMFs successfully. That means, by the coupling of thresholding of period gram and an iterative estimation of every mode fulfilling a self-consistency nature in this decomposition method has settled that missing example issue prominently. After iterative threshold VMD, a signal B (t) can be expressed as in equation 5,

$$B(t)=\sum_{k=1}^{N} z_k(t)+y_n, k=1,2,.....N \quad (5)$$

To precisely extricate the missing example in the flag, consider the ceasing rule is used for the foundation of combination for some $\delta\rangle 0$ which are given beneath,

$$\sum_{K}\frac{\left\|z_k^{(m+1)}-z_k{}^m\right\|_2^2}{\left\|z_k^m\right\|_2^2}\langle\delta \quad (6)$$

Where is its decomposed IMF of the signal, is the uproarious signal. Here the iterative threshold VMD decides the mono segment of the first signal independently however there is an all-inclusive need to analyze each mono-component separately dependent on its instantaneous recurrence, so the viable features can be gotten from each individual of vibrated signal and likewise what's more this decomposition technique forms results with some constrained measure of clamor, which is added to be handled. The above expressed issues are handled by taking the Hilbert Haung transform which impacts a specific preprocessing of the useful signal features acquired. Accordingly, HHD is connected to the two sides of Eq. (5), the Hilbert Haung range of B (t), Q (ω, t), might be done by the accompanying condition (6):

$$Q(\omega,t)=S\,\mathrm{e}\sum_{i=1}^{n} B_i(t)e^{l\int\omega_i(t)dt} \quad (7)$$

Where Se is the operator of the real part, $B_i(t)$ and $\omega_i(t)$ denotes the functions of the amplitude and instantaneous frequency, respectively. Note that the residual term y_n in Eq. (5), which involves next to no vitality of the signal, is overlooked that is the commotion is maintained a strategic distance from which is equal to preprocessing of the signal by using the RHSM nature dependent on eqn (7) , which thus diminishes the time factor for extraction of faulty features.

Accordingly, the marginal spectrum of Hilbert–Huang transform which extract RKC features can be defined by an integrated spectrum with respect to time as in equation (8)

$$Q(\omega)=\int_0^T Q(\omega,t_{(RMS,Crest\ Factor,Kurtosis)})dt \quad (8)$$

Where T is no of the features of $B(t)$ and $Q(\omega,t)$ can precisely define the extraction of efficacious faulty features such as RMS, Kurtosis, and Crest of each IMF mono-component with characterized time interim. The benefit of our extraction strategy with the assistance of HHT is to acquire all components extracted within instantaneous frequency restrict without clamor presence.

To such an extent that ascertaining IMF and Residue utilizing RHSM, the machine commotions are destroyed and crude information are molded naturally with the assistance of RHSM Training. In this way all in all element Extraction by Robust Harmonized Swan Machine (RHSM) is the procedure of definitely removing required features for the faulty flag from the crude vibrated flag. The general design of Robust Harmonized Swan Machine is depicted underneath fig 2.

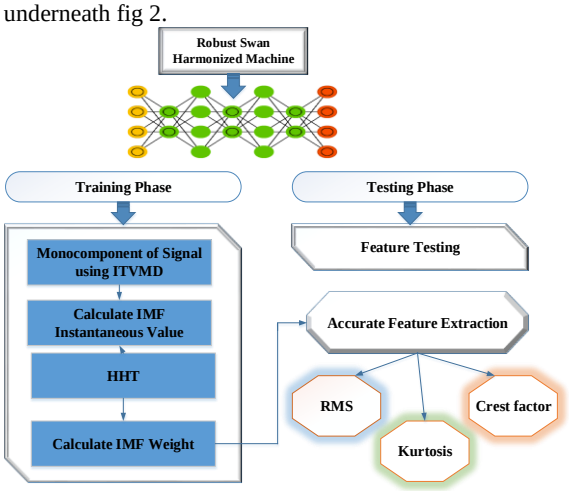


Figure 2: Robust Harmonized Swan Machine schematic architecture

Ordinarily, the factors, for example, RMS, Kurtosis, and Crest factor are believed to be noticeably anticipated criteria of a faulty signal when contrasted and typical vibrated signals. Along these lines by breaking down the RKC features it is less demanding for thumping out of the faulty signals without random-looking which is the purpose behind deeply learning the three-parameter RKC. This RKC translates as the way to pass judgment on the faulty signals that happened. The depiction of RKC Features are given underneath

3.2.1. RMS

The root mean square (RMS) estimation of a vibration signal is a time analysis feature that is the proportion of the power content in the vibration signature. This feature regards track the general clamor level, yet it won't give any data on which segment is fizzling. It very well may be exceptionally compelling in identifying a noteworthy out-of-balance in pivoting frameworks. The following is the condition that is utilized to ascertain the root mean square estimation of a prepared information arrangement, u(t)n over length N.

$$RMS=\sqrt{\frac{1}{N}}\times\sum_{n=1}^{n}B(t)_n^2 \quad (9)$$

3.2.2. Kurtosis

Kurtosis is characterized as the conveyance and measures the relative peakedness or levelness of a dissemination when contrasted with a typical dispersion. Kurtosis gives a proportion of the measure of the tails of appropriation and is utilized as a marker of significant crests in an arrangement of information. As a rigging wears and breaks this feature should signal a blunder because of the expanded level of vibration. The condition for kurtosis is given by:

$$Kurtosis = \frac{\sum_{n=1}^{N}\left[B(t)-\mu\right]^4}{N\times(\sigma^2)^2} \quad (10)$$

where $v(t)$ is the raw time series at point n, μ is the mean of the trained data, $\sigma^2$is the variance of the data, and N is the total number of data points.

3.2.3. Crest Factor

A superior measure is to utilize "crest factor" which is characterized as the proportion of the pinnacle level of the info signal to the RMS level. In this manner, tops in the time arrangement signal will result in an expansion in the crest factor esteem. The following is the condition for the crest factor:

$$Crestfactor = \frac{Peak\ Level}{RMS} \quad (11)$$

At last features are extricated consummately without commotion and excess. The fundamental preferred standpoint of our proposed highlight extraction system is the flexibility of all signal kind because of the joint probability distribution parameter of RBM. After getting features, an astute example order is expected to satisfy the defective conclusion naturally by the gathering of each one of those separated features. In this work Stalwart TRIPPY Algorithm proposed to satisfy the blame finding naturally by gathering every one of those removed monochrome features which is accomplished by the predictable ensemble nature of the classifier and is proposed underneath.

3.3. Fault Diagnosis Using Stalwart Trippy Algorithm

The Precise classification is accomplished in our work by using most spearheading nature motivated classifier absolved as stalwart trippy classifier dependent on the profound looking and exact prescient nature of trippy angle though this classifier offers a profound learning or seeking by methods for the randomized character of RF (Random forest) additionally prescient capacity by methods for DENFIS with information based earlier judging. In this manner, once the scholarly flawed highlights are fed to this classifier which are investigated profoundly and offer precise determination which in turn isolates the characterized faulty signals.

Primarily the harvest outcomes L_i (1) and L_i (2) of the trained HSM should be fed into the trippy classifier which compromises deep learning by means of RF. The data fusion training of the faulty features is defined as

$$\left\lfloor \begin{matrix} {}^{\circ(1)}_{i} \quad arq_i \max K_i^{(1)} = arq_i \max j(\omega_i) \\ {}^{\circ(2)}_{i} \quad arq_i \max K_i^{(2)} = arq_i \max j(\omega_i) \end{matrix} \right\rfloor \quad (12)$$

Where i=1, 2, 3… m,

$$where\left\lfloor j(\omega_i) = j(\omega_1) + j(\omega_2) + ... + j(\omega_m) \right\rfloor \quad (13)$$

The Learning Features $K_i(\ j(\omega_i))$ is imported from equation 7 that consist of RKC, which is utilized for diagnosing the fault.

$$K_i = \left[K_1^{(1)}, K_2^{(2)}, ..., K_m^k \ \right] \quad (14)$$

The above equation (14) is utilized to develop a tree with various bootstrap test from unique information utilizing a tree classification calculation. Where 'm' is the quantity of highlights which are removed from RHSM. After the timberland is framed, protest that should be prepared is put down under every one of the tree in the forest for preparing. The preparation highlights are depicted in the equation (15)

$$\omega_i = K_i \max j(\omega_i) \quad (15)$$

Following, precise forecast procedure of the Stalwart trippy classifier after profound learning is finished by using the Neurofuzzy interference system which uses information based example recognizable proof for the expectation of vibrated blame signal. Here we are using a DENFIS demonstrate which utilizes a dynamic Takagi-Sugeno fuzzy inference system. Notwithstanding powerfully making and refreshing fuzzy rules amid the learning procedure, the fluffy guidelines that take part in the derivation for each new info vector are progressively browsed the current fuzzy principle set contingent upon the situation of the current information vector in the input space. DENFIS is made out of m fuzzy standards showed as pursues:

if x1 is Rm1 and x2 is Rm2 and … and xq is Rmq, then y is fm(x1, x2, …, xq) (16)

where “xj is Rij”, i = 1, 2, … m; j = 1, 2, … q, the overall output will be

$$y^{o} = \frac{\sum_{i=1}^{m} w_i\, f_i\left(x_1^{o}, x_2^{o}, ...x_q^{o}\right)}{\sum_{i=1}^{m} w_i} \quad (17)$$

Where,

$$w_i = \prod_{j=1}^{q} S_{ij}\left(x_j^{0}\right); i = 1, 2, ...m,\ j = 1, 2, ...q$$

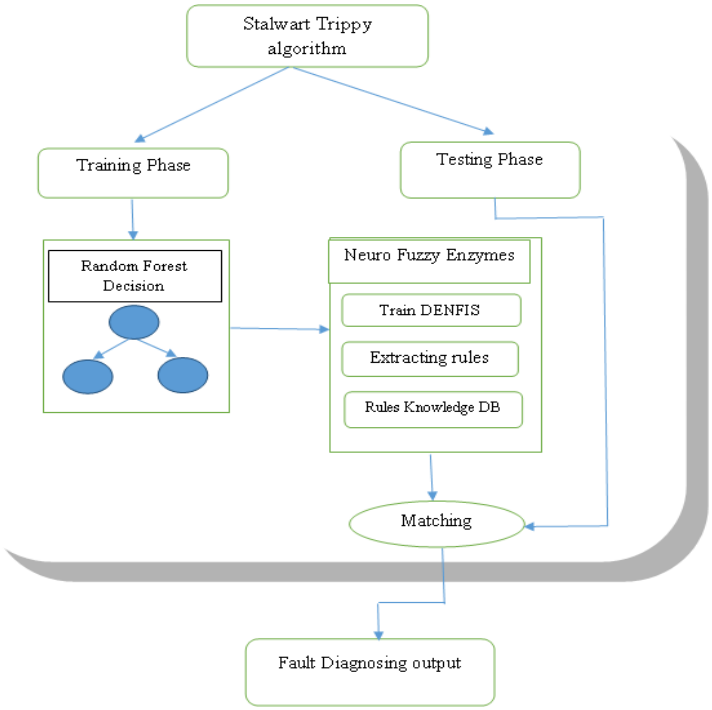


Figure 3. Stalwart Trippy Classifier for fault Diagnosing

In this manner by the use of proficient feature extraction, and by using RHSM and the keen pattern recognition by methods for Stalwart trippy classifier correctly characterizes blame with the help of learning based feature extraction. This is the sufficiently bright process because of the expellation of preprocessing and precise prediction process which wipes out the regarded use of time. The outcome approval in the underneath segment will be an additional verification for the productivity of the work. Thus efficient outputs in fault diagnosis of vibrating signals are done by utilizing this framework and the results are discussed below:

4. RESULTS AND DISCUSSION

The proposed technique is described in previous section 3 and in this section the detail explanation and its performance is analyzed.

4.1. Materials and methods

The proposed method is implemented in the working platform of MATLAB with the following system specification.

Platform : MATLAB 2015a
OS : Windows 8
Processor : Intel core i5
RAM : 8 GB RAM

Dataset was collected on a single stage reciprocating type air compressor placed at the Department of Electrical Engineering Workshop. The setup consists of a single stage reciprocating air compressor, in which the common faults were simulated. The data acquisition process were carried out using a piezoelectric accelerometer connected to NI Data acquisition (NIDAQ)system, where analog data was acquired by fast Fourier transform method as digital data and the processed data were used to classify the different faults acquiring the statistical features from the sensor data. Specifications of the air compressor are as follows:

- Air Pressure Range: 0-500 lb/m2, 0-35 Kg/cm2
- Induction Motor: 5HP, 415V, 5Am, 50 Hz, 1440rpm
- Pressure Switch: Type PR-15, Range 100-213 PSI

The computation part of the experiment was done on a desktop computer having an i7 3770 processor, and 8 GB of RAM. Acoustic recordings were taken from the air compressor in 8 different designated states. These 8 states include a healthy state, and 7 faulty states: Leakage Inlet value (LIV) fault, leakage outlet valve (LOV) fault, non-return valve (NRV) fault, piston ring fault, flywheel fault, rider belt fault, and bearing fault. To get recordings from all these states, a similar environment was simulated by seeding faults into the air compressor. Details of the different air compressor states are given below [28].

Table 1. Appendix for datacollection

NIDAQ	National Instruments Data Aquisition
LIV	Leakage Inlet value
LOV	Leakage Outlet valve
NRV	non-return valve

4.2. Simulation result

The vibrated bearing fault raw dataset are plotted in the given below

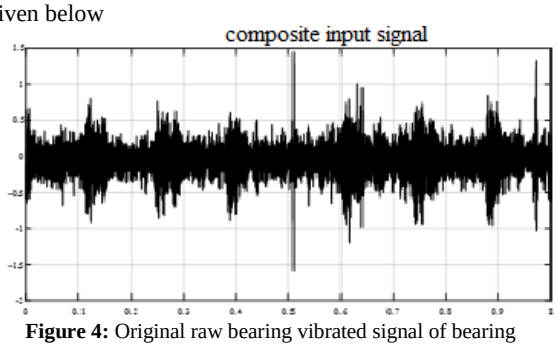


Figure 4: Original raw bearing vibrated signal of bearing

The vibrated signal is concentrating on the feature extraction like RMS, kurtosis, Crest factor. In this Paper Robust Harmonized Swan Machine is used for exact extraction of feature. ITVMD decides the mono part of unique signal. Our proposed natural mode capacity of the original signal is depicted in fig 5, 6 and 7.

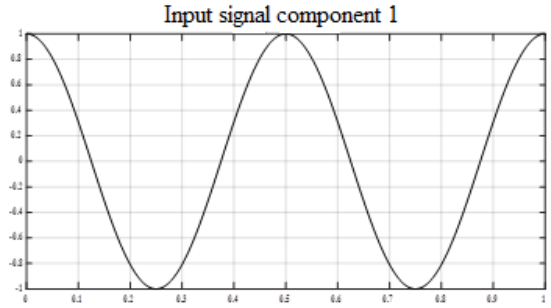


Figure 5: IMF 1 values of original signal

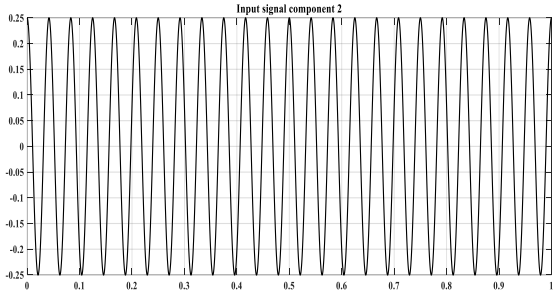


Figure 6: IMF 2 values of original signal

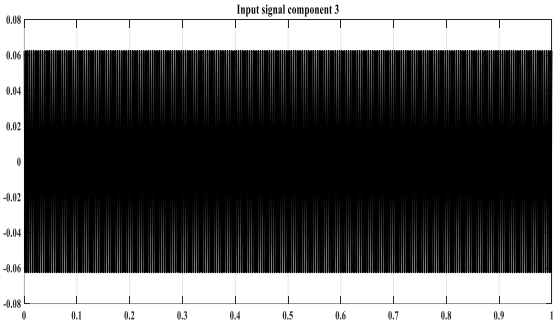


Figure 7: IMF 3 values of original signal

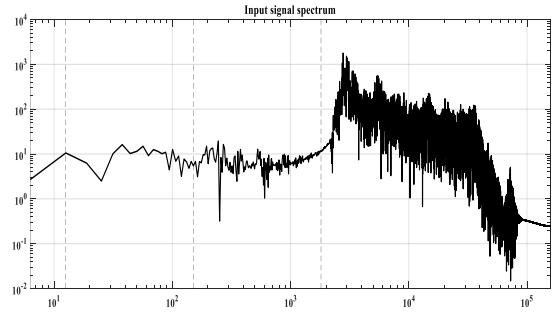


Figure 8: Spectrum of the input signal

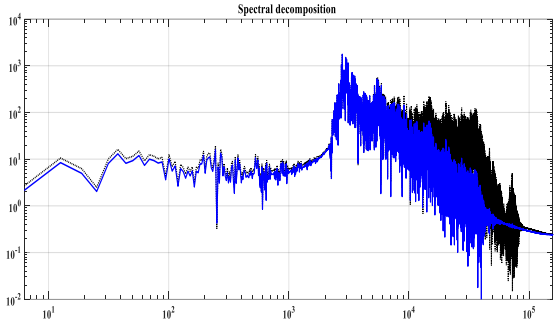


Figure 9: ITVMD spectral decomposition

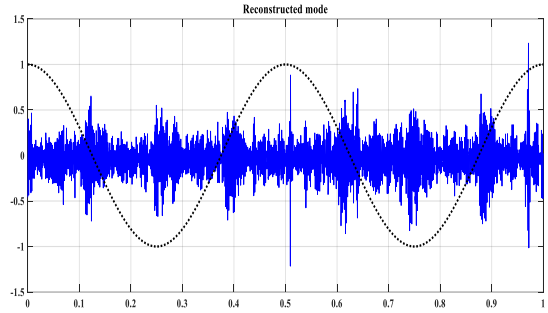


Figure 10: ITVMD Reconstructed modes

After ITVMD technique the instantaneous frequency is calculated using HilbertHaung Transform. Below, figure 10 shows that the spectrogram of HHT of instantaneous frequency signal.

Figure 11: ITVMD based Hilbert-Huang transform

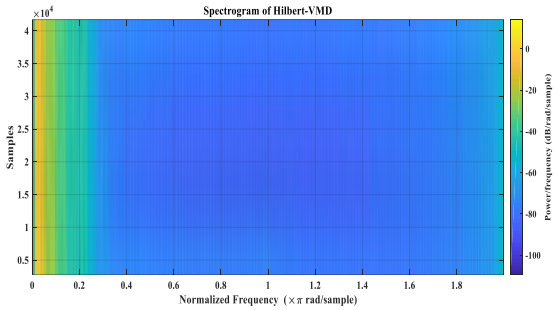


Figure 12: Spectrogram of ITVMD- Hilbert Haung Transform

After that procedure some flawed signal are accessible in the above data. With the end goal to evade this issue RHSM is used for profound learning. The learning conduct of RHSM is utilized to extricate the required feature for blame finding in exact path with the assistance of IMF computation dependent on the idea of RHSM which is broadly utilized monochrome of individual recurrence signal. At long last exact commotion features are disposed

of and our required features, for example, RKS are precisely separated inside the less time. After this procedure our proposed feature extraction are portrayed in the underneath fig 13, 14, 15.

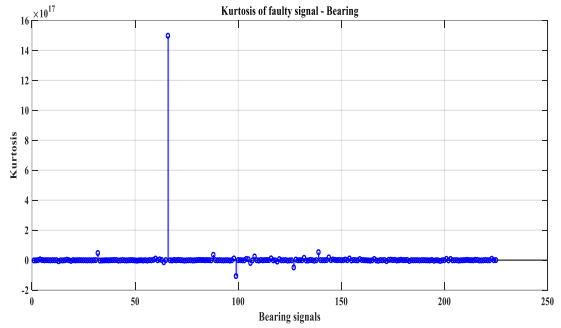


Figure 13: Kurtosis of bearing

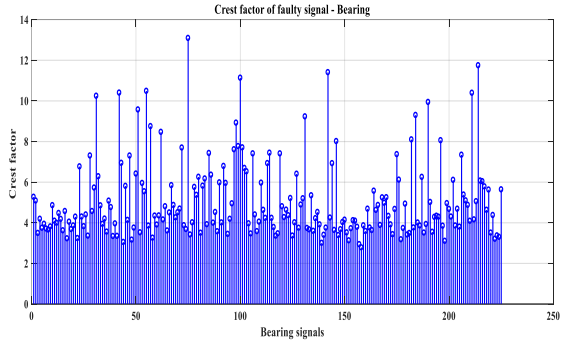


Figure 14: Crest factor of bearing

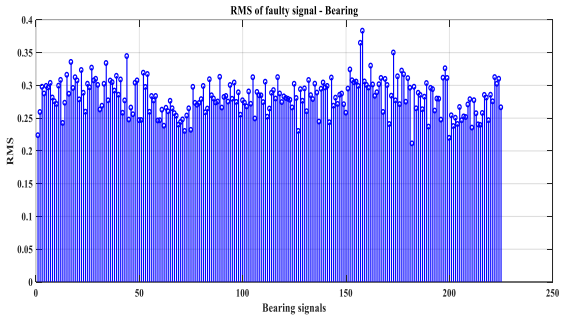


Figure 15: RMS of bearing

With the end goal to anticipate the precise vibrated signal, the insightful based example order is required, so in this examination paper, it centers around Stalwart TRIPPY for predicting the exact defective signal consequently. The random forest decision tree doesn't anticipate the exact signal since it has apriori learning less nature

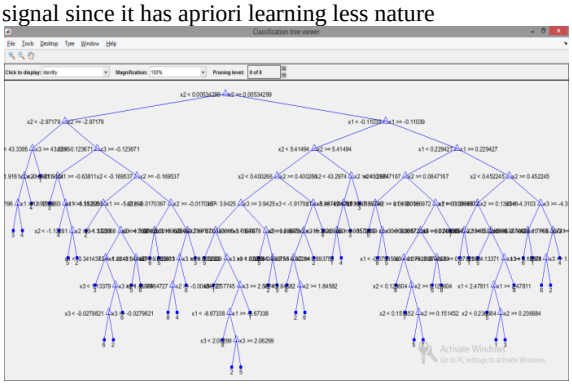


Figure 16: Stalwart Trippy algorithm decision Tree

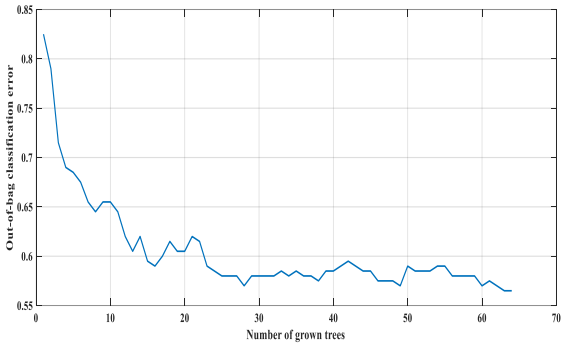


Figure 17: Proposed accurate faulty signal detection based on decision Tree

In the above fig 17 depicts the faulty signal detection in the revolution machine utilizing Stalwart TRIPPY Classifier. Finally the precise conduct of faulty state are distinguished in best way. Our proposed calculation has information based standard programmed detection is accustomed to diagnosing the faulty signal.

4.3. Comparision Analysis

Comparison was made by proper analysis of computation time, accuracy,Diagonosis Accuracy,Testing Prediction Time and Diagonosis Noise are described below section.

4.3.1. Computation time

Computation time amid feature extraction is characterized as the time required for separating the fundamental features from crude vibrated information. The calculation equations are depicted underneath.

$$\text{Computation Time} = \textit{Starting Time of featureextraction} - \text{Ending Time of featureextraction}$$

Table 2: computation Time of proposed transform and existing Transform

Sl.no	Algorithm	Run Time
1	TDSP	0.034
2	Fast flourier Transform	0.013
3	Morlet wave let Transform	0.072
4	Discrete wave let transform	0.092
5	Short time Fourier transform	0.648
6	Wigner wile Distribution	0.008
7	Pseudo- Wigner wille Distribution	2.120
8	Auto correlation	0.028
9	Updated more let Transform	0.029
10	Convolution with sine	0.012
11	S-transform	5.328
12	Proposed	0.019

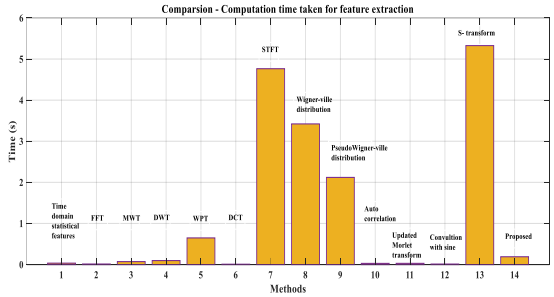


Figure 18: comparison of proposed and existing transform for feature extraction

In the above, fig 18 exhibits the normal calculation time per recording for processing each change and its comparing highlights. These qualities were found by averaging the calculation time more than 1000 chronicles, which depicts the computation time for existing change for dissecting the chosen includes, those utilizing FFT, WPT, DCT, WVD, STFT the vales are portrayed in table 1 and our proposed highlight extraction which use preparing with ITVMD based HHT achieve the lower computation time when contrasted with existing works.

4.3.2. Accuracy

The accuracy of a test is its capacity to separate the faulty and healthy condition effectively. To gauge the accuracy of a test, we ought to figure the extent of genuine positive and genuine negative in all assessed cases. Numerically, this can be expressed as:

$$Accuracy = \frac{a+d}{a+d+c+b} \quad (18)$$

Where, True positive (a) = the number of features correctly identified as faults,

False positive (c) = the number of features incorrectly identified as faults,

True negative (d) = the number of features correctly identified as Normal,

False negative (b) = the number of features incorrectly identified as Normal.

Table 3: Accuracy comparison of various bearing and proposed bearing

Sl.no	Algorithm	Total (sec)
1	KNN	86.67
2	PNN	90
3	RBN	96.67
4	PSO-SVM	96.67
5	PROPOSED	**98.675**

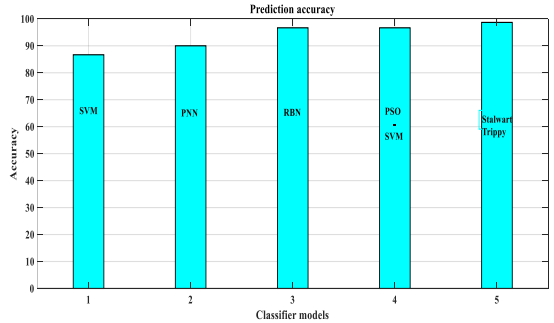


Figure 19: Accuracy comparison for various bearing method of existing and proposed

In the above, fig 19 shows the contributions of the Accuracy prediction time per recording for processing each change and its relating fault diagnosis, which depicts the accuracy for existing and breaking down the diagnosing the fault, those utilizing KNN, PNN, RBN, PSO-SVM, and proposed the qualities are portrayed in table 2 and our proposed stalwart Trippy algorithm, which achieve the higher accuracy when contrasted with existing algorithm.

4.3.3. Testing prediction Time

The testing prediction time is defined as the time taken to predict the precise fault.

Table 4: overall prediction using existing and proposed

Sl.no	Algorithm	Total (sec)
1	KNN	0.51
2	PNN	0.064
3	RBN	20.6307
4	PSO-SVM	0.05033
5	PROPOSED	0.061

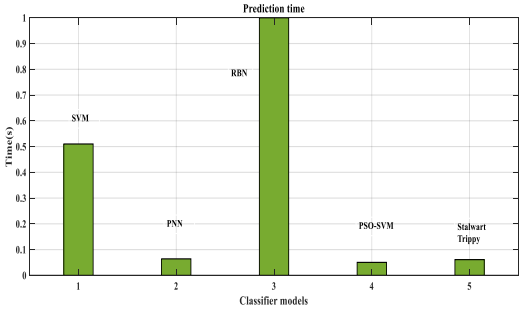


Figure 20: overall prediction time using existing and proposed

In the above, fig 20 portrays the general prediction examination of our proposed algorithm and existing algorithm, for example, KNN, PNN, RBN, PSO-SVM and proposed esteems are plotted in the Table:3. At long last our proposed Stalwart Trippy classifier accomplishes less prediction time when contrasted with all other existing algorithm, and furthermore it accurately determines the fault to have 0.061 secs which is very higher. This time is worthy since we do preparations twice and we accept this less time.

4.3.4. Comparison of Diagnosis Accuracy

The Diagnosis Accuracy is characterized as the general likelihood that a fault will be effectively arranged dependent on the learning test informational collection. The Diagnosis Accuracy is portrayed underneath

$$Diagnosis\ Accuracy = \frac{a+d}{a+b+d+c} \quad (19)$$

Table 5: Comparison of Diagnosis Accuracy Proposed and Exiting Method

Sl. No	Features Algorithm	Classifier Algorithm	No.of training sampes	No.of testing samples	No. of classes	Diagnosis accuracy
1	LCD-SVD	CRO-SVM	240	80	4	100
2	TDF-FDF	PNN	240	80	4	94.38
3	TDF	Random forest	200	200	4	98.04
4	HSM	Trippy	225	50	8	98.675

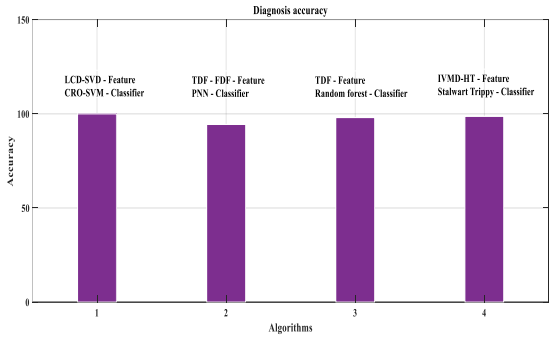


Figure 21: Diagnosis Accuracy of different and our proposed Method

Table 5 depicts the preparation and testing tests for accurate feature selection utilizing distinctive classes for features extraction. By and large point of view our proposed highlight determination procedure achieves the higher accuracy for choosing the component on the grounds that the preparation informational index are immense contrasted with existing works, for example, LCD-SVD, TDF-FDF and TDF which is effectively feasible to separate the element. In any case, our proposed highlight such RMS, Crest factor and Kurtosis, which are not effectively achievable to remove the component but rather in this work our proposed calculation to extricate the precise required element effortlessly with the assistance of stochastic nature. At long last our proposed diagnosing accuracy esteem is being expanded 98.28 when contrasted with all other existing classifier like CRO, PNN, and Random forest.

4.3.5. Comparison of diagnosis noise

Diagnosis Noise is most easily defined via the mean squared error (MSE). Given a noise-free m×n monochrome feature I and its noisy approximation K, MSE is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-k(i,j)]^2 \quad (20)$$

PSNR (in dB) is defined as,

$$PSNR = 20.\log_{10}(MAX_I) - 10.\log_{10}(MSE) \quad (21)$$

Here, MAX_I is the greatest conceivable blunder estimation of the vibrated signal. At the point when the issues are spoken to utilizing 8 bits for each example, this is 255. All the more for the most part, when tests are spoken to utilizing direct PCM with B bits per test.

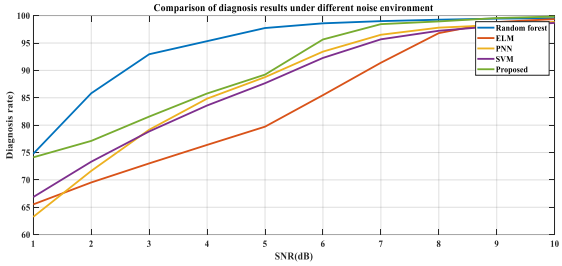


Figure 22: Diagnostic results of employed classifiers with different SNRs.

In the above fig 22, it is clear that Stalwart Trippy Classifier accomplishes a factually critical prevail upon the relative techniques when dealing with the test tests with various levels of commotion. As indicated esteems in Table 4, for ELM, PNN ,SVM and stalwart Trippy , a conspicuous descending pattern along these lines shows up when the SNR is lower than 22 dB, while stalwart Trippy still performs fundamentally well in a more extensive scope of SNR attributable to the solid enemy of clamor capacity of our proposed calculation.

Table 6: Comparison of diagnosis results under different noise environment

Diagonsis Classifier	12dB	13 dB	14 dB	15 dB	16 dB	18 dB	20 dB	22 dB	24 dB	26 dB
Random Forest	74.78	85.82	92.95	95.34	97.74	98.60	99.00	99.26	99.46	99.53
ELM	65.53	69.53	73.01	76.39	79.72	85.48	91.40	96.81	98.83	99.37
PNN	63.25	71.64	79.16	84.84	88.77	93.43	96.51	97.81	98.28	98.87
SVM	66.87	73.32	78.85	83.59	87.66	92.28	95.68	97.24	98.11	98.67
Proposed	74.12	77.12	81.56	85.78	89.23	95.65	98.45	98.95	99.45	99.84

4.3.6. Time consumption of prediction in different classifiers

Time consumption of prediction is the proportion of aggregate time taken for expectation to the time taken for finishing.

$$TimeConsumption = \frac{Time\ taken\ \ to\ predict\ the\ \ features}{Total\ time\ taken\ \ for\ \ the\ \ completion\ \ of\ \ the\ \ process}$$

Table 7: Time consumption for prediction in different classifiers

Sl.no	Algorithm	Prediction Time
1	Random forest	5.08 ms
2	ELM	0.60 ms
3	PNN	8.44 ms
4	SVM	0.50 ms
5	Proposed	0.45ms

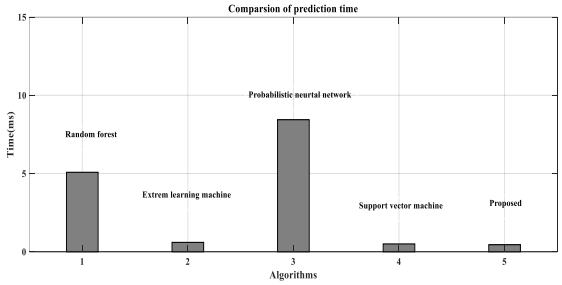


Figure 23: Time consumption of prediction in different classifiers

In the above, fig 23 portrays the general diagnosing forecast time contrasted with existing work and our proposed Stalwart Trippy Classifier which accomplishes less than the expectation time utilizing extensive informational collection for preparing and testing.

4.4. Discussion

Initially the proposed Robust Harmonized Swan Machine extracts the exact features such as RMS, kurtosis and crest factor from the input vibrational signal. The efficiency of our proposed classifier is to diagnosing blame in the turning apparatus in slightest time expectation. In table 6 esteems showed the expectation time examination of various technique, for example, Random Forest accomplishes the forecast esteem 5.08ms, ELM strategy achieve 0.60ms, PNN forecast time utilization esteem is 8.44ms,the forecast of blame diagnosing time SVM technique is 0.50ms and our proposed technique RHSM with Stalwart Trippy Algorithm accomplishes the less expectation time because of the specific prescient nature of trippy classifier, which achieve 0.45 ms forecast time when contrasted with existing our proposed methodology possesses accomplish the better expectation energy for blame determination. In addition the ITVMD based HHT achieve the lower computation time when contrasted with existing works.Thus our proposed calculation consequently identify the blame in the pivoting vibrated machine in beginning period and furthermore spares industry from substantial misfortunes happening because of machine breakdowns.

5. CONCLUSION

This work builds up a methodology for fault diagnosis for pivoting hardware by using RHSM based Stalwart Trippy classifier. RHSM investigates the precise powerful fault features though the stalwart trippy classifier predicts blame event specifically by methods for extracted faulty features. With the use of proficient feature extraction by using RHSM and the keen pattern recognition by Stalwart trippy classifier correctly characterizes blame with the help of learning based feature extraction. The proposed RHSM with Stalwart trippy classifier technique uses an iterative mechanism thus efficiently overcomes the missing sample problem with less learning time and memory constraints. Formally, the above techniques are effectively executed on an air blower and can distinguish the assigned shortcomings with fantastic exactness. Indeed, the blame acknowledgment was finished by examining vibrated signals with computation time of 0.019 sec, 98.675% accuracy and prediction time of 0.045ms utilizing our organized feature extraction and grouping structure. This work can be extended to completely cure the hidden problems of mechanical systems. Therefore, future research should first regard the mechanical equipment as a multi-layered, non-linear complex whole determining the root cause of system failure and the primary failure occurs, thereby resulting in the complete curing the hidden trouble of mechanical system.

REFERENCES

[1] Riaz, Saleem, Hassan Elahi, Kashif Javaid, and Tufail Shahzad, “Vibration feature extraction and analysis for fault diagnosis of rotating machinery-A literature survey,” Asia Pacific Journal of Multidisciplinary Research, Vol. 5, No. 1, pp. 103-110,2017.

[2] Zhang, Xiao Li, Wei Chen, Bao Jian Wang, and Xue Feng Chen, "Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization,” Neurocomputing, Vol. 167,pp. 260-279,2015.

[3] El-Thalji, Idriss, and Erkki Jantunen, “A summary of fault modelling and predictive health monitoring of rolling element bearings,” Mechanical Systems and Signal Processing, Vol. 60, pp. 252-272,2015.

[4] V.N. Patel, N. Tandon, and R. K. Pandey, “Vibrations generated by rolling element bearings having multiple local defects on races,” Procedia Technology, Vol. 14, pp. 312-319,2014.

[5] P.E. William, M.W. Hoffman, “Identification of bearing faults using time domain zero-crossings. Mech. Syst,” Signal Process., Vol. 25, pp. 3078–3088,2011.

[6] X. Zhang, Y. Liang, J. Zhou, Y. Zang, “A novel bearing fault diagnosis model integrated permutation entropy,” ensemble empirical mode decomposition and optimized svm. Measurement, Vol. 69, pp. 164–179,2015.

[7] Rodriguez, Nibaldo, Guillermo Cabrera, Carolina Lagos, and Enrique Cabrera. "Stationary wavelet singular entropy and kernel extreme learning for bearing multi-fault diagnosis." *Entropy* 19, no. 10 (2017): 541.

[8] R. Yuan, Y. Lv, G. Song, “Multi-fault diagnosis of rolling bearings via adaptive projection intrinsically transformed multivariate empirical mode decomposition and high order singular value decomposition,” Sensors, Vol. 18,pp. 1210,2018.

[9] R. Tiwari, V.K. Gupta, P.K. Kankar, "Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier," J. Vib. Control, Vol. 21, pp. 461–467,2013.

[10] Lu, C., Wang, Z.Y., Qin, W.L. and Ma, J., 2017. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, *130*, pp.377-388.

[11] H. Gao, L. Liang, X. Chen, G. Xu, "Feature extraction and recognition for rolling element bearing fault utilizing short-time fourier transform and non-negative matrix factorization," Chin. J. Mech. Eng., Vol. 28, pp. 96–105,2015.

[12] Wu, S.D., Wu, P.H., Wu, C.W., Ding, J.J. and Wang, C.C., 2012. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy*, *14*(8), pp.1343-1356.

[13] Wang, C.C., Kang, Y., Shen, P.C., Chang, Y.P. and Chung, Y.L., 2010. Applications of fault diagnosis in rotating machinery by using time series analysis with neural network. *Expert Systems with Applications*, *37*(2), pp.1696-1702.

[14] Santos, Pedro, Jesús Maudes, and Andres Bustillo. "Identifying maximum imbalance in datasets for fault diagnosis of gearboxes." *Journal of Intelligent Manufacturing* 29, no. 2 (2018): 333-351.

[15] Vapnik, N. Vladimir, "The nature of statistical learning theory," IEEE Trans. Neural Netw., Vol. 38, pp. 409,1997.

[16] Shenghan Zhou, Silin Qian, Wenbing Chang, Yiyong Xiao, and Yang Cheng, A Novel Bearing Multi-Fault Diagnosis Approach Based on Weighted Permutation Entropy and an Improved SVM Ensemble Classifier. Sensors, 18(6), 2018

[17] S. Bououden, M. Chadli, H.R. Karimi, "An ant colony optimization-based fuzzy predictive control approach for nonlinear processes," Inf. Sci., Vol. 299,pp.143–158,2015.

[18] S. Bououden, M. Chadli, F. Allouani, S. Filali. "A new approach for fuzzy predictive adaptive controller design using particle swarm optimization algorithm," Int. J. Innovative Comput. Inf. Control., Vol. 9,pp. 3741–3758,2013.

[19] L. Li, M. Chadli, S.X Ding, J. Qiu, Y. Yang, "Diagnostic observer design for t–s fuzzy systems: Application to real-time-weighted fault-detection approach," IEEE Trans. Fuzzy Syst., Vol. 26,pp. 805–816,2018.

[20] T. Youssef, M. Chadli, H. R Karimi, R. Wang, "Actuator and sensor faults estimation based on proportional integral observer for ts fuzzy model," J. Franklin Inst., Vol. 354,pp. 2524–2542,2017.

[21] Xie, Yuan, and Tao Zhang, "Fault Diagnosis for Rotating Machinery Based on Convolutional Neural Network and Empirical Mode Decomposition," Shock and Vibration, 2017.

[22] Lv, Yong, Rui Yuan, Tao Wang, Hewenxuan Li, and Gangbing Song, "Health degradation monitoring and early fault diagnosis of a rolling bearing based on CEEMDAN and improved MMSE," Materials, Vol. 11, No. 6, pp. 1009,2018.

[23] Liang, Mingliang, Dongmin Su, Daidi Hu, and Mingtao Ge, "A Novel Faults Diagnosis Method for Rolling Element Bearings Based on ELCD and Extreme Learning Machine," Shock and Vibration, 2018.

[24] Pang, Bin, Guiji Tang, TianTian, and Chong Zhou, "Rolling Bearing Fault Diagnosis Based on an Improved HTT Transform," Sensors, Vol. 18, No. 4, pp. 1203, 2018.

[25] Li, Chuanhao, Wei Zhang, Gaoliang Peng, and Shaohui Liu, "Bearing fault diagnosis using fully-connected winner-take-all autoencoder," IEEE Access, Vol. 6, pp. 6103-6115,2018.

[26] Lu, C., Wang, Z.Y., Qin, W.L. and Ma, J., 2017. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, *130*, pp.377-388.

[27] Shao, H., Jiang, H., Zhao, H. and Wang, F., 2017. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. Mechanical Systems and Signal Processing, 95, pp.187-204.

[28] Nwakpang, I.A., Lebele-Alawa, B.T. and Nkoi, B., 2019. Performance Assessment of a Two-Stage Reciprocating Air Compressor. *European Journal of Engineering Research and Science*, *4*(4), pp.74-82.

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665
Pages 237 - 247

Jordan Journal of Mechanical and Industrial Engineering

Designing Food Packages to Attract Customers: A Systematic Approach

Sa'Ed M. Salhieh

Faculty of Engineering & IT, The British University in Dubai, Dubai, United Arab Emirates

Received February 10 2020 *Accepted June 4 2020*

Abstract

Packages are considered an influential tool used to sell food products by communicating certain marketing messages to attract customers and influence their purchasing decision. This paper presents a systematic approach to design food product packages capable of delivering marketing messages intended to attract customers. The approach begins by identifying the main marketing messages that need to be communicated to customers. Next, perceptual maps relating customers' preferences toward existing or new package designs are constructed to assess the package's effectiveness in delivering the marketing messages. After that, package design characteristics capable of embodying the marketing messages are identified and used as design factors within a conjoint analysis study to design new packages. The applicability of the approach was tested by designing new flavored milk packages. The results of the study showed that the approach allowed package designers to systemize their analysis of the effectiveness of food package designs in delivering the intended marketing messages, and to identify main design factors to be considered when designing a new package, and to use these design factors to develop package designs capable of attracting customers.

Keywords: food package design, new product development, package marketing messages, flavored milk, willingness to buy;

1. Introduction

Food packages play a major role in marketing through the messages communicated by the package allowing consumers to instantly recognize products through distinctive shapes, branding, and labeling. Thus, food packages are considered an important communication tool that increases product visibility and sets it apart from its competition. A package can also lead to marketplace differentiation due to its ability to convey relevant product information that influences consumer's perception, preference, and choice (Becker et al, 2011; Sundar & Noseworthy, 2014). Packaging has an important role in influencing in-store purchase, especially for food products where purchase decisions are characterized by low involvement and impulsive processes (Rebollar et. al, 2012).

The marketing function of food packages implemented through its visual appearance is well recognized by practitioners who investigated the effect of several design attributes related to visual appearances such as the material used which was found to have an influence on the way in which consumers perceive the product and the ideas that they generate about its characteristics (Mutsikiwa&Marumbwa, 2013, Magnier&Schoormans, 2015). The image displayed on the product package is another key element of its visual appearance and was found to have a significantly greater effect on influencing the evaluation of products than other elements (Chrysochou&Grunert, 2014; Ampuero& Vila, 2006). Additionally, the images displayed have a strong effect on emotional responses (Lang et. Al., 1993), unconscious responses (Larsen et al., 2003), and affect a range of consumer reactions, such as attitudes about the package and beliefs about some sensory attributes such as taste (Underwood & Klein, 2002). The verbal cues used to communicate a message or an idea to the consumer is considered a key element in the design of a food package (Machiels&Karnal, 2016; Kauppinen-Räisänen et. al., 2012). The typeface used was found to influence consumer responses and perceptions (Henderson. et. al., 2004, McCarthy and Mothersbaugh, 2002). Package color also affects consumers' willingness-to-buy (Robollar et. al 2012) as it was found capable of conveying different messages, such as price, quality, gender, and age (Klimchuk&Krasovec, 2012; Plasschaert; 1995), color was also found capable of differentiating products (Garber et. al., 2000), grab consumers' attention (Klimchuk&Krasovec, 2012), and is related to consumers' emotional responses (Gao & Xin, 2006). The effect of the interaction and the combination of these design attributes has been studied by several researchers. For example, some studies addressed the effect of the combination of color and shape (Becker et., 2011). The effect of these design attributes and their interaction may vary according to the product which the package contains and the target customer base it addresses.

The existing literature provides food product package designers with a wealth of useful information to consider when designing a new package. However, using this information relies to some extent on the designer experience

* Corresponding author e-mail: saed.salhieh@buid.ac.ae.

and prior knowledge with respect to the effect of some visual characteristics of the package. The two major approaches adopted by researchers when investigating the effect of package design on consumers can be characterized as being atomistic or holistic. The atomistic approach deals with packaging as a group of individual elements, such as color, imagery, shape, size, and typeface (Underwood, 2003). Consumers are thought to evaluate each individual packaging element separately, which in turn affects their overall response towards packaging or evaluations towards the product (Becker et al., 2011). The holistic approach views packaging as a bundle of elements that are blended into a holistic design (Orth &Malkewitz, 2008). Researchers in both approaches usually investigate a handful of attributes at a time, since increasing the number of attributes will increase the number of factors and levels per factor that needs to be analyzed which complicates conducting the required experiments and may lead to uninterpretable results. Additionally, the current literature does not include a systematic approach that can be used by food package designers to identify the most important package attributes that affect customer choice and predict customer willingness-to-buy.

This paper proposes a systematic approach that builds on existing methods and techniques to elicit the messages intended to be delivered to consumers through food package design and assesses the package design success in communicating the elicited messages. In addition, the proposed approach allows the designers to simulate the effect of varying some package design attributes on customer willingness-to-buy. In summary, the aim of this work is to provide a systematic approach for food product package designers that can be used to:

- Analyze package design attributes' effect on consumer willingness-to-buy.
- Identify actionable package design factors capable of communicating the messages intended by package designers or product owners.
- Propose new product package designs that have a high willingness-to-buy.

2. New Food Package Design Approach

A four-step approach is introduced to design new food packages. The approach (Figure 1) starts with determining the marketing messages that need to be delivered to customers, then the effectiveness of current or new package design concepts in delivering the marketing messages are assessed in step 2. Package design factors capable of embodying the intended marketing messages will be selected in step 3 and used in step 4 to design new food product packages.

Step 1: Determine Marketing Messages

Marketing messages are a set of important information which are believed to be capable of attracting customers and influencing their purchasing behavior. Determining marketing messages requires understanding the target audience's needs by conducting market studies to investigate the factors influencing customers' purchasing decisions. Market studies usually result in determining a set of product attributes used by customers to evaluate different alternative products. Product developers need to ensure that these attributes are incorporated into the product design and select an appropriate communication method to convey their existence to customers. For example, customers considering food products could be interested in the product's nutritional value, thus product developers may deliver this message by displaying the nutritional value information using text. Product developers can use different methods to conduct market studies and determine the marketing messages intended to be delivered to customers through package design (Ulrich and Eppinger, 2015). These methods include – but are not limited to-:

- Customer market surveys: customer's needs and reasons for purchase are investigated by conducting interviews, or using questionnaires.
- Expert panels: the opinions of market experts and product sellers are investigated and used to determine the main marketing messages.
- Reviewing major and relevant trade magazines: product reviews and critiques published in trade magazines can form the basis for determining the main marketing message.

The result of this step is a list of marketing messages that need to be delivered to customers through the package design.

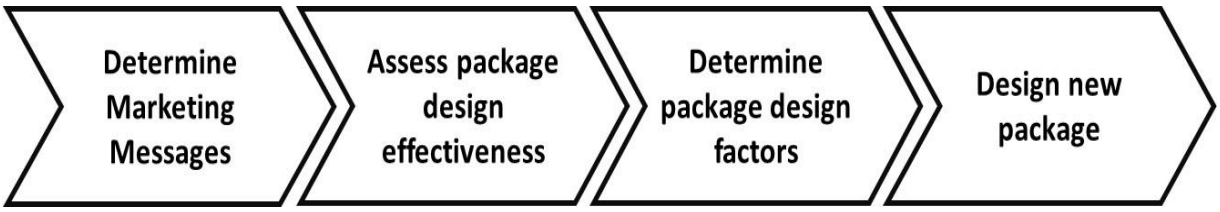


Figure 1. Food package design approach

Step 2: Assess Package Design Effectiveness in Delivering the Marketing Messages

Product developers may consider various design concepts to deliver the required marketing messages. These concepts could be based on existing product package designs or novel ideas to deliver the messages. In both cases, it is essential to assess the package design effectiveness in delivering the marketing message. It should be noted here that a product package design could communicate several marketing messages, and these messages might be well received by customers, but this does not guarantee that customers will purchase the product. Thus, it is equally important to determine the impact of these marketing messages on customers' purchasing decisions. This step will focus on determining the effectiveness of the package design in delivering the marketing messages, and the impact that these marketing messages have on customers' purchasing decisions.

Marketing messages identified in the previous step will be structured using the Semantic Differential Method (Morrow et. al. 2015), which consists of two pairs of bipolar adjectives (opposite meaning) anchoring both sides of a Likert Scale, to build a semantic space based on the subjective perceptions of customers thus facilitating the process of exploring how well the message intended is received by target customers. In addition; a preference attribute representing the willingness-to-buy will be included in the assessment. This analysis will be conducted by developing a "joint-space" attribute-based perceptual map that combines both customer perceptions and preferences (Lilien et. al., 2017).

A joint-space map (Figure 2) is a perceptual map showing the relative position of competing objects plotted in Euclidean space (Kardes 2010, Kwon et. al. 2011). Each point on the map represents one object. Two points near to each other indicate that there are similarities between the objects. The axes of the map are assumed to denote the underlying dimensions used by respondents to form perceptions and preferences for an object. The perceptual map will be diagrammed using Correspondence Analysis; a multidimensional scaling interdependence technique for dimension reduction and perceptual mapping (Malhotra et. al. 2017, Greenacre 2016, Beh& Lombardo, 2014). The input data for the analysis are in the form of a contingency table indicating a qualitative association between the rows (e.g. attributes) and columns (e.g. objects). Correspondence analysis scales the rows and columns in corresponding units so that each can be displayed graphically in the same low-dimensional space. These spatial maps provide insights into (1) similarities and differences within the rows with respect to a given column category; (2) similarities and differences within the column categories with respect to a given row category; and (3) relationship among the rows and columns. Correspondence analysis results in the grouping of categories (activities, brands, or other stimuli) found within the contingency table. The results are interpreted in terms of proximities among the rows and columns of the contingency table. Categories that are closer together are more similar in the underlying structure. Joint-space maps can be generated by two methods: averaged ideal-point and averaged vector models (Lilien et. al., 2017). In the averaged ideal-point map, a hypothetical ideal object is added to the set of alternative objects that customers are asked to rate. That is, customers are asked to indicate where their ideal object would fall in terms of the different attributes; the farther an object is from the ideal object, the less it would be preferred by customers. The averaged vector model adds a preference attribute to the set of attributes that will be rated by customers and will appear as a vector on the perceptual map indicating the direction of increasing preference. The farther an object appears along the preference vector, the more it is preferred, and other attributes that are closely aligned with the preference vector can be interpreted as drivers of explanations for customer preference.

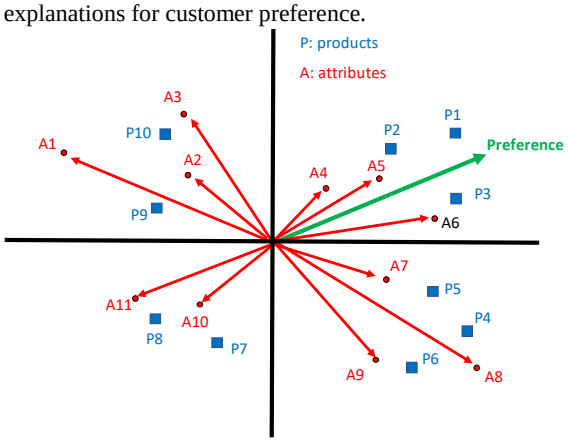


Figure 2. Joint-Space Perceptual Map

This step will generate a joint-space perceptual map showing customer perceptions with regards to the marketing messages communicated through product package designs, and the preferences that customers have with respect to each product rated. For example, Figure 2 shows a joint-space perceptual map showing customer perceptions towards 10 products with respect to 11 attributes and a preference attribute. The map can be interpreted to identify the following:

- Product similarity assessed through product proximity. For example, products P1 and P2 have close proximity indicating that they are perceived as similar products by customers, while products P1 and P7 are farther apart indicating that they are perceived as dissimilar.
- Attribute correlations assessed through the angle between the attribute vectors. Acute angles represent a positive correlation between the attributes, while right angles indicate a lack of correlation and obtuse angles indicate a negative correlation. Correlated attributes could be combined into one attribute. For example, attributes A4, A5, and A6 are correlated attributed and could be combined into one attribute if needed.
- Product performance with respect to the different attributes as indicated by the position of the product with respect to the attribute vector. The farther the product on the attribute vector, the better the perceived performance of the product. For example, product P1 has a better performance with respect to attribute A5 than product P2.

Preference analysis as represented by the preference vector indicating the direction of the increasing preference. For example, product P1 is preferred more than product P2. Additionally, the correlation between the preference vectors and other attribute vectors sheds light into what are the main attributes that affect the preference. For example, attribute A5 is highly correlated with the preference attribute, while

attribute A3 is not correlated, and attribute A1 is negatively correlated.

The main outcome of this step will be the identification of the influential marketing messages that affect customers' preference and leads to increasing the willingness-to-buy.

Step 3: Determine Package Design Factors

The goal of this step is to identify the package design characteristics that embody the marketing messages in order to use them in the design of new packages. This will be accomplished by analyzing the product package designs along with the marketing messages delivered successfully as identified in the previous step. Common characteristics among the product packages will be considered as the design embodiment features. These features will be selected such that they would be modifiable or actionable by package designers.

The set of actions or modifications applicable to each design features are considered to be the range of values that this feature can assume. That is, the design features identified will be treated as design factors with multiple levels; each level represents a possible variation or action that can be applied to the feature. Thus; this step will result in identifying the set of design factors along with their levels that need to be considered when designing a new product package that can deliver the intended marketing messages.

Step 4: Design New Product Package

The objective of this step is to determine the best product package design capable of delivering the marketing messages. The new package design will be conceptualized using conjoint analysis to assess different combinations of the design factors identified in the previous step and find the optimal combination to maximize customer willingness-to-buy. The procedure followed in this step will be as follows:

1. Develop product bundles

In conjoint analysis, product bundles represent product concepts constructed by varying the combinations of design factor levels of the product under study (Vithala, 2014). Customers' overall evaluation of the bundles will be obtained and will be decomposed into the part-worth that each customer attaches to each level of each factor. The full-profile approach, where full or complete profiles are constructed for all the factors, can be used when the number of factors/levels is small. But once the number of factors increases, it will not be possible nor will it be feasible to ask customers to evaluate all the possible combinations. Thus; the number of product bundles is reduced by using a special class of fractional design called orthogonal arrays which allows for the efficient estimation of all main effects (Lilien et.al, 2017). Orthogonal arrays permit the measurement of all main effects of interest on an uncorrelated basis. These designs assume that all interactions are negligible. Orthogonal designs can be generated automatically by computer packages such as SPSS where two sets of data are obtained: the 'estimation set' used to calculate the part-worth functions for the factor levels, and the 'holdout set' used to assess reliability and validity.

Customer evaluations of the product bundles will be gathered by presenting the customers with a description of each bundle. This description is usually based on listing the attributes and their corresponding levels on cards (i.e. text description), or by showing customers rendered images representing the product bundle along with some description of the features available. Customers will be asked to rate the likelihood of buying each product bundle on a scale of 0 to 100.

2. Determine design factors' relative importance

The part-worth model (Malhotra et. al., 2017) will be used to determine the utility score or part-worth for each factor level. The utility scores are analogous to regression coefficients and provide a quantitative measure of the preference for each factor level where larger values indicate a greater preference. Part-worths have common units and can be added together to give total utility or overall preference for any combination of factor levels. The part-worths can be used for predicting the preference of new product profiles. The goodness of fit of the estimated model can be evaluated using the value of R^2 which will indicate the extent to which the model fits the data.

3. Finalize the new product package design

One of the advantages of using conjoint analysis is its ability to predict the total utility or preference of new designs, this allows designers to simulate customer response to the improvement or design changes made to the current design. New designs will be proposed in this step by selecting a combination of factor levels that will maximize the overall utility or preference of the new product concept while taking into account the feasibility of producing the product economically. The performance of the new product concepts can also be verified by asking the customer to rate the new product concepts (if possible). If these ratings are gathered, then it can be used to double-check the model validity by comparing the results of the simulation with actual consumer responses.

3. Illustrative Example:

The use of the methodology is illustrated using an example for the design of flavored milk packages as shown in the following sections.

Step 1: Determine Marketing Messages

The marketing messages intended to be delivered to customers through the flavored milk package design were identified by a panel of experts consisting of three package designers, two product producers, and two marketing managers. The panel analysis resulted in identifying 12 messages as shown in Table 1. Each marketing message was given a short indicative label so as to be used as attributes in later stages.

Table 1. Marketing Messages

No	Attribute	Marketing Message Description
1	Natural	The package indicates that the product is made from natural ingredients.
2	Rich in Flavor	The package suggests that the product has an intense flavor.
3	Healthy	The package shows that the product contains healthy ingredients.
4	Tasty	The package indicates that the product has a satisfying and indulging taste.
5	Fresh	The package indicates that the product is made from fresh ingredients.
6	Thick	The package reflects that the product is full of ingredients and is not diluted with water.
7	Energetic	The package suggests that the product provides the energy needed for activities.
8	Fun	The package indicates that the product is enjoyable.
9	Attractive	The package makes the product stand out and look appealing.
10	Elegant	The package reflects a stylish appearance.
11	High Quality	The package reflects a high-quality product image.
12	Luxurious	The package indicates that the product reflects a sense of self- indulgence, and pleasure.

Step 2: Assess Package Design Effectiveness in Delivering the Marketing Messages

The marketing messages (attributes) and the willingness-to-buy were structured using the Semantic Differential Method and a set of 19 flavored milk packages commonly found in the local market were selected to be assessed with respect to their performance in terms of delivering the intended marketing messages.

The analysis was conducted by utilizing a survey distributed to a sample consisting of parents who had at least one child aged four years or more. The sample had one hundred and thirty-two participants (approximately 65% female and 35% male, 26 – 46 years old). The participants were shown pictures of existing product packages (see Appendix A). The pictures were edited using Adobe Photoshop® to hide the brand name. Additionally, participants were informed that all products had the same size, the same price, and the same brand. This was done to ensure that these factors will not be taken into account when making the rating. Product package pictures were shown to participants in random sequential order and the participants were given a questionnaire (Table 2) to evaluate the product packages with respect to the selected attributes including the willingness-to-buy on a scale of 1 to 9. The participants were also given the list of marketing messages as described in Table 1 without any further explanation or examples of designs in order not to bias their responses. Participants were interviewed while shopping in three different major shopping malls.

Table 2. Questionnaire

	Preference Rating									
	1	2	3	4	5	6	7	8	9	
Artificial										Natural
Lacking flavor										Rich in Flavor
Unhealthy										Healthy
Insipid										Tasty
Not fresh										Fresh
Thin										Thick
Dull										Energetic
Boring										Fun
Repugnant										Attractive
Poky										Elegant
Low Quality										High Quality
Cheap										Luxurious
Willingness to buy										Willingness to buy
	Absolutely Not								Absolutely Yes	

A perceptual map of the product packages and attributes (i.e. messages conveyed by the package) was constructed using correspondence analysis. The resulting map (Figure 3) showed the perceived similarities/dissimilarity between different product packages along with the attributes associated with these product packages. Initial examination of the map revealed that the attributes could be clustered into three groups:

- Group A contains attributes Fun, Energetic, Attractive, and Elegant.
- Group B contains Natural, Fresh, and Healthy, and
- Group C contains Luxurious, Tasty, Intense, High Quality, and Rich in Flavor.

On the other hand, the product packages formed two clusters. The first cluster was close to attributes in Group A and it had packages with a cartoon image displayed. The second cluster was associated with the attributes in Group B and had a text stating the vitamin content. The rest of the product packages were not clearly clustered into groups which indicate the need for further analysis to identify distinguishing factors. This could be due to the difficulty of analyzing a map with thirteen attributes (conveyed messages) which proved to be cumbersome and even not practical if a designer wanted to identify specific design factors to use. Factor reduction is needed to identify an appropriate number of factors (attributes) capable of retaining most of the information within the map. This was assessed by the proportion of inertia accounted for by each retained dimension or factor as shown in Table 3, were three factors accounted for 70.2% of the inertia. Next, correlated attributes are merged into a single attribute or factor. The merged attributes are named by either using the name of the attribute that has the largest distinguishing power (the farthest point from the origin) or introduce a name that could represent all the correlated attributes. The following factors names were proposed and will be used in the rest of the study:

- Attributes Fun, Energetic, Attractive, and Elegant were combined into one factor named “Childish” to indicate that it is appealing to children.

- Attributes Natural, Fresh, and Healthy are combined into the factor "Natural" indicating the importance of the healthy content which is expected to appeal to parents.
- Attributes Luxurious, Tasty, Intense, High Quality, and Rich in Flavor are combined into one factor called "Rich in Flavor" to indicate the quality and intensity of the taste.

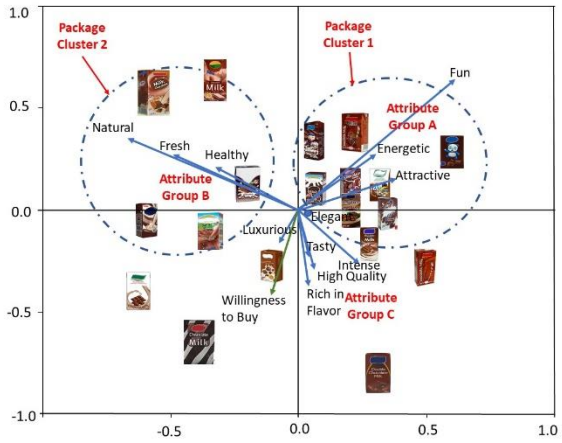


Figure 3. Perceptual map

Table 3. Inertia analysis of the perceptual map

Dimension	Singular Value	Inertia	Proportion of Inertia	
			Accounted for	Cumulative
1	0.103	0.011	0.308	0.308
2	0.091	0.008	0.237	0.545
3	0.074	0.005	0.157	0.702
4	0.054	0.003	0.086	0.787
5	0.052	0.003	0.079	0.866
6	0.040	0.002	0.046	0.912
7	0.035	0.001	0.036	0.947
8	0.026	0.001	0.020	0.967
9	0.022	0.000	0.014	0.980
10	0.020	0.000	0.011	0.992
11	0.015	0.000	0.006	0.998
12	0.008	0.000	0.002	1.000
Total		0.035	1.000	1.000

The data also showed some products (packages 2, 3, 7, 11, 13, 16) had a low willingness-to-buy (less than four), these products will be excluded from the study to reduce the burden of data gathering. The three main factors or attributes representing the marketing messages conveyed by the packages, and the willingness-to-buy attribute will be used to assess the packages once more. This was done by conducting a survey that asked the participants to rate the product packages that had a willingness-to-buy of more than four with respect to the three attributes in addition to the willingness-to-buy using the questionnaire shown in Table 4, and the resulting map was as depicted in Figure 4. The proportion of inertia (Table 5) clearly shows that using three factors accounted for all proportions of inertia.

Table 4. Revised Questionnaire

	Preference Rating									
	1	2	3	4	5	6	7	8	9	
Artificial										Natural
Adult										Childish
Lacking Flavor										Rich in Flavor
Willingness to buy										Willingness to buy
	Absolutely Not							Absolutely Yes		

The resulting map (Figure 4) shows that the products were almost separated into three distinct groups, each group revolved around one attribute. All three attributes (Childish, Natural, and Rich in Flavor) are perpendicular to each other indicating that they are uncorrelated. This highlights that the selected attributes can be used to differentiate products and that each set of products had within them features or characteristics that can successfully deliver the message intended and measured by the attribute. Product packages grouped around the attribute "Childish" had a cartoon image displayed on them, while products grouped around "Natural" had a text that displayed the vitamin content. As for products near to "Rich in Flavor"; they had an image of chocolate bars with a splash of milk. Another characteristic apparent from the map is that the only non-brown product package was separated into its own group, this indicates that brown color influenced consumer choice which is logical since many people associate chocolate with brown color. Finally, willingness-to-buy was found to be correlated with "Childish", which means that consumers would be more willing to buy flavored milk products that convey a message of "Childish".

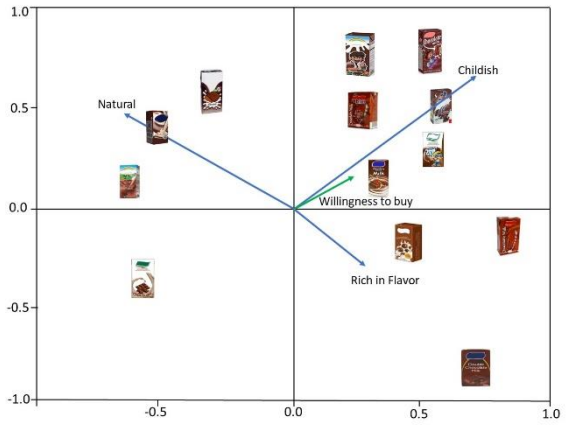


Figure 4. Revised perceptual map

Table 5. Inertia analysis for the revised perceptual map

Dimension	Singular Value	Inertia	Proportion of Inertia	
			Accounted for	Cumulative
1	0.126	0.016	0.548	0.548
2	0.098	0.010	0.333	0.881
3	0.059	0.003	0.119	1.000
Total		0.029	1.000	1.000

Step 3: Determine Package Design Factors

The joint-space perceptual map generated in the previous step was analyzed in this step to identify the package design characteristics that embodied the marketing

messages. This was done by examining the package designs and identifying the common package characteristics with respect to the attribute. Doing so revealed that:

- Packages clustered near to "*Childish*" had a cartoon image displayed,
- Packages near to "*Rich in Flavor*" had dark brown color (chocolate color), and chunks of chocolate bars image displayed.
- Packages near to "*Natural*" had a text stating the vitamin content, and the word "Healthy" or "Made with Fresh Milk" displayed on the package.

This analysis provides designers with the main design factors that they need to consider when designing the product package. That is, designers need to decide on: a) which image to display, b) which text to write, and c) which color to use. This means that there are three main design factors to consider when designing the product package, i.e Image, Text, and Color. Next, the levels associated with each factor are determined. The levels are the set of alternatives that each design factor can assume. The factor/level combinations (Table 6) were determined to be as following:

- Image: two images are considered here; a cartoon image and an image displaying chocolate bars, in addition to two more levels which include displaying both images or none.
- Text: a sentence stating "with Fresh Milk" or "Enriched with Vitamin A & D" in addition to displaying both text sentences or none.
- Color: only Brown color was considered since the package deals with chocolate-flavored milk. The levels considered are: Dominant Brown where most of the package will have a brown color, Moderate Brown where about 50% of the package will have a brown color, and Minor Brown where about 25% of the package is brown.

It should be mentioned here that all other attributes or features of the package are kept fixed, this includes the location of text or image when present on the package.

Table 6. Package Design Factors

Design Factor	Levels
Image	Cartoon, Chocolate Bars, Both, None
Text	With Fresh Milk, Enriched with Vitamin A&D, Both, None
Color	Dominant Brown, Moderate Brown, Minor Brown

Step 4: Design New Product Package

1. Develop product bundles

The full-profile approach was used in this step and a total of 48 profiles (product bundles) were needed. The number of product bundles was reduced using the orthogonal design method to 20 in order to reduce respondents' fatigue. The product bundles (Table 7) included 16 bundles used for estimation and 4 holdouts used to assess the reliability and validity of the model. Images to represent each product bundle were created using Adobe Photoshop® as shown in Appendix B.

Table 7. Product bundles for the conjoint analysis

No	Image Displayed	Text Displayed	Package Color
1	Cartoon	Fresh	Brown 100%
2	Both	Fresh	Brown 25%
3	Choc Bars	Both	Brown 100%
4	None	Vitamin Content	Brown 50%
5	Cartoon	Vitamin Content	Brown 25%
6	Cartoon	None	Brown 100%
7	Both	Both	Brown 100%
8	Both	None	Brown 50%
9	Choc Bars	None	Brown 25%
10	Choc Bars	Vitamin Content	Brown 100%
11	None	Both	Brown 25%
12	None	Fresh	Brown 100%
13	Choc Bars	Fresh	Brown 50%
14	Cartoon	Both	Brown 50%
15	None	None	Brown 100%
16	Both	Vitamin Content	Brown 100%
17[a]	Choc Bars	Fresh	Brown 100%
18[a]	Cartoon	Fresh	Brown 50%
19[a]	Cartoon	None	Brown 25%
20[a]	Both	Vitamin Content	Brown 25%

[a] holdouts

Product bundles' images were presented to ninety-seven participants (approximately 60% female and 40% male, 25 – 41 years old) in random order. Participants were asked to evaluate each product bundle by answering one question: "On a scale of 0 to 100, how likely are you to purchase this product?". Participants were informed that all products had the same brand, size, and price.

2. Determine design factors' relative importance

The part-worth utility scores and relative importance of attributes were estimated using the conjoint function in SPSS®. The validity of the model was assessed using Pearson's R and Kendall's tau which are considered acceptable measures for assessing attribute-based conjoint analysis data (Shan et. al, 2017; Jaeger et. al, 2013). Part-worth utilities of factor levels (i.e. the relative preference score computed for each factor level), and the relative importance of factors were estimated for each participant, and the mean values are summarized in Table 8. The high values observed for Pearson's R and Kendall's tau suggest that the conjoint analysis outcome fits the data well.

Table 8. Conjoint analysis results

Attribute	Attribute Level	Purchase Intention	
		Utility Estimate	Relative Importance
Image Displayed	Cartoon	4.239	63.440 %
	Chocolate Bars	-7.173	
	Both	31.050	
	None	-28.116	
Text	Fresh Milk	-5.253	24.260 %
	Vitamin Content	3.067	
	None	-10.028	
	Both	12.214	
Color	Dominant Brown	-4.823	12.300 %
	Moderate Brown	6.076	
	Low Brown	-1.252	
Constant		58.510	
Goodness of fit of conjoint analysis		Pearson's R = 0.988	
		Kendall's tau = 0.867	
		Kendall's tau for Holdouts = 1.000	
		(n = 93)	

The results show that the image displayed on the package had the highest importance, followed by the text, and lastly the color. This shows the importance that the designers need to give to these factors when designing a new flavored milk package. Additionally, the part-worth utility estimates for each attribute level can be used to assess the preference of new product designs as will be shown in the next section.

3. Finalize the new product package design

Three new package designs were proposed using the design factors as shown in Figure 5. The expected customer preference and the probability of purchase were simulated using the conjoint model using SPSS® and the results were as shown in Table 9. The results show that package A had the highest utility, followed by package B and package C. The simulation results were checked by conducting an experiment with only 20 participants. Each participant was shown images of the three designs and was asked: "On a scale of 100, how likely is it to purchase the product". The experiment results are listed in Table 9, which confirms the ability of the conjoint model to simulate customer preferences and the probability of purchase. This adds to the validity of the procedure used throughout this study.

This step shows how product package developers can simulate consumer preference towards new package designs allowing the developer to test new concepts before committing to any design.

		Package A	**Package B**	**Package C**
				
Design Factor	Image	Both	None	Cartoon
	Text	Both	None	Vitamin Content
	Color	Moderate Brown	Minor Brown	Dominant Brown

Figure 5. New product package designs

Table 9. New package designs simulation and experimental results

	Experiment	**Simulation**			
			Preference Probabilities		
	Mean Preference	**Preference**	**Maximum Utility**	**Bradley-Terry-Luce**	**Logit**
Package A	95	107.851	100.0%	57.4%	100.0%
Package B	15	19.114	0.0%	10.1%	0.0%
Package C	70	60.992	0.0%	32.4%	0.0%

4. Conclusion

Asystematic approach to design packages for food products capable of attracting customers and increasing the willingness-to-buy was presented in this paper. The approach integrates several methods and techniques in a systematic fashion to facilitate the development process. The approach is composed of four major steps. The marketing messages that need to be communicated to customers are determined by a panel of experts in the first step. Next, the effectiveness of existing package designs and/or new packages concepts in delivering the marketing messages to customers are assessed by constructing perceptual maps in the second step. The perceptual map provides a graphical depiction of customers' perceptions of the messages delivered by each package design along with an analysis of the customers' preference towards the different packages assessed. The third step of the approach addresses the identification of the package design characteristics that embodied the marketing messages. These characteristics are treated as design factors that can be varied by package designers to develop new concepts. The design factors are used in step four to develop a new package design using conjoint analysis. Conjoint analysis is also used to predict the probability of purchasing new package designs.

The approach can be used by package designers to assess the effect of package design attributes on consumers' willingness-to-buy, thus allowing the designers to focus on major attributes that have a high impact on customer purchasing decisions. Additionally, the approach systemizes the process of identifying the actionable package design factors used to communicate the influential marketing messages. Also, the developed approach allows package designers to simulate the effect of varying package design attributes on customer willingness-to-buy and to propose new package designs with a high willingness-to-buy.

The study performed on the package design of flavored milk identified that the images displayed on the package had a greater effect in communicating marketing messages than color or text. The images within the context of flavored milk packages were able to deliver a message that the product is childish and had a rich flavor. While the text was found better at communicating health-related messages such as vitamin content. The color was used to deliver the message that the milk had a chocolate flavor.

Although this paper mainly addressed food product packages, the approach can be extended to design other product packages that need to communicate marketing messages to customers through its visual appearance.

References

[1] Ampuero, O., & Vila, N. (2006). Consumer perceptions of product packaging. Journal of Consumer Marketing, 23(2), 100–112.

[2] Becker, L., van Rompay, T. J. L., Schifferstein, H. N. J., &Galetzka, M. (2011). Tough package, strong taste: The influence of packaging design on taste impressions and product evaluations. Food Quality and Preference, 22(1), 17–23

[3] Beh, Eric J. and Lombardo, Rosaria (2014). Correspondence Analysis: Theory, Practice and New Strategies. Wiley.

[4] Chrysochou, P., &Grunert, K. G. (2014). Health-related ad information and health motivation effects on product evaluations. Journal of Business Research, 67(6), 1209–1217.

[5] Gao, X. P., & Xin, J. H. (2006). Investigation of human's emotional responses on colors. Color Research & Application, 31(5), 411–417.

[6] Garber Jr., L. L., Burke, R., & Jones, J. (2000). The role of package appearance in consumer purchase consideration and choice. In, Marketing Science Institute Working Paper Series. Boston: MSI.

[7] Greenacre, Michael (2016). Correspondence Analysis in Practice. Third Edition, Chapman & Hall/CRC Interdisciplinary Statistics.

[8] Henderson, P. W., Giese, J. L., & Cote, J. A. (2004). Impression management using typeface design. Journal of Marketing, 68(4), 60–72.

[9] Jaeger, S. R., Mielby, L. H., Heymann, H., Jia, Y. L., & Frost, M. B. (2013). Analysing conjoint data with OLS and PLS regression: A case study with wine. Journal of the Science of Food and Agriculture, 93(15), 3682–3690.

[10] Kardes, Frank, 2010. Consumer Behavior. Cengage Learning.

[11] Kauppinen-Räisänen, H., Owusu, R. A., &AbeekuBamfo, B. (2012). Brand salience of OTC pharmaceuticals through package appearance. International Journal of Pharmaceutical and Healthcare Marketing, 6(3), 230–249.

[12] Klimchuk, M. R., &Krasovec, S. A. (2012). Packaging design: Successful product branding from concept to shelf. New Jersey: John Wiley & Sons.

[13] Kwong, C.K., Chen, Y. & Chan, K. Y. (2011) A methodology of integrating marketing with engineering for defining design specifications of new products, Journal of Engineering Design, Vol. 22, No. 3, March 2011, 201–213.

[14] Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology, 30(3), 261–273.

[15] Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. Psychophysiology, 40(5), 776–785.

[16] Lilien G. L., Rangaswamy A., De Bruyn A., (2017). Principles of Marketing Engineering and Analytics. DecisionPro, Inc.

[17] Machiels, C. J. A., & Karnal, N. (2016). See how tasty it is? Effects of symbolic cues on product evaluation and taste. Food Quality and Preference, 52, 195–202

[18] Magnier, L., &Schoormans, J. (2015). Consumer reactions to sustainable packaging: The interplay of visual appearance, verbal claim and environmental concern. Journal of Environmental Psychology, 44, 53–62.

[19] Malhotra, Naresh., Nunan, Dan., and Birks, David. (2017). Marketing Research: An applied approach, 5th Edition. Pearson.

[20] McCarthy, M. S., &Mothersbaugh, D. L. (2002). Effects of typographic factors in advertising-based persuasion: A general model and initial empirical tests. Psychology and Marketing, 19(7–8), 663–691.

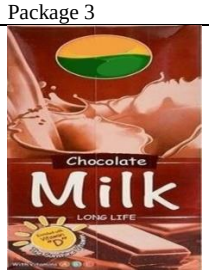
[21] Morrow J. R., Mood D. P., Disch J. G., Kang M., (2015). Measurement and Evaluation in Human Performance, Human Kinetics, Inc.

[22] Mutsikiwa, M., &Marumbwa, J. (2013). The impact of aesthetics package design elements on consumer purchase decisions: A case of locally produced dairy products. Journal of Business and Management, 8(5), 64–7.

[23] Orth, U. R., &Malkewitz, K. (2008). Holistic package design and consumer brand impressions. Journal of Marketing, 72(3), 64–81.

[24] Plasschaert, J. (1995). The meaning of colour on packaging – A methodology for qualitative research using semiotic principles and computer image manipulation. Decision Making and Research Action, 48th ESOMAR Marketing Research Congress (pp. 217–232). Amsterdam: ESOMAR.
[25] Rebollar, R., Lidón, I., Serrano, A., Martín, J., & Fernández, M. J. (2012). Influence of chewing gum packaging design on consumer expectation and willingness to buy. An analysis of functional, sensory and experience attributes. Food Quality and Preference, 24, 162–170.
[26] Shan, Liran C., De Brún, Aoife., Henchion, Maeve., Lid, Chenguang., Murrin, Celine., Wall, Patrick G., & Monahan, Frank J. (2017). Consumer evaluations of processed meat products reformulated to be healthier – A conjoint analysis study. Meat Science, 131, 82-89.
[27] Sundar, A., & Noseworthy, T. J. (2014). Place the logo high or low? Using conceptual metaphors of power in packaging design. Journal of Marketing, 78(5), 138–151.
[28] Ulrich K., Eppinger S., Product Design and Development, McGraw-Hill Education, 2015.
[29] Underwood, R. L. (2003). The communicative power of product packaging: Creating brand identity via lived and mediated experience. Journal of Marketing Theory and Practice, 11, 62–76.
[30] Underwood, R. L., & Klein, N. M. (2002). Packaging as brand communication: Effects of product pictures on consumer responses to the package and brand. Journal of Marketing Theory and Practice, 10(4), 58–68.
[31] Vithala R. R. (2014). Applied conjoint analysis. Springer.

Appendix A: Existing Product Packages

Package 1	Package 2	Package 3	Package 4	Package 5
				
Package 6	Package 7	Package 8	Package 9	Package 10
				
Package 11	Package 12	Package 13	Package 14	Package 15
				
Package 16	Package 17	Package 18	Package 19	
				

Appendix B: Product Package Design for Conjoint Analysis Study

Package 1	Package 2	Package 3	Package 4	Package 5
				
Package 6	Package 7	Package 8	Package 9	Package 10
				
Package 11	Package 12	Package 13	Package 14	Package 15
				
Package 16	Package 17	Package 18	Package 19	Package 20
				

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665
Pages 249 - 255

Jordan Journal of Mechanical and Industrial Engineering

Adaptive Backstepping Position Controller for PMSM Drive with Uncertainties of Mechanical Parameters

Salam Mahmoud, Mohammad Merei*

Power Systems engineering Department, The University of Damascus, Damascus, Syria

Received December 4 2019 *Accepted June 15 2020*

Abstract

The permanent magnet synchronous motors (PMSMs) are widely used in various industrial applications because of their numerous advantages. However, the performance of conventional controllers (PID) is insufficient in PMSM nonlinear drive systems, which requires high performance. In this paper, we investigate the possible enhancements in performance by using an adaptive backstepping position controller that is designed based on the Lyapunov stability theory. The backstepping technique has been successfully applied for nonlinear systems with external disturbances. The results of the simulation indicate an improved performance due to the designed controller ability to track the position reference signal precisely unlike the conventional controller. The proposed adaptive backstepping controller has improved performance and effectiveness of the PMSM drive systems.

Keywords: Lyapunov stability theory, adaptive backstepping position controller, the permanent magnet synchronous motor(PMSM);

1. Introduction

The permanent-magnet synchronous motor (PMSM) is extremely popular in the industrial applications due to its many attractive features, such as high efficiency nearly no rotor losses and its power density, high torque/power density and high-performance reliability. It is suitable for many applications due to advances in permanent magnetic materials, microelectronics, and modern control technologies for example machine tools, automobiles, robotics, renewable power generation, aeronautics and aerospace domains. [1,2, 3] The traditional proportional integral derivative (PID) controllers are still commonly applied in industrial control processes. Because of the controller design's simplicity and good performance in different cases of operating conditions. Although (PID) controllers can achieve good response characteristics with the linear system, it is not easy for a nonlinear system to get satisfying results. PMSMs represent the system's non-linear function of mechanical parameters and unknown disturbances, therefore the conventional (PID) controllers and linear control strategy are ineligible for designing a controller for the drive system.[4, 5, 6] Recently nonlinear systems are more applicable to the development of control theory and in control design, such as robust theory [7], robust H ∞ controller[8],

Sliding mode control [9],LQR method [10] and backstepping technique [11] have been introduced into the control system of PMSM. The adaptive Backstepping is a systematic procedure and recursive design methodology, which is often applied to design a controller for nonlinear lower triangular systems. Hence ensuring the stability of the systems. [12, 13] the adaptive backstepping control was developed depending on the Lyapunov stability theorem for position tracking control of Permanent-magnet synchronous motor (PMSM) using MATLAB/Simulink software. The Simulation results of the adaptive backstepping controller are compared with PI controller results.

2. MATHEMATICAL MODEL OF PMSM:

The PMSM is a synchronous machine with three-phase winding in stator and uses a permanent magnet in rotor separated by air gap. The PMSMs are classified into :(a) Surface Mounted magnets type Fig.1(a), (b) Inset magnets typeFig.1(b) according to the position of magnets on rotor [14, 15].

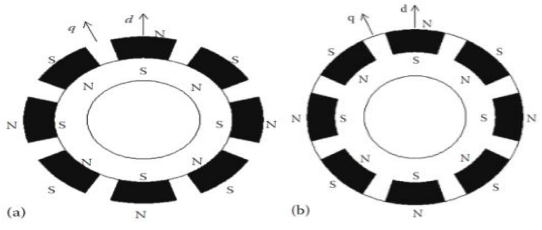


Figure 1. PMSM rotor permanent magnets layout: (a) Surface permanent magnets, (b) Inset permanent magnets

* Corresponding author e-mail: muhammed.h.merei@gmail.com.

The mathematical model of PMSM in (d, q) reference frame according to vector control can be expressed as: [3,6].

$$\frac{di_d}{dt} = \frac{-R_s}{L_d} i_d + \frac{PL_q}{L_d} \omega_r i_q + \frac{1}{L_d} u_d \quad (1)$$

$$\frac{di_q}{dt} = \frac{-R_s}{L_q} i_q - \frac{PL_d}{L_q} \omega_r i_d - P\frac{\Phi_f}{L_q} \omega_r + \frac{1}{L_q} u_q \quad (2)$$

$$\frac{d\omega_r}{dt} = \frac{3P\Phi_f}{2J} i_q + \frac{3P}{2J}(L_d - L_q) i_d i_d - \frac{B}{J}\omega_r - \frac{T_L}{J} \quad (3)$$

$$\frac{d\theta}{dt} = \omega_r \quad (4)$$

With

Rs –Stator resistance.
L_d, L_q - dq-axis inductances.
Φ_f - Permanent-magnet flux linkage.
i_d, i_q -stator currents.
vd, vq -stator voltages.
ω_r- Rotor speed.
θ - Rotor angular position.
J -the moment of inertia.
B -viscous friction coefficient.
p - Number of pole pairs.
T_L -load torque.

3. Controller Design:

The adaptive backstepping controller design is a mix of backstepping algorithm and adaptive Laws. The function of the controller is to track the rotor-desired position in engineering applications. This system can be divided into some sorts of sub-systems, with a self-controller design procedure, which contains several stages with three steps in each one. In the First step, the design is started from the first sub-system. By determining the error between the reference input signal and the actual one. The second step depends on the definition of the Lyapunovcontrol function of the first sub-system. The third step is achieved by calculating the first virtual control, which provides global stability. the first virtual control is considered as a references signal for the second sub-system , the procedures in these steps are repeated until all calculation are fulfilled for all the subsystems .Eventually we get the last virtual control signal, which is the real control input signal of the system .then by calculating the others virtual control laws in a recursive procedure in the same way, until we get the last virtual control which considers as the real control input signal of the system. [16, 17, 18, 19]

The proposed adaptive backstepping controller is described in detail systematically as follows: For height subsystem, we define the position tracking and its time derivative as

$$e_1 = \theta^* - \theta \quad (5)$$

$$\dot{e_1} = \dot{\theta}^* - \dot{\theta} = \dot{\theta}^* - \omega_r \quad (6)$$

Defining a Lyapunov function

$$V_1 = \frac{1}{2} {e_1}^2 \quad (7)$$

Differentiating (7) and substituting (5) and (6) in (8)

$$\dot{V_1} = e_1 \dot{e_1} = e_1 (\dot{\theta}^* - \omega_r) \quad (8)$$

Adding and subtracting $k_1 e_1$:

$$\dot{V_1} = -k_1 {e_1}^2 + e_1 (\dot{\theta}^* - \omega_r + k_1 e_1) \quad (9)$$

Where:
k_1 Is a positive constant.

The function$\dot{V_1}$should be negative according to Lyapunovtheory to ensure the system's stability. The desired speed for position control, which is indicated by the stabilizing function (virtual control) from (9), are represented as follows:

$$\dot{\theta}^* - \omega_r + k_1 e_1 = 0 \quad (10)$$

$${\omega_r}^* = \dot{\theta}^* + k_1 e_1 \quad (11)$$

The tracking error for the speed reference signal ${\omega_r}^*$ is:

$$e_2 = {\omega_r}^* - \omega_r \quad (12)$$

Substituting (11) in (12)

$$e_2 = \dot{\theta}^* + k_1 e_1 - \omega_r \quad (13)$$

Substituting (6) in (13)

$$e_2 = k_1 e_1 + \dot{e_1} \quad (14)$$

By using (14) the position error dynamics can be given as

$$\dot{e_1} = -k_1 e_1 + e_2 \quad (15)$$

Dynamic Speed error is expressed as

$$\dot{e_2} = \dot{{\omega_r}^*} - \dot{\omega_r} \quad (16)$$

$$\dot{e_2} = (\theta^{(2)*} + k_1 \dot{e_1}) - (\frac{3P\Phi_f}{2J} i_q + \frac{3P}{2J}(L_d - L_q) i_d i_q - \frac{B}{J}\omega_r - \frac{T_L}{J}) \quad (17)$$

Let,

$$A = \frac{3P\Phi_f}{2J}, B = \frac{3P}{2J}(L_d - L_q), \quad C = \frac{T_L}{J}, D = \frac{B}{J} \quad (18)$$

Thus, (17) can be expressed as

$$\dot{e_2} = k_1 \dot{e_1} + \theta^{(2)*} - (A i_q + B i_d i_q - C - \omega_r D)$$

$$\dot{e_2} = k_1(-k_1 e_1 + e_2) + \theta^{(2)*} - A i_q - B i_d i_q + C + \omega_r D$$

$$\dot{e_2} = -{k_1}^2 e_1 + k_1 e_2 + \theta^{(2)*} - A i_q - B i_d i_q + C + \omega_r D \quad (19)$$

Defining a new Lyapunov function

$$V_2 = \frac{1}{2} {e_1}^2 + \frac{1}{2} {e_2}^2 \quad (20)$$

Differentiating (20)

$$\dot{V_2} = e_1 \dot{e_1} + e_2 \dot{e_2} \quad (21)$$

Substituting (15) and (19) in (21)

$$\dot{V_2} = e_1(-k_1 e_1 + e_2) + e_2(-{k_1}^2 e_1 + k_1 e_2 + \theta^{(2)*} - A i_q - B i_d i_q + C + \omega_r D) \quad (22)$$

Adding and subtracting $k_2 e_2$

$$\dot{V_2} = -k_1 {e_1}^2 - k_2 {e_2}^2 + e_2[(1 - {k_1}^2) e_1 + k_1 e_2 + k_2 e_2 + \theta^{(2)*} - A i_q - B i_d i_q + C \omega_r D] \quad (23)$$

To give the Eq. (23) a negative value, the third term should be eliminated, and this is accomplished by defining

reference currents i_d^* and i_q^* for the state variables i_d and $i_q$respectively

$$\left[\left(1-{k_1}^2\right)e_1+\left(k_1+k_2\right)e_2+\ \theta^{(2)*}-Ai_q-\right.$$
$$\left.Bi_di_q+C+\omega_rD\right]=0 \quad (24)$$

$$i_d^*=0 \quad (25)$$

$$i_q^*=\frac{1}{A}\left[\left(1-{k_1}^2\right)e_1+\left(k_1+k_2\right)e_2+\ \theta^{(2)*}+C+\omega_rD\right] \quad (26)$$

Then, (23) becomes

$$\dot{V_2}=-k_1{e_1}^2-k_2{e_2}^2 \quad (27)$$

Wherek_1,k_2 are positive design constants, the Eq. (27) is given a negative value when the backstepping gains $k_1$and $k_2$have high values which ensures stability. The system parameters can change according to the variation of working conditions, which affect the controller's performance. The position and speed control performance of PMSM system is sensitive to slight deviations in the mechanical parameters because the nonlinear decoupling control is dependent on those parameters. Therefore, it is required to reflect the parametric uncertainties in the mechanical system model equations to compensate the external disturbances and achieve adaptive backstepping control. Hence the parameters C and D in Eq. (24) and (26) are considered variables $(\hat{C}.\hat{D})$ and not constant values. Therefore, Eq. (26) is presented as:

$$\hat{\imath}_q^*=\frac{1}{A}\left[\left(1-{k_1}^2\right)e_1+\left(k_1+k_2\right)e_2+\ \theta^{(2)*}+\hat{C}+\omega_r\hat{D}\right] \quad (28)$$

Since i_d and i_q are considered as manipulating variables for position control, their error functions are given by

$$e_3=\hat{\imath}_q^*-i_q \quad (29)$$

$$e_4=\hat{\imath}_d^*-i_d \quad (30)$$

Using Eq. (29) and (30), dynamicspeed error (19) can be rewritten as

$$\dot{e_2}=-{k_1}^2e_1+k_1e_2+\ \theta^{(2)*}-A\left(\hat{\imath}_q^*-e_3\right)-Bi_q(\hat{\imath}_d^*-e_4)+C+\omega_rD$$

$$\dot{e_2}=-{k_1}^2e_1+k_1e_2+\ \theta^{(2)*}-Ae_3-Bi_qe_4+C+\omega_rD-A\hat{\imath}_q^*$$

$$\dot{e_2}=-{k_1}^2e_1+k_1e_2+\ \theta^{(2)*}-Ae_3-Bi_qe_4+C+\omega_rD-A\frac{1}{A}\left[\left(1-{k_1}^2\right)e_1+\left(k_1+k_2\right)e_2+\ \theta^{(2)*}+\hat{C}+\omega_r\hat{D}\right] \quad (31)$$

$$\dot{e_2}=-e_1-k_2e_2+Ae_3+Bi_qe_4-\tilde{C}-\tilde{D}\omega_r \quad (32)$$

Where:

$\tilde{C}=\hat{C}-C$

$\tilde{D}=\hat{D}-D$

Thus, dynamic current error is given

$$\dot{e_3}=\dot{\hat{\imath}_q}^*-\dot{\imath_q} \quad (33)$$

$$\dot{e_3}=\frac{1}{A}\left[\left(1-{k_1}^2\right)\dot{e_1}+\left(k_1+k_2\right)\dot{e_2}+\ \theta^{(3)*}+\dot{\hat{C}}+\dot{\hat{D}}\omega_r+\hat{D}\dot{\omega_r}\right]-\dot{\imath_q}$$

$$\dot{e_3}=\frac{1}{A}\left[\left(1-{k_1}^2\right)(-k_1e_1+e_2)+(\ k_1+k_2)(-e_1-k_2e_2+Ae_3+Bi_qe_4-\tilde{C}-\tilde{D}\omega_r)+\ \theta^{(3)*}+\dot{\hat{C}}+\dot{\hat{D}}\omega_r+\hat{D}(Ai_q+Bi_di_q-\hat{C}-\hat{D}\omega_r)\right]+\frac{R_s}{L_q}i_q+\frac{PL_d}{L_q}\omega_ri_d+P\frac{\Phi_f}{L_q}-\frac{1}{L_q}u_q \quad (34)$$

$$\dot{e}_4=\dot{\hat{\imath}_d}^*-\dot{\imath_d} \quad (35)$$

$$\dot{e}_4=0-\left[-\frac{R_s}{L_d}i_d+\frac{PL_q}{L_d}\omega_ri_q+\frac{1}{L_d}u_d\right] \quad (36)$$

$$=\frac{R_s}{L_d}i_d-\frac{PL_q}{L_d}\omega_ri_q-\frac{1}{L_d}u_d \quad (37)$$

The final Lyapunov function is given by

$$V_3=\frac{1}{2}\left({e_1}^2+{e_2}^2+{e_3}^2+{e_4}^2+\frac{1}{\gamma_1}\tilde{C}^2+\frac{1}{\gamma_2}\tilde{D}^2\right) \quad (38)$$

Differentiating (38)

$$\dot{V_3}=\left(e_1\dot{e_1}+e_2\dot{e_2}+e_3\dot{e_3}+e_4\dot{e_4}+\frac{1}{\gamma_1}\tilde{C}\dot{\tilde{C}}+\frac{1}{\gamma_2}\tilde{D}\dot{\tilde{D}}\right) \quad (39)$$

Where:

$\gamma1, \gamma2$ are positive adaptation gains

Substituting alldynamic error in (39)

$$\dot{V_3}=\left(e_1(-k_1e_1+e_2)+e_2(-e_1-k_2e_2+Ae_3+Bi_qe_4-\tilde{C}-\tilde{D}\omega_r)+e_3\left\{\frac{1}{A}\left[\left(1-{k_1}^2\right)(\ k_1+k_2)(-e_1-k_2e_2+Ae_3+Bi_qe_4)-(\ k_1+k_2)\left(\tilde{C}+\tilde{D}\omega_r\right)+\dot{\hat{C}}+\dot{\hat{D}}\omega_r+\hat{D}Ai_q+\hat{D}Bi_di_q-\hat{D}\hat{C}-\hat{D}^2\omega_r\right]+\frac{R_s}{L_q}i_q+\frac{PL_d}{L_q}\omega_ri_d+P\frac{\Phi_f}{L_q}-\frac{1}{L_q}u_q\right\}+e_4\left\{\frac{R_s}{L_d}i_d-\frac{PL_q}{L_d}\omega_ri_q-\frac{1}{L_d}u_d\right\}+\frac{1}{\gamma_1}\tilde{C}\dot{\tilde{C}}+\frac{1}{\gamma_2}\tilde{D}\dot{\tilde{D}}\right) \quad (40)$$

$$\dot{V}_3 = -k_1 {e_1}^2 - k_2 {e_2}^2 - k_3 {e_3}^2 - k_4 {e_4}^2 + A e_2 e_3 + B e_2 e_4 i_q + e_3 \left\{ \frac{1}{A} \left[\left(1 - {k_1}^2\right)(k_1 + k_2)(-e_1 - k_2 e_2 + A e_3 + B i_q e_4) - (k_1 + k_2)\left(\tilde{C} + \tilde{D}\omega_r\right) + \dot{\hat{C}} + \dot{\hat{D}}\omega_r + \hat{D} A i_q + \hat{D} B i_d i_q - \hat{D}\hat{C} - \hat{D}^2 \omega_r \right] + \frac{R_s}{L_q} i_q + \frac{P L_d}{L_q} \omega_r i_d + P \frac{\Phi_f}{L_q} - \frac{1}{L_q} u_q + k_3 e_3 \right\} + e_4 \left\{ \frac{R_s}{L_d} i_d - \frac{P L_q}{L_d} \omega_r i_q - \frac{1}{L_d} u_d + k_4 e_4 \right\} - \left(\tilde{C} + \tilde{D}\omega_r\right) e_2 - e_3 \frac{1}{A}(k_1 + k_2)\left(\tilde{C} + \tilde{D}\omega_r\right) + \frac{1}{\gamma_1} \tilde{C}\dot{\tilde{C}} + \frac{1}{\gamma_2} \tilde{D}\dot{\tilde{D}} \quad (41)$$

Rearranging (41):

$$\dot{V}_3 = -k_1 {e_1}^2 - k_2 {e_2}^2 - k_3 {e_3}^2 - k_4 {e_4}^2 - A e_2 e_3 + \tilde{C}\left\{ \frac{1}{\gamma_1} \dot{\hat{C}} - e_2 - \frac{1}{A}(k_1 + k_2) e_3 \right\} + \tilde{D}\left\{ \frac{1}{\gamma_2} \dot{\hat{D}} - e_2 \omega_r - \frac{1}{A}(k_1 + k_2) e_3 \omega_r \right\} + e_3 \left\{ \frac{1}{A} \left[\left(1 - {k_1}^2\right)(-k_1 e_1 + e_2) + (k_1 + k_2)(-e_1 - k_2 e_2 + A e_3) + \theta^{(3)*} + \dot{\hat{C}} + \dot{\hat{D}}\omega_r + \hat{D} A i_q + \hat{D} B i_d i_q - \hat{D}\hat{C} - \hat{D}^2 \omega_r \right] + \frac{R_s}{L_q} i_q + \frac{P L_d}{L_q} \omega_r i_d + P \frac{\Phi_f}{L_q} - \frac{1}{L_q} u_q + k_3 e_3 \right\} + e_4 \left\{ B e_2 i_q + B \frac{i_q e_3}{A}(k_1 + k_2) + \frac{R_s}{L_d} i_d - \frac{P L_q}{L_d} \omega_r i_q - \frac{1}{L_d} u_d + k_4 e_4 \right\} \quad (42)$$

The system's stability is accomplished when $\dot{V}_3 \to 0$ this condition is ensured as the positive terms of Eq. (42) decreasing to zero. Hence, the parameter adaptation laws are given by $\dot{\hat{C}} = \gamma_1 \left(e_2 + \frac{1}{A}(k_1 + k_2) e_3 \right)$ (43)

$$\dot{\hat{D}} = \gamma_2 \left(e_2 \omega_r + \frac{1}{A}(k_1 + k_2) e_3 \omega_r \right) \quad (44)$$

Control voltages are defined as

$$v_q{}^* = L_q \left\{ k_3 e_3 + \frac{1}{A} \left[\left(1 - {k_1}^2\right)(-k_1 e_1 + e_2) + (k_1 + k_2)(-e_1 - k_2 e_2 + A e_3) + \theta^{(3)*} + \dot{\hat{C}} + \dot{\hat{D}}\omega_r + \hat{D} A i_q + \hat{D} B i_d i_q - \hat{D}\hat{C} - \hat{D}^2 \omega_r \right] + \frac{R_s}{L_q} i_q + \frac{P L_d}{L_q} \omega_r i_d + P \frac{\Phi_f}{L_q} \omega_r \right\} \quad (45)$$

$$v_d{}^* = L_d \left\{ k_4 e_4 + B i_q \left[e_2 + \frac{e_3}{A}(k_1 + k_2) \right] + + \frac{R_s}{L_d} i_d - \frac{P L_q}{L_d} \omega_r i_q \right\} \quad (46)$$

Substituting (43)-(46) into (42)

$$\dot{V}_3 = -k_1 {e_1}^2 - k_2 {e_2}^2 - k_3 {e_3}^2 - k_4 {e_4}^2 + A e_2 e_3 \quad (47)$$

The system is considered stable when the backstepping gains k_2 and k_3 have high values and then $\dot{V}_3$ is a negative definite.

4. EXPERIMENTAL RESULTS:

The structure diagram of adaptive backstepping control of PMSM is shown in Fig (2). The rotor position and the three-phase currents are measured by sensors. That are used as feedback signals for the controller. The position error signal is acquired by comparing the actual position signal with the reference Position signal. The adaptive backstepping control uses position error signal, current and voltages signals and parameter adaptation laws (43),(44) to determine the control voltages which present the actual control inputs. However these signals are in (d-q) reference frame, thus they must be transformed into the three-reference frame to give the inverter voltages, which supply and drive the PMSM motor.

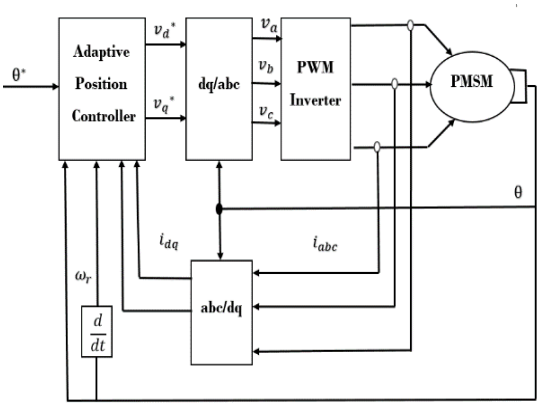


Figure 2. Block diagram of the Adaptive backstepping controller for position Control of PMSM Drive.

The nominal parameters of PMSM motor that using in the simulation in this paper are given in the **Table 1**

The adaptive backstepping control gains are selected as k1 =200, k2 =60, k3=95, k4 =950, γ_1=0.02, γ_2=0.005.

Table 1.PMSM PARAMETERS [20]

PMSM parameters	Nominal values (unit)
$\boldsymbol{StatorresistanceR_s}$	0.02 (Ω)
d − axis inductance $\boldsymbol{L_d}$	$1.25(mH)$
q − axis inductance $\boldsymbol{L_q}$	$1.25(mH)$
$\boldsymbol{polepairsP}$	4
$\boldsymbol{MotorinertiaJ}$	$0.089(Kg.m^2)$
$\boldsymbol{FrictioncoefficientB}$	$0.005(N.m.\frac{sec}{rad})$
$\boldsymbol{Ratedpower}$	40 KW
$\boldsymbol{Ratedvoltage}$	$380(V)$
$\boldsymbol{Ratedcurrent}$	70 (A)
Magnetic flux constant $\boldsymbol{\Phi_f}$	$0.381(V.\frac{sec}{rad})$

We studied the performance of the proposed control, in these two cases:

Case 1: A sine-wave trajectory is taken as reference position signal θ^*=10sin (2t) rad and the load torque of $T_l = 10\ N.M$ is suddenly applied at $t = 1.5\ s$, the simulation results are given in fig (3-5). Fig (3) shows the curve of position reference and actual position response, zooming in fig (4), the blue line represents the position reference signal and the red line represents actual position signal. It can be observed from the figures that the adaptive backstepping controller has a Good tracking effect, as it shows the actual position signal tracks the reference position signal closely. It can be noted from fig.5 that the speed returns to its reference value after a short period of slight deviation whenever a load is added.

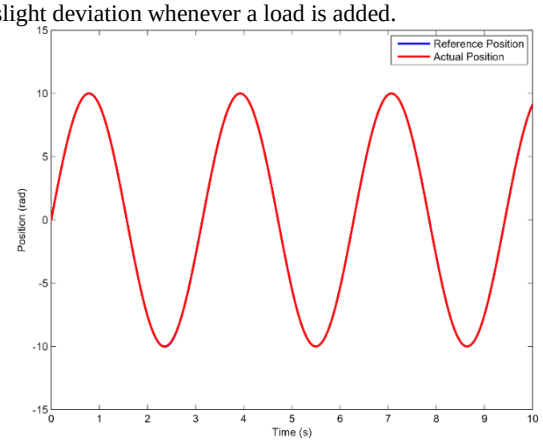


Figure 3. Position Tracking Curve.

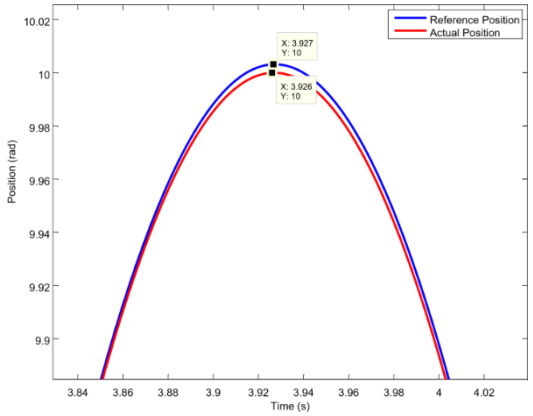


Figure 4. Zoom in on Position Tracking Curve

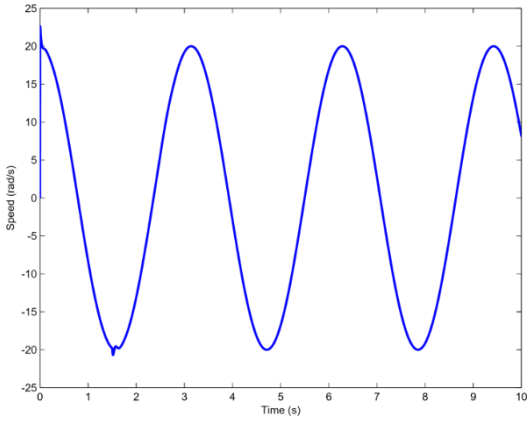


Figure 5. Speed Curve.

Case 2: A triangular -wave trajectory is taken as reference position signal and the load torque of $T_l = 10\ N.M$ is suddenly applied at $t = 1.5\ s$, the simulation results are given in Fig 7(a)-(e).and compared with the simulation results of the convention control (PI) for PMSM drive Fig. 6(a)-(e). The gain values of (PI) controller, which is used for the simulation, are selected as, Speed controller Kp =0.6835 and Ki=0.27 and Current controllers Kp=0.25 and Ki=0.8, position controller K=15.Fig. (a) Shows the comparative simulation results between backstepping controller and conventional controller with triangular position reference. The measured position tracks the reference signal closely by using adaptive backstepping control with driving system. While there is a deviation in the reference position tracking using conventional controller (PI) in driving system, as shown.Figure (b) shows the speed curves of both the driving systems with the same triangular -position reference signal. According to the figure, the proposed controller has high fast response with small stability error at marginal changes for (1 ms) compared with the conventional controller. Figure (c) shows the q-axis current component of both the driving systems with the triangular -position reference signal. According to the figure, the fast transient state occurs, then the q-axis current component returns to a value that achieves the demanded load torque. In contrast, the driving system with conventional controller has large fluctuations in the transient state period, which can cause damage in the driving system. According to vector control technique, the d-axis current component (id) is regulated quite to zero. Both conventional and backstepping controllers could achieve the main control condition, as show in figure (d).According to figures, the response curves of the adaptive backstepping controller have a smaller steady-state error than the PI controller. The proposed controller is featured by fast response and has a shorter settling time than the PI controller has. By selecting positive values for the backstepping gains, the stability of the system is achieved depending on the Lyapunov Stability Theory which is not achieved by selecting the parameter of the PI controller. As a result, the proposed controller has a better performance than the traditional controller does.

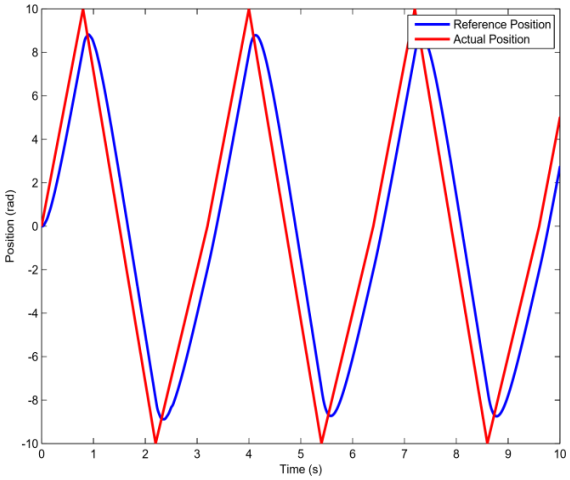


Figure 6 (a). Position Tracking Curve under control of PI controller.

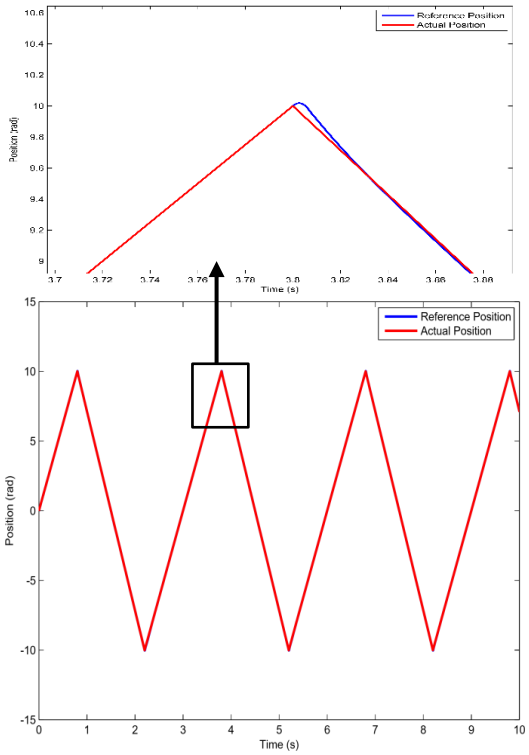


Figure 7(a). Position Tracking Curve under control of Adaptive Back stepping controller.

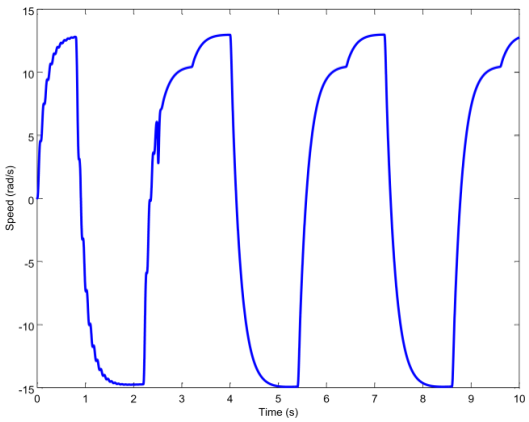


Figure6(b). Speed curve under controlof PI controller

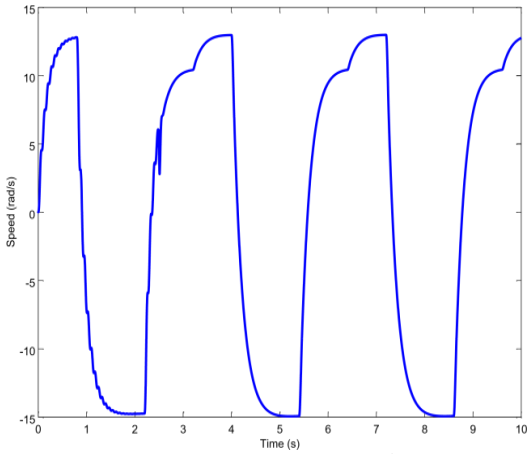


Figure 6 (b) . Speed curve under control of PI controller

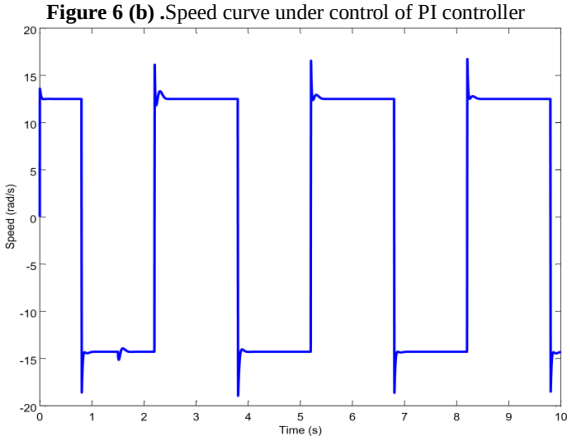


Figure 7(b). Speed curve under control of Adaptive Back stepping controller.

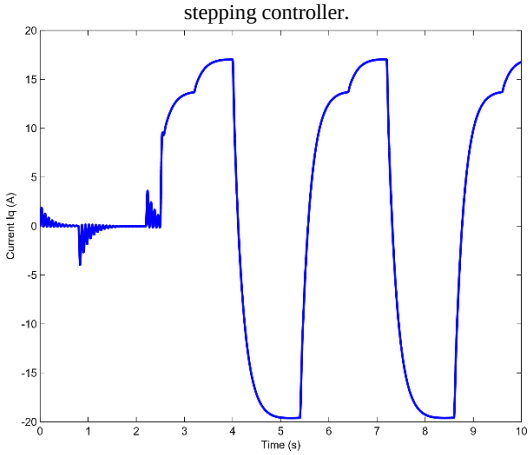


Figure6(c). Q-axis current response curve under control of PI controller.

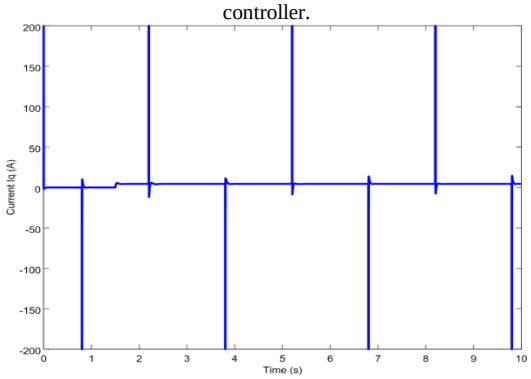


Figure 7(c). Q-axis current response curve under control of Adaptive Back stepping controller.

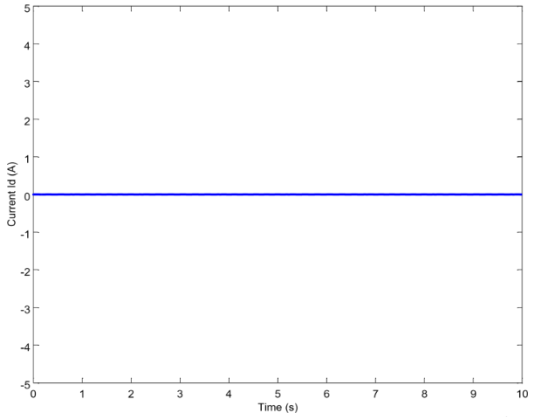


Figure6(c). D-axis current response curve under control of PI controller.

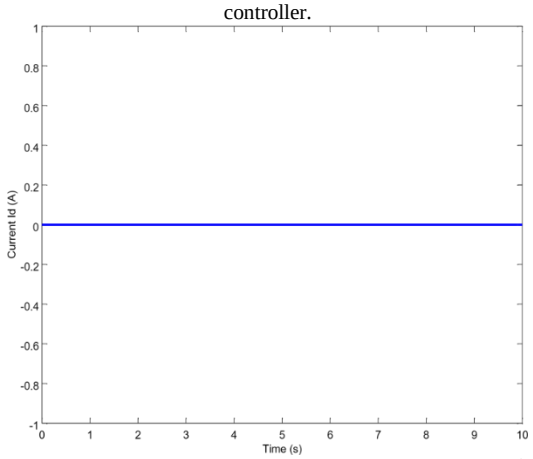


Figure 7(c). D-axis current response curve under control of Adaptive Back stepping controller.

5. Conclusion:

In this paper, an adaptive backstepping controller has been proposed for the control of a PMSM drive system. This controller can overcome the performance problems of the drive system with the traditional controller (PID). In addition, it estimates the varying mechanical system parameters, such as load torque and inertia. The simulation results show that using the proposed controller can ensure the rotor position tracking the reference curves precisely, and achieved fast response.

References

[1] Y. Yang, X. Guo, and Z. Mi, "Adaptive Robust Backstepping Control of Permanent Magnet Synchronous Motor Chaotic System with Fully Unknown Parameters and External Disturbances," Mathematical Problems in Engineering, vol. 2016, Article ID 3690240, 14 pages, 2016.

[2]Y. H. Lan, and L. Zhou, "Backstepping Control with Disturbance Observer for Permanent Magnet Synchronous Motor," Journal of Control Science and Engineering, vol. 2018, Article ID 4938389, 8 pages, 2018.

[3] C. Liu, W. Zhang, "Adaptive Backstepping Speed Control for IPMSM With Uncertain Parameters". International Conference on Information and Automation (ICIA), Macau SAR, China, 2017.

[4] J. Chen, W. Yao, Y. Ren, R. Wang, L. Zhang, and L. Jiang, "Nonlinear adaptive speed control of a permanent magnet synchronous motor: A perturbation estimation approach, " Control Engineering Practice , 85 (2019) 163–175.

[5] M. Karabacak, H. I. Eskikurt, "Speed and current regulation of a permanent magnet synchronous motor via nonlinear and adaptive backstepping control," Mathematical and Computer Modelling , 53 (2011) 2015–2030.

[6] M. Yang, X. Wang, K. Zheng, "Adaptive Backstepping Controller Design for Permanent Magnet Synchronous Motor". Proceedings of the 8th World Congress on Intelligent Control and Automation, Jinan, China, 2010.

[7]L. Sheng, G. Xiaojie, W. Yuchao, "Robust Nonlinear Control for Six-phase Permanent Magnet
Synchronous Motor with Intelligent Uncertainty Observer". Proceedings of the 36th Chinese Control Conference, Dalian, China, 2017.

[8] X. Cao, X. Li, "Backstepping-based Robust H ∞ Tracking Controller Design for a Class of Nonlinear Systems". Proceedings of the 37th Chinese Control Conference, Wuhan, China, 2018.

[9] Q. Jihong , L. Zihao, W. Hongyan, "Adaptive Sliding Mode Control with Nonlinear Disturbance Observer for Uncertain Nonlinear System Based on Backstepping Method". Proceedings of the 10th
World Congress on Intelligent Control and Automation, Beijing, China, 2012.

[10] M. Ababneh,"Controlling of Chaos Synchronization,"Jordan Journal of Mechanical and Industrial Engineering, Vol. 9, No. 2, 2015, 75-84.

[11]Z. Yin, B. Wang, C. Du, "Barrier-Lyapunov-Function-Based BacksteppingControl for PMSM Servo System with Full State Constraints". International Conference on Electrical Machines and Systems (ICEMS), China, 2019.

[12] S. J. Rai, "Control of a Permanent Magnet
Synchronous Motor (PMSM) with constraints,"Mscdiss,University ofNorwegian, 2017.

[13] H. Pang, X. Zhang, J. Chen, K. Liu, "
Design of a coordinated adaptive backstepping tracking control for nonlinear uncertain active suspension system," Applied Mathematical Modelling, 76 (2019), 479–494.

[14]Krishnan R. Permanent Magnet Synchronous and Brushless DC Motor Drives. Virginia: Taylor & Francis Group; 2010.

[15]JACEK F G.PERMANET MAGNET MOTOR TECHNOLOGY DESIGN AND APPLICATION.3rd ed. New York: Taylor & Francis Group; 2010.

[16]]L. SONNEVELDT,"Adaptive Backstepping
Flight Control for Modern Fighter Aircraft," PhD diss., University of Delft, 2010.

[17]O. Härkegård, "Flight Control Design Using Backstepping,"Mscdiss,University of Linköpings, 2001.

[18] Krstić M, Ioannis k, and Kokotović P. NONLINEAR AND ADAPTIVE CONTROL DESIGN. The United Stat of America: New York; 1995.

[19] KHALIL, H K. Nonlinear Systems. 3rd ed. New Jersey: PRENTICE HALL; 2002.

[20]] M.Yang, X. Wang, and K. Zheng,"Nonlinear Controller design for Permanent Magnet Synchronous Motor Using Adaptive Weighted PSO".American Control Conference Marriott Waterfront, Baltimore, MD, USA, 2010.

JJMIE

Volume 14, Number 2, June. 2020
ISSN 1995-6665
Pages 257 - 270

Jordan Journal of Mechanical and Industrial Engineering

Design of a Vendor Managed Inventory Model for Impulse Purchase Products in a Two-level Supply Chain

David García*, Daniel Palencia, Cristian Solano, Adel Mendoza

Department of Industrial Engineering, Universidad del Atlántico, Puerto Colombia, Colombia

Received April 11 2020 *Accepted June 7 2020*

Abstract

Although there are multiple methodologies to carry out collaborative practices of inventory management, none are set up for impulse purchase products. This is a disadvantage because with the opening of new markets and the proliferation of consumer culture, the economic importance of buying products on impulse always remains relevant. In this paper, a Vendor Managed Inventory model was designed based on the direct participation of a vendor and a buyer (two-level supply chain), in order to agree on the procurement operations of a portfolio of impulse purchase products. For this proposal, a mathematical model based on classical optimization was designed to minimize inventory costs. Subsequently, a case study was conducted comparing the economic impact of the model with respect to a traditional supply agreement in a non-cooperative scenario. The results reflected positive economic effects in the implementation of the model related to the economies of scale to exploit fixed costs present in the agreement. Additionally, the conditions under which the implementation of this model grants individual and global benefits to the participating companies were validated.

Keywords: Vendor Managed Inventory, Supply chain management, Classical optimization, Impulse purchase products, Logistics;

1. Introduction

Proactive management and timely distribution in an organization's supply chain usually translate into greater savings and other benefits in its operational processes. Various inventory management and order preparation and delivery methodologies - e.g., JIT (Just-in-Time), ECR (Efficient Customer Response), or VMI (Vendor Managed Inventory)- have been proven to increase supply chain competitiveness through cost reduction. Foremost among these practices, the VMI often translates into a win-win situation for both parties: buyers save on storage costs by not having to allocate labor and space in managing overstocked inventories, and vendors save on distribution costs by coordinating shipments to different buyers [1], [2]. Therefore, once an integrated planning and inventory collaboration is established, buyers using a VMI supply chain are able to work together efficiently with their suppliers to optimize inventory replenishment [3]–[5].

The Council of Supply Chain Management Professionals (CSCMP) defines VMI as the practice of companies making suppliers responsible for determining order size and timing, usually based on receipt of retail point of sale (POS) and inventory data from their customer [6]. The evolution of VMI research has been directed towards an interdisciplinary environment where not only are the impacts on inventory policies quantified but also specialized models are designed for different types of products or business sectors, along with the possibility of including risk restrictions or preferences for each of the members of the agreement. These designs have used a wide range of technical tools. Consequently, the selection of an appropriate approach depends on the objectives set in a research project and the availability of necessary data and resources. The background is explored further in the literature review section.

Related to the above paragraph, although there are multiple methodologies to carry out this practice, none are set up for impulse purchase products. This is a disadvantage because with the opening of new markets and the proliferation of consumer culture, the economic importance of buying products on impulse always remains relevant [7]–[9]. This kind of merchandise can be defined as those products that a consumer acquires suddenly and immediately without a plan prior to purchase [10]. Impulse buying behavior has been described as a novelty or escape purchase that breaks the normal buying pattern[11]. Generally, these items are strategically displayed in hot spots (areas with a large circulation of people), such as near checkouts in retail stores. Along these lines, the previous research that has been carried out does not take into account the particularities of these products, leading to arbitrary or generalized models that are used for the management of the collaborative inventory of these goods.

* Corresponding author e-mail: davidagarcia@mail.uniatlantico.edu.co.

Therefore, the purpose of this project was to design a VMI inventory model for a two-level supply chain, which would represent an adequate logistics operations scheme for a portfolio of impulse purchase products between a vendor and a buyer. The model design was based on a three-stage methodology: involving gathering information, formulation, and validation of the model. For this purpose, a product characterization was carried out, collecting data, and diagnosing the key elements that this model should have. Subsequently, it was mathematically formulated using classical or unconstrained optimization; the choice of this technique among different alternatives was supported by an expert judgment. At last, in the validation stage, which is defined as a test stage where the model is executed to evaluate its response with respect to a real scenario; A case study was carried out with data from a retail company (buyer) and a sugar confectionery vendor, in order to evaluate the results for eight impulse purchase products in a period of five months.

In this way, this research contributes to the literature on supply chain management of impulse purchasing products and designs a VMI model for this purpose. This article is organized as follows: The related literature is reviewed in Section 2; Section 3 describes the methodology used in this model; Section 4 explains each of the stages of model design; Section 5 concludes this paper and give some recommendations for future works.

2. Literature review

From the review of past and recent literature, there were no studies related to VMI models focused on the management of impulse purchase product inventories. However, as for the study of variables related to the trend of impulse buying. Darrat et al. [12] and Badgaiyan et al.[7] studied impulsive buying tendency and validate it by examining its association with other relevant variables. Also, studies related to the role of the store in consumer decisions and psychology were consulted. Flamand et al.[13] studied the optimization of store-wide shelf-space allocation in order to maximize the visibility of products to consumers; Bellini et al.[10] explored the determinants of impulse buying in a context of more planning and preparation for shopping. In the same manner, Wei et al.[14] studied the relationship between flow experience, perceived transaction value, positive effect, shopping motivation, and impulse buying behavior. There are also authors, such as Leong et al.[15], Chen et al.[16], and Sundström et al.[17], who explored how online shopping affects impulsive buying behavior.

As for the design of VMI models, one of the most implemented techniques corresponds to mathematical modeling based on classical optimization. Lee et al.[18]examined VMI systems with stock out-cost sharing between a supplier and a customer using an EOQ model with shortages allowed under limited storage capacity. Additionally, Cai et al.[19] designed a two-echelon supply chain that markets two substitutable brands of a product with uncertain demand. Lee and Cho [3], examined (z, Z)-type contracts for VMI. A (z, Z) VMI contract stipulates minimum and maximum inventory levels and their corresponding under- and over-stocking penalties. Bai et al.[20] formulated an optimization model for the centralized system, in order to investigate the effects of carbon emission reduction on a supply chain with one manufacturer and two competing retailers for deteriorating products under VMI. Other authors proposed to integrate an optimization in transport costs or distribution routes. Rahim and Aghezzaf[21] optimized the inventory holding costs and the transportation costs for a two-stage supply chain. Similarly, Mateen and Chatterjee [22] developed analytical models for various approaches through which a single vendor-multiple retailer system may be coordinated through VMI. They also highlight the savings that can be derived in the transportation cost in a VMI setting. Stellingwerf et al.[1] quantified both the economic and environmental benefits of implementing cooperation via Joint Route Planning (JRP) and VMI, optimizing routing and inventory planning decisions simultaneously. Additionally, Saif-Eddine et al.[23] formulated a mathematical model to minimize the total supply chain cost considering the Inventory Location Routing Problem (ILRP) while adopting the VMI strategy.

Furthermore, it should be mentioned that linear and non-linear programming have played an important role in generating new research proposals for VMI in recent years. Park et al.[24] constructed a mixed-integer linear programming model for the vendor-managed inventory routing problem with lost sales, while maximizing the supply chain profit over a planning horizon. Other authors developed models based on non-linear programming (NLP). Hariga et al.[25] formulated a mixed integer nonlinear program that minimizes total supply chain costs and allows unequal shipment frequencies to the retailers. They considered a supply chain where a vendor manages its multiple retailers' stocks under a VMI contract. Diabat [26] addressed the issue of VMI by considering a two-echelon single vendor/multiple buyer supply chain network. The model finds the optimal sales quantity by maximizing profit, given as a nonlinear and non-convex objective function. In addition, Verma and Chatterjee [27] developed a nonlinear mixed-integer programming model to compute the optimal replenishment frequency and quantity for each of the retailer, such that the total system cost is minimized.

It is worth mentioning other approaches that may be considered. Sadeghi and Niaki[28] designed a bi-objective VMI model with a single vendor and multiple retailers, in which the demand is fuzzy, and the vendor manages the retailers' inventory in a central warehouse. Akbari Kaasgari et al.[29] formulated a VMI supply chain for perishable products by considering discount. Then, a genetic algorithm and a particle swarm optimization algorithm are developed for solving it. In the same way, Chen[30] considered a new decision issue for perishable products in production inventory with pricing and promotion for a single-vendor multi-buyer system comprising one manufacturer and multiple retailers. He developed a centralized decision model with VMI control system under a just-in-time shipment policy. Also, Filho et al.[31] presented a case study supported in the development of a system VMI attached to the philosophy of Customer Relationship Management whose goal was to map the buying behavior of customers who purchase low-volume products.

Finally, it is essential to mention the studies in the field of Game Theory. These add enormous value to the design of VMI models. Torres et al. [32] studied the evolution of

individual strategies of the producer and the buyer by a formalism derived from the theory of evolutionary games. Tsao et al.[33] developed a multi-player retailer Stackelberg game to model the interaction between retailer and manufacturers. In this model, a retailer maximizes profit by taking the manufacturers' trade allowance response into account.Yang et al.[34] formulated joint configuration of a product catalogue and its supply chain as a leader-follower Stackelberg game that is enacted through a bi-level hierarchical optimization mechanism to model the coordination. Nishi and Yoshida [35] addressed the optimization of multi-period bilevel supply chains under demand uncertainty. The decentralized supply chain planning problem was modelled as a multi-period non-cooperative game. In a recent work, Chen and Xiao [36]developed game models for a two-echelon supply chain with one supplier and multiple competing retailers. They studied the pricing decision and the replenishment policy for each member.

3. Methodology

Three key stages were considered for the design of the model: information gathering, model formulation, and model validation. During the information gathering stage, the main logistic needs or requirements that companies have in relation to the inclusion of impulse purchase products within their VMI models were diagnosed. In addition, the nature of these products, the type of companies that commercialize them, and the ideal characteristics that the designed model should have were investigated. This first stage has, as its objective, analyzed the information that will determine the basic characteristics of the model and the industry to which it will be directed., based on current market trends. This stage included activities, such as the product characterization (Section 4.1) and the needs assessment (Section 4.2).

Subsequently, at the formulation stage, the model was designed using a technique that would allow for a better representation of the logistical operations of impulse purchase products. Relevant variables and assumptions were, therefore, established. This stage included activities, such as selecting the modeling technique (Section 4.3) and the mathematical formulation (section 4.4). Finally, the model was subjected to a validation stage. The objective of this stage was to evaluate the economic impact that would arise from its implementation. For this purpose, a real case study was carried out with the real data of a retail company and a supplier of sugar confectionery, to evaluate the results of the model for eight impulse purchase products in a period of five months. This stage included model validation (section 4.5). There is a more detailed explanation for each stage in the next section.

4. Model development

4.1. Product characterization

The Council of Supply Chain Management Professionals (CSCMP) defines product characterization as all of the elements that define a product's character, such as size, shape, weight, etc.[6] Therefore, a product characterization process can be defined as the establishment of the attributes for a given product. In this first design stage, impulse purchase products were characterized after various interactions with primary and secondary information sources. It is important to mention that two big retail companies supplied data from their product catalogs. Impulse buying products were characterized as products that cost little money, are quickly consumed, and require little time for purchasing decisions (i.e. chocolate bars, cookies, razors, chewing gum, candy, etc.). In addition, they are usually strategically located in the hot spots throughout a store [37], [38].

In terms of commercial and logistical characteristics, it was found that the main distribution channels for these products reproducers and wholesale distributors [39]. Given that the consumer sector generally requires greater intermediation to diversify the market. With regard to the marketing channel, the selection of the store is aimed at retailers that sell products directly to the customers, such as department stores, supermarkets, convenience stores, etc. Thus, although these products along the supply chain can be classified as retail or non-retail trade item, only those units that pass through the point of sale and are purchased by the final consumer can be denoted as impulse purchase products [40]. Consequently, the scope of this investigation takes into account those companies whose target market is represented by consumers who purchase these products through a retail channel. It is imperative to mention that other characteristics, such as types of packaging, transport, and storage management were also consulted.

Furthermore, companies generally resort to reducing dependence on forecasts and require increasing the frequency of delivery to reduce inventory shortage. Then, it is imperative to recognize that lead time analysis of this type of product over time establishes a specific analysis point for any particular SKU (Stock Keeping Unit). However, in general, these products have a short lead-time [41], since they are usually included in frequent orders, with a shorter, faster and less risky forecast horizon.

Finally, it was proposed that the model work with an aggregate demand as a fundamental input for its operation. This is in order to design a realistic model and minimize the statistical uncertainty that can arise when working with individual stores. Otherwise, it would be an impractical model because a company could have hundreds of stores distributed in a territory. This is explained in more detail within the formulation stage.

4.2. Needs assessment

A needs assessment stage was proposed to establish the elements that the model should have in order to avoid the mistake of failing to consider the current context of the research. This was carried out through a review of an academic and business landscape, where the current models of national and international companies such as Wal-Mart or Carrefour were consulted. In addition to the above, a review of academic articles from major scientific journals was conducted, taking into account journals from the last five years. The final synthesis constituted an extensive procedure in its own right. This was carried out through group consensus among the research team where it was decided that the following key points were to be considered:

- *A value chain vision.* It was necessary to have a proposal that helps to break down the existing barriers in the cooperation between the levels of the supply chain for this type of products, leaving aside individualistic benefits and seeking global efficiency as a value chain.
- *Order cycle.* Order cycle time is an important aspect, a VMI agreement must establish an appropriate balance related to costs and decision making on lot size.
- *Stockouts.* Stockouts must be considered in the model, due to the demand variability in relation to the customer service level, safety stocks, and customer service.
- *Case study*. In the literature review, most of the research carried out in recent years does not present a clear or real validation of their VMI models, most of them are only limited to numerical analysis. Therefore, it is necessary a contribution of a real case that illustrates the results of the designed model.

Additionally, the main factors that should be considered in the development of a model, and in the establishment of the guidelines of this collaborative agreement, were investigated. These can be seen as reflected in the set of variables and assumptions in the formulation of the model (section 4.4). It is important to mention that, in relation to the optimization of the model, it was recommended to orient it to minimize inventory costs.

4.3. Modeling technique selection

Due to the number of current modeling techniques and the nuances of each one, it was necessary to select the technique best suited to the characteristics that the model intended to achieve. The selection was made through an expert judgment based on an analytic hierarchy process (AHP).This activity had a group of 19 experts, whose professional research included the study of Production and Logistics Systems, and Operations Research. Having the most adept committee was crucial to reduce the error and uncertainty in the selection of the technique.

When multiple objectives are important to a decision maker, it may be difficult to choose between alternatives. Given a large number of judgments that can be made by expert staff, it was imperative to solve a Multi-Criteria Decision-Making problem. The analytic hierarchy process (AHP) is a method of measurement that relies on pair wise comparisons and the judgments of experts to derive priority scales. This method has been one of the most widely used decision-making techniques by decision-makers and researchers [42]–[44]. Thus, it was perfectly adjusted to the requirements of this problem, to carry out the selection of a modeling technique. The application of the AHP involves an extensive procedure for describing it[45], [46]. However, the fundamental aspects that were taken into account and the results obtained are detailed below.

A set of alternatives and a set of criteria are required to carry out an AHP. In the selection of alternatives, three modeling techniques were considered that were widely used by the authors of the scientific literature consulted, for the design of VMI models: 1) classical optimization, 2) non-linear programming, and 3) game theory. In respect to classical optimization, it is effective in obtaining the optimum solution of unconstrained and constrained continuous and differentiable functions. Besides, analytical methods make use of differential calculus in finding the optimum solution assuming that the function is differentiable concerning the design variables. Nevertheless, although analytical methods with essential and sufficient conditions are easier to use, these methods are difficult to apply for functions that are not continuous and/or not differentiable. Similarly, nonlinear programming continues to be an effective tool for supply chain modeling. The main advantage of a nonlinear programming approach is the guaranteed solution of a well-formulated problem and the ability to vary the supply chain parameters to understand the behavior of the system under various settings. Also, it make it possible to solve large-scale problems. However, nonlinear programming models are analytically hard to solve, and applying these methods to bigger cases can increase computation time. As a consequence, efficient algorithms and solution techniques could be adopted to find approximately optimal solutions and reduce the calculation times.

In addition, game theory has been recently applied to models for supply chain coordination. Given the current industry environment in which cooperative relations are becoming more prevalent in supply chains, a mutually beneficial approach addresses the coordination issues for vendor–buyer interactions. Thus, a VMI agreement can be modeled as either a dynamic cooperative or non-cooperative game concerning the overall supply chain. Also, once the whole game settles into an equilibrium, none of the chain members will be able to improve its payoff or profits by acting unilaterally without negatively affecting the performance of the other players. Despite these advantages, some drawbacks--mainly related to the fact that games that include multiple products that exist among multiple retailers are not easy to model--are present. As to which forms of games and roles are suitable for formulating coordinated decisions, these factors mainly depend on the competitive advantage of products in markets, as well as the organizational forms of their supply chain. In this respect, an individual entity's share of the market, and thus, negotiating power, also has significant effects as to the game's outcome.

Consequently, it was possible to select a recommended approach and contribute to aspects of it that had not yet been analyzed or explored. On the other hand, the criteria were selected by consensus, which was reached by defining the needs or requirements of the proposed model. The criteria for choosing a technique were as follows: 1) Ability to model complex systems, 2) Ease of replication, 3) Flexibility and 4) Variety of solutions. Each of these is explained in detail below:

- *Ability to model complex systems*. The selected technique should be useful to represent the study problem correctly, delineate the operational needs, and clearly define the expected outputs.
- *Ease of replication*. The selected technique should be easy to replicate, execute, and manipulate analytically. In order to reduce implementation costs, it is necessary to design a useful but not over-simplified model.
- *Flexibility.* As long as the established assumptions are met, the technique should allow the model to be applied in various scenarios
- *Variety of solutions*. The technique should serve to generate a variety of solutions that allow different

aspects or scenarios of the same case study to be analyzed.

In this way, Figure 1 shows the hierarchy proposed for this expert judgment. Note that the first level of the hierarchy is the goal (Modeling technique selection); The second level in the hierarchy is constituted by the criteria experts used to decide the modeling technique. The third level consists of the alternatives.

Thus, the second step in the AHP process was to derive the relative priorities or weights for the criteria. Evidently, the importance or weight of each criterion was different and because of this, experts first were required to derive by pairwise comparisons the relative priority of each criterion with respect to each of the others using a numerical scale for comparison. Although the exact method will not show in detail here, the general idea is simple. Next, once judgments were entered, it was necessary to check that they were consistent. For this purpose, AHP calculates a consistency ratio (CR) comparing the consistency index (CI) of the matrix in question (the one with our judgments) versus the consistency index of a random-like matrix (RI). In AHP, the consistency ratio is defined as CR = CI/RI.Saaty (1985)[47] has shown that a consistency ratio (CR) of 0.10 or less is acceptable to continue the AHP analysis. If the consistency ratio is greater than 0.10, it is necessary to revise the judgments to locate the cause of the inconsistency and correct it.

It is important to mention that the consistency index of the consensual valuations was calculated. This value did not reflect inconsistencies. Then, judgments about the alternatives were consistent and there was no contradiction in any of them. The result, called priority vector of alternatives constituted the solution of the expert judgment. This vector presented a preference percentage for each of the alternatives. Table 1 shows the results.

Table 1: Priority vector of alternatives

Alternatives	Ranking
Classical or unconstrained optimization	0.491
Game Theory	0.238
Nonlinear programming	0.271
Total	1.0

According to the results, it was clear from the results that the technique that best complied with the requirements – based on the consistency of all experts' judgments – was Classical Optimization. Its success rate was 49.1% higher than Game Theory and Nonlinear Programming, which has success rates of 23.8% and 27.1% respectively. Due to the above, it is possible to emphasize that execution process of expert judgment was correctly carried out for the selection of an appropriate technique that fulfilled a set of criteria and distinguished alternatives.

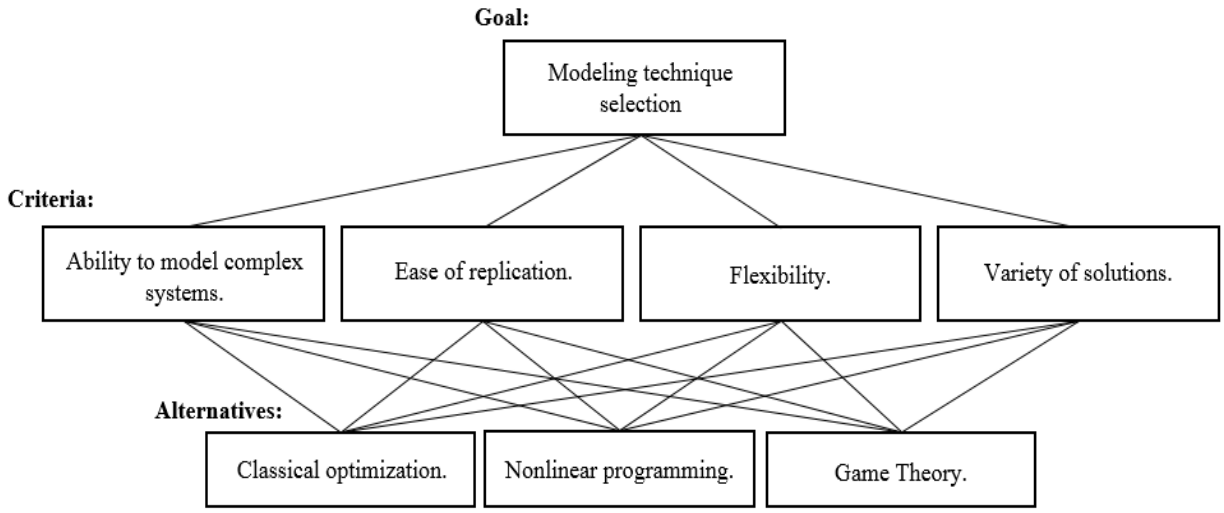


Figure 1: Hierarchical structure of AHP.

4.4. Model formulation and assumptions

A VMI model was designed for impulse purchase products in a two-level supply chain with a vendor and a buyer. The development of the model was established as a cooperative agreement between both parties, in which the buyer shares the sales and inventory information of a set of products with the vendor. Subsequently, the vendor places a suggested order based on the information received. This will be a consolidated order, which means that it will group together the supply needs of all stores. In simple terms, in the traditional inventory management, a buyer makes their own decisions regarding the order size while in VMI, a buyer shares their sale and inventory data with a vendor such that the vendor can determine the order size for both. This policy can prevent stocking undesired inventories and hence provides greater visibility to inventory replenishment and supply planning. The inclusion of a safety stock and stock out costs were also considered. Table 2 presents the mathematical notation.

Table 2: Notations of parameters and variables

Index:	
$i = 1,2,..m$	Index for products.
$j = 1,2..w$	Index for stores.
Parameters:	
D_i	Forecasted demand.
P_i	Replenishment rate.
C_{oi}	Cost per item.
S	Fixed ordering cost (buyer).
C_{pi}	Variable ordering cost (buyer).
H	Order cost (vendor): order preparation.
C_{Pi}	Order cost (vendor): shipment.
G_i	Order cost (vendor): packaging.
C_{si}	Cost per item short.
$E[x_i]$	Expected number of shortages per order.
T	Time horizon.
C_{Hi}	Vendor's holding cost.
C_{hi}	Buyer's holding cost.
C_{Vi}	Pertain to the cost of holding stock in the store display shelf: cost of shelf space.
V_{Ai}	Pertain to the cost of holding stock in the store display shelf: cost per volume.
Decision variables:	
n	Common ordering frequency: number of orders per unit time.
n^*	Optimal ordering frequency: optimal number of orders per unit time
Q_i	Order size (items/order).
Q_i^*	Optimal order size.
L	Lead Time.
I_{buyer}	Average inventory (buyer).
I_{vendor}	Average inventory (vendor).
s_i	Reorder point.
t_1	Replenishment cycle.
SS_i	Safety stock.
TC_V	Total Cost (Vendor).
TC_B	Total Cost(Buyer).
TC_{SC}	Total Cost(Supply Chain).
TRC_{SC}	Total Relevant Cost.

Also, in this model the following assumptions were defined:

- *Single vendor and single buyer with m products (multiproduct systems). The buyer can have w stores.* VMI agreements are mostly implemented in a two-level supply chain. Furthermore, a collaborative agreement between two companies contains key parameters related to specific policies. This model then involves a single vendor serving a set of buyer's stores. This one-to-many model is not only ubiquitous, but it also describes the distribution activities of many companies while keeping the analytical complexity at a tractable level.
- *The information of the buyer's replenishment decision parameters is available to the vendor***.** Sharing sales and inventory information is an indispensable assumption in a VMI agreement.
- *Planning horizon of one period*. Considering an uncertain demand, it is necessary to contemplate a planning horizon of one period, which allows to reach a higher level of accuracy in the results.
- *The flow of information between the levels of the supply chain is automatic and in real time*. This is a condition that must be met to ensure that shared information is consistent, real and error-free.
- *The vendor has demand visibility of their buyer.* A major assumption of VMI is the transfer of information between the vendor and the buyer. The retailing industry in particular is sensitive to the vicissitudes of consumer demand. The uncertain nature of the demand is one of the motivations to consider aggregating multiple products in a single order. A forecasted demand by the vendor is considered.
- *Shortage cost is the loss of sales revenue from not meeting the demand*. This assumption allows us to consider certain real inventory replenishment policies that involve goals related to the level of service.
- *Quantity discounts are not permitted.* Quantity discounts is not an important aspect in the formulation of the model because the model seeks to optimize the replenishment cycle of the products.
- *The holding cost is same for all stores.* This assumption allows an adequate modeling of the holding cost in the stores, and at the same time, an appropriate simplification of reality is achieved.
- *The model assumes constant lead time*.
- *Safety Stock is required*. A safety stock is necessary to reduce the probability of stock-outs.
- *The model assumes that demand is uncertain and follows a normal distribution*: Although empirical probabilities can be used in this study, they are inconvenient for many reasons. Firstly, they require maintaining a record of the demand history for every product. This can be costly and unwieldy. Secondly, the distribution must be expressed as different probabilities for each of the past values, and products may have an even wider range of past values. Finally, it is more difficult and impractical to compute optimal inventory policies with empirical distributions. For these reasons, in practice, it is common and most popular for inventory application, to assume that demand follows a normal distribution. One reason is the frequency with which it seems to accurately model demand fluctuations. Another is its convenience and the fact that, according to the central limit theory, for sufficiently large samples ($n \geq 30$), the sample means will be distributed around the population mean approximately in a normal distribution.

4.4.1. Buyer's total cost

When analyzing the different processes in a VMI model, it is clear that the purchase of products by the buyer is one of the main features. The purchase cost depends on each product $i$and the quantities ordered of that product. Each product $i$is assigned to a SKU (Stock-keeping unit) and has a unit purchase cost(C_{oi}), which is assumed by the buyer. At the beginning of each cycle, there is a forecasted demand for each product (D_i), which must be met in n_i orders. Then, the order size (Q_i) to meet this demand is defined as:

$$Q_i = \frac{D_i}{n_i} \tag{1}$$

In this way, the purchase cost will be the multiplication of the order size (Q_i), the purchase unit cost (C_{oi}), and the number of orders (n_i) in the cycle, for each of the m products. That is:

$$\text{Purchasing cost} = \sum_{i=1}^{m} n_i C_{oi} Q_i = \sum_{i=1}^{m} n_i C_{oi} \left(\frac{D_i}{n_i}\right) = \sum_{i=1}^{m} C_{oi} D_i \tag{2}$$

Also, an order cost is generated that depends on the characteristics of the product i. This cost has two components: a fixed cost (S) and a variable unit order cost (C_{pi}).The fixed cost (S)corresponds to the costs related to the administration and reception of the order.On the other hand, it was considered that for each type of product i, there is a variable unit cost of ordering (C_{pi}), which is associated with the follow-up costs that depend on the special requirements of these products. Hence, the cost of ordering is the sum of the fixed (S) and variable cost (C_{pi}) incurred in placing an order, multiplied by the number of orders $n_{i,}$:

$$\text{Ordering cost} = \sum_{i=1}^{m} n_i (S + C_{pi}) \tag{3}$$

Inventory management seeks to minimize stock out costs that occur when demand is greater than anticipated and cannot be met immediately. The process of estimating this cost involves an overview of non-quantitative variables including, but not limited to, customer perception, the long-term reliability of these perceptions, and the loss of consumer loyalty[48]. Subsequently, the stock-out costs consists of three elements. The first is $n_i$which is the ordering frequency, or the total number of orders placed over the whole length of time. The second term is C_{si}, which is the cost incurred due to the shortage of each product i. And then the last term is the Expected Unit Short $E[x_i]$, which is equal toσ_{DL} times $G_u(k)$, where σ_{DL} is the standard deviation of the demand over lead time and $G_u(k)$ is the unit normal loss function. The mathematical procedure for calculating the Expected Unit Shortis well-known, hence it can be entirely omitted.

$$Stock\ out\ costs = \sum_{i=1}^{m} n_i C_{si} E[x_i] = \sum_{i=1}^{m} n_i C_{si} \sigma_{DL} G_u(k) \tag{4}$$

In addition, it is important to estimate the buyer's holding cost which depends on inventory levels over time. The inventory level is represented in Figure 3, where Q is the order size, and SS is the safety stock.

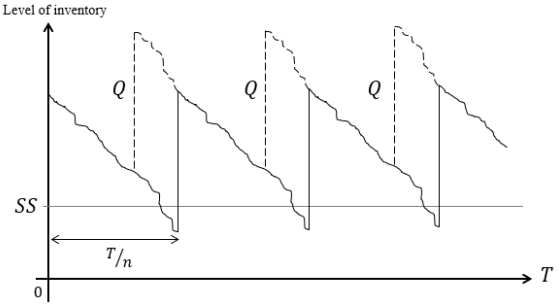


Figure 2: Estimating the buyer's holding cost.

According to the Figure 2, the inventory level can be approached to a linear function of the form $y = mx + b$, where m is equal to $-Qn/T$, and $b = Q$. Therefore, the average inventory (I_{buyer})can be readily obtained as follows (according to the linearity assumption adopted),

$$I_{buyer} = \frac{QTn}{2nT} + SS = \frac{Q}{2} + SS \tag{5}$$

Thus, the quantity $T(Q + SS)/2n$ is the area of the triangle with height $Q + SS$ and base T/n, which is divided by the cycle length T/n to calculate the average inventory over the cycle. The cost of holding stock in the store j for a specific product is the sum of theaverage inventory I_{buyer}, multiplied by the cost of holding that product in that facility j (C_{hj}). Therefore, since each store has an order quantity Q_j, it is possible to estimate the cost of holding stock per unit of a certain product in the store j as:

$$C_{hj}\left(\frac{Q_j}{2} + SS_j\right) \tag{6}$$

Next, under the assumption that the cost of holding stock of a certain product in the store$j$ is the same for all stores and that there are w stores, this condition translates to:

$$C_h\left(\sum_{j=1}^{w} \frac{Q_j}{2} + \sum_{j=1}^{w} SS_j\right) \tag{7}$$

In addition, since it is possible to express the consolidated units of a product for all stores in terms of two variables Q and SS, That is,

$$\sum_{j=1}^{w} Q_j = Q \quad ; \sum_{j=1}^{w} SS_j = SS \tag{8}$$

Then, according to Eq. (7) and Eq. (8), after rearranging terms, the cost of holding stock in all buyer's stores for a particular product would be:

$$C_h\left(\frac{Q}{2} + SS\right) \tag{9}$$

Substituting into Eq. (9) the value Q according to Eq. (1), this condition translates to:

$$C_h\left(\frac{D}{2n} + SS\right) \tag{10}$$

Finally, using the fact that the VMI agreement included several products, Eq. (10) can be generalized to the case in which there are m products, as follows:

$$\sum_{i=1}^{m} C_{hi}\left(\frac{D_i}{2n_i} + SS_i\right) \tag{11}$$

Therefore, according to Eq.(2), Eq. (3), Eq. (4), and Eq. (11), the buyer's total cost(TC_B)would be:

$$TC_B = \sum_{i=1}^{m} C_{oi}D_i + \sum_{i=1}^{m} n_i(S + C_{pi}) + \sum_{i=1}^{m} n_i C_{si} E[x_i] + \sum_{i=1}^{m} C_{hi}\left(\frac{D_i}{2n_i} + SS_i\right) \quad (12)$$

4.4.2. Vendor's total cost

The vendor's order cost has three components. First, a fixed component (H) that is determined by the cost of preparing n_i orders. It typically includes fees for placing the order, and all kinds of clerical costs related to invoice processing, accounting, or communication. Second, a shipping cost (C_{Pi}) that depends on the type of product being shipped. Thus, it is multiplied by the number of orders. Finally, a packaging cost (G_i) associated with pre-consolidation or package formation includes weighting, labeling and packaging. This cost is dependent on the number of units to be shipped (D_i/n_i). Thus, the vendor's order cost is defined as:

$$\text{Vendor's ordering cost} = \sum_{i=1}^{m} n_i\left(H + C_{Pi} + \frac{D_i}{n_i}G_i\right) = \sum_{i=1}^{m} n_i(H + C_{Pi}) + \sum_{i=1}^{m} D_i G_i \quad (13)$$

Furthermore, in this vendor-managed inventory (VMI) model for impulse purchase products, the vendor's holding cost has two components. The first of these two components is the cost of holding stock in the storage facilities that is incurred before serving a set of buyer's stores. The second component is the cost of holding stock in the store display shelf that is incurred by allowing the vendor to display their product inside the buyer's store. Each of these components is discussed below:

Assuming a replenishment or production rate (P), the vendor's holding cost incurred in the storage activities is equal to the multiplication of the average inventory (I_{vendor}) and the cost of holding stock in their facility (C_H).The average inventory was calculated according to Figure3, where T is the total time horizon, n is the number of orders, and t_1 is the time period required for the vendor to replenish (or produce) an entire batch quantity$Q$ at a rate P.

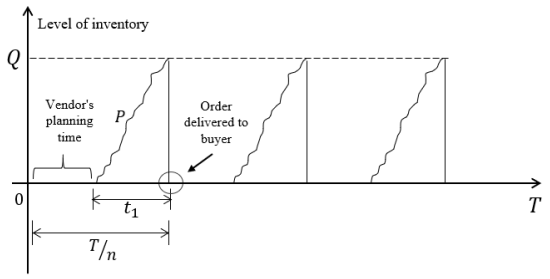


Figure 3. Estimating the vendor's holding cost.

According to the Figure 2, the inventory level can be approximated to a linear function of the form$y = mx + b$,where$m$is equal tothe replenishment or production rate $\left(Q/t_1\right)$,and$b = y - mx = Q - Q/t_1\,(T/n)$, since the inventory level is Q, after T/n units of time.The vendor's average inventory (I_{vendor}) can be readily obtained as follows (according to the linearity assumption adopted). The quantity $Qt_1/2$ is the area of the triangle with height Q and base t_1, which is divided by the cycle length T/n to calculate the average inventory over the cycle.

$$I_{vendor} = \frac{nQt_1}{2T} \quad (14)$$

Thus, the vendor's holding cost stock incurred in the storage activities can be expressed as:

$$C_H I_{DC} = C_H\left(\frac{nQt_1}{2T}\right) \quad (15)$$

Using the fact that $t_1 = Q/P$ and $Q = D/n$, which are substituted into Eq. (15), the cost of holding stock in the vendor's distribution center would be:

$$C_H\left(\frac{D^2}{2nPT}\right) \quad (16)$$

Eq. (16) can be generalized to the case in which there are m products, as follows:

$$\sum_{i=1}^{m} C_{Hi}\left(\frac{{D_i}^2}{2n_i P_i T}\right) \quad (17)$$

Secondly, the cost of holding stock in the store display shelf depends on the shelf space to display the product in the store and it is defined as the multiplication of the product volume (V_{Ai}) and the shelf space cost (C_{Vi}). Using the fact that there are m products on display, this condition translates into:

$$\sum_{i=1}^{m} V_{Ai} C_{Vi} \quad (18)$$

Therefore, the vendor's total cost is the sum of Eq. (13),Eq.(17), and Eq.(18):

$$TC_V = \sum_{i=1}^{m} n_i(H + C_{Pi}) + \sum_{i=1}^{m} D_i G_i + \sum_{i=1}^{m} C_{Hi}\left(\frac{{D_i}^2}{2n_i P_i T}\right) + \sum_{i=1}^{m} V_{Ai} C_{Vi} \quad (19)$$

4.4.3.Supply Chain

Because we are looking for a general optimization with the VMI agreement, the total cost of the supply chain (CT_{SC}) will be equal to the sum of the total costs of each party (see Eq.(12) and Eq.(19)):

$$TC_{SC} = TC_B + TC_V = \left[\sum_{i=1}^{m} C_{oi}D_i + \sum_{i=1}^{m} n_i(S + C_{pi}) + \sum_{i=1}^{m} n_i C_{si}E[x_i] + \sum_{i=1}^{m} C_{hi}\left(\frac{D_i}{2n_i} + SS_i\right)\right] + \left[\sum_{i=1}^{m} n_i(H + C_{\boldsymbol{P}i}) + \sum_{i=1}^{m} D_i G_i + \sum_{i=1}^{m} C_{\boldsymbol{H}i}\left(\frac{{D_i}^2}{2n_i P_i T}\right) + \sum_{i=1}^{m} V_{Ai}C_{Vi}\right] \quad (20)$$

To optimize supply chain costs, it was proposed to exploit fixed costs by aggregating multiple products in a single order. In other words, the same amount of orders is placed for all products purchased from the same vendor. In this way, the objective function was defined as follows:

$$TC_{SC} = \left[\sum_{i=1}^{m} C_{oi}D_i + n\sum_{i=1}^{m}(S + C_{pi}) + n\sum_{i=1}^{m} C_{si}E[x_i] + \frac{1}{2n}\sum_{i=1}^{m} C_{hi}D_i + \sum_{i=1}^{m} C_{hi}SS_i\right] + \left[n\sum_{i=1}^{m}(H + C_{\boldsymbol{P}i}) + \sum_{i=1}^{m} D_i G_i + \frac{1}{2nT}\sum_{i=1}^{m} C_{\boldsymbol{H}i}\left(\frac{{D_i}^2}{P_i}\right) + \sum_{i=1}^{m} V_{Ai}C_{Vi}\right] \quad (21)$$

Note that in Eq.(21) the sub-index for variable n has been removed. In this way, the decision variable becomes the number of orders to be placed. Therefore, by eliminating those terms that do not depend on n (since they are not relevant for optimization), the supply chain cost function is simple function of n. This function was called Total Relevant Costs(CTR_{SC}), as shown below:

$$TRC_{SC} = \left[n\left(\sum_{i=1}^{m} C_{pi} + S\right) + n\sum_{i=1}^{m} C_{si}E[x_i] + \frac{1}{2n}\sum_{i=1}^{m} C_{hi}D_i\right] + \left[n\left(\sum_{i=1}^{m} C_{\boldsymbol{P}i} + H\right) + \frac{1}{2nT}\sum_{i=1}^{m} C_{\boldsymbol{H}i}\left(\frac{{D_i}^2}{P_i}\right)\right] \quad (22)$$

From the fact that classical optimization was selected to formulate the model, the purpose was to minimize the total relevant cost function by finding an extreme point solution. Using calculus, the derivative of the total relevant cost function (Eq.(22)) was taken and set equal to zero and solve for n. The total relevant costs curve was convex i.e. curvature is upward then a minimizer was obtained. In addition, taking into account that all function values were greater than 0, which indicated that $f''(n) > 0$, it was concluded that the function has a point where it takes a minimum value. By setting the first derivative equal to 0 and solving for n, the optimum number of orders is given by:

$$n^* = \sqrt{\frac{\sum_{i=1}^{m} C_{hi}(D_i) + \frac{1}{T}\sum_{i=1}^{m} C_{\boldsymbol{H}i}\left(\frac{{D_i}^2}{P_i}\right)}{2\left(\sum_{i=1}^{m} C_{pi} + \sum_{i=1}^{m} C_{si}E[x_i] + \sum_{i=1}^{m} C_{\boldsymbol{P}i} + S + H\right)}} \quad (23)$$

Similarly, based on Eq.(1) and the n^* definition, it is possible to represent the optimal order lot size as follows:

$$Q_i^* = \frac{D_i}{n^*} \quad (24)$$

In this way, benefits could be expected by optimizing the total inventory costs of the supply chain by exploiting the fixed costs present in the VMI agreement operations. The reduced fixed cost of receiving makes it optimal to reduce the lot size ordered for each product, thus reducing cycle inventory. Aggregation of orders for different products also brings advantages associated with administrative and implementation activities. In addition, the inventory policy has another parameter that needs to be estimated, the reorder point, s_i. This variable is important when trying to find what amount of inventory is sufficient to handle all the expected demand over the lead time for each product i. As previously mentioned, adding a safety stock is necessary because the demand is variable, and how much safety stock to add is determined by how badly the buyer would like to avoid stockout. Thus, the policy will be order Q_i^* when inventory position is less than or equal to s_i, the reorder point for a product i. Therefore, it can be assumed that $s_i = \mu_{DL} + SS_i = \mu_{DL} + k\sigma_{DL}$, where the reorder point is equal to the forecast mean demand over lead time μ_{DL} (or the expected demand over the lead time period), plus the safety stock SS_i. And SS_i is simply, k, the safety stock factor, times σ_{DL}, the standard deviation of the error of the forecast over the lead time, or the root mean square error.

Now, considering a cost-minimization approach, as this approach is most commonly used in large and more sophisticated organizations, it is possible to analyze the value of k. Using the total cost equation and taking the first-order condition to minimize the total cost, the only relevant costs are the safety stock and the stock out cost because they are the only ones with the variable k(Eq. 4 and Eq. 11). By taking a first-order condition and setting it equal to 0, it ends with this condition: k is optimal for the minimum total relevant cost if the probability of the demand x greater than$k$ is equal to Q_i^* times the cost of holding C_{hi}, divided by the demandD_i, times the cost of shortage C_{si}: $P[x \geq k] = Q_i^* C_{hi}/D_i C_{si}$. Notice that it makes sense that the ratio of C_{hi} over C_{si} would help determine how much a party would want to stock, because it is just a tradeoff between having too much and having too little. Then, the decision rule--when assuming cost of shortage--first has to make sure that the following expression is less than or equal to 1:$Q_i^* C_{hi}/D_i C_{si}$. And if it is, then it can be determined that the probability of stockout is equal to $P[x \geq k]$. Otherwise, k should be set as low as management allows.

In this way, it is valuable to summarize both why and how the management of impulse purchase products differs

from other types of products as well as how this was reflected in the proposed model. Firstly, it is necessary to start by recognizing that VMI agreements are common among companies marketing impulse purchase products; that is, the companies that have a greater implementation of these types of collaborative models are big-box and department store retailers that offer a wide breadth of products. Second, the model assumed a planning horizon of one period because considering that impulse purchase products are characterized by an uncertain demand, it is necessary to contemplate a planning horizon that allows for the reaching higher levels of accuracy in the results. In relation to the above mentioned, in practice, shortage cost is the loss of sales revenue from not meeting the demand in the market place for impulse purchase products, and inventory management should minimize stock out costs that occur when demand is greater than anticipated and cannot be met immediately. All these aspects were considered in the model, even including some important variables by modelling the respective cost of inventory and stock-outs, like ordering frequency, cost incurred due to the shortage of each product, expected unit short, etc. Also, the safety stock was added to reduce the probability of stock-out because many buying decisions for these products are based on impulse and at the point of purchase.

It should also be noted that impulse buying products are characterized as products that cost little and are quickly consumed. Additionally, in general terms, an assumption related to constant lead time, it is an appropriate simplification of reality. Usually, a short lead time is a relatively common occurrence in the impulse purchase products market, as information distortion is magnified if replenishment lead times between stages are volatile or long. Then, by decreasing the replenishment lead time, companies and managers can minimize the uncertainty of demand during the lead time. Another critical aspect of these products is the way to calculate holding costs in practice. Commonly, managers consider the cost of holding stock in the storage facilities that are incurred before serving a set of buyer's stores, and the cost of holding stock in the display shelf defined as the multiplication of the product volume and the shelf space cost (remember that generally, these items are strategically displayed in hot spots), which depends on the shelf space to display the product in the store. Each of these cost components was included in the model. Last but not least, firms in this sector may order in large lots because the presence of fixed costs associated with ordering, quantity discounts in product pricing, and short-term promotions, encourages different stages of a supply chain to exploit economies of scale and order in large lots, not mention that another main advantage of ordering all products jointly is that a VMI agreement would be easy to administer and implement.

4.5. Validation and sensitivity analysis

To validate the model, a case study was conducted with two major companies that market impulse purchase products, where the cost reduction of the model was compared against a conventional form of inventory replenishment policy in a traditional supply chain. The vendor was a sugar confectionery producer company, and the buyer was a national retail company that is currently working on collaborative agreements with this impulse purchase products vendor. A copy of the data used in this study is available on request for any interested researcher. Please contact the corresponding author. The current scenario (traditional supply chain) was compared with two scenarios based on the proposed VMI model. Beyond having as it is purpose to execute the model in a real-life scenario, the main objective was to learn the economic impact as a key indicator of its implementation. It is also important to mention that the model was executed with eight products that are part of the VMI agreement, and historical data from the two previous years (24 months) were used to make key inferences from the model. The model measured performance over a five-month period (months 25, 26, 27, 28, and 29).

- *Scenario 1: VMI model with economies of scale to exploit fixed costs.* The VMI model was analytically executed with actual input data from both companies over a five-month period, to calculate and compare the total costs. In this scenario, the vendor and buyer have a VMI agreement for eight impulse purchase products, and the optimal number of orders(n^*)was calculated to reduce fixed costs in supply chain operation (Eq.(23)). In other words, products were ordered jointly (aggregate orders).
- *Scenario 2: VMI model without economies of scale to exploit fixed costs.* The VMI model was analytically executed with actual input data from both companies over a five-month period, to calculate and compare the total costs. In this scenario, the vendor and buyer have a VMI agreement for eight impulse purchase products, However, unlike scenario 1, the optimization proposal associated with the exploitation of fixed costs in the agreement, was not applied here. This means that products were ordered independently.
- *Scenario 3: traditional supply chain.* The total costs of both companies were calculated in a traditional scenario over a five-month period. In thistraditional supply chain, the buyer is responsible for tracking inventory levels at theirlocations and create a purchase order to make appropriate inventory replenishmentdecisions. The vendor has no information aboutfuture demand or inventory levels at theirbuyer's facilities and so has no prior knowledge about the quantity and time of thepurchase order.

The results were compared for each of the five individual periods (months 25, 26, 27, 28, and 29) and for the entire period (from month 25 to month 29). These results can be found in the Table 3.

In the vendor's case, it was noted that the greatest economic benefit is achieved in Scenario 1, with a savings of 35.52%, and 47.55% compared to Scenarios 2 and 3 respectively.As for the buyer's total relevant cost, scenario 1 showed beneficial results as well, with a savings of 32.41%, and 42.78% compared to Scenarios 2 and 3 respectively. Comparing the total relevant cost for the supply chain, the results showed a reduction in cost of 45.13% in the proposed model compared to the traditional scheme of both companies (Scenario 3), and savings of 33.91% compared to Scenario 2. This demonstrated that the model proposed in this case study obtained the best results in each of the scenarios for each of the parties to the agreement, as well as for the supply chain in general. In addition, for this case study, the exploitation of fixed costs was beneficial in the agreement.

Although it is true that the model works with an aggregate demand (the sum of the demand of all the products), it was relevant to analyze the impact of variations in demand on the results. Therefore, knowing how well the model adapts to variations in demand and how these variations affect the optimal order size were important questions. By replacing in Eq. (24), the value of n^*established by Eq. (23), the following equation was obtained:

$$Q_i^* = \frac{D_i}{\sqrt{\frac{\sum_{i=1}^{m} C_{hi}(D_i) + \frac{1}{T}\sum_{i=1}^{m} C_{\boldsymbol{H}i}\left(\frac{D_i^2}{P_i}\right)}{2\left(\sum_{i=1}^{m} C_{pi} + \sum_{i=1}^{m} C_{si}E[x_i] + \sum_{i=1}^{m} C_{\boldsymbol{P}i} + S + H\right)}}} \quad (25)$$

Then, assuming a percentage change denoted as α in demand of a product i, the variation in optimal lot size $(Q_i^*)$is given by:

$$Q_i^* = \frac{(1+\alpha) * D_i}{\sqrt{\frac{(1+\alpha) * \sum_{i=1}^{m} C_{hi}(D_i) + \frac{(1+\alpha)^2}{T} * \sum_{i=1}^{m} C_{\boldsymbol{H}i}\left(\frac{D_i^2}{P_i}\right)}{2\left(\sum_{i=1}^{m} C_{pi} + \sum_{i=1}^{m} C_{si}E[x_i] + \sum_{i=1}^{m} C_{\boldsymbol{P}i} + S + H\right)}}} \quad (26)$$

From Eq. (26), it can be inferred that once an optimal order size has been calculated, a new optimal order size can be calculated according to a variation in demand of a product i. Also, if all products have the same variation in demand, the percentage variation will also be the same. Then, the increase or decrease in demand will be directly proportional to the size of the order.

To prove this, an analysis of the demand variation was carried out for all products. The results showed that the percentage change in order size was the same for each product. Variations were applied in an interval between ± 30%, the order size showed a variation between -9.48% and 6.52%, which indicates that the proposed model was adapted to fluctuations in demand without greatly varying the results in the period analyzed (months 25, 26, 27, 28, and 29), as shown in Figure 4.

Similarly, the total relevant costs showed a decrease at each level of the supply chain, which translates into a reduction in overall total relevant costs. Applying the same demand variation criterion used previously (variations in the order of ±30%), it was noticeable that the increase in costs at each level was proportional to the percentage increase in demand. This is illustrated in Figure 5. The results indicated that, although the total relevant cost of the supply chain was largely dependent on demand, the exploitation of fixed costs helped mitigate the impact due to large variations.

It is possible to point out that the validation carried out demonstrated positive benefits in relation to the economic impact that the implementation of the proposed model could bring about, provided that the set of assumptions and guidelines set forth throughout its formulation are complied with. Also, the exploitation of fixed costs and their inherent advantages were correct in regard to the financial terms within the agreement. Finally, with the development of the sensitivity analysis, it was noted that the research may lead us to inferences related to the generation of strategies in environments of uncertain demand. The model seems to adjust itself correctly to fluctuations in demand as a result of the variations in the optimal lot size, which would be a function of the percentage increase in demand and not of the associated costs.

Table 3: TotalRelevant cost savings (%): A comparison analysis

Savings percentage	Month	Scenario 1 vs. Scenario 2	Scenario 1 vs. Scenario 3
Total Relevant Cost: Vendor	25	34.79%	44.96%
	26	36.24%	52.83%
	27	35.40%	50.06%
	28	35.31%	43.14%
	29	36.04%	47.67%
	Entire period (5 months):	35.52%	47.55%
Total Relevant Cost: Buyer	25	31.86%	41.30%
	26	32.80%	45.37%
	27	32.25%	43.97%
	28	32.64%	40.79%
	29	32.51%	42.71%
	Entire period (5 months):	32.41%	42.78%
Total Relevant Cost: Supply Chain	25	33.29%	43.11%
	26	34.38%	48.98%
	27	33.73%	46.94%
	28	33.98%	41.97%
	29	34.21%	45.14%
	Entire period (5 months):	33.91%	45.13%

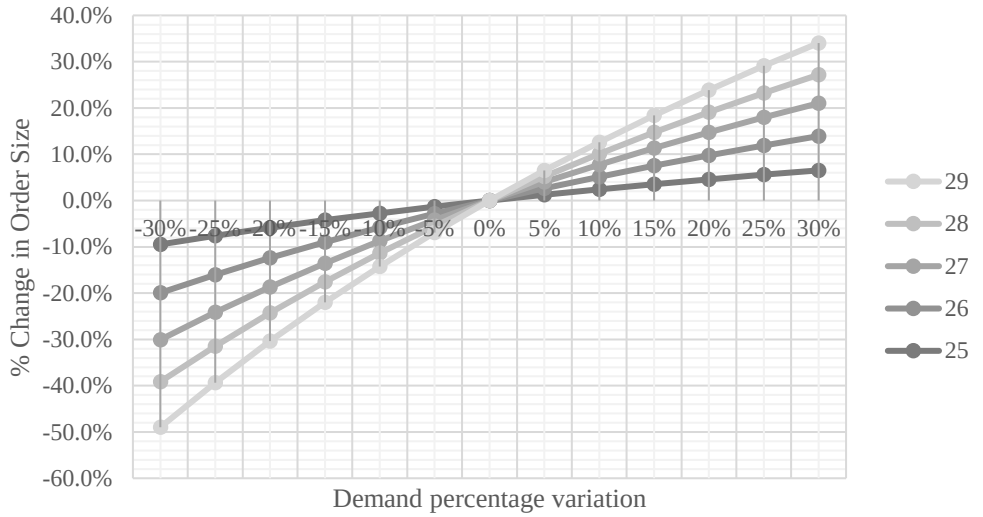


Figure 4. % Change in Order Size.

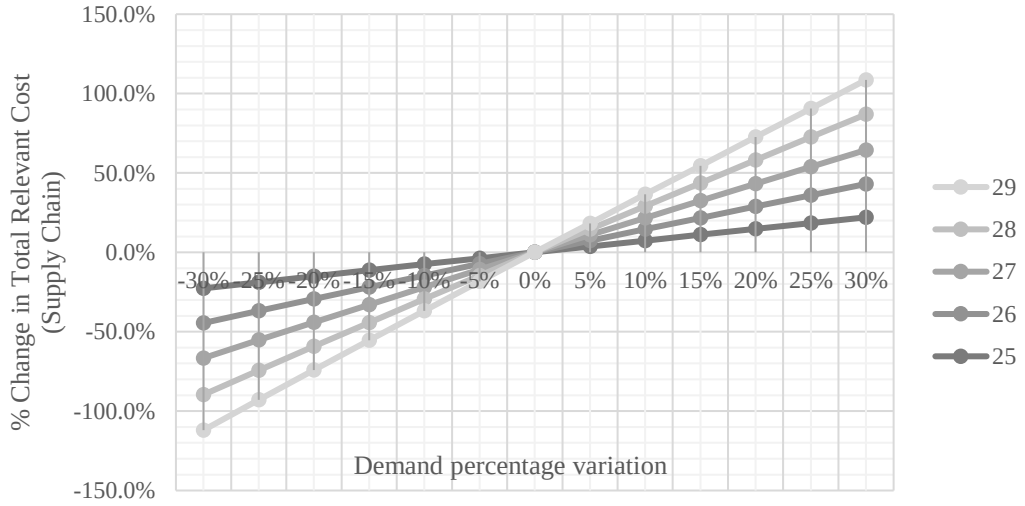


Figure 5. % Change in Total Relevant Cost (Supply Chain).

5. Conclusion

The proposed model mathematically represents a scheme that achieves a coordinated solution in a two-tier supply chain for impulse purchase products. Also, managing economies of scale in the model was formulated by exploiting fixed costs present in the supply chain. Hence, products are ordered and delivered jointly, and the model will distribute the ordering cost over a number of items. In this way, the model brings economic and administrative benefits related to the management of impulse purchase products for both parties, minimizing total costs and optimizing logistics operations. A lower cycle inventory is better because the average flow time is lower, inventory holding costs are lower, and working capital requirements are lower. Via contracts, the vendor and buyer are willing to operate with the same order sizes.

In the same way, the validation stage provides evidence that managing economies of scale that exploit fixed costs brings advantages in financial terms by reducing the ordering costs of the vendor and the buyer. Satisfactory results were obtained by comparing an executed scenario with the proposed model (scenario 1) against a scenario that did not consider the management of economies of scale (scenario 2) over a period of five months, obtaining savings of 35.52%, 32.41% and 33.91%, for the vendor, buyer, and supply chain respectively. Similarly, by comparing the economic impact of model execution (scenario 1) against a scenario based on a traditional supply chain without any non-cooperative agreement (scenario 3) for five months, the proposed model obtained the best measurement results in terms of costs, with an improvement of approximately 47.55%, 42.78%, and 45.13% for the vendor, buyer and supply chain respectively.

Finally, the validation phase of the model can be complemented with additional real cases. This is an aspect that would provide an enormous benefit to ensure greater credibility. Through the execution of case studies that contain different sets of variables, which may alter the behavior for demand, such as, the product portfolio or the

agreement guidelines. In addition, it is possible to use this research as a starting point to move towards much more elaborate models, which reflect greater complexity of the real system, and better representation of operational characteristics along the supply chain. As far as future research is concerned, this model presents several aspects to take into account, mainly its implementation, since it would be interesting to see its development, impact, and response with respect to other business environments.

References

[1] H. M. Stellingwerf, G. Laporte, F. C. A. M. Cruijssen, A. Kanellopoulos, and J. M. Bloemhof, "Quantifying the environmental and economic benefits of cooperation: A case study in temperature-controlled food logistics," Transp. Res. Part D Transp. Environ., vol. 65, pp. 178–193, 2018.

[2] A. Sainathan and H. Groenevelt, "Vendor managed inventory contracts – coordinating the supply chain while looking from the vendor's perspective," Eur. J. Oper. Res., vol. 272, no. 1, pp. 249–260, 2019.

[3] J.-Y. Lee and R. K. Cho, "Optimal (z, Z)-type contracts for vendor-managed inventory," Int. J. Prod. Econ., vol. 202, pp. 32–44, 2018.

[4] A. Shaban, F. Costantino, G. Di Gravio, and M. Tronci, "A new efficient collaboration model for multi-echelon supply chains," Expert Syst. Appl., vol. 128, pp. 54–66, 2019.

[5] F. Neves-Moreira, B. Almada-Lobo, J.-F. Cordeau, L. Guimarães, and R. Jans, "Solving a large multi-product production-routing problem with delivery time windows," Omega, vol. 86, pp. 154–172, 2019.

[6] K. Vitasek, "Supply Chain Management Definitions and Glossary," Council of Supply Chain Management Professionals (CSCMP), 2013. [Online]. Available: http://cscmp.org/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms/CSCMP/Educate/. [Accessed: 05-Apr-2017].

[7] A. J. Badgaiyan, A. Verma, and S. Dixit, "Impulsive buying tendency: Measuring important relationships with a new perspective and an indigenous scale," IIMB Manag. Rev., vol. 28, no. 4, pp. 186–199, 2016.

[8] C. Horváth and M. van Birgelen, "The role of brands in the behavior and purchase decisions of compulsive versus noncompulsive buyers," Eur. J. Mark., vol. 49, no. 1/2, pp. 2–21, Jan. 2015.

[9] J. J. Kacen, J. D. Hess, and D. Walker, "Spontaneous selection: The influence of product and retailing factors on consumer impulse purchases," J. Retail. Consum. Serv., vol. 19, no. 6, pp. 578–588, 2012.

[10] S. Bellini, M. G. Cardinali, and B. Grandi, "A structural equation model of impulse buying behaviour in grocery retailing," J. Retail. Consum. Serv., vol. 36, no. February, pp. 164–171, 2017.

[11] S. A. Sofi and S. A. Najar, "Impact of personality influencers on psychological paradigms: An empirical-discourse of big five framework and impulsive buying behaviour," Eur. Res. Manag. Bus. Econ., vol. 24, no. 2, pp. 71–81, 2018.

[12] A. A. Darrat, M. A. Darrat, and D. Amyx, "How impulse buying influences compulsive buying: The central role of consumer anxiety and escapism," J. Retail. Consum. Serv., vol. 31, pp. 103–108, 2016.

[13] T. Flamand, A. Ghoniem, and B. Maddah, "Promoting impulse buying by allocating retail shelf space to grouped product categories," J. Oper. Res. Soc., vol. 67, no. 7, pp. 953–969, Jul. 2016.

[14] W. J. Wei, Z. J. Ma, and Y. W. Chen, "Empirical study of the relationship between flow experience, perceived transaction value and impulse buying behavior," in 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2017, pp. 2246–2250.

[15] L.-Y. Leong, N. I. Jaafar, and S. Ainin, "The effects of Facebook browsing and usage intensity on impulse purchase in f-commerce," Comput. Human Behav., vol. 78, pp. 160–173, 2018.

[16] Y. Chen, Y. Lu, B. Wang, and Z. Pan, "How do product recommendations affect impulse buying? An empirical study on WeChat social commerce," Inf. Manag., vol. 56, no. 2, pp. 236–248, 2019.

[17] M. Sundström, S. Hjelm-Lidholm, and A. Radon, "Clicking the boredom away – Exploring impulse fashion buying behavior online," J. Retail. Consum. Serv., vol. 47, pp. 150–156, 2019.

[18] J. Y. Lee, R. K. Cho, and S. K. Paik, "Supply chain coordination in vendor-managed inventory systems with stockout-cost sharing under limited storage capacity," Eur. J. Oper. Res., vol. 248, no. 1, pp. 95–106, 2016.

[19] J. Cai, P. R. Tadikamalla, J. Shang, and G. Huang, "Optimal inventory decisions under vendor managed inventory: Substitution effects and replenishment tactics," Appl. Math. Model., vol. 43, pp. 611–629, 2017.

[20] Q. Bai, Y. (Yale) Gong, M. Jin, and X. Xu, "Effects of carbon emission reduction on supply chain coordination with vendor-managed deteriorating product inventory," Int. J. Prod. Econ., vol. 208, pp. 83–99, 2019.

[21] M. K. I. A. Rahim and E.-H. Aghezzaf, "Effectiveness of vendor managed inventory approach in a two-stage supply chain when demand rates are static," AIP Conf. Proc., vol. 1648, no. 1, pp. 1–4, Mar. 2015.

[22] A. Mateen and A. K. Chatterjee, "Vendor managed inventory for single-vendor multi-retailer supply chains," Decis. Support Syst., vol. 70, pp. 31–41, 2015.

[23] A. S. Saif-Eddine, M. M. El-Beheiry, and A. K. El-Kharbotly, "An improved genetic algorithm for optimizing total supply chain cost in inventory location routing problem," Ain Shams Eng. J., vol. 10, no. 1, pp. 63–76, 2019.

[24] Y. B. Park, J. S. Yoo, and H. S. Park, "A genetic algorithm for the vendor-managed inventory routing problem with lost sales," Expert Syst. Appl., vol. 53, pp. 149–159, 2016.

[25] M. Hariga, M. Gumus, and A. Daghfous, "Storage constrained vendor managed inventory models with unequal shipment frequencies," Omega (United Kingdom), vol. 48, pp. 94–106, 2014.

[26] A. Diabat, "Hybrid algorithm for a vendor managed inventory system in a two-echelon supply chain," Eur. J. Oper. Res., vol. 238, no. 1, pp. 114–121, 2014.

[27] N. K. Verma and A. K. Chatterjee, "A multiple-retailer replenishment model under VMI: Accounting for the retailer heterogeneity," Comput. Ind. Eng., vol. 104, pp. 175–187, 2017.

[28] J. Sadeghi and S. T. A. Niaki, "Two parameter tuned multi-objective evolutionary algorithms for a bi-objective vendor managed inventory model with trapezoidal fuzzy demand," Appl. Soft Comput. J., vol. 30, pp. 567–576, 2015.

[29] M. Akbari Kaasgari, D. M. Imani, and M. Mahmoodjanloo, "Optimizing a vendor managed inventory (VMI) supply chain for perishable products by considering discount: Two calibrated meta-heuristic algorithms," Comput. Ind. Eng., vol. 103, pp. 227–241, 2017.

[30] Z. Chen, "Optimization of production inventory with pricing and promotion effort for a single-vendor multi-buyer system of perishable products," Int. J. Prod. Econ., vol. 203, pp. 333–349, 2018.

[31] J. Z. Filho, F. Dias, and A. Moura, "Application of a Vendor Managed Inventory (VMI) System Model in an Animal Nutrition Industry," Procedia CIRP, vol. 67, pp. 528–533, 2018.

[32] F. Torres, C. Garcia-Diaz, and N. Rakoto-Ravalontsalama, An Evolutionary Game Theory Approach to Modeling VMI Policies, vol. 47, no. 3. IFAC, 2014.
[33] Y. Tsao, J. Lu, N. An, F. Al-khayyal, R. W. Lu, and G. Han, "Int . J . Production Economics Retailer shelf-space management with trade allowance : A Stackelberg game between retailer and manufacturers," Int. J. Prod. Econ. J., vol. 148, pp. 133–144, 2014.
[34] D. Yang, J. Jiao, Y. Ji, G. Du, P. Helo, and A. Valente, "Joint optimization for coordinated configuration of product families and supply chains by a leader-follower Stackelberg game," Eur. J. Oper. Res., vol. 246, no. 1, pp. 263–280, 2015.
[35] T. Nishi and O. Yoshida, "Optimization of Multi-period Bilevel Supply Chains under Demand Uncertainty," Procedia CIRP, vol. 41, pp. 508–513, 2016.
[36] K. Chen and T. Xiao, "Pricing and replenishment policies in a supply chain with competing retailers under different retail behaviors," Comput. Ind. Eng., vol. 103, pp. 145–157, 2017.
[37] B. Verplanken and A. Herabadi, "Individual differences in impulse buying tendency: feeling and no thinking," Eur. J. Pers., vol. 15, no. S1, pp. S71–S83, Nov. 2001.
[38] J. C. Barruezo, Gestión moderna del comercio minorista: el enfoque práctico de las tiendas de éxito, 2nd ed. Madrid: ESIC, 2003.
[39] S. Seshanna, Retail Management, 1st ed. Nueva Delhi: McGraw Hill Education (India) Private Limited, 2015.
[40] S. Miquel, Distribución comercial, 6th ed. Madrid: ESIC Editorial, 2008.
[41] V. . Bagad, Management Science, 1st ed. Technical Publications, 2008.
[42] R. D. F. S. M. Russo and R. Camanho, "Criteria in AHP: A systematic review of literature," Procedia Comput. Sci., vol. 55, no. Itqm, pp. 1123–1132, 2015.
[43] F. R. Lima-Junior and L. C. R. Carpinetti, "Quantitative models for supply chain performance evaluation: A literature review," Comput. Ind. Eng., vol. 113, no. July, pp. 333–346, 2017.
[44] A. Khaira and R. K. Dwivedi, "A State of the Art Review of Analytical Hierarchy Process," Mater. Today Proc., vol. 5, no. 2, pp. 4029–4035, 2018.
[45] T. Al-Hawari, S. Al-Bo'ol, and A. M. Momani, "Selection of Temperature Measuring Sensors Using the Analytic Hierarchy Process," Jordan J. Mech. Ind. Eng., vol. 5, no. 5, pp. 451–459, 2011.
[46] D. Dalalah, F. AL-Oqla, and M. Hayajneh, "Application of the analytic hierarchy process (AHP) in multi-criteria analysis of the selection of cranes," Jordan J. Mech. Ind. Eng., vol. 4, no. 5, pp. 567–578, 2009.
[47] T. Saaty, Decision Making for Leaders, vol. SMC-15. 1985.
[48] S. Nahmias and T. L. Olsen, Production and Operations Analysis: Seventh Edition, 7th ed. Waveland Press, 2015.

المملكة الأردنية الهاشمية

الجامعة الهاشمية

المجلة الأردنية
للهندسة الميكانيكية والصناعية

JJMIE

مجلة علمية عالمية محكمة

تصدر بدعم من صندوق البحث العلمي

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

ISSN 1995-6665

المجلة الأردنية للهندسة الميكانيكية والصناعية

المجلة الأردنية للهندسة الميكانيكية والصناعية: مجلة علمية عالمية محكمة تصدرعن الجامعة الهاشمية بالتعاون مع صندوق دعم البحث العلمي والابتكار- وزارة التعليم العالي والبحث العلمي في الأردن.

هيئة التحرير

فريـق الدعـم

ترسل البحوث إلى العنوان التالي

رئيس تحرير المجلة الأردنية للهندسة الميكانيكية والصناعية

الجامعة الهاشمية
كلية الهندسة
قسم الهندسة الميكانيكية
الزرقاء - الأردن

هاتف : 3903333 5 00962 فرعي 4147

Email: jjmie@hu.edu.jo

Website: www.jjmie.hu.edu.jo