

SCHOLARLY PUBLICATIONS

*A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT*

@DTU

(24th Edition)

DECEMBER 2014

BY: CENTRAL LIBRARY

DELHI TECHNOLOGICAL UNIVERSITY

(FORMERLY *DELHI COLLEGE OF ENGINEERING*)

GOVT. OF N.C.T. OF DELHI

SHAHBAD DAULATPUR, MAIN BAWANA ROAD

DELHI 110042

PREFACE

This is the Twenty fourth Issue of Current Awareness Bulletin started by Delhi Technological University, Central Library. The aim of the bulletin is to compile, preserve and disseminate information published by the faculty, students and alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of Delhi Technological University (DTU) as a whole to the academia.

The bulletin contains information resources available in the internet in the form of articles, reports, presentations published in international journals, websites, etc. by the faculty and students of DTU. The publications of faculty and student which are not covered in this bulletin may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose.

The learned faculty and students are requested to provide their uncovered publications to the library either through email or in CD, etc to make the bulletin more comprehensive.

This issue contains the information published during December 2014. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A multiple regression model for urban traffic noise in Delhi N. Garg¹, **S. Maji*, S. K. Mangal & P. Dhiman, Mechanical Engg, DTU
2. A systematic review of machine learning techniques for software fault prediction, **Ruchika Malhotra*, Software Engineering, DTU
3. Advanced Adaptive Particle Swarm Optimization based SVC Controller for Power System Stability based, Poonam Singhal, **S. K. Agarwal*, **Narendra Kumar*, YMCA Univ. of Science and Technology, DTU /Electronics Engg, Faridabad, Delhi, India
4. Ag-GaAs-Ag quantum nano-lenses system for plexcitonic interaction, **Venus Dillu* and **R K Sinha*, Applied Physics, DTU.
5. All Pass Network based MSO using OTRA, **Rajeshwari Pandey*, **Neeta Pandey*, **Romita Mullick*, **Sarjana Yadav* and Rashika Anurag, ECE, DTU.
6. An Empirical Study on Market Research of Organized Students' Housing Industry in India, Mayank Garg **Kunal Gupta*, Production and Industrial Engg. **Rakshit Jha*, Dept of Mathematics, DTU.
7. Analyzing Software Change in Open Source Projects using Artificial Immune System Algorithms, **Ruchika Malhotra* and **Megha Khann*, Software Engineering, DTU.
8. APPLICATION OF FUZZY BASED METHODOLOGY IN SELECTION OF VEHICLE ACCORDING TO FUEL TYPE, Kiran Pal, *@Surendra Tyagi*, Manish Jain, Computer Science, DTU
9. CHATSEP: Critical Heterogeneous Adaptive Threshold Sensitive Election Protocol for Wireless Sensor Networks, **R.K. Yadav* and **Arpan Jain*, Computer Engineering, DTU

10. Circuit Analysis Using Matlab/Simulink, **Pooja Verma*, Electrical Engineering, DTU
11. Cross Project Change Prediction Using Open Source Projects, **Ruchika Malhotra* and **Ankita Jain Bansal*, Software Engineering, DTU
12. Customizing Agile Methods using Genetic Algorithms, Rinky Dwivedi and **Daya Gupta*
13. Determining the Water Quality Index of an Urban Water Body Dal Lake, Kashmir, India @*Raj Shailesh Kanakiya*, **S.K.Singh* and J.N.Sharma, Environmental Engineering, DTU.
14. Effect of phase separation induced super cooling on magneto transport properties of epitaxial, $\text{La}_{5/8}\text{-yPr}_{\text{y}}\text{Ca}_{3/8}\text{MnO}_3$ ($\text{y} \approx 0.4$) thin film, **Sandeep Singh*, Geetanjali Sharma, P. K. Siwach, **Pawan Kumar Tyagi*, K. K. Maurya and H. K. Singh, Applied Physics, DTU
15. Proceeding Of International Conference on Mechatronics and Mechanical Engineering (ICMME 2014), (*APPLICATION OF FUZZY BASED METHODOLOGY IN SELECTION OF VEHICLE ACCORDING TO FUEL TYPE*), (page 17 onwards) Kiran Pal, @*Surendra Tyagi*, Manish Jain, Computer Science, DTU.
16. Enhanced Heuristic Approach for Travelling Tournament Problem based on Extended Species Abundance Models of Biogeography, **Daya Gupta*, **Lavika Goel* and **Ashish Chopra*, Computer Engineering, DTU
17. Environmental Management in Mass Gatherings: A Case Study of Maha Kumbh Mela 2013 at Prayag, India, **Prof. S.K. Singh* and **Anshika Bisht*, Environmental Engineering, DTU

18. Fragment based group QSAR and molecular dynamics mechanistic studies on arylthioindole derivatives targeting the α - β interfacial site of human tubulin, Chetna Tyagi, **Ankita Gupta*, Sukriti Goyal, Jaspreet Kaur Dhanjal and Abhinav Grover, Biotechnology, DTU
19. INTERACT-O-FINDER: A Tool for Prediction of DNA-Binding Proteins Using Sequence Features, **Monika Samant*, Minesh Jethva and **Yasha Hasija*, Biotechnology, DTU
20. Method Configuration Process for Agile Methods, **Daya Gupta* and **Rinky Dwivedi*, Computer Engineering Department, DTU
21. Methods for Improving Thermodynamic Performance of Vapour Compression Refrigeration Systems using Thirteen Ecofriendly Refrigerants in Primary Circuit and Tio₂ Nano Particles Mixed with R718 used in Secondary Evaporator Circuit for Reducing Global Warming and Ozone Depletion, **R.S Mishra*, Department of Mechanical Engineering, DTU
22. Micro structured-Core Photonic Crystal Fiber with All Normal Dispersion for Super continuum Generation, Than Singh Saini, Ajeet Kumar and **Ravindra Kumar Sinha*, Applied Physics, DTU
23. Modified f (R,T) gravity theory and scalar field cosmology, **Vijay Singh* and **C.P. Singh*, Applied Mathematics, DTU
24. Myconanosomes: Antibacterial, antifungal and immunomodulatory activities of silver nanoparticles synthesized from Alternaria brassicae (KF934409), Faria Fatima, Neelam Pathak, Sarika Singh, **Smita Rastogi Verma* and **Preeti Bajpai*, Biotechnology, DTU
25. Novel Hybrid Fuzzy-Intelligent Water Drops Approach for Optimal Feeder Multi Objective Reconfiguration by Considering Multiple-Distributed Generation, H. Bagheri Tolabi, M. H. Ali, and **M. Rizwan*, Department of Electrical Engineering, DTU

26. On the Applicability of Evolutionary Computation for Software Defect Prediction, **Ruchika Malhotra* and **Nakul Pritam*, Software Engineering, DTU
27. On the Applicability of Evolutionary Computation for Software Defect Prediction, **Ruchika Malhotra* and **Nakul Pritam*, Software Engineering, DTU
28. Performance analysis of ball end magnetorheological finishing using sintered magnetic abrasive based magnetorheological polishing fluid, **Mahendra Singh Niranjana* and **Sunil Jha*, Mechanical Engineering, DTU
29. Performance Optimization of Four Stage Cascade Refrigeration Systems using Energy-Exergy Analysis in the R1234ze & R1234yf in High Temperature Circuit and Ecofriendly Refrigerants in Intermediate Circuits and Ethane in the Low Temperature Circuit for Food, Pharmaceutical, Chemical Industries, **R.S. Mishra*, Mechanical Engg, DTU
30. Secure Code Assignment to Alphabets Using Modified Ant Colony Optimization along with Compression, **Hitesh Hasija* and **Rahul Katarya*, Information Technology, DTU
31. Simultaneous Reconfiguration, Optimal Placement of DSTATCOM, and Photovoltaic Array in a Distribution System Based on Fuzzy-ACO Approach, Hajar Bagheri Tolabi, Mohd Hasan Ali, and **M. Rizwan*, Electrical Engineering, DTU
32. Slow light enhanced slotted photonic crystal waveguide Preeti Rani, Yogita Kalra and **R. K. Sinha*, Applied Physics, DTU
33. Slow light generation in single-mode tellurite fibers, Than Singh Sainia, Ajeet Kumara & **Ravindra Kumar Sinha*, Applied Physics, DTU
34. Stimulated Brillouin Scattering Based Tunable Slow Light in Tellurite Photonic Crystal Fiber, Than Singh Saini, Ajeet Kumar, and **Ravindra Kumar Sinha*, Applied Physics, DTU

35. Study of Flexural, Tensile, Impact properties and Morphology of Potato Starch/ Polypropylene blends. **A.P. Gupta* and **Aftab Alam*, Applied Chemistry and Polymer Technology, DTU

36. Test Suite Optimization using Mutated Artificial Bee Colony, **Ruchika Malhotra* and Manju Khari, Software Engineering,, DTU

37. Text Categorization – A Review, **Rajni Jindal*, *@Shweta Taneja*, Dept. of Computer Engineering, DTU

38. Thermodynamic Irreversibilities Analysis of Industrial Ice Plant, **R. S. Mishra* and **S. S. Kachhwaha*, Mechanical Engineering, DTU

39. Total Quality Management in Technical Educational System, **R. S. Mishra* and **Ashok Kumar Madan*, Mechanical Engineering, DTU

40. Use of Hydrocarbons in Low Temperature Circuit in Terms of First Law and Second Law Efficiency of Four Stage Cascade Refrigeration of Semen Preservation, **R.S. Mishra*, Mechanical Engineering, DTU

***Faculty**

@Research Scholar

\$PG Scholar

#Alumni

£Undergraduate Student

~ Ex Faculty

! Teaching-cum-Research Fellow <Ex Research Scholar

>Ex PG Scholar

A multiple regression model for urban traffic noise in Delhi

N. Garg¹, S. Maji², S. K. Mangal³ & P. Dhiman⁴

Abstract – This paper reviews the strategies so far recommended for modeling road traffic noise in India. An analytical model is developed to predict road traffic noise for busy roads of Delhi, India. Equivalent continuous sound pressure level, L_{AeqT} is analyzed at different busy road locations of Delhi. A multiple linear regression analysis is conducted to predict the noise metrics L_{Aeq} , TNI and NPL in terms of traffic flow rate, percentage of heavy vehicles, and average traffic speeds. The model so developed is validated with actual experimental data.

Keywords:— Traffic noise, Equivalent continuous sound pressure level, L_{AeqT} , Traffic Noise Index, TNI.

Introduction

In Indian scenario, the vehicular population is increasing at an enormous rate as compared to the road infrastructure. In National capital region, the vehicular population is increasing at a rate of 135.6% while the road network has increased by 16.5%. The number of vehicles both cars and two wheelers are rising at a rate of 8% per annum. For a city of 17 million, Delhi has close to 8 million registered vehicles more than Mumbai, Kolkata put together [1]. Although, the introduction of MRTS (Mass rapid transit system) and BTS (Bus transit system) has proven to be a convenient means of transport, yet the noise and vibration induced due to these sources has to be controlled for their betterment and mass use by community. Thus, it is imperative to conduct studies pertaining to the increased noise levels due to heavy vehicular density for planning suitable effective measures to control it. There has been an exhaustive study carried out in developed nations in this regard. Every nation has its own validated road traffic noise model for conducting EIA studies [2]. However, there are lot of uncertainties involved in modelling traffic noise. As a result, no model can be considered as an 'ideal model'. The source noise characteristics are also varying in different countries which thus create a need for developing a specific area based or country based model rather than a harmonized approach like that in CNOSSOS approach followed in EU [3]. The present work is an attempt to focus on this critical issue and tries to use multiple linear regression approach in modelling the sound levels generated by vehicular traffic correlating the parameters viz., vehicular flow, percentage heavies and average speed of vehicles. The accuracy of model so developed is ascertained by statistical tests.

Studies done in India

The objective is to develop a regression based model for traffic noise prediction and its forecasting with definite accuracy so that it could be helpful in carrying out noise impact assessments in metropolitan cities like Delhi. It can also be helpful for planning and control especially for urban development bodies. Various models have been developed for different Indian cities in past few years. Rao *et al.* developed a regression equation for modeling L_{A10} as a function of traffic density [4]. In urban areas, most of the traffic flow is often interrupted by traffic signals and thus interrupted traffic flow conditions on urban roads create substantially different noise characteristics from highways to expressways [5 & 6]. Rajakumara *et al.* [6] developed a regression noise prediction model for both acceleration and deceleration lanes. Agarwal [7] introduced equivalent number of light and heavy vehicles for the calculation of L_{eq} values. Light motor vehicles have been analyzed to be the major culprit in noise pollution. The recent investigations of Kalaiselvi [8] introduces horn noise component into account. It has been observed that horn noise event with frequency of 16 per minute raises L_{eq} by 12 dB (A). There have been some studies that have reported a new factor i.e. tendency to blow horn in conventional Federal highway Administration (FHWA) model which could predict the equivalent noise level, L_{eq} to an accuracy of ± 3 dB(A) [9]. In Indian context, horn noise is an important issue that seriously affects the prediction accuracy of L_{AeqT} . There have been only a few studies reported on modelling the horn noise. Kalaiselvi study using the noise nomogram approach is one such example. A recent study by Paras *et al.* shows the application of ANN in modelling traffic noise for field noise monitoring data gathered in Patiala city [10]. Kumar *et al.* reported a road traffic noise prediction model using regression analysis based on Calixto model [11 & 12].

Similarly a recent study made by Sharma *et al.* (2014) models equivalent traffic noise in terms of four input variables i.e. equivalent volume, equivalent vehicle speed, distance and honking [13]. Various studies have been reported based on acoustic equivalence approach, wherein correlation have been developed based on acoustic equivalence of different vehicles with respect to light cars [14].

¹Corresponding author, Sr. Scientist, CSIR-National Physical Laboratory, New Delhi, India,

² Principal, G. B. Pant Engg. College, New Delhi – 110 020,

³Associate Professor, Deptt. of Mechanical Engineering, PEC University of Technology, Chandigarh, 160012,

⁴Graduate Student, Deptt. of Mechanical Engineering, PEC University of Technology, Chandigarh, 160012,

Table 1. Recommended empirical formulations in various Indian studies for modeling traffic noise [5-13]

Researcher	Area of study	Empirical formulation recommended
Paras <i>et al.</i> 2014	Patiala, Punjab	$L_{eq} = 67.6 + 5.8 \log Q + 0.27 P - 6.58 \log V$
Sharma <i>et al.</i> , 2014	Nagpur, Maharashtra	$L_{eq} = 59.44 + 4.461 \log Q_e + 0.1983 S_e - 0.3771d + 0.582h$ for $S_e > 30$ km/h & honking > 0 , where Q_e is the equivalent traffic flow, S_e is equivalent vehicle speed, d is the distance and h is total honking per minute.
Kalaiselvi <i>et al.</i> , 2012	Chennai	$L_{eq, 15m} = 34.31 + 13.9 \log Q_i + 9.2 \log V$ where Q_i is the equivalent traffic flow
Kumar <i>et al.</i> , 2011	NH 58 Highway	$L_{eq} = 19.92224 \log [Q(1+0.1VP)] + 12.59764$ where VP is percentage of heavy vehicles
Agarwal <i>et al.</i> , 2010	Jaipur	$L_{eq} (Lv) = 8.43 \log (x) + 39.522$ $L_{eq} (Hv) = 12.67 \log (x) + 27.579$, where x is equivalent no of vehicles
Rajakumara <i>et al.</i> , 2009	Bangalore	Acceleration Lane Model: $L_{eq} = 59.21 + 0.043 S_E + 5.71 \log Q_E - 0.0197 R$ Deceleration Lane Model: $L_{eq} = 65.12 + 0.061 S_E + 2.14 \log Q_E + 0.923 \log L - 0.041 R$ where R is position of sound level meter & L is queue of waiting vehicles on deceleration lane
Nirjar <i>et al.</i> , 2003	Delhi	$L_{eq} = 81.715 + 0.361 S_n + 2.412 \log V_n - 0.189 S_f - 0.015 \log V_f - 1.532 D_g$ where S_n is mean speed of traffic on near side of observer, S_f is mean speed of traffic on far side of observer, V_n is volume of traffic for near side, V_f is volume of traffic on far side and D_g is geometric mean of road side section.

This factor denoted by E or $PCNE$ (in some studies) is calculated as $E = 10^{\left(\frac{SPL_T - SPL_C}{10}\right)}$ wherein SPL_T is sound pressure level of heavy vehicles and SPL_C is sound pressure level of a car.

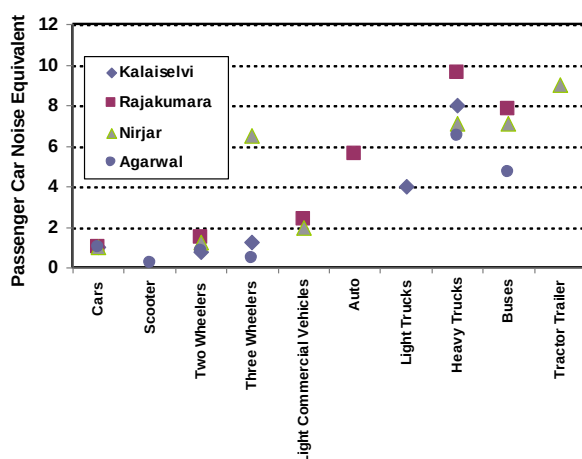


Fig 1. Passenger Car Noise Equivalent ($PCNE$) for different vehicles in various research studies carried out in India.

Fig 1 shows the value of $PCNE$ from various studies in India. A comprehensive study conducted by Rajakumara *et al.* [6] from 2.249 data sets reveals that the noise level produced by

one truck is equivalent to noise emitted by 9.63 car/jeep/vans. Similarly, noise produced by one bus, auto, LCV/MB and two wheelers is equivalent to 7.80, 5.60, 2.39 and 1.48 car/jeep/van respectively. The $PCNE$ values quantified by Nirjar *et al.* [5] for Delhi city is 1.2 for two wheelers, 2 for light commercial vehicles, 7.08 for Buses and Trucks, 6.5 for three-wheelers and 9 for tractor-trailor. The recent study reported by Sharma *et al.* [12] shows that one heavy vehicle produce 9.63 times the noise produced by medium vehicle, whereby the auto and light vehicles produce 5.60 times and 1.48 times the noise produced by a single medium vehicle (MV). It may be noted that generally the reference speed of measurement is chosen as 50 km/h. Consequently, an adjustment factor should be added to reference emission level as, $\Delta L_v = A(V_{eq}-50)$, where A is a constant and V_{eq} is number of equivalent vehicles [14 & 15]. There have been few more studies reported such as those by Kalaiselvi *et al.* [8] and Agarwal *et al.* [9]. In view of these approaches, it is imperative to conduct a revised survey for Delhi vehicles especially when noise and air pollution norms are strictly enforced and practised by Central Pollution Control Board (CPCB), India [16]. The exhaustive literature review shows that the majority of studies followed in India are based on multiple regression analysis of experimental data gathered from traffic noise measurements. This approach has been also followed in some countries. To [17] reported a multiple regression model for predicting L_{10} noise level due to road traffic. Statistical tests using the analysis of variance (ANOVA) and Tukey test reveals that the total traffic flow and the number of heavy vehicles are the most significant factors of

urban traffic noise. Li *et al.* [15] developed an integrated noise-GIS system for China based on a similar approach. Givargis *et al.* [18] work on modifying the U.K CORTN model (Calculation of Road Traffic Noise) to a model capable of calculating $L_{Aeq,1h}$ for Tehran roads is an excellent illustration of calibration of road traffic noise models for country or area based usage.

Experimental Data Acquisition & Analysis

The precision digital sound level meter used for the study was kept at a distance of 0.4 m from the body and at a height of 1.4 m from the ground level and 3-4 m from the road for avoiding any reflections from road side barriers. The measurement of sound pressure in dB (A) was done along with the monitoring of the average speed of vehicles with the speed gun (Make: Bushnell) and numbers of vehicles were counted manually. Precision digital sound level meter helped in measuring L_{eq} and statistical parameters *e.g.* L_{10} , L_{50} , L_{90} etc. Short term, L_{AeqT} measurements ranging from 15 minutes to 60 minutes were undertaken at different busy road locations of Delhi. The measurements were conducted using a Norsonic, Nor 118 sound level analyzer and B&K 2250 sound level meter. While conducting the measurements, it was ensured that there is no reflection from the adjoining building facades or wall. It may be noted here that the recent studies conducted by Maruyama *et al.* shows that a minimum 170 number of vehicles is enough for obtaining a reliable L_{AeqT} during measurement time interval T [19]. The equivalent traffic flow is calculated as:

$$Q_{eqv} = Q_{Car} + Q_{2W} \times E_{2W} + Q_{MCV} \times E_{MCV} + Q_{3W} \times E_{3W} + Q_{Bus} \times E_{Bus} + Q_{Truck} \times E_{Truck} \quad (1)$$

where Q_{Car} , Q_{2W} , Q_{MCV} , Q_{3W} , Q_{Bus} and Q_{Truck} are the total volume of car, two-wheelers, medium commercial vehicles, three-wheelers, bus and trucks respectively, while the values of E_{2W} , E_{MCV} , E_{3W} , E_{Bus} and E_{Truck} are measured as 1.13, 1.67, 2.36, 7.33 and 8.97 respectively as shown in Fig 1.

The equivalent vehicle speed is calculated as:

$$S_{eqv} = (Q_{Car} V_{Car} + Q_{2W} \times E_{2W} \times V_{2W} + Q_{MCV} \times E_{MCV} \times V_{MCV} + Q_{3W} \times E_{3W} \times V_{3W} + Q_{Bus} \times E_{Bus} \times V_{Bus} + Q_{Truck} \times E_{Truck} \times V_{Truck}) / Q_{eqv} \quad (2)$$

where V_{Car} , V_{2W} , V_{MCV} , V_{3W} , V_{Bus} and V_{Truck} are average speed of car, two-wheelers, medium commercial vehicles, three-wheelers, bus and trucks respectively.

The objective function based upon the three parameters is thus defined as:

$$L_{eq} = A + B \times \log Q_{eqv} + C \times \log V_{eqv} + D \times (V_{eqv} - 50) \quad (3)$$

where L_{eq} is the predicted equivalent continuous sound pressure level and Q_{eqv} , V_{eqv} are equivalent traffic volume and equivalent traffic speed. A , B , C and D are the constants whose value is to be computed.

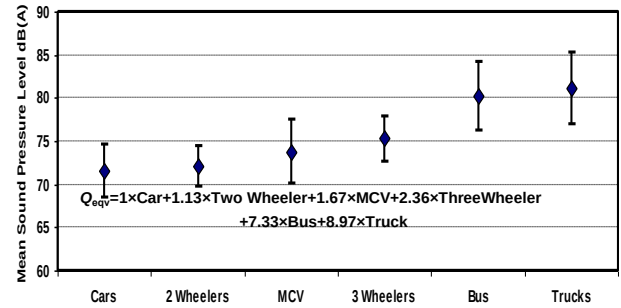


Fig. 2. Mean sound pressure level emitted by various vehicles in Delhi roads for acoustic equivalence of various vehicles in terms of Passenger Car Noise Equivalent, PCNE (or E).

Multiple Regression Analysis

Multiple linear regression analysis is an extension of simple linear regression. It is used to predict the value of a variable based on the value of two or more other variables. The variable desired to be predicted is called the dependent variable. Additionally, multiple linear regression approach also facilitates the determination of the overall fit of the model and the relative contribution of each of the predictors to the total variance explained. Standardized regression coefficients (constants) are a measure of how strongly each predictor variable influences the criterion variable. It can be generalized as:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \varepsilon \quad (4)$$

where β is coefficient of each term, k is number of independent variables and ε is error. The coefficient of multiple determination is calculated as:

$$R^2 = 1 - \frac{SS_e}{SS_T} \quad (5)$$

where SS_e is Sum squared error and SS_T is sum squared total error. R^2 indicates the proportion of the variability in the observed responses that can be attributed to changes in the predictor variables [20].

The analysis of experimental data shows that sound pressure level L_{AeqT} varied from 72 to 81 dB (A) at various locations. Predicted values were computed by the empirical formula using software tool (SPSS Downloaded version v 20). No constraints were considered for estimating parameters A , B , C and D and iteration rate is taken as 500. The non-linear regression parameters estimation is run. Table 2 represents the results of non-linear parameters estimates in L_{Aeq} determination in terms of Q_{eqv} . The final empirical

formulations are derived as equations (6) to (9) in terms of equivalent traffic flow, Q_{eqv} .

$$L_{Aeq} = 67.969 + 4.165 \log Q_{eqv} - 3.857 \log V_{eq} + 0.077(V_{eq}-50) \quad (6)$$

$$L_{10} = 71.639 + 3.627 \log Q_{eqv} - 4.176 \log V_{eq} + 0.024 (V_{eq}-50) \quad (7)$$

$$NPL = 73.454 + 5.532 \log Q_{eqv} - 6.881 \log V_{eq} + 0.069(V_{eq}-50) \quad (8)$$

$$TNI = 55.382 + 6.785 \log Q_{eqv} - 8.971 \log V_{eq} + 0.047(V_{eq}-50) \quad (9)$$

Table 2. Results of non-linear parameter estimation in L_{Aeq} determination in terms of Q_{eqv}

Parameter	Estimate	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
A	67.969	14.320	39.640	96.299
B	4.165	0.398	3.377	4.953
C	-3.857	8.707	-21.083	13.369
D	.077	0.093	-0.107	0.261

The model goodness of fit is ascertained by paired t -test [10, 21, 22]. The test statistic (t -stat) value is compared with the t -critical and in case the t -stat value is within $\pm t$ -critical value for the two tailed test, then there is no significant difference between the two samples (accept the null hypothesis).

It can be observed from table 3 that the t -stat value is less than and far away from the critical value and is within the non-rejection region, which shows that the predicted traffic noise level fits well with the experimental data at 5 % significance level. Fig 3 shows the comparison of measured and predicted noise levels for the multiple regression model in terms of Q_{eqv} . The predicted sound pressure levels are within ± 2 dB(A) as compared to the measured values for the model.

Table 3. Paired t -Test for Measured & Predicted $L_{Aeq,T}$

Parameters	Measured Values	Predicted Values
Mean	76.46	76.46
Variance	1.76	1.07
Observations	134	134
Pearson Correlation		0.74
Hypothesized Mean Difference		0
Df		133
t Stat		0.002
P(T<=t) one-tail		0.499
t Critical one-tail		1.656
P(T<=t) two-tail		0.998
t Critical two-tail		1.98

The analysis of predicted data in comparison to the measured sound pressure level shows that the difference from measured value is within ± 2 % with average difference of 0.26 % for 40 samples. It may be noted here that the coefficient of

determination, R^2 for the predicted model is 0.54 owing to the uncertainties induced due to honking noise, ground reflections, reflection from building facades, meteorological parameters, road surface corrections etc. Also the concept of acoustic equivalence may induce some uncertainties due to variable vehicular speed owing to the intermittent or free-flowing characteristics. It may be noted here that average vehicular flow calculated with $E=1$ for each vehicle also has some limitations as it is observed that the percentage heavies ranges from 1.5 % to 12.4 %. Consequently, the average speed is more influenced by light vehicles as the light vehicles are the major contributor of traffic flow, which has been also reported in previous studies [7]. However, the present model in conjunction with the sound propagation part as followed in ISO 9613-2 [23] can be used to predict the traffic noise levels with an accuracy of ± 2 dB(A).

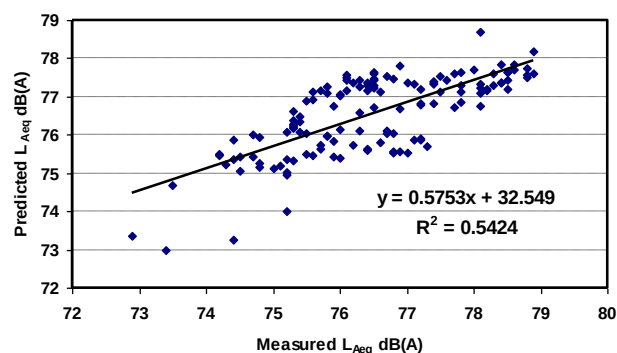


Fig 3. Predicted versus Measured $L_{Aeq,T}$ for multiple regression model in terms of Q_{eqv} .

Conclusions

A traffic noise survey is conducted on various sites in Delhi and a multiple linear regression approach is used to fit the data for development of a model. The experimental data was collected from various locations at various times during the day. The traffic density, average speed of vehicles and percentage heavies were simultaneously monitored along with the traffic noise. A multiple linear regression approach was utilized to develop a model using software tool SPSS (Downloaded free version v-20). The present model follows a generalized approach of modelling traffic noise without considering the geometrical propagation of sound waves and receiver locations, which may induce uncertainty in predictions. This model in conjunction with distance correction, road surface gradient correction, noise barrier correction, road surface condition correction, ground absorption correction can be thus a good substitute for modelling traffic noise and reliable predictions within an accuracy of ± 2 dB(A). It is imperative to further extend these studies for developing a validated model which can be very helpful in predicting the noise levels rampant at various locations and identification of hot spots so as to devise and plan suitable measures to control the traffic noise in Delhi city.

Acknowledgments

Author thanks M. Tech scholars of NIT Kurukshetra, Mr. Manish, Mr. Manoj, and Mr. Rahul for helping in noise monitoring around various sites. Authors also thank Mr. Anil Kumar, Head, Mass Standards, CSIR-NPL for his constant support and guidance. The financial grant allocation by CSIR-NPL under project NWP-45 & OLP 071132 for procurement of noise monitoring facility and softwares is acknowledged.

References

- [1] Hindustan Times, New Delhi, Where all roads lead to chaos, May 07, 2013.
- [2] Garg N and Maji S, A critical review of principal traffic noise models: Strategies and implications, Environmental Impact Assessment Review, 46 (2014) 68-81.
- [3] Kephelopoulous S., Paviotti M., F Anfosso-Lèdèe F., 2012. Common noise assessment methods in Europe (CNOSSOS-EU), 1-180.
- [4] Rao P and Rao S., Prediction of L_{AT} Traffic Noise Levels in the City of Visakhapatnam, India, Applied Acoustics, (1991) 101-110.
- [5] Nirjar R. S., Jain S. S., Parida M., Katiyar V. S. and Mittal N., A study of transport related noise pollution in Delhi, Journal of the Institute of Engineers (I)- Environment, 84, 6-15.
- [6] Rajakumara H. N. and Gowda R. M. M., Road traffic noise pollution model under interrupted traffic flow condition, Environmental Monitoring Assessments, 14, 2009, 251-257.
- [7] Agarwal S. and Swami S. L., Comprehensive approach for the development of traffic noise prediction model for Jaipur City, Environmental Monitoring Assessment, 2011, 113-120.
- [8] Kalaiselvi R. and Ramachandraiah A., A model for traffic noise prediction in heterogeneous traffic conditions, International Journal of Current Research, 4, 2012, 180-184.
- [9] Agarwal S., Swami B. L. & Gupta A. B., Development of a noise prediction model under interrupted traffic flow conditions: A case study for Jaipur city, Noise and Health, 11, (2009) 189-93.
- [10] Kumar P., Nigam S. P. and Kumar N., Vehicular traffic noise modelling using artificial neural network approach, Transportation Research Part C, 40 (2014) 111-122.
- [11] Kumar K, Katiyar V K, Parida M and Rawat K, Mathematical modeling of road traffic noise prediction, International Journal of Applied Maths and Mech, 7 (2011) 21-28.
- [12] Calixto A, Diniz F B and P H T Zannin, The statistical modelling of road traffic noise in an urban setting, Cities, 20 (2003) 23-29.
- [13] Sharma A, Bodhe G L and Schimak G, Development of a traffic noise prediction model for an urban environment, Noise Health, 16 (2014) 63-67.
- [14] Rahmani S, Mousavi S M and Kamali M J, Modeling of road-traffic noise with the use of genetic algorithm, Applied Soft Computing, 11 (2011) 1008-1013.
- [15] Bengang Li, Shu T, Dawson R W, Cao J and Lam K, A GIS based road traffic noise prediction model, Applied Acoustics 63 (2002) 679-691.
- [16] Central Pollution Control Board (CPCB), Noise limits for vehicles applicable at manufacturing stage, April, 2005. <http://cpceb.nic.in/divisions/pci2/Moise-vehicle.pdf>.
- [17] To W M, Ip R C W, Lam G C K and Yau C T H, A multiple regression model for urban traffic noise in Hong Kong, Journal of Acoustical Society of America, 112 (2002) 551-556.
- [18] Givargis Sh and Mahmoodi M, Converting the UK calculation of road traffic Noise (CORTN) to a model capable of calculating $L_{Aeq, 1h}$ for the Tehran's roads, Applied Acoustics, 69 (2008) 1108-1113.
- [19] Maruyama M, Kuno K and Sone T, The minimum measurement time for estimating L_{AeqT} of road traffic noise from the number of vehicle pass-bys, Applied Acoustics, 74, 2013, 317-324.
- [20] <http://www.public.iastate.edu/~maitra/stat501/lectures/MultivariateRegression.pdf>.
- [21] Pamanikabud P, Vivitjinda P, Noise prediction for highways in Thailand, Transportation Res. Part D, 7 (2002) 441-449.
- [22] Montgomery D C, Runger G C, Applied Statistics and Probability for Engineers, J. Wiley and Sons Inc., USA, Fifth Edition, 2011, Chapter 10, 351-394.
- [23] ISO 9613-2: Acoustics-Attenuation of sound during propagation outdoors-Part 2: General method of calculation, 1996.



Mr. Naveen Garg is working as a Senior scientist in CSIR-National Physical Laboratory, New Delhi, India. He has been working in CSIR-NPL since year 2004 in the field of maintenance of primary standards of sound pressure and vibration amplitude, sound transmission loss and absorption characteristics of acoustical materials and noise and vibration abatement and control and played a leading role in getting the NPL, India Calibration and Measurement capabilities (CMCs) for parameters 'sound' and 'vibration' published in Key comparison data base (BIPM) website. He has been involved in various consultancy projects pertaining to noise and vibration abatement and control. He has also spent some time in PTB, Germany, NIMT, Thailand, KIM-LIPI, Indonesia for getting training in acoustic and vibration metrology and has published many papers in national and international journals.



Prof. Dr. Sagar Maji is Principal of G. B. Pant Engg. College, New Delhi. He is B. Tech, M.E and PhD in Mechanical Engg. Earlier he had been working as a Professor in Mechanical Engg, Deptt. in Delhi Technological University (DTU), Delhi and also served as Dean, IRD in DTU. He is an eminent academician and researcher and has published many papers in national and international journals and also guided several PhDs. He is active member of many academic bodies in India and has provided consultancies in field of performance and emission characteristics of IC engines.



Sanjay Kumar Mangal is working as Associate Professor in Mechanical Engg. Department, PEC University of Technology Chandigarh. He has received his B.E. degree in Production Engg. from Punjab Engineering College, Chandigarh in 1988 and M. E. in Mechanical Engg. from IIT, Roorkee in 1990. He has obtained his Doctoral of Philosophy in Mech. Engg. from I.I.T. Kanpur in 2000. His field of interest is FEM and vibration control. He has 22 years of teaching and research experience and published various papers in international and national journals. He has guided more than 15 M. Tech thesis. He is a life member of Indian society for technical education (I.S.T.E.), India and Associate member of Institution of Engineers (A.M.I.E) Kolkata, India.



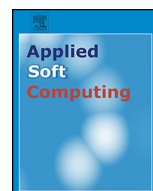
Puneet Dhiman is currently pursuing his ME in Mechanical Engineering from PEC University of Technology, Chandigarh. He has received his B. Tech degree in Mechanical Engineering from NIT Hamirpur in 2012. His academic interest areas includes Mathematical Modelling and Optimization, Manufacturing Processes and Artificial Neural Networks.



Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc



A systematic review of machine learning techniques for software fault prediction

Ruchika Malhotra

Department of Software Engineering, Delhi Technological University, Bawana Road, Delhi, India

ARTICLE INFO

Article history:

Received 31 December 2013
Received in revised form 8 October 2014
Accepted 22 November 2014
Available online xxx

Keywords:

Machine learning
Software fault proneness
Systematic literature review

ABSTRACT

Background: Software fault prediction is the process of developing models that can be used by the software practitioners in the early phases of software development life cycle for detecting faulty constructs such as modules or classes. There are various machine learning techniques used in the past for predicting faults. **Method:** In this study we perform a systematic review of studies from January 1991 to October 2013 in the literature that use the machine learning techniques for software fault prediction. We assess the performance capability of the machine learning techniques in existing research for software fault prediction. We also compare the performance of the machine learning techniques with the statistical techniques and other machine learning techniques. Further the strengths and weaknesses of machine learning techniques are summarized.

Results: In this paper we have identified 64 primary studies and seven categories of the machine learning techniques. The results prove the prediction capability of the machine learning techniques for classifying module/class as fault prone or not fault prone. The models using the machine learning techniques for estimating software fault proneness outperform the traditional statistical models.

Conclusion: Based on the results obtained from the systematic review, we conclude that the machine learning techniques have the ability for predicting software fault proneness and can be used by software practitioners and researchers. However, the application of the machine learning techniques in software fault prediction is still limited and more number of studies should be carried out in order to obtain well formed and generalizable results. We provide future guidelines to practitioners and researchers based on the results obtained in this work.

© 2014 Published by Elsevier B.V.

1. Introduction

The growing complexity and dependency of the software has increased the need for delivering high quality, maintainable software at lower costs. Software fault prediction is very important and essential activity in order to improve the software quality and reduce the maintenance effort before the system is deployed [1]. Early detection of faults may lead to timely correction of these faults and delivery of maintainable software. There are various software metrics available in the literature. These software metrics and fault data can be used to construct models that can be used for predicting faulty modules/classes in the early phases of software development life cycle.

Software fault prediction (SFP) can be done by classifying the modules/classes as fault prone and not fault prone. The software metrics and fault data obtained from a similar project or previous

release can be used to construct SFP models. The model constructed can be subsequently applied to the modules/classes of current software projects for classifying them as fault prone or not fault prone. The software practitioners can then target the available testing resources on the fault prone areas of the software in the early phases of software development. For example, if only 30% of testing resources are available, the knowledge of the weaker areas will help the testers in focusing the available resources on fixing the classes/modules that are more prone to faults. Hence, a low cost, high quality and maintainable software can be produced in the given time and budget.

Soft computing techniques are well suited for real life problems that use methods to extract useful information from complex and intractable problems in less time. They are tolerant to data that is imprecise, partially incorrect or uncertain. One of the important components of the soft computing techniques includes the machine learning techniques. The machine learning (ML) techniques have been used in the literature in order to predict models for estimating the faulty modules/classes. For example, Singh et al. [2] have

E-mail address: ruchikamalhotra2004@yahoo.com

<http://dx.doi.org/10.1016/j.asoc.2014.11.023>
1568-4946/© 2014 Published by Elsevier B.V.

employed decision trees and artificial neural networks in order to predict classes at various severity levels of faults. There are several advantages of using these classification techniques for predicting faulty modules/classes.

To facilitate the use of ML techniques in software fault prediction it is necessary to systematically summarize the empirical evidence obtained on these techniques from the existing literature and studies. The existing literature reviews aim to answer various research questions [3–5], but to the best of the authors knowledge there is no systematic review that focus on the ML techniques for the software fault prediction i.e. whether a module/class is faulty or not. This paper extensively reviews all the studies between the period of 1991 and 2013. In order to perform the review we have identified seven categories of the ML techniques. The aim of this systematic literature review is to summarize, analyze and assess the empirical evidence regarding: (1) ML techniques for SFP models (2) performance accuracy and capability of ML techniques for constructing SFP models (3) comparison between the ML and statistical techniques (4) comparison of performance accuracy of different ML techniques (5) summarize the strength and weakness of the ML techniques. We further provide future guidelines to software practitioners and researchers regarding the application of the ML techniques in SFP. In order to achieve this aim we extensively searched through seven digital libraries and identified 64 studies to answer the research questions pertaining to the use of the ML techniques for the SFP. The primary studies were selected according to the quality assessment of the studies and relevance.

The rest of the paper is organized as follows: Section 2 presents the research questions that are addressed in this systematic review and the research criteria followed in this study for selection of primary studies. Section 3 presents the answers to the research questions identified in this work. Section 4 provides the limitation of this work and Section 5 provides conclusions and future directions obtained from this systematic review.

2. Method

The planning, conducting and reporting of the systematic review carried out in this paper is done by following the procedure given by [6]. The process is depicted in Fig. 1. In the planning stage we developed the review protocol that included the following steps: research questions identification, search strategy design, study selection criteria, study quality assessment, data extraction process and data synthesis process. After formation of the review protocol a series of steps were carried out in the review. In the first step we formed the research questions that addressed the issues to be answered in the systematic literature review (SLR). In the second step, we described the search strategy including identification of search terms and selection of sources to be searched in order to identify the primary studies. The third step involves determination of relevant studies based on the research questions. This step also determines the inclusion and exclusion criteria for each primary study. In the next step, we identify quality assessment criteria by forming the quality assessment questionnaire in order to analyze and assess the studies. The second last step involves the design of data extraction forms to collect the required information in order to answer the research questions and in the last final step we devise methods for data synthesis.

Development of review protocol is an important step in a SLR as it reduces the possibility and risk of research bias in the SLR. We established the review protocol in this work by frequently holding meetings and group discussions in the group formed comprising of senior Assistant Professors. In the following subsections we describe the research questions and the steps followed while conducting the SLR.

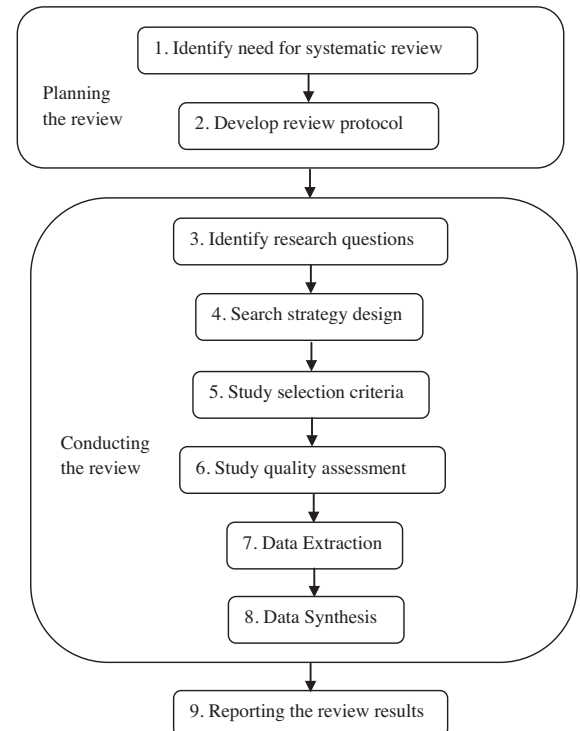


Fig. 1. Systematic review process.

2.1. Research questions

The aim of this SLR is to provide and assess the empirical evidence obtained from the studies using the ML techniques for SFP in the literature. Table 1 presents six research questions addressed in this SLR. From the primary studies we identified the ML techniques used for predicting software faults (RQ1). We analyzed these studies using the ML techniques for SFP to answer RQ2, RQ2.1, RQ2.2, RQ2.3, RQ2.4, RQ2.5, RQ2.6 and RQ2.7. In the next research question, we analyzed the performance of the ML techniques for SFP (RQ3). This question focuses on the values of the performance measures for SFP. RQ4 compares the performance of ML techniques with statistical techniques. The aim of this question is to determine whether ML techniques are better than the statistical techniques. RQ5 assesses which ML techniques outperformed the other ML techniques in order to determine which ML technique is consistently better than the other ML techniques. The final question (RQ6) determines the strengths and weakness of different ML techniques to provide software practitioners guidance regarding the selection of an appropriate ML technique.

2.2. Search strategy and study selection

We formed the sophisticated search terms by incorporating alternative terms and synonyms using Boolean expression 'OR' and combining main search terms using 'AND'. The following general search terms were used for identification of primary studies:

Software AND (fault OR defect OR error) AND (proneness OR prone OR prediction OR probability) AND (regression OR machine learning OR soft computing OR data mining OR classification OR Bayesian network OR neural network OR decision tree OR support vector machine OR genetic algorithms OR random forest)

The terms on ML were derived from textbooks and research papers on ML [7,8]. After identifying the search terms, the relevant and important digital portals were selected. The selection was not restricted by the availability of digital portals at the home

Table 1
Research questions.

RQ#	Research questions	Motivation
RQ1	Which ML techniques have been used for SFP?	Identify the ML techniques commonly being used in SFP.
RQ2	What kind of empirical validation for predicting faults is found using the ML techniques found in RQ1?	Assess the empirical evidence obtained.
RQ2.1	Which techniques are used for feature reduction for SFP?	Identify techniques reported to be appropriate for selecting and extracting relevant features.
RQ2.2	Which metrics are commonly used for SFP?	Identify metrics commonly used for SFP.
RQ2.3	Which metrics are found useful for SFP?	Identify metrics reported to be appropriate for SFP.
RQ2.4	Which metrics are found not useful for SFP?	Identify metrics reported to be inappropriate for SFP.
RQ2.5	Which data sets are used for SFP?	Identify data sets reported to be appropriate for SFP.
RQ2.6	What is the fault distribution in the data sets used for SFP?	Identify the fault distribution in the data sets used for SFP.
RQ2.7	Which performance measures are used for SFP?	Assess the performance of the ML techniques for SFP.
RQ3	What is the overall performance of the ML techniques for SFP?	Investigate the performance of the ML techniques for SFP.
RQ4	Whether the performance of the ML techniques better than the statistical techniques?	Investigate the performance of the ML techniques over statistical techniques for SFP.
RQ5	Are there any ML techniques that significantly outperform other ML techniques?	Investigate the performance of the ML techniques over the other ML techniques for SFP.
RQ6	What are the strengths and weaknesses of the ML techniques?	Determine the information about ML techniques.

universities. The following seven electronic databases were used for searching the primary studies:

- (1) IEEE Xplore
- (2) ScienceDirect
- (3) ACM Digital Library
- (4) Wiley Online Library
- (5) Google scholar
- (6) SpringerLink
- (7) Web of Science

The search was made on the above seven electronic databases using the target search string. In addition to searching the above electronic databases the relevant journals and conferences listed on <http://web.engr.illinois.edu/taoxie/seconferences.htm> were searched.

We restricted the search from 1991 to 2013 as the machine learning techniques were launched in the 1990s. After determining which electronic databases to search, an initial search to identify the candidate primary studies was performed. After performing an initial search, the relevant studies were determined by obtaining the full text papers following the inclusion and exclusion criteria described in the next section. We also included those studies that were found important from the reference section of the relevant studies.

We included the empirical studies using the ML techniques for SFP in the SLR. We identified 96 primary studies for inclusion in the SLR. After scanning the reference lists of the relevant study further 26 studies were included. Hence, total number of 122 studies [1,2,20–139] were identified for further processing and analysis. The candidate studies were selected after following the inclusion and exclusion criteria given below:

Inclusion criteria:

- Empirical studies using the ML techniques for SFP.
- Empirical studies combining the ML and non ML techniques.
- Empirical studies comparing the ML and statistical techniques.

Exclusion criteria:

- Studies without empirical analysis or results of use of the ML techniques for SFP.
- Studies based on dependent variable other than fault proneness.
- Studies using the ML techniques in context other than SFP.

- Similar studies i.e. studies by same author in conference as well extended version in journal. However, if the results were different in both the studies they were retained.
- Review studies

The above inclusion and exclusion criteria were tested by two researchers independently and they reached a common decision after detailed discussion. In case of any doubt, full text of the study was reviewed and a final decision regarding the inclusion/exclusion was made. In addition, the quality of the studies was determined by their relevance to the research questions. The similar studies with the same results by a particular author were removed.

By applying the above selection criteria, 79 studies were selected. Finally, the quality assessment criteria given in the following section was used to obtain the final studies.

2.3. Quality assessment criteria

We formed a quality questionnaire for assessing the relevance and strength of the primary studies. The quality questionnaire was developed by considering the suggestions given by Wen et al. [9]. The quality assessment is used for weighing the studies. Table 2 presents the quality assessment questions. The questions are ranked 1 (yes), 0.5 (partly) and 0 (no). The final score is obtained after adding the values assigned to each question. A study could have maximum score of 15 and minimum score of 0.

Two independent researchers ranked the quality questions for each primary question and consulted other researchers in case of any disagreement. Finally after thorough reviews, discussions and brain storming sessions a final decision about the inclusion/exclusion for each study was made.

2.4. Data extraction and data synthesis

We filled a form for each of the selected primary study (step 7). The purpose of using data extraction form is to determine which research question was satisfied by a primary study. We summarized author name, title, publishing details, data set details, independent variables (metrics) and the ML techniques. The details of which specific research questions answered by each primary study were present in the data extraction card. These data extraction cards were used to collect information from the primary studies. Two independent researchers collected the information required for each primary study from the data extraction card. The two researchers then matched their results and if there was any disagreement among the two, other researchers were consulted to

Table 2
Quality assessment questions.

Q#	Quality questions	Yes	Partly	No
Q1	Are the aims of the research clearly stated?			
Q2	Are the independent variables clearly defined?			
Q3	Is the data set size appropriate?			
Q4	Is the data collection procedure clearly defined?			
Q5	Are metrics sub selected?			
Q6	Are the ML techniques justified?			
Q7	Are the ML techniques clearly defined?			
Q8	Are the performance measures used to assess the SFP models clearly defined?			
Q9	Are the results and findings clearly stated?			
Q10	Are the limitations of the study specified?			
Q11	Is the research methodology repeatable?			
Q12	Is there any comparative analysis conducted (statistical vs ML)?			
Q13	Is there any comparative analysis conducted (ML vs ML)?			
Q14	Does the study contribute/add to the literature?			
Q15	Does the study have adequate number of the average citation count per year?			

resolve these disagreements. The resultant data was saved into a excel file for further use during the data synthesis process.

The basic objective while synthesizing data is to accumulate and combine facts and figures from the selected primary studies in order to formulate a response and resolve the research questions [9]. Collection of a number of studies which state similar and comparable viewpoints and results helps in providing research evidence for obtaining conclusive answers to the research questions. We scrutinized and evaluated both the quantitative data which includes values of various performance metrics like AUC, prediction accuracy and qualitative data which includes strengths and weaknesses of the ML methods, categorization of various ML methods, feature sub-selection methods and data sets used. We utilize a number of techniques to synthesize data collected from our primary studies. In order to answer the research questions we used visualization techniques such as line graph, box plots, pie charts and bar charts. We also used tables for summarizing and presenting the results.

3. Results and discussion

This section presents the results obtained from primary studies selected for SFP. First, we present the overview of each primary study along with their corresponding ranking based on the scores assigned to the quality assessment questions. Second, the answers to the review research questions are provided in each sub section. Then we provide the discussion and interpretation of the results in the view of the facts obtained from each primary study.

3.1. Description of primary studies

In this section, we provide the brief description of the selected primary studies. We have selected 64 primary studies, out of 79 studies (based on the quality questions), which use the ML techniques for SFP. The studies were mostly based on public or open source data sets.

Table 3
Summary of top publications.

Publication name	Type	Number	Percent
IEEE Transactions on Software Engineering	Journal	9	16
Journal of Systems and Software	Journal	6	10
Empirical Software Engineering	Journal	6	10
Information and Software Technology	Journal	3	5
IEEE International Symposium on Software Reliability (ISSRE)	Conference	3	5
International Conference on Predictor Models in Software Engineering (PROMISE)	Conference	2	4
Software Quality Journal	Journal	2	4
Automated Software Engineering	Journal	2	4
Information Sciences	Journal	2	4
Expert Systems with Applications	Journal	1	2
Engineering Applications of Artificial Intelligence	Journal	1	2
Product Focus Software Process Improvement	Journal	1	2
IEEE Software	Journal	1	2
International Journal of Systems Assurance Engineering and Management	Journal	1	2

3.1.1. Publication source

Table 3 summarizes the details of the publications in top journals and conferences along with the number and percentage of primary studies in the corresponding journal. The major publications were in IEEE Transactions on Software Engineering, Information and Software Technology, Journal of Systems and Software, Software Quality Journal, Empirical Software Engineering, Information Sciences and so on. The statistics show that 56% of the primary studies were present in journals and 44% of the primary studies were present in conferences. Hence, the number of studies published in journals (36) was slightly more than published in conferences (28). The top 14 sources consisted of 72% of the primary studies. Table 3 shows that 16% of studies were published in IEEE Transactions on Software Engineering and 10% each in Information and Software Technology and Empirical Software Engineering.

3.1.2. Quality assessment questions

The scores were assigned to the quality assessment questions. We divided the scores into various categories very high ($13.5 \leq \text{score} \leq 15$), high ($9.5 \leq \text{score} \leq 13$), medium ($5 \leq \text{score} \leq 9$) and low ($0 \leq \text{score} \leq 4.5$). The maximum score that could be assigned to a study was 15 and minimum 0.

Table 4 provides a unique identifier to each selected primary study along with the reference. These unique identifiers will be used in all subsequent sections to refer to their corresponding selected primary study. We also wanted to assess the studies based on the visibility they have gained in the community. However, recognizing that the recent papers published in 2013 may not have yet obtained high citations, we assessed three papers published in 2013 (SF62–SF64) excluding quality question Q15. Studies SF12, SF4, SF23, were the top three rankers in terms of average citations count per year. It is also worth mentioning that all of these papers have been published in the IEEE Transactions on Software Engineering. Five studies (SF10, SF23, SF27, SF45, SF64) obtained maximum score. Hence, the interested readers may take these publications into consideration for further reading. Studies SF7, SF9, SF20, SF23, SF25, SF29, SF32, SF36, SF37, SF44, SF45, SF48, SF49, SF57, SF59, SF60, SF61, SF63, SF64 that compared the performance of the ML techniques with statistical techniques were selected in the SLR. The studies with score less than 9.5 were excluded from the analysis. Hence, after quality assessment further 15 studies

Table 4
Selected primary studies.

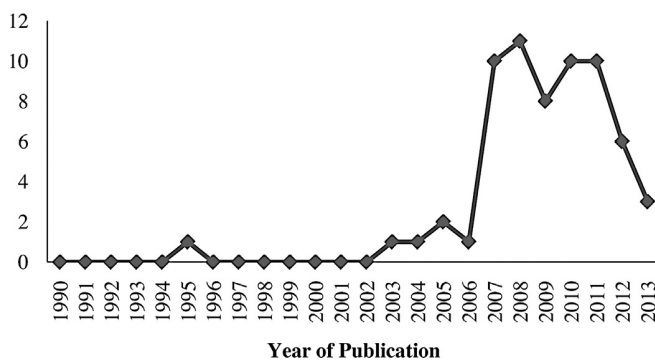
Study no.	Paper	Ref no.	Study no.	Paper	Ref no.
SF1	Sherer (1995)	[20]	SF33	Singh (2009b)	[51]
SF2	Guo (2003)	[21]	SF34	Tosun (2009)	[52]
SF3	Guo (2004)	[22]	SF35	Zimmermann (2009)	[53]
SF4	Gyimothy (2005)	[23]	SF36	Afzal (2010)	[54]
SF5	Koru (2005)	[24]	SF37	Arisholm (2010)	[55]
SF6	Zhou (2006)	[25]	SF38	Carvalho (2010)	[56]
SF7	Arisholm (2007)	[26]	SF39	Liu (2010)	[57]
SF8	Catal (2007)	[27]	SF40	Malhotra (2010)	[58]
SF9	Jiang (2007)	[28]	SF41	Ostrand (2010)	[59]
SF10	Kanmani (2007)	[29]	SF42	Pendharkar (2010)	[60]
SF11	Li (2007)	[30]	SF43	Seliya (2010)	[1]
SF12	Menzies (2007)	[31]	SF44	Singh (2010)	[61]
SF13	Ma (2007)	[32]	SF45	Zhou (2010)	[62]
SF14	Pai (2007)	[33]	SF46	Azar (2011)	[63]
SF15	Turhan (2007)	[34]	SF47	Diri (2011)	[64]
SF16	Turhan (2007a)	[35]	SF48	Malhotra (2011)	[65]
SF17	Carvalho (2008)	[36]	SF49	Martino (2011)	[66]
SF18	Elish (2008)	[37]	SF50	Mishra (2011)	[67]
SF19	Gondra (2008)	[38]	SF51	Misirh (2011)	[68]
SF20	Jiang (2008)	[39]	SF52	Ricca (2011)	[69]
SF21	Kaur (2008)	[40]	SF53	Rodriguez (2011)	[70]
SF22	Kim (2008)	[41]	SF54	Song (2011)	[71]
SF23	Lessmann (2008)	[42]	SF55	Twala (2011)	[72]
SF24	Menzies (2008)	[43]	SF56	Chen (2012)	[73]
SF25	Moser (2008)	[44]	SF57	Malhotra (2012)	[74]
SF26	Turhan (2008)	[45]	SF58	Okutan (2012)	[75]
SF27	Vandecruys (2008)	[46]	SF59	Yang (2012)	[76]
SF28	Bener (2009)	[47]	SF60	Yu (2012)	[77]
SF29	Catal (2009)	[48]	SF61	Zhou (2012)	[78]
SF30	Menzies (2009)	[49]	SF62	Cahill (2013)	[79]
SF31	Singh (2009)	[50]	SF63	Chen (2013)	[80]
SF32	Singh (2009a)	[2]	SF64	Dejaeger (2013)	[81]

[96,100–102,108,118–121,125,127,132–135] were removed from the analysis.

The details of the studies addressing each specific research question are provided in Table A.1 in [Appendix](#).

3.1.3. Publication year

[Fig. 2](#) presents the distribution of the studies from the year 1991 to 2013. [Fig. 2](#) shows that before 2007 out of 64 only 6 studies examined the use of the ML techniques for SFP. Most of the studies were conducted from 2007 onwards. The number of studies in the years 2007, 2008, 2009, 2010 and 2011 were 10, 11, 8, 10 and 10, respectively accounting about 90% of the studies between 2007 and 2013. For year 2013, data is not complete as the search was started in October 2013 and only few studies were online till then. The availability of the public domain data sets from the promise repository from 2005 made a significant impact on the increase in the studies. Thus, since year 2007 the studies in the literature have started examining the effectiveness of the ML techniques for SFP. This confirms the inclusion of recent and relevant studies in this

**Fig. 2.** Year-wise distribution of studies.

SLR. This also shows that the studies have recently started using the ML techniques for SFP and this research domain is relatively new.

3.1.4. RQ1: which ML techniques have been used for SFP?

This section provides the details of the ML techniques used in the primary studies selected in this SLR. Based on the selected studies we categorized the ML techniques used for defect prediction as follows:

- Decision trees (DT)
- Bayesian learners (BL)
- Ensemble learners (EL)
- Neural networks (NN)
- Support vector machines (SVM)
- Rule based learning (RBL)
- Evolutionary algorithms (EA)
- Miscellaneous

[Fig. 3](#) presents the classification taxonomy of ML techniques for SFP used in this SLR.

[Table 5](#) presents the number and percentage of the studies with respect to the ML techniques. Among the above listed categories of

Table 5
Distribution of studies across ML techniques based on classification.

Method	# of Studies	Percent
DT	31	47.7
NN	17	26.16
SVM	18	27.7
BL	31	47.7
EL	12	18.47
EA	8	12.31
RBL	5	7.7
Misc.	16	24.62

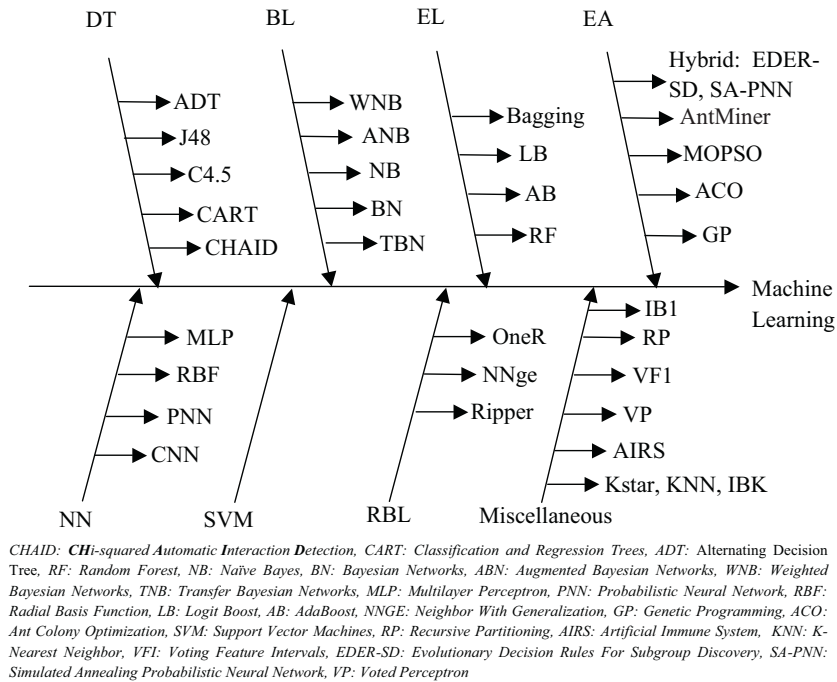


Fig. 3. Classification of ML Techniques for SFP. CHAID: Chi-squared automatic interaction detection, CART: classification and regression trees, ADT: alternating decision tree, RF: random forest, NB: naïve Bayes, BN: Bayesian networks, ABN: augmented Bayesian networks, WNB: weighted Bayesian networks, TBN: transfer Bayesian networks, MLP: multilayer perceptron, PNN: probabilistic neural network, RBF: radial basis function, LB: logit boost, AB: AdaBoost, NNGE: neighbour with generalization, GP: genetic programming, ACO: ant colony optimization, SVM: support vector machines, RP: recursive partitioning, AIRS: artificial immune system, KNN: K-nearest neighbour, VF1: voting feature intervals, EDER-SD: evolutionary decision rules for subgroup discovery, SA-PNN: Simulated annealing probabilistic neural network, VP: Voted perceptron

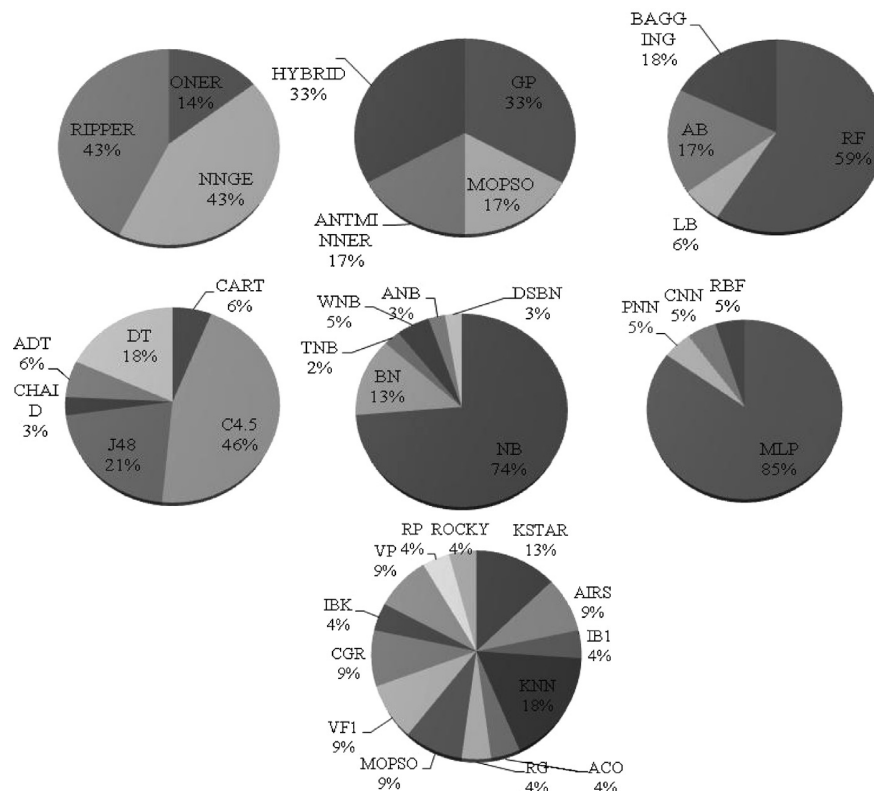


Fig. 4. Distribution of sub categories in (a) DT, (b) BL, (c) NN, (d) RL, (e) EA, (f) EL, (g) Misc.

the ML techniques the most frequently used techniques were from categories DT, BL, SVM and NN analyzed in 47.7%, 47.7%, 27.7%, 26.16% studies, respectively. These ML techniques were used for SFP. The primary studies consisted of two types of studies: (1) studies comparing statistical and ML techniques (2) studies comparing ML techniques. The distribution of the ML techniques for SFP with respect to each identified category is presented in Fig. 4. The most frequently used the ML techniques for SFP were C4.5 in DT category (46%), NB in BL category (74%), MLP in NN category (85%) and RF in EL category (59%). The top five techniques included C4.5, NB, MLP, SVM and RF which were assessed in 15, 28, 17, 19, 10 studies, respectively.

3.2. RQ2: what kind of empirical validation for predicting faults is found using ML techniques found in RQ1?

This section determines the techniques used for metrics reduction, commonly used metrics, useful and not useful metrics, nature of data sets used, and the performance measures used in the selected primary studies.

3.2.1. Which techniques are used for feature reduction for SFP using ML techniques?

Over the years, a number of techniques have been used in order to reduce the dimensionality of features and develop models with better results. There are two basic kinds of techniques for feature reduction: feature extraction and feature selection. While feature selection helps in selection of the most relevant features, feature extraction involves combination of a set of relevant features into a new smaller feature set. However, only 49% of the selected primary studies used a feature reduction technique to aid its model development.

The most commonly used feature selection technique in the selected studies was co-relation based feature selection (CFS) [10] (used by SF18, SF21, SF37, SF38, SF48, SF62, SF53, SF55, SF58). The CFS technique eliminates noisy and redundant features and retains only those features, which are highly co-related with the fault proneness attribute. Apart from CFS, some other commonly used feature selecting techniques were infogain method (used by SF12, SF15, SF31, SF50, SF51), χ^2 based filter (used by SF27, SF50), correlation analysis (used by SF14) and wrapper attribute selection (used by SF53, SF54, SF62). Some less commonly used feature selection techniques include Markov Blanket feature selection (used by SF64), consistency based selection (used by SF53, SF62), chi-square feature selection (SF22) and random forests (used by SF3). While majority of the studies used a feature selection technique, few studies used principal component analysis for feature extraction (SF10, SF15, SF19, SF28, SF44). Another technique called Isomap was also used for feature extraction by SF15.

3.2.2. Which metrics are commonly used in SFP?

In order to predict fault proneness, the primary studies use software metrics as independent variables. There are a number of metrics used in software engineering domain to quantify the characteristics of a software. We define the categorization of the selected primary studies based on the type of metrics used by these studies for predicting fault proneness as follows. Similar categorization is given in Radjenovic et al. [3].

- **Studies using procedural metrics:** These studies use metrics that include the traditional static code metrics which are defined by Halstead [11] and McCabe [12] along with the size metrics such as the LOC (lines of code) metric. Studies which are included in this category are SF2, SF3, SF11, SF12, SF13, SF15, SF16, SF18, SF19, SF20, SF22, SF23, SF24, SF25, SF27, SF28, SF29, SF30, SF31, SF39,

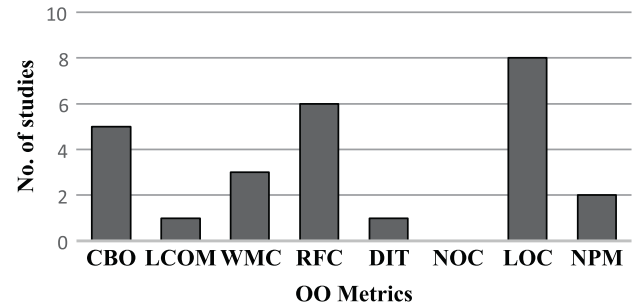


Fig. 5. Useful OO metrics for SFP.

SF42, SF43, SF44, SF47, SF50, SF51, SF53, SF54, SF55, SF56, SF60, SF62, SF64.

- **Studies using object-oriented metrics:** These studies use metrics that measure various attributes of object-oriented (OO) software like cohesion, coupling and inheritance for an OO class [13–15]. Though, LOC is a traditional metric, many studies also include LOC or SLOC (source lines of code) while using OO metrics. For the purpose of simplicity, we include size metrics in both the categories (procedural as well as OO). Studies which are included in this category are SF4, SF5, SF6, SF7, SF8, SF10, SF14, SF17, SF21, SF32, SF33, SF40, SF45, SF46, SF48, SF49, SF57, SF58.
- **Studies using hybrid metrics:** Some studies (SF26, SF38, SF47, SF59, SF61) used both OO metrics as well as procedural metrics for fault proneness prediction.
- **Miscellaneous metrics:** These studies include metrics such as requirement metrics (SF9), change metrics (SF25), network metrics extracted from dependency graph (SF34), churn metrics (SF35), fault slip through metrics (SF36), process metrics (SF37), file age, size, changes and faults in previous versions metrics (SF41), elementary design evolution metrics (SF52), and other miscellaneous metrics which cannot be grouped as either procedural or OO metrics.

The procedural metrics are most commonly used metrics (51%) in the selected primary studies. A number of OO metrics have also been used for SFP and the most commonly used metrics suite was defined by Chidamber and Kemerer [13] which was used in all the 28% studies which use OO metrics.

3.2.3. Which metrics are found useful for SFP?

Certain studies (SF7, SF8, SF14, SF18, SF21, SF27, SF37 SF48 etc.) report OO metrics which are highly correlated to fault proneness. Fig. 5 reports the OO metrics which are found useful for SFP. According to the figure, CBO (coupling between objects), RFC (response for a class) and LOC are highly useful metrics for SFP using feature selection methods. However, studies using feature extraction techniques have not clearly specified useful metrics for SFP.

As far as the usefulness of procedural metrics is concerned, the studies do not yield a conclusive result as the primary studies do not compare the usefulness of specific procedural metrics. Only SF38 comments about unique operand and EV(g) as useful indicators for SFP.

3.2.4. Which metrics are not found useful for SFP?

While CBO, RFC and LOC are good predictors for fault proneness, there are certain metrics which were not useful for SFP. The results obtained from the primary selected studies indicate that the NOC (number of children) and DIT (depth of inheritance tree) as not useful metrics for SFP. Studies which support this are SF8, SF14 and SF58. The primary studies do not give any conclusive results for procedural metrics which are not useful for SFP as these studies have not stated and analyzed usefulness of specific procedural metrics.

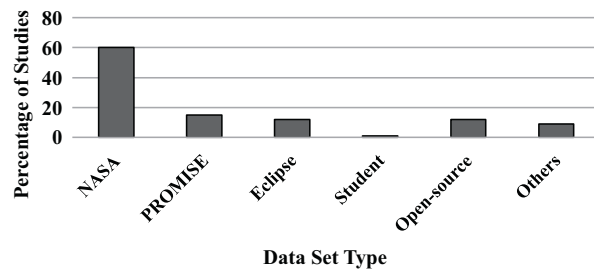


Fig. 6. Data sets for SFP.

3.2.5. Which data sets are used for SFP?

A variety of data sets have been used in the SFP studies. Fig. 6 depicts the percentage of studies using different data sets. The data sets used can be publically available or private in nature. Public data sets are freely available while private data sets are not shared by researchers and thus results on such data sets cannot be verified and such studies are not repeatable. The major data sets used and their categorization are as follows:

- **NASA data sets:** These data sets are publically available in the NASA repository by NASA Metrics Data Programme and are the most commonly used data sets for SFP. They are used in 60% of the selected primary studies.
- **PROMISE repository data sets:** These data are freely publically available in the PROMISE repository (<https://code.google.com/p/promisedata/>) and include AR data sets, Jedit software data sets amongst other data sets. They are used in 15% of the selected primary studies.
- **Eclipse data set:** Around 12% of our selected primary studies used eclipse data set, which is an open source project whose defect data set is publically available. This project is an integrated development environment developed in Java language.
- **Open source project data sets:** This categorization involves studies which use other open source projects such as Lucene, Xylan, Ant, Apache, POI, etc. amongst others used by about 12% of selected primary studies.
- **Student data set:** This is an academic software developed by students and is used by SF10.
- **Others:** This category includes some industrial and private data sets like that of a commercial bank, commercial Java application, telecommunication company data set etc. They are used by about 9% of selected primary studies.

There are very few studies (about 6%) that use industrial datasets to examine the effectiveness of the ML techniques for SFP.

3.2.6. What is the fault distribution in the data sets used for SFP?

The fault distribution of the consistently used data sets used in the selected primary studies is shown in Fig. 7. In Fig. 7, the

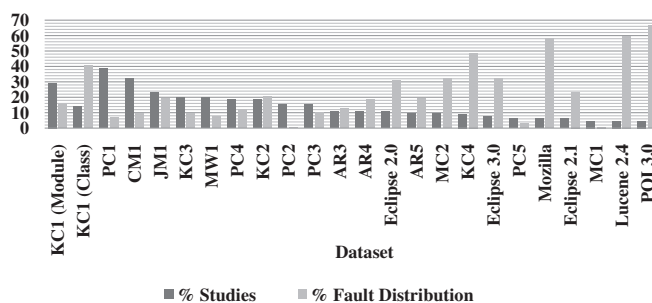


Fig. 7. Distribution of faults in studies.

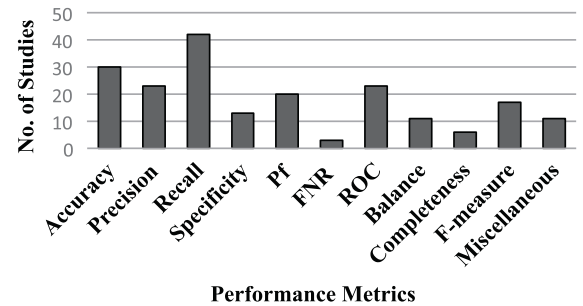


Fig. 8. Studies using different performance measures for SFP.

percentage of faulty classes/modules in the widely used data sets (used at least in 3% of studies) is shown. For example, KC1 is the most widely used in NASA data set (both module and class level) which is applied in 43% of studies and has 15.4% faulty modules and 40.6% faulty classes. Similarly, PC1 data set is used in 40% of studies and contains only 6.8% of faulty modules.

The results show that 8% of datasets used in studies contain 0–4% of faulty classes/modules, 17% of datasets used in studies contain 5–9% of faulty classes/modules, 17% of datasets used in studies contain 10–14% of faulty classes/modules, and 14% of datasets used in studies contain 15–19% of faulty classes/modules. Hence, 60% of datasets used in studies contain less than 20% of faulty classes/modules which shows that very less amount of faulty classes/modules were present in data sets in more than half of the selected primary studies. The data having non proportionate ratio of faulty and not faulty modules/classes is often known as unbalanced data in the literature. The details on fault distribution in datasets corresponding to the studies are given in Appendix, Table A.2.

3.2.7. Which performance metrics are used for SFP?

A number of metrics are used for gauging the performance of different SFP models. These performance metrics are used for comparing and evaluating models developed using various ML and statistical techniques. Table 6 below depicts the performance metrics, their definitions and the studies in which they are used.

A depiction of the number of studies using each performance metric is used in Fig. 8. According to the figure, the most commonly used performance metric is Recall which is closely followed by Accuracy, Precision and AUC measures. Specificity, Pf and F-measures are other commonly used metrics. Some less commonly metrics are grouped in the miscellaneous category namely H-measure, Precision-recall curve, error rate, G-mean, G-mean1 and G-mean2.

There has been debate on the use of performance measures given the imbalanced nature of datasets in literature. Zhang et al. [16] criticize the use of recall and Pf for construction of prediction models. Menzies et al. [17] criticize the use of precision and pf, however recent studies (Lessmann et al. [42]; Menzies et al. [31]) advocate the use of AUC as performance measure for construction of SFP models. The detailed discussion on the use of performance measures is presented in Section 3.4.

3.3. RQ3: what is the overall performance of ML techniques for SFP?

In this section we summarize the results obtained by examining the results of SFP using the ML techniques. We provide the results of the ML techniques that were assessed in at least five out of 64 selected primary studies. The results are provided using two frequently used performance measures in the 64 primary studies (refer Section 3.2.5). The results in Section 3.2.5 show that Accuracy,

Table 6
Performance measures used in studies.

Performance metric	Definition	Studies
Accuracy	It represents the proportion of total number of correct predictions amongst total number of correct as well as incorrect predictions. It is also referred as <i>Precision</i> in certain studies.	SF2, SF3, SF4, SF6, SF8, SF11, SF13, SF14, SF18, SF20, SF21, SF22, SF25, SF27, SF31, SF32, SF35, SF37, SF38, SF40, SF41, SF42, SF43, SF46, SF48, SF49, SF53, SF57, SF60, SF63.
Precision/ correctness	It represents the proportion of correctly classified fault prone classes amongst total number of classified fault prone classes.	SF4, SF5, SF6, SF7, SF8, SF10, SF13, SF18, SF20, SF21, SF22, SF33, SF34, SF35, SF37, SF38, SF44, SF49, SF50, SF51, SF53, SF59, SF63.
Recall (sensitivity, Pd, true positive rate (TPR))	It is the proportion of correctly predicted fault prone classes amongst all actual fault prone classes.	SF2, SF3, SF5, SF7, SF8, SF9, SF12, SF13, SF14, SF15, SF16, SF18, SF20, SF21, SF22, SF24, SF25, SF26, SF27, SF28, SF30, SF32, SF33, SF34, SF35, SF36, SF37, SF38, SF40, SF43, SF44, SF48, SF49, SF50, SF51, SF53, SF56, SF57, SF59, SF60, SF62, SF63.
Specificity (true negative rate (TNR))	It is the proportion of correctly predicted non-fault prone classes amongst all actual non-fault prone classes.	SF2, SF3, SF8, SF14, SF20, SF27, SF32, SF33, SF40, SF43, SF44, SF48, SF53, SF57, SF60.
Pf (false positive rate (FPR))	It is the ratio of all non-fault prone classes which are incorrectly predicted as fault prone.	SF2, SF3, SF9, SF12, SF14, SF15, SF16, SF24, SF25, SF26, SF28, SF29, SF30, SF34, SF36, SF43, SF50, SF51, SF56, SF62.
FNR (false negative rate)	It is the ratio of faulty classes which are classified as non-fault prone.	SF14, SF29, SF43.
AUC (area under the curve)	ROC (receiver operating characteristic curve) is plotted with recall values on the y-axis and the 1-specificity values on the x-axis. The effectiveness of the model is classified by measuring the area under the ROC curve.	SF7, SF13, SF17, SF20, SF21, SF23, SF32, SF33, SF36, SF37, SF38, SF40, SF43, SF44, SF45, SF48, SF49, SF54, SF56, SF57, SF58, SF63, SF64.
Balance	It represents the optimal cutoff point of (pf, sensitivity) i.e. the normalized euclidian distance from which from (0,1) point in the ROC curve.	SF12, SF15, SF16, SF24, SF26, SF28, SF30, SF34, SF51, SF54, SF61.
Completeness	It is the sum of all faults in classified fault prone classes over the total number of actual faults in a system.	SF4, SF6, SF10, SF32, SF33, SF40.
F-measure	It is the harmonic mean of precision and sensitivity.	SF5, SF8, SF13, SF18, SF20, SF21, SF22, SF38, SF43, SF49, SF50, SF52, SF53, SF56, SF59, SF61, SF63.
Other miscellaneous measures	They include H-measure, Precision-Recall Curve, Error Rate, G-mean1 and G-mean2.	SF8, SF13, SF20, SF29, SF31, SF39, SF43, SF55, SF64.

F-measure, Precision Recall and AUC are the most frequently used performance measures in the selected primary studies. However, we provide the summarized results of two performance measures Accuracy and AUC. This is due to the fact that Menzies et al. [17] has criticized the use of F-measure and precision due to their unstable nature. In addition, most of the ML techniques being examined evaluate the effectiveness of the models using AUC.

Table 7
Results of SFP models using RF, J48, MLP and NB techniques.

Method	Count	Performance measure	Min.	Max.	Mean	Median	Std.
RF	7	Acc.	55	93.4	75.63	75.94	15.66
	35	AUC	0.66	1	0.83	0.82	0.09
J48	8	Acc.	78	92.562	83.02	81.7	4.54
MLP	10	Acc.	64.02	93.44	82.23	83.46	9.44
	40	AUC	0.54	0.95	0.78	0.77	0.09
NB	18	Acc.	61.92	92.56	80.57	81.4	7.63
	47	AUC	0.64	0.95	0.78	0.78	0.08

We have used box plots to gain insight about the prediction accuracy of the ML techniques. In Fig. 9 we present the box plots for AUC and accuracy values of the ML techniques. The outliers were removed from the analysis before providing the statistics.

Tables 7 and 8 present the minimum (Min.), maximum (Max.), mean, median and standard deviation (Std.) values for two selected performance measures. The higher values of the performance measures indicate more accurate prediction. In terms of AUC, the RF performed the best with 0.83 value, followed by the MLP, NB and BN models. The SVM models performed the worst with 0.7 value of AUC. The results indicate that the mean values of AUC for all the seven ML techniques range from 0.7 to 0.83. Similarly the mean values of accuracy range from 75% to 85%. This shows that the ML techniques have reasonable prediction capability (the values of AUC and accuracy indicate acceptable levels).

3.4. RQ4: whether the performance of ML techniques better than the statistical techniques?

The models predicted using the ML techniques have been compared with the models predicted using the traditional logistic regression (LR) technique, which is the most widely used statistical technique in the literature. In order to answer RQ4 we compared the LR models with ML models. The ML models showing above 2% improvement in the values of AUC were considered to outperform the corresponding LR models under comparison. It was observed that only 29% of studies compared LR and ML models (19 out of 64). We analyzed the results of those techniques which were compared with LR technique in at least three studies and five data sets.

AUC computed from ROC analysis is widely used in medical diagnosis since the past many years, and its use is increasing in the field of data mining research. Carvalho et al. [56] advocated AUC to be the relevant criterion for dealing with unbalanced and noisy data as AUC is insensitive to the changes in distribution of class. He and Garcia [18] have recommended the use of AUC for dealing the issues of imbalanced data as with regard to class distributions, it provides a visual representation of the relative trade-offs between the advantages (represented by true positives) and costs (represented by false positives) of classification. Further, recent studies (Lessmann et al. [42]; Menzies et al. [31]) have pointed out the existence of imbalanced data and use of AUC for construction of fault prediction models. We have seen the presence of imbalanced data in various studies in literature (refer Section 3.2.5), hence AUC

Table 8
Results of SFP models using BN, SVM and C4.5 techniques.

Method	Count	Performance measure	Min.	Max.	Mean	Median	Std.
BN	4	Acc.	74.89	81.43	76.88	75.6	3.06
	25	AUC	0.62	0.9	0.78	0.79	0.08
SVM	24	Acc.	63.34	97.6	80.77	80.41	9.34
	17	AUC	0.5	0.94	0.7	0.71	0.17
C4.5	11	Acc.	68.81	93.59	84.79	84.1	7.48
	20	AUC	0.5	0.99	0.77	0.77	0.12

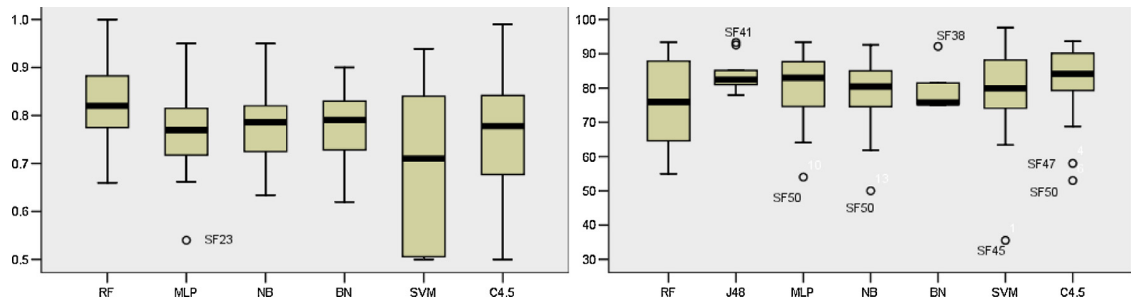


Fig. 9. Box plots for AUC and accuracy (outliers are labelled with study identification nos.).

is used for comparing the performance of models predicted using ML and LR techniques in this SLR.

The Fig. 10 indicates the percentage of the ML models (shown by bars above the zero-line) having value greater than the LR models with respect to datasets. The results show that 65% of the ML models outperform in terms of AUC as compared to LR models with respect to the datasets. The comparison results show that the RF, NB, C4.5, MLP and SVM are the most commonly used techniques for comparison in the literature. The RF (80%) and Bagging (100%) models outperform the LR models in most of the datasets. The NB model outperforms the LR models in 66% of studies followed by the MLP and C4.5 in 60% of studies. The LR models outperform the SVM models in 60% of studies. The models predicted using the J48 technique is compared with the LR models in only four studies (SF9, SF20, SF25, SF60). No more than three studies compared the AB (SF49, SF58, SF64) and Bagging (SF20, SF49, SF58) models with the LR model. The remaining ML techniques are compared in only one or two studies. Hence, the results obtained from the comparison of the ML techniques with the LR technique in one or two studies are not assessed and shown in this section as the less number of comparisons limits the possibility of generalization of results.

We may conclude that the ML models outperform the LR models in general. However, there is a threat to the validity to this conclusion. Due to lack of the studies comparing the performance of the LR and ML models (only 29%), it is difficult to obtain and generalize the conclusions.

Thus, more number of studies comparing ML models and LR models for SFP should be carried out in order to obtain well-formed and generalized results.

3.5. RQ5: are there any ML techniques that significantly outperform other ML techniques?

For the purpose of comparing the results of ML models with respect to other ML models we adopted the same methodology as in RQ4. The AUC was used to compare the results of ML models with other ML models. Table 9 presents the summary of average AUC,

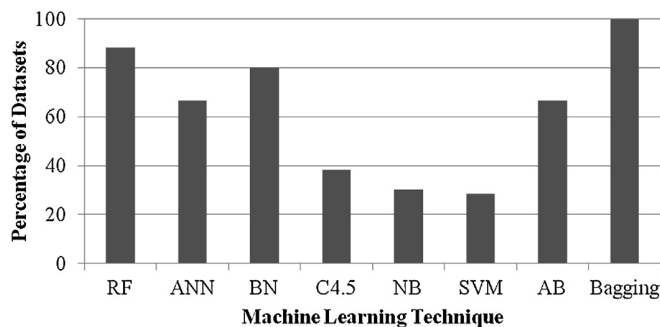


Fig. 10. Comparison of AUC between ML models and LR models.

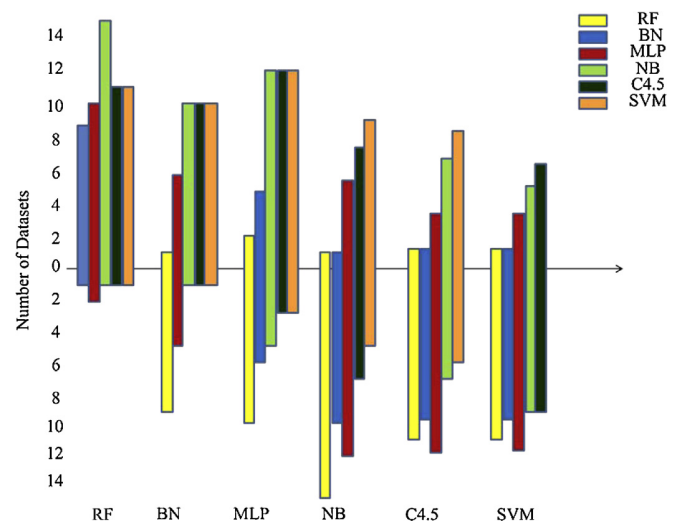


Fig. 11. Comparison of AUC between ML models vs ML models.

the best technique and the percentage of faulty modules/classes corresponding to the datasets. The results show that irrespective of the percentage of faulty data points in the datasets the AUC is greater than 0.6.

Fig. 11 shows the comparison results of ML models with other ML models with respect to the AUC value. The bars above the

Table 9
Average AUC corresponding to datasets.

Best technique	Dataset	Faulty %	Average AUC
RF	CM1	9.49	0.810
SVM	Cos	50.00	0.793
RF	Eclipse 2.0a	14.49	0.820
RF	Eclipse 2.1 a	10.83	0.740
RF	Eclipse 3.0a	14.80	0.770
ANN	Eclipse 2.0	30.80	0.662
ANN	Eclipse 2.1	23.48	0.710
NB	Eclipse 3.0	32.26	0.649
BN	Ivy	11.36	0.846
BN	Jedit	2.24	0.658
RF	JM1	19.34	0.747
RF	KC1	15.42	0.800
SVM	KC1 (class)	40.68	0.890
BN	KC2	20.38	0.850
RF	KC3	9.36	0.860
RF	KC4	48.8	0.850
RF	MC1	0.72	0.915
RF	MW1	7.67	0.810
RF	PC1	6.86	0.850
ANN	PC2	0.41	0.885
RF	PC3	10.23	0.807
ANN	PC4	12.20	0.945
RF	PC5	3.00	0.970

Table 10
Strengths and weaknesses of ML techniques.

Technique name	Strengths	Supporting studies
C4.5	The model developed is cost-effective and simple.	SF7, SF31
	Easy to build and apply.	SF7, SF27
	Comprehensive capability.	SF27, SF36
RF	It can handle large data and are consistent performers.	SF3, SF21
	The method is robust to noisy and missing data.	SF3, SF21
	Fast to train, robust towards parameter setting.	SF2, SF23, SF21
	Provide understandable model.	SF23, SF21
	Comprehensive capability.	SF23, SF36
	Runs efficiently on large data sets.	SF3, SF48, SF57
NB	Helps in identifying most important independent variables.	SF3, SF48, SF57
	It is robust in nature.	SF29, SF47
	It is easy to interpret and construct.	SF23, SF64
SVM	Computationally efficient.	SF23, SF64
	Good tolerance for high-dimension space and redundant features.	SF49, SF62, SF63
	Robust in nature.	SF23, SF62
	It can handle complex functions.	SF23, SF40
	It can handle non linear problems.	SF40, SF49
ANN	It can infer complex non-linear I/O transformation.	SF19, SF32
Technique name	Weaknesses	Supporting studies
NB	Does not consider feature correlation.	SF28, SF20
	Performance of model dependent on attribute selection technique used.	SF50, SF64
	Unable to discard irrelevant attributes.	SF50, SF64

zero-line depict the percentage of datasets in which the ML technique outperforms the other ML techniques. The bars below the zero-line were only used when the ML technique did not outperform the other ML techniques in any of the dataset. For example, the RF models outperform BN models in 9 datasets and underperforms BN model in 1 dataset. The RF models outperform MLP models in 10 datasets and underperform MLP models in 2 datasets. The RF models outperform NB models in 15 datasets and underperform NB model in 1 dataset. Similarly, the RF model outperform C4.5 and SVM models in 11 datasets an underperform C4.5 and SVM models in 1 dataset.

The results show that the RF models outperformed in most of the studies in comparison to the models predicted using the MLP, NB, BN, C4.5 and SVM techniques. The NB models outperform the MLP and SVM models in all the studies and the BN models outperformed the MLP, C4.5 and SVM models. Except the BN model, the SVM models did not outperformed any other ML technique.

3.6. RQ6: what are the strengths and weaknesses of the ML techniques?

In this section we outline the strengths and weaknesses exhibited by the ML techniques as recorded by researchers. These are representative of author's opinion and may not be reliable. Hence,

the strengths and weaknesses of the ML techniques supported by more than one study are presented in this section. The NB technique has been reported to exhibit good performance in defect prediction problems and is also good in handling multiple datasets with varying properties. At the same time this technique does not consider the correlation among features in the input and the performance of the NB technique varies heavily with the attribute selection technique used. The SVM has been applauded for its excellent ability to deal with the redundant features and high-dimensionality. The C4.5 algorithm is simple in implementation and provides good results, so is a good way of including the power of ML in one's research without much overhead. Table 10 summarizes the strengths and weaknesses of the ML techniques for SFP.

Summarily, different techniques have different pros and there is not a one size fits all solution to the problem of SFP in form of a ML technique. It all depends on the domain of application; in a situation where false positives are fatal would favour some techniques in comparison to a situation where cost is a factor.

4. Limitations

This systematic review considered a number of primary studies to evaluate and assess the performance of various ML techniques amongst themselves and with the LR technique. A limitation in this review is that only 19 out of the 64 primary studies compare LR and ML techniques. Thus, the LR vs ML comparison is not definitely conclusive. Moreover, while comparing ML techniques, each primary study may use different experimental settings which include data sets, feature reduction methods, weighing methods and pre-processing methods [9]. Though we have exhaustively searched all the stated digital search libraries, there still may be a possibility that a suitable study may be left out. Also this review does not include any unpublished research studies [19]. We have assumed that all the studies are impartial, however if this is not the case then it poses a threat to this study.

5. Conclusion and future guidelines

In this paper we perform a systematic literature review in order to analyze and assess the performance of the ML techniques for software fault prediction (SFP). First, after a thorough analysis by following a systematic series of steps and analysing the quality of the studies we identified 64 primary studies (1991–2013). Second, we summarized the characteristics of the primary studies based on metrics reduction techniques, metrics, data sets and performance measures. Third, we assessed the performance of the ML techniques for software fault prediction. Fourth, we compared the performance of models predicted using the ML techniques with the models predicted using the logistic regression technique. Then, we analyzed the performance of these predicted models using the ML techniques with other ML techniques. Finally, we summarized the strengths and weaknesses of the ML techniques. The main findings obtained from the selected primary studies are:

- The ML techniques were broadly categorized into Decision Tree, Bayesian Learning, Ensemble Learning, Rule Based Learning, Evolutionary Algorithms, Neural Networks and Miscellaneous. The most frequently used ML techniques for Software Fault Prediction were C4.5, Naive Bayes, Multilayer Perceptron, Support Vector Machines and Random Forest.
- The most commonly used technique for sub selection of metrics in the studies was Correlation based Feature Selection. The procedural metrics were found to be the most commonly used metrics in the literature. Among Object Oriented metrics CBO, RFC and LOC were found to be useful. The results show that the NASA data

set was the most frequently used data set in the literature. However, it was observed that very few studies use industrial datasets for evaluating the effectiveness of the ML techniques. The results depict that Accuracy, precision, Recall, AUC, F-Measure are the most commonly used performance measures in the primary studies.

- The ML techniques have acceptable prediction capability for estimating software fault proneness. The average values of AUC ranges from 0.7 to 0.83 and the mean values of accuracy ranges from 75% to 85%.
- The ML models outperformed the LR models for SFP in general.
- The models predicted using the RF technique outperformed all the other ML models in all the studies. The models predicted using the NB and BN techniques outperformed the other ML models in most of the studies.

The following are the guidelines for the researchers and software practitioners for carrying out future research on software fault prediction using the ML techniques:

- (1) More number of studies for software fault prediction should be carried out using the ML techniques in order to obtain generalizable results. There are few studies that compare the ML techniques with the LR techniques. Hence, more number of studies should compare and assess the performance of the ML techniques with LR technique.
- (2) It is observed that there are very few studies that examine the predictive capability of RBL, LB, AB, Bagging, RBF, ADT. Some ML techniques have never been applied in FP. Studies should be conducted in order to assess the prediction accuracy of these rarely used techniques.
- (3) There are very few studies that examine the effectiveness of evolutionary algorithms such as GP, ACO for software fault

prediction. The future studies may focus on the predictive accuracy of evolutionary algorithms for software fault prediction.

- (4) The data sets should be made available in public so that more number of studies can be carried out using the freely available data sets.
- (5) The ML techniques should be carefully selected based on their properties. Before making the decision of selection of a ML technique, the characteristics of the ML technique should be understood completely by the researcher.
- (6) In order to assess the true potential of a given ML technique, it should be compared with other ML/statistical techniques.
- (7) Most of the studies were based on the NASA or publicly available open source software. Hence, studies that explore the effectiveness of the ML techniques on industrial data sets should be conducted in order to provide practical usefulness of the ML techniques.
- (8) The study should clearly specify the parameters values for the respective ML techniques used so that the framework in the studies can be successfully repeated by the software community.

Acknowledgements

The author would like to thank Prof. Yogesh Singh for his guidance, constant input and suggestions throughout the work. The author is indebted to him for his advice and reviews on selection of primary studies. The author would also acknowledge the support of Megha Khanna and Nikita Jain for assisting in collecting and extracting the required data.

Appendix.

See [Tables A.1 and A.2.](#)

Table A.1
RQ's addressed in individual studies.

Study no.	RQ1	RQ2	RQ3	RQ4	RQ5	RQ6	Study No.	RQ1	RQ2	RQ3	RQ4	RQ5	RQ6
SF1	✓	✓	✓				SF33	✓	✓	✓			
SF2	✓	✓	✓		✓	✓	SF34	✓	✓	✓		✓	
SF3	✓	✓	✓		✓	✓	SF35	✓	✓	✓			
SF4	✓	✓	✓		✓		SF36	✓	✓	✓	✓	✓	✓
SF5	✓	✓	✓				SF37	✓	✓	✓	✓	✓	
SF6	✓	✓	✓		✓		SF38	✓	✓	✓		✓	
SF7	✓	✓	✓	✓	✓	✓	SF39	✓	✓	✓		✓	
SF8	✓	✓	✓		✓		SF40	✓	✓	✓		✓	
SF9	✓	✓	✓	✓	✓		SF41	✓	✓	✓		✓	✓
SF10	✓	✓	✓		✓		SF42	✓	✓	✓		✓	
SF11	✓	✓	✓		✓		SF43	✓	✓	✓		✓	
SF12	✓	✓	✓		✓	✓	SF44	✓	✓	✓	✓	✓	
SF13	✓	✓	✓		✓		SF45	✓	✓	✓	✓	✓	
SF14	✓	✓	✓			✓	SF46	✓	✓	✓		✓	
SF15	✓	✓	✓				SF47	✓	✓	✓		✓	
SF16	✓	✓	✓		✓		SF48	✓	✓	✓	✓	✓	✓
SF17	✓	✓	✓		✓		SF49	✓	✓	✓	✓	✓	✓
SF18	✓	✓	✓		✓		SF50	✓	✓	✓		✓	✓
SF19	✓	✓	✓		✓	✓	SF51	✓	✓	✓		✓	
SF20	✓	✓	✓	✓	✓	✓	SF52	✓	✓	✓			
SF21	✓	✓	✓		✓	✓	SF53	✓	✓	✓		✓	
SF22	✓	✓	✓				SF54	✓	✓	✓		✓	
SF23	✓	✓	✓	✓	✓	✓	SF55	✓	✓	✓		✓	
SF24	✓	✓	✓		✓		SF56	✓	✓	✓			
SF25	✓	✓	✓	✓	✓		SF57	✓	✓	✓	✓	✓	✓
SF26	✓	✓	✓		✓		SF58	✓	✓	✓		✓	
SF27	✓	✓	✓	✓	✓	✓	SF59	✓	✓	✓	✓	✓	
SF28	✓	✓	✓		✓	✓	SF60	✓	✓	✓	✓	✓	
SF29	✓	✓	✓			✓	SF61	✓	✓	✓	✓	✓	
SF30	✓	✓	✓				SF62	✓	✓	✓		✓	✓
SF31	✓	✓	✓		✓	✓	SF63	✓	✓	✓	✓	✓	✓
SF32	✓	✓	✓	✓	✓	✓	SF64	✓	✓	✓	✓	✓	✓

Table A.2
Fault distribution in studies.

Range of % fault distribution	# Datasets	Datasets with studies
0–4	9	PC2 (SF3, SF12, SF15, SF16, SF23, SF24, SF28, SF54, SF62, SF64), MC1 (SF24, SF54, SF64), SP3 (SF43), SP4 (SF43), PC5 (SF20, SF24, SF54, SF64), ARGO (SF52), Camel 1.0 (SF59), Inventory System 13 (SF41), SP2 (SF43)
5–9	17	Inventory System 14 (SF41), CCCS12 (SF43), Inventory System 17 (SF41), Inventory System 16 (SF41), SP1 (SF43), Ivy 1.4 (SF59), PC1 (SF3, SF5, SF11, SF12, SF13, SF15, SF16, SF17, SF18, SF20, SF23, SF24, SF27, SF28, SF30, SF38, SF39, SF43, SF51, SF53, SF54, SF55, SF56, SF62, SF64), Inventory System 12 (SF41), Inventory System 15 (SF41), Inventory System 11 (SF41), AR1 (SF44, SF54), MW1 (SF12, SF15, SF16, SF23, SF24, SF28, SF30, SF39, SF43, SF53, SF54, SF56, SF62), Inventory System 10 (SF41), Datatrieve (SF42), KC3 (SF12, SF15, SF16, SF18, SF23, SF24, SF28, SF39, SF43, SF53, SF54, SF56, SF62), CM1 (SF3, SF5, SF9, SF12, SF13, SF15, SF16, SF18, SF20, SF23, SF24, SF28, SF30, SF38, SF43, SF51, SF53, SF54, SF56, SF62), CCCS8 (SF43).
10–14	17	Argo uml 0.26.2 (SF52), Synapse 1.0 (SF59), PC3 (SF12, SF15, SF16, SF23, SF24, SF28, SF51, SF54, SF62, SF64), Arc (SF48), Inventory System 7 (SF41), Eclipse 2.1a (SF64), Ant 1.5 (SF59), Ivy 2.0 (SF59, SF58), Inventory System 9 (SF41), POI 2.0 (SF59), PC4 (SF12, SF15, SF16, SF23, SF24, SF27, SF28, SF51, SF54, SF60, SF63, SF64), Inventory System 8 (SF41), AR3 (SF26, SF29, SF30, SF34, SF51, SF54, SF56), Inventory System 6 (SF41), Eclipse 2.0a (SF64), Eclipse 3.0a (SF64), AR6 (SF54, SF51).
15–19	15	Xalan 2.4 (SF59), Xerces 1.3 (SF59), Inventory System 4 (SF41), KC1 (SF3, SF9, SF12, SF15, SF18, SF20, SF23, SF24, SF27, SF30, SF38, SF39, SF43, SF53, SF54, SF56, SF60, SF62, SF64), Inventory System 2 (SF41), Ant 1.3 (SF59), Xerces 1.2 (SF59), Inventory System 5 (SF41), Camel 1.4 (SF59), AR4 (SF26, SF29, SF30, SF34, SF51, SF54, SF56), JM1 (SF3, SF5, SF9, SF11, SF13, SF19, SF20, SF23, SF38, SF39, SF42, SF43, SF54, SF55, SF62, SF64), AR5 (SF26, SF29, SF30, SF34, SF51, SF56), Camel 1.6 (SF59), CCCS4 (SF43), Inventory System 3 (SF41).
20–24	7	KC2 (SF2, SF3, SF5, SF12, SF13, SF20, SF24, SF30, SF38, SF39, SF43, SF53, SF56), Ant 1.7 (SF58, SF59), Jedit 4.3 (SF58), Ant 1.4 (SF59), Eclipse 2.1 (SF25, SF34, SF35, SF45), Jedit 4.0 (SF49, SF59), Eclipse Component: SWT (SF61).
25–29	3	Ant 1.6 (SF59), Synapse 1.1 (SF59), CCCS2 (SF43).
30–34	6	Eclipse 2.0 (SF25, SF34, SF35, SF61, SF52, SF50, SF45), Eclipse 3.0 (SF25, SF35, SF45, SF52, SF61), MC2 (SF20, SF24, SF30, SF53, SF54, SF56), Jedit 3.2 (SF59, SF21), Synapse 1.2 (SF58, SF59), Velocity 1.6 (SF58, SF59).
35–39	2	Camel 1.2 (SF60), Inventory System 1 (SF41).
40–44	2	KC1 (SF5, SF6, SF8, SF14, SF32, SF33, SF38, SF40, SF50), Library Management System (SF10).
45–49	5	Xalan 2.6 (SF59, SF61), Lucene 2.0 (SF59), Xerces init (SF59), Xalan 2.5 (SF58, SF59), KC4 (SF15, SF16, SF20, SF23, SF28, SF54).
50–54	2	Cos (SF7, SF37), Eclipse Component: JDT Core (SF61)
55–59	5	Ivy 1.1 (SF59), Mozilla (SF1), Lucene 2.2 (SF59), POI 1.5 (SF59), Lucene 2.4 (SF58, SF59, SF61)
60–64	1	POI 2.5 (SF59).
65–70	2	Velocity 1.5 (SF59), POI 3.0 (SF57, SF58, SF59).
>70	5	Xerces 1.4 (SF59), Velocity 1.4 (SF59), Tomcat 6.0 (SF35, SF58), Jedit1 (SF49, SF46), Jedit2 (SF49, SF46).

Bold text signifies the names of the data sets used in the respective studies depicted in brackets.

References

- [1] N. Seliya, T.M. Khoshgoftaar, J. Van Hulse, Predicting faults in high assurance software, in: IEEE 12th International Symposium on High-Assurance Systems Engineering, 2010, pp. 26–34.
- [2] Y. Singh, A. Kaur, R. Malhotra, Empirical validation of object-oriented metrics for predicting fault proneness models, *Softw. Qual. J.* 18 (2009) 3–35.
- [3] D. Radjenovic, M. Hericko, R. Torkar, A. Zivkovic, Software fault prediction metrics: a systematic literature review, *Inform. Softw. Technol.* 55 (2013) 1397–1418.
- [4] C. Catal, Software fault prediction: a literature review and current trends, *Expert Syst. Appl.* 38 (2011) 4626–4636.
- [5] C. Catal, B. Diri, A systematic review of software fault prediction studies, *Expert Syst. Appl.* 36 (2009) 7346–7534.
- [6] B.A. Kitchenham, Guidelines for performing systematic literature review in software engineering, Technical report EBSE-2007-001, UK, 2007.
- [7] T.M. Mitchell, Machine Learning, McGraw Hill, Maidenhead, UK, 1997.
- [8] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Elsevier, USA, 2005.
- [9] J. Wen, S. Li, Z. Lin, Y. Hu, C. Huang, Systematic literature review of machine learning based software development effort estimation models, *Inform. Softw. Technol.* 54 (2012) 41–59.
- [10] M.A. Hall, Correlation-based feature selection for discrete and numeric class machine learning, in: Proceedings of the Seventeenth Conference on Machine Learning, 2000, pp. 359–366.
- [11] M. Halstead, Elements of Software Science, Elsevier, New York, 1977.
- [12] T. McCabe, A complexity measure, *IEEE Trans. Softw. Eng.* 2 (1976) 308–320.
- [13] S.R. Chidambler, C.F. Kemerer, A metrics suite for object oriented design, *IEEE Trans. Softw. Eng.* 20 (1994) 476–493.
- [14] K.K. Aggarwal, Y. Singh, A. Kaur, R. Malhotra, Empirical study of object-oriented metrics, *J. Obj. Technol.* 5 (2006) 149–173.
- [15] M. Lorenz, J. Kidd, Object-oriented Software Metrics, Prentice Hall, Englewood Cliffs, NJ, 1994.
- [16] H. Zhang, X. Zhang, Comments on 'data mining static code attributes to learn defect predictors', *IEEE Trans. Softw. Eng.* 33 (2007) 635–636.
- [17] T. Menzies, A. Dekhtyar, J. Distefano, J. Greenwald, Problems with precision: a response to comments on 'data mining static code attributes to learn defect predictors', *IEEE Trans. Softw. Eng.* 33 (2007) 637–640.
- [18] H. He, E.A. Gracia, Learning from imbalanced data, *IEEE Trans. Knowl. Data Eng.* 21 (2009) 1263–1284.
- [19] W. Afzal, R. Torkar, On the application of genetic programming for software engineering predictive modelling: a systematic review, *Expert Syst. Appl.* 38 (2011) 11984–11997.
- [20] S.A. Sherer, Software fault prediction, *J. Syst. Softw.* 29 (1995) 97–105.
- [21] L. Guo, B. Cucik, H. Singh, Predicting fault prone modules by the Dempster-Shafer belief networks, in: 18th IEEE International Conference on Automated Software Engineering, 2003, pp. 3–6.
- [22] L. Guo, Y. Ma, B. Cucik, H. Singh, Robust prediction of fault-proneness by random forests, in: 15th International Symposium on Software Reliability Engineering, 2004, pp. 417–428.
- [23] T. Gyimothy, R. Ferenc, I. Siket, Empirical validation of object-oriented metrics on open source software for fault prediction, *IEEE Trans. Softw. Eng.* 31 (2005) 897–910.
- [24] A.G. Koru, H. Liu, Building effective defect-prediction models in practice, *IEEE Softw.* 22 (2005) 23–29.
- [25] Y. Zhou, H. Leung, Empirical analysis of object-oriented design metrics for predicting high and low severity faults, *IEEE Trans. Softw. Eng.* 32 (2006) 771–789.
- [26] E. Arisholm, L.C. Briand, M. Fuglerud, Data mining techniques for building fault-proneness models in telecom java software, in: 18th IEEE International Symposium on Software Reliability, 2007, pp. 215–224.
- [27] C. Catal, B. Diri, B. Ozumut, An artificial immune system approach for fault prediction in object-oriented software, in: 2nd International Conference on Dependability of Computer Systems, 2007, pp. 1–8.
- [28] Y. Jiang, B. Cucik, T. Menzies, Fault prediction using early lifecycle data, in: 18th IEEE International Symposium on Software Reliability, 2007, pp. 237–246.
- [29] S. Kanmani, V.R. Uthararaj, V. Sankaranarayanan, P. Thambidurai, Object-oriented software fault prediction using neural networks, *Inform. Softw. Technol.* 49 (2007) 483–492.
- [30] Z. Li, M. Reformat, A practical method for the software fault-prediction, in: IEEE International Conference on Information Reuse and Integration, 2007, pp. 659–666.
- [31] T. Menzies, J. Greenwald, A. Frank, Data mining static code attributes to learn defect predictors, *IEEE Trans. Softw. Eng.* 33 (2007) 2–13.
- [32] Y. Ma, L. Guo, B. Cucik, A statistical framework for the prediction of fault-proneness, *Adv. Mach. Learn. Appl. Softw. Eng. IGI Global* (2007) 237–263.
- [33] G.J. Pai, J.B. Dugan, Empirical analysis of software fault content and fault proneness using bayesian methods, *IEEE Trans. Softw. Eng.* 33 (2007) 675–686.

- [34] B. Turhan, A. Bener, A multivariate analysis of static code attributes for defect prediction, in: 7th International Conference on Quality Software, 2007, pp. 231–237.
- [35] B. Turhan, A. Bener, Software defect prediction: heuristics for weighted naïve bayes, in: 2nd International Conference on Software and Data Technologies, 2007, pp. 244–249.
- [36] A.B. De Carvalho, A. Pozo, S. Vergilio, A. Lenz, Predicting fault proneness of classes through a multiobjective particle swarm optimization algorithm, in: 20th IEEE International Conference on Tools with Artificial Intelligence, 2008, pp. 387–394.
- [37] K.O. Elish, M.O. Elish, Predicting defect-prone software modules using support vector machines, *J. Syst. Softw.* 81 (2008) 649–660.
- [38] I. Gondra, Applying machine learning to software fault-proneness prediction, *J. Syst. Softw.* 81 (2008) 186–195.
- [39] Y. Jiang, Y. Ma, B. Cukic, Techniques for evaluating fault prediction models, *Empir. Softw. Eng.* 13 (2008) 561–595.
- [40] A. Kaur, R. Malhotra, Application of random forest in predicting fault-prone classes, in: International Conference on Advanced Computer Theory and Engineering, 2008, pp. 37–43.
- [41] S. Kim, E.J. Whitehead, Y. Zhang, Classifying software changes: clean or buggy? *IEEE Trans. Softw. Eng.* 34 (2008) 181–196.
- [42] S. Lessmann, B. Baesens, C. Mues, S. Pietsch, Benchmarking classification models for software defect prediction: a proposed framework and novel findings, *IEEE Trans. Softw. Eng.* 34 (2008) 485–496.
- [43] T. Menzies, B. Turhan, A. Bener, G. Gay, B. Cukic, Y. Jiang, Implications of ceiling effects in defect predictors, in: 4th International Workshop on Predictive Models in Software Engineering – PROMISE’08, 2008, pp. 47–54.
- [44] R. Moser, W. Pedrycz, G. Succi, A comparative analysis of the efficiency of change metrics and static code attributes for defect prediction, in: 30th International Conference on Software Engineering, 2008, pp. 181–190.
- [45] B. Turhan, G. Kocak, A. Bener, Software defect prediction using call graph based ranking (CGBR) framework, in: 34th Euromicro Conference on Software Engineering and Advanced Applications, 2008, pp. 191–198.
- [46] O. Vandercruys, D. Martens, B. Baesens, C. Mues, M. D Backer, R. Haesen, Mining software repositories for comprehensible software fault prediction models, *J. Syst. Softw.* 81 (2008) 823–839.
- [47] A. Bener, B. Turhan, Analysis of naïve bayes assumptions on software fault data: an empirical study, *Data Knowl. Eng.* 68 (2009) 278–290.
- [48] C. Catal, U. Sevim, B. Diri, Clustering and metrics thresholds based software fault prediction of unlabeled program modules, in: Sixth International Conference on Information Technology: New Generations, 2009, pp. 199–204.
- [49] T. Menzies, B. Turhan, J. D Stefano, A.B. Bener, On the relative value of cross-company and within-company data for defect prediction, *Empir. Softw. Eng.* 14 (2009) 540–578.
- [50] P. Singh, S. Verma, An investigation of the effect of discretization on defect prediction using static measures, in: International Conference on Advances in Computing, Control, and Telecommunication Technology, 2009, pp. 837–839.
- [51] Y. Singh, A. Kaur, R. Malhotra, Software fault proneness prediction using support vector machines, in: Proceedings of the World Congress on Engineering, 2009, pp. 240–245.
- [52] A. Tosun, B. Turhan, A. Bener, Validation of network measures as indicators of defective modules in software systems, in: Proceedings of 5th International Conference on Predictive Models Software Engineering, 2009.
- [53] T. Zimmermann, N. Nagappan, H. Gall, E. Giger, B. Murphy, Cross-project defect prediction: a large scale experiment on data vs domain vs process, in: 7th Joint Meeting of European Software Engineering Conference. ACM SIGSOFT Symposium Foundations of Software Engineering, 2009, pp. 91–100.
- [54] W. Afzal, Using faults-slip-through metric as a predictor of fault-proneness, in: Asia Pacific Software Engineering Conference, 2010, pp. 414–422.
- [55] E. Arisholm, L.C. Briand, E.B. Johannessen, A systematic and comprehensive investigation of methods to build and evaluate fault prediction models, *J. Syst. Softw.* 83 (2010) 2–17.
- [56] A.B. Carvalho, A. Pozo, S.R. Vegilio, A symbolic fault-prediction model based on multiobjective particle swarm optimization, *J. Syst. Softw.* 83 (2010) 868–882.
- [57] Y. Liu, T.M. Khoshgoftaar, N. Seliya, Evolutionary optimization of software quality modeling with multiple repositories, *IEEE Trans. Softw. Eng.* 36 (2010) 852–864.
- [58] R. Malhotra, Y. Singh, A. Kaur, Empirical validation of object-oriented metrics for predicting fault proneness at different severity levels using support vector machines, *Int. J. Syst. Assur. Eng. Manag.* 1 (2010) 269–281.
- [59] T.J. Ostrand, E.J. Weyuker, R.M. Bell, Comparing the effectiveness of several modeling methods for fault prediction, *Empir. Softw. Eng.* 15 (2010) 277–295.
- [60] P.C. Pendharkar, Exhaustive and heuristic search approaches for learning a software defect prediction model, *Eng. Appl. Artif. Intell.* 23 (2010) 34–40.
- [61] Y. Singh, A. Kaur, R. Malhotra, Prediction of fault-prone software modules using statistical and machine learning methods, *Int. J. Comput. Appl.* 1 (2010) 6–13.
- [62] Y. Zhou, B. Xu, H. Leung, On the ability of complexity metrics to predict fault-prone classes in object-oriented systems, *J. Syst. Softw.* 83 (2010) 660–674.
- [63] D. Azar, J. Vybihal, An ant colony optimization algorithm to improve software quality prediction models: case of class stability, *Inform. Softw. Technol.* 53 (2011) 388–393.
- [64] B. Diri, C. Catal, U. Sevim, Practical development of an eclipse-based software fault prediction tool using Naïve Bayes algorithm, *Expert Syst. Appl.* 38 (2011) 2347–2353.
- [65] R. Malhotra, Y. Singh, On the applicability of machine learning techniques for object-oriented software fault prediction, *Softw. Eng. Int. J.* 1 (2011) 24–37.
- [66] S. Martino, F. ferrucci, C. Gravino, F. Sarro, A genetic algorithm to configure support vector machine for predicting fault prone components, in: Proceedings of the 12th International Conference on Product Focus Software Process Improvement, 2011, pp. 247–261.
- [67] B. Mishra, K.K. Shukla, Impact of attribute selection on defect proneness prediction in OO software, in: 2nd International Conference on Computing, Communication and Technology, 2011, pp. 367–372.
- [68] A.T. Misirlı, A.B. Bener, B. Turhan, An industrial case study of classifier ensembles for locating software defects, *Softw. Qual. J.* 19 (2011) 515–536.
- [69] S. Kpodjedo, F. Ricca, P. Galinier, Y.-G. Guéhéneuc, G. Antoniol, Design evolution metrics for defect prediction in object oriented systems, *Empir. Softw. Eng.* 16 (2011) 141–175.
- [70] D. Rodríguez, R. Ruiz, J.C. Riquelme, J.S. Aguilar-Ruiz, Searching for rules to detect defective modules: a subgroup discovery approach, *Inform. Sci.* 191 (2011) 14–30.
- [71] Q. Song, Z. Jia, M. Shepperd, S. Ying, J. Liu, A general software defect-proneness prediction framework, *IEEE Trans. Softw. Eng.* 37 (2011) 356–370.
- [72] B. Twala, Software faults prediction using multiple classifiers, in: 3rd International Conference on Computing Research and Development, vol. 4, 2011, pp. 504–510.
- [73] A. Chen, Y. Ma, X. Zeng, G. Luo, Transfer learning for cross-company software defect prediction, *Inform. Softw. Technol.* 54 (2012) 248–256.
- [74] R. Malhotra, A. Jain, Fault prediction using statistical and machine learning methods for improving software quality, *J. Inf. Process. Syst.* 8 (2012) 241–262.
- [75] A. Okutan, O.T. Yıldız, Software defect prediction using Bayesian networks, *Empir. Softw. Eng.* 19 (2012) 154–181.
- [76] Y. Yang, Z. He, F. Shu, M. Li, Q. Wang, An investigation on the feasibility of cross-project defect prediction, *Autom. Softw. Eng.* 19 (2012) 167–199.
- [77] L. Yu, An evolutionary programming based asymmetric weighted least squares support vector machine ensemble learning methodology for software repository mining, *Inform. Sci.* 191 (2012) 31–46.
- [78] M. Li, H. Zhang, R. Whu, Z. Zhou, Sample-based software defect prediction with active and semi-supervised learning, *Autom. Softw. Eng.* 19 (2012) 201–230.
- [79] J. Cahill, J.M. Hogan, R. Thomas, Predicting fault-prone software modules with rank sum classification, in: 22nd Australian Software Engineering Conference, 2013, pp. 211–219.
- [80] N. Chen, S.C.H. Hoi, X. Xiao, Software process evaluation: a machine learning framework with application to defect management process, *Empir. Softw. Eng.* 19 (2013) 1531–1564.
- [81] K. Dejaeger, T. Verbraken, B. Baesens, Toward comprehensible software fault prediction models using bayesian network classifiers, *IEEE Trans. Softw. Eng.* 39 (2013) 237–257.
- [82] G.K. Upadhyay, S. Das, U. Chandra, A.J. Paul, Modelling the investment casting process: a novel approach for view factor calculations and defect prediction, *Appl. Math. Model.* 19 (1995) 354–362.
- [83] M. Evett, B. Raton, E. Allen, GP-based software quality prediction, in: Proceedings of 3rd Annual Genetic Programming Conference, 1998, pp. 60–65.
- [84] E. Baisch, T. Liedtke, Comparison of conventional approaches and soft-computing approaches for software quality prediction, in: IEEE International Conference on Systems, Man and Cybernetics, Comput. Cybern. Simul. 2 (1997) 1045–1049.
- [85] A.R. Gray, S.G. MacDonell, A comparison of techniques for developing predictive models of software metrics, *Inform. Softw. Technol.* 39 (1997) 425–437.
- [86] M. Zhao, C. Wohlin, N. Ohlsson, M. Xie, A comparison between software design and code metrics for the prediction of software fault content, *Inform. Softw. Technol.* 40 (1998) 801–809.
- [87] T. Kamiya, S. Kusumoto, K. Inoue, Prediction of fault-proneness at early phase in object-oriented development, in: 2nd IEEE International Symposium on Object-Oriented Real-Time Distribution and Computing, 1999, pp. 253–258.
- [88] L.C. Briand, J. Wust, J.W. Daly, D.V. Porter, Exploring the relationships between design measures and software quality in object-oriented systems, *J. Syst. Softw.* 51 (2000) 245–273.
- [89] X. Yuan, T.M. Khoshgoftaar, E.B. Allen, K. Ganesan, An application of fuzzy clustering to software quality prediction, in: 3rd IEEE Symposium on Application-Specific Systems and Software Engineering Technology, 2000, pp. 85–92.
- [90] G. Denaro, P. Milano, An empirical evaluation of fault-proneness models, in: 24th International Conference on Software Engineering, 2002, pp. 241–251.
- [91] T.M. Khoshgoftaar, N. Seliya, Tree-based software quality estimation models for fault prediction, in: 8th IEEE Symposium on Software Metrics, 2002, pp. 203–214.
- [92] N.J. Pizzi, A.R. Summers, W. Pedrycz, Software quality prediction using median-adjusted class labels, in: International Joint Conference of Neural Network, 2002, pp. 2405–2409.
- [93] M.D. Ambros, M. Lanza, R. Robbes, An extensive comparison of bug prediction approaches, in: 7th IEEE Working Conference on Mining Software Repositories, 2003, pp. 31–41.
- [94] T.M. Khoshgoftaar, N. Seliya, Fault prediction modeling for software quality estimation: comparing commonly used techniques, *Empir. Softw. Eng.* (2003) 255–283.
- [95] K. Kaminsky, G. Boetticher, Building a genetically engineerable evolvable program (GEEP) using breadth-based explicit knowledge for predicting software

- defects, in: IEEE Annual Meeting on Fuzzy Information and Process, 2004, pp. 10–15.
- [96] T. Menzies, J.S. Di Stefano, How good is your blind spot sampling policy? in: 8th IEEE International Symposium on High Assurance System Engineering, 2004, pp. 129–138.
- [97] T.S. Quah, M.M.T. Thwin, Prediction of software development faults in PL/SQL files using neural network models *Inform. Softw. Technol.* 46 (2004) 519–523.
- [98] Q. Wang, B. Yu, J. Zhu, Extract rules from software quality prediction model based on neural network, in: 16th IEEE International Conference on Tools with Artificial Intelligence, 2004, pp. 191–195.
- [99] G. Canfora, L. Cerulo, Impact analysis by mining software and change request repositories, in: 11th IEEE International Software Metrics Symposium, 2005, pp. 20–29.
- [100] V.U.B. Challagulla, F.B. Bastani, I.L. Yen, R.A. Paul, Empirical assessment of machine learning based software defect prediction techniques, in: 10th IEEE International Workshop on Object-Oriented Real-Time Dependable Systems, 2005, pp. 263–270.
- [101] F. Xing, P. Guo, M.R. Lyu, A novel method for early software quality prediction based on support vector machine, in: 16th IEEE International Symposium on Software Reliability Engineering, 2005, pp. 213–222.
- [102] V.U.B. Challagulla, F.B. Bastani, I.L. Yen, A unified framework for defect data analysis using the MBR technique, in: 18th IEEE International Conference on Tools with Artificial Intelligence, 2006, pp. 39–46.
- [103] T.M. Khoshgoftaar, N. Seliya, N. Sundaresh, An empirical study of predicting software faults with case-based reasoning, *Softw. Qual. J.* 14 (2006) 85–111.
- [104] P. Knab, M. Pinzger, A. Bernstein, Predicting defect densities in source code files with decision tree learners, in: Proceedings of International Workshop on Mining Software Repositories, 2006, pp. 119–125.
- [105] M. Mertic, M. Lenic, G. Stiglic, P. Kokol, Estimating software quality with advanced data mining techniques, in: International Conference on Software Engineering Advances, 2006, pp. 19–22.
- [106] M.E.R. Bezerra, A.L.I. Oliveira, S.R.L. Meira, A constructive RBF neural network for estimating the probability of defects in software modules, in: Proceedings of International Joint Conference on Neural Networks, 2007, pp. 2869–2874.
- [107] W. Li, R. Shatnawi, An empirical study of the bad smells and class error probability in the post-release object-oriented system evolution, *J. Syst. Softw.* 80 (2007) 1120–1128.
- [108] N. Seliya, T.M. Khoshgoftaar, Software quality estimation with limited fault data: a semi-supervised learning perspective, *Softw. Qual. J.* 15 (2007) 327–344.
- [109] P. Tomaszewski, J. Håkansson, H. Grahm, L. Lundberg, Statistical models vs. expert estimation for fault prediction in modified code – an industrial case study, *J. Syst. Softw.* 80 (2007) 1227–1238.
- [110] B. Yang, L. Yao, H.-Z. Huang, Early software quality prediction based on a fuzzy neural network model, 3rd International Conference on Natural Computing 2 (2007) 760–764.
- [111] T. Zimmermann, R. Premraj, A. Zeller, Predicting defects for eclipse, in: Third International Workshop on Predictive Models in Software Engineering, 2007, pp. 9–15.
- [112] W. Afzal, R. Torkar, A comparative evaluation of using genetic programming for predicting fault count data, in: 3rd International Conference Software Engineering Advances, 2008, pp. 407–414.
- [113] K.K. Aggarwal, Y. Singh, A. Kaur, R. Malhotra, Empirical analysis for investigating the effect of object-oriented metrics on fault proneness: a replicated case study, *Softw. Process Improv. Pract.* 14 (2008) 39–62.
- [114] S. Bibi, I. Vlahavas, G. Tsoumakas, I. Stamelos, Regression via classification applied on software defect estimation, *Expert Syst. Appl.* 34 (2008) 2091–2101.
- [115] A. Marcus, D. Poshvanyk, Using the conceptual cohesion of classes for fault prediction in object-oriented systems, *IEEE Trans. Softw. Eng.* 34 (2008) 287–300.
- [116] S. Shafi, S.M. Hassan, A. Arshaq, M.J. Khan, S. Shamil, Software quality prediction techniques: a comparative analysis, in: 4th International Conference on Emerging Technologies, 2008, pp. 242–246.
- [117] W. Afzal, R. Torkar, R. Feldt, Search-based prediction of fault count data, in: 1st International Symposium on Search Based Software Engineering, 2009, pp. 35–38.
- [118] Y. Jiang, B. Cukic, Misclassification cost-sensitive fault prediction models, in: Proceeding of 5th International Conference on Predictive Models in Software Engineering, 2009, pp. 20–30.
- [119] A. Kaur, P.S. Sandhu, A.S. Brar, Early software fault prediction using real time defect data, in: 2nd International Conference on Machine Vision, 2009, pp. 242–245.
- [120] T.M. Khoshgoftaar, K. Gao, Feature selection with imbalanced data for software defect prediction, in: International Conference on Machine Learning Applications, 2009, pp. 235–240.
- [121] Y. Singh, A. Kaur, R. Malhotra, Prediction of software quality model using gene expression programming, *Prod. Softw. Process Improv.* 32 (2009) 43–58.
- [122] W. Afzal, R. Torkar, R. Feldt, G. Wikstrand, Search-based prediction of fault-slip-through in large software projects, in: 2nd International Symposium on Search Based Software Engineering, 2010, pp. 79–88.
- [123] L. Hribar, D. Duka, Software component quality prediction using KNN and fuzzy logic, in: Proceedings of 33rd International Convention (MIPRO), 2010, pp. 402–408.
- [124] T. Illes-Seifert, B. Paech, Exploring the relationship of a file's history and its fault-proneness: an empirical study, in: Testing: Academic and Industrial Conference on Practical and Research Techniques, 2008, pp. 13–22.
- [125] P.S. Sandhu, A.S. Brar, R. Goel, J. Kaur, S. Anand, A model for early prediction of faults in software systems, in: 2nd International Conference on Computer Automation Engineering, vol. 4, 2010, pp. 281–285.
- [126] N. Nagappan, A. Zeller, T. Zimmermann, K. Herzig, B. Murphy, Change bursts as defect predictors, in: IEEE 21st International Symposium on Software Reliability and Engineering, 2010, pp. 309–318.
- [127] N.H. Chiu, Combining techniques for software quality classification: an integrated decision network approach, *Expert Syst. Appl.* 38 (2011) 4618–4625.
- [128] M. Masud, J. Gao, L. Khan, J. Han, B.M. Thuraisingham, Classification and novel class detection in concept-drifting data streams under time constraints, *IEEE Trans. Knowl. Data Eng.* 23 (2011) 859–874.
- [129] P. Bangcharoensap, A. Ihara, Y. Kamei, K. Matsumoto, Locating source code to be fixed based on initial bug reports – a case study on the Eclipse project, in: 4th International Workshop on Empirical Software Engineering Practice, 2012, pp. 10–15.
- [130] P.S. Bishnu, V. Bhattacherjee, Software fault prediction using quad-tree based *k*-means clustering algorithm, *IEEE Trans. Knowl. Data Eng.* 24 (2012) 1146–1150.
- [131] M.D. Ambros, M. Lanza, Evaluating defect prediction approaches: a benchmark and an extensive comparison, *Empir. Softw. Eng.* 17 (2012) 531–577.
- [132] L. Pelayo, S. Dick, Evaluating stratification alternatives to improve software defect prediction, *IEEE Trans. Reliab.* 61 (2012) 516–525.
- [133] R.G. Ramani, S.V. Kumar, S.G. Jacob, Predicting fault-prone software modules using feature selection and classification through data mining algorithms, in: International Conference on Computational Intelligence & Computing Research, 2012, pp. 1–4.
- [134] S.S. Rathore, A. Gupta, Investigating object-oriented design metrics to predict fault-proneness of software modules, in: CSI Sixth International Conference on Software Engineering, 2012, pp. 1–10.
- [135] P. Singh, S. Verma, Empirical investigation of fault prediction capability of object oriented metrics of open source software, in: International Joint Conference on Computer Science and Software Engineering, 2012, pp. 323–327.
- [136] A. Mahaweerawat, P. Sophatsathit, C. Lursinsap, Adaptive Self-organizing Map Clustering for Software Fault Prediction, 2007, pp. 35–41.
- [137] A. Mahaweerawat, P. Sophatsathit, C. Lursinsap, Software Fault Prediction using Fuzzy Clustering and Radial Basis Function Network (2002) 304–313.
- [138] G. Mauša, Search Based Software Engineering and Software Defect Prediction, 2010, <https://www.fer.unizg.hr/download/repository/Kvalifikacijski.-Goran.Mausa.pdf>.
- [139] A. Monden, T. Hayashi, S. Shinoda, K. Shirai, J. Yoshida, M. Barker, K. Matsumoto, Assessing the cost effectiveness of fault prediction in acceptance testing, *IEEE Trans. Softw. Eng.* 39 (2013) 1345–1357.

Advanced Adaptive Particle Swarm Optimization based SVC Controller for Power System Stability

Poonam Singhal

YMCA Univ. of Science and Technology/Electrical Engg, Faridabad, India
Email: ymcapoonamsinghal@gmail.com

S. K. Agarwal, Narendra Kumar

YMCA Univ. of Science and Technology, DTU /Electronics Engg, Electrical Engg, Faridabad, Delhi, India
Email: {sa_3264@yahoo.co.in, dnn_1963@yahoo.com}

Abstract— The interconnected systems is continually increasing in size and extending over whole geographical regions, it is becoming increasingly more difficult to maintain synchronism between various parts of the power system. This paper work presents an advanced adaptive Particle swarm optimization technique to optimize the SVC controller parameters for enhancement of the steady state stability & overcoming the premature convergence & stagnation problems as in basic PSO algorithm & Particle swarm optimization with shrinkage factor & inertia weight approach (PSO-SFIWA). In this paper SMIB system along with PID damped SVC controller is considered for study. The generator speed deviation is used as an auxiliary signal to SVC, to generate the desired damping. This controller improves the dynamic performance of power system by reducing the steady-state error. The controller parameters are optimized using basic PSO, PSO-SFIWA & Advanced Adaptive PSO. Computational results show that Advanced Adaptive based SVC controller is able to find better quality solution as compare to conventional PSO & PSO-SFIWA Techniques.

Index Terms— Small Signal Stability SMIB (single machine on infinite bus bar), SVC (Static Var Compensator), Particle Swarm Optimization (PSO), PSO with Shrinkage Factor & Inertia Weight Approach (PSO-SFIWA), PSO with Time Varying Acceleration Coefficients (PSO-TVAC)

I. INTRODUCTION

Small signal stability analysis is a study about the dynamic stability limits of synchronous machines resulted from poorly damped rotor oscillations under small perturbations or disturbances [1] [2]. These disturbances continuously occur on the power system because of small variation in load and generation. The perturbations are considered sufficiently small, in order to make system equations linearized. The dynamic performance of the machine can be improved by suitable control methods. The instability condition may result in two forms:

- a) Steady increase in rotor angle due to lack of sufficient synchronizing torque.
- b) Rotor oscillations of increasing amplitude due to the lack of sufficient damping torque [1].

In today's practical power system, small signal stability is a problem of insufficient damping of oscillations. Today, the fast acting static excitation

systems are available for the modern power system and they have a large impact on the small signal stability of power systems. At the range of 0.1 Hz to 2.5 Hz in electromechanical oscillation frequencies of the machine, these excitation systems introduce negative damping [2], so that they make the system unstable under local and inter area modes of oscillations. And when, we have a weak system with the weak tie lines, a small disturbance can also make the system unstable. So the dynamic behavior of the synchronous machine under small disturbance is studied to determine small signal stability & there is a need of positive damping in order that generators remain in synchronism under small disturbances. Recent advances in power electronics lead to the introduction of Flexible Alternating Current Transmission (FACT) devices. These devices are capable of improving the power system stability [1]. Tuning of control parameters of FACT devices is complex exercise. Various conventional techniques like gradient descent, Ziegler Nichols, mathematical programming & eigen value analysis are employed pertaining to their design problems. These conventional methods are time consuming, having slow convergence rate and during the search operation they might trap in local minima, due to which optimal solution may not be achieved [6]. Researchers study reveals that AI based FACTS devices are more efficient for power system stability than conventional power system stabilizers. Various computational intelligence techniques have been widely used in solving power system stability problems. Amongst the popular techniques are Evolutionary Programming (EP), Particle Swarm Optimization (PSO), Genetic Algorithm, Artificial Immune System (AIS), Bacterial Foraging Optimization and Hybrid BF-PSO. These algorithms are heuristic population-based search methods that used both random variation and selection. In this paper Advanced Adaptive PSO tuned PID damped SVC (Static Var Compensator) controller is proposed. The simulation results justify the optimization efficacy of the proposed Advanced Adaptive PSO over basic PSO & PSO-SFIWA.

The present paper is laid out as follows:

Section I - Introduction, section II describes about the Static Var Compensator (SVC). Section III gives the

review of work done by different researchers in various fields using a number of evolutionary techniques. Section IV gives a brief description of Particle Swarm Optimization Technique. Section V describes about PSO-SHRINKAGE Factor with time varying Inertia weight approach (PSO-TVIWA) for global search of Controller parameters. Section VI proposes the implementation of Advanced Adaptive (PSO-TVAC) Technique for optimizing the parameters of SVC Controller for better global search. Section VII describes the formulation of an objective function to be minimized. Section VIII describes the mathematical modeling of SVC Controller. Study system has been described in section IX. Section X gives the description of proposed PID-damped SVC Controller connected to an SMIB system. Simulation results have been shown in section XI and finally the conclusion in section XII.

II. STATIC VAR COMPENSATOR (SVC) DESCRIPTION

SVC's are first generation FACTS controllers which are expected to revolutionize power transmission in future. SVC provides an excellent source of rapidly controllable reactive shunt compensation for dynamic voltage control through its utilization of high-speed thyristor switching controlled devices. SVC is typically made up of coupling transformer, thyristor valves, reactors, capacitance (often tuned for harmonic filtering)[4]. They are used to control the voltage profile under load variations, increase power transfer capability and improve system stability[5]. They can be used for damping power system oscillations incorporating some auxiliary signals.

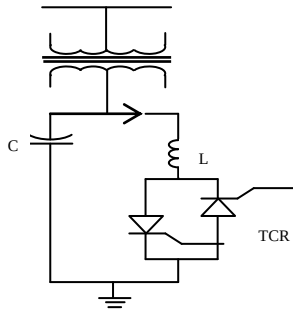


Fig. 1. Configuration of SVC Controller

III. ABOUT PAST STUDIES

A number of evolutionary techniques like BFO, GA, PSO, hybrid BF-PSO, BF with PSO biased evolution Ant Colony, water droplet optimization technique & many more has been proposed by many researchers [6-11]. Wang has applied advanced no velocity PSO for grid scheduling [12]. Attractive and Repulsive Particle Swarm Optimization has been used for orienteering problem in paper [13]. Rehab [14] has applied Genetically Improved PSO Algorithm for Efficient Data Clustering. Paper [7] has suggested centroid based PSO technique. Paper [15] has applied Hybrid PSO which is based on the fusion of the PSO, Clonal Selection Algorithm, and Mind

Evolutionary Computation. An Ant Colony Optimization based clustering algorithm was introduced by Shelokar [10]. Jin-Zhu Hu has explained the algorithm of PSO with dynamic inertia weight in paper [18]. Complex ELD problem has been solved by BF with PSO biased evaluation in paper reference [17].

IV. OVERVIEW OF PARTICLE SWARM OPTIMIZATION

PSO was introduced by Eberhart and Kennedy in 1995 [20]. It is a heuristic & stochastic based optimization technique. PSO can be used on optimization problem that are partially irregular, noisy, change over time, etc. It is developed from swarm intelligence and is based on the research of bird and fish flock movement behavior. The particle swarm optimization consists of swarm of particles which are initialized with a population of random candidate solution in the multidimensional search space. During their flying movement they follow the trajectory according to their own best flying experience (pbest) & best flying experience of the group (gbest). During this process, each particle modifies its position and velocity according to shared information to follow the best trajectory leads to optimum solution & this technique is simple & very few parameters need to be determined. The choice of PSO parameters can have a large impact on optimization performance. Selecting PSO parameters that yield good performance has therefore been the subject of much research. The PSO algorithm consists of just three steps, which are repeated until some stopping condition is met.

1. Initialize the number of particles, initial particle position & velocity

$$v_{id}^k = v_{min} + (v_{max} - v_{min}) * rand() \quad (1)$$

Where,

$$v_{max} = 0.1 * (m_{max} - m_{min})$$

$v_{min} = -0.1 * (m_{max} - m_{min})$, while m_{max} and m_{min} are maximum and minimum limits for the d-dimension search space.

$$x_{id}^k = m_{min} + (m_{max} - m_{min}) * rand() \quad (2)$$

Thereafter evaluate the fitness of each particle and define pbest & gbest.

2. Update velocity and position of each particle.

$$v_{id}^{k+1} = v_{id}^k + c1 * rand1(pbest_i^k - x_{id}^k) + c2 * rand2(gbest^k - x_{id}^k) \quad (3)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (4)$$

$c1$ and $c2$ represents the acceleration factors, and represents distributed random numbers between (0,1). First part of (1) depicts the previous velocity of the particle, the second part is a positive cognitive component & third part is a positive social component as described in [25].

3. Update individual (pbest) and global best (gbest) fitness and position.

v_{id}^k and x_{id}^k defines separately the velocity & position of the particle i at its k times iteration and the d -dimension quality of its position:

$pbest_{id}^k$ represents the d -dimension quality of the individual i at its best optimized position at its k times.

$gbest_d^k$ represents the d -dimension quality of the swarm at its best optimized position.

V. ABOUT PSO-SHRINKAGE FACTOR WITH TIME VARYING INERTIA APPROACH

The conventional PSO has the disadvantage of premature convergence & problem of stagnation. This can be overcome by incorporating additional factor called inertia weight factor w , concept of which is explained by Shi & Ehbhart [21] in order to achieve the balance between local & global search during the iteration process of an optimization problem. Inertia weight plays a vital role in controlling the exploitation and exploration capability of algorithm. For enhancement of performance index of conventional PSO algorithm, one can adjust the inertia weight w , the value of which is linearly reduced during the iterations. Typical values of w_{max} & w_{min} are 0.9 & 0.4 respectively. Since there is a need of vast random search in multi-dimensional search space initially & at later stage when problem is converging to best solution, fine tuning is required. Due to dynamic behaviour of w , this technique can be named as PSO-TVIWA as described in paper [20]. The inertia weight w is a function of iteration time and mathematically,

$$w = w_{max} - ((w_{max} - w_{min}) / iter_{max}) \cdot iter \quad (5)$$

This technique had some limitations of not converging towards global minima. In order to converge efficiently and to avoid explosion of the particle swarm, an additional new parameter called shrinkage factor/constriction factor K has been proposed by Clerc et al [23, 24]. Mathematical representation of updated particle velocity and position using shrinkage factor is given as follows:

$$v_{id}^{k+1} = S \left(w * v_{id}^k + c1 * rand1(pbest_i^k - x_{id}^k) + c2 * rand2(gbest^k - x_{id}^k) \right) \quad (6)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (7)$$

x_{id}^k is the previous position of the particle and v_{id}^{k+1} is the updated particle velocity.

Mathematically, Shrinkage factor has been represented as

$$S = \frac{2}{|2 - \theta - \sqrt{\theta^2 - 4\theta}|}$$

$$\theta = c1 + c2 \text{ and } \theta > 4$$

Normally, values of $c1$ and $c2$ lie between 0.5 to 2.05 for $c1=c2=2.05$, the computed value of $S=0.73$.

These values can be changed according to the optimization problem.

VI. IMPLEMENTATION OF ADVANCED ADAPTIVE (PSO-TVAC) TECHNIQUE FOR SVC CONTROLLER

This strategy has the advantage of both fast convergence as well as fine tuning to global minima [26]. Both cognitive part and social part of velocity plays an important role in optimising to global minima. In order to increase the randomness at the initial stage, value of $c1$ is made larger than $c2$. This allows particle to search the optima in the whole m -dimensional search space. Lower value of $c1$ can lead to trapping into local minima due to premature convergence. Whereas, higher value of $c2$ than $c1$ is desired at later stage to enhance the social interaction and hence converges efficiently to global minima. It is mentioned here that this algorithm is used with time varying inertia weight factor. Mathematically updated particle velocity and position is represented as:

$$v_{id}^{k+1} = K \left(w * v_{id}^k + c1' * rand1(pbest_i^k - x_{id}^k) + c2' * rand2(gbest^k - x_{id}^k) \right) \quad (8)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (9)$$

where modified $c1'$ and $c2'$ are mathematically represented as [25, 27]:

$$c1' = ((c1f - c2i) * k / K_{max}) + c1i, \quad c1f < c1i$$

$$c2' = ((c2f - c1i) * k / K_{max}) + c2i, \quad c2f > c2i$$

$c1i$, $c2i$, $c1f$ and $c2f$ are initial and final values of $c1$ and $c2$ respectively. Normally $c1$ varies from 2.5 to 0.5 over the full range of search.

VII. FORMULATION OF AN OBJECTIVE FUNCTION

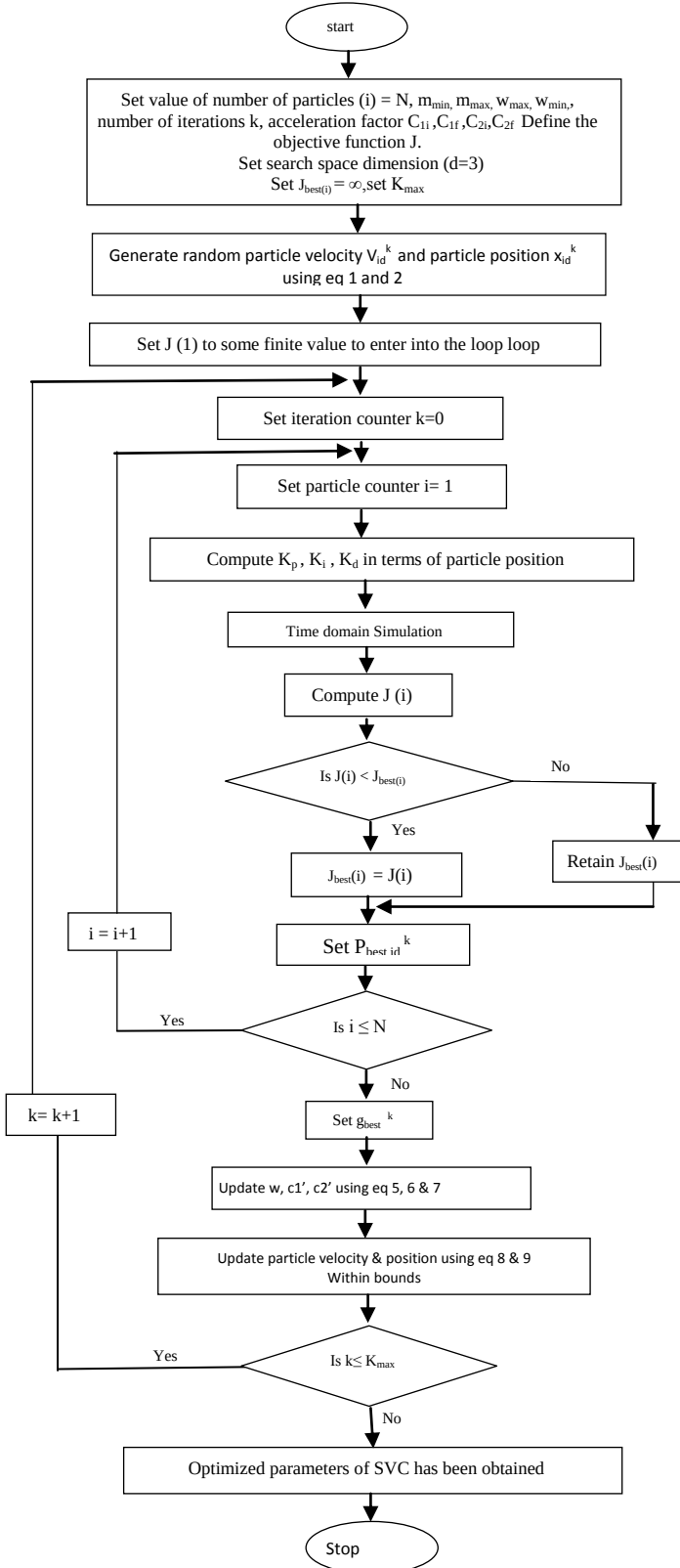
Objective function formulated is based on the optimization parameters. It is worth mentioning that the PID damped SVC controller is designed to minimize the power system oscillations after a disturbance so as to improve steady state stability. These oscillations are reflected in the deviations in the generator rotor speed Δw and deviation in terminal voltage Δv_t . In the present study the objective function J is formulated as the minimization of

$$J = \int_0^t [t(\Delta\omega(t, x))^2 + t(\Delta v_t(t, x))^2] dt$$

In the above equations, $\Delta w(t, x)$ denotes the rotor speed deviation and $\Delta v_t(t, x)$ denotes the variation in terminal voltage for a set of controller parameters x and (note that, here x represents the parameters to be optimized i.e k_p , k_i and k_d , the parameters of PID controller) and t is the time range of the simulation. With the variation of the parameters x , $\Delta w(t, x)$ and $\Delta v_t(t, x)$ will also be changed. For objective function calculation, the time-domain simulation of the power system model is

carried out for the simulation period. It is aimed to minimize this objective function in order to improve the system response in terms of the settling time and overshoots.

A. Flow chart of proposed algorithm for controller design



VIII. MATHEMATICAL MODELLING

A. Machine Modeling

The synchronous generator is represented by 1.1 model, i.e. with field circuit and one equivalent damper winding on q axis. The machine equations are [2, 3]:

$$\frac{d\delta}{dt} = \omega_B (S_m - S_{m0})$$

$$\frac{dS_m}{dt} = \frac{1}{2H} [-D(S_m - S_{m0}) + T_m - T_e]$$

$$\frac{dE'_q}{dt} = \frac{1}{T'_{do}} [-E'_q + (x_d - x'_d)i_d + E_{fd}]$$

$$\frac{dE'_d}{dt} = \frac{1}{T'_{qo}} [-E'_d + (x_q - x'_q)i_q]$$

The electrical torque is expressed in terms of variables E'_d, E'_q, i_d and i_q as:

$$T_e = E'_d i_d + E'_q i_q + (x_d - x'_d)i_d i_q$$

For a lossless network, the stator algebraic equations and the network equations are expressed as:

$$E'_q + x'_d i_d = v_q$$

$$E'_d - x'_q i_q = v_d$$

$$v_q = -x_e i_d + E_b \cos \delta$$

$$v_d = x_e i_q - E_b \sin \delta$$

Solving the above equations, the variables i_d and i_q can be obtained as:

$$i_d = \frac{E_b \cos \delta - E'_q}{x_e + x'_d}$$

$$i_q = \frac{E_b \sin \delta - E'_d}{x_e + x'_d}$$

The notation of above parameters and variable used are defined in nomenclature [1] [2] [3].

B. Dynamic Modeling of SVC

The basic SVC model [22] comprises of a measurement module, voltage regulator module & thyristor susceptance control module for simulation of SVC control loop stability & SVC response time. This model is a fundamental frequency model & suitable for continuous control of SVC. SVC has a reactive/current/susceptance regulator so that SVC can return back to a desired steady state operating point.

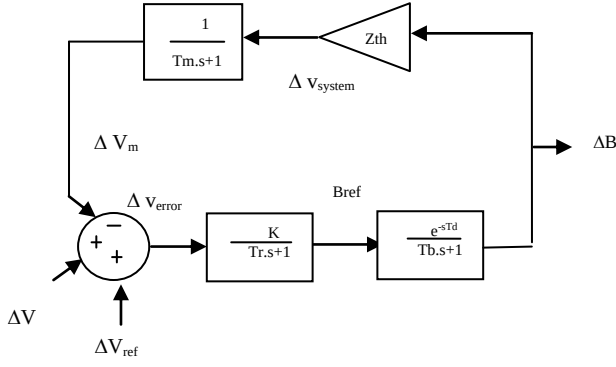


Fig. 2. Model of Static Var Compensator

1. Measurement Block

The line voltage & current measured are converted into quasi-dc control signal & is proportional to the magnitude of fundamental frequency component of the measured parameter & the resulting model is a single low frequency filter

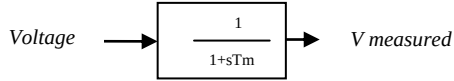


Fig. 3 Measurement module

2. Voltage Regulator Block

It is basically a proportional type regulator where K_R is the gain of the regulator & is reciprocal of the slope setting & which can be set between 5% & 1% on the SVC base. T_R is the regulator time constant which is taken between 20 & 150 ms. PI controller is being used in some voltage regulator circuit. The generic AVR control block is defined by the transfer function:

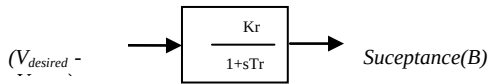


Fig. 4 Voltage regulator block

3. Thyristor Susceptance Control Module

The CIGRE model showing Thyristor firing delays is being used in this module where T_d is gating transport delay & T_b is effect of Thyristor firing susceptance control (3-6 ms). B_{ref} & TCR firing angle has nonlinear relationship [22]. The transfer function of thyristor susceptance control is:

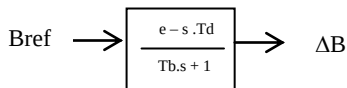


Fig. 5 Thyristor susceptance control block

The Thyristor susceptance block gives the incremental change in the susceptance value when a firing angle delay is given to it. The Z_{th} is the thevenin impedance of SVC controller, which is generally specified as a constant. SVC with additional PID controller is shown in Fig.7 is used, as this introduces an additional damping in the system and damps the rotor mechanical low frequency oscillations quickly. They are placed in the supplementary control signal of SVC.

IX. STUDY SYSTEM

In this paper, the proposed study system consists of four 555MVA, 24Kv, 60 Hz thermal units which are connected to infinite bus system through two parallel transmission lines. The generators are to be modeled as a single equivalent generator represented by the classical model with all the parameters expressed in p.u on 2220 MVA, 24 Kv base. The resistance of the transmission line is neglected in this paper. The linearized mathematical modeling of single machine power system is carried out by linearizing the equations around the operating point [3]. Phillips Heffron model, a linearized model of SMIB system is considered for simulation. SVC is placed at the generator busbar. The machine data is given in the appendix. The network reactances are in p.u on 2220MVA, 24 Kv base. The operating conditions are: $P_g=0.5$ p.u, $V_t=1$ p.u and $E_b=0.995$ p.u. The initial conditions are computed from the operating points [3] and K_1 to k_6 constants of Phillips Heffron model are then derived using the expressions given in the appendix. A small perturbation of 0.1 p.u is then given to examine the steady state stability.

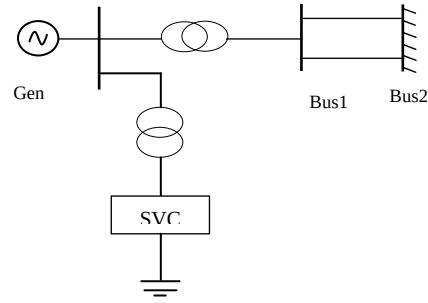


Fig. 6 Single machine infinite bus with SVC

X. PID DAMPED SVC CONTROLLER

PID damped SVC controller shown in Fig.7 introduces an additional damping in the system and damps the rotor mechanical low frequency oscillations quickly. They are placed in the supplementary control signal of SVC.

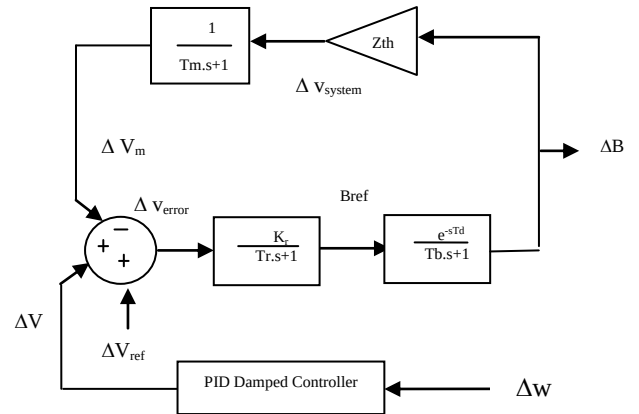


Fig. 7. PID Damped SVC controller

PID damped SVC controller placed at the generator bus is used to improve the dynamic response as well as to

reduce or eliminate the steady-state error. Unified Phillips Heffron model with PID damped SVC controller is shown in fig. (8). The parameters of PID controller of SVC are optimized firstly by a conventional method and then using these computed parameters as base, the control parameters are computed by PSO-SFIWA and by PSO-TVAC Techniques

XI. SIMULATION

Simulation is carried out both in MATLAB coding & MATLAB Simulink by giving small disturbance $\Delta T_m = 0.1$ p.u. Four cases have been analyzed:

1. Without SVC controller.
2. With SVC controller.
3. With PSO-SFIWA based PID damped SVC Controller.
4. With PSO-TVAC based PID damped SVC Controller.

The simulation results of the system for all the cases mentioned above with respect to variation in rotor angle, rotor speed deviation & change in terminal voltage are shown in fig. 9 (a) to 9 (i).

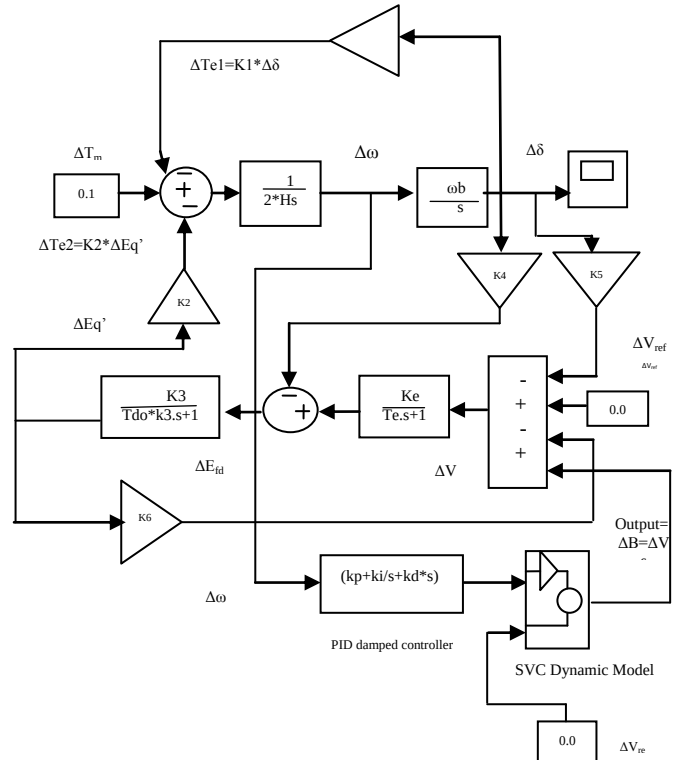


Fig. 8. Phillips Heffron model of SMIB system with PID damped SVC controller

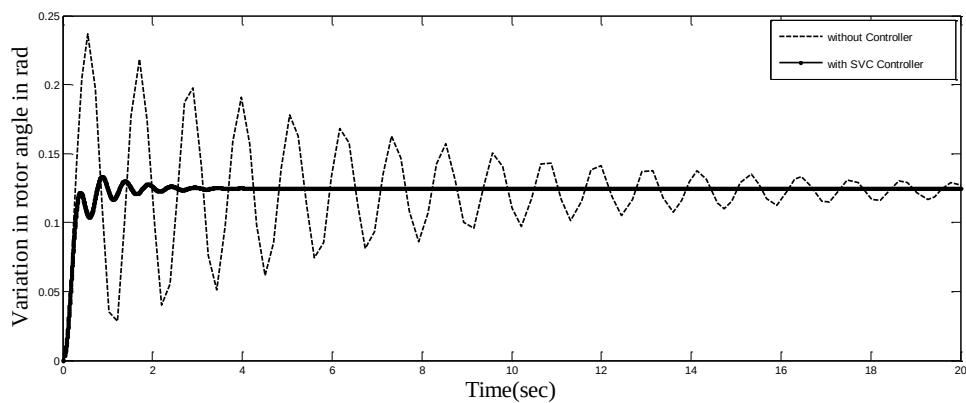


Fig.9a Time Response of Variation in rotor angle

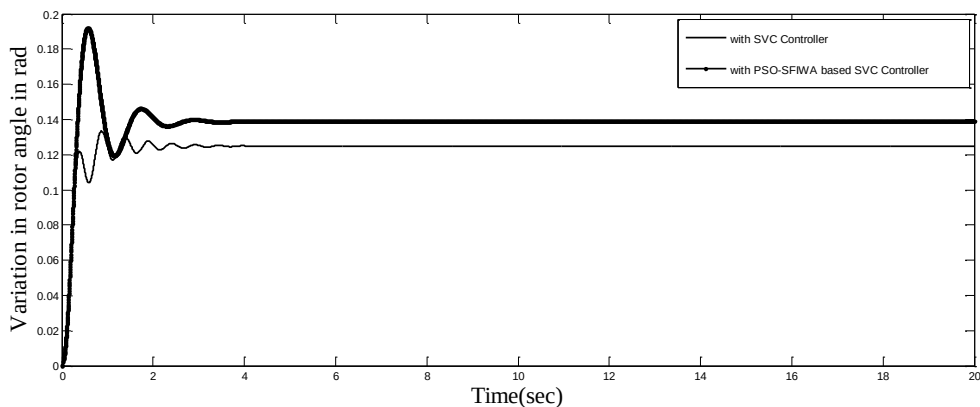


Fig. Fig. 9b Time Response of Variation in rotor angle

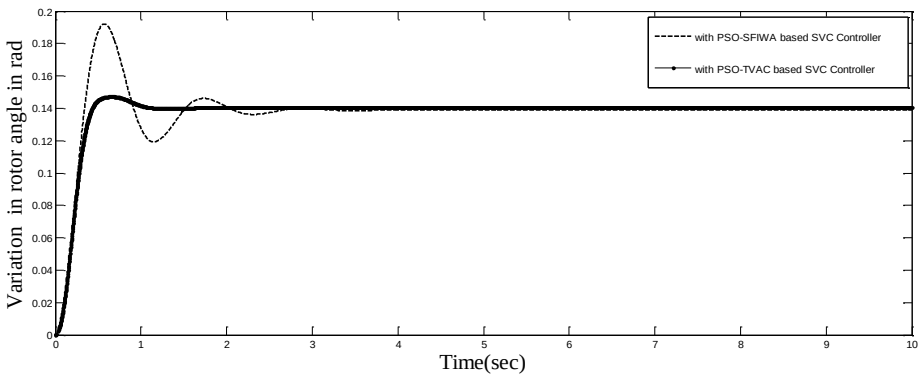


Fig. 9c Time Response of Variation in rotor angle

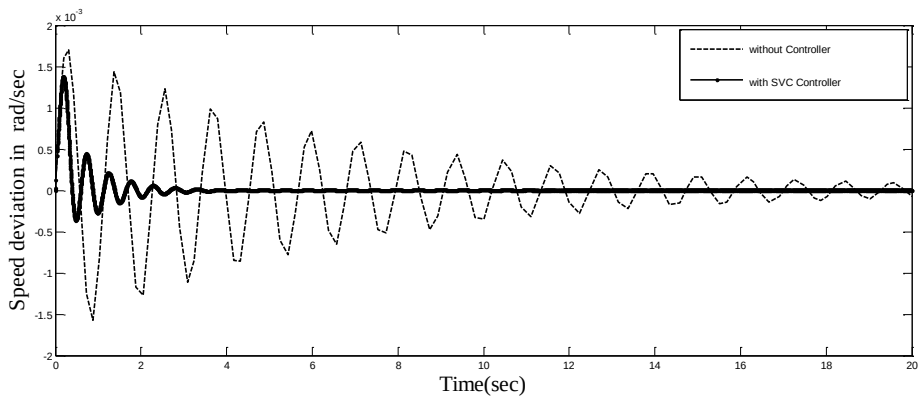


Fig. Fig. 9d Time response of speed deviation

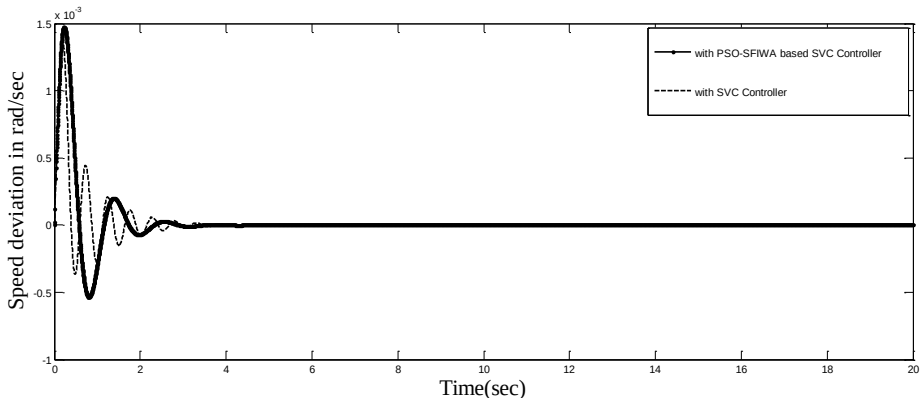


Fig. 9e Time response of Speed deviation

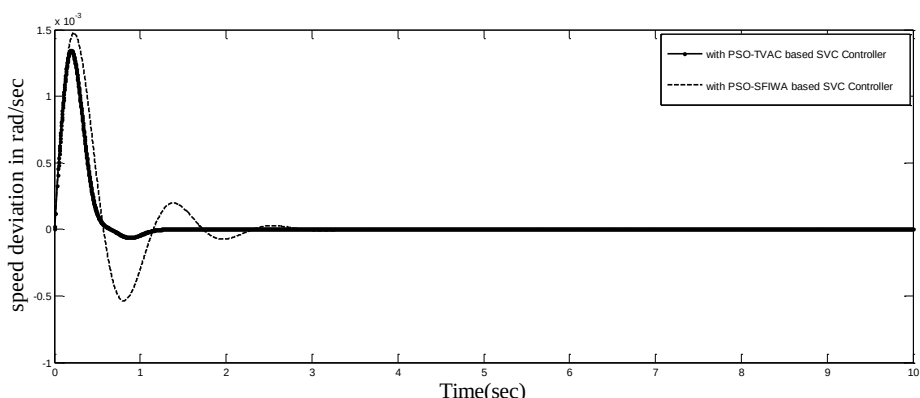


Fig. 9f Time Response of speed deviation

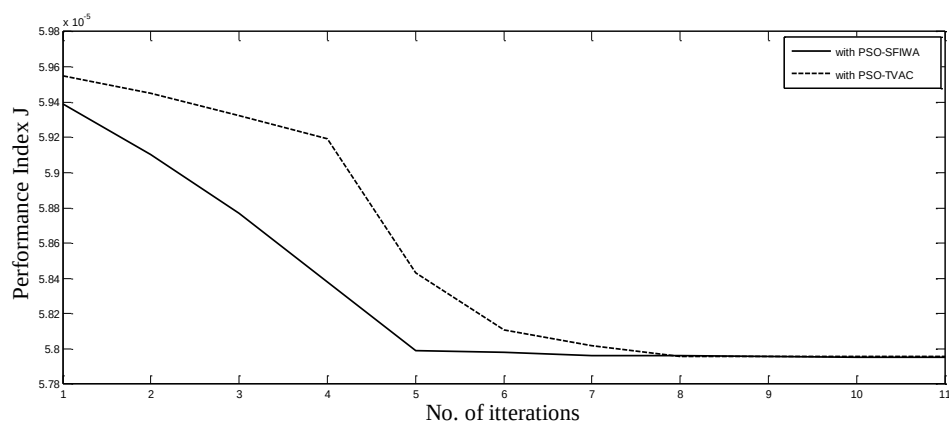
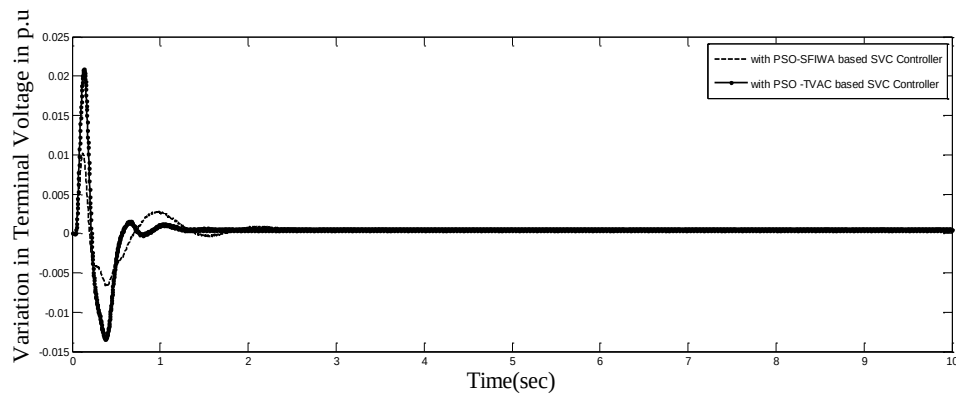
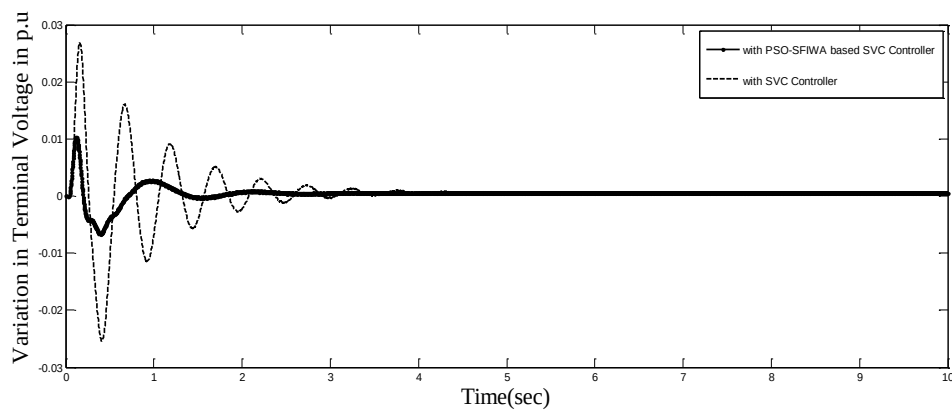
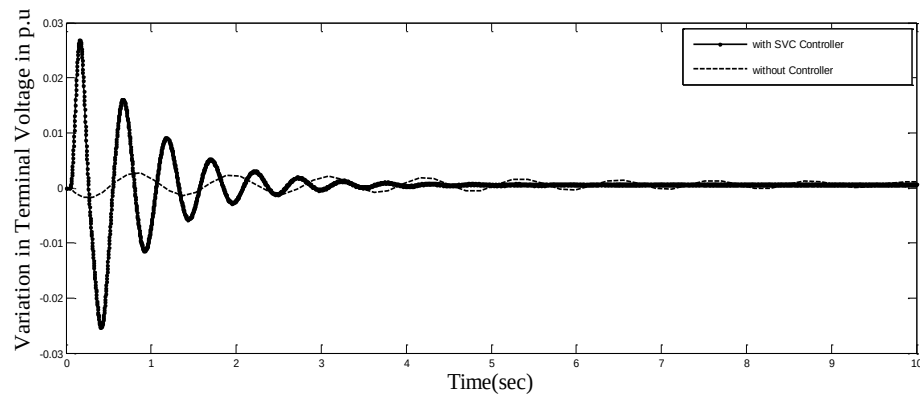


Table 1. Comparison of settling time of oscillations

Cases	$\Delta \delta$	$\Delta \omega$	Δv_t
-	Settling time	Settling time	Settling time
Without SVC Controller	57sec	56sec	55sec
With SVC Controller	5.3sec	5.4sec	7sec
Parameters of PID damped SVC Controller using PSO-SFIWA approach.	4.2sec	4sec	4sec
Parameters of PID damped SVC Controller using PSO-TVAC approach	1.6sec	1.7sec	1.8sec

XII. CONCLUSION

AI based PID damped SVC controller is highly effective in improving the dynamic stability of power system i.e. they damp out the rotor mechanical low frequency oscillations effectively. Hence, they can be utilized to replace the existing conventional power system stabilizers in future and can be applied for the real-time control of power system. The simulation results show Advanced Adaptive PSO is more effective than basic PSO and PSO-SFIWA. However the system gets stabilized to a new value of machine angle, but stabilizing the rotor angle to a new position is not of much concern. The main concern is to damp out the speed oscillations quickly to enhance the steady state stability of power system.

APPENDIX

Generator Data:

$x_d=1.6$, $x_q=1.55$, $x_d'=0.32$, $T_{d0}'=6$ sec, $H=5.0$, $f=60$ Hz, $x_e=0.4$, $D=0$

Expressions for computing Phillips Heffron model constants:

$$K1 = \frac{E_b E_{qo} \cos \delta_o}{(x_e + x_q)} + \frac{(x_q - x_d)}{(x_e + x_d')} E_b i_{qo} \sin \delta_o$$

$$K2 = \frac{x_e + x_q}{(x_e + x_d')} i_{qo} = \frac{E_b \sin \delta_o}{(x_e + x_d')}$$

$$K3 = \frac{x_e + x_d'}{x_d + x_e}$$

$$K4 = \left(\frac{x_d - x_d'}{(x_d' + x_e)} \right) E_b \sin \delta_o$$

$$K5 = \frac{-x_q v_{do} E_b \cos \delta_o}{(x_e + x_q) V_{to}} - \frac{x_d' v_{qo} E_b \sin \delta_o}{(x_e + x_d') V_{to}}$$

$$K6 = \frac{x_e}{(x_e + x_d')} \left(\frac{v_{qo}}{V_{to}} \right)$$

Exciter data:

$K_E=200$, $T_E=0.05$ sec

SVC Controller data:

$K_R = 80$, $T_R=0.025$ sec, $T_d = 0.0371$ sec, $T_b=0.0375$ sec
 $Z_{th} = 0.02$ p.u, $T_m = 0.000795$ seconds

PSO Parameters:

Population=50, $c1=2.05$, $c2=2.05$, $itermax=10$, $c1i=2.5$, $c2f=0.5$, $c2i=0.5$, $c2f=2.5$, $w_{max}=0.9$, $w_{min}=0.4$, $k_{min}=[0.5 -0.5 0.01]$, $k_{max}=[0.7 -0.3 0.02]$

Table 2. PID controller parameters:

CASES	k_p	k_i	k_d
Without SVC Controller	----	----	----
With SVC Controller	----	----	----
Parameters of PID damped SVC Controller using PSO-SFIWA approach.	0.3171	-.459	.0147
Parameters of PID damped SVC Controller using PSO-TVAC approach	0.6969	-0.5	0.02

REFERENCES

- [1] N. G. Hingorani, L. Gyugyi, "Understanding FACTS", IEEE Press, 2001.
- [2] P. Kundur, "Power System Stability and Control", Tata McGraw-Hill, 2006.
- [3] K. R. Padiyar, *Power System Dynamics Stability and Control*, BS Publications, 2nd Edition, Hyderabad, India, 2002
- [4] Y. H. Song, A. T. Johns, "Flexible AC Transmission Systems (FACTS)", IET, 2009
- [5] Poonam Singhal, S.K. Aggarwal and Narender Kumar, 'Transient Stability Enhancement using UPFC.' IREMOS, 2013.
- [6] Poonam Singhal, S.K. Aggarwal, Narender Kumar & Monika, "Stability Enhancement of a multi-machine power system using Static Var Compensator" IJAIR, volume 2, issue 2, February 2013.
- [7] Y. Mishra, S. Mishra and Fangxing Li 'Coordinated Tuning of DFIG-Based Wind Turbines and Batteries Using Bacteria Foraging Technique for Maintaining Constant Grid Power Output'. Published in Systems Journal, Volume:6, Issue: 1, pp.no.16-26, 2012 IEEE
- [8] Shengli Song Li Kong and Jingjing Cheng, "A Novel Particle Swarm Optimization model with Centroid and its application" I.J. Intelligent Systems and Applications, 2009, 1, pp.no. 42-49
- [9] X. Wang, X. Z. Gao and S. J. Ovaska, "A Hybrid Particle Swarm Optimization Method", IEEE International Conference on Systems, Man, and Cybernetics, Taipei, Taiwan, October 2006

- [10] Mohamed Zellagui and Abdelaziz Chaghi, "Impact of TCSC on Distance Protection Setting based Modified Particle Swarm Optimization Techniques". I.J. Intelligent Systems and applications 2013,06, pp.no.12-24.
- [11] P. S. Shelokar, V. K. Jayaraman, and B. D. Kulkarni, 'An ant colony approach for clustering,' *Analytica Chimica Acta* Vol. 509(2): pp.187-195, 2004
- [12] Shah-Hosseini, H. (2009) 'The intelligent water drops algorithm: a nature-inspired swarm-based optimisation algorithm', *Int. J.Bio-Inspired Computation*, Vol. 1, Nos. 1/2, pp.71-79.
- [13] Wang Meihong, Zeng Wenhua*, Wu Keqing. Grid task scheduling based on advanced no velocity PSO'.
- [14] Herby Dallard and Sarah S. Lam 'Solving the Orienteering Problem Using Attractive and Repulsive Particle Swarm Optimization', *Proceedings in IEEE International Conference on Information Reuse and Integration, IRI 2007*, pp. no. 13-15, Las Vegas, Nevada, USA. Heuristic Optimization and Search
- [15] Rehab F. Abdel Kader," Genetically Improved PSO Algorithm for Efficient Data Clustering", *Second International Conference on Machine Learning and Computing (icmlc)*, pp.71-75, 2010.
- [16] Wael M. Korani, Hassen Taher Dorrah, Hassan M.Emara,"Bacterial Foraging Oriented by Particle Swarm Optimization strategy for PID Tuning".
- [17] Jiao Wei, Liu Guang-bin," An improved Particle Swarm Optimization Algorithm with Immunity". Published in *Intelligent Computation Technology and Automation*, 2009. ICICTA '09. Second International Conference on 10-11 Oct. 2009.
- [18] Hardiansyah, Junaidi, Yohannes Load Dispatch using Bacterial Foraging Technique with Particle Swarm Optimization Biased Evolution". I.J. Intelligent Systems and applications 2012,12, pp. 12-18
- [19] JinJin-Zhu Hu,Jia-qiao Wang, Ting Xu," Research on Particle Swarm Optimization with dynamic inertia weight'. Published in 2010 International Conference on Computer Design and Applications (ICCD 2010).
- [20] D.Harikrishna, N.V.Srikanth,"Unified Phillips-Heffron Model of Multi Machine Power System equipped with PID Damping Controlled SVC for Power Oscillation Damping." Published in *India Conference (INDICON)*, 2009 Annual IEEE, 18-20 Dec. 2009
- [21] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc .IEEE Int. Conf. Neural Networks*, 1995, pp. 1942-1948.
- [22] Y.Shi and Russel Eberhart," Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization". Published in: *Evolutionary Computation*, *Proceedings of the 2001*
- [23] IEEE special stability controls working group, Static VAR compensator models for power flow and Dynamic performance simulation. *IEEE Transaction on power systems*, Vol. 19, No. 1, February 1994, pp. 229 - 239.
- [24] Clerc, "The swarm and the queen: Toward a deterministic and adaptive particle swarm optimization," in *Proc. IEEE Int. Congr .Evolutionary Computation*, vol. 3, 1999, p. 1957.
- [25] M. Clerc and J.Kennedy, "The particle swarm—Explosion, stability, and convergence in a multi-dimensional complex space," *IEEE Trans. Evol .Comput.*, vol. 6, pp. 58-73, Feb. 2002.
- [26] Asanga Ratanvweera, Saman K. Halgamuge,Harry C.watson, "Self Organising Hierarchical Particle Swarm Optimizater with Time-Varying Acceleration

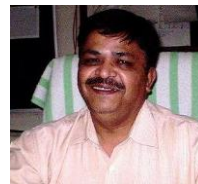
Coefficient."IEEE Transaction on Evolutionary Computation, 2004

- [27] Rajendraprasad Narne and Prafulla Chandra Panda,"Coordinated Design PSS of with Multiple FACTS Controllers using Advanced Adaptive PSO". *International Journal on Electrical Engineering and Informatics - Volume 5,Number 3, September 2013*

Author's Profile



Poonam Singhal was born in 1964. She is currently Associate Professor in YMCA University of Science & Technology, Faridabad; Haryana (India). She did her B.Sc Engg from REC Rourkela, Odhisa (India) & M.Tech (Power System & Drives) from YMCA University of Science & Technology, Faridabad, Haryana (India). Her main area of interest is FACTS, Power System Operation control & Simulation.



S. K. Agarwal was born in 1961. He is currently Professor in Department of Electronics Engineering; YMCA University of Science & Technology, Faridabad, and Haryana (India). He did his B.Tech from REC Calicut, Kerela (India) & M.E. (Controls & Instrumentation) from Delhi Technological University Delhi (India) & Ph.D from Jamia Milia Islamia University Delhi (India). Prof. S. K. Agarwal has many publications in National/International Journals & presented papers in International conferences. Prof. Agarwal's main area of interest is Power System & Controls.



Narender Kumar was born in Aligarh (India). He received his B.Sc. Engg. & M.Sc. Engg in Electrical Engineering from A.M.U. Aligarh in 1984 & 1986 respectively. Presently, he is Professor of Electrical Engg. in Delhi Technological University, Delhi. Prof. Narender Kumar has published various papers in national and international conferences and journals. Prof. Narender Kumar's main area of interest are Power System Operation Control & Simulation, SSR, AGC etc.

Ag-GaAs-Ag quantum nano-lenses system for plexcitonic interaction

Venus Dillu and R K Sinha

TIFAC- Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Department of Applied Physics, Delhi Technological University (Formerly Delhi College of Engineering) Bawana Road, Delhi-110042, India
dr_rk_sinha@yahoo.com

Abstract: We propose plexcitonic system formed by multi-layered silver-gallium arsenide-silver (Ag-GaAs-Ag) quantum nano-lenses and study optical properties using finite difference time domain method. Plane wave expansion method reveals hybrid band-structure. Cavity defect support strongly confined mode.

OCIS codes: (250.5403) Plasmonics; (240.6680) Surface plasmon.

1. Introduction

Advancement in nanotechnology has allowed miniaturization of metal-dielectric particles to nano dimensions which leads to extraordinary optical response unlike microstructures. System of hybrid metal dielectric nanostructures provides a large number of permutations that can be explored for interesting tunable electromagnetic properties [1-5]. Interaction between plasmons and excitons from semiconductor quantum dots (QDs) give rise to plexcitons which results in enhanced emission, allows controlled energy transfer and can also modify spontaneous emission in QDs. Such structures are capable of manipulating light at nanoscales but require extensive research to be able to interpret unique properties of the metal-dielectric at nanoscale. In this paper we investigate the silver-gallium arsenide-silver (Ag-GaAs-Ag) quantum nano-lens sandwiched plexcitonic system for its optical response.

2. Ag-GaAs-Ag hybrid quantum nanodisks system

The proposed plexcitonic system comprises of an array of layered Ag-GaAs-Ag in hexagonal arrangement with GaAs quantum nano-lenses sandwiched between the two layers of Ag nano-lenses as shown in Figure 1 (a). The three layered system of lenses touch at the convex edge and are arranged in hexagonal lattice with lattice constant $a = 100$ nm and radius $r = 40$ nm. This metal-semiconductor-metal system of quantum nano-lenses excite plexcitonic coupling and can have important applications in lasing, nano-antennae, Fano devices etc.

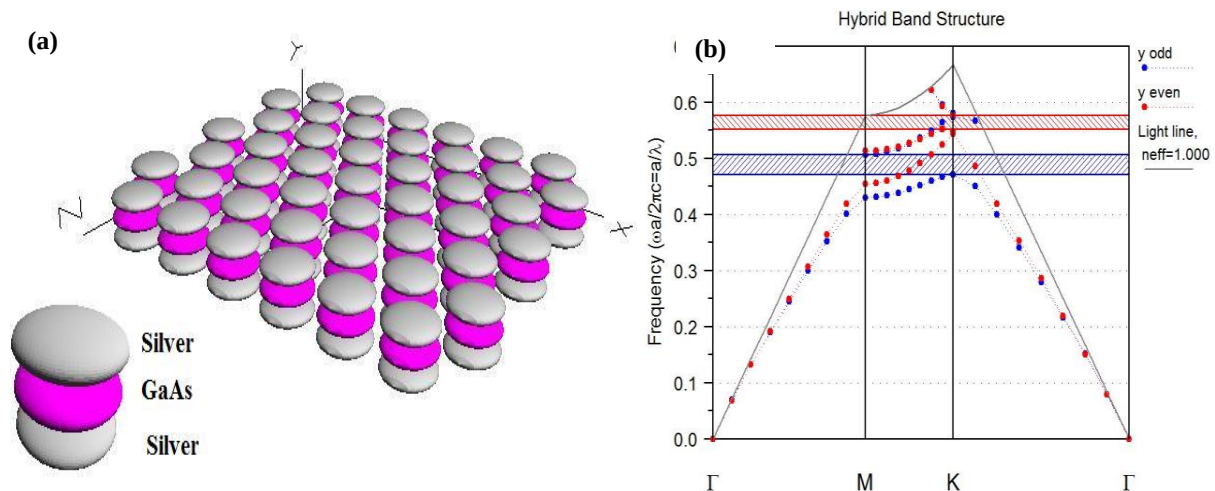


Figure 1 (a). Three layered system of Ag-GaAs-Ag nano-lenses in hexagonal arrangement. (b) Hybrid band structure of the proposed plexcitonic system with odd parity for TE gap and even parity for TM gap.

3. Result and Discussion

We study the proposed system by using 3D plane wave expansion (PWE) method to get hybrid band structure as given in Figure 1 (b). It is observed from Figure 1 (b) that the system offers hybrid band structure giving transverse

electric (TE-odd parity) and transverse magnetic (TM-even parity) gap in ultraviolet regime of the electromagnetic spectrum. Further, to investigate its optical response, we employ finite difference time domain (FDTD) method.

We introduce a cavity defect at the center of the system. Transmission characteristics of the cavity are obtained as given in Figure 2 with the monitor value representing the transmission in the graphs. It is observed that for TM wave (Figure 2 (a)) the peak value of the cavity is 0.4 (blue curve) with respect to the source with peak value of 0.6 (green curve in *inset*), hence resonating with more than 66%. Contrary to this we observe that for TE wave the peak intensity is very low (Figure 2 (b)), that is of the order of 10^{-10} and is negligible. Although the hybrid band structure (Figure 1 (b)) reveal gap for TE wave as well, but due to the presence of silver metal nano-lenses, the plexcitonic system enhances the TM mode which is the characteristic property of plasmonic system. This observation was verified by using the optimized value for the input of 193nm (Figure 2 (c)) yielding sharp rise for the output signal (blue) with respect to the input (green peak in the inset of Figure 2 (c)) i.e. the system increases the output 10 times. The output signal shifts with respect to the input due to the plexcitonic coupling. The 3D contour for electric field E_y , for the cavity defect is given in Figure 2 (d) showing strongly confined intense cavity resonance.

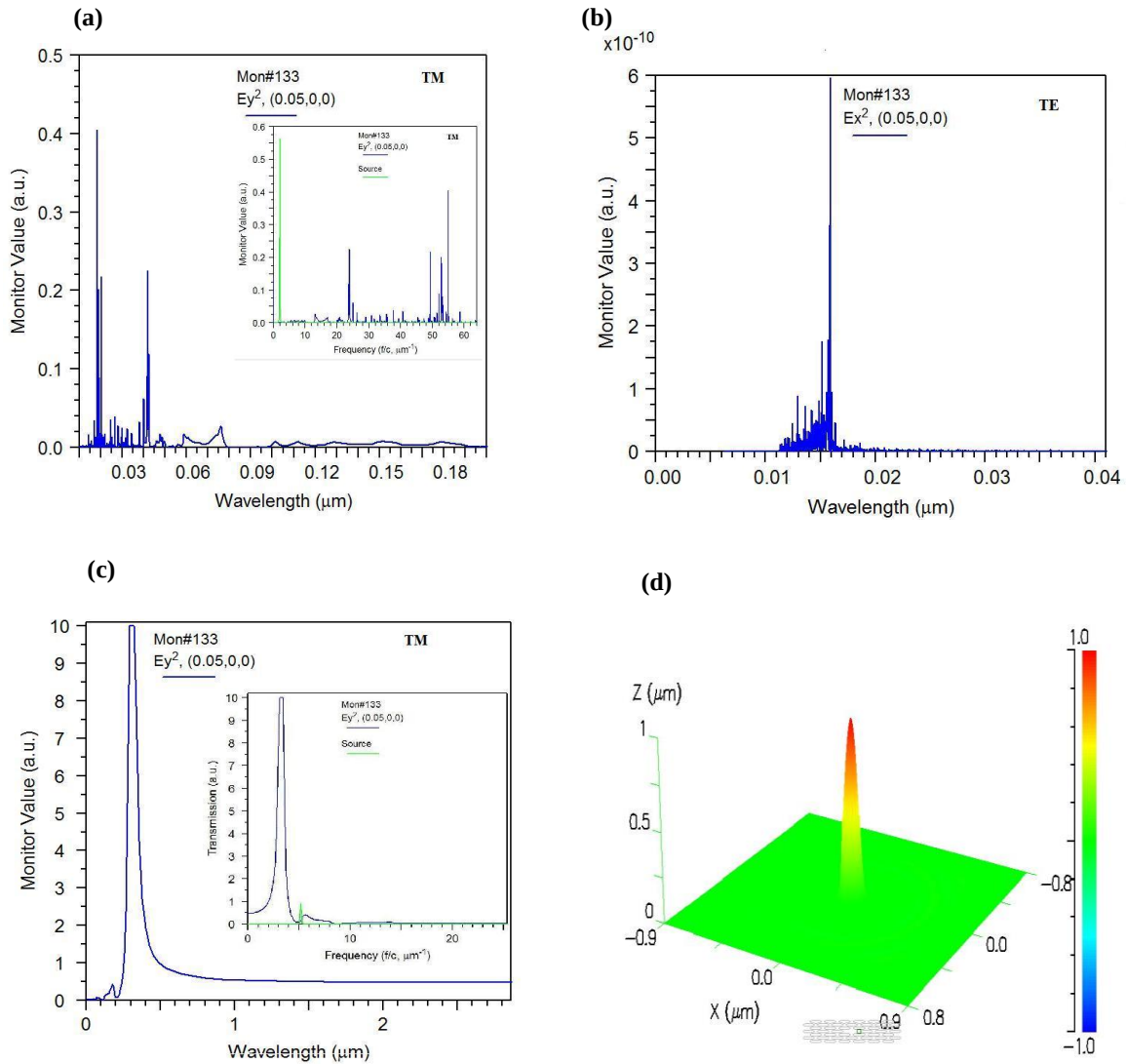


Figure 2 (a) and (b) Transmission characteristics of the Ag-GaAs-Ag nano-lenses plexcitonic system for TM and TE waves respectively (Monitor value represents transmission). (c) Strongly resonating peak is obtained for the input of 193nm with enhanced signal with respect to the input. (d) 3D contour for electric field E_y , for the cavity defect showing strongly confined intense cavity resonance

5. Conclusion

The proposed multi-layered Ag-GaAs-Ag quantum nano-lenses plexcitonic system results in strongly confined and enhanced resonance for cavity defect. Hence the system has applications in lasing, switching, sensing etc. As the cavity mode resonates in ultra-violet regime, it can be an alternative for excimer lasers.

Acknowledgement

The authors gratefully acknowledge (i) the initiatives and support towards establishment of “TIFAC Centre of Relevance and Excellence in Fiber Optics and Optical communication at Delhi College of Engineering, Delhi” through Mission Reach program of Technology Vision2020, Government of India and (ii)UGC sponsored major research project in the area of Photonic Crystal Fibers for sensing and telecom applications

References

- [1] A. E. Schlather, N. Large, A. S. Urban, P. Norlander and N. J. Halas, “Near field mediated plexitonic coupling and giant Rabi splitting in individual metallic dimmers,” *Nano Letters* **13**, 3281-3286 (2013).
- [2] J. B. Li, N. C. Kim, M. T. Cheng, L. Zhou, Z. H. Hao and Q. Q. Wang, “Optical bistability and nonlinearity of coherently coupled exciton plasmon system,” *Optics Express* **20**, 1856-1861 (2012).
- [3] D. Melnikau, D. Savateeva, A. Susha, A.L. Rogach and Y. P. Rakovich, “Strong plasmon-exciton coupling in a hybrid system of gold nanostars and J-aggregates,” *Nanoscale research letters* **8**, 1-6 (2013).
- [4] B. G. DeLacy, W. Qiu, M. Soljacic, C. W. Hsu, O. D. Miller, S. G. Johnson and J. D. Joannopoulos, “Layer-by-layer self-assembly of plexcitonic nanoparticles,” *Optics Express* **21**, 19103-19112 (2013).
- [5] A. Manjavacas, F. J. G. de Abajo and P. Norlander, “Quantum plexcitonics: Strongly interacting plasmons and excitons,” *Nano Letters* **11**, 2318-2323 (2011).

All Pass Network based MSO using OTRA

Rajeshwari Pandey¹, Neeta Pandey², Romita Mullick³, Sarjana Yadav⁴, Rashika Anurag⁵

^{1,2,3,4} Dept. of Electronics and Communication Engineering, Delhi Technological University,
Delhi -110042, India

⁵Department of Electronics and Communication Engineering, JSS Academy of Technical
Education, C-20/1 Sector 62, NOI DA, Uttar Pradesh, 201301, India

¹rajeshwaripandey@gmail.com, ²n66pandey@rediffmail.com,
³romitamullick.93@gmail.com, ⁴sarjanayadav@gmail.com, ⁵raashika@rediffmail.com

Abstract. This paper presents multiphase sinusoidal oscillators (MSO) using Operational Trans-Resistance Amplifier (OTRA) based allpass networks. Both even and odd phase oscillations of equal amplitudes which are equally spaced in phase can be produced using single allpass section per phase. The propose MSOs provide voltage output and can readily be used for driving voltage input circuits without increasing component count. The effect of nonideality of OTRA on the circuit performance is also analysed. The functionality of the proposed circuit is verified through PSPICE simulations.

1. INTRODUCTION

Multiphase sinusoidal oscillators (MSO) provide multiple outputs of same frequency, equally spaced in phase and find extensive application in the field of communications, instrumentation and power electronics. In Communications MSO circuits are used in single-sideband generators, phase modulators and quadrature mixers [1]. Selective voltmeters and vector generator are common applications of MSOs in the field of instrumentatation [2]. In power electronics three-phase MSOs are frequently utilized in for PWM converters [3] and inverters [4].

A large number of MSO realizations using various analog building blocks (ABB) [2, 5-25] are available in literature. These MSO are based on a basic design philosophy of forming closed loop using n ($n \geq 3$) cascaded phase shifting networks thereby producing n equally spaced phases. For phase shifting either first order low pass networks (LPNs) [5-7, 9-21, 24,25] or first order all pass networks (APNs) [2,8,22,23] are used. These reported structures provide either voltage [5-16] or current [2,17-24] outputs.

The MSOs of [5-8] are realized using operational amplifiers (op-amps). However due to constant gain-bandwidth product and lower slew rate of the op-amps, their high frequency operations are limited. Additionally the active R implementations of [5] and [6] lack tunability as these strucures make use of the opamp's parasitic capacitance. The Current Feedback Operational Amplifier (CFOA) based MSO structure [9] is capable of producing high frequencies but requires an accessible compensation terminal of a CFOA. The MSOs of [10, 11] are OTA based electronically tunable structures, however, provide voltage output at high impedance making a buffer necessary to drive the voltage input circuits. In addition, for both the

structures the output voltage is temperature sensitive too. The MSO configurations presented in [12-15] are CC based designs and provide voltage output at high impedance. Further the MSO of [14] being active R structure, lacks tunability. Three topologies of OTRA based MSOs are proposed in [16] and are designed using LPNs. The structures proposed in [2, 17-25] provide current outputs which need to be converted back to voltage for circuits requiring voltage inputs, which would considerably increase the component count. A detailed comparison of these structures is given in Table 1 which suggests that OTRA based MSO is most suitable choice for voltage output.

Table 1 Detailed comparison of available MSOs

Reference	ABB	Phase shift network	Output	Output Impedance	Tunability
[2]	CCCDTA	APN	Current	High	Yes
[5]	Opamp	LPN	Voltage	Low	No
[6]	Opamp	LPN	Voltage	Low	No
[7]	Opamp	LPN	Voltage	Low	Yes
[8]	Opamp	APN	Voltage	Yes	Yes
[9]	CFOA	LPN	Voltage	No	Yes
[10]	OTA	LPN	Voltage	High	Yes
[11]	OTA	LPN	Voltage	High	Yes
[12]	CCII	LPN	Voltage	High	Yes
[13]	CCII	LPN	Voltage	High	Yes
[14]	CCII	LPN	Voltage	High	No
[15]	CCII	LPN	Voltage	High	Yes
[16]	OTRA	LPN	Voltage	Low	Yes
[17]	Current follower	LPN	Current	High	Yes
[18]	CCII	LPN	Current	High	Yes
[19]	CCII	LPN	Current	High	Yes
[20]	CCII	LPN	Current	High	Yes
[21]	CDTA	LPN	Current	High	Yes
[22]	CDTA	APN	Current	High	Yes
[23]	CDTA	APN	Current	High	Yes
[24]	CC	LPN	Current	High	Yes
[25]	CFTA	LPN	Current	High	Yes
Proposed work	OTRA	APN	Voltage	Low	Yes

In this paper authors aim at presenting OTRA based MSO, designed using APNs. The proposed circuit utilizes n ($n \geq 3$) APNs to produce n phase oscillations of equal amplitudes with a phase difference of $(360/n)^\circ$. The APN can produce a phase shift up to 180° as against a maximum of 90° produced by LPN. As a result the APN based scheme can be used to implement an even phase system for $n = 4$ also whereas $n = 6$ is the minimum for the systems designed using LPNs [8].

2. Circuit Description

In this section the generalized APN based MSO scheme [8] is described first which is followed by the design adaption using OTRA.

2.1 The APN based MSO design Scheme

The APN based MSO structure consists of n cascaded first-order APN blocks. The output of the n^{th} stage is fed back to the input of the first stage thus forming a closed loop as shown in Fig. 1(a). The output of n^{th} stage is fed back directly for odd phase system whereas for even phase system it is inverted, for sustained oscillations. In case of even phased system the structure can be modified for obtaining n even phased oscillations by replacing $n/2$ APN with $n/2$ inverters as shown in Fig. 1(b)

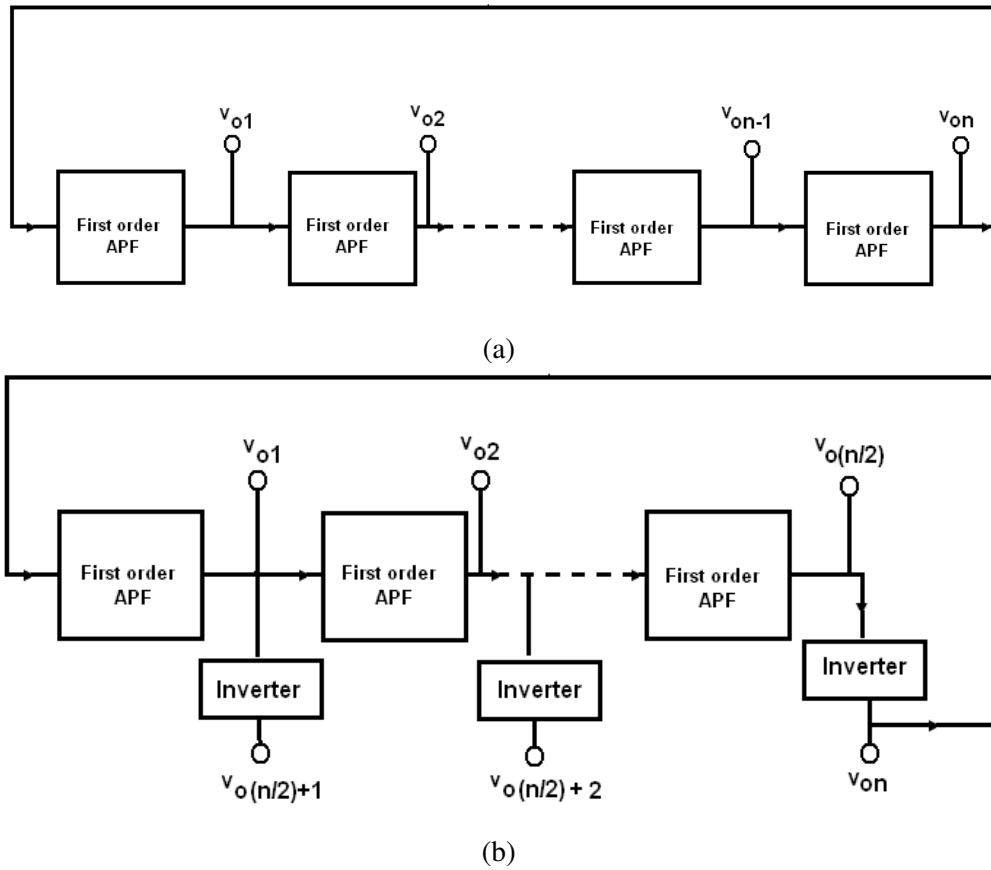


Fig.1. Generalized APN based MSO structure for producing n phase oscillations.
(a) Odd phased oscillations (b) Even phased oscillations

The transfer function $G(s)$ of each APN block is given by

$$G(s) = K \frac{(1 - s\tau)}{(1 + s\tau)} \quad (1)$$

where K represents the gain and time constant τ determines the corner frequency of the APN. The phase (ϕ) of each APN block is computed to be

$$\phi = -2 \tan^{-1}(\omega\tau) \quad (2)$$

Thus an APN can introduce signal phase lag from 0 to 180° as frequency ω varies from zero to infinity. From Fig.2 the open loop gain $L(s)$ can be expressed as

$$L(s) = K^n \left(\frac{(1-s\tau)}{(1+s\tau)} \right)^n \quad (3)$$

The Barkhausen criterion [27] for sustained oscillations at a frequency ω_0 , can be expressed by (4) and (5)

$$|L(j\omega_0)| = K^n \left(\frac{1-j\omega_0\tau}{1+j\omega_0\tau} \right)^n = 1 \quad (4)$$

$$\angle L(j\omega_0) = n\phi = -2\pi \quad (5)$$

Since $\left(\frac{1-j\omega_0\tau}{1+j\omega_0\tau} \right)^n = 1$, (4) results in condition of oscillation (CO) as

$$CO : K = 1 \quad (6)$$

Substituting from (2) in (5) total phase shift of the loop can be computed as

$$n(-2 \tan^{-1}(\omega_0\tau)) = -2\pi \quad (7)$$

Equation (7) will converge only for values of n such that $n \geq 3$. This results in frequency of oscillation (FO) as

$$FO : f_0 = \frac{1}{2\pi\tau} \tan\left(\frac{\pi}{n}\right) \quad (8)$$

The spacing between different phases is given by

$$\phi = \frac{2\pi}{n} \quad (9)$$

Thus the circuit gives rise to equally spaced oscillations having a phase difference of $(360/n)^\circ$.

2.2 The OTRA based MSO implementation

The OTRA is a high gain, current input voltage output ABB. The circuit symbol of OTRA is shown in Fig. 2 and the port characteristics in matrix form are given by (10), where R_m is transresistance gain of OTRA. For ideal operations the R_m of OTRA

approaches infinity and forces the input currents to be equal. Thus OTRA must be used in a negative feedback configuration [25, 28].

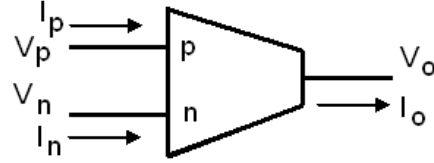


Fig.2. OTRA circuit Symbol

$$\begin{bmatrix} V_p \\ V_n \\ V_o \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ R_m & -R_m & 0 \end{bmatrix} \begin{bmatrix} I_p \\ I_n \\ I_o \end{bmatrix} \quad (10)$$

The MSO scheme outlined in section 2.1 can be implemented using OTRA based first order all pass sections. The OTRA based APN proposed in [28] and shown in Fig. 3 is used for MSO implementation.

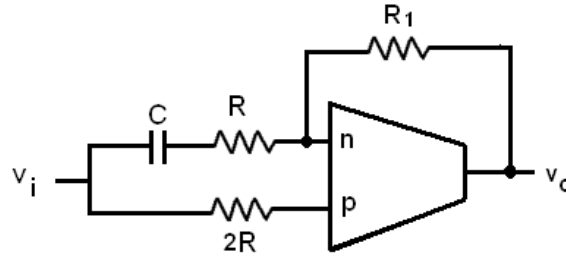


Fig. 3 The OTRA based APN [28].

The transtfer function of the APN can be written as

$$G(s) = \frac{V_o(s)}{V_i(s)} = K \frac{(1-s\tau)}{(1+s\tau)} \quad (11)$$

where

$$K = \frac{R_1}{2R} \quad \text{and} \quad \tau = \frac{1}{CR} \quad (12)$$

and the phase relation is expressed as

$$\phi = -2 \tan^{-1}(\omega CR) \quad (13)$$

The OTRA based odd and even phased MSO structures are shown in Fig. 4(a) and (b) respectively. The loop gain can be written as

$$L(s) = \left(\frac{R_1}{2R} \right)^n \left(\frac{(1 - \frac{s}{RC})}{(1 + \frac{s}{RC})} \right)^n \quad (14)$$

CO and FO can be expressed as

$$CO : R_1 = 2R \quad (15)$$

$$FO : f_0 = \frac{1}{2\pi RC} \tan\left(\frac{\pi}{n}\right) \quad (16)$$

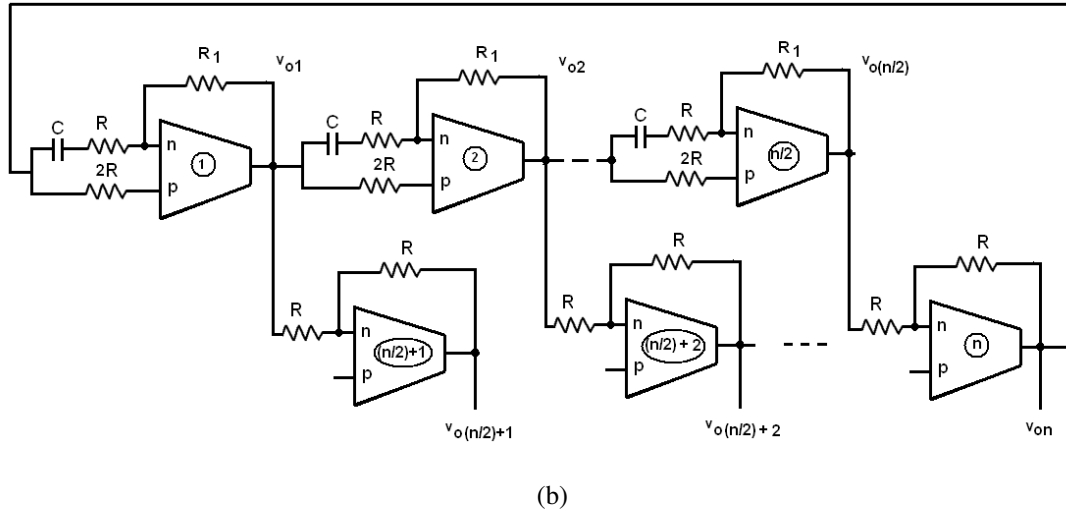
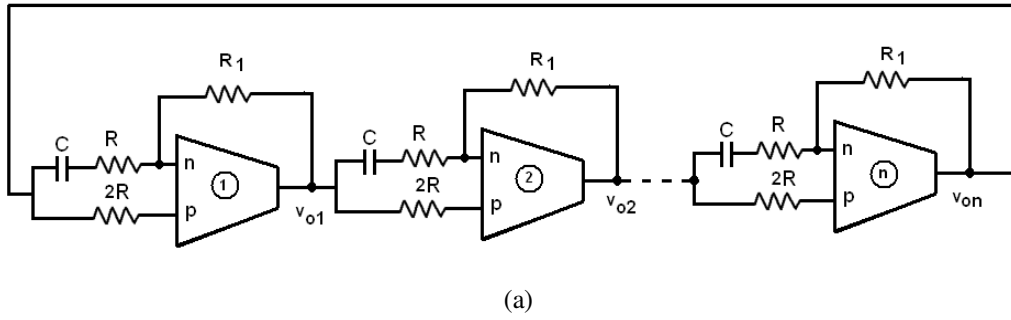


Fig.4. The OTRA based MSO circuits. (a) Odd phased (b) Even phased.

3. NON IDEALITY ANALYSIS

Ideally the trans-resistance gain R_m is assumed to approach infinity. However, practically R_m is a frequency dependent finite value. Considering a single pole model for the trans-resistance gain, R_m can be expressed as

$$R_m(s) = \left(\frac{R_0}{1 + s/\omega_0} \right) \quad (17)$$

For high frequency applications the trans-resistance gain, $R_m(s)$ reduces to

$$R_m(s) = \left(\frac{1}{sC_p} \right) \text{ where } C_p = \left(\frac{1}{R_0\omega_0} \right) \quad (18)$$

C_p is known as parasitic capacitance of OTRA. Taking the effect of C_p into account (11) modifies to

$$G(s) = \frac{R_1}{2R} \left(\frac{(1-sCR)}{(1+sCR)(1+sC_pR_1)} \right) \quad (19)$$

Thus the C_p of OTRA results in introduction of another pole having pole frequency $\omega_p = 1/R_1C_p$. However the value of C_p being small (typically 5pF) the parasitic pole frequency would be far off from the operating frequency of the APN and would not effect the MSO operation.

4. SIMULATION RESULTS

The proposed circuits are simulated using PSPICE to validate the theoretical predictions. The CMOS realization of OTRA presented in [26] and reproduced in Fig. 5 is used for simulation. The output of the MSO of Fig.5 (a) for $n = 3$, with component values $R = 2.5 \text{ K}\Omega$, $R_1 = 5 \text{ K}\Omega$ and $C = 0.1\text{nF}$, is depicted in Fig. 6. The steady state output is shown in Fig. 6(a) while the frequency spectrum is depicted in Fig. 6(b). The simulated frequency of oscillations is observed to be 1MHz against the calculated value of 1.1MHz.

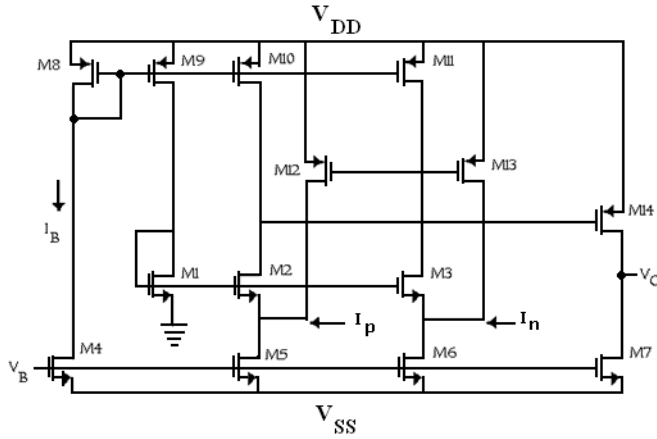
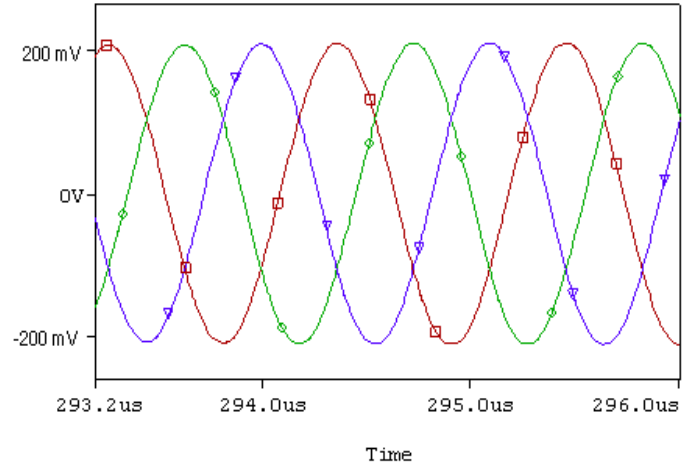
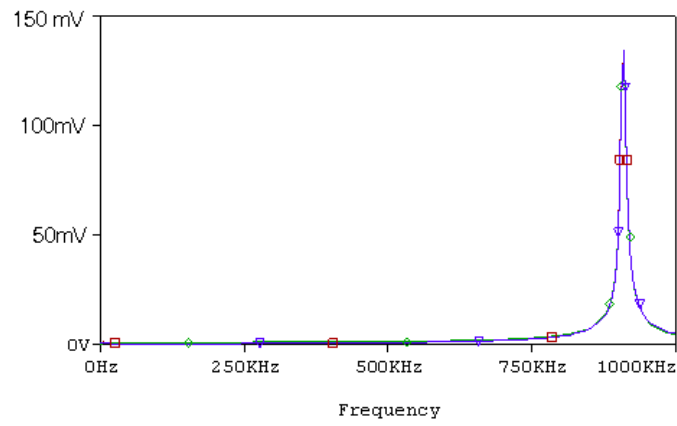


Fig. 5. The CMOS implementation of OTRA [26].



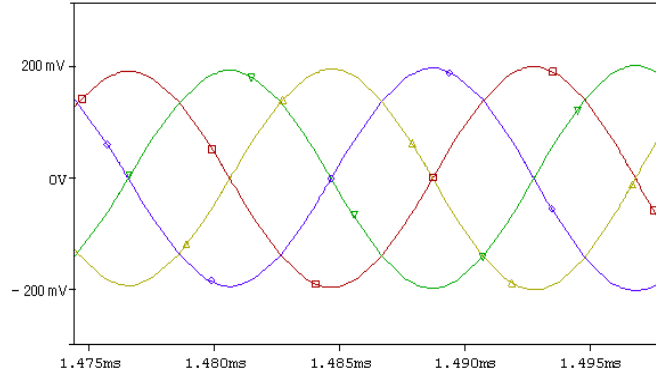
(a)



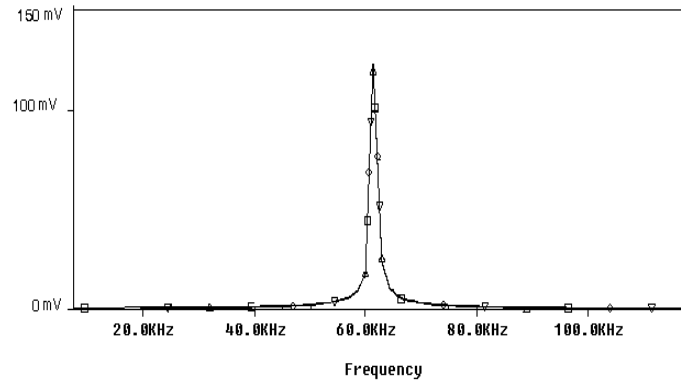
(b)

Fig. 6. Odd phased MSO output for $n=3$ (a) Steady state output (b) Frequency spectrum

Simulation results for $n = 4$, with component values $R = 2.5 \text{ K}\Omega$, $R_1 = 5 \text{ K}\Omega$ and $C = 1 \text{ nF}$ are depicted in Fig. 7. The simulated frequency is found to be 61.69 KHz while the theoretical calculation yields an FO of 63.69 KHz.



(a)



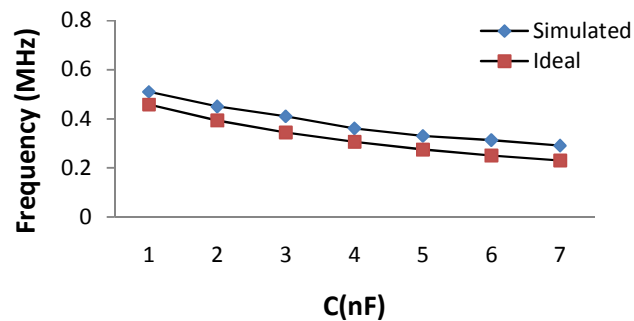
(b)

Fig. 7. Even phased MSO output for $n = 4$ (a) Steady state output (b) Frequency spectrum

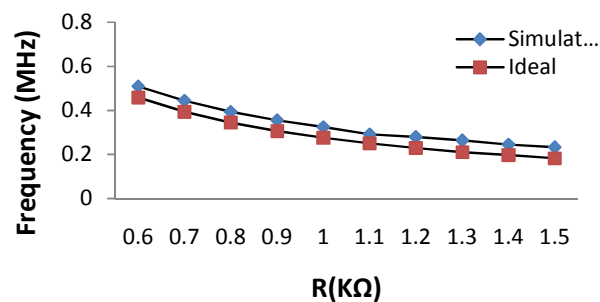
It may be observed from (8) that the FO can be varied either through R or by changing C . Variation of FO with respect to R while keeping $C = 1$ nF has been depicted in Fig. 8(a) whereas tuning with C is shown in Fig.8 (b) with $R = 0.5$ K Ω . It shows that the simulated values closely follow the theoretically calculated values.

8. CONCLUSION

In this work OTRA based MSO circuits, designed using first order all pass networks are presented. The proposed structures produce ' n ' phase oscillations of equal amplitudes which are equally spaced in phase. These circuits provide voltage output at low impedance thus can readily be used to drive voltage input circuits without increasing component count. The proposed circuits are very accurate in providing the desired phase shift. The workability of the circuit has been demonstrated through SPICE simulations.



(a)



(b)

Fig.8 Frequency tuning of MSO (a) with R (b) With C

References

- [1] TOMASI, W., Electronic Communications System, New Jersey: Prentice-Hall Inc., 1998.
- [2] JAIKLA, W., PROMMEE, P., Electronically tunable current-mode multiphase sinusoidal oscillator Employing CCCDTA-based allpass filters with Only Grounded Passive Elements, *Radioengineering*, vol. 20, 2011, p. 594-599
- [3] LEE, D.C., KIM, Y.S., "Control of single-phase-to-three-phase AC/DC/AC PWM converters for induction motor drives," *IEEE Transaction on Industrial Electronics*, vol.54, 2007, pp.797-804.
- [4] RAMAMURTI, V. P., RAMASWAMI, B., A novel three-phase reference sine wave generator for PWM invertors, *IEEE Transaction on Industrial Electronics*, vol. IE-29, pp. 235-240, 1982.
- [5] ABUELMA'ATTI, M. T., ALMANSOURY, W. A., Active-R multiphase oscillators, *Proc. Inst. Elect. Eng. pt. G*, vol. 134, 1987, pp. 292-293.
- [6] STIURCA, D., On the multiphase symmetrical active-R oscillators, *IEEE Transactions on Circuits and Systems II*, vol. 41, 1994, pp. 156-158.
- [7] GIFT, S. J. G., Multiphase sinusoidal oscillator system using inverting-mode operational amplifiers, *IEEE Tran. Instrum. Meas.*, vol. 47, 1998, pp. 986-991.

- [8] GIFT, S. J. G., The application of all-pass filters in the design of multiphase sinusoidal systems, *Microelectronics Journal*, 2000, vol. 31, pp.9–13.
- [9] WU, D.S., LIU, S. I., HWANG, Y. S., WU, Y. P., Multiphase sinusoidal oscillator using the CFOA pole, *Proc. Inst. Elect. Eng. Circuits, Devices Syst.*, vol. 142, pp. 37–40, 1995.
- [10] HOU, C. L., WU, J. S., HWANG, J., LIN, H. C., OTA-based even-phase sinusoidal oscillators, *Microelectronics Journal*, 28, 1997, pp. 49-54.
- [11] KHAN, I. A., AHMED, M. T., MINHAI, N., Tunable OTA-based multiphase sinusoidal oscillators, *International Journal of Electronics*, vol. 72, 1992, pp. 443-450.
- [12] WU, D.S., LIU, S.I., HWANG, Y.S., WU, Y.P., Multiphase sinusoidal oscillator using second-generation current conveyors, *International Journal of Electronics*, vol. 78, 1995, pp. 645–651.
- [13] HOU, C. L. SHEN, B., Second-generation current conveyor-based multiphase sinusoidal oscillators, *International Journal of Electronics*, vol. 78, pp. 317–325, 1995.
- [14] ABUELMA'ATTI, M. T., AL-QAHTANI, M. A., A grounded resistor Current conveyor based active-R multiphase sinusoidal oscillator, *Analog Integrated Circuits Signal Processing*, vol.16, 1998, pp.29–34.
- [15] SKOTIS, G. D., PSYCHALINOS, C., Multiphase sinusoidal oscillators using second generation current conveyors. *International Journal of Electronics and Communication (AEU)* , 2010, vol. 64, p. 1178–1181.
- [16] PANDEY, R., PANDEY, N., BOTHRA, M., PAUL, S. K., Operational transresistance amplifier based multiphase sinusoidal oscillators, *Journal of Electrical and Computer Engineering* Volume 2011, Article ID 586853, 8 pages doi:10.1155/2011/586853.
- [17] ABUELMA'ATTI, M. T., Current-mode multiphase oscillator using current followers. *Microelectronics Journal*, vol. 25, 1994, pp. 457–461.
- [18] ABUELMA'ATTI, M.T., AL-QAHTANI, M.A., A new current-controlled multiphase sinusoidal oscillator using translinear current conveyor, *IEEE Transactions on Circuits and Systems II*, vol. 45, 1998, pp. 881–885.
- [19] ABUELMA'ATTI, M.T., AL-QAHTANI, M.A., Low-component second-generation current conveyor-based multiphase sinusoidal oscillator, *International Journal of Electronics*, vol. 84, 1998, pp. 45–52,
- [20] LOESCHARATARAMDEE, C., KIRANON, W., SANGPISIT, W., YADUM, W., Multiphase sinusoidal oscillators using translinear current conveyors and only grounded passive components, *Proceedings of the 33rd Southeastern Symposium on System Theory*, pp. 59-63, 2001.
- [21] TANGSRIRAT, W., TANJAROEN, W., Current-Mode Multiphase Sinusoidal Oscillator Using Current Differencing Transconductance Amplifiers, *Circuits System Signal Processing*, 27, 2008, pp. 81–93.
- [22] TANGSRIRAT, W., TANJAROEN, W., PUKKALANUN, T., Current-mode multiphase sinusoidal oscillator using CDTA-based allpass section, *International Journal of Electronics Communication (AEU)*, 63, 2009, pp. 616 – 622.

- [23] JAIKLA, W., SIRIPRUCHYAUN, M., BIOLEK, D.,BIOLKOVA, V. High-output-impedance current-mode multiphase sinusoidal oscillator employing current differencing transconductance amplifier-based allpass filters, *International Journal of Electronics*, vol.97, 2010, pp. 811–826.
- [24] KUMNGERN, M., DEJHAN, K., CHANWUTITUM, J., Electronically tunable multiphase sinusoidal oscillator using translinear current conveyors *Analog Integrated Circuit Signal Processing*, vol. 65, 2010, pp. 327–334.
- [25] UTTAPHUT, P., Realization of Electronically Tunable Current Mode Multiphase Sinusoidal Oscillators Using CFTAs, *World Academy of Science, Engineering and Technology*, vol. 6, 2012, pp.657-660
- [26] MOSTAFA, H., SOLIMAN, A. M., A modified realization of the OTRA, *Frequenz*, vol.60, 2006, pp.70-76.
- [27] SEDRA, A. S., SMITH, K. C., *Microelectronic Circuits*, Oxford University Press, New York, 2004.
- [28] KILINC, S. , CAM, U., Cascadable allpass and notch filters employing single operational transresistance amplifier, *Computers and Electrical Engineering*, vol. 31, 2005, pp. 391–401.

An Empirical Study on Market Research of Organized Students' Housing Industry in India

Mayank Garg

Dept. of Electrical and Electronics
Birla Institute of Tech. and Science
Pilani, India
mayank.gargbits@gmail.com

Kunal Gupta

Production and Industrial Eng.
Delhi Technological University
New Delhi, India
kunalgupta.k1@gmail.com

Rakshit Jha

Dept of Mathematics
Delhi Technological University
New Delhi, India
jha.rakshit@gmail.com

ABSTRACT: *The purpose of this paper is to examine the scope of the organized student housing industry in one of the largest and most complicated markets of the world, where rate of increase in supply of new student housing facilities is much below the rate of increase in demand, which is pushing more students to opt for alternate student housing options. This study follows a qualitative approach of survey by systematically collecting, analyzing and interpreting the required data. The data pertaining to city overview, student profile, facilities and price points offered by alternative housing options, government building norms, best practices across the globe and key parameters for establishing student housing facilities is collected through secondary research on internet, analyzed and interpreted for the given case study in order to propose positioning, facilities, location, USP and potential tie-ups for a new housing facility. This study identifies huge business potential which the Indian student housing industry offers to the organized firms. At present, the students are deeply unsatisfied with the unorganized student housing alternatives which is not able to cater to their rising expectations of low cost and easy availability. This provides firms with a long term vision to dominate this industry which at present is in its pristine stage. This study is the first attempt to examine the scope of student housing industry in India by analyzing data collected according to the research methodology and propose a facility in the given city.*

KEYWORDS: *Student housing industry, student accommodation, market research, organized sector, business potential.*

I. INTRODUCTION

One of the most highlighting trends in real estates over the past decade have surrounded around the student housing industry. This is due to a variety of factors such as, generational variation in tastes, competition among colleges, broader trends in the

overall housing market and steady transformation in the nature of student housing on and near the institutions. The growth of student housing industry in India is primarily driven by availability of good education facilities. The quality of the present market competitors like paying guests, rented flats, private hostels and the college hostels have not been up to satisfactory level. The very need for a better more developed and affordable housing is still in rising demand.

A. FACTORS AFFECTING NEED OF STUDENT HOUSING

The need of the student housing strives on factors related to college/enrolment, student and infrastructure. College/enrolment related factors are establishment of new educational institutes, increase in capacity of established institutions and increase in student migration to education hubs to seek for better quality of education.

Student related factors include safety/security in wake of increasing crime against women, affordability of the hostel facility since hostels are expected to provide a variety of facilities at lower cost. Infrastructure related factors are the facilities which the hostel offer to the students which may not be available in alternate housing options such as outdoor sports facilities and recreation facilities, location/transportation, where students have to maintain a strategic balance between rental costs, daily transportation costs and personal convenience. Thus, on the basis of these factors, need for organized student housing options can be identified.

Table1: Best practices of market players from India and Abroad

	Unite Group PLC	CRM Students	Asset Campus Housing Inc.	American Campus Communities	The PAD	Unilodge	Manipal Integrated Services	Proposed Facility
Country	UK	UK	USA	USA	Australia	Australia	India	India
Average Room Size (sqft.)	137	120	171	150	120	159	200-300	150-300
Number of Facilities	120	54	92	200	5	35	3	NA
Total beds	42,000	19,000	60,000	125,000	1,700	10,000	3,500 (Proposed 20,000)	1,000
INR/Room/Month	NA	NA	NA	NA	NA	NA	4,500-16,000	4,000-8,000*
Basic Furniture	✓	✓	✓	✓	✓	✓	✓	✓
Air Conditioning	✓	✓	✓	✓	✓	✓	✓	✓
Heating Facility	✓	✓	✓	✓	✗	✗	✗	✗
Attached Bathroom	✓	✓	✓	✓	✓	✓	✓	✓
Telephone	✓	✓	✓	✓	✗	✗	✗	✗
Bills Included	✓	✓	✓	✓	✓	✓	✓	✓
Appliances	✓	✓	✓	✓	✗	✓	✗	✓
Wi-Fi	✓	✓	✓	✓	✓	✓	✓	✓
Security/CCTV	✓	✓	✓	✓	✓	✓	✓	✓
Common Room/TV Room	✓	✓	✓	✓	✓	✓	✓	✓
Laundry	✓	✓	✗	✓	E	✓	✗	E
Swimming Pool	✗	✓	✓	✓	✗	✓	✗	✗
Fitness Facility/Gym	✗	✓	✓	✓	✓	✓	✓	E
Theatres	✗	✓	✓	✓	✓	✗	✗	✗
Housekeeping	✗	✓	✗	✗	✗	✓	✓	✓
Indoor Sports	✗	✓	✗	✓	✓	✓	✓	✓
Outdoor Sports	✗	✓	✗	✓	✓	✓	✓	✓
Others	✗	✓	✓	✓	✓	✓	✓	✓

B. EXISTING PLAYERS

In countries like US, UK and Australia, student housing industry is dominated by organized players. The success of their business models has inspired many firms to enter into this industry in India since it has high potential for new players who can make a mark with the rising demand in many of the cities which are emerging as educational hubs. Table 1. Shows best practices of some market players from India and abroad.

C. GOVERNMENT SUPPORT AND REGULATIONS

Given the present condition of college hostels in India, Central and several State Governments have introduced various schemes to establish, develop, maintain and manage hostels. Assistance is extended to NGOs, public sector undertakings and educational institutions through grants, funding and land at subsidized or minimal rates. For instance, 20 new IIT's are proposed to be set up with hostel facilities through the financial assistance of the Central and State government. Government of Gujarat has supported 12 engineering colleges with hostel facilities by leasing land at nominal rentals and providing financial assistance up to INR 10 crores. A pilot project to construct girls' hostel is developed by the Government of Madhya Pradesh at an estimated project cost of INR 50 crore or USD 10.4 million. These are some of the projects which reflect growing momentum in the student housing industry across the country.

At present government is introducing a number of schemes in order to alleviate the housing related problems faced by the students. For instance, to promote building of hostels for students in Maoist-hit areas- The Chhattisgarh Government has introduced a scheme to construct hostels in Naxalite affected areas with the CSR funds of corporate operating in the state. For construction of hostels for women, colleges affiliated with the UGC will get financial assistance of INR 60 lakh to 2 crore. A centrally sponsored scheme provides financial assistance to construct hostels for housing of OBC college

students. Babu Jagjivan Ram Chhatrawas Yojna is a scheme which provides funds for constructing new hostel buildings and renovating existing hostels. Since private players cannot directly avail the benefits of these government schemes, it is suggested that they partner with academic institutions such as colleges, NGOs and universities who are eligible to do so.

D. FUTURE OF THIS INDUSTRY

Student housing is one of the most vibrant Indian real estate markets in the foresee future. Dense student populations that exist around prominent colleges positively affect the demand for residential spaces as well restaurants and small retail spaces. In fact, the future of this sector in India is extremely promising with the country's top 10-15 corporate schools having plans to develop a total of 8000-1000 schools over the next decade, with the aim to deliver quality education to far-flung parts of the country. The corporate higher education industry is expected to build close to 10million sq.ft. of educational institution spaces over the next few years-and this does not include the rapid expansion plans that are underway to build the IIT's, IIM's, SPA's and other reputed educational institutions. The demand for the right kind of student housing that such a scenario presents can well be imagined.

1. LITERATURE REVIEW

Students going to live away from their parents for higher educations in college campuses have a greater impact on their personalities and psychology. This is emphasized by Owens (2010) [4] where his research study identified that both race and extracurricular involvement influenced the psychosocial development of first-year students who lived on-campus.

Research conducted in the United States has shown that students living in residence have greater gains in areas such as intellectual development, and are more likely to stay in university and complete their degrees, than students who live off-campus. In a study done by J. Paul Grayson (1997) of York

University [1], a large commuter university, it is shown that place of residence does affect student involvement and first year marks. He states that students living in college residential are more likely to score more than students living in home as they come from more affluent families and students living in college hostels are mostly males and most of the girls came from homes. In his study, he found the average co-curricular involvement scores calculated for students in double and single rooms are 22 percentile points higher than those living with parents.

There is an adequate need for the improvement of the quality of housing accommodation for tertiary education students. For instance Olotuah (2000), in a study [3] of the suburban housing situation in Akure, Ondo State of Nigeria, indicated that in spite of the evident poor housing conditions in some areas of that capital city, residents did not complain because of their low expectations with respect public social services.

The importance of Student Accommodation Quality (SAQ) can be well understood in the study made by Nimako and Bondinuba (2013) [5], where he argues that students priority of SAQ is common among the higher education students where the availability of basic facilities like water and electricity, followed by residential infrastructure, security, environmental conditions, toilet, distance to institute, bedroom and bathroom situation, expenses, rules and regulation is followed by support dimension of SAQ being garbage disposal. The study concludes that student housing providers and institution's hostel managements should try to focus more of their resources to the priorities which accommodated students prioritize. Generally, giving more of emphasis on the SAQ dimension to ensure effective and efficient hostel accommodation services that promote students and induce larger demand in the market.

A policy study done by Gilroy, Davis, Anzia and Segal (2007) tries to come up with a solution [2] to this problem. They compare the living condition of college hostel to that of dingy dorm. With lack of good quality food, sports facilities overcrowded and shortage of residential hostels, some universities have

called upon the expertise of private housing providers to develop comprehensive solutions to these problems. In return, they have endeavor the rewards of lower cost, higher-quality student housing. Yet despite these success stories, privatization of student housing remains underutilized. The far most suitable example of the success story of the privatization of accommodation is that of the U.S. military housing model. The MHPI has proven to be a remarkable success since its creation in 1996. In less than a decade, the military services have made impressive strides in utilizing private housing developers to replace inadequate on-base housing. The private sector has proved that it can build, renovate, operate, and maintain higher-quality housing communities at less cost than traditional military construction methods.

The advent of more educational institutions in major cities has attracted many big players from real estate into picture. Established players in the market are on a buying spree in hopes of remaining competitive among bi-name new comers. In year 2012, a record \$3.7 billion of student accommodation properties trades took place, up nearly by 95% following 2011, according to ARA student housing. There are growing concerns of over-building. ARA tracked 71 projects under way in year 2013 up by nearly 129% from previous year. A big reason why companies are screeching into student accommodation is that they believe it is recession free. Kayne Anderson Real Estate Advisors, a private-equity investor with 15,000 beds, posted annual returns that exceeded 20% between 2007 and 2012, despite of economic slowdown. The firm isn't worried of oversupply.

II. SCOPE AND NEED OF STUDY

There has been very less research done in this industry and almost nil in India. To the extent that this industry can grow might lead to intensive investment in this industry particularly in India shadowing some foreign countries. Pertaining to the success and increase of interests in some countries, there is a strategic approach needed to get into the Indian market with high potential.

III. RESEARCH METHODOLOGY

This research is conducted by collecting and analyzing data pertaining to city overview, student profile, competitor's profile, hostel building norms set by the government, issues presently faced by students, best practices across the globe shown in Table 1 and key parameters for establishing student housing facilities shown in Table 2. The research takes place by obeying the following steps:-

Step 1- Develop overview of the selected city by collecting data pertaining to profile of currently operating higher education institutes and the areas in which they are located.

Step 2- Develop student's profile by collecting and analyzing data pertaining to percentage of outstation students in total enrollment and split of student's residing in rented housing options and its respective price points.

Step 3- Understand the supply of rented housing options in the vicinity of the education institutes such as college hostels, private hostels, paying guests services (PG) and rented flats by collecting and

analyzing data pertaining to price points and facilities offered by each option.

Step 4- Study hostel building norms set by the development authority of the selected city.

Step 5- Analyze the housing issues presently faced by the students and suggest pragmatic solutions.

Step 6- Considering the analysis conducted in the above steps, best practices across the globe (as shown in Table 1) and key parameters for establishing student housing facilities (as shown in Table 2), we propose positioning, key facilities, location, USP and potential tie-ups for the new housing facility.

Table 2: Key Parameters for Establishing Student Housing Facilities

Price per month paid for housing	Affordability is a key consideration for students who usually have limited earnings while studying, thereby getting financial support from their families
Distance from College/University	Distance from education center is of primary importance as it results in saving time. Also, easy access to a developed public transport system acts as a driver for the location
Amenities provided	Students prefer housing that provides access to a range of services and modern facilities like hygienic food, Wi-Fi, furnished rooms, etc. This meets the essential, leisure and community needs of the students.

IV. CASE STUDY- BANGALORE, INDIA

A. OVERVIEW OF BANGALORE

According to the All India Survey of Higher Education by the Ministry of HRD, Government of India, the city of Bangalore has the highest number of educational institutions (898 institutes with enrolments over 5, 12,000). City home to institutes including Indian Institute of Science, National Law School, Indian Institute of Management, and National Institute of Design cover varied fields of comprising Aerospace Engineering, Agriculture, Biotechnology,

Business Management, Law, Sociology, Economics, Nanotechnology a plethora of other departments. In addition to being an IT hub, Bangalore is an education center that attracts students from all over India and abroad with 10,000 foreign students accounting for 37% of all those coming to India. Figure 1 discusses the composition of external and local students, exhibiting the diversity of the city. From a total of 898 institutes 301 are formal education institutes, 529 technical or professional institutes and 68 are of other types. These institutes are mainly located in 17 areas as shown in Table 3.

Table 3: Main locations of institutes

	1. Kolar	6. VV Puram	11. Jayanagar	16. Banashankari
	2. RT Nagar	7. Basavanagar	12. Christ School Road	17. City Road
	3. Malleswaram	8. Nagarbhavi	13. Koramangala	
	4. Rajajinagar	9. Vasant Nagar	14. Hosur Road	
	5. Vijay Nagar	10. JB Campus	15. BTM Layout	

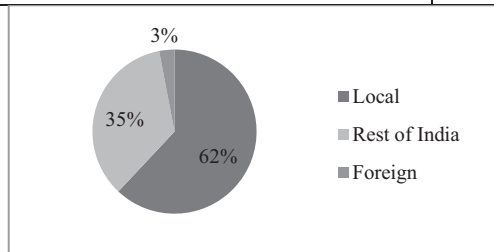


Figure no 1: Student profile residing in Bangalore

A. STUDENT PROFILE OVERVIEW

Bangalore not only attracts students from the southern states of Andhra Pradesh, Tamil Nadu and Kerala but also from other North and Eastern states of India. Students from different countries like Sudan, Germany, Africa, Iran, Japan, Thailand, Korea, Malaysia, the Middle East, and neighbouring Bangladesh are enrolled in city institutes. Total number of enrolment is estimated at 512, 215 out of which 65% is the estimated percent of outstation students, that is, 332,940 students are outstation students. Of the total number of outstation students, only 15% receive housing in college hostels. The housing for the remaining outstation students is met through the unorganized sector, with paying guest being the main provider. The pricing for different categories of housing vary from INR 1,200 – 7,000 based on location and facilities. And the split of students in rented housing, as shown in Figure 2, signifies that Bangalore offers huge market potential

for organized student housing where currently 85% market is dominated by unorganized modes of student housing.

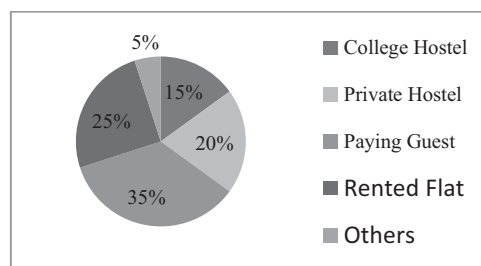


Figure no 2: Split of students in rented housing

B. SUPPLY OF STUDENT HOUSING OPTIONS

• College Hostels

With the number of outstation students coming to Bangalore increasing, getting a hostel room has become a challenge for students. Most hostels give rooms on first come-first serve basis, and some colleges give preference to toppers in the first year. Admission into hostel in the second year depends on marks scored in the second semester. Conditions in these college hostels are poor with several news reports outlining their poor infrastructure and students face several problems of overcrowding, poor quality food, unhygienic conditions and lack of infrastructure. Facilities offered by college hostels are mentioned in Table 4. The rentals of college hostels are generally in the range of INR 850 – 2,500 per student per month. Some of the key areas for college

hostels and its respective price points are shown in Table 5.

Table 4: Facilities offered by college hostels

Included in rent	Student Need*
Furniture	Medium
Dining Hall	Medium
Water Supply	High
Electricity charges	High
Internet	Medium
Security	Medium

Library	Low
Study Hall	Low
Indoor sports room	Low
Outdoor sports facility	Low
Medical facilities	Medium
Housekeeping	Medium
Available at additional cost	Student Need*
Mess services	Medium
Laundry	Medium
Air Conditioning	Medium

Table 5: Key areas and price points of college hostels

Area	Name of College	Price/student/month (INR)
JB Campus	Bangalore University	850-1,500
Kolar	Bangalore University Hostel	1,300-2,000
City Campus	Bangalore City College	1,750-2,500
Christ School Road	Christ University	1,200-2,500
Jayanagar	City Group Institutions	1,200-2,500

Table 6: Key areas and price points of private hostels

Area	Names of Private Hostels	Price/student/month (INR)
Basavanagudi	Mahratta Hostel, Deepak Women's Hostel	1,500 – 3,000
Banashankari	KIMS boys hostel, CVMS Hostel	3,000 – 4,000
Rajajinagar	B.S Siddappa Memorial Men's Hostel, Sriraksha Women's Hostel	3,000 – 5,000

and its respective price points are discussed in Table 6.

- Private Hostel

Several private hostels have sprung up to cater to the large student population seeking housing but the quality is poor in most hostels with students face problems with bad food and lack of basic amenities. Facilities offered by private hostels are mentioned in Table 7. Generally, private hostels house 100-150 students with the size of the rooms from 100 sq.ft to 300 sq.ft rents lying between INR 1,500 – 5,000 per student per month. Recently, Manipal Integrated Services acquired Bangalore-based Woodstock Ambience, with an objective to build and manage student housings in Bangalore and has completed a hostel complex with 1000 rooms to provide housing to around 2000 students. Key areas for private hostels

Table 7: Facilities offered by private hostels

Included in rent	Student Need [#]
Furniture	Medium
Dining Hall	Medium
Water supply (hot and cold)	High
Filtered drinking water	High
Security	High
Library	Low
TV Room	Low
Computer room	Low
Medical facilities	Medium
Housekeeping	Medium
Available at additional cost	Student Need [#]
Mess services	Medium
Laundry	Medium
Air conditioning	Medium
Internet	Medium
Electricity and power back up	High

- Paying Guest

Most Bangalore colleges, particularly newer ones, lack hostel facilities, and many of the older hostels offer limited amenities, resulting in demand for PG housing. Paying Guest housing is also a popular choice with young IT professionals in the city. Housing is available in a wide price range and owners also offer facilities such as kitchen and ease of restrictions on timing etc. Facilities offered by Paying Guests is mentioned in Table 9. There have been reports stating that Paying Guest housing in Bangalore have lax security and this being a major disadvantage for female students. The rentals of PG in Bangalore are generally higher than college hostels and private hostels with range varying between INR 2,000 – 8,000. Key areas for Paying Guests and its respective price points are shown in Table 8.

Table 8: .Key areas and price points of Paying Guests

Area	Price/student/month (INR)
Shanthi Nagar	2,750 – 8,000
Vijay Nagar	3,000 – 5,000
Malleswaram	3,500 – 5,500
V V Puram	2,000 – 4,000
Jayanagar	5,000 – 7,000

Table 9: Facilities offered by private hostels

Included in rent	Student Need [#]
Furniture	Medium
Water supply	High
Security	High
Kitchen facility	Medium
Common Room	Low
Ensuite	Medium
Available at additional cost	Student Need [#]
Food	Medium
Laundry	Medium

Included in rent	Student Need [#]
Furniture	Medium
Dining Hall	Medium
Water supply (hot and cold)	High
Filtered drinking water	High
Security	High
Library	Low
TV Room	Low
Computer room	Low
Medical facilities	Medium
Housekeeping	Medium
Available at additional cost	Student Need [#]
Mess services	Medium
Laundry	Medium
Air conditioning	Medium
Internet	Medium
Electricity and power back up	High
Air conditioning	Low

Internet	Medium
Electricity and power back up	High
Filtered drinking water	High
Housekeeping	Medium

- **Rented Flats**

Lack of facilities and strict rules in hostels and paying guest housing have resulted in students to pool resources and share an apartment/ rented flat. Facilities offered by rented flats are mentioned in Table 11. Many flat owners do not want to rent to single girls and prefer to rent to international students. However, flexibility in terms of budget and facilities offered is a key driver for students to opt for this mode of housing. The rentals are in the range of INR 5,000 – 12,000 per student per month. Key areas for rented flats and its respective price points is shown in Table 10.

Table 10: Key areas and price points of rented flats

Area	Price per student per month (INR)
Kormangla	8,000 – 10,000
BTM Layout	5,000 – 7,000
Nagarbhavi	6,000 – 8,500
RT Nagar	8,000 – 12,000
Vasanth Nagar	5,000 – 9,000

Table 11: Facilities offered by rented flats

Included in rent	Student Need [#]
Kitchen facility	Medium
Ensuite	Medium
Available at additional cost	Student Need [#]
Laundry	Medium
Air conditioning	Medium
Internet	High
Electricity and power back up	High
Water charges	High
Security	High

- **Hostel Building Norms for Bangalore**

Hostel building norms in Bangalore is set by the Bangalore Development Authority. An overview of select guidelines for hostels in Bangalore is shown below in Table 12. These norms would vary, based on the site selection, classification of hostels and its corresponding the land use proposed as per the development Plan of the city.

- **Issues presently faced by students**

An analysis of the problems faced by the students with the present housing facilities is done in order to incorporate the solution suggested in the proposed facility. The analysis is shown in Table 13.

Table 12: Hostel Building Norms for Bangalore

Activity	Minimum Area required	Road width in mtr.	Definitions	Additional developments permitted
Residential Use Zone				
Hostel	Minimum 500 sq.mt.	12 mtr and above		Plotted residential developments, Villas, Semi-detached houses, Apartments, Hostels, Dharamshalas, Multi Dwelling Housing, Service Apartments, Group housing.
Commercial Use Zone				
Technical College Hostel	Hostel facility area is required to be 'adequate'	30 meters and above	A premises having educational and playing facilities for students of undergraduate and post graduate courses under a university	College, Residential Flat, Institutional hostel, Retail Shops (general, grocery, books & chemists, barber, launderer), Canteen, Auditorium, Indoor games hall, swimming pool, playground
Medical College Hostel	There must be hostel provision for 75% of total intake. The size of the room must not be less than 27 sq. meters for a sharing room and 9 sq. meters for single room	30 meters and above		
Industrial				
Hostel	Hostels in Bangalore are prohibited to be located in the Industrial Zones			

V. PROPOSED FACILITY

- It is suggested that the proposed facility/hostel be located at either Koramangala or Adugodi.
- This offers possibility of potential tie-ups with private and government universities like Christ University, Bangalore University and the National Law School of India.
- The accommodation cherishes the transportation facilities that it gets right at its door, including metro station, railway station, bus stand and international airport.
- The proposed expense is to be laid out with nominal rates ranging from INR 4,000- 6,000 per month.

- The USP of the proposed facility will be modern facilities, strategic location, proximity to public transports and key facilities, as proposed for individual rooms and overall facility respectively in Table 15.
- Depending upon the analysis done above, positioning of the proposed facility is shown in Table 14.

Table 13: Issues presently faced by students and suggested solution

Table 14: Margin specifications

Required bed facility	1000 beds
Suggested room size	200-250 sqft
Students willing to pay	INR 4,000 - 6,000 per month

Table 15: Proposed facilities for individual rooms and overall facility

Individual Rooms	
Furniture	✓
Bathroom (shared)	✓
Electric Appliances	✓
Wi-Fi	✓
Overall Facility	
Security	✓
Cafeteria/ Common Area	✓
Housekeeping/ Laundry	✓
Power Backup	✓

VI. CONCLUSION

In a billion plus Indian society where college education is becoming a trivial, the sector has the potential to bolster a plethora of related business with student housing one of them. With the increasing intake of students the colleges are finding it tough to accommodate them in the present facilities, pushing students to opt for alternate student housing options. In the absence a proper research the paper identifies the problem deeply, comparing with the present international scenario and identifying the roots causes of the issue. To further understand the student housing problem in India a case study is taken for Bangalore, a pioneer city in education. Student profiling is done for the city and each alternatives are study in detail. Finally after in-depth analysis a profitable business model is suggested for the city

Parameters	Issues	Suggested solution
Condition	Almost all the college hostels across the cities have deplorable and poor hostel infrastructure	Hostels with well-designed interiors and best-in-class equipment
Size	Generally the size of rooms provided in all the housing is small. These small rooms also have an occupancy of 2 or more students	Room sizes ranging between 150-300 sq.ft., varying across cities with provisions for opting Single/Double Occupancy
Facilities	Facilities provided in these hostels are generally inadequate and unhygienic, Majority of hostels only provide basic amenities like cot, almirah and quality of food is poor	Hygienic meals, dining space, cafeteria, medical Facility, Wi-Fi access, laundry, indoor and outdoor sports facilities, etc.
Security	Security and safety is a major concern , especially for female students	Installation of CCTV cameras, recruitment of security personnel and wardens
Restrictions	Students face a lot of problems with rigid rules and regulations imposed by hostel authorities	Flexibility in rules and regulations relating to hostel timings, visitors, and overnight guests
Unorganized Hostels	Majority of the hostels across the selected cities are unorganized and at times in unauthorized buildings	Planned and organized facilities (e.g. Manipal Group, a pioneer has established YOHO Hostels in 3 cities in India)

VII. REFERENCES

- [1] J. Paul Grayson, *“place of residence, student involvement, and first year marks”*, The Canadian Journal of Higher Education La revue canadienne d'enseignement supérieur Volume XXVII-1, 1997, pages 1-24.
- [2] Leonard C. Gilroy, Laura J. Davis, Sarah F. Anzia, Geoffrey F. Segal, *“privatization of student housing, a financing alternative”* Kathy King-Griswold Warner School University of Rochester EDU 496 Spring, 2013.
- [3] Olotuah, A. O. (2000). *“Evaluation of suburbia housing conditions in akure, nigeria”*, African Journal of Development Studies.2 (1&2). 165, 172.
- [4] Owens, Justin (2010). *“The impact of university housing construction type on psychosocial development of first-year students”* Paper 672.
- [5] Simon Gyasi Nimako, Francis Kwesi Bondinuba *“relative importance of student accommodation quality in higher education”* Current Research Journal of Social Sciences 5(4): 134-142, 2013 ISSN: 2041-3238.

Analyzing Software Change in Open Source Projects using Artificial Immune System Algorithms

Ruchika Malhotra

Department of Software Engineering
Delhi Technological University
Delhi- 110042, India
ruchikamalhotra2004@yahoo.com

Megha Khanna^{*,+}

^{*}Acharya Narendra Dev College, University of Delhi
⁺Delhi Technological University
Delhi, India
meghakhanna86@gmail.com

Abstract—Development of software change prediction models, based on the change histories of a software, are valuable for early identification of change prone classes. Classification of these change prone classes is vital to yield competent use of limited resources in an organization. This paper validates Artificial Immune System (AIS) algorithms for development of change prediction models using six open source data sets. It also compares the performance of AIS algorithms with other machine learning and statistical algorithms. The results of the study indicate, that the models developed, are effective means of predicting change prone classes in the future versions of the software. However, AIS algorithms do not perform better than machine learning and other statistical algorithms. The study provides conclusive results about the capabilities of AIS algorithms and reports whether there are any significant differences in the performance of different algorithms using a statistical test.

Keywords— *Artificial Immune System algorithms; Change proneness; Open source projects; Object- Oriented metrics; Software Quality.*

I. INTRODUCTION

It is necessary for a software system to change progressively in order for it to be useful and functional during its entire lifecycle. A primary challenge for the software practitioners is to understand and maintain these changing software systems effectively. A change in a software could be corrective, adaptive, perfective or preventive in nature [1]. More than the nature of the change, it is crucial to identify classes which would probably change in future. Identification of these change prone classes of a software would help in efficient maintenance by assigning more resources to these classes. Moreover, determining change prone classes would also help develop better quality software as these classes are probable sources of defects and enhancements and thus stringent testing activities on such classes would result in less number of changes and defects in future.

A number of researchers have developed prediction models which efficiently predict change prone classes of a software using Object- Oriented (OO) metrics [2-6]. OO metrics are a representative of various OO attributes like coupling, cohesion, inheritance etc [7-8]. Values of these OO metrics are indicative of software design and help in early assessment of probable

change prone classes. Researchers have also evaluated a number of algorithms for efficient development of change prediction models which incorporate statistical as well as Machine Learning (ML) algorithms. Different algorithms work differently and may yield contrasting results on various software data sets. Hence, it is important to evaluate and assess a number of algorithms for the task of change prediction.

Artificial Immune System (AIS) algorithms are a new set of algorithms, which emulate the human immune system. The human immune system classifies a particular pathogen and learns the antibodies to be produced in order to protect human from a disease. This process is similar to reinforcement learning where the system memorizes the structure of a pathogen to provide quicker responses in future [9]. These capabilities of “memory” and “learning” are used by the AIS algorithms to solve complex problems. Moreover, AIS algorithms have distributed control and are adaptive, self-optimizing and evolutionary in nature [9]. These specific characteristics of the AIS algorithms make them a potential candidate for exploration in change prediction tasks. Though these algorithms have been successfully used for defect prediction [10], optimization [11], anomaly detection problems [12] and computer security [13], to the best of author’s knowledge, they have not yet been evaluated for change prediction tasks.

The main aim of this study is to assess the effectiveness of AIS algorithms for developing change prediction models. We further compared the results of these models with other ML algorithms and statistical algorithms. The research questions addressed in the study are as follows: (1) Are AIS algorithms suitable for change prediction tasks? (2) What is the comparative performance of AIS algorithms with other ML algorithms or logistic regression (statistical algorithm)? In order to explore the above research questions, we developed change prediction models using three AIS algorithms (Artificial Immune Recognition System (AIRS), IMMUNOS99 (IM99), Clonal Selection Algorithm (CLG)), four ML algorithms (Bagging (BG), Adaptive Boosting (AB), Random Forests (RF), Multilayer Perceptron (MLP)) and a statistical algorithm, Logistic Regression (LR). The results were analysed over six open source data sets, two of which were developed using C++ language while other four were developed using Java language.

The paper is organized as follows Section II states the related work and Section III describes the research background. Section IV presents an overview of various algorithms used in the study and Section V presents the analysis and results of the study. Section VI presents the threats to validity of the study and Section VII states conclusions and future directions.

II. RELATED WORK

Evaluating the change prone nature of a class is important in order to plan effective utilization of constraint resources and develop good quality maintainable software at low costs.

A number of researchers have tried to assess the change prone nature of a class by developing models with various OO metrics as independent variables. Lu et al. [4] used statistical meta-analysis technique and validated 62 OO metrics for their capability for predicting change prone classes on 102 Java systems. They found that size metrics have moderate predictive power for classifying change prone classes, followed by coupling and cohesion metrics and the inheritance metrics were shown to have the least predictive power. A study by Zhou et al. [5] investigated whether class size exhibits a confounding effect on the relationship between OO metrics and change proneness. They found an over estimation of this relationship due to confounding effect of size. A recent study by Elish et al. [3] evaluated a new suite of metrics called evolution based metrics for predicting change. They found that evolution based metrics along with widely used Chidamber and Kemerer (CK) metric suite [8] gave the best results while evaluating change. Malhotra et al. [2] assessed 22 OO metrics for predicting change prone classes. They found ML algorithms to be better than statistical methods for change prediction. Koru et al. [6] analysed that the modules with high values of cohesion, coupling, size and inheritance measures are different from the ones having highest number of changes.

Catal et al. [10] validated CK metric suite along with certain method level metrics in order to predict faults using AIRS algorithm. The validation was done on NASA data set. Afzal et al. [14] used AIRS and four other techniques for determining the number of faults which were not recognized in unit, function, integration or testing phases. AIRS performed better in unit and function testing phases than regression algorithm. Though, AIS algorithms have been used for predicting faults in the literature, to the best of author's knowledge there is no study till date which assesses the effectiveness of these algorithms for predicting change prone classes. This study will help researchers and software developers to assess the applicability of AIS algorithms for change prediction.

III. RESEARCH BACKGROUND

This sections states the dependent and independent variables, the data collection procedure and the hypothesis to be tested in this study.

A. Dependent and Independent Variables

The dependent variable used in the study is change proneness. It is the probability that a class will change after the software has gone into operational phase [2]. The independent

variables used in the study are CK metrics which include Coupling Between Objects (CBO), Lack of Cohesion amongst methods (LCOM), Weighted Methods of a Class (WMC), Depth of Inheritance Tree (DIT), Number of Children (NOC) and Response for a Class (RFC) [8]. Apart from these OO metrics we also included Source Lines of Code (SLOC), a metric for measuring length of the code. The SLOC gives the count of all the lines containing source code of the class which excludes comments. All these metrics were computed using Understand (www.scitools.com) software.

B. Data Collection

We used six software data sets for empirically validating and comparing the capability of AIS algorithms with other ML and statistical algorithms. In order to collect data, we first downloaded two versions each of all the six software data sets from <http://sourceforge.net> or <https://apache.org>. The six software data sets and the corresponding versions analyzed were Art Of Illusion (2.7, 2.9.2), HyperSQL Database (2.3.0, 2.3.1), JMeter (2.8, 2.9), Abra (0.9.8, 0.9.9) Celestia (1.4.1, 1.6.1) and Glest (1.0.10, 3.2.2). While the Art of Illusion (AOI) software is a 3D modelling, rendering and animation studio, HyperSQL (HSQL) database is a small multithreaded relational database engine, JMeter is a tool developed by Apache to perform load testing and Abra is a tool to create and share brain mappings. Celestia is an application for 3D space visualization while Glest is a 3D real-time customizable game.

TABLE I. DATA COLLECTION STATISTICS

Data Set	AOI	HSQL	Jmeter	Abra	Celstia	Glest
Lang.	Java	Java	Java	Java	C++	C++
SLOC	85,728	1,43,870	88,558	14,660	80,002	13,764
Data Points	434	510	335	180	254	108
Change Classes (%)	30%	41%	60%	18%	58%	66%

We first compute various OO metrics (independent variables) of the previous versions of all the data sets using Understand software. The next step is to prepare comparable versions for each data set. In order to so, we extract common classes from both the versions of the data set. All the classes with the same name which were present in both the versions of the data sets were extracted. If a class was newly added in the latest version, or deleted from the previous older version, it was discarded. We also exclude all library classes and interface files. Now we compare the comparable versions of a data set to find the number of lines added, deleted or modified in a class. The Understand tool helps us in this computation too, as it highlights changes in two files. The total lines added, deleted or modified in a class are summed up to get TOTAL CHANGE of a class. An ALTER variable is calculated for each class which is "yes" if the TOTAL CHANGE of a class is greater than zero otherwise it is "no". Finally, we combine the OO metrics, change statistics (number of lines added, deleted, modified and total change) and ALTER variable to produce data points. The number of data points in each data

set are actually the number of common classes in both the versions of corresponding data set. Table I shows the number of data points obtained from each data set, the percentage of changed classes (% of changed classes), the language (Lang.) in which the data set was developed and the size of the data set in SLOC. The procedure for data collection was similar to that followed by Malhotra et al. [2] and Zhou et al. [5].

C. Hypothesis

In order to assess the capability of AIS algorithms for predicting change prone classes, we test the following hypothesis.

H1 Alternate Hypothesis: The performance of AIRS algorithm is better than BG, AB, RF, MLP and LR algorithms for prediction of change prone classes. *Null Hypothesis:* The performance of AIRS algorithm is worse than BG, AB, RF, MLP and LR algorithms for prediction of change prone classes.

H2 Alternate Hypothesis: The performance of IM99 algorithm is better than BG, AB, RF, MLP and LR algorithms for prediction of change prone classes. *Null Hypothesis:* The performance of IM99 algorithm is worse than BG, AB, RF, MLP and LR algorithms for prediction of change prone classes.

H3 Alternate Hypothesis: The performance of CLG algorithm is better than BG, AB, RF, MLP and LR algorithms for prediction of change prone classes. *Null Hypothesis:* The performance of CLG algorithm is worse than BG, AB, RF, MLP and LR algorithms while for prediction of change prone classes.

IV. RESEARCH METHODOLOGY

This section gives a brief description of all the algorithms used in the study. The parameter settings used for each algorithm are the default settings of the WEKA tool [15] and the WEKA plugin used for AIS algorithms [16]. They are presented in Table II.

A. AIS Algorithms

Artificial Immune Recognition System (AIRS) is a supervised learning algorithm whose objective is to construct a pool of memory cells which are good representatives of training data such that they can easily classify unseen data [17]. AIRS uses resource limited approach. The algorithm consists of five steps where the first step involves preparation of data set for training and is called initialization. The next step is antigen training which is followed by competition of limited resource in the Artificial Recognition Ball (ARB) pool. In the next step, the memory cells with the maximum stimulation score are chosen for candidate memory cells. Finally, the memory cell pool can be used for classification and K-nearest neighbor approach is applied in this step.

Immunos99 (IM99) is an AIS algorithm used for classification where the data is divided into different antigen groups according to the classification labels. Now for each antigen group, an initial B-cell population is created. The next

step involves generation and mutation of different antibodies by calculating fitness scores, population pruning, affinity maturation and random selection. The last step involves the final pruning of each B-cell population [18].

Clonal Selection Algorithm (CLG) is based on the clonal selection principle which states that only the cells which efficiently recognize the antigens are selected for reproduction in successive generations. These selected cells are further exposed to affinity maturation process to improve their affinity to specific antigens [19].

B. Machine Learning Algorithms

Bagging (BG) involves creation of a number of versions of training sets. It incorporates training a new function for each set. While predicting the result, the class output by majority is the final output. It is successful because of the instability of the training method [20]. Adaptive Boosting (AB) accumulates a number of weak performing classifiers to generate a strong performing classifier. This can be done by building a number of models and gaining expertise on previously incorrectly classified instances. AB uses voting method and assigns higher weights to better performing models [21].

TABLE II. PARAMETER SETTINGS FOR ALL ALGORITHMS

Algorithm	Parameters
AIRS	Affinity threshold: 0.2; ARB initial pool size: 1; Clonal rate: 10.0; Hyper mutation rate: 2.0; knn: 3; Memory initial pool size: 1; Mutation rate: 0.1; Stimulation value: 0.9; Total resources: 150.
IM99	Minimum fitness threshold: -1.0; Seed: 1; Seed population percentage: 0.2; Total generations: 1.
CLG	Antibody pool size: 30; Clonal factor: 0.1; No. of generations: 10; Remainder pool ratio: 0.1; Selection pool size: 20.
BG	Bag sie percent: 100; Seed: 1; No. of iterations:10.
AB	Seed: 1; No. of iterations: 10; Weight Threshold: 100.
RF	No. of trees: 10; seed: 1.
MLP	Learning rate: 0.3; Momentum: 0.2; Seed: 0; Training time: 500; Validation Threshold: 20.

Random Forests (RF) consists of a number of tree predictors where each tree is dependent on a random vector. For each constituent tree in the forest, the random vector is sampled independently and with the same distribution. The mode of all the classes output by individual trees is the output class of RF [22]. Multilayer Perceptron (MLP) simulates complex relationships between input and output layer with the help of a number of intermediate layers. Back-propagation is used in order to train MLP networks where the weights of the network are adjusted in such a way so as to minimize the difference between desired and actual output [23]. Apart from the above described algorithms we also used logistic regression (LR), a statistical algorithm for developing change prediction models.

V. ANALYSIS AND RESULTS

This section states the results and briefly describes the performance measures used to assess the model. The models were created using ten-fold cross validation. The procedure

involves division of all the data points into ten subsets. We perform ten iterations by using nine subsets for training each time while using the tenth subset for validation. The procedure when repeated ten times gives the prediction results of the entire data set [24]. We perform ten runs for ten-fold cross validation. The performance measures used in this study were Accuracy and Receiver Operating Characteristic (ROC) analysis [25]. While accuracy is a measure of percentage of correctly predicted change prone and not change prone classes, ROC analysis plot a curve between sensitivity and 1-specificity values. The Area Under the ROC curve (AUC) is an indication of model's effectiveness.

A. Descriptive Statistics

We analyse the values of independent variables (OO metrics) for each data set and make the following observations.

- The size of the classes of software data sets used in this study vary from 4 – 3,053 SLOC (AOI), 4 – 2,801 SLOC (HSQL), 3 – 876 SLOC (JMeter), 2 – 932 SLOC (Abra), 5 – 3,242 SLOC (Celestia) and 10 – 1,972 SLOC (Glest).
- Inheritance property was not much used in these software data sets as the mean values of NOC (AOI - 0.46, HSQL – 0.56, JMeter – 0.33, Abra – 0.73, Celestia – 0.06 and Glest – 0.11) and DIT (AOI - 1.77, HSQL – 2.14, JMeter – 1.83, Abra – 1.81, Celestia – 0.89 and Glest – 0.74) are low. Other studies [2, 26] have also shown similar results.
- The LCOM metric which represents the number of classes without any common attribute usage has high values (Max. – 100) in all the data sets under study.

B. Ten-fold Cross Validation Results

In order to eliminate redundant and noisy features from the data set, we first applied Correlation Based Feature Selection (CFS) technique [27]. Table III states the independent variables selected in each data set after applying the CFS technique. According to the table, the SLOC metric is selected in five data sets indicating that it is a good predictor of change. The CBO metric selected by four data sets, the RFC metric and the WMC metric selected by three data sets are also good indicators of change.

TABLE III. CFS RESULTS

Dataset	OO Metrics Selected
AOI	CBO, SLOC, LCOM, WMC
HyperSQL	CBO, NOC, RFC, SLOC, LCOM, WMC
JMeter	CBO, NOC, SLOC
Abra	CBO, LCOM
Celestia	RFC, SLOC
Glest	RFC, SLOC, WMC

Figure 1 shows the mean accuracy values of each algorithm after applying ten runs of ten-fold cross validation on each of the six data sets used in the study. The accuracy is calculated

for each fold in each run of ten-fold cross validation and then the mean value is computed. The bars under each algorithm represent the mean accuracy value (in percentage) of the corresponding algorithm when applied on a particular data set. For example, AIRS algorithm gave an accuracy value of 67% on AOI data set, an accuracy value of 62% on HSQL data set. Similarly it gave accuracy values of 55%, 60%, 72% and 72% on Jmeter, Celestia, Glest and Abra data sets respectively. The accuracy values of all other algorithms on all other data sets can be easily seen in Figure 1. As seen in the figure, the mean accuracy values of AIS algorithms range from 43% to 78%. Thus, they can be effectively used for change prediction tasks.

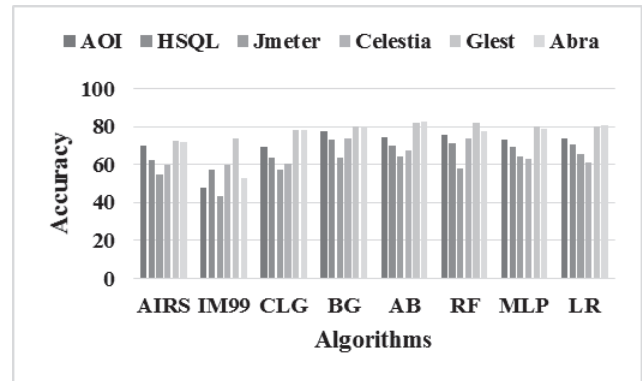


Fig. 1. Mean Accuracy values of all models

Figure 2 shows the mean AUC values after performing ten runs of ten-fold cross validation for each algorithm on the six data sets of the study. The AUC is calculated for each fold in each run of ten-fold cross validation and then the mean value is computed. The individual bars represent the AUC values of a particular algorithm on a specific data set. For example, LR algorithm gave an AUC value of 0.76 on AOI data set, and an AUC value of 0.75 on HSQL data set. Similarly it gave AUC values of 0.67, 0.71, 0.87 and 0.76 on Jmeter, Celestia, Glest and Abra data sets respectively. The mean AUC values of AIS algorithms range from 0.50 to 0.77. Thus, AIS algorithms can be effectively used for predicting change prone classes.

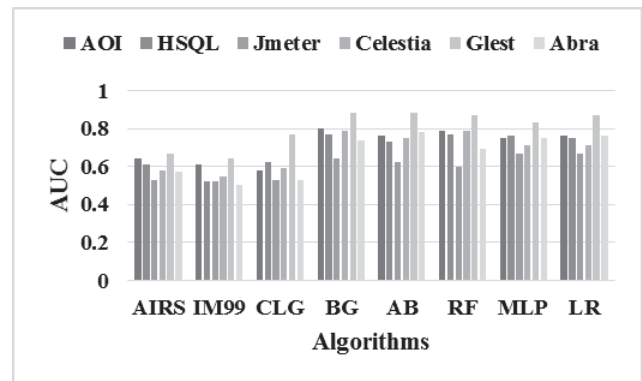


Fig. 2. Mean AUC values of all models

C. T-test Results for comparative analysis

This section states the comparative results of all the algorithms after application of paired t-test. Paired sample t-test is used in order to compare the means of the ten-fold cross

validation results for all the ten runs on each data set for two different algorithms. The differences between values of means of two different algorithms is computed at 0.05 significance level. The test calculates whether differences in performance measures among different pairs of algorithms is significant or not significant [28]. Table IV and V state the results of t-test using accuracy and AUC performance measure respectively. Each row of the tables states the pair of two algorithms compared and the columns represent the data set on which the results were generated. We use four symbols: 'S+', 'S-', 'NS+' and 'NS-' in these tables. 'S+' indicates that the performance of the first algorithm is significantly better than the performance of the second algorithm in the pair, 'S-' represents that the performance of the first algorithm in the pair is significantly poorer than the second algorithm of the pair. However, 'NS+' indicates that the performance of the first algorithm is better than the performance of the second algorithm in the pair but it is not significant statistically, 'NS-' represents that the performance of the first algorithm in the pair is poorer than the second algorithm of the pair but the not statistically significant.

TABLE IV. T-TEST RESULTS FOR AIRS USING ACCURACY

Algorithms	AOI	HSQL	Jmeter	Celestia	Glest	Abra
AIRS:BG	S-	S-	S-	S-	NS-	S-
AIRS:AB	NS-	S-	S-	NS-	S-	S-
AIRS:RF	S-	S-	NS-	S-	S-	NS-
AIRS:MLP	NS-	S-	S-	NS-	NS-	NS-
AIRS:LR	NS-	S-	S-	NS-	NS-	S-
IM99:BG	S-	S-	S-	S-	NS-	S-
IM99:AB	S-	S-	S-	S-	S-	S-
IM99:RF	S-	S-	S-	S-	S-	S-
IM99:MLP	S-	S-	S-	NS-	NS-	S-
IM99:LR	S-	S-	S-	NS-	NS-	S-
CLG:BG	S-	S-	NS-	S-	NS-	NS-
CLG:AB	NS-	S-	S-	S-	NS-	NS-
CLG:RF	S-	S-	NS-	S-	NS-	NS-
CLG:MLP	NS-	NS-	NS-	NS-	NS-	NS-
CLG:LR	NS-	S-	S-	NS-	NS-	NS-

The first five rows of Table IV and V are used to test hypothesis H1. In Table IV (rows 1-5), there are 18 cases in which the model developed by the AIRS algorithm gave significantly poorer accuracy values than the ones developed using other algorithms. The other 12 cases which compare AIRS algorithm with other algorithms also show that AIRS algorithm gives poor accuracy values. However, these cases are not statistically significant. The first five rows of Table V show that 29 times models developed using other (BG, AB, RF, MLP, LR) gave significantly better AUC values than AIRS models. This leads to acceptance of null hypothesis. These results indicate that the performance of the AIRS algorithm is not better than BG, AB, MLP and LR algorithms when we use the accuracy or AUC performance measures.

Rows 6-10 of Table IV and V are used to test hypothesis H2 which state that the results of IM99 and all

other compared algorithms (BG, AB, RF, MLP and LR) are significantly different. According to Table IV, there are 25 such cases in which the accuracy of the models developed using IM99 algorithm is significantly lower than the ones developed by other algorithms. Similarly, Table V shows in all the 30 cases, the models developed using other algorithms gave significantly higher AUC values. As all the cases show that IM99 models yield poor accuracy and AUC values, we conclude that the performance of IM99 algorithm is outperformed by all the other compared algorithms when we use accuracy or AUC as the performance measure of change prediction models. Thus, we accept the null hypothesis H2.

TABLE V. T-TEST RESULTS FOR AIRS USING AUC

Algorithms	AOI	HSQL	Jmeter	Celestia	Glest	Abra
AIRS:BG	S-	S-	S-	S-	S-	S-
AIRS:AB	S-	S-	S-	S-	S-	S-
AIRS:RF	S-	S-	NS-	S-	S-	S-
AIRS:MLP	S-	S-	S-	S-	S-	S-
AIRS:LR	S-	S-	S-	S-	S-	S-
IM99:BG	S-	S-	S-	S-	S-	S-
IM99:AB	S-	S-	S-	S-	S-	S-
IM99:RF	S-	S-	S-	S-	S-	S-
IM99:MLP	S-	S-	S-	S-	S-	S-
IM99:LR	S-	S-	S-	S-	S-	S-
CLG:BG	S-	S-	S-	S-	S-	S-
CLG:AB	S-	S-	S-	S-	S-	S-
CLG:RF	S-	S-	NS-	S-	S-	S-
CLG:MLP	S-	S-	S-	S-	NS-	S-
CLG:LR	S-	S-	S-	S-	NS-	S-

In order to test hypothesis H3 we examine the last five rows of Table IV and V. According to Table IV, in 19 cases, the accuracy results of the CLG algorithm are not significantly worse than the results of the models developed using other algorithms. The compared results show that the accuracy values yielded by the MLP algorithm is not significantly better than the CLG algorithm in any of the data set. In fact, no algorithm gave significantly better accuracy results than the CLG algorithm on Glest and Abra data sets. However, Table V shows that in 27 cases, the models developed using other algorithms gave significantly better AUC values than CLG algorithm. Thus, we accept null hypothesis H3.

D. Discussion of Results

This section discusses the various results presented in the study. As the most commonly selected metric for inclusion in change prediction model was the SLOC metric, which was selected for five of the six data sets after applying the CFS technique, it is an efficient indicator of change proneness. Further, the CBO metric, the RFC metric and the WMC metric are also highly correlated with change.

The mean of the accuracy measures of the change prediction models using AIS algorithms ranged from 43% to 82% while the mean of the AUC measures of the change prediction model ranged from 0.50 to 0.77. These results show that

though change prediction models using AIS algorithms may not always give good results, however in some cases they may yield appropriate change prediction models with good accuracy and AUC measures. However, when compared, other algorithms showed accuracy values from 57%-82% and AUC values from 0.6-0.8. This shows only slightly poor performance of AIS algorithms. This could be because of negative selection process, where the algorithms determines and eradicates cells that may attack self-tissues. Moreover, these algorithms are based on Darwinian principals and are evolutionary in nature. Thus, they may take time evolve and provide effective results due to project's changing characteristics while evolution. Thus, AIS algorithms should be explored by a number of other researchers in future studies to effectively determine the type of data sets they perform best on for development of a change prediction model.

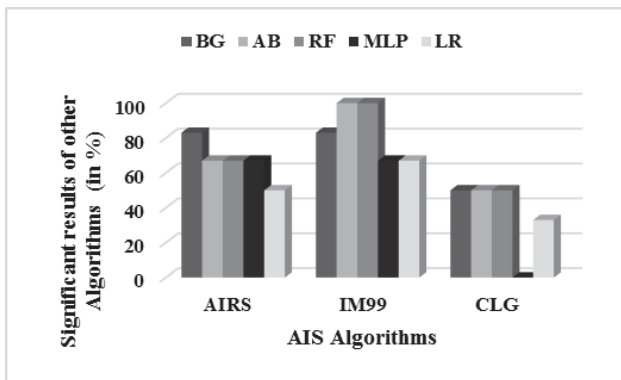


Fig. 3. Summary of Hypothesis using Accuracy measure

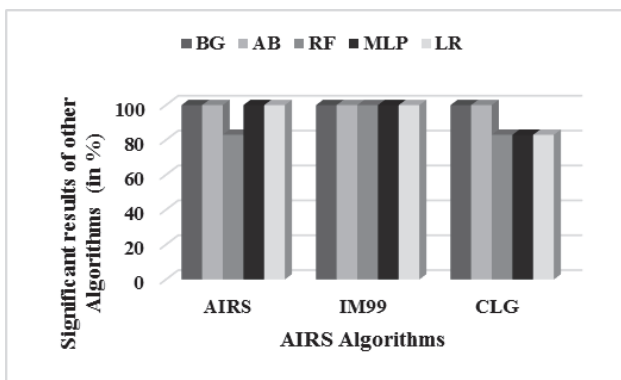


Fig. 4. Summary of Hypothesis using AUC measure

We performed paired t-test in order to compare the performance of AIS algorithms with other ML and statistical algorithms using accuracy and AUC measures. The models were compared at 0.05 significance threshold. Figure 3 and 4 show the summary of all the hypothesis results using accuracy and AUC measure respectively. In Figure 3, each bar represents the percentage of cases in which a specific algorithm (ML or statistical) significantly outperformed a specific AIS algorithm. For example, the BG algorithm outperformed the AIRS algorithm in 80% of the cases where these two algorithms were compared using Accuracy values. Similarly in Figure 4, it can be seen that the BG algorithm

performs better than AIRS algorithm in all the cases (100%), when these two algorithms are compared using AUC measure. As already stated in the Section V.C, all ML and statistical algorithms give better results (accuracy and AUC) than the AIS algorithms. The insignificant results of the remaining cases also show ML and statistical algorithms to be better than AIS algorithms. Thus, the t-test results conclude that the ML and statistical methods (BG, AB, RF, MLP and LR) outperform the AIS algorithms (AIRS, IM99 and CLG) while developing change prediction models and these results are significantly different from AIS algorithms. Thus, we accept the null hypothesis (H1, H2 and H3) of the study. However, a number of other studies using a varied size of data sets should be conducted in order to obtain generalized results about AIS algorithms. Catal et al. [10] demonstrated AIRS as a good algorithm for developing fault prediction models. This study shows AIS algorithms as effective tools for change prediction models. Out of the three explored AIS algorithms, the best results were shown by AIRS with accuracy values ranging from 60% to 72% and AUC values from 0.53 to 0.67. Though the results of CLG algorithm were lower than AIRS algorithm in majority of data sets but the CLG outperformed the other AIS algorithms in Glest data set with a mean AUC value of 0.77 and an accuracy of 78.12%. Similarly, the CLG algorithm showed exceptional results in Abra data set with an accuracy of 78.33%.

Apart from evaluating the algorithms according to their performance, we also evaluated the time taken by a specific algorithm to build change prediction model. The results showed that the MLP algorithm took the largest time to develop models, followed by AIRS algorithm. IM99 and CLG algorithms were faster than AIRS and in some cases competitive with faster algorithms like LR, AB, BG and RF for model development.

VI. THREATS TO VALIDITY

This section briefly accounts for the various threats to the validity of the study. The threat to construct validity in this study is minimized as various OO metrics have been validated in previous studies [29] as to whether they accurately measure the OO concepts like coupling, cohesion etc. However, the study does not take into account the "causal effect" of other factors except the investigated OO metrics on change prone nature of a class. Thus, the threat to internal validity exists in the study [5]. The threat to external validity is minimized to an extent by taking six software data sets (small size) developed in two languages (Java and C++). Thus the results presented are generalized to a certain extent. However, more studies should be conducted using varied size data sets (medium and large sized) in order to further reduce the threat to external validity.

VII. CONCLUSIONS AND FUTURE WORK

The objective of the study was to analyze the effectiveness of AIS algorithms for predicting change prone classes. In order to assess the capability of AIS algorithms we used six open source data sets developed in Java and C++ languages to develop change prediction models. The models were

developed by using CK and SLOC metrics as the independent variables and the change proneness attribute as the dependent variable. We further compared the results of AIS algorithms with other ML (BG, AB, RF and MLP) and statistical (LR) algorithms. The differences amongst various algorithms was statistically evaluated using t-test with 0.05 significance level. The conclusions of the study are summarized below:

- The SLOC metric is a primary indicator of change proneness followed by the CBO metric, the WMC metric and the RFC metric. These metrics were selected in most of the data sets for development of change prediction model after applying the CFS technique.
- The results show that the AIS algorithms (AIRS, IM99, CLG) are outperformed by other ML (BG, AB, RF and MLP) and statistical (LR) algorithms while predicting change proneness and using accuracy or AUC as the performance measure.
- Among the AIS algorithm, the best change prediction models was developed by the AIRS algorithm. The CLG algorithm gave very good results in only a few data sets. IM99 algorithm did not give encouraging results for development of change prediction models.

Change prediction models can be effectively used by software developers to prioritize maintenance and testing resources on change prone parts of the software and therefore yield good quality and highly maintainable software at low costs. In future, a number of studies should be conducted to analyze the effectiveness of AIS algorithms using medium sized and large sized data sets. This would help researchers and software practitioners to derive conclusive results over a wide array of software data sets. Moreover, a number of other bio-inspired and evolutionary algorithms should be evaluated for their effectiveness in developing change prediction models. Apart from accuracy measure and AUC measure, future studies may assess change prediction models using other performance measures such as f-measure, precision, g-measure etc. and perform paired t-test on these performance measures.

REFERENCES

- [1] Y. Singh and R. Malhotra, *Object-Oriented Software Engineering*. PHI Learning Pvt. Ltd. New Delhi, 2012.
- [2] R. Malhotra and M. Khanna, "Investigation of Relationship between Object-oriented Metrics and Change Proneness," *Int. J. Mach. Learn. & Cyber.*, 4, pp. 273-286, 2013.
- [3] M. O. Elish and M. A. Al-Khiaty, "A suite for quatifying historical changes to predict future change-prone classes in object-oriented software," *Journal of Software: Evolution and Process*, vol. 25, no. 5, pp. 407-437, 2013.
- [4] H. Lu, Y. Zhou, B. Xu, H. Leung and L. Chen, "The ability of Object-Oriented metrics to predict Change-proneness: a meta-analysis," *Empirical Software Eng.*, vol. 17, no. 3, pp. 200-242, 2012.
- [5] Y. Zhou, H. Leung and B. Xu, "Examining the Potentially Confounding Effect of Class Size on the Associations between Object Metrics and Change proneness," *IEEE Trans. Softw. Eng.*, vol. 35, no. 5, pp. 607-623, 2009.
- [6] A.G. Kori and J. Tian, "Comparing High-Change Modules and Modules with the Highest Measurement Values in two Large-Scale Open-Source Products," *IEEE Trans. Softw. Eng.*, vol. 31, no. 8, pp. 625-642, 2005.
- [7] K.K. Aggarwal, Y. Singh, A. Kaur and R. Malhotra, "Empirical Study of Object-oriented Metrics," *Journal of Object-Technology*, vol. 5, no. 8, pp. 149-173, 2006.
- [8] S. Chidamber and C. Kemerer, "A Metric Suite for Object- Oriented design," *IEEE Transactions on Software Engineering*, vol. 20, no. 6, pp. 476-493, 1994.
- [9] L. N. De Castro, and F. J. Von Zuben, "Artificial Immune Systems: Part I- Basic Theory and Applications," Technical Report, No. 01/99, 1999.
- [10] C. Catal, B. Diri, B. Ozumut, "An Artificial Immune System Approach for Fault Prediction in Object-Oriented Software," *Proceedings of Second Interantional Conference on Dependability of Computer Systems*, pp. 238-245, 2007.
- [11] L. N. De Castro, and F. J. Von Zuben, "Learning and Optimization using Clonal Selection Principle," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239-251, 2002.
- [12] L. Hong, "Artificial Immune System for Anomaly Detection," *Proceedings of Knowledge Acquisition and Modeling Workshop*, pp. 340-343, 2008.
- [13] K.A. Al-Shestawi, H.M. Abdul-Kader, N.A. Ismail, "Artificial Immune Clonal Selection Classification Algorithms for Classifying Malware and Benign processes using API call sequences," *International Journal of Computer Science and Network Security*, vol. 10, no. 4, pp 31-39, 2010.
- [14] W. Afzal, R. Torkar, R. Feldt, G. Wikstrand, " Search-based prediction of fault-slip-through in large software projects," *Proceedings of Second International Symposium on Search based Software Engineering*, pp. 79-88, 2010.
- [15] Weka 3: Software Mining in Java. Available online at <http://www.cs.waikato.ac.nz/ml/weka/>
- [16] Weka Plugin for AIS algorithms in Java. Available online at <http://weka.classalgos.sourceforge.net>.
- [17] J. Brownlee, "Artificial Immune Recognition System: A review and Analysis, Technical Report, No. 1-02, Swinburne University of Technology, 2005.
- [18] J. Brownlee, "IMMUNOS-81 The Misunderstood Artificial Immune System," Technical Report, No. 3-01, Swinburne University of Technology, 2005.
- [19] J. Brownlee, "Clonal Selection Algorithm & CLONALG: The Clonal Selection Classification Algorithm (CSCA), Technical Report, No. 2-02, Swinburne University of Technology, 2005.
- [20] L. Brieman, "Bagging Predictors," *Machine Learning*, vol. 24, pp. 123-140, 1996.
- [21] I.H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, USA 2005.
- [22] L. Brieman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [23] S. Haykin, *Neural networks: A comprehensive Foundation*, 2nd Edition, Pearson Education, Delhi, 2004.
- [24] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of Royal Statistics and Society*, vol. 36, pp. 111-147, 1974.
- [25] T. Fawcett, "An introduction to ROC Analysis," *Journal of Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2005.
- [26] CartWright, M. and Shepperd, M., 2000. An Empirical Investigation of an Object-Oriented Software System, *IEEE Trans. Softw. Eng.*, vol. 26, no. 8, 786-796.
- [27] M.A. Hall, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning," *Proceeding of the Seventeenth International Conference on Machine Learning*, pp. 359-366, 2000.
- [28] R. Malhotra, A. Kaur and Y.Singh, " Comparative Analysis of Ranom Forests with Statistical and Machine Learning Methods in predicting Fault-prone classes," *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies*, IGI-Global, pp- 428-449, 2012.
- [29] L. Briand, J. Daly, and J. Wust, "A unified framework for coupling measurement in object-oriented systems", *IEEE Trans. Softw. Eng.*, vol. 25, no. 1, pp. 91-121, 1999.



APPLICATION OF FUZZY BASED METHODOLOGY IN SELECTION OF VEHICLE ACCORDING TO FUEL TYPE

Kiran Pal¹, Surendra Tyagi², Manish Jain³

¹Assistant Professor, Department of Mathematics, Delhi Institute of Tool Engineering, Delhi,

²Research Scholar, Department of Computer Science, Delhi Technological University, Delhi,

³Department of Mechanical Engineering, RJIT, BSF Academy, Takenpur, Gwalior, India

Email: kiranpaldite@gmail.com, surendratyagi2000@gmail.com, manish_jn@rediffmail.com

Abstract: Number of Private vehicles in India is growing very rapidly. Poor public transportation system in most of the urban cities has enforced many middle class to switch to own a vehicle for their daily commuting of family commuting. This paper is an attempt to choose suitable vehicle according to fuel type using fuzzy based methodology. Time is changing rapidly; climatic condition due to vehicular pollution is going to worst. So in addition to economy a number of other important factors should be considered at the time of buying vehicle. Using the proposed methodology a new customer can easily choose a proper fuel type for his individual requirement.

Index term: Fuzzy, Hybrid, Petrol, Fuel, Diesel

I. INTRODUCTION

Our fossil fuel sources are limited and depleting quickly. Petrol and diesel will eventually get all used up. Atop this are concerns regarding burning exhaust gases and particulate matter from vehicles adding to air pollution. Considering all these factors, the recent focus has shifted towards the use of alternative fuel sources for mobility. India still falls short in terms of awareness though. A number of options are at hand, Electric Vehicles, Hybrid technology and Hydrogen fuel cells. Hydrogen fuel cell powered vehicles are the most sensible

source, but unfortunately it will take years to install the infrastructure in place. Hybrids and electric cars are the only clever option, however due to the complexities of engine construction they cost a fortune [1]

II. CONCEPT OF FUZZY LOGIC

Fuzzy logic has rapidly become one of the most successful of today's technologies for complex decision making applications. The reason for which is very simple. Fuzzy logic addresses such applications perfectly as it resembles human decision making with an ability to generate precise solutions from certain or approximate information. While other approaches require accurate equations to model real-world behaviors, fuzzy design can accommodate the ambiguities of real-world in human language and logic. Although genetic algorithms and neural networks can perform just as well as fuzzy logic in many cases, fuzzy logic has the advantage that the solution to the problem can be cast in terms that human operators can understand, so that their experience can be used in the design of the controller. This makes it easier to mechanize tasks that are otherwise already successfully performed by humans. [2] In a broad sense, fuzzy logic refers to fuzzy sets - a set with unclear boundaries. Examples of fuzzy sets are "hot," "tall," "medium," etc. In a narrow sense, fuzzy logic is a logical system that aims to formalize approximate reasoning. In fuzzy logic a fuzzy symbol can take any truth

values from the closed set $[0, 1]$ of real numbers thus generalizing the Boolean truth values. As the technology was further embraced, fuzzy logic was used in more useful applications.

III. FUZZY SETS BACKGROUND

The precision of mathematics owes its success in large part to the efforts of Aristotle and the philosophers who preceded him. In their efforts to devise a concise theory of logic, and later mathematics, the so-called "Laws of Thought" were posited. One of these, the "Law of the Excluded Middle," states that every proposition must either be True or False. Even when, Parmenides proposed the first version of this law around 400 B.C.E. There were strong and immediate objections: for example, Heraclitus proposed that things could be simultaneously, True and not True. It was Plato who laid the foundation for what would become fuzzy logic, indicating that there was a third region (beyond True and False) where these opposites "tumbled about." Other, more modern philosophers echoed his sentiments, notably Hegel, Marx, and Engels.[3] But it was Lukasiewicz who first proposed a systematic alternative to the bi-valued logic of Aristotle.

In the early 1900's, Lukasiewicz described a three-valued logic, along with the mathematics to accompany it. The third value he proposed can best be translated as the term "possible" and he assigned it a numeric value between True and False. Eventually, he proposed an entire notation and axiomatic system from which he hoped to derive modern mathematics. Later, he explored four-valued logics, five-valued logics, and then declared that in principle there was nothing to prevent the derivation of an infinite-valued logic.

Lukasiewicz felt that three- and infinite-valued logics were the most intriguing, but he ultimately settled on a four-valued logic because it seemed to be the most easily adaptable to Aristotelian logic.

IV. FUZZY BRIEFS

The mechanisms which make fuzzy logic machines work, it is important to realize what fuzzy logic actually is. Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth-values between "completely true" and

"completely false". As its name suggests, it is the logic underlying modes of reasoning which are approximate rather than exact. The importance of fuzzy logic derives from the fact that most modes of human reasoning and especially common sense reasoning are approximate in nature. The essential characteristics of fuzzy logic are as follows.

Knuth proposed a three-valued logic similar to Lukasiewicz's, from which he speculated that mathematics would become even more elegant than in traditional bi-valued logic.[4] His insight, apparently missed by Lukasiewicz, was to use the integral range 12 alternative failed to gain acceptance, and has passed into relative obscurity. It was not until relatively recently that the notion of an infinite-valued logic took hold. In 1965 Lotfi A. Zadeh published his seminal work "Fuzzy Sets" which described the mathematics of fuzzy set theory, and by extension fuzzy logic [5] This theory proposed making the membership function (or the values False and True) operate over the range of real numbers $[0.0, 1.0]$. New operations for the calculus of logic were proposed, and showed to be in principle at least a generalization of classic logic. It is this theory which we will now discuss.

- In fuzzy logic, exact reasoning is viewed as a limiting case of approximate reasoning.
- In fuzzy logic everything is a matter of degree.
- Any logical system can be fuzzified
- In fuzzy logic, knowledge is interpreted as a collection of elastic or, equivalently, fuzzy constraint on a collection of variables.
- Inference is viewed as a process of propagation of elastic constraints.

The third statement hence, defines Boolean logic as a subset of Fuzzy logic.

V. FUZZY SETS THEORY

Fuzzy Set Theory was formalised by Professor Lofti Zadeh at the University of California in 1965. [6] What Zadeh proposed is very much a paradigm shift that first gained acceptance in the Far East and its successful application has ensured its adoption around the world. A paradigm is a set of rules and regulations which

defines boundaries and tells us what to do to be successful in solving problems within these boundaries. For example the use of transistors instead of vacuum tubes is a paradigm shift - likewise the development of Fuzzy Set Theory from conventional bivalent set theory is a paradigm shift. Bivalent Set Theory can be somewhat limiting if we wish to describe a 'humanistic' problem mathematically

VI. SELECTION OF VEHICLE BASED ON FUELTYPE

To buy a vehicle is an important time to general person. He should keep each and every fact in

mind. Present day demand of petrol and gas cars is increasing and demand of diesel car is going to decrease because of likely same prices of diesel and petrol. [6] Figure 1 and 2 shows the comparisons of diesel and petrol vehicles market share. And it is clear from these figures but there are many factors which affect the choice for a vehicle according to fuel type. These factors are discussed in table 1 and 2. These factors may be suitably modified according to their applicability in a particular situation

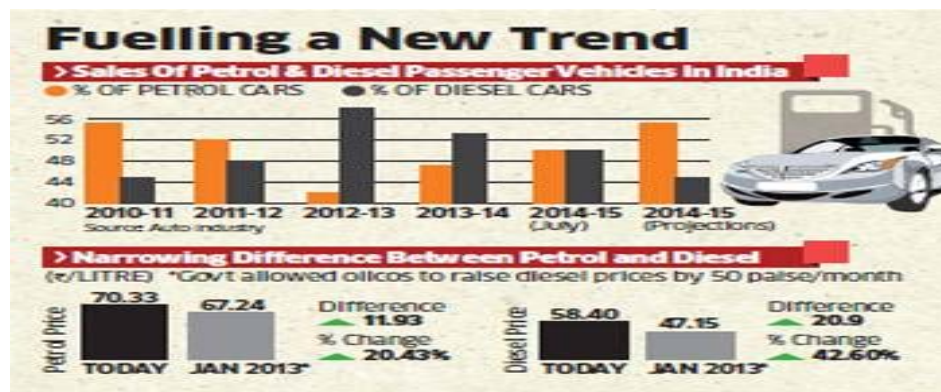


Figure 1: Sales trend of Petrol and diesel cars [7]

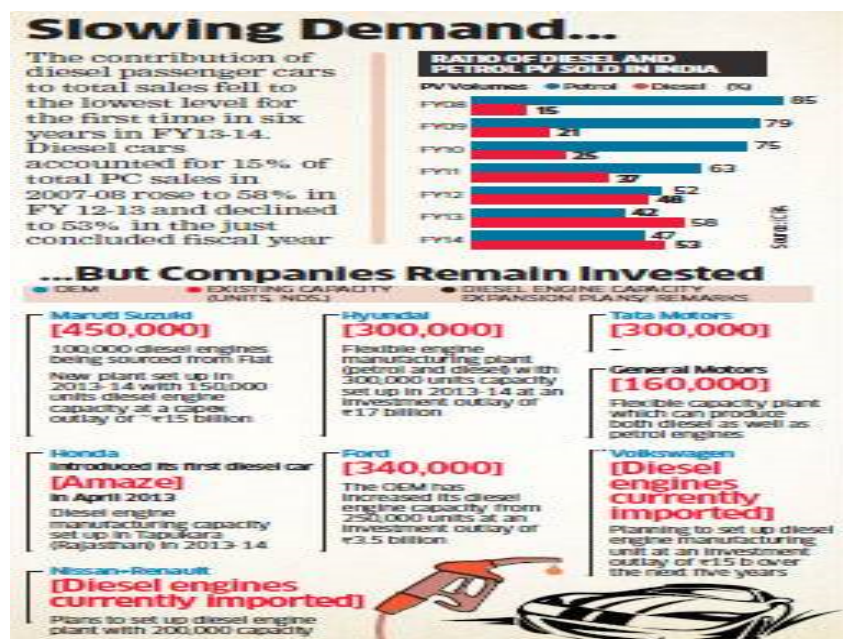


Figure 2: Reduction of the demand of Diesel Cars in India [8]

VII. SELECTION CRITERIA:

Selection of vehicle for a person is an important factor. In this paper, there are following 7 representative factors have been considered, though nit can be modified to incorporate any other factor.

1. Price
2. Environment Impact
3. Availability of fuel
4. Daily Running requirement
5. Load
6. Climate condition
7. Maintenance

Price: if a person does not have money, he cannot arrange his requirements but has to manage all things for his family in limited resources. In case of buying vehicle, he has to see all the aspects related to money and post expenses in fuel so price of fuel is important factor for a person who is buying vehicle.

Environment Impact: To keep clean environment is our responsibility. We should keep in mind this factor, when we do any work. So Environment Impact factor is also important in present scenario while pollution has become more discussion issue.

Availability of fuel: Friends indeed, friends in need, whether a fuel why not a so good but this is not available, it is meaningless for us. So we have to choose that fuel which is easily available.

Daily Running requirement: A person travel daily from one place to other, his requirement depends on this.

Load: According his purpose, vehicle has to bear that load.

Climate condition: Temperature, pressure, road condition are also factor , which has to keep in mind while buying vehicle.

Maintenance: To buy a vehicle is easy but, it is very difficult of its maintenance.

Table 1:Fuel Selection criteria and respective weight factors

S. No.	Selection factor	Notation	Weight factor	Person's Rating	Weighted rating
1	Price	P	W_1	R_1	$W_1 * R_1$
2	Environment Impact	E	W_2	R_2	$W_2 * R_2$
3	Availability of fuel	A	W_3	R_3	$W_3 * R_3$
4	Daily Running requirement	R	W_4	R_4	$W_4 * R_4$
5	Load	L	W_5	R_5	$W_5 * R_5$
6	Climatic condition	C	W_6	R_6	$W_6 * R_6$
7	Maintenance	M	W_7	R_7	$W_7 * R_7$
	Total weightage		$\sum W_i$	$\sum R_i$	$\sum R_i * W_i$

$$Membership\ value = \sum(R_i * W_i) / (\sum W_i * R_{max})$$

Illustrative example

Table 2: Illustrative example

			S.I.(Petrol & CNG)		C.I.(Diesel)		EL(SOLAR & HYBRID)	
S. No.	Selection factor	Weight factor	Person's Rating for C1	Weighted rating C1	Person's Rating for C2	Weighted rating C2	Person's Rating for C3	Weighted rating C3
1	Price	0.55	6	3.3	8	4.4	9	4.95

2	Environment Impact	0.9	6	5.4	5	4.5	9	8.1
3	Availability of fuel	0.7	7	4.9	7	4.9	5	3.5
4	Daily Running requirement	0.85	6	5.1	8	6.8	4	3.4
5	Load	0.75	7	5.25	8	6	4	3
6	Climatic condition	0.7	6	4.2	4	2.8	7	4.9
7	Maintenance	0.5	7	3.5	9	4.5	8	4
	Total	4.95		31.65		33.9		31.85
	Membership value		0.6393 94		0.684 848		0.643 434	

VIII. CONCLUSION:

Selection of the fuel type for the vehicle is a very complex problem for the first time vehicle buyers. Present fuzzy based approach is an attempt to help the vehicle buyer in the selection of type of fuel, It takes care of the environmental, availability of the fuel, climatic factors etc. Thus it is a suitable method for this problem.

IX. REFERENCES:

- [1] http://www.business-standard.com/article/news-cd/union-budget-why-should-electric-and-hybrid-cars-get-more-subsidies-114070801032_1.html
- [2] http://www.dortmundintelligence.com/index.php/technology_ci_fuzzy_systems.html
- [3] Fukumi S., Mizurnoto M., Tanaka K., **Some considerations of fuzzy conditional inference**, Fuzzy Sets and Systems, 4 (1980), pp. 243–273
- [4] Zadeh, L.A. (1971), "Similarity relations and fuzzy ordering", Information Sciences, 3, 177--200.
- [5] Zadeh, L.A. (1965), "Fuzzy sets", Information and control, 8, 338--353.
- [6] http://articles.economictimes.indiatimes.com/2014-08-20/news/53029171_1_petrol-cars-petrol-models-two-fuels
- [7] http://articles.economictimes.indiatimes.com/2014-07-05/news/51092090_1_diesel-vehicles-diesel-engines-new-diesel-plant
- [8] http://www.moneycontrol.com/news/business/buyingcar-why-petrol-vehicles-make-more-economic-sense_1180157.html.

CHATSEP: Critical Heterogeneous Adaptive Threshold Sensitive Election Protocol for Wireless Sensor Networks

R.K. Yadav

Department of Computer Engineering
Delhi Technological University
Delhi, India
rkyadav@dce.edu

Arpan Jain

Department of Computer Engineering
Delhi Technological University
Delhi, India
er.arpanj@gmail.com

Abstract— Networking together of hundreds to thousands of limited energy battery powered nodes derive a Wireless Sensor Network (WSN). The use and applicability of WSNs has increased in vivid areas like vehicular movement, weather monitoring, security and surveillance, industry applications etc. The limited powered nodes in WSNs sense the environment and send the desired information to a processing centre (base station) either directly or via a mechanism for optimization.

In this paper we propose CHATSEP, a clustering protocol for reactive networks with threshold sensitive heterogeneous sensor nodes. It includes an adaptive characteristic which helps the base station to be aware of the status of the nodes in situations when the nodes are idle from a long time and hence helping the base station to analyze the information network dynamically and efficiently. It also incorporates Critical Threshold which is any information of utmost importance in the network, and when such information is sensed it has to be sent to base station with highest priority. Our proposed protocol with adaptive nature and criticality of information is observed to perform better than the conventional clustering protocols like LEACH, SEP and TSEP in terms of stability period, network lifetime and throughput for a temperature sensing application.

Keywords— Network lifetime; Critical Threshold; Stability, Heterogeneity; Adaptive meter

I. INTRODUCTION

Recent advances in technology have led to a move from wired to wireless domain. Wireless sensor networks (WSNs) are derived of many limited on board storage capacity tiny battery powered sensor nodes [1]. These sensor nodes are deployed randomly or statically in the area of interest to perform desired operations as monitoring environmental and physical conditions like temperature, pressure, motion, sound etc. WSNs has great potential and is employed in variety of environments, however it comes with some restrictions as the sensor nodes used in WSNs have low power battery, limited processing capability, low bandwidth for communication and less memory space. Most of the energy of the sensor nodes is used in the transmission of the information to the sink. The

sensor nodes coordinate with each other which results in a communication network like a multi-hop network or a hierarchical network with several clusters and cluster heads (CHs) [2]. Cluster formation increases the performance of the WSNs by making efficient use of the energy of sensor nodes [3, 4, 5]. The CH selection amongst the nodes is done based on the probability or on the basis of remaining energy of the nodes. After CH selection nodes near to CH are associated with CH and transmit data to CH instead of base station (BS) to reduce transmission energy. CH then transmits only non-redundant necessary information to the BS which increases the network lifetime [6, 7]. Classification of WSNs, based on the target application can be understood as:

a. Proactive Networks: Sensor nodes transmit data periodically in such networks. After a fixed time quantum called frame time, nodes turn on the sensors to sense the environment and transmit the data to the CH or BS at regular intervals. They are suitable and used when information is required to be monitored periodically [8].

b. Reactive Networks: They are application specific protocols where nodes keep on sensing the environment continuously and transmit the information to CH or BS only when it observes a drastic change in the environment as per the defined threshold parameters. It reduces the transmission by not sending the data periodically and hence is efficient for time critical applications [8].

II. BACKGROUND

Clustering protocols can be classified broadly in two types on the basis of energy of the nodes; homogeneous and heterogeneous WSNs. In homogeneous networks all the nodes have same initial energy and they share equal participation in CH selection process whereas heterogeneous networks have small fraction of nodes with higher energy to play a greater role in CH selection process to increase the overall performance of the WSNs with little increase in the energy and hardware complexity. In this section some concerned protocols are described.

Direct Transmission, a traditional approach where sensors switch on their transmitters and send the sensed information directly to BS. Nodes which are near the BS may survive but nodes far from the base die quickly [3]. In MTE (Minimum Transmission Energy) protocol “the nodes transmit data to the BS with the help of multi hop communication via closest path to the BS which results in early depletion in the energy of the nodes near the BS” [1]. In SPIN protocol [9] “high level data descriptors called meta-data are used for efficient dissipation of information among nodes. It uses meta-data negotiations to eliminate the redundant transmissions in the network”.

A. LEACH (Low Energy Adaptive Clustering Hierarchy)

It is a self organising homogeneous proactive clustering protocol proposed in [10]. Sensor nodes with equal limited initial energy are used for randomized distribution of load in the network. The nodes are organised into local clusters with one CH. Nodes transmit data to their associated CH which then aggregates the data and sends only the required irredundant data to the BS, optimising the energy. “LEACH is categorized in two phases; cluster setup phase and steady state phase and to minimize the overhead, duration of steady state phase is higher than cluster setup phase” [10].

In the setup phase, a probabilistic threshold $T(n)$ is used to decide if the node can become a CH. Sensor node chooses a random number between 0 and 1 and if number falls lower than $T(n)$, the node can become the CH. $T(n)$ can be obtained as

$$T(n) = \begin{cases} \frac{P}{1 - P * \left(r \bmod \frac{1}{P} \right)} & \text{If } n \in G \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where P is the optimal percentage of CHs, G is group of nodes which have not become cluster head in $1/P$ rounds and r is the current round. CH is decided in a way that only once a node can become a CH in an epoch of $1/P$ rounds. Cluster formation is done by CHs advertising to other nodes which then associate themselves to nearest CH based on the signal strength of the advertisement. After local clusters are formed CH informs all the nodes in its cluster about when to send the information based on a TDMA schedule. Then in steady state phase, nodes switch on their sensors and transmit the data to CH periodically as per the parameters assigned by CH. CH aggregates the data and send it to BS. Steady state continues for certain period of time and then network goes to setup phase again with CH selection and local cluster formation followed by steady state. After LEACH was introduced, many protocols were proposed which imposed some improvements over LEACH and are generally termed as LEACH family of protocols. E-LEACH, TL-LEACH, V-LEACH are some of the protocols from LEACH family of protocols. E-LEACH [11] was LEACH with consideration of residual energy in the selection of CH. TL-LEACH is introduction to two hops in the LEACH protocol in consideration of reduction in the network traffic. V-LEACH proposed in [12] handled the cases of

failure of CHs by electing a vice-CH to handle the failures. Proactive nature of LEACH decreases its efficiency in terms of energy and is not suitable for time critical applications and less efficient in large areas.

B. SEP (Stable Election Protocol)

SEP is a heterogeneous proactive protocol with two levels of heterogeneity. “It improves the stable region of the clustering process using the parameters of heterogeneity, which are fraction of advanced nodes (m) and the additional energy factor (α) between advanced and normal nodes” [13]. SEP maintains balanced constraint energy of the network by assigning high probability for advanced nodes to become CH which is equal to the fairness constraint on energy consumption of the nodes. Optimal probability (P_{opt}) of the node to become CH remains the same as in [10] as heterogeneous network does not change spatial density of WSN. Weighted election probabilities of the heterogeneous nodes, P_{nrm} and P_{adv} can be calculated as:

$$P_{nrm} = \frac{P_{opt}}{1 + \alpha \cdot m} \quad (2)$$

$$P_{adv} = \frac{P_{opt}}{1 + \alpha \cdot m} * (1 + \alpha) \quad (3)$$

Cluster election in SEP is done randomly using weighted election probabilities of the advanced and normal nodes similar to [10]. After clusters are formed all the members of the cluster send data to CH which then processes the data and sends only the required irredundant information to the BS after processing the data. SEP performs better in respect to stability period and network lifetime when compared to family of LEACH protocols and uses the heterogeneity efficiently. However SEP is not suited for time critical applications because of its periodic nature.

C. TSEP (Threshold Sensitive Stable Election Protocol)

As described in [14] TSEP is a reactive protocol with three levels of heterogeneity. In TSEP transmissions are based on drastic changes in the environment which are monitored by the thresholds decided by the BS as per the application. Threshold Sensitive Environment reduces the transmissions to the BS and saves a lot of energy of the nodes. The nodes are categorised as normal nodes, intermediate nodes and advanced nodes based on the energy of the nodes. Energy of intermediate node lies between normal nodes and advanced nodes. The nodes are grouped into clusters and the CH election is done based on the weighted election probabilities of the nodes. Election probabilities are decided so as to maintain stable energy constraint in the network and hence increase the energy efficiency. It is based on two thresholds for transmitting the information to the CH, hard threshold and soft threshold. After CH selection and cluster formation is done, the nodes in the cluster keep on sensing the environment. When sensed value (SV) reaches equal to or greater than the defined hard threshold, nodes turn on the transmitters and send the sensed information to the associated CH. Then they transmit the data to the CH only when the difference between

the current sensed value and previously SV is greater than or equal to the soft threshold. Thus heterogeneity and thresholds optimizes the energy and increases the performance of the protocol.

In LEACH, parameters cannot be changed at cluster selection phase while it can be done in SEP and TSEP. SEP improves the performance of WSNs but heterogeneity increases throughput which affects the efficiency of the network. TSEP is threshold sensitive protocol which reduces the transmission to increase the performance but it does not take care of critical information. Also TSEP has a drawback that if the SV of nodes is not satisfying the hard and soft thresholds criteria, nodes will not transmit and BS will not get the information whether one or all nodes are alive or dead. Our proposed protocol provides a solution by taking into consideration the importance of critical information and introducing adaptive nature in all the nodes of the network to help the BS analyse and utilize the information network efficiently.

III. PROPOSED PROTOCOL

In this section we discuss our proposed protocol, CHATSEP (Critical Hybrid Adaptive Threshold Sensitive Election Protocol). CHATSEP comes with two special characteristics of “Criticality aware transmission” which gives utmost priority to any sensed information that is specified important by the network and “Adaptive nature” which dynamically keeps the BS aware of the status of the network in case if there is no transmission from long time due to reactive nature of the network. It helps BS to efficiently monitor the network with great ease and flexibility.

To get a clear understanding of the protocol we start with the first order radio energy model used. Energy consumed by the radio sensors for transmission of k bit information to a node at distance d with acceptable signal to noise ratio (SNR) is given by

$$ETx(k, d) = \begin{cases} E_{elect} * k + \mathcal{E}_{amp} * k * d^2 & \text{if } d \leq d_0 \\ E_{elect} * k + \mathcal{E}_{amp} * k * d^4 & \text{otherwise} \end{cases} \quad (4)$$

Where “ E_{elect} ” is the energy dissipated per bit to run the transmitter or receiver circuit, $\mathcal{E}_{amp}$ and $\mathcal{E}_{fs}$ are decided base upon the amplifier model” [10]. Optimal distance d_0 can be obtained as

$$d_0 = \sqrt{\frac{\mathcal{E}_{fs}}{\mathcal{E}_{amp}}}.$$

To receive a k bit information energy expended by the node is given by

$$ERx(k) = E_{elect} * k \quad (5)$$

Our protocol uses three levels of heterogeneity with normal, intermediate and advanced nodes in terms of initial energy of the nodes as in [14]. Initial energy of a normal node is E_0 and that of advanced nodes is given by $E_{adv} = E_0(1 + \alpha)$ where α is the fraction of higher energy for the advanced nodes from the normal nodes. Energy of intermediate nodes is between that of normal and advanced node and is given by $E_{int} = E_0(1 + \mu)$

where $\mu = \alpha/2$. Thus collective energy of the network will be $E_{tot} = n.E_0(1 + b\mu + m\alpha)$ with n as the total number of nodes, b as the fraction to total nodes which are intermediate and m as the proportion of total nodes that are advanced. The value of optimal probability is considered as in [10]. A CH selection is done based on the weighted election probabilities of the nodes. In our proposed network election probabilities are different based on the three levels of heterogeneity in the network. The weighted election probabilities [14, 15] are calculated as:

$$P_{nrm} = \frac{P_{opt}}{1 + \alpha * m + b * \mu} \quad (6)$$

$$P_{int} = \frac{P_{opt} * (1 + \mu)}{1 + \alpha * m + b * \mu} \quad (7)$$

$$P_{adv} = \frac{P_{opt} * (1 + \alpha)}{1 + \alpha * m + b * \mu} \quad (8)$$

CH selection is done based on the thresholds. Each node generates a random number between 0 and 1 and if this value is less than its particular threshold then node becomes a probable candidate for the position of CH. Calculation of thresholds as explained in [14, 15] is based on the probabilities of the nodes which can be obtained as:

$$T_{nrm} = \begin{cases} \frac{P_{nrm}}{1 - P_{nrm} * \left(r \bmod \frac{1}{P_{nrm}} \right)} & \text{If } n_{nrm} \in G' \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

$$T_{int} = \begin{cases} \frac{P_{int}}{1 - P_{int} * \left(r \bmod \frac{1}{P_{int}} \right)} & \text{If } n_{int} \in G'' \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$T_{adv} = \begin{cases} \frac{P_{adv}}{1 - P_{adv} * \left(r \bmod \frac{1}{P_{adv}} \right)} & \text{If } n_{adv} \in G''' \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Where G' is the group of normal nodes which have not taken the responsibility of becoming CH in last $1/P_{nrm}$ rounds, G'' is the group of intermediate nodes that were not the CH in last $1/P_{int}$ rounds and G''' are the group of advanced nodes that were not the CH in last $1/P_{adv}$ rounds. Average numbers of CH is same as that of LEACH, SEP and TSEP. During cluster formation phase in CHATSEP, the elected CHs first broadcast the following parameters.

1) *Attributes (A)*: These are physical parameters in which user or BS is interested in obtaining the desired information.

2) *Hard Threshold (H_T)*: It is an absolute value which can trigger the transmission of the nodes. If the SV of the cluster node becomes equal to or greater then hard threshold it will

switch on the transmitter and will send the sensed information to the associated CH.

3) *Soft Threshold (S_T)*: It is the small component of the thresholds. If difference between SV and current value is equal to or greater then S_T then node will turn on its transmitter and transmit the information to the CH.

4) *Critical Threshold (C_T)*: This is a new parameter we have introduced in our proposed protocol to give highest priority to information of importance in the network. For a temperature application, priority of information can be decided by C_T . If SV of any node is equal to or greater than C_T , it checks if it has sufficient energy to send the information directly to the BS or not. If yes then it transmits directly to the BS else information goes to BS via CH. Thus it gives highest priority to critical information by responding instantly and decreasing the response time.

5) *Adaptive Meter (A_M)*: It is another parameter we have introduced to keep the BS aware of the status of the network in case of lots of continuous rounds with no transmissions in the network. If a node has not transmitted data consecutively for A_M number of rounds then the node initiates a transmission irrespective of the thresholds and informs the base station about its existence. Optimal value of A_M depends on the application. After these parameters are broadcasted by CH, working of our protocol can be understood with the help of the following flow diagram:

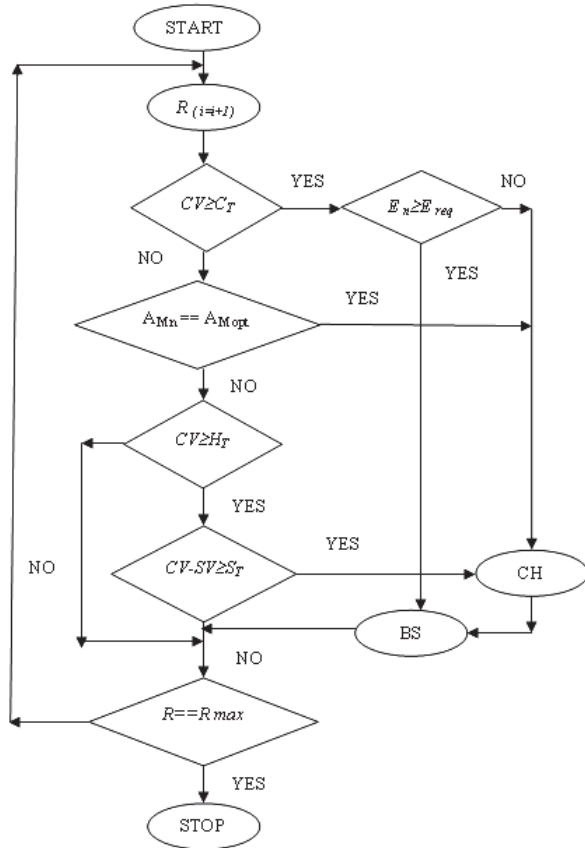


Fig.1. Working flow of CHATSEP

If the current sensed value (CV) is equal to or greater than C_T , nodes switch on their transmitter and send the data directly to the BS or CH based upon the remaining energy of the node. Thus gives highest priority to the critical information. Otherwise the member nodes check if they have not transmitted the data for contiguous A_M opt rounds. If yes then they initiate a transmission to the CH to make BS aware of the status of the nodes irrespective of the sensed values of the nodes. If both the aforementioned criteria's are not satisfied then node transmits only when the SV is greater than H_T and the difference between the CV and previous SV is greater than S_T . Hence it not only reduces the transmissions using the thresholds but also maintains the information network intact for the BS and ensures to give highest priority to critical information.

Each round (R) comprises of cluster setup phase followed steady state phase as in [10] and it also includes the critical information check and A_M opt based transmission. The transmission of sensed information in every round is adaptive in nature and is dependent on various aforementioned broadcasting parameters. R_{max} is the count of maximum number of rounds required for the particular WSN and is application dependent. Some of the features of CHATSEP are:

1) Critical information is communicated instantly to the BS.

2) BS is always dynamically aware of the network structure and status of the nodes which are idle.

3) User can modify the broadcasting parameters like A_M , C_T etc after each cluster set up phase as per the requirements of the application and can enhance the lifetime of the network.

Although transmissions based on A_M increases little complexity i.e. it requires a flag to keep the count of the number idle rounds. But this is a reasonable trade-off and provides additional versatility and flexibility for the network. C_T gives highest priority to data specified important by the application. With BS at the centre, C_T based transmission is observed to be efficient than usual reactive transmission as when CH is not involved E_{Rx} and data aggregation energy at CH is saved which is more than required for transmission to BS and hence increases the performance of CHATSEP.

IV. SIMULATIONS AND RESULTS

Matlab is the tool we used for simulation and performance evaluation. Our aim with the simulation was to compare LEACH, SEP, TSEP with our CHATSEP protocol in respect to energy and network lifetime. Performance metrics we used are:

- 1) *Alive Nodes*: No. of alive nodes in each round.
- 2) *Throughput*: No. of packets transmitted to BS from CH.
- 3) *A_M Index*: No. of adaptive transmissions initiated to keep the BS informed.
- 4) *C_T Scale*: No. of nodes sensing critical information per round.
- 5) *Network Lifetime*: Sum of stability period (time period till all the nodes are alive) and instability period (time period from the death of first node till all nodes die).

TABLE I
PARAMETER SETTING

Parameters	Value
E_{elect}	50nJ/bit
E_{DA}	5nJ/bit/message
$\mathcal{E}_{fs}$	10pJ/bit/m ²
$\mathcal{E}_{amp}$	0.0013pJ/bit/m ⁴
E_o	0.5J
K	4000
N	100
P_{opt}	0.1
A	1
M	0.1
A_M	10,15
C_T	170,175

100 immovable nodes are deployed randomly in the 100*100 field dimension network with BS at the centre. The field has a varying temperature in different regions with temperature range (50⁰, 200⁰) with optimal values of the broadcasting parameters mentioned in Table 1.

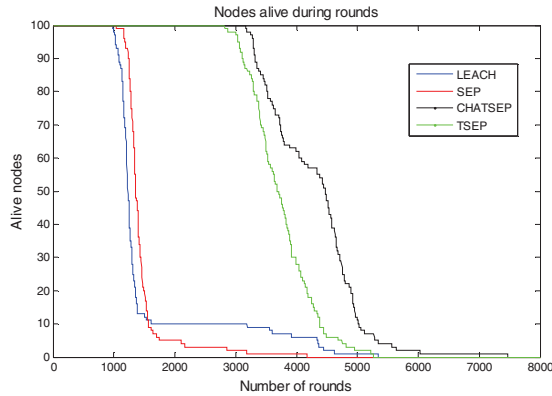


Fig. 2 No. of alive nodes

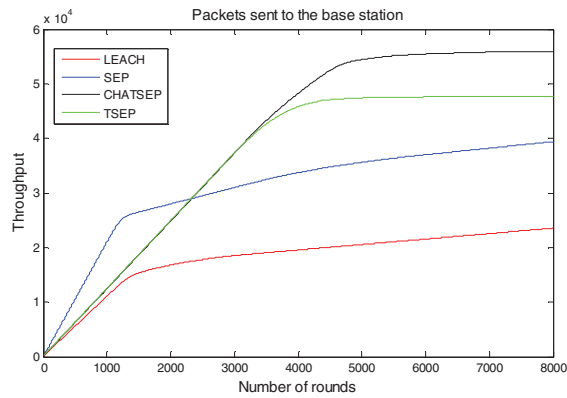


Fig. 3 Packets transmitted to BS from CHs

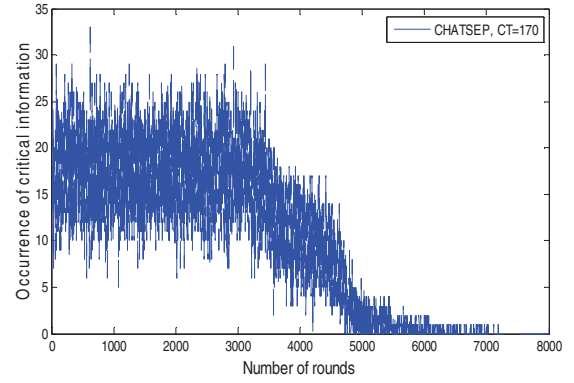


Fig. 4 No. of critical transmissions per round

We have simulated the proposed protocol for different values of C_T and A_M and kept the energy parameters (α, μ) and fraction of nodes (m, b) constant to observe the change in the network lifetime, C_T frequency and A_M based transmissions for different values of C_T and A_M . For first case (fig.2-5) we have taken ($C_T=170, A_M=10$) and for second case (fig. 6-9) we have taken ($C_T=175, A_M=15$). As $P_{opt}=0.1$, the weighted election probabilities are taken from (6), (7) and (8) whereas thresholds for CH selection are based on (9), (10) and (11).

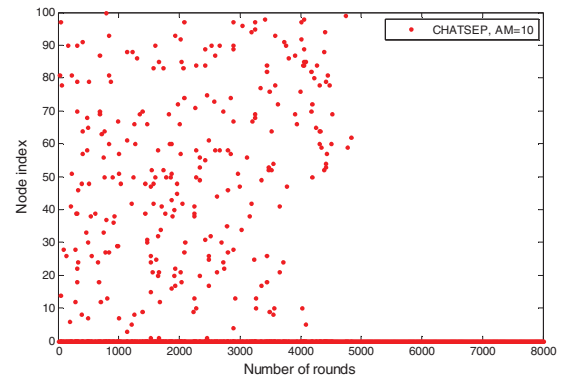


Fig. 5 A_M based adaptive transmissions

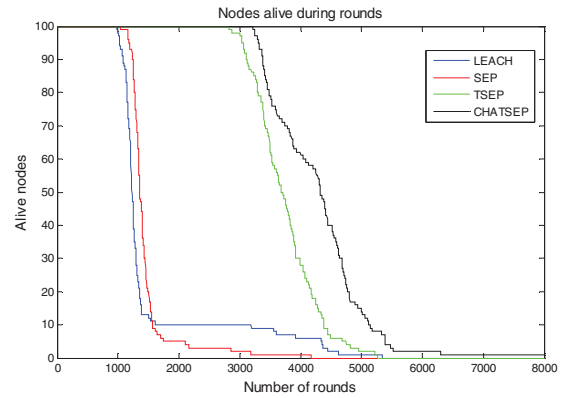


Fig. 6 No. of alive nodes

Fig. 2 and fig. 6 shows comparison between LEACH, SEP, TSEP and CHATSEP based on the number of alive nodes in the network and the network lifetime. Simulations are performed for 8000 cycles. SEP and LEACH with heterogeneous behaviour results in close stability period. If homogeneous LEACH and SEP are compared, SEP has much higher stability period. TSEP is observed to perform better than LEACH and SEP as it is threshold driven protocol with three levels of heterogeneity. CHATSEP with C_T , A_M is observed to perform better than aforementioned protocols because of its threshold sensitive mode of transmission which decreases the frequency of transmissions and C_T based transmissions which are observed to be efficient for a network with BS at the centre when CHs are not involved. Three levels of heterogeneity also contribute in enhancing the performance of CHATSEP when compared to LEACH and SEP. By varying the values of C_T (170 to 175) and A_M (10 to 15) it can be seen that as A_M increases stability period also increases (3173 to 3220) as the number of transmissions are reduced. Optimal values of C_T and A_M are based on the application and user's efficiency requirements.

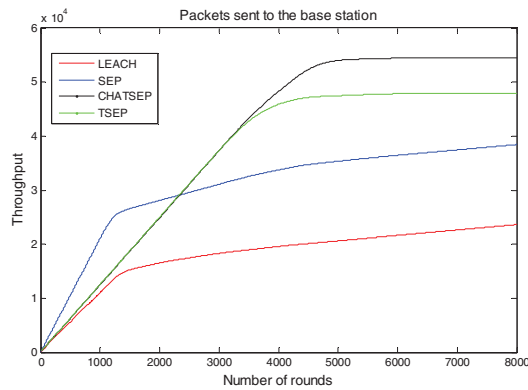


Fig. 7 Packets transmitted to BS from CHs

Fig. 3 and 7 represents throughput of the network, the number of packets sent to the BS by the CHs. Throughput of CHATSEP and TSEP is higher than LEACH and SEP because of three levels of heterogeneity. Throughput of CHATSEP is more than TSEP because although both protocols are threshold sensitive and has three levels of heterogeneity but introduction of C_T and A_M increases the communication for achieving dynamic behaviour in knowing about the status of idle nodes. Also as the network lifetime increases the throughput also increases. By varying the values of C_T (170 to 175) and A_M (10 to 15) we observe a decrease in the throughput (5.95×10^4 to 5.45×10^4) because of reduction in the overall transmissions.

Fig. 4 and fig. 8 shows the occurrences of critical information in each round. We have observed that there are instances when some information is of great importance to the BS and our protocol sends such information to BS with highest priority which enhances the communication network. We observed a decrease in C_T transmissions from $C_T=170$ to $C_T=175$ and the

value of C_T is dependent upon the requirements of the application and can be changed in every round of CH selection. Thus all the critical sensed information is sent to the BS with utmost priority.

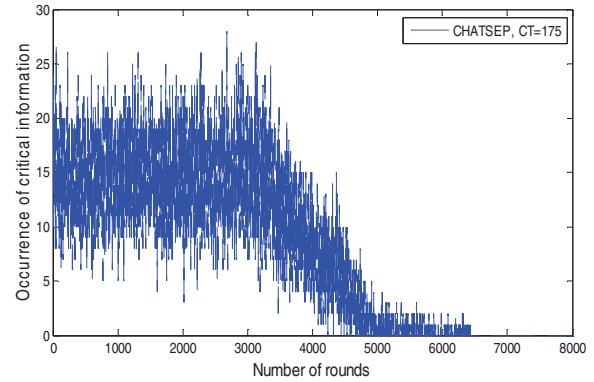


Fig. 8 No. of critical transmissions per round

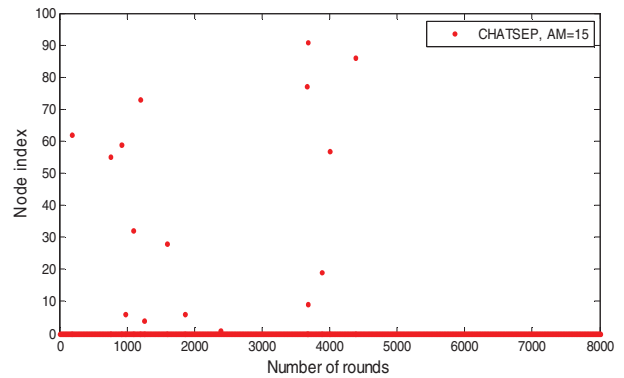


Fig. 9 A_M based adaptive transmissions

Fig. 5 and fig. 9 show the adaptive characteristic of CHATSEP for $A_M=10$ and $A_M=15$. These figures represent which node (indexed from 1 to 100) is performing A_M based transmission with respect to simulation rounds. It can be seen that number of adaptive transmissions decrease as the value of A_M increases as there are more idle nodes which have not transmitted information from 10 consecutive rounds as compared to those with 15 consecutive rounds. Increase in the value of A_M increases the network lifetime. Hence value of A_M is assigned on how often the BS needs to know about the status of the network.

From the simulations performed we have observed that stability period and network lifetime of CHATSEP is higher than related protocols. CHATSEP has a higher throughput as compared to other protocols because of adaptive transmissions and higher lifetime which also maintains network information structure intact. Critical and adaptive characteristics are implemented and are observed to be efficient and useful in time critical and safety aware applications where data is not required periodically but the information sensed is prioritized and dynamic status of the network is needed along with threshold sensitive monitoring.

V. CONCLUSION

In this paper we introduced CHATSEP, a heterogeneous threshold sensitive election protocol with critical information awareness and adaptive characteristics. The features introduced are flexible to be monitored in any round of CH selection depending on user's requirement. Our protocol is observed to perform better than TSEP, SEP and LEACH in terms of network lifetime and also keeps the BS aware of the status of the network along with transmission of the critical information with highest priority.

REFERENCES

- [1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks", *IEEE Communications Magazine*, vol. 40, no. 8, August 2002, pp. 102–114.
- [2] R. Rajagopalan, and P.K. Varshney, "Data-Aggregation techniques in sensor networks: a survey", *IEEE communications surveys*, vol. 8, no. 4, 2006.
- [3] W. R. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan "An application-specific protocol architecture for wireless microsensor networks", *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, October 2002.
- [4] F.A. Aderohunmu, and J.D. Deng, "An Enhanced Stable Election Protocol (SEP) for Clustered Heterogeneous WSN", *Department of Information Science*, University of Otago, New Zealand.
- [5] G. Wang et al., "The Clustering Algorithm of Wireless Sensor Networks Based on Multi-hop between Clusters", *Computer Science and Information Engineering*, vol. 3, March-April 2009, pp. 177–181.
- [6] N.S. Patil, and P.R. Patil, "Data aggregation in wireless sensor network", *IEEE Int Conf. on Computational Intelligence and Research*, 2010.
- [7] N. Kimura, S. Latifi, "A Survey on Data Compression in Wireless Sensor Networks", *Proc. Int. Conf. on Information Technology: Coding and Computing (ITCC '05)*, 2005..
- [8] A. Manjeswar and D.P. Agrawal, "TEEN: A protocol for enhanced efficiency in wireless sensor networks", *Proc. 1st Int. Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing*, San Francisco, CA, USA, 2001.
- [9] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks", *Proc. 5th Annual ACM/IEEE Int. Conf. on Mobile Computing and Networking*, Seattle, Washington, USA August, 1999, pp.174-185.
- [10] W. R. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "Energy efficient communication protocol for wireless microsensor networks", *Proc. 33rd Hawaii Int. Conf. on System Sciences (HICSS-33)*, January 2000.
- [11] X. N. Fan and Y. L. Song, "Improvement on LEACH protocol of wireless sensor network", *Proc. Int. Conf. on Sensor Technologies and Applications, Sensor Comm*, 2007, pp. 260-264.
- [12] M. B. Yassein et al. "Improvement on LEACH Protocol of Wireless Sensor Network (VLEACH)", *Int. Journal of Digital Content Technology and its Applications*, Vol. 3, No. 2, June 2009
- [13] G. Smaragdakis, I. Matta, A. Bestavros, "SEP: A Stable Election Protocol for clustered heterogeneous wireless sensor networks", *2nd Int. Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004)*, 2004.
- [14] A. Kashaf, N. Javaid, Z. A. Khan and I. A. Khan, "TSEP: Threshold-sensitive stable election protocol in wireless sensor network", *10th IEEE Int. Conf. on Frontiers of Information Technology*, 2012.
- [15] Femi A. Aderohunmu, Jeremiah D. Deng, "An Enhanced Stable Election Protocol (SEP) for Clustered Heterogeneous WSN", *Discussion Paper Series*, No. 2009/07, October 2009, ISSN: 1177-455X, Department of Information Science, University of Otago, (2010).

Circuit Analysis Using Matlab/Simulink®

Pooja Verma

Electrical Engineering Department, Delhi Technological University (formerly DCE),
Main Bawana Road, Delhi-110042, India

Abstract: Circuits having various combinations of R, L and C are the building blocks for many important areas of study in electrical engineering and also used extensively in communication systems for design and development of filters and oscillators. Hence, effective simulation (or prediction) of such systems is imperative. This paper explores the ability of MATLAB/Simulink® to achieve this feat with relative ease- either by writing MATLAB code commands or via Simulink® for linear Initial Value Problems (IVPs). The applicability of the MATLAB/Simulink® also finds use in the integrated electrical drives.

Keywords: MATLAB/Simulink®, RLC Circuits, Circuit Response Analysis, ODE Solutions

1. INTRODUCTION

An LC circuit, also called a resonant circuit, tank circuit, or tuned circuit, consists of an inductor, represented by the letter L, and a capacitor, represented by the letter C. When connected together, they can act as an electrical resonator, an electrical analogue of a tuning fork, storing energy oscillating at the circuit's resonant frequency. A major application of tuned circuits can be seen in television and Radio sets. Many different signals reach the antenna of a radio receiver at the same time. Within the radio or TV receiver, the actual "selecting" of the desired signal and the rejecting of the unwanted signals are accomplished by the use of tuned circuits. Another combination forms the class of filters. Filters are technical realizations of given system functions, which affect the spectral characteristics of an input signal in the main (Frequency selection). In the context of electro-technology the realizations with electrical networks interest as analog and digital circuits. Matlab/Simulink® can play a significantly important role in the pre-implementation stages of such circuits by providing a platform for modeling real world systems with appreciable accuracy.

2. ELECTRICAL ENGINEERING APPLICATION

2.1. Series RL circuit

The circuit consists of a resistance R and an inductor coil L connected in series. Fig.1. below gives a pictorial view of the typical circuit arrangement in consideration.

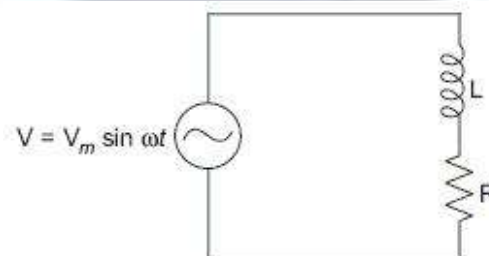


Fig.1. Typical RL series circuit

The first order ordinary differential equation² that describes a simple series electrical circuit with a resistor, inductor and sinusoidal voltage source is as follows:

$$\frac{di}{dt} + \frac{1}{\tau_{RL}} i = \frac{1}{L} v_s$$

$$\tau_{RL} = \frac{L}{R}$$

For this example, the inductance, L is 1 Henry and the resistance R is 10 Ω . The voltage source is sinusoidal with a peak voltage equal to 10 volts and an angular velocity of 150 rad/sec. The initial conditions have been assumed zero, i.e. $i(0)=0$.

The MATLAB statement written to solve (1)

symbolically is as follows:

```
c=dsolve('Dc=-10*c+10*sin(150*t)', 'c(0)=0');
```

```
ezplot(c, [0 0.5]), grid
```

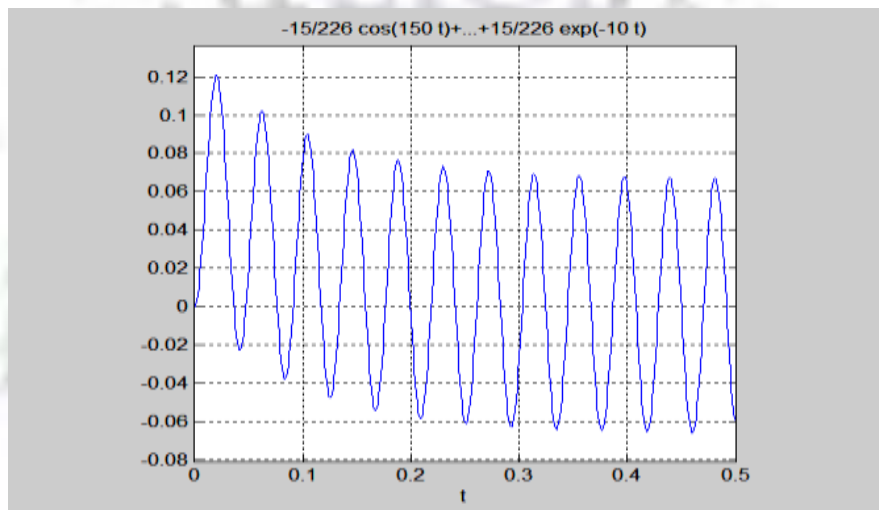


Fig. 2. Graphical result of the RL circuit dynamics

```
dsolve('eq1,eq2,...','cond1,cond2,...','v')
```

Symbolically solves the ordinary differential equation(s) specified by (1), (2), using v as the independent variable and the boundary and/or initial condition(s) specified by $\text{cond1}, \text{cond2}$. The default independent variable is t . The letter 'D' denotes differentiation with respect to the independent variable; with the primary default, this is d/dx . A 'D' followed by a digit denotes repeated differentiation. For example, $D2$ is d^2/dx^2 . Any character immediately following a differentiation operator is a dependent variable. For example, $D3y$ denotes the third derivative of $y(x)$ or $y(t)$. Initial/boundary conditions are specified with equations like $y(a) = b$ or $Dy(a) = b$, where y is a dependent variable and a and b are constants. If the number of initial conditions specified is less than the number of dependent variables, the resulting solutions will contain the arbitrary constants $C1, C2, \dots$. The output from the dsolve function is the symbolic solution to the differential equation. Next, the differential equation describing the RL electrical circuit is modeled using Simulink®. Simulink® is a block diagram environment for multi-domain simulation and Model-Based Design. It supports system-level design, simulation, automatic code generation and continuous test & verification of embedded systems. With Simulink®, the differential equation is described using blocks from Simulink® library. The complete Simulink® model for the electrical circuit is depicted in Fig. 3. Note that a time clock is added to the Simulink® model to enable the exportation of the simulation time to Matlab workspace for accessibility to plot simulation time against the current.

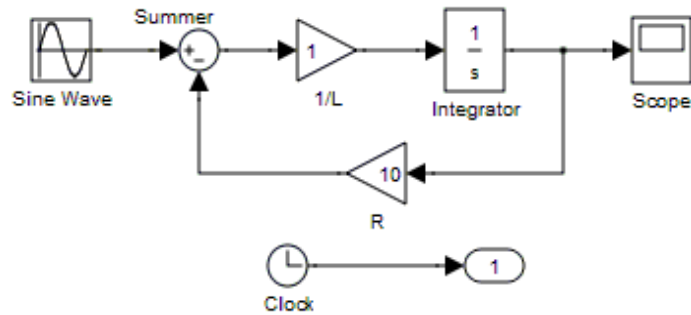


Fig. 3. Simulink® model of the RL series circuit

The results obtained from the Simulink® model of the RL circuit are presented below:

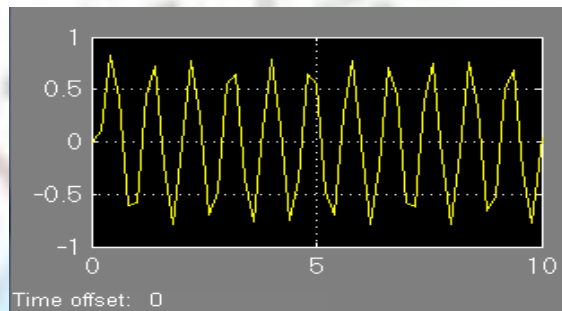


Fig. 4. Obtained result from Simulink® model of RL circuit

2.2. Series RC circuit

The circuit consists of a resistance R and an capacitor C connected in series. Fig. 4 below gives a pictorial view of the typical circuit arrangement in consideration.

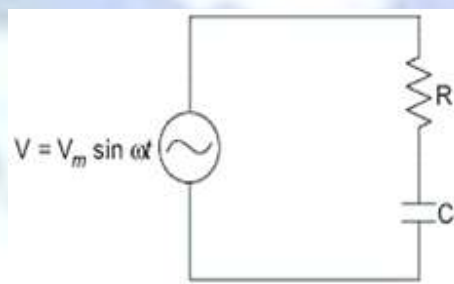


Fig.5. Typical RC series circuit

The first order ordinary differential equation that describes a simple series electrical circuit with a resistor, inductor and sinusoidal voltage source is as follows:

$$\tau_{RC} \frac{dv_C}{dt} + v_C = v_s$$

$$\tau_{RC} = RC$$

For this example3, the capacitance C is 4mF and the resistance R is 10 Ω . The voltage source is sinusoidal with a peak voltage equal to 85 volts and an angular velocity of 150 rad/sec. The initial charge on the capacitor has been assumed to be -0.05C, i.e.

$q(0) = -0.05C$.

The MATLAB statement written to solve (3) symbolically (i.e., obtaining exact solution) is as follows:
`q=dsolve('Dq=-25*q+8.5*cos(150*t)','q(0)=-0.05');`

`i=diff(q);`

`ezplot(q, [0 0.5]), grid`

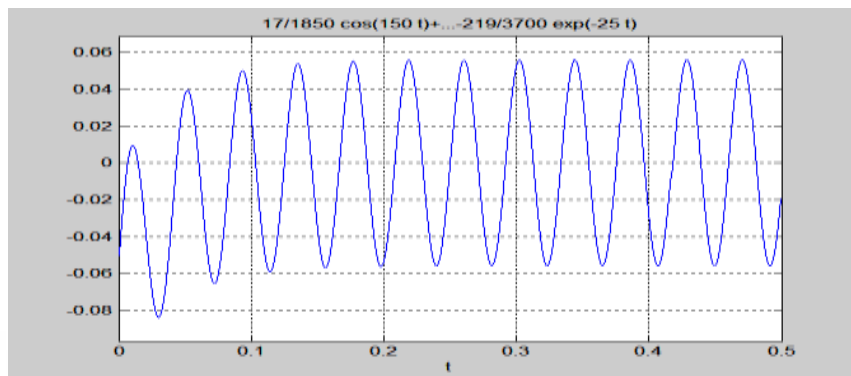


Fig. 6. Graphical representation of RC circuit dynamics Next, the differential equation describing the RC electrical circuit is modeled using Simulink®.

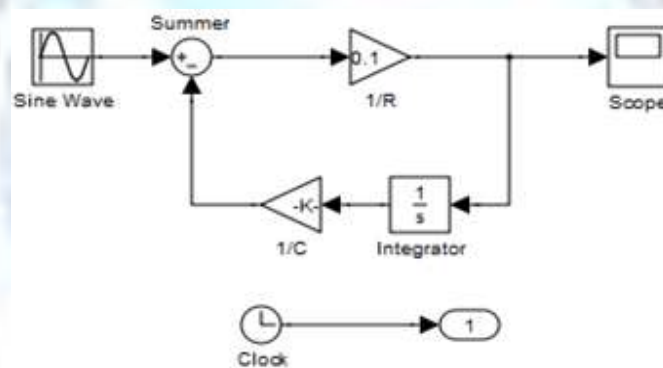


Fig. 7. Simulink® model of the RC series circuit

The following results are obtained when the said simulation is run:

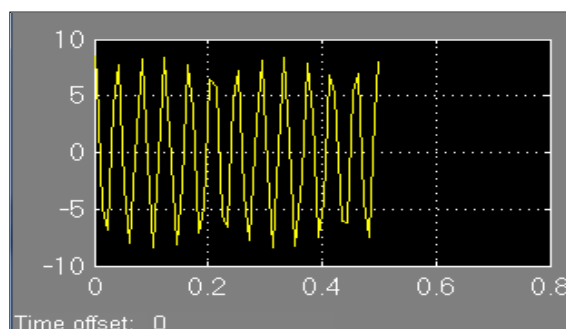


Fig. 8. Obtained result from Simulink® model of RC circuit

2.3. Series RLC circuit

The circuit consists of a resistor, inductor and capacitor, all connected in series. The initial conditions are assumed to be zero. Fig. 9 below gives a pictorial view of the typical circuit arrangement in consideration

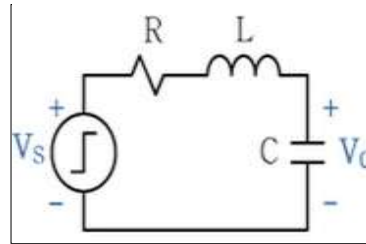


Fig. 9. Typical RLC series circuit

The first order ordinary differential equation that describes a simple RLC series⁴ electrical circuit with a resistor, inductor, capacitance and sinusoidal voltage source is as follows:

$$L \frac{d^2 q}{dt^2} + R \frac{dq}{dt} + \frac{q(t)}{C} = e(t),$$

where $i(t) = \frac{dq}{dt}$ and $e(t) = E_0 \cos(\omega t)$

For this example, the inductance is 1mH, the capacitance C is 1mF and the resistance R is 40 Ω . The voltage source supplies a step input of 1V.

The MATLAB statement written to solve (5) symbolically (i.e., obtaining exact solution) is as follows:

```
w=10000;
t=0:0.00001:0.005;
vc=1-(cos(w*t)+.2*sin(w*t)).*exp(-2000*t);
plot(t, vc, 'blue'), grid
```

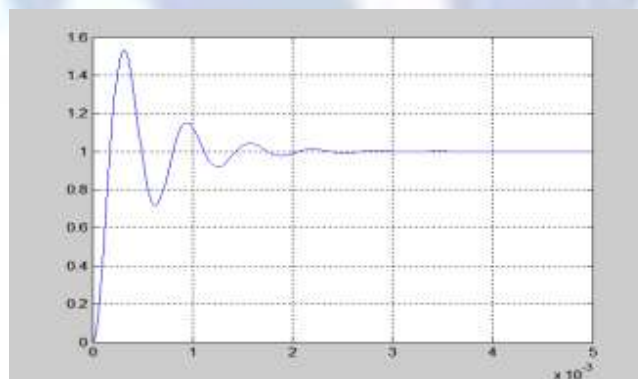


Fig. 10. Showing response of RLC series circuit

Next, the differential equation describing the RLC series electrical circuit is modeled using Simulink®.

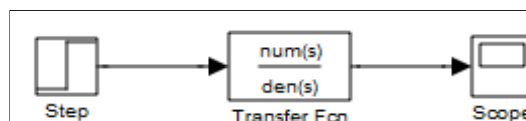


Fig. 11. Simulink® model of the RLC series circuit

$$\frac{V_C(s)}{V_S(s)} = \frac{1}{LCs^2 + RCs + 1}$$

Fig. 12. Transfer function of series RLC circuit

3. MECHANICAL ENGINEERING APPLICATION

Given the following rotating mechanical system with gears⁵:

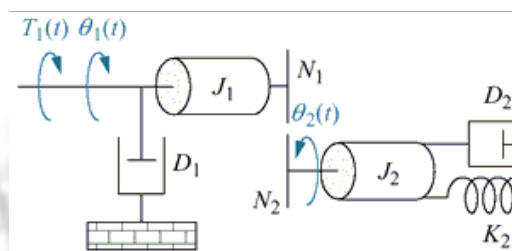


Fig. 13. Rotating mechanical system with gears

We will not explore the solution in its entirety. But we are giving a brief idea about how such a problem could be approached.

The Laplace transformed model of the above setup when modeled in Simulink®:

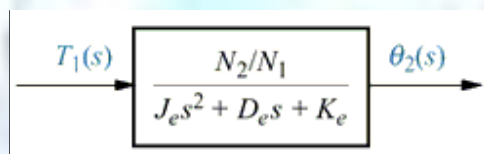


Fig. 14. Transfer function of mechanical system

$$D_e = D_1 \left(\frac{N_2}{N_1} \right)^2 + D_2 \quad (6)$$

$$J_e = J_1 \left(\frac{N_2}{N_1} \right)^2 + J_2 \quad (7)$$

$$K_e = K_2 \quad (8)$$

CONCLUSION

In this paper, the authors presented three circuits with two methods each, using MATLAB and Simulink® to solve Ordinary Differential Equations (ODEs) of RL & RC filters and RLC Tuning Circuits. Example problem sets for each of two engineering disciplines have been provided. With the RL, RC and RLC circuits, we were able to show that using Matlab to analyze these circuits saves time and provides accurate results with minimal prerequisite knowledge of programming. And by using Simulink®, we were able to analyze the circuit using its transfer function representation. Thus, identification of the transfer function of the system is the very first step in obtaining its solution. We solved the circuits, using the MATLAB symbolic solver (dsolve) for analytical solutions, using the block diagram programming language-Simulink®. With less effort, these techniques allow scientists to solve engineering problems with an excellent graphical picture of the result. The purpose of comparing simulation results with analytical one is to ensure that the numerical approximation gives an acceptable result using the analytical one as a benchmark.

REFERENCES

- [1]. The MathWorks, Inc., MathWorks Documentation-MATLAB Version 7.6.0.324 (R2008a). Symbolic Math Toolbox, 2008.
- [2]. L. F. Shampine, I. Gladwell, S. Thompson, 'Solving ODEs with MATLAB', USA, 2003.
- [3]. M.E. Van Valkenburg, 'Network Analysis', Prentice Hall of India, 1974.
- [4]. William H. Hyat, Jack E. Kemmerly, Steven M Durbin, 'Engineering Circuit Analysis', McGraw-Hill.
- [5]. Aliyu B. Kisabo, C. A Oshoku, Adetoro M.A Lanre, Aliyu Funmilayo A, 'Ordinary Differential Equations: MATLAB /Simulink® Solutions', International Journal of Scientific & Engineering Research, August-2012.



Cross Project Change Prediction Using Open Source Projects

Ruchika Malhotra

Software Engineering Department
Delhi Technological
Delhi, India

Ruchikamalhotra2004@yahoo.com

Ankita Jain Bansal

Software Engineering Department
Delhi Technological University
Delhi, India

Ankita.bansal06@gmail.com

Abstract— Predicting the changes in the next release of software, during the early phases of software development is gaining wide importance. Such a prediction helps in allocating the resources appropriately and thus, reduces costs associated with software maintenance. But predicting the changes using the historical data (data of past releases) of the software is not always possible due to unavailability of data. Thus, it would be highly advantageous if we can train the model using the data from other projects rather than the same project. In this paper, we have performed cross project predictions using 12 datasets obtained from three open source Apache projects, Abdera, POI and Rave. In the study, cross project predictions include both the inter-project (different projects) and inter-version (different versions of same projects) predictions. For cross project predictions, we investigated whether the characteristics of the datasets are valuable for selecting the training set for a known testing set. We concluded that cross project predictions give high accuracy and the distributional characteristics of the datasets are extremely useful for selecting the appropriate training set. Besides this, within cross project predictions, we also examined the accuracy of inter-version predictions.

Keywords—Cross Project, Inter-version prediction, Change prediction, Object oriented paradigm, Machine learning, Metrics

I. INTRODUCTION

Change prediction deals with detecting change prone modules and thus, helps in allocating the resources (time, money and manpower) judiciously. Identification of change prone modules will bring some insights on the design of the software, such as it can indicate about the extent of coupling, cohesion, inheritance etc. used in the system. Thus, based on these insights, if required we can alter the design. There are various object oriented metrics proposed in literature ([1], [2], [3], [4], [5], [6], [7], [8]) that help in the prediction of change prone classes. Several empirical studies are done ([9],[10],[11],[12],[13],[14],[15],[6],[17]) to develop various change proneness models. But these studies have predicted the model by training it using the historical data from the same project. The predicted model was used to further

identify change prone modules in the upcoming or future releases. However, prediction of change prone models using the training data from the same project is not always feasible as it might be possible that the adequate training data is not available or is not well collected ([18],[19],[20]). The possibility of non-availability of training data has led to the idea of performing cross project predictions. There have been few studies that have carried out cross project prediction for defect proneness ([18], [19], [20], [21], [22], [23]). Zimmermann, Nagappan and Gall [18] ran 622 cross-project defect predictions and found that only 21 predictions worked successfully (all Precision, Recall, and Accuracy are greater than 75%). The authors, Turhan, Menzies and Bener ([19], [23]) also conducted cross-company defect prediction and concluded that cross-company predictions increase the probability of defect detection at the cost of increasing false positive rate. Watanabe, Kaiya and Kaijiri [21] found that data characteristics of the datasets are important for cross-project defect prediction.

But to the best of our knowledge, the work of cross project predictions for change proneness has not been done yet. Hence, this has motivated us to conduct cross project predictions for change proneness. Cross project change prediction is defined as predicting the changes by training the models using the historical data of some other projects ([18], [21]). In other words, to predict changes in a project B, we train the models using the data of the other project A, and then test them on the project B. We can observe that training and testing are performed on the different projects, instead of using the same project. In this study, under cross project predictions, we have considered two categories: inter-versions predictions and inter-project predictions. Inter-version predictions refer to the predictions carried on two different versions of the same project. Whereas, inter-project predictions refer to the predictions carried on two different projects.

In this work, we tried to investigate the following two issues or research questions (RQ):

RQ1: How does the accuracy of the model differ when we train the model using the different versions of the same project

(i.e. performing inter-version validation) or when we use different project (i.e. performing inter-project)?

RQ2: Are characteristics of data sets valuable for selecting suitable training data for cross-project defect prediction?

Aiming at the above questions, we conducted an experiment on 12 public datasets from 3 projects. All the three projects are Apache projects (Apache POI, Apache Rave and Apache Abdera) and we considered 4 latest releases for each project, thus making a total of 12 projects. We considered all the possible combinations (A, B) to perform cross project predictions for change proneness. We employed a machine learning technique, logitboost to construct the prediction models (to measure the performance of each cross project predictions) with the help of the open source data mining tool WEKA. To determine the effectiveness of distributional characteristics in selecting the training data, we constructed a 'Classification and Regression Tree' (CRT) and generated various classification rules.

The different sections of the paper are: Following this section is the section 2 which explains the research background; the variables used in the study and various performance measures used. In next section, we explain our datasets along with the data collection process. Section 4 describes the basic four steps involved in this study. We discuss the results in section 5 and conclude the study in section 6.

II. RESEARCH BACKGROUND

In this section, we explain the independent and dependent variables used in this study. We also explain the various performance measures that are used to evaluate the performance of the models.

A. Variables Used

We have used the famous object oriented metrics proposed by Chidamber and Kemerer [5] as independent variables. Along with the metrics proposed by Chidamber and Kemerer, one additional metrics is used, SLOC. Each of these metrics is defined below:

Coupling between objects (CBO): Number of other classes to which a class is coupled and vice versa

Response for a class (RFC): Number of methods in the class including the methods that are called by class's methods

Lack of cohesion (LCOM): For each data field in a class, the percentage of the methods in the class using that data field; the percentages are averaged and then subtracted from 100%

Number of children (NOC): Number of immediate subclasses of a class in a hierarchy

Depth of inheritance (DIT): Maximum number of steps from the class node to the root

Weighted methods per class (WMC): Count of sum of complexities of all methods in a class

Source lines of code (SLOC): Number of lines of code

Changes in software are inevitable and therefore, predicting the classes that are change prone in the future release of the software is highly beneficial. We have used dependent variable as change proneness which is defined as the probability of prediction of change in the next version of the software. This prediction is based on the changes in the current or present version. We have compared the classes of the two versions and measured changes in terms of number of lines of code added, deleted or modified and modified in the class of first (previous) version with respect to the class of the same name in the current version.

B. Performance Measures

We have constructed a number of prediction models (for each possible train/test (A, B) combination) using a machine learning technique, logitboost. We evaluated each of these models using two performance measures explained in this section. Based on certain values of these performance measures, we have categorized each prediction as 'successful/valid' or 'not valid'.

There can be following categories of results:

- TNCP: number of classes correctly predicted to be 'not change prone'.
- FNCP: number of classes that are incorrectly predicted to be 'not change prone'
- FCP: number of classes that are incorrectly predicted to be 'change prone'
- TCP: number of classes that are correctly predicted to be 'change prone'

Precision: In terms of above categories, we define precision mathematically as $TCP / (TCP + FCP)$. The ideal value of precision is 1.0.

Area under the Receiver Operating Characteristics Curve (ROC): ROC curve is defined as a plot of sensitivity and 1-specificity on the y and x coordinates respectively ([24],[25]). We calculate sensitivity and specificity at different cut-off points between 0 and 1. We calculate area under the ROC curve which is a measure of overall performance of the predicted model. Higher the area under the curve, higher is the accuracy of correct prediction [26].

III. EMPIRICAL DATA COLLECTION

We have used 4 releases each of 3 open source Apache projects, Abdera, POI and Rave. Each release is considered as a different dataset, thus making a total of 12 datasets. Apache Abdera (<http://abdera.apache.org/>) is an implementation of the Atom Syndication Format and Atom Publishing Protocol which are used for creating, editing, and publishing various web

sources. Apache POI (<http://poi.apache.org/>) is used for making and modifying different file formats based upon OOXML and Microsoft's OLE 2. We can read/ write Word, Excel and PowerPoint files using Java. It is used in text extraction applications such web spiders etc. Apache Rave (<http://rave.apache.org/>) provides an extensible platform for using, integrating and hosting various OpenSocial and W3C Widget related features, technologies and services. Rave is popularly used as engine for internet and intranet portals.

The releases that we have considered are as follows (1) for Abdera: Abdera 1.0, Abdera 1.1, Abdera 1.1.1 and Abdera 1.1.2, (2) for POI: POI 3.0, POI 3.6, POI 3.7 and POI 3.9, (3) for Rave: Rave 0.19, Rave 0.20.1, Rave 0.21.1 and Rave.22. The details of each of these releases which include the release year, total number of classes and the number of classes changed in each release are shown in table I. We have calculated change between every two successive versions. For example, change is calculated for the common classes between the releases Abdera 1.0 and Abdera 1.1. We briefly describe the change collection process as follows:

For collection of the data, we used a tool named Defect Collection and Reporting System (DCRS), developed in Java by the students of Delhi Technological University, Delhi, India. The tool is used to generate the change reports for open source software hosted at *Git repository*. The tool functions as follows: Two consecutive versions of software are taken and their source code is analysed. Using Git commands, we process the source code and retrieve Change-Logs. A Change Log contains all the information concerned with the modifications that have been incurred from time to time in the source code. The modification in the source code could be due to various reasons such as defect-fixing, function/feature enhancements etc. Each modification is stored as an individual change record in the Change log. An individual change record at the Git repository consists of: a) Change timestamp, b) Unique change identifier, c) Description of the change including the reason why the change has occurred, d) Source code files that are modified, along with the LOC changes for each modified file.

Finally, these Change records are mapped to source code Classes. Only java source code files are considered and other files such as xml and other resources are ignored. The change- report generated consists of the following fields:

- Name of the Java source file (Class name).
- A binary variable to indicate whether a class has been modified, i.e. 'yes' else 'no'.
- Corresponding data for Metrics for each class.

TABLE I. DETAILS OF THE PROJECTS

Project	Written in / License	Releases	Release Date	No. of classes	No. of Classes Changed
Abdera	Java / Apache 2.0	Abdera 1.0	May 2010	679	624
		Abdera 1.1	July 2010	677	8
		Abdera 1.1.1	Dec. 2010	686	634
		Abdera 1.1.2	Jan. 2011	685	635
POI	Java / Apache 2.0	POI 3.0	July 2007	1515	1276
		POI 3.6	Dec. 2009	2088	1988
		POI 3.7	Oct. 2010	2472	2212
		POI 3.9	Dec. 2012	2786	2706
Rave	Java / Apache 2.0	Rave 0.19	Dec. 2012	607	32
		Rave 0.20.1	Mar. 2013	642	628
		Rave 0.21.1	Apr. 2013	676	648
		Rave 0.22	May 2013	685	225

IV. RESEARCH METHODOLOGY

Here, we explain the basic four steps involved in this study. The first step is the creation of train/test combination or generation of datapoints. This is followed by measuring distributional characteristics for each train/test combination. The third step is the construction or prediction of machine learning model to determine the accuracy of cross project predictions. Finally, we construct a Classification and Regression tree (CRT) to judge the suitability of characteristics of data sets for selecting the training data for cross-project defect prediction. All the steps are explained diagrammatically in figure 1.

A. Creation of Train/Test Combinations

We have used 3 Apache projects (Apache POI, Apache rave, Apache Abdera), with four releases of each. A project with 4 releases is considered as 4 separate datasets and therefore, 3 projects will comprise of 12 datasets. To perform cross project validation, one dataset among the 12 datasets is taken as training set and some other dataset is taken as the testing set. We have considered the possible combinations of training and testing sets (A, B) if: 1) A and B belong to different projects or 2) A and B belong to same project, then B should be the later version than A [18]. In other words, we want the testing set to be the later version than the training version as it is illogical to perform testing on the former version and training on the later version. For example, we used Abdera 1.1.1 to predict the later version, Abdera 1.1.2 and not the former version, Abdera 1.1.

Therefore, the number of possible combinations with A,B,C.....N projects, having number of releases as a,b,c.....n respectively:

$$P = \sum_{i=A}^T \frac{v_i * (v_i - 1)}{2}$$

Where, T= number of datasets = a+b+c....+n

V_i = number of releases of i_{th} project. For example, number of releases for A project is a.

Thus, based on the above formula, the total possible combinations in this study are 114.

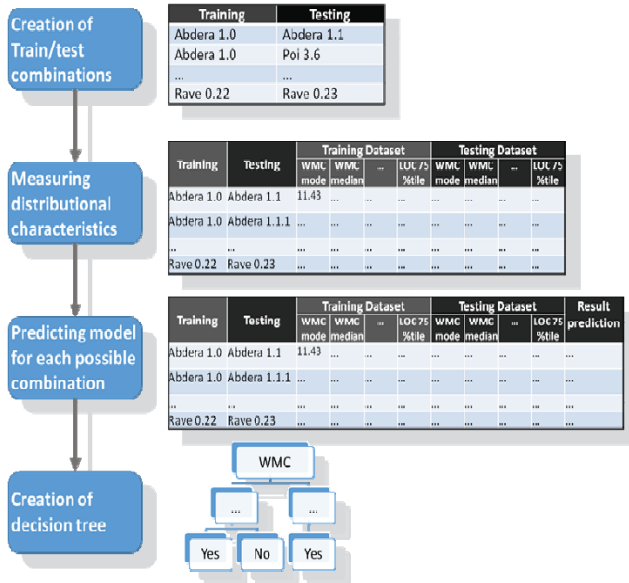


Fig. 1. Basic Research Methodology

B. Measuring Distributional Characteristics

We have used three open source projects with 4 releases of each. For these 12 different datasets, we calculate various descriptive statistics which measure the distributional characteristics of each independent variable (number of independent variables are 7). We measure 11 distributional characteristics (indicators) for each variable. As discussed, we have 114 possible train-test combinations. We create 144 attributes for each combination (or datapoint) as follows:

1. We combine the distributional characteristics of each variable of training set to get 77 (7 independent variables * 11 indicators) attributes
2. We combine the distributional characteristics of each variable of testing set to get 77 (7 independent variables * 11 indicators) attributes
3. 77 attributes of training set and 77 attributes of testing set are further combined to yield a total of 144 attributes.

The various indicators that we used to measure the distributional characteristics are Mode, Median, Mean (to describe the central tendency of attributes), standard deviation, variance, skewness, kurtosis, minimum, maximum, sum, first quartile and third quartile.

C. Predicting Model for each Possible Prediction

We have used a machine learning algorithm, logitboost to build the model for each valid training-testing pair (A, B). We predicted the model using the dataset A and tested it on the dataset B. Thus, in this study, we have 114 model predictions. For each prediction, we say whether the prediction is valid or not valid based on the values of performance measures, precision and area under the ROC curve. If the value of area under the ROC curve is greater than 0.65 and precision is greater than 50%, we have considered the prediction as valid or successful. This we call as the 'result' of each prediction, which is discrete or binary in nature. In this study, for each prediction, we have defined a datapoint having three parts: 1) consists of distributional characteristics of training set, 2) consists of distributional characteristics of testing set, 3) the result of each prediction (i.e. valid or not valid). Thus, we have 114 datapoints, each having 145 attributes (77 number of attributes comprising of distributional characteristics of training set, 77 number of attributes comprising of distributional characteristics of testing set and 1 attribute comprising of the result of the prediction).

D. Verify the relationship between Distributional Characteristics of data sets and Selection of Training Set

Our one of the objectives is to determine whether the distributional characteristics of training and testing set have an impact on the cross project predictions. The studies by Zimmermann, Nagappan & Gall [18] and Watanabe, Kaiya & Kaijiri [21] show that the distributional characteristics are essential for identifying successful cross-project defect prediction. In this work, we construct a decision tree on the dataset generated above (consisting of 114 datapoints) to check the accuracy of the decision tree in finding the valid cross-project predictions. The independent variables are the distributional characteristics of training and testing set (77+77=144 attributes) and the binary dependent variable is the result of each prediction (valid/not valid). The validation technique used is 10-cross validation. If the decision tree predicts the valid predictions with high accuracy, then we conclude that the cross project predictions have an impact on the distributional characteristics of the training and testing sets.

V. RESULT ANALYSIS

In this section, we address the two research questions stated in section 1 and intend to answer them

5.1 How does the accuracy of the model differs when we train the model using the different versions of the same project (i.e. performing inter-version validation) or when we use different project (i.e. performing inter-project validation)?

For each of the possible train/test combination (A, B), we predicted the machine learning model (logitboost) using dataset A and tested it on dataset B. We measured the performance in

terms of precision and area under the ROC curve. We are interested in knowing the number of successful cross project predictions. We refer a prediction as successful if the precision is more than 50% and area under the ROC curve is more than 0.65. We observed that out of a total of 114 predictions, 41 satisfy our criteria and thus are referred as successful predictions (i.e. the success rate is 36%). This is quite high when we compare it with the success rate obtained by [18]. Zimmermann, Nagappan & Gall [18] ran 622 cross-project predictions and found only 3.4% actually worked. Within the successful predictions, we want to compare the number of successful inter-project and inter-version predictions. Inter-version predictions are the predictions where the training and testing sets belong to the same project, but different versions. Whereas, in inter-project predictions, training and testing sets belong to different projects. It would seem that the different releases of same project will predict each other well, as they are developed under same environment, have similar characteristics etc. However, there is a difference of opinion in this issue by previous studies. Some of the previous studies ([28],[29],[30],[31]) have shown that different releases from the same project have predicted each other well. However, the latest study by He, Shu, Yang Li and Wang [32] had shown unexpected result, i.e. decreased prediction accuracy for the same (ranging from 0.32% to 4.67%).

In table II, we analyze the number of successful inter-project and inter-version predictions. If two datasets A and B belong to the same project, then the total possible train/test combinations (A,B : B should be later version than A), are 18 (6 combinations for each Abdera, POI and Rave). In other words, we say 18 is the total number of inter-version combinations. Out of 18 combinations, we observe that the number of successful predictions is 12, i.e. approximately 67% success rate. On the other hand, if two datasets A and B belong to the different project, then the total number of train/test combinations (A,B) are 96 (114-18). Thus, the number of inter-project combinations is 96. Out of 96, the number of successful predictions is 29, i.e. approximately 30% success rate. This is significantly lower than the success rate of inter-version predictions. Thus, in our study, we conclude that the inter-project predictions give higher accuracy than the inter-project predictions.

Table III shows the range of area under the ROC curve (AUC) for successful predictions. For Abdera as testing project, we see that using Abdera as training project (different version of same project), the range of AUC is 0.67 to 0.78. We can observe that the range of AUC when training set is Poi and Rave is also approximately same as when the training set is Abdera. This shows that the accuracy of the inter project and inter version predictions is approximately the same. Same observation is concluded when the testing project is Poi and Rave.

Table II. NO. OF SUCCESSFUL PREDICTIONS

TRAIN/TEST	Abdera	POI	Rave	Total
Abdera	5	5	5	15
POI	10	4	0	14
Rave	9	0	3	12
Total	24	9	8	41

Table III. AUC OF SUCCESSFUL PREDICTIONS

Testing Project	Training Project	Min. AUC	Max. AUC
Abdera	Abdera	0.67	0.78
	Poi	0.67	0.71
	Rave	0.67	0.71
Poi	Abdera	0.66	0.72
	Poi	0.67	0.69
	Rave	-	-
Rave	Abdera	0.65	0.72
	Poi	-	-
	Rave	0.66	0.68

5.2 Are characteristics of data sets valuable for selecting suitable training data for cross-project defect prediction?

We generated a total of 114 possible predictions and the number of successful predictions is 41. To determine whether the characteristics of the datasets help in selecting the suitable training set, we applied decision tree (CRT) on the dataset of 114 datapoints having 145 attributes (144 attributes consisting of distributional characteristics of training and testing sets + 1 attribute for the result of prediction). CRT trained from the train/test instances consisted of 15 nodes, out of which 8 are leaf nodes. The leaf nodes will contain the value of the result of prediction (dependent variable). We observed 3 nodes resulted in the value 'yes', from which we generated the classification rules. These classification rules can be used to select a suitable training set for a given test set. We also performed 10-cross validation on this dataset and obtained precision as 85.6 % and the area under the ROC curve as 0.98. Taking into consideration success rate for inter-project predictions (67%) and the success rate for inter-version predictions (30%), we can say that the performance of CRT is very high. Thus, we can conclude that the distributional characteristics of the datasets play an important role in selecting the suitable training set for a given test set.

VI. CONCLUSION

Many a times, in various commercial organizations, the availability of historical data becomes questionable as it is unavailable or is insufficient. Generally, the models for change prediction are learned using the historical data of some project. Then, using this trained/learned model, we predict the changes

in the upcoming releases of that project. But non-availability of historical data poses a difficult scenario. In such cases, we may use the data from some other project rather than the same project. This is referred to as cross project prediction, i.e. training and testing the model on different projects. Careful selection of training data from other projects is very important. Thus, cross project prediction is gaining wide importance for projects with insufficient or no data to build prediction models. Despite its importance, to the best of our knowledge, no work has been done in the field of cross project change prediction. In this paper, we have used 4 releases each of 3 open source projects, Apache Abdera, POI and Rave. We have total 12 datasets (4*3) and we made all the possible combinations (A, B) of training/testing set. We calculated the total number of possible combinations is 114 and thus, we ran 114 cross project predictions using a machine learning technique, logitboost. Following are the main conclusions that can be drawn from the study:

1. We have considered the successful predictions as those cross project predictions where precision is more than 0.5 and area under the ROC curve is more than 0.65. Using this criterion, we found a success rate of 36%.
2. We conducted both cross project (different projects) and inter version (different versions of same projects) predictions. Among the successful number of predictions, we found that the success rate of inter-version (different versions of the same project) is 67% and the success rate of cross project predictions (different projects) is 30%. From this, we concluded that inter-version predictions are more accurate than the inter-project predictions.
3. We constructed a CRT decision tree to judge the suitability of characteristics of the datasets in selecting the training set for a given test set. We found high values of precision, recall and area under the ROC curve. This shows that selection of training set depends on the distributional characteristics of the datasets. For a given test set, we can select an appropriate training set using the classification rules (that produced a 'yes') of CRT.

In our future work, we would like to experiment our methodology on more number of projects. This would help us to generalize the results. Besides this, we would try to find out if cross-project change prediction is symmetrical. In other words, we will see if project A predicts project B well, does it imply that vica-versa is also true (i.e., if project B also predicts project A well). Finally, we would propose a well defined training data selection technique. For a given testing set, we would propose a technique to find its suitable training set.

REFERENCES

- [1] L. Briand, W. Daly and J. Wust, "Unified framework for cohesion measurement in object-oriented systems," *Empirical Software Engineering*, vol. 3, pp. 65–117, 1998
- [2] L. Briand, W. Daly and J. Wust, "A unified framework for coupling measurement in object-oriented systems," *IEEE Transactions on Software Engineering*, vol. 25, pp. 91–121, 1999.
- [3] J. Bieman and B. Kang, "Cohesion and reuse in an object-oriented system", in *Proceedings of the ACM Symposium on Software Reusability (SSR'94)*, pp. 259–262.
- [4] M. Cartwright and M. Shepperd, "An empirical investigation of an object-oriented software system," *IEEE Transactions on Software Engineering*, vol. 26, pp. 786–796, 1999.
- [5] S. Chidamber and C. Kemerer, "A Metrics Suite for Object-Oriented Design," *IEEE Trans. Software Eng.*, vol. 20, pp. 476–493, 1994.
- [6] W. Li and W. Henry, "Object-oriented Metrics that Predict Maintainability," *Journal of Software and Systems*, vol. 23, pp. 111–122, 1993.
- [7] M. Lorenz and J. Kidd, *Object-Oriented Software Metrics*, Prentice-Hall, 1994.
- [8] J. Bansiya, and C. Davis, "A Hierarchical Model for Object-Oriented Design Quality Assessment," *IEEE Trans. Software Eng.*, vol. 28, pp. 4–17, 2002.
- [9] M. Askari and R. Holt, "Information Theoretic Evaluation of Change Prediction Models for Large-Scale Software," in *ACM: Proceedings of international workshop on Mining software repositories*, pp.126–132, 2006.
- [10] A.R. Han, S.U. Jeon, D.H. Bae and J.E. Hong, "Behavioural Dependency Measurement for Change-proneness Prediction in UML 2.0 Design Models," *Annual IEEE International Computer Software and Applications Conference*, 2008.
- [11] A.G. Koru and J. Tian, "Comparing High-Change Modules and Modules with the Highest Measurement Values in Two Large-Scale Open-Source Products. *IEEE Transactions on Software Engineering*," vol. 31, pp. 625–642, 2005.
- [12] A. Guñes, Koru and H. Liu, "Identifying and characterizing change-prone classes in two large-scale open-source products," *The Journal of Systems and Software*, vol. 80, pp. 63–73, 2007.

- [13] D. Romano and M. Pinzger, "Using Source Code Metrics to Predict Change-Prone Java Interfaces," Tech. Rep. ISSN 1872-5392.
- [14] M. Lindvall, "Are large C++ classes change-prone? An empirical investigation," *Software Pract Ex.*, vol. 28, pp. 1551–1558, 1998.
- [15] E. Arisholm, L.C. Briand and A. Føyen, "Dynamic Coupling Measurement for Object-Oriented Software," *IEEE Transactions on Software Engineering*, vol. 30, pp. 491–506, 2004.
- [16] Y. Zhou, H. Leung and B. Xu, "Examining the Potentially Confounding Effect of Class Size on the Associations between Object-Oriented Metrics and Change-Prone," *IEEE Transactions on Software Engineering*, vol. 35, pp. 607–623, 2009.
- [17] H. Lu, Y. Zhou, B. Xu, H. Leung and L. Chen, "The ability of object-oriented metrics to predict change-proneness: a meta-analysis," *Empir Software Eng.*, vol. 17, pp. 200–242, 2011.
- [18] T. Zimmermann, N. Nagappan and H. Gall, "Cross-project defect prediction: a large scale experiment on data vs. domain vs. process," In: *Proceedings of the 7th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering*, pp. 91–100, 2009.
- [19] B. Turhan, T. Menzies and A. Bener, "On the relative value of cross-company and within company data for defect prediction," *Empir. Softw. Eng.*, vol. 14, pp. 540–578, 2009.
- [20] M. Jureczko and D. Spinellis, "Using object-oriented design metrics to predict software defects," In: *Proceedings of the 5th International Conference on Dependability of Computer Systems*, pp. 69–81, 2010.
- [21] S. Watanabe, H. Kaiya and K. Kaijiri, "Adapting a fault prediction model to allow inter language reuse," In: *Proceedings of the International Workshop on Predictive Models in Software Engineering*, pp. 19–24, 2008.
- [22] N. Nagappan, T. Ball, and A. Zeller, "Mining metrics to predict component failures," In: *Proceedings of the 28th International Conference on Software Engineering*, pp. 452–461, 2006.
- [23] B. Turhan, A. Bener and T. Menzies, "Regularities in learning defect predictors," In: *The 11th International Conference on Product Focused Software Development and Process Improvement*, pp. 116–130, 2010.
- [24] K. El Emam, S. Benlarbi, N. Goel and S. Rai, "A validation of object-oriented metrics," NRC Tech. rep. ERB-1063, 1999.
- [25] Y. Singh, A. Kaur and R. Malhotra, "Empirical validation of object-oriented metrics for predicting fault proneness," *Softw Qual J*, vol. 18, pp. 3–35, 2010.
- [26] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.
- [27] R. Malhotra and M. Khanna, "Investigation of relationship between object-oriented metrics and change proneness," *Int. J. Mach. Learn. & Cyber.*, vol. 4, pp. 273–286, 2013.
- [28] A. Tosun, B. Turhan and A. Bener, "Practical considerations in deploying AI for defect prediction: a case study within the Turkish telecommunication industry," In: *Proceedings of the 5th International Conference on Predictor Models in Software Engineering*, pp. 1–9, 2009.
- [29] T.J. Ostrand, E.J. Weyuker and R.M. Bell, "Predicting the location and number of faults in large software systems," *IEEE Trans. Softw. Eng.*, vol. 31, pp. 340–355, 2005.
- [30] E.J. Weyuker, T.J. Ostrand and R.M. Bell, "Comparing the effectiveness of several modeling methods for fault prediction," *Empir. Softw. Eng.*, vol. 15, pp. 277–295, 2009.
- [31] E.J. Weyuker, T.J. Ostrand and R.M. Bell, "Do too many cooks spoil the broth? Using the number of developers to enhance defect prediction models," *Empir. Softw. Eng.*, vol. 13, pp. 539–559, 2008.
- [32] Z. He, F. Shu, Y. Yang, M. Li and Q. Wang, "An investigation on the feasibility of cross-project defect prediction," *Autom Softw Eng*, vol. 19, pp. 167–199, 2012.



Customizing Agile Methods using Genetic Algorithms

Rinky Dwivedi¹ and Daya Gupta²

¹ Maharaja Surajmal Institute of Technology, New Delhi, INDIA
rinkydwivedi@gmail.com

² Delhi Technological University, New Delhi, INDIA
daya_gupta2005@yahoo.co.in

Abstract— Method customization for satisfying organisation configuration definitions has become an important activity for many modern organisations. Method engineers have become more and more concerned about accurately customizing the method for the situation-in-hand. In the last few decades many method customization models have been developed but no model has proved to be successful at effectively deliver the customized light-weight methods fulfilling the organisation requirements. In this paper we propose the use of genetic algorithms for forming situational method using method configuration. This technique randomly selects the configured methods, evaluate their fitness and then if required second generation of population is produced, thereby converging the population towards the requirement. The agile methods used in our study are XP, scrum, DSDM and FDD. Experimental results of XP are given to show the effectiveness of the proposed approach.

Index Terms— Method Configuration, Genetic Algorithms, Agile Methods, Situational methods.

I. INTRODUCTION

Genetic Algorithm is an evolutionary based search optimization technique. Every search operation is supported by logic, the logic is based on the “Survival of the Fittest” concept which is basically Darwin’s theory-the best will survive. Genetic algorithms have proven to be useful for the domains such as software testing, finding shortest path, natural language processing and computer vision. In these applications there is a large solution space containing implicit regularities to be discovered and the solution must be adapting to changing conditions. The basic approach to genetic algorithm is to generate a set of solutions and devise mechanism based on the feedback, to evaluate the goodness of each solution and select the one which is close to optimal. Far-from-optimal solutions are discarded and close to optimal candidates are allowed to mutate with each other and generate next generation of solutions.

Optimal solutions to several engineering problems have been derived through Genetic algorithm [2]. In our study we have come across many problems in developing situational method like gathering project characteristics, retrieving the candidate method and assemble the method chunks to form the situated method. These problems need to be explored through genetic algorithm so that an optimum route can be found to reach situational method. Every progressive iteration will direct the process to global optimum situational method.

Presently the customization or engineering of light-weight methods is an open area in software engineering domain, it specifies that there is a need for changing and adapting methods, not only necessary for organizations that employ formal/mature software processes, but also for those that use agile methods[1]. In the present research, we are interested in investigating the use of genetic algorithms to configure an optimized agile method for organisation need.

The paper is organized as: Section 2 briefly presents the previous research in the area. Section 3 foresees the general idea behind the method optimization and genetic algorithms. Section 4 describes the method configuration process for forming situational methods. Section 5 describes the actual process, needs to follow to satisfy organisational requirements, details of encoding and initial population generation and evaluation of initial population corresponding to the organisational configuration definition. The paper closes with the production of new solutions, by the process of crossover and their fitness results.

II. STATE OF ART

It has now been commonly accepted that there is no single method that fits all situations. Projects differ in various dimensions, for example, with respect to development context, situation, complexity and granularity. Different proposals for how to create a situational method that fits the unique project have been put forth, ranging from formalized meta-methods [11] to architectural based [9] and further extended to more formal guidelines [10].

These types of method engineering approaches are anchored with several assumptions major ones are, it is possible for project members to explicitly specify the required situational method upfront and successfully communicate these requirements to the method engineer and these requirements do not change over the lifetime of a project. However, in practice requirements are often evolutionary and vague; a commonly cited method for these system requirements is the user-centred agile approach that helps in bridging between end users and systems developers. Present research reveals that how agile methods can be configured to form situated methods. Studies reveal that agile methods are not anti-methods and can be tailored as any other traditional methods [14].

When talking about ‘configuration’, we refer to a particular variant of engineering methods, tailored in compliance to a specific development situation—a situational method derived from one specific base method [12, 13]. In other words, Method configuration means to adapt a particular method to various situated factors.

III. OPTIMIZATION USING GENETIC ALGORITHMS

This section presents a brief discussion on optimization of methods and how genetic algorithms are used as a potential solution to find an optimal solution for the project-in-hand. Optimization is an iterative process where initial solution is processed to improve the performance till an optimal solution is attained. However, Goldberg alleges that this is a highly “calculus” oriented view of optimization and has low natural emphasis.

In general, goodness is judged by relative comparison for example the best scientist, cricket, poet, tennis player and so on cannot be defined they can only be judged relatively. In the engineering domain the best test suite formed is the one that is capable of covering all constraints and boundary values for a system better than other test suites. The basic process of optimization is to evaluate each solution based on the characteristics of the problem. In the present research we use genetic algorithms for finding an optimized situation method.

Before proceeding, a basic explanation of genetic algorithm is required and how they are used to effectively solve the engineering problems. Genetic algorithm maintains a pool of solutions rather than just one. The principle behind the process is shown below the process is inherited from the evolution of species.

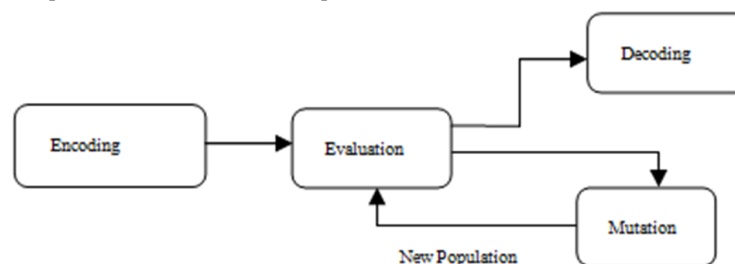


Figure 1: Genetic algorithm process

As shown in figure1 the basic genetic algorithm consists of the four steps, however complexity of each step depends upon the problem in hand:

- *Encoding and population generation*- For every problem need to be solved through GA, a suitable encoding is found for the solution to the problem so that each possible solution has a unique code in the form of a string. The initial population is then selected either randomly or through heuristics approach.
- *Evaluation*- The fitness of each individual in the population is then computed; that is, how well the individual fits the problem and whether it is near the optimum as compared to the other individuals in the population.
- *Mutation*- The fitness evaluated in the previous step is used to find the individual's probability of crossover. If an individual has a high probability (which indicates that it is significantly closer to the optimum than the rest of its generation) then it is more likely to be chosen to crossover. Once, a new generation has been formed and the process is repeated until population converges towards the required method.
- *Decoding*- At the end solution closest to the optimum is decoded and the process is complete.

IV. METHOD CONFIGURATION

Since Method services are not one size fit all. Method customization comes into play; and method Configuration is like a rethink about method customization. Method configuration supports to build a requirement specific method while overcoming the classic problem of customization.

Configuration approach makes possible for an organization to configure and extend agile methods to meet their specific needs; it allows the organization to make changes in the methods without losing the purpose to stay on the method.

So the approach is different instead of providing open ended toolkit, a framework is provided to extend the Organization configuration definitions. The framework supports the principle of essentiality for each method. Which states The Essentiality attribute can take two values either common or variable. Common are ones which if omitted; the method will lose its identity. So the variability in the proposal lies in the variable part of the method [5, 6].

A. Essentiality in Agile Methods

17 experts from various software development fields met to discuss the improved ways to build up the quality software. After an in-depth analysis and discussion agile manifesto comes into existence in year 2001. The manifesto contains core values and principles of agile methodology. For an in-depth history of agile kindly refer [4, 3].

The four core principles or statements defined in the manifesto are:

- Individuals and interactions over processes and tools
- Working software over comprehensive documentation
- Customer collaboration
- Responding to change

These principles states that the project managers have to let the development team and user cooperate together or collaborate together. Projects needs to concentrate on delivery and not on documentation however, comprehensive documentation is valuable, don't abandon the documentation a balance need to be created between delivery and documentation. Customer collaboration requirements cannot be fully collected at the beginning of the software cycle therefore, continual customer or stakeholder involvement is very important, responding to change.

Table 1 shows the most popular agile methods and their practices corresponding to the core principles of the agility. To preserve these principles, the practices corresponding to them are treated as essential to the method; however they can be replaced by other practices falling under same domain as well. The goal is to preserve the agility of the method.

V. GENERATING CONFIGURED SITUATIONAL SPECIFIC AGILE METHOD

A method is a set of practices; these practices are the distinguished features of a method. Sometimes, similar or same type of practices can be inherited by more than one method. But the entire set of practices for two different methods cannot be same.

TABLE I: - AGILE PRACTICES CORRESPONDING TO THE CORE AGILE PRINCIPLES FOR THE POPULAR AGILE METHODS

Practices	XP	Scrum	FDD	ASD	DSDM
Individual and interaction Over process.	The planning game	Daily scrum meetings	Domain object modelling	Adaptive management model	Active user involvement
Working software over Comprehensive Documentation.	Short releases	Sprints	Developing by feature	Developing by components	Frequent product delivery
Customer collaboration Over contract negotiation.	On-site customers	Product Backlog	Domain object modelling	Joint application development	Collaboration and cooperation
Responding to change over Following a plan	Metaphor	Sprint Review	Configuration management	Customer focus group reviews	Reversible changes

Method: {set of practices}

Encoding process includes following steps

- Scan the entire set of practices and remove the redundant ones.
- Presence of a practice in a method is represented by HIGH bit and a LOW represents the absence of a practice.
- No. of bits use to represent a method depends on the number of entries in the set of practices
For example, $M_1 = \{1001010010\}$ and $M_2 = \{1001101101\}$

Presently number of agile software development methods are used in practice, some of them are; Extreme programming [6, 7], scrum [8, 9], dynamic system development method [10] and feature driven development [11]. A collection of complete set of all agile practices is beyond the scope of this paper, for our research work we included the major practices of the most popular agile methods; the set is given in table 1.

A. Encoding and Initial population generation

Any genetic algorithm begins by the generation of a suitable initial population. The goal in this process is to create a set of solutions which have sufficient diversity. The process is of two types Unconstraint representation and constraint representation. Unconstraint is an indirect representation where the population generated is completely random. However, if this is not the case, the generation of initial population becomes more subtle. In the present research, the population is generated for Information System Domain methods. These methods have their own constraints and dependencies making the task of population generation to be more complex. For the current problem the population generated is a constraint representation and by exploring various paths we found method configuration as the potential solution to generate the initial set of solutions for information system domain methods.

Initial population of XP

Based on the above explanation, the initial population of XP [7] are given in table 2. The XPCM1 is the first configured method of XP; it has only the essential part of the method all the variables are external to it. XPCM2 is the second configured method contains essential part as well as a variable from the set of practices. Similarly, other configured methods are formed.

TABLE II: CONFIGURED METHODS OF XP

Practices	CM1	CM2	CM3	CM4	CM5	CM6	CM7	CM8	CM9
Planning game	1	1	1	1	1	1	1	1	1
On-site customer	1	1	1	1	1	1	1	1	1
Short release	1	1	1	1	1	1	1	1	1
Metaphor	1	1	1	1	1	1	1	1	1
Collective ownership	0	1	0	0	0	0	0	0	0
Pair Programming	0	0	1	0	0	0	0	0	0
Continuous Integration	0	0	0	1	0	0	0	0	0
Coding Standard	0	0	0	0	1	0	0	0	0
40- Hours Week	0	0	0	0	0	1	0	0	0
Refactoring	0	0	0	0	0	0	1	0	0
Simple Design	0	0	0	0	0	0	0	1	0
Testing	0	0	0	0	0	0	0	0	1

The initial population for other agile methods can also be generated, to provide a flavour, the set of initial solutions for Scrum has also been provided in table 3. The complete set of initial population for all the agile method can be taken from the authors.

TABLE III:-INITIAL POPULATION OF SCRUM

Practices	SCRCM1	SCRCM2	SCRCM3	SCRCM4
Daily scrum meetings	1	1	1	1
Product backlog	1	1	1	1
Sprint	1	1	1	1
Sprint review	1	1	1	1
Sprint planning	0	1	0	0
Scrum teams	0	0	1	0
Daily scrum meetings	0	0	0	1

B. Evaluation of initial population

The initial population is evaluated with respect to the current Organizational requirements. Every organization is different in terms of requirements they have, processes and facilities they have. Here process extension comes into play. The numerical value of the objective $f(n)=Y-I$ corresponds to the concept of fitness in genetics, where Y is the organizational requirement definition and I is the initial solution. After trial solutions are selected, a new generation is produced by selection, using stochastic principles; the “fittest” parents are selected to produce “children” among the trial solutions.

$Y = \{\text{set of organisation's requirement}\}$.

For the illustration purpose, Y is taken as $Y = \{111101010100\}$. To evaluate, a population of $2n$ to $4n$ trial design vectors are selected rather than just one and is used initially in the problem with n design parameters (GA works with the set of solution). Since, $n=1$ the four arbitrarily chosen design vectors are used as initial set of solution.

The calculated results after scaling and rearranging the design vectors are shown in table 4

TABLE IV: CALCULATING FITNESS NUMBER

String Number	Design vectors	Initial Population(I)	Fitness Number $F(n)=Y-I$
1.	111101000000	3.904	0.02
2.	111100010000	3.856	0.068
3.	111100000100	3.844	0.08
4.	111100000000	3.840	0.084

Plotting the above findings in figure 2

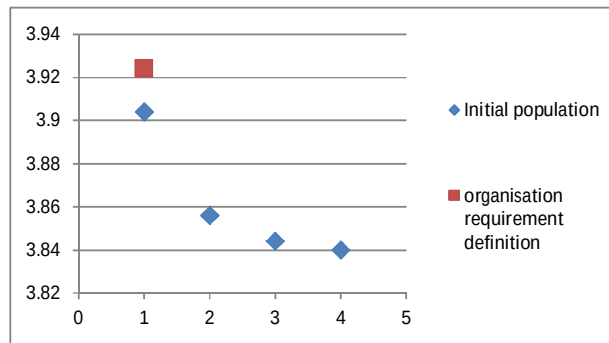


Figure 2: Graph shows the desired method (Y) and the randomly selected initial population of XP

This shows the fittest candidate is ‘1’ followed by ‘2’ followed by ‘3’ and so on. Now the mating pool has to be selected. The candidate ‘4’ has less relative fitness so is not worth to include it in the mating pool.

Good string(s) (“fittest”) in a population are selected and assigned a large number of copies to form an effective mating pool. Candidates selected for the mating pool along with their actual counts are shown in table 5.

TABLE V:- THE MATING POOL SELECTED FOR THE MUTATION

Initial Population	Relative fitness	Actual Count	Mating Pool
111101000000	0.02	2	111100010000
111100010000	0.068	1	111100000100
111100000100	0.08	1	111101000000

C. Mutation

New strings are created by exchanging information among strings of the mating pool. After which bit crossover is done. In mutation or exchange the portion of strings of each of the two parents generate new solutions. Calculate “Y” for new values of $x_1, x_2, \dots, x_n$ and put them in ascending order to find out maximum Y is shown in table 6.

TABLE VI:- NEW POPULATION

Initial population		New population (I_{new})	Fitness of new population (F_{new})
Mate 1	111101000000	111101010000	3.920
Mate 2	111100010000		
Mate 1	111101000000	111101000100	3.908
Mate 2	111100000100		
Mate 1	111100010000	111100010100	3.924
Mate 2	111100000100		

Plotting these new values of population, the results are shown in figure 3

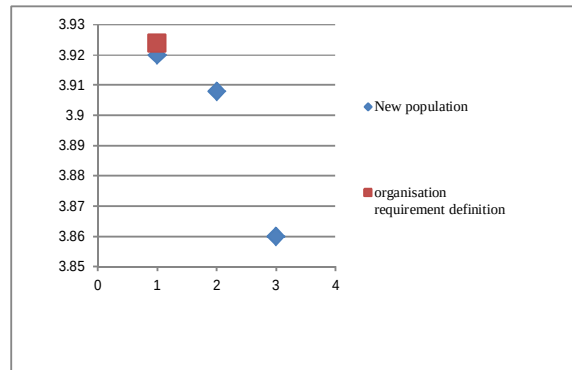


Figure 3: New generated population

Graph shows that after the crossover the second level population formed is converging towards the organisation configuration definition. Hence, an optimized method as per the requirement can be formed.

VI. CONCLUSION

Forming organisational specific agile method is a challenging task for both the industrial and academic communities. The accurate customization of the method for an organisation can greatly benefit the development team. In this paper agile methods are configured to form situated methods. The user-centred approach and prototyping design forms the basis of the entire process hence bridging the gap between systems developers and method engineers understanding of and expectations on a situational method.

Presently project requirement elicitation process depends on Method Requirement Specification languages, these languages are formed on a method engineer perspective rather than a method user one. However, the use of traditional elicitation methods like interviews tends only to capture the initial method requirements and fails to address the dynamic project requirements.

Our future work is to provide the set of project characteristics for the agile methods. These project characteristics must be capable enough to efficiently capture the negotiated users requirements based on several perspectives. The purpose is to design a simple and efficient elicitation process for handling dynamic project requirements.

REFERENCES

- [1] B. Fitzgerald, N. Russo and O' Kane T, "Software development method tailoring at Motorola", Communication of ACM, vol.46, pp.64-70, 2003.
- [2] Md. SH, Md. AM and Md. ANB, "An implementation of intrusion detection system using genetic algorithm". International journal of network security and its application vol.4, No.2, March 2012.
- [3] O. Nierstrasz and M. Lungu "Agile software assessment- program comprehension" IEEE 20th international conference on ICPC, Paesau, 11-13 june 2012.
- [4] F.Karlsson and P.J.Ågerfalk, "Exploring Agile Values in Method Configuration", European Journal of Information Systems 18(4), (2009), 300-316.
- [5] D. Gupta and R. Dwivedi, "Method Configuration from Situational Method Engineering", ACM SIGSOFT, Software Engineering Notes, 37(3), (2012), 1-11.
- [6] D. Gupta and R. Dwivedi, "A Step towards Method Configuration from Situational Method Engineering", Software Engineering: An International Journal (SEIJ), 2(1), (2012), 51-59.
- [7] Beck, K., "Embracing change with Extreme programming". IEEE computer society press 32(10):70-77, 1999(a).
- [8] Beck, K., "Extreme programming explained: Embrace change" Reading, Mass.,Addison-wessley, 1999(b).
- [9] S. Moaven et.al. "Towards an Architectural-Centric Approach for Method Engineering", IASTED conference on Software Engineering, Austria, pp. 74-79, 2008.
- [10] N. Prakash et.al "An Intention Driven Method Engineering Approach", International Conference on research challenges in Information sciences, 281-288, 2008.
- [11] D. Gupta and N. Prakash "Engineering Methods from their Requirements Specification", Requirements Engineering Journal, 3, pp.133– 160, 2001.
- [12] F. Karlsson, P.J. Ågerfalk, P.J. "Method Configuration: Adapting to Situational Characteristics While Creating Reusable Assets", Information and Software Technology 46, 619-633, 2004.
- [13] B. Fitzgerald, G. Hartnett and K. Conboy "Customizing agile methods to software practices at Intel Shannon", European Journal of Information Systems, 15(2), pp. 197-210, 2006.
- [14] Dwivedi R., "Configuring Issues and Effort for configuring agile approaches-Situational based method engineering" International journal of computer application, vol. 16(17), January 2013.

Determining the Water Quality Index of an Urban Water Body Dal Lake, Kashmir, India

¹Raj Shailesh Kanakiya, ²S.K.Singh, ³J.N.Sharma

1-Research Scholar, Environmental Engineering Department Delhi Technological University, Delhi, India

2-Professor & Head, Environmental Engineering Department, Delhi, India

3-J.N.Sharma,Supretendent Engineer, J&K pollution Control Board, J&K India

Abstract: A Water Quality Index (WQI) is a useful statistical tool for simplifying, reporting and interpreting complex information obtained for any water body. A simple number given by any WQI model explains the level of water contamination. A water quality index carried out for Dal lake based on very important parameters can provide a simple indicator of water quality. The present study deals with the monitoring of variation of seasonal water quality index of Dal lake. The index is used to improve the comprehension of general water quality issues, communicates water quality status and illustrates the need for and the effectiveness of protective practices. It is found that in all cases the change in WQI value follow a similar trend throughout the study period. Water Quality Index of Dal lake for all four basins were (102.66,120.33,109.35,101.35) for all seasons summer, winter, monsoon thus it is found to be unfit for drinking.

Keywords: Dal lake, Water quality index, Water quality parameters

I. Introduction

Dal Lake is perceived as vast expanse of water in a pristine landscape where one goes for recreation. It is a place of experiencing nature by way of boating, camping, fishing, swimming, bird watching, etc. however, when one mentions 'urban lakes' the picture is soon demystified. While urban lakes are different from the common perception of lakes in general they too have value and functions, both ecosystem functions and social values.

The word "Lake" is used loosely to describe many types of water bodies – natural, manmade and ephemeral including wetlands.

The biodiversity of lake and pond ecosystems is currently threatened by a number of human disturbances, of which the most important include increased nutrient load, contamination, acidification, and invasion of exotic species (Bronmark & Hansson, 2002). The ecological stress of the system is reflected by deterioration of water quality and increased levels of biological productivity.

Hydrologic change is the most visible impact of urbanization. Hydrology concerns the quality, duration, rates, frequency and other properties of water flow. Urbanization typically increases runoff peak flows and total flow volumes and damages water quality and aesthetics. Pollutants reach wetlands mainly through runoff. Urbanized watersheds generate large amounts of contaminants, including eroded soil from construction sites, toxic metals and petroleum from roadways, industrial and commercial areas, and nutrients and bacteria from residential areas. By volume, sediment is the most important non-point pollutant. At the same time that urbanization produces large quantities of pollutants, it reduces water infiltration capacity, yielding more surface runoff. Pollutants from urban land uses are, therefore, more vulnerable to transport by surface runoff than pollutants from other land uses.

The organic and inorganic pollutant load in the Dal has accelerated the macrophytic growth which in turn has reduced the water quality and biological oxygen demand (BOD) of the lake and hence has reduced the recreational and aesthetic appeal of the lake

The chemistry of lake water and sediment is a cumulative reflection of catchment geology, weathering and erosional processes as well as anthropogenic inputs. The chemical degradation of silicate and carbonate minerals is various processes, such as dissolution, hydrolysis, oxidation, and reduction (Gupta and Subramanian, 1998).

The basic chemical reactions take place among silicates, carbonates, and rainwater, congruently or incongruently, gives rise to various ions and clay minerals (Freeze and Cherry, 1979). Being an urban type lake, municipal and domestic effluents have altered the surface water composition of Dal Lake, leading to increased eutrophication (Hutchinson, 1999). Moreover, excessive sedimentation rates enhanced by extensive soil erosion due to deforestation and an encroachment by surrounding population have dramatically reduced the lake volume (Chakrapani, 2002). The lake serves as a resource of drinking water, irrigation, fisheries, recreation, tourism, etc.

A Water Quality Index (WQI) is a useful statistical tool for simplifying, reporting and interpreting complex information obtained from any body of water. A simple number given by any WQI model explains the level of water contamination. A water quality index based on some very important parameters can provide a simple indicator of water quality. In general, water quality indices incorporate data from multiple water quality parameters into a mathematical equation that rates the health of a waterbody with number.

Study Area

Dal Lake (Lat. 34° – 6' N, 74°-45' E, alt.1583m) situated in the heart of Srinagar city, the summer capital of Jammu Kashmir State is under tremendous anthropogenic pressure. The myriad is ways in which people use the lake along with the numerous pollutant generating activities have stressed the lake ecosystem in diverse ways.

The lake is divided into four basins

Hazratbal, Boddal, Nageen and Gagribal The lake is mainly fed by a large perennial inflow channel, Telbal nala, which drains the largest subcatchment area of about 145 km² and contributes to about 80% of the total inflow to the lake (Zutshi and Vass 1978; Trisal 1987, Jellani 2006) as well as a number of small streams, viz, Peshpaw nala, Shalimar nala, Merakhsha nala, Harshikul, etc, around the shore line besides some contribution from groundwater. Within the lake basin itself there are number of springs (Kundangar et al., 1995) which act as permanent water source to the lake. The Nageen basin is the deepest basin (maximum depth of about 6 m), and Gagribal basin the shallowest (maximum depth 2.5m). Of the total area of the lake, 4.1 km² is under floating garden or cultivation, 1.51 km² is submerged land and 2.25 km² under marshy conditions.

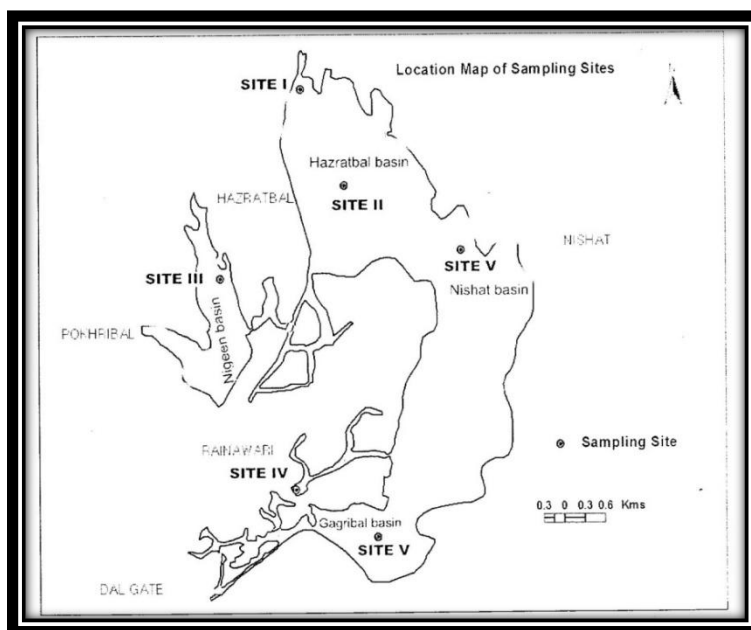
II. Materials And Methods

The Survey of the lake and its catchment area is done to know the sampling locations/ sites in the six sub-basins of the lake. The Sampling Sites is shown in the map indicated below. Samples have been collected from these Sites and the Physico –Chemical Analysis is being done in the Boards Laboratory for 12 parameters which are pH, conductivity, Hardness, DO, BOD, TDS, Total Alkalinity, Nitrite, Sulphate, Chlorides, Calcium, Magnesium.

Various physico-chemical parameters analyzed for water and sediment samples included those for pH (Digital pH meter DPH 504), Electrical conductivity (EC) (Digital EC meter DEM900). Total dissolved solid (TDS) was determined by oven dry method. Biological Oxygen Demand (BOD) was estimated by Azide modification of Winkler method. Total Alkalinity (TA) as HCO₃⁻, Calcium (Ca²⁺), Magnesium (Mg²⁺) total hardness (TH), Chloride (Cl), was estimated by standard titrimetry. Sulphate (SO₄²⁻) by turbidometry, Nitrate (NO₃²⁻) by Brucine method all using HITACHI UV-VIS spectrophotometer.

Table.1 Description of the Sites

Sr.No	Site	Description
1	Bod-dal	Central Char Chinar
2	Nageen Lake	Central part of Nageen Lake
3	Hazratbal Lake	Near outlet of Hazratbal STP
4	Gagribal Lake	Near Nehru Park



In this study, for the calculation of water quality index, twelve important parameters were chosen. The WQI used here are taken from as recommended by the standards of drinking water quality Bureau of Indian Standards (BIS).

The weighted arithmetic index method (Brown et. al.,) has been used for the calculation of WQI of the waterbody. Further, quality rating or sub index (qn) was calculated using the following expression.

$$Q_n = 100 * [V_n - V_{io}] / [S_n - V_{io}]$$

(Let there be n water quality parameters and quality rating or subindex (q_n) corresponding to n th parameter is a number reflecting the relative value of this parameter in the polluted water with respect to its standard permissible value.)

q_n = Quality rating for the n th Water quality parameter

V_n = Estimated value of the n th parameter at a given sampling station.

S_n = Standard permissible value of the n th parameter.

V_{io} = Ideal value of n th parameter in pure water. (i.e., 0 for all other parameters except the parameter pH and Dissolved oxygen (7.0 and 14.6 mg/L respectively))

Unit weight was calculated by a value inversely proportional to the recommended standard value S_n of the corresponding parameter.

$$W_n = K / S_n$$

W_n = unit weight for the n th parameters.

S_n = Standard value for n th parameters

K = Constant for proportionality.

The overall Water Quality Index was calculated by aggregating the quality rating with the unit weight linearly.

$$WQI = \sum q_n W_n / \sum W_n$$

Table 2. Water Quality Index (WQI) and status of water quality (Chatterji and Raziuddin 2002)

Sr.No	Water Quality Index	Water Quality Status
1	0-25	Excellent water quality
2	26-50	Good water quality
3	51-75	Poor water quality
4	76-100	Very Poor water quality
5	>100	Unsuitable for drinking

Table 3. Drinking Water standards recommending Agencies and unit weights. (All values except pH and Electrical Conductivity are in mg/L)

Sr.No	Parameters	Standard permissible value	Unit Weight
1	pH	8.5	0.2190
2	Electrical Conductivity	300	0.371
3	Total Dissolved Solids	500	0.0037
4	Total alkalinity	120	0.0155

Determining the Water Quality Index of an Urban Water Body Dal lake, Kashmir, India

5	Total hardness	300	0.0062
6	Calcium	75	0.025
7	Magnesium	30	0.061
8	Chlorides	250	0.0074
9	Nitrate	45	0.0412
10	Sulphate	150	0.01236
11	Dissolved oxygen	5	0.3723
12	Biological oxygen demand	5	0.3723

Table.No.4. Physico-chemical Characteristics of Sampling sites in all seasons

Sr.No	Parameter	Central Char Chinar			Nageen Central			Hazratbal STP			Nehru Park		
		Feb	Jun	Jul	Feb	Jun	Jul	Feb	Jun	Jul	Feb	Jun	Jul
1	pH	8.20	8.40	8.40	8.10	8.20	8.70	7.50	7.90	8.20	7.40	8.00	8.20
2	Conductivity	220.0	220.0	209.0	350.0	301.0	220.0	285.0	265.0	282.0	143.0	128.0	138.0
3	TDS	143	116	152	192	192	149	184	198	201	83	85	91
4	T.Alkalinity	160.0	212.0	225.0	190.0	309.0	235.0	225.0	218.0	256.0	68.0	76.0	70.0
5	T.Hardness	392	112	142	228	126	120	366	156	184	268	82	104
6	Calcium	36.00	37.00	36.00	38.00	40.00	39.00	38.00	43.00	42.00	45.69	22.40	21.64
7	Magnesium	4.50	3.60	3.80	4.00	5.00	5.00	4.00	6.00	7.00	7.42	6.30	12.15
8	Chlorides	12.0	15.0	16.0	20.0	18.0	18.0	17.0	21.0	24.0	21.0	16.0	12.0
9	Nitrates	450	395	390	480	450	539	435	490	590	420	456	580
10	Sulphates	11.0	7.0	12.0	9.0	19.0	16.0	16.0	14.0	18.0	4.35	11.7	19.20
11	DO	6.8	6.5	6.8	5.4	5.5	5.0	6.0	5.8	5.5	4.3	6	4.3
12	BOD	3.0	2.0	2.0	2.5	2.8	2.0	2.2	1.0	1.0	2.0	2	1.0

Table.No.5 Calculation for Quality Rating

Sr.No	Parameter	Central Char Chinar			Nageen Central			Hazratbal STP			Nehru Park		
		Feb	Jun	Jul	Feb	Jun	Jul	Feb	Jun	Jul	Feb	Jun	Jul
1	pH	80.0	93.3	93.3	73.3	80.0	113.3	33.3	60.0	80.0	26.7	66.7	80.0
2	Conductivity	73.3	73.3	69.7	116.7	100.3	73.3	95.0	88.3	94.0	47.7	42.7	46.0
3	TDS	28.6	23.2	30.4	38.4	38.4	29.8	36.8	39.6	40.2	16.6	17.0	18.2
4	T.Alkalinity	133.3	176.7	187.5	158.3	257.5	195.8	187.5	181.7	213.3	56.7	63.3	58.3
5	T.Hardness	130.7	37.3	47.3	76.0	42.0	40.0	122.0	52.0	61.3	89.3	27.3	34.7
6	Calcium	48.0	49.3	48.0	50.7	53.3	52.0	50.7	57.3	56.0	60.9	29.9	28.9
7	Magnesium	15.0	12.0	12.7	13.3	16.7	16.7	13.3	20.0	23.3	24.7	21.0	40.5
8	Chlorides	4.8	6.0	6.4	8.0	7.2	7.2	6.8	8.4	9.6	8.4	6.4	4.8
9	Nitrates	1000.0	877.8	866.7	1066.7	1000.0	1197.8	966.7	1088.9	1311.1	933.3	1013.3	1288.9
10	Sulphates	7.3	4.7	8.0	6.0	12.7	10.7	10.7	9.3	12.0	2.9	7.8	12.8
11	DO	156.0	162.0	156.0	184.0	182.0	192.0	172.0	176.0	182.0	206.0	172.0	206.0
12	BOD	60.0	40.0	40.0	50.0	56.0	40.0	44.0	20.0	20.0	40.0	40.0	20.0
Σ	Qn	1737.1	1555.6	1566.0	1841.4	1846.1	1968.6	1738.8	1801.6	2102.9	1513.2	1507.4	1839.0

Table.No.6 Formation of Water Quality Index

Sr.No	Parameter	Central Char Chinar			Nageen Central			Hazratbal STP			Nehru Park		
		Feb	Jun	Jul	Feb	Jun	Jul	Feb	Jun	Jul	Feb	Jun	Jul
1	pH	17.52	20.44	20.44	16.06	17.52	24.82	7.30	13.14	17.52	5.84	14.60	17.52
2	Conductivity	27.21	27.21	25.85	43.28	37.22	27.21	35.25	32.77	34.87	17.68	15.83	17.07
3	TDS	0.11	0.09	0.11	0.14	0.14	0.11	0.14	0.15	0.15	0.06	0.06	0.07
4	T.Alkalinity	2.07	2.74	2.91	2.45	3.99	3.04	2.91	2.82	3.31	0.88	0.98	0.90
5	T.Hardness	0.81	0.23	0.29	0.47	0.26	0.25	0.76	0.32	0.38	0.55	0.17	0.21
6	Calcium	1.20	1.23	1.20	1.27	1.33	1.30	1.27	1.43	1.40	1.52	0.75	0.72
7	Magnesium	0.92	0.73	0.77	0.81	1.02	1.02	0.81	1.22	1.42	1.51	1.28	2.47
8	Chlorides	0.04	0.04	0.05	0.06	0.05	0.05	0.05	0.06	0.07	0.06	0.05	0.04
9	Nitrates	41.20	36.16	35.71	43.95	41.20	49.35	39.83	44.86	54.02	38.45	41.75	53.10
10	Sulphates	0.91	0.58	0.99	0.74	1.57	1.32	1.32	1.15	1.48	0.36	0.96	1.58
11	DO	58.08	60.31	58.08	68.50	67.76	71.48	64.04	65.52	67.76	76.69	64.04	76.69
12	BOD	22.34	14.89	14.89	18.62	20.85	14.89	16.38	7.45	7.45	14.89	14.89	7.45
Σ	WnQn	172.38	164.66	161.28	196.36	192.91	194.83	170.04	170.90	189.83	158.51	155.36	177.82
Σ	Wn	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62
	(WnQn)/Wn	106.54	101.77	99.68	121.36	119.23	120.41	105.09	105.62	117.32	97.97	96.02	109.90

Graphical Representation of Seasonal Variation of Water Quality in Dal Lake

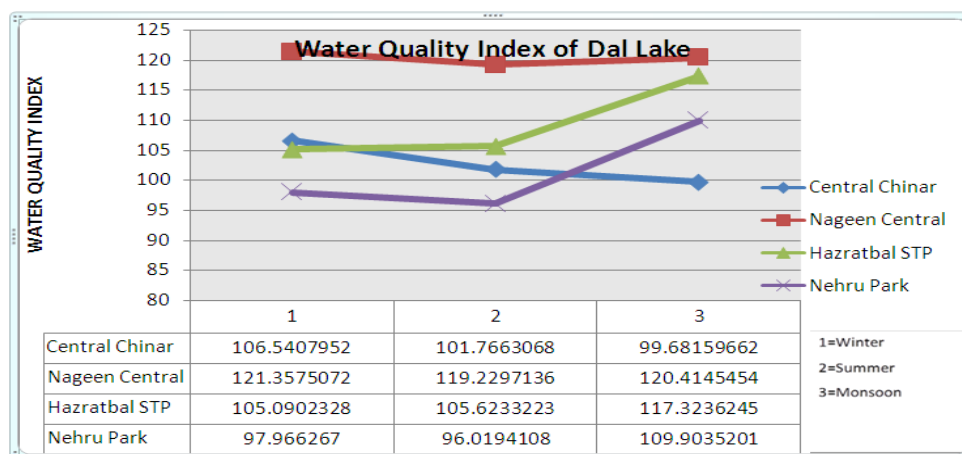


Table.No.8 Average Water Quality Index of Sampling Sites

Sr.No	Sampling Site	Average Water Quality Index
1	Central Chinar	102.66
2	Nageen Central	120.33
3	Hazratbal STP	109.35
4	Nehru Park	101.30

III. Results & Discussion

Water quality Index of the Dal lake is calculated for all seasons viz, rainy season, winter season and summer season . The values of various water related physico-chemical parameters for calculation of Water quality index are presented in Table 3. Season wise

Water Quality Index calculations are depicted in the Table 5,6 and 7.

The Water Quality Index is calculated for four sites Char Chinar,NageenCentral,Hazratbal STP,i.e, are 102.66,120.33,109.35,101.3 respectivelyas per Table.No.8 thus it indicates the water unfit for drinking (Chatterji and Raziuddin 2002).

1. PH

Amongst the 12 parameters selected for forming Water Quality Index for Dal lake, pH is an important parameter which determines the suitability of water for various purposes. In the present study pH ranged between 7.4 &8.7. Overall average values for three seasons are taken into account the waterbody was found to be slightly alkaline in nature.(Ambasht (1971), Petre (1975), Shardendu and Ambasht (1988), have also made similar observations in their studies on different waterbodies.

2. Conductivity

Conductivity is a measure of capacity of substance or solution to conduct electric current. Conductivity determines the total dissolved solids in the water. Here,the range of Electro Conductivity is average for all other lakes except Nageen where it can be inferred that there the Conductivity is attributed to high salinity and high mineral content.

3. Total Alkalinity

According to BIS the maximum permissible limit is 120 mg/L. The observed average value of total alkalinity was in greater in range of 160-390 mg/l except for Nehru Park in Gagribal basin. Totalalkalinity values in the study indicates that the water was very hard. Higher values of alkalinityregistered during summer might be due to the presence of excess of free CO₂product as a resultof decomposition process coupled with the mixing of sewage and domestic waste. The low alkalinityduring rainy season may be due to dilution. Jain et. al (1996) also reported similar finding in the study of the Halali Reservoir.

4. Chloride

Chloride is one of the most important parameter in assessing the water quality. Munawar (1970) is of the opinion that higher concentrations of chlorides indicate higher degree of organic pollution. In the present study the concentration of chloride fluctuated between 12 mg/l to 24mg/l .

5. Nitrogen

Nitrogen is an important nutrient for plant and algae growth in the lake. Nitrogen may enter a lake from surface runoff or groundwater sources. Sediments clearly cause nitrogen to undergo a number of changes.

Nitrogen (rather than phosphorus) limits algae growth. The concentration of Nitrogen here is quite high in this study it is in range of 435 mg/l to 580 mg/l.

6. D.O and B.O.D

The measurement of the DO is primary parameter in water pollution studies as it indicates aerobic or anaerobic nature of biological activities in water bodies (Trivedi, 1995).

The dissolved oxygen concentration depends on the physical, chemical and biochemical activities in the water body, and its measurement provides a good indication of water quality. The present study suggests that the concentration of dissolved oxygen fluctuated between 4.3mg/l and 5 mg/l. Seasonally, the concentration of dissolved oxygen was more during monsoon and least during summer. Thus indicating variation of dissolved oxygen according to temperature and quantity of water supply. This observation is in conformity with the observations of Reddy et.al., (1982), Ghosh and George (1989), Swarnalatha and Narasingarao (1993) and Venkateswarlu (1993). Bio-chemical oxygen demand is a parameter to assess the organic load in a waterbody.

The BOD concentration ranged between 2 mg/l to 3mg/l indicating the fact that the water body is eutrophic. Seasonally, it was high during summer, being in conformity with the observation of Chatterjee (1992). From all these foregoing observations of the physico-chemical parameters, it can be concluded that the water body shows the characters of eutrophication.

Low dissolved oxygen, high bio-chemical oxygen demand and high nitrate concentration these overall conditions indicate that highly eutrophic condition for the Dal lake.

7. Calcium & Magnesium

The observed average value of calcium was in range of 22 to 40 mg/l, which is below standard permissible limit of 75 mg/l prescribed by BIS. The quantities of calcium in natural water depends on geology, types of rock present in the catchment.

While the observed average value of magnesium was in range of 4 to 12 mg/l. Magnesium hardness particularly associated with the sulphate ion has laxative effect on persons unaccustomed to it (Khursid, 1998).

8. Sulphate

Sulphate in lake water is primarily related to the types of minerals found in the watershed and to acid rain. Industries and utilities that burn coal release sulphur compounds into the atmosphere that are carried into lakes by rainfall. In water depleted of oxygen (anaerobic water), sulphate can be reduced to hydrogen sulphide (H_2S). The range of Sulphate is 4.35mg/l to 19.2 mg/l.

9. Algal Bloom

In Dal lake the with the advent of time and due to increased and unabated human incursions within and lake peripheries have witnessed frequent algal blooms. Recurrence of such blooms has become a regular phenomenon in the various basins of the lake. A close relationship was observed between chloride and nitrates which almost coincided with those of high euglenoid population. The reason behind algal bloom can be inferred as lack of water flushing, enrichment of nutrient and building up of free carbon dioxide were the possible causes of the euglenoid bloom in Dal lake.

10. Hardness

Hardness in Dal Lake is higher than permissible. Hardness concentration of freshwater bodies of Kashmir Valley has been associated with thick population of plankton, especially Cyanophyceae (Bhat and Pandit, 2003). In this study in range is 300 to 400 mg/l in certain time of year. The values of total hardness are much fluctuating from that of summer months to winter months.

IV. Conclusion

At the outset, the study clearly indicates that the lake water is not fit for drinking consumption. It shall undergo treatment for it to be used for drinking purpose. Thus through Water Quality Index (WQI) we can have overall judgement whether that what purpose does the lake water suffice. In this method we use more systematic approach which gives comparative evaluation of the water quality of sampling stations within a water body. Water Quality index is helpful for public to understand the quality of water as well as being a useful tool in many ways in the field of water quality management.

These results are important so that the local authorities may implement preventive measures to reduce the threat of domestic and industrial discharges as well as agricultural activities discharges and have a check over the land use pattern in the Dal catchment.

References

- [1]. Ambasht, R.S., 1971. Ecosystem study of a tropical pond in relation to primary production of different vegetation zones. *Hydrobiologia* 12 : 57-61.
- [2]. BIS 1983. Standards for Water for Drinking and other purposes, Bureau of Indian Standards, New Delhi.
- [3]. Bhat SA, Pandit AK (2003). Phytoplankton Dynamics in Anchar Lake, Kashmir. *J. Res. Dev.* 3:71-96
- [4]. Brönmark C, Hansson L-A (2002) Environmental issues in lakes and ponds: current state and perspectives. *Environmental Conservation*.
- [5]. Brown, R.M., McClelland, N.J., Deininger, R.A. and O'Connor, M.F. (1972): A Water Quality Index – Crossing the Psychological Barrier (Jenkins, S.H., ed.) *Proc. Int. Conf. on Water Poll. Res.*, Jerusalem, Vol.6, pp.787-797.
- [6]. Chakrapani GJ (2002) Water and sediment geochemistry of major Kumaun Himalayan lakes, India. *Environ Geol* 43, 99–107.
- [7]. Chatterjee, A.A., 1992. Water quality of Nandakananlake. India., *J. Environ. Hlth.* 34(4): 329-333.
- [8]. Chatterjee, C. and Raziuddin, M. 2002. Determination of water quality index (WQI) of a degraded river in Asanoli Industrial area, Raniganj, Burdwan, West Bengal. *Nature, Environment and pollution Technology*, 1(2):181-189.
- [9]. Gupta LP and Subramanian V (1998) Geochemical factors controlling the chemical nature of water and sediments in the Gomti river. *India Env. Geol.* 36 (1–2), 102–108
- [10]. Ghosh, A. and J.P. George, 1989. Studies on the abiotic factors and zooplankton in a polluted urban reservoir Hussain Sagar, Hyderabad: Impact on water quality and Embryonic Development of Fishes. *Indian J. Environ. Hlth.* 31(1): 49-59
- [11]. Hutchinson G. E., *A Treatise on Limnology*, Vol. II, Introduction to Lake Biology and the Limnoplankton, John Wiley and Sons, Inc. New York (1967)
- [12]. Freeze, R.A., and Cherry, J.A., 1979, *Groundwater*: Englewood Cliffs, NJ, Prentice-Hall, 604 p.
- [13]. Jain S. M., Meenakshi Sharma and Ramesh Thakur. Seasonal variation in Physico-chemical parameters of Halali reservoir of Vidisha district, *Indian Journal of Ecobiology* 8:3, 81 – 188 (1996).
- [14]. Jeelani, G, Shah AQ (2006) Geochemical characteristics of water and sediment from the Dal Lake, Kashmir Himalaya: constraints on weathering and anthropogenic activity. *Environ Geol.* 50, 12–23.
- [15]. Khursid, S, Zaheeruddin and A. Basheer. *Ind. J. Env. Prot.* 18(4): 246-249 (1998)
- [16]. Munawar, M., 1970. Limnological studies on fresh water ponds of Hyderabad, India-II, *J. Hydrobiologia.* 35:127-162.
- [17]. Kundankar MRD, Sarwar SG, Shah MA (1995). Limnological characteristics of Hazratbal basin of Dal Lake 1992-93. Technical Report-submitted to Government of Jammu and Kashmir, India.
- [18]. Reddy, K.R.P.D. Sacco, D.A Graetz, K.L. Campbell and L.R. Sinclair. 1982. Water treatment by aquatic ecosystem: Nutrient removal by reservoirs and flooded fields. *J. Environmental Management*, 6(3): 261-271.
- [19]. Shardendu and R.S. Ambasht, 1988. Limnological studies of a rural pond and an urban tropical aquatic ecosystem: oxygen forms and ionic strength. *J. Tropical Ecology.* 29 (2): 98-109.
- [20]. Swarnalatha, N. and A. Narasingrao, 1993. Ecological investigation of two lentic environments with reference to cyanobacteria and water pollution. *Indian J. Microbial. Ecol.*, 3:41-48.

Enhanced Heuristic Approach for Travelling Tournament Problem based on Extended Species Abundance Models of Biogeography

Daya Gupta
Delhi Technological University
Computer Engineering Department
Delhi, India
Email: dgupta@dce.ac.in

Lavika Goel
Delhi Technological University
Computer Engineering Department
Delhi, India
Email: goel.lavika@gmail.com

Ashish Chopra
Delhi Technological University
Computer Engineering Department
Delhi, India
Email: ashish231039@yahoo.com

Abstract—This paper shows a heuristic approach of enhanced simulated annealing based on extended species abundance models of biogeography in order to obtain optimal solution for Travelling Tournament Problem. We upgrade the migration step of BBO by using probabilistic measures and hybridize it with simulated annealing to solve the TTP problem and avoid the problem of local minima. Our proposed hybrid approach converges to an optimal solution for TTP. There is negative impact of non deterministic problems on the TTP solution. We considered all these non-deterministic problems as noise. The physical significance of noise in our algorithm is any existing parameter which can affect the fitness of the habitat. We also calculate the overall cost of TTP for various extended species abundance models of BBO (Linear and Non linear models) to achieve desirable results. We compare the performance of our approach with other methodologies like ACO and PSO.

Keywords: BBO; Extended Species Models; TTP; Enhanced simulated annealing; ACO and PSO

1. INTRODUCTION

TTP (Travelling Tournament Problem) is a sport league scheduling problem where we need to organize the game for participating teams either at home or away and need to calculate the cost required for the schedule produced. We need to form an effective schedule, so that it attracts several viewers and thereby increase in revenue. For the development of this schedule we need an algorithm. We used biogeography based optimization technique for schedule generation and hybridize it with enhanced simulated annealing to obtain the effective schedule. BBO (Biogeography Based Optimization) is global optimization method which represents organism distribution in our biological system in terms of a mathematical model. BBO has defeated techniques ACO (Ant colony Optimization) and PSO (Particle Swarm Optimization) for various problem like land cover feature extraction, path planning and TSP (Travelling Salesman Problem). Due to combinatorial explosive nature of problems involved in TTP motivated us to apply BBO on it. In section 2 we talked about

related concepts like TTP and BBO. In this section we discussed our problem statement and BBO working procedure. In section 3 we discussed our proposed methodology for obtaining the better solution for TTP. In this portion we discussed the block diagram of our approach. We also discussed enhanced simulated annealing approach and its various operations. At the end of section 3 we explained our proposed algorithm. In section 4 we discussed the results of our approach based on extended species abundance model of BBO to the previous results. We also plot some graphs to compare the performance of our approach with the existing ones. In section 5 we discussed the conclusion of our paper.

2. RELATED CONCEPTS

A. Travelling Tournament Problem

TTP is the problem of obtaining schedule in which teams that are participating are listed and corresponding to that there are rounds which shows the location of team where they are going to play all of their matches. Our goal is to optimize this schedule by reducing the cost which is organizing cost at particular location. We use two constraints for the TTP problem:

1. Maximum: No team is allowed to play consecutively 3 games in home or away. Let n_i be the number of games by team i then three cases have to be considered.

Case 1: If $(n_{T_i} > n_{T_j})$

1.) If T_i played its last game in its home location then our assignment strategy should assign next game in T_j home location.

2.) Otherwise schedule game in T_i home location.

Case 2: If $(n_{T_i} < n_{T_j})$

1.) If T_j played its last game in its home location then our assignment strategy should assign next game in T_i home location.

2.) Otherwise schedule game in T_j home location.

Case 3: If ($n_{T_i} < n_{T_j}$)

1.) If T_i played its last game at away location and T_j played its last game at home, then schedule the next match between them at T_i home location.

2.) If T_i played its last game at home location and T_j played its last game at away, then schedule the next match between them at T_j home location.

3.) Otherwise, use random function to assign either at T_i or T_j

2. Without repetition: If T_i and T_j played a match in T_i location then second match between T_i and T_j is not going to be played at T_j home location.

B. Working Procedure of BBO

BBO is an evolutionary algorithm whose working principle is based upon migration mechanisms of species from one habitat to other depending upon the fitness of the habitat which are favorable to them. The habitat which have high HSI (high suitability index) have high value of species count. Therefore habitat which has high value of HSI have high emigrating rate i.e. it is ready to send its SIV to other habitat, while the habitat having low value of HSI have low value of species count and their immigrating rate is high, i.e. it is ready to accept species towards itself. HSI of a habitat can be affected on the basis of SIV (suitability index variables) which are independent variables. Firstly the question arises why there is need of modification in the current algorithm. We are dealing with TTP i.e. Np-hard problem, which is dynamic in nature so we have to modify our solution which can take care all dynamic constraints like problem of earthquake, floods etc. of nature.

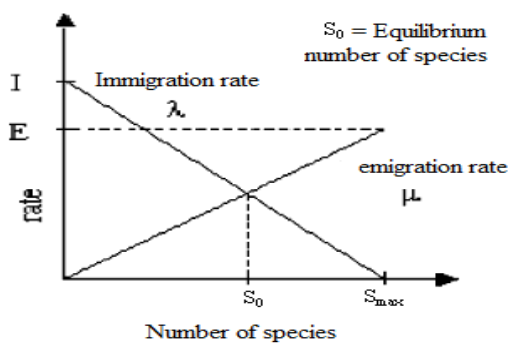


Fig.1 BBO working principle

The above diagram illustrates the basic mechanism and relation between immigration curve and emigration curve. Here I is maximum Immigration rate, E is maximum emigration rate, s_0 is equilibrium number of species, μ is emigration rate and λ is Immigration rate.

Let we consider two habitats HB_1 and HB_2 . These habitats have their fitness's as f_1 as immigrating and f_2 as emigrating habitat. Let noise involved in two habitats is n_1 and n_2 . Due to affect of noise the measured fitness is $f - n$ instead of f . If we consider f_1 has more fitness than f_2 and let n_1 has huge value than n_2 and both high value than f_1 and f_2 . Therefore the overall fitness becomes:

$$f_1 > f_2 \quad (1)$$

$$n_1 > n_2 \quad (2)$$

$$f_1 - n_1 < f_2 - n_2 \quad (3)$$

Therefore HB_1 accepts the SIV from HB_2 , therefore condition of BBO gets satisfied as immigrating habitat fitness is less than emigrating habitat. But population of HB_1 is already high due to its high HSI. The BBO migration procedure will corrupt. Its measured fitness is more if don't consider noise, so this immigration should not be done. That's why we need to modify it.

In order to calculate the uncertainties, we use the concept of noisy BBO [5].

$$U = U_{prev} + V_1 \quad (4)$$

$$E = E_{prev} + \frac{U_{prev}}{U_{prev} + V_2} (z - E_{prev}) \quad (5)$$

$$U = \frac{U_{prev} \times V_2}{U_{prev} + V_2} \quad (6)$$

Where U is the uncertainty of the state estimate, m is the estimated fitness, z is the measured fitness, V_1 is the variance of the process noise, and V_2 is the variance of the observation noise. The uncertainty U_{prev} and the estimated fitness E_{prev} are the values from the previous iteration step before the most recent fitness measurement is updated. The process noise is assumed to be zero, therefore the uncertainty U is only related U_{prev} and V_2 .

$$U = \frac{U_{prev} \times V_2}{U_{prev} + V_2} \quad (7)$$

$$U = \frac{U_{prev}}{\frac{U_{prev}}{V_2} + 1} \quad (8)$$

Because $\frac{U_{prev}}{V_2} > 0$, now $\frac{U_{prev}}{V_2} + 1 > 1$. Therefore $\frac{U_{prev}}{V_2} + 1 < U_{prev}$. With each step in the above modified algorithm, the uncertainty U will be reduced according to V_2 and U_{prev} .

Small value of uncertainty leads to high accuracy of estimated fitness.

If limit tends to infinity, our purposed approach gives an estimate value of the fitness which is equal to the real value.

3. PROPOSED METHODOLOGY

Our problem is to deal with dynamic problems, to refine the schedule, minimized the cost of the schedule for TTP and to solve the problem of local optima. For solving these issues we proposed an approach to add some features to the immigration step of BBO and then apply enhanced simulated annealing to the seeds produced by BBO and apply this approach to all extended species abundance models[2] and compare our results to the previous obtained solutions.

A. Architecture

The architecture for our approach is represented in fig 2. Our input is a distance matrix of $N \times N$, where N is team size. This team matrix contains the cost of travelling from one location to other as teams have to move away to play the matches at challengers venue. Our algorithm is evolutionary algorithm that works on migration mechanisms of species, so we need to map given input to this problem. First we generate random permutation of size N by applying fast constructive heuristic. We use roulette wheel selection method to make initial the schedule. We calculate the HSI for each habitat. We map the total cost which is given to us in terms of matrix to HSI by using mapping function stated in Table 4. We need to calculate species count because it is the parameter for calculating the growth rate and decline rate. We implemented both linear models and non linear models of extended species abundance models to check the correctness of our algorithm. We compare our approach with the existing ones by testing it on matlab platform. We did simulation to calculate performance of our approach over 100 generations.

$$\text{Growth Rate } \sigma = I(1-k/n) + c/k; \quad (9)$$

$$\text{Decline Rate } \rho = E\phi k/2 + 2\sqrt{nk}; \quad (10)$$

k = Species count, n = Total number of species, ϕ is dependency factor, I = Maximum immigration rate, E = Maximum emigration rate.

After getting the growth rate we go for modification phases to select the SIV which is going to be replaced. We use the decline rate in order to select the habitat which is going to send its SIV. After the implementation of our algorithm to the problem, we hybridize it with efficient simulated annealing to obtain the efficient schedule and to solve the problem of local minima. Enhanced simulated annealing has various stages. First it produces the neighborhood for by using swapping moves. Then we apply standard simulated annealing algorithm in order to obtain feasible schedule. After those concepts of reheat and strategic oscillation is used in order to improve results.

B. Efficient Migration Mechanism of BBO

We add some feature to the migration step of BBO by using some probabilistic measures. Our added features composed of three phases for optimizing the results.

1) Compare with the other optimal selected SIV

In this phase we consider two habitats h_1 as immigrating habitat and h_2 act as emigrating habitat. We consider two instances of h_1 as h_{1x} and h_{1y} . Firstly h_{1x} is going to accept optimal SIV from h_2 and then h_{1y} accept another best suitable SIV from h_2 and after that their performance gets measured based on the probabilistic measures as:

$$P_1 = 1 - (1 - P(f_{1x,t+1} > f_{1x,t})) (1 - P(f_{1y,t+1} > f_{1y,t})) \\ = 1 - (1 - P(f_{1,t+1} > f_{1,t}))^2 \quad (11)$$

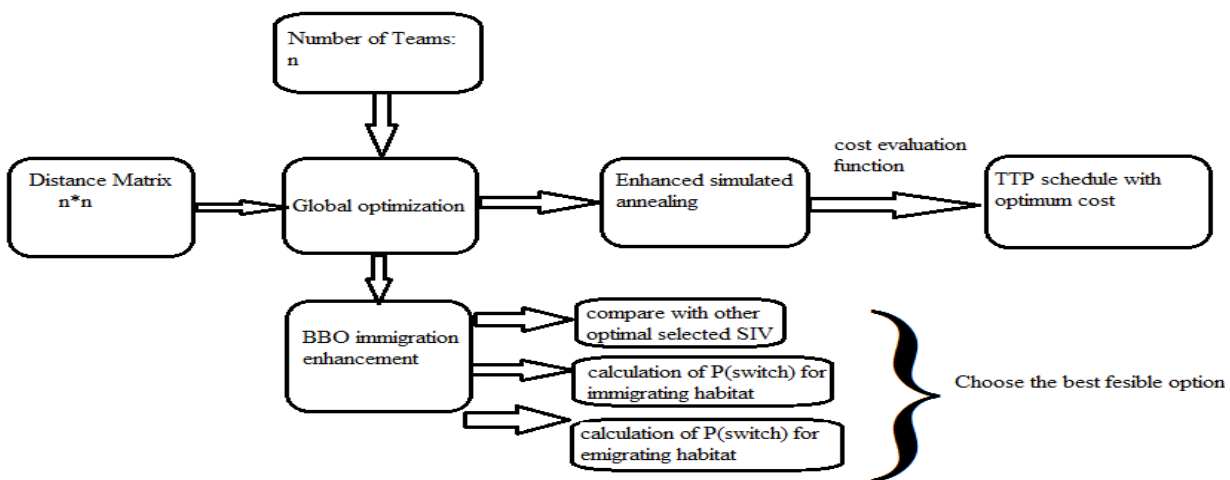


Fig 2. Proposed Architecture

$P(f_{1,t+1} > f_{1,t})$: It tells the probability of a habitat h_1 with fitness $f_{1,t+1}$ after accepting selected SIV greater than fitness $f_{1,t}$.

2) *Calculation of $P(\text{switch})$ for Immigrating Habitat*
In this phase we calculate $p(\text{switch})$ from equation (15) which can be calculated as a integration to the function of $f_p(z1), f_q(z2)$ from [16] and [17] for immigrating habitat. These are two probability density functions whose value is computed by equation [15]. If the new estimated fitness of immigrating habitat after re-evaluation is still worse than emigrating habitat, migration is going to occur, else migration don't exists.

$$P_2 = p_{\text{new}}(f_{1,t+1} > f_{1,t}) \quad (12)$$

3) *Calculation of $P(\text{switch})$ for emigrating habitat*

In this phase we calculate $p(\text{switch})$ for the emigrating habitat

$$P_3 = p_{\text{new}}(f_{1,t+1} > f_{1,t}) \quad (13)$$

From the above phases we choose the best option for the immigration step.

Secondly we map this approach to all the variants of extended species abundance models of BBO and implement it on TTP problem. We modified the immigration step and apply this modification approach to all the linear and non linear models of BBO to check whether we are able to achieve the optimal results or not. We test our algorithm to obtain various results which provide optimal solution for TTP problem

C. Equations Used

$$(P(f_{1,t+1} > f_{1,t} | f_{2,t} > f_{1,t})) = \frac{P(iSIV > rSIV, f_{2,t} > f_{1,t})}{P(f_{2,t} > f_{1,t} | m_2 > m_{1,t})} \quad (14)$$

The above equation is generalization of Baye's rule.

$(P(f_{1,t+1} > f_{1,t} | f_{2,t} > f_{1,t}))$: Probability of a habitat with fitness $f_{1,t+1}$ after accepting selected SIV greater than fitness $f_{1,t}$ given that $f_{2,t} > f_{1,t}$.

$P(f_{2,t} > f_{1,t} | m_2 > m_{1,t}) = 1 - P(\text{switch})$ where $P(\text{switch})$ is given by :

$$P(\text{switch}) = \int_0^{2y} \int_0^{2z} f_p(z1) f_q(z2) dz_1 dz_2 \quad (15)$$

$\frac{-y(-y+4x)}{24x^2}$; Where y, x, z are noise and fitness range parameters

The PDF of p is as follows

$$f_p(z1) = \begin{cases} \frac{2x+z1}{4x^2}, z1 \in [-2x, 0] \\ \frac{2x-z1}{4x^2}, z1 \in [0, 2y] \end{cases} \quad (16)$$

The PDF of q is as follows.

$$f_q(z2) = \begin{cases} \frac{2y+z2}{4x^2}, z2 \in [-2x, 0] \\ \frac{2y-z2}{4y^2}, z2 \in [0, 2y] \end{cases} \quad (17)$$

D. Fitness Function Used

The fitness function is used to calculate the fitness of the habitats on which our equations depends.

$$\text{fitness} = ((x_1^2 + x_2^2 + 100)^{\frac{1}{4}} - 3.6552) * 10$$

Such that: $x_1, x_2 \in [0, 10]$ where x_1 is the first SIV, x_2 is the second SIV and their range is between 1 to 10

E. Enhanced Simulated Annealing

We also apply simulated annealing to seed produced by above algorithm and obtain neighborhood by using mentioned five moves:

1) Obtaining Neighborhood

a) *Swap Homes (S, T_i, T_j)*: This is the first step to make a change in the initial schedule. What we have to do is the schedule which we obtained after applying approach to BBO we need to swap home locations of team T_i, T_j together to make an upgradable in obtained schedule.

b) *Swap rounds (S, r_a, r_b)*: Here we swap rounds. These are the r_a, r_b two rounds which we need to swap. Here we need to use $O(n^2)$ moves to proceed further.

c) *Swap teams (S, T_i, T_j)*: Swap team T_i, T_j .

d) *Partial Swap rounds (S, T_i, r_a, r_b)*: For the given team T_i we need to swap r_a, r_b partially i.e. only few values of a selected round gets changed.

e) *Partial Swap Teams (S, T_i, T_j, r_a)*: This is the last swapping that we did to our schedules.

We make the initial schedule represented in table 1 which is produced after roulette wheel selection method on a random permutation of 1 to 6 numbers. We apply roulette wheel selection method five times to generate the first half of the schedule.

Table 1 Initial schedule:

T/R	First Half					Second Half				
	1	2	3	4	5	6	7	8	9	10
1	6	-2	4	-3	-5	-6	-2	-4	3	5
2	5	1	-3	6	4	-5	-1	3	-6	-4
3	-4	5	2	1	6	4	-5	-2	-1	-6
4	3	6	-1	5	-2	-3	-6	1	-5	2
5	-2	-3	6	-4	1	2	3	-6	4	-1
6	-1	-4	-5	-2	-3	1	4	5	2	3

Applying the above swapping methods to our initial schedule we generate the neighborhood. The optimal neighborhood is:

Table 2 Optimal neighborhood

	First Half					Second Half				
<i>T/R</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
<i>1</i>	6	-2	2	3	-5	-4	-3	5	4	-6
<i>2</i>	5	1	-1	-5	4	3	6	-4	-6	-3
<i>3</i>	-4	5	4	-1	6	-2	1	-6	-5	2
<i>4</i>	3	6	-3	-6	-2	1	5	2	-1	-5
<i>5</i>	-2	-3	6	2	1	-6	-4	-1	3	4
<i>6</i>	-1	-4	-5	4	-3	5	-2	3	2	1

Second half of the schedule is just the mirrored image of first half in provided a constraint i.e. the team which plays home in first half, plays away in second half and vice versa.

2) Simulated Annealing Algorithm:

Now we apply simulated annealing algorithm to the above neighborhood in order to converge our solution.. Previously it has been done in three steps as discussed in [7], but by doing it in five steps and collaborate it with all extended abundance species models we are able to produce better results.

3) Strategic Oscillation

This strategy is used in various local search approaches. In this we include the weight parameter ω during the search operation. Here we multiply the parameter ω to the constant $\delta > 1$ we produce infeasible schedule and we divide by constant $\theta > 1$ if we are able to produce feasible schedule.

4) Reheats

This is used in the simulating annealing approach to avoid the problem of local minima. When simulated annealing goes to very low temperature, we are unable to avoid the problem of local minima because the probability of accepting increasing moves is much less. By applying this approach we increase the temperature of TTSA, their by solve the problem of local minima.

5) Objective function

$$Cost(s): \begin{cases} cost(S); \text{ if } S \text{ is feasible} \\ \sqrt{cost(s)^2 + [w.f(nbv(S))]^2} \text{ otherwise} \end{cases}$$

$Cost(s)$: It is the cost evaluation function in order to calculate the cost of the schedule.

F. Proposed Algorithm

We proposed the algorithm in table 3 which shows our work. In first step we calculate the random permutation p for generating the value of maxgen. This population size gets generated using fast constructive heuristic method. After that

we calculate value of HSI for each habitat. We need to map the given cost to species count. We can achieve this by species count function used in table 4. We calculate decline rate and growth rate stated in [2]. In next step we use the calculated growth rate for selection of SIV of immigrating habitat. Then we use the modification phases to check that which SIV is the best for replacement. Let SIV in immigrating habitat which gets selected for the replacement after getting output from modification phases has index j . The j th indexed SIV of immigrating habitat gets stored in temporary location. On this j th index, x th SIV on j th index from emigrating habitat gets replaced. The value stored in temporary variable gets restored back in immigrating habitat on the location where we found x th SIV value, except at j th location because it is the replaced ones. This will work until the loop terminates. We apply enhanced simulated annealing as discussed in section. E n the schedule produced by modified BBO.

Table 3 Efficient algorithm

1. Generate random permutation equal to team size P by calling fast constructive heuristic P times.
2. Calculate HSI of each habitat.
3. For $i = 1$ to $Maxgen$
4. Use $species\ count()$ described in (Table 4) to map cost to species count and calculate species count k and total number of species n .
5. Calculate evolution rate and extinction rate [2].
6. Calculate growth rate σ and decline rate ρ from the above specifications.
7. Use decline rate ρ to select an emigrating habitat.
8. For $j = 1$ to $length\ of\ population\ member(team\ size)$
9. Choose σ to select SIV from immigrating habitat.
10. Go to the three modification phases for choosing the best option:
 - a.) Choose the other optimal selected SIV.
 - b.) Calculate $P(switch)$ for immigrating habitat.
 - c.) Calculate $P(switch)$ for emigrating habitat.
11. From the above option choose the best probability by using optimal selection procedure.
12. If SIV of immigrating habitat gets selected, store j (its index) to replace in index $array[]$.
13. End for.
14. Value at j th index of immigrating habitat is stored in temporary variable and value of x th SIV at j th index of emigrating habitat gets selected and stored in j th index of immigrating habitat and we search where value of x th SIV exists in immigrating habitat excepts at j th location because it is the replaced ones, at that location store temporary variable value.
15. End for
16. Use the generated schedule as input to enhanced simulated annealing and perform its operations as discussed in section 3.E on it.

Table 4 Species count (): Mapping function

1. **for** $i = 2$ to $size\ of(Population)$
2. $present\ Cost = present\ Cost - Population(i).cost$
3. $Population(i).species\ Count = present\ Cost$
4. **end for**

4. COMPUTATIONAL RESULTS

In this section we simulate our approach on matlab over 100 generations. We implement our hybridized heuristic to extended species abundance models for various team size and obtain optimal results in table 5. There are two types of extended species abundance models: linear models and non-linear models. For each model we produce a separate optimal schedule and calculate cost relative to it. We show the output of result produced for NL6 team for linear models in table 3 and for non linear models in table 4.

A. Parameter Specifications

1. BFC (best feasible cost): This is the overall cost of the schedule by applying cost function to it.
2. BFS (best feasible schedule): This is the feasible schedule produced after implementation of our hybrid approach with enhanced simulating annealing.
3. BIC(best infeasible cost):
4. BIS(best infeasible schedule):
5. NovBF(number of violations in best feasible):
6. NovBI(number of violations in best infeasible):

B. Matlab Simulation over 100 generations

In this section we did simulation to verify our results on matlab. We compare the performance evaluation of our approach against the existing ones. We plot a graph in fig 3 which represents cost computation calculated for linear models by our approach and standard BBO under noisy conditions for National League 6 team over 100 generations. In fig 4 we plot a graph between the comparisons of cost evaluation for non linear models between our approach and standard BBO for NL6 team. This graph sows the cost computation calculated in order to solve the TTP. In table 5we represent the comparative study of our approach with existing ones for various team sizes.

Table 3 Linear model results for NL6 team

Models	BFC	BFS	BIC	BIS	Nov BF	Nov BI
a. Constant Immigration and Linear Emigration Model	28757	<6*10 double>	2.1739e+04	<6*10 double>	0	16
b. Linear Immigration and Constant Emigration Model	29745	<6*10 double>	2.0143e+04	<6*10 double>	0	16
c. Linear Immigration and Linear Emigration Model	27701	<6*10 double>	2.0653e+04	<6*10 double>	0	16

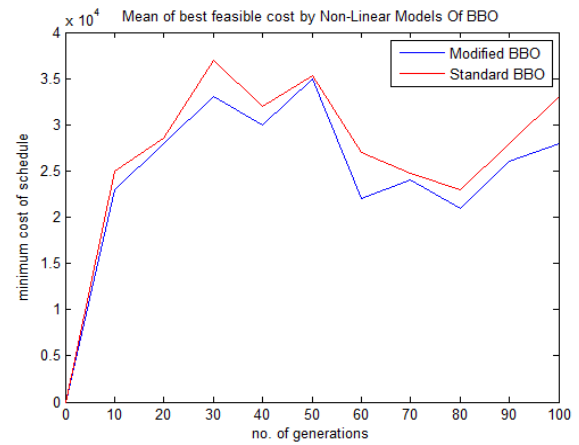


Fig 3 Comparison of cost produced by linear model for Modified BBO and for Standard BBO

Table 4 Non Linear Models for NL6 Team

Model	BFC	BFS	BIC	BIS	Nov BF	Nov BI
a. Trapezoidal Migration Model	28965	<6*10 double>	2.0537e+04	<6*10 double>	0	18
b. Quadratic Migration Model	27912	<6*10 double>	2.1003e+04	<6*10 double>	0	18
c. Sinusoidal Migration Model	28708	<6*10 double>	2.1893e+04	<6*10 double>	0	18

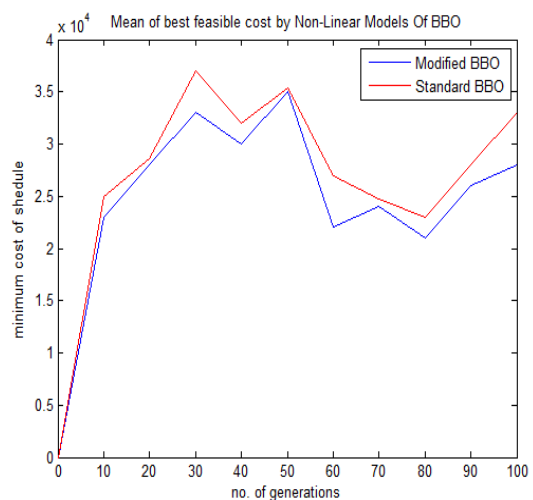


Fig 4 Comparison of cost produced by Non linear model For Modified BBO and for Standard BBO.

Table 5 Comparative Study

Instances	MBBO /ESA	ACO	PSO	BBO/SA
NL8	39721	44112	43805	42568
NL10	61608	67264	66331	66121
NL12	119608	121981	121070	120040
NL14	199363	208086	210132	207848
NL16	278305	290188	291394	281963
NL24	500756	511256	513543	510095

C. Performance comparison with other Algorithms:

We did simulation over 100 number of generations for cost produced by modified BBO and by standard known algorithm like ACO(Ant colony optimization) and PSO(particle Swarm Optimization).These simulations have been done on matlab platform. We calculate the minimum cost obtained by all Algorithms for TTP ACO, PSO and by our proposed approached for BBO and plot a graph between their cost which they calculated to obtain as a solution for Travelling Tournament Problem. Our calculated cost is less than their cost, so our performance is better.

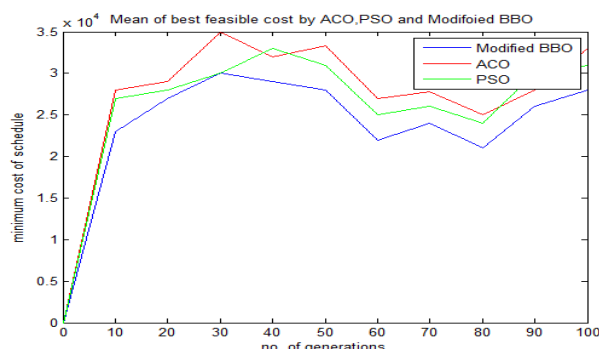


Fig 5 Comparison Between the cost evaluated by ACO, PSO, Modified BBO

5. CONCLUSION

We proposed an approach which enhances the correctness of immigration step of standard BBO algorithm by using probabilistic measures and hybridized it with enhanced simulated annealing which affects the overall cost for TTP problem in order to optimize the schedule, so that best feasible cost is less. To achieve the optimal results and in order to solve the problem of local minima we implemented enhanced simulated annealing instead of Local search or Tabu search . Our algorithm converges to local minima very quickly and accurately. We did matlab simulation for our approach to the various extended species abundance models of BBO to measure the performance of our algorithm with the existing

ones over 100 generations. We also compare the performance of ACO, PSO and our proposed algorithm in order to obtain a solution for TTP problem.

References

- [1] Daya Gupta, Lavika Goel and Vipin Aggarwal, "A hybrid Biogeography Based optimization on Traveling tournament problem" at third international conference on contemporary computing pp 325-331 in 2013,
- [2] Daya Gupta, Lavika Goel, V.K. Panchal "Extended Species Abundance models of Biogeography Based Optimization" at fourth international conference on computational intelligence, modeling and simulation pp-332-338 in 2012
- [3] G. Kendall, S. Knust, C. Riberio and S. Urrutia, "Scheduling in sports: An annotated bibliography," Computers and Operations Research, vol. 37, pp. 1-19, 2010.
- [4] D. Simon, "Biogeography-based optimization," IEEE Transactions on Evolutionary Computation, vo. 12 (6), pp. 702-713, 2008
- [5] Dan simon and Haiping ma "Noisy biogeography based optimization" IEEE conference pp 201-208, 2010.
- [6] D. Palupi Rini, S. M. Shamsuddin and S.S.Yuhaniz"Particle Swarm Optimization Technique, System and Challenges," in "International Journal of Computer Applications", no. 1, vol. 14, pp. 19-27 in 2011.
- [7] L. Goel, D. Gupta, V.K. Panchal, "Information Sharing In Swarm Intelligence Techniques: A Perspective Application for Natural Terrain Feature Elicitation in Remote Sensing Images", International Journal of Computer Applications (IJCA), vol. 32, 2011.
- [8] R. Poli "Analysis of the Publications on the Applications of Particle Swarm Optimization",in Journal of Artificial Evolution and Applications, pp. 1-10, 2008.
- [9] L. DiGasparo and A. Schaerf "A composite-nighbourhood tabu search approach to the traveling tournament problem," in Journal of Heuristics, vol. 13, pp. 189-207, 2007.
- [10] F. Costa, S. Urrutia and C. Riberio on "An ILS heuristic for the traveling tournament problem with fixed venues,"at 7th International conference on the theory of automatic timetabling, Montreal, 2008.
- [11] J. Dinitz, E. Lamken and W. Wallis on "Scheduling a tournament," in Handbook of Combinatorial Designs, CRC Press, pp. 578-584, 1995.
- [12] Tajbakhshl, K. Eshghi and A. Shamsi, "A hybrid PSO-SA algorithm for the Traveling Tournament Problem," in International conference on Computers & Industrial Engineering, Troyes, 2009.
- [13] K. Cheung, "Solving mirrored traveling tournament problem benchmark instances with eight teams," DiscreteOptimization, vol. 5, pp. 138-143, 2008.
- [14] L. Tseng and S. Liang, A Hybrid Metaheuristic for " Computational Optimization and Applications, Vol. 34, Issue 1, pp. 85- 113, 2006.
- [15] P. Stroud, \Kalman-extended Genetic Algorithm for Search in Nonstationary Environments with Noisy Fitness Evaluations," IEEE Transactions on Evolutionary Computation, Vol. 5, Issue 1, pp. 66/77, February 2001.
- [16] K. Easton, G. Nemhauser, M. Trick, "Solving the Travelling Tournament Problem: A Combined Integer Programming and Constraint Programming Approach", E. Burke and P.D. Causmaecker (Ed.), Practice and Theory of Automated Timetabling IV, Lecture Notes in Computer Science, vol. 2740, Springer, pp. 100-109, 2003
- [17] Lavika Goel, Daya Gupta, V.K. Panchal, "Dynamic model of blended biogeography based optimization for land cover feature extraction", Springer CCIS 306 (LNCS), pp. 8-19, IC3, 2012.

Environmental Management in Mass Gatherings: A Case Study of Maha Kumbh Mela 2013 at Prayag, India

Prof. S.K. Singh

*Professor & Head of the Department
Department of Environmental Engineering
Delhi Technological University, Delhi*

Anshika Bisht

*Master of Technology
Department of Environmental Engineering
Delhi Technological University, Delhi*

Abstract

A mass gathering is an essential important aspect of modern society. There are various types of gatherings where thousands of people actively participate. A mass gathering is an effortless way to interact and socialize among the various groups of people. There are various challenges that usually come across while conducting an event which leaves an undesirable impact on the environment. This paper discusses the environmental management in a mass gathering by taking a case study of Maha Kumbh 2013 at Prayag, India. By analyzing the data and information collected regarding the event, a conclusion can be drawn that although such gatherings can entail significant health risks, but there are ways of curbing the ill effects. Also some technological advancement in the field of treatment has made, which can be implemented in other rallies, political events, sporting events, conferences, award ceremonies, other religious gatherings like Haj, Rath Yatra, fairs and fates etc.

Keywords: Environmental Management, Maha Kumbh, Health Risks, Technological Advancements.

I. INTRODUCTION

A public event is an event addressed to a large number of public. Mass gathering is though an assemblage of more than 1000 person at a particular place for a certain period. It is a preplanned event held for a definite period of time at a particular location. It generally includes number of events like political rallies, sporting events like cricket match, basketball match, conferences, religious gatherings like Rath Yatra, Haj, Kumbh Mela, etc, exhibitions, carnivals etc.

There are various roles & responsibilities while conducting any public event. Any public event begins as a concept being proposed by any individual or organization. Firstly a pre-event planning is done in which various issues are discussed like when or where the event will be held, type of event, number of audience interested, duration of the event etc. But the basic and most important aspect of any event conduction is the effects that it can create on the environment. So a detailed study of the largest gathering in the Earth, in terms of identification of beneficial practices, and the failures and success stories of the event is the vital theme of this paper.

Kumbh mela is a mass gathering of Hindus along the banks of the holy river Ganga in India. It is held by turn in four different cities of India i.e. Allahabad, Haridwar, Ujjain and Nashik. The venue depends on the position of the stars and planets. It is held every third year at one of the four venues by rotation: Haridwar, Allahabad, Nashik, Ujjain. The Ardh(i.e. half) Kumbh fair is celebrated in every six years only at Haridwar and Allahabad. While Maha Kumbh takes place once in twelve years only at Allahabad.

Kumbh is an ancient word that is mentioned in The Vedas according to the Hindu mythology which means pitcher and Mela means fair in Hindi. The pilgrimage is held for about one and a half months at each of the above four places where it is believed in Hinduism that drops of nectar fell from Kumbh carried by Gods after sea was churned. The festival is considered as the “biggest mass gathering on the Earth” and attendees from all over the world can be seen during the fair. There is no precise method of finding out the number of pilgrims even approximately but it is estimated that on auspicious days about two to eight millions pilgrims attend the Kumbh Mela.

Recently Kumbh Mela was held at Allahabad in 2013. To manage this event, the organizing authorities had set up various facilities for the attendees like temporary roads, toilets, hospital and ambulance facilities, fire stations, water and sewerage facilities, etc. There were numerous difficulties, unforeseen conditions along the way. But event was overall success.

For more specification, Maha kumbh 2013 held at Prayag, Allahabad is the main area of study. The main objective of the research is to focus on all the environmental challenges been confronted by the organizing authority of Kumbh 2013 and the measures that could be taken to further enhance the effectual management. The emphasis is basically paid on the introduction of the environmental friendly ways to conduct any large event. We can also relate the conclusions been drawn by the research to the other major mass events held all over the world.

Following are figures showing the rapidly growing number of attendees, according to *The Imperial Gazetteer Of India*:

YEAR	ATTENDANCE
1998	10 million
2001	40 million
2007	70 million
2013	100 million

II. ORGANIZATION OF MAHA KUMBH MELA 2013 AT PRAYAG, INDIA:

The Maha Kumbh held at Prayag in 2013 was started on 14 January 2013 and ended on 10 March 2013. There were approximately 30 million people who visited the ending ceremony. While the total estimates of people visited is quite complex to identify but a rough idea of about 100 million can be estimated for 55 days festival.

A. Study Area:

The study area is the city of Allahabad in the state of Uttar Pradesh, India. It lies on coordinate 25.4500 °N, 81.8500 °E. Located in southern Uttar Pradesh, the city's metropolitan area covers 63.07km². It is connected by road to prominent cities of country like Jhansi, Kanpur, Lucknow, Jaipur, Bhopal, New Delhi etc. Buses operated by Uttar Pradesh and other State Road transport Corporation as important means of road transport.

Allahabad Junction is one of the main railway junctions in the Northern India and headquarters of the North Central Railway Zone. The four main stations in Allahabad are Prayag Station, City Station at Rambagh, Daraganj station and Allahabad Junction. The city is connected to most of the Uttar Pradesh cities and other major cities like Chennai, Vishakhapatnam, New Delhi etc.

Allahabad is served by Allahabad Airport, which is 12 kilometres from the city centre. Air India and Spice Jet connects Allahabad to Delhi and Mumbai. Inland waterways Authority of India connect Allahabad to Haldia.

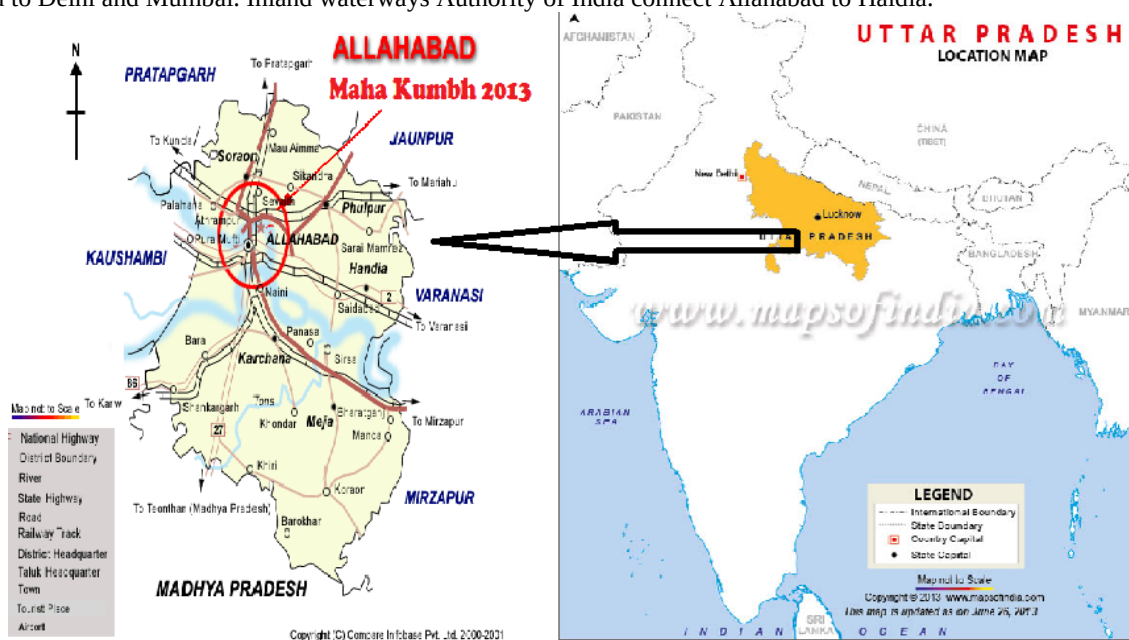


Fig. 1: Geographical Map of Allahabad

B. Background of problems:

India is world's second largest populated country with country's population near 1.2 billion according to 2011 census. With such a huge population, the religion, faith, thinking and beliefs vary differently for different lot. Every sect of people has their own festivals and their ways of celebration. Generally these celebrations end up becoming an event of mass get-together. These events are observed with large number of turnouts with limited span and space. With this increasing number of events, it becomes imperative for the Government to get involved and make required arrangements to keep the functioning of the event up to date. The authorities need to get prepared for all kinds of casualties, security, disease outbreak, health and sanitation etc. Proper planning is essential for conducting any kind of event.

Kumbh Mela is generally acknowledged to be an event that is organized in religiously traditional method. With almost 10 million people gathering in a limited area, it becomes difficult for the organizing committee to take care of needs of each and every individual. This is similar to establishing and running a whole new city with all the basic amenities in a short duration.

C. Planning at Maha Kumbh 2013:

Kumbh Mela is a massive event attracting millions of human population from all over the world. Every time there are new challenges. It was indeed a tough job for the planners to assess the data based on the previous year's Kumbh Mela and extrapolate the current estimated people travelling down by considering the factors such as population growth, infrastructure, road connectivity etc. the following data provides consideration for comparison:

- (1) Maha Kumbh 2001 was for 44 days while Maha Kumbh 2013 was for 55 days
- (2) India's Population was 102.87 crores in 2001 which has mounted to 121.02 crores in Census 2011.
- (3) The state population of Uttar Pradesh was 16.61 crores in 2001 which has raised by 20% in 2011.
- (4) The population of Allahabad district in 2001 was 49 lakhs and in 2011 it was 59 lakhs.

Estimated pilgrims during Kumbh Mela as estimated by Allahabad Nagar Nigam:

Table - 1
Important Days and Estimated Visitors At Maha Kumbh 2013

S.No.	Bathing Days	Kumbh 2001		Kumbh 2013	
		Dates	Estimated Visitors	Dates	Projected Visitors
1.	Makar Sankranti	14.1.2001	100 lacs	14.1.2013	110 lacs
2.	Paush Purnima	09.1.2001	50 lacs	27.1.2013	55 lacs
3.	Mauni Amawasya	24.1.2001	276 lacs	10.2.2013	305 lacs
4.	Vasant Panchami	29.1.2001	175 lacs	15.2.2013	193 lacs
5.	Maghi Purnima	08.2.2001	150 lacs	25.2.2013	165 lacs
6.	Maha Shivratri	21.2.2001	50 lacs	10.3.2013	55 lacs

Various amenities provided by the Government were: road, water supply, land area, toilets, roadways, railways, fire stations, hospitals etc. as mentioned in Table 2:

Table - 2
Services Provided By Allahabad Government

S.No.	SERVICES	UNITS	2001	2013
1	Area	Hectares	1495.31	1936.56
2	Toilets	Number		
2.a	Individual	Number	20481	35000+
2.b	Sulabh Complex (10 seaters)	Number	20	340
2.c	Trench Pattern	Number	17100	7500 (PRAI type)
2.d	Non Conventional	Number	0	1000+
3	Buses (Special)	Number	2824	3608
4	Trains	Number	600	750
5	Water supply	Kilolitres	56000	80000+
6	Roads laid	Kilometres	96.40	156.20
7	Hospitals			
7.a	Allopathic	Number	14	14
7.b	Homeopathy	Number	07	12
7.c.	Ayurvedic	Number	10	12

PRAI: Planning Research and Action Institute

III. VARIOUS ENVIRONMENTAL CONCERN IN KUMBH MELA 2013

There are various factors which are been affected by the gathering of masses which are enlisted as follows:

- (1) Water supply and quality
- (2) Public health
- (3) Sanitation
- (4) Waste water management
- (5) Noise Pollution
- (6) Air pollution

Each of the above issues is quite important regarding the environmental approach of event management. Following are some challenges faced while conducting Kumbh Mela 2013:

A. Water supply and Quality

Each day more than 2.9 billion litres of untreated waste water is dumped in Ganga by nearby industrial areas of Allahabad. One of the major challenges at the Kumbh is managing the water flow for the millions of people who come to bathe. The clean water availability could not meet the requirements of the huge mass, predominantly on the special days. At Sangam, Allahabad, BOD of Yamuna and Ganga is generally less than 6mg/l which is generally due to the discharge of effluent of pulp and paper industries into the tributaries of Ganga (Ram Ganga and kali). Defecation in or nearby Ganga had lead to various diseases. Dips and the materials used for carrying the holy rituals like candles, flowers, etc further degrade the water quality of the river.



Fig. 2: Saints Performing Rituals During Kumbh 2013
Source: National Geographic Website

B. Public health & sanitation

Public health refers to all organized measures (whether private or public) to prevent communicable as well as non communicable diseases, promote health and prolong life among the population as whole. Its activities aim to provide conditions in which mass can stay healthy. The focus of public health intervention is to improve health and quality of life through prevention and treatment of diseases. A modern health practice requires multi disciplinary tea, of professionals and public health workers. In selecting a site for conducting an event, an analysis should be made of any potential health hazards in the area. Hazards vary as powerlines which could be brought down by storms, extremes of temperature, pollens, particulate matter, dust storms, pests, vectors etc, The basic problems associated with public health are:

- (1) People do not use toilets and just go off in the open in a mass event basically in developing Country like India.
- (2) Manual Scavenging still prevails. There are about 9000 sweepers whose job is to sweep the feces into the pits.
- (3) The absence of an effective public health network in densely populated areas has resulted in an extraordinary high disease burden.
- (4) Breeding of mosquitoes and growth of pathogens followed by odor nuisance are basic characteristics of human feces.
- (5) Various waterborne disease was spread in recent Maha Kumbh 2013 having following numbers and time relationship:

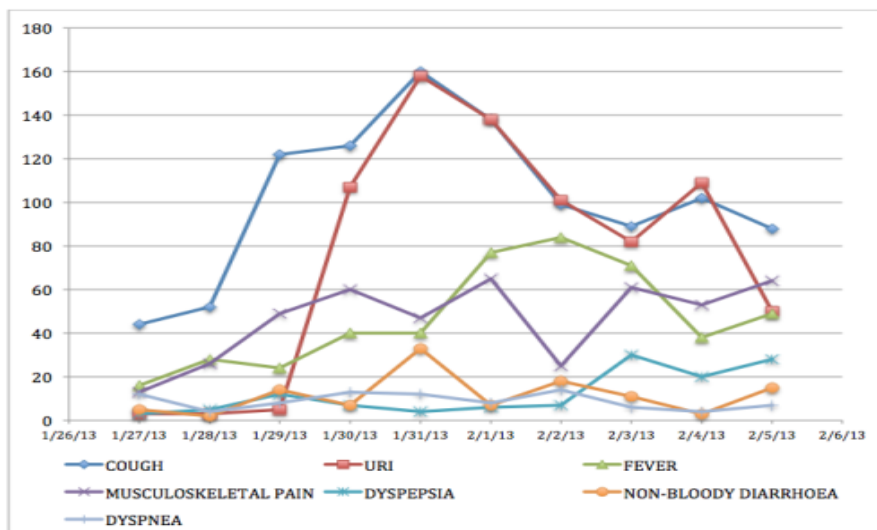


Fig. 3: Analysis of Diseases In Maha Kumbh 2013

C. Solid waste management

The basic important issues related to solid waste are: Damage to environment and innumerable health issues. The following were the common sites at Kumbh Mela:

- (1) Trash decomposes in open which attracted flies and rats.
- (2) Trash is dumped in non biodegradable materials like polyether bags thus exist forever contaminating soil and the groundwater.
- (3) Animals often die from ingesting harmful materials like plastic, polythene bags etc.
- (4) The solid waste is swept into Ganga and her tributaries where large amount of biodegradable waste takes a lot of oxygen from water to decompose leading to death, illness and BOD level at certain stretch stoop down to unfavorable limits.
- (5) Non biodegradable materials just float & block off the free flow of water and slowly leech their harmful chemicals into water itself.
- (6) Lack of education and awareness among the ordinary people.

D. Air Pollution

Air pollution has reached such a critical stage where it adversely affects every creature in the planet. The major sources are just stem out of the human activities like burning of fossil fuels, oil and natural gas, factories emitting smoke, automobiles, power houses, engines etc. These leads to the emissions of various gases like sulphur dioxide, nitrogen dioxide, carbon monoxide etc.

In Maha Khumbh the major cases of respiratory illness in approximately 15,000 patients were the smoke from choolahs used for cooking and shredding away the coldness of winter seasons. Also the priests and the rural people lit up their cigars, hookas and even cigarettes smoking was th common site in the gathering. Smoking is the huge issues in a public gathering in India. People just litter away the cigarettes, cigars, bidis etc. it raises the level of dangerous particulate matter linked to heart diseases, lung cancer, mouth and throat disorder etc.

Dust caused by large crowds or vehicle movements created annoyance and had caused problems for performers, patrons, and even the audience especially affecting the sensitive group of people.

E. Sewerage and wastewater

Event site without a sewerage system can never be approved for conducting any kind of event. A relevant health authority or agency takes control of the waste water treatment and disposal system from food stalls, ablution blocks, etc. The Ganga and Yamuna are already beyond the acceptable levels of pollution. The clean water is taken ruthlessly from the rivers and sewage, offerings, wastewater are directly dumped in these rivers prior any kind of treatment. It is clear that for the occasion of Maha Kumbh, the government has made huge efforts to clean the two rivers but with a little success. There were certain steps implemented:

- (1) More water is allowed to flow in the river. This is critical because without dilution there will no assimilative capacity in the river.
- (2) Allahabad has built sewerage treatment plants. But it do not have proper system to intercept the sewage and transport it to the treatment units. In this way plants are not used effectively. The system remains underutilized for its capacity.
- (3) The city tried the experiments with affordable ways of treating sewage by using bioremediation system
- (4) The government had put a strict regulation on the polluting industries like distilleries, tanneries etc for discharging their waste in Ganga. This step provided temporary relief against pollution.

IV. SUSTAINABLE EFFORTS

Mass events always require special management measures including non routine investments public resources in order to mange and fulfill short term benefits keeping in mind the sustainable utilization of resources. But the condition of rapid fulfillment of the demands like building up of whole new arena for Maha Kumbh causes imbalance. The challenges were to conduct the Maha Kumbh by overcoming the imbalance and properly managing the resources in best possible way.

This makes the role of technology to come into play. Few technological methods and techniques could make the event more successful:

A. Hazard identification and risk Assessment

The initial step of any mass event is to recognize any kind of hazard in the vicinity of the event area. The nature of hazard may vary according to location, but general natural hazards like floods, droughts, fire breakdown, wind storms, cyclones etc. can always occur without prior alarm. Many socio-political, economical and environmental prerequisites are considered for adequate risk assessment:

- (1) Adequate ventilation, illumination
- (2) No slip and leveled flooring
- (3) Safe access to equipments
- (4) Illuminated area layout to put in display at regular noticeable places
- (5) Fire extinguishers, fire blankets and fire alarms provision

- (6) A emergency team formation to counter the emergencies
- (7) Skilled personals to check and operate the equipments.

B. Public Awareness

Public awareness should be emphasized if the proper planning has to be implemented. Many efforts should be made by the organizing committees to aware the general public concerning the necessary information through media, newspapers, electronic display boards, hoardings etc. For the very grass root announcements, pamphlets can be distributed among the public. Billboards and hoardings can also be used for mentioning help lines and other information. DO's & DON'T's information could also be mentioned in the public hoardings other information containing document. The entry and exit gate's information is also must in case of stampede.

Apart from above facilities, a Help Desk can also be setup for addressing all type of queries, complaints, requests etc. having staffs of both the sexes.

C. Water Supply

Water is required for survival. It's the most important duty of an organizing committee to provide safe and portable drinking water. The water supplied should be uniform and the constant supply is enhanced by laying proper pipelines carrying drinking water from Water Board Supplies. The pressure should be sufficient enough to meet the required demands at all outlet points. Water should be readily accessible at all food premises, ablutions, toilets, laundry, designated standpipes, fire fighting, first aid posts and all the other required area. But it is strongly be available by the recognized water authority supply.

If event is arranged in a smaller scale, tanker facilities could be provided. But for the large scale events lasting from few days to months, spigots could be installed along with the water purification facility.

The quality of water for drinking and washing should be different as to retain the good quality water resources so that it cannot be wasted for washing and cleaning purposes. But the non potable supplies are to be clearly identified as non drinking purposes. The recycled water should be used for cleaning, gardening, washing, flushing etc. The recycled water can also be used for fire fighting, dust suppression and even sometimes for luxuries activities like spraying fans, coolers etc.

D. Health & Sanitation

Public health is the most significant area of concern. There are numerous ways to create a clean environment for the mass event like regular fogging for pest control, cleaning of stagnant water pits to control the breeding of mosquitoes and flies, hospitals and health centres should be setup with qualifies doctors and nurses to take proper care of the patients, the medicines and other important equipments for the treatments should be present, an ambulance facility should be provided in case of emergency.

There are various methods by which the sanitation of the mass gathering event can be progressed. By technological advancements, different kind of toilets have been developed which can temporarily provides effective and clean sanitation facility:

1) Ecological Sanitation

These are also known as eco-san, it is a form of sanitation that usually involves urine diversion and the recycling of water and nutrients contained within human wastes back into local environment. Ecosan is based on an overall economical, environmental, sustainable wastewater management system which has been developed to fulfill the need of users as well the local conditions. It provides a flexible framework where the centralized elements can be combined with the decentralized ones, waterborne with dry sanitation high-tech and low-tech etc. by considering a much larger range of options, optimal and economic solutions can be developed for particular situation.

The basic advantages of ecosan are:

- (1) Improvement of health by minimizing the introduction of pathogens from human excreta into the water cycle.
- (2) Promotion of safe, hygienic recovery and use of nutrients, organics, trace elements, water and energy.
- (3) Preservation of soil fertility
- (4) Contribution to conservation of resources through lower water consumption, substitution of mineral fertilizers and minimization of water pollution
- (5) Improvement of agricultural productivity and food security
- (6) Preference for modular, decentralized partial flow-systems for more appropriate cost-efficient solutions adapted to the local situations
- (7) Promotion of holistic, interdisciplinary approach
- (8) Material flow cycle instead of disposal of valuable resources

2) Portable toilets

These are simple portable enclosures containing a chemical disinfectant instead of water. It is a temporary toilet used for large gatherings with minimal uses of water. Most portable toilets have black open front U-shaped toilet seats with a cover. They are often constructed out of light weight molded plastics.

– Advantages:

Portable toilets have several significant benefits mostly related to their portability; it can be drained, cleaned, disinfected and deodorized on a regular basis. It's cheaper as they are not plumbed, they don't clog. An average portable toilet is able to hold enough sewage for 10 people during the course of a 40 hour work week before the hold reaches unsanitary conditions.

– Disadvantages:

Since they are not plumbed, they keep waste inside the bathroom; this leads to smelly conditions, if not cleaned properly and periodically. This is also considered as an eyesore in most communities, some of which prohibit its use without permission of municipalities.

3) Bio-digester Toilets

These are based on anaerobic biodegradation of organic wastes by unique microbial consortium was conceived and developed by DRDO for defense forces deployed at higher altitudes areas. In this system, fecal matter is decomposed in bits and converted into water and little carbon dioxide & methane gas. It is totally maintenance free system and doesn't require any sewerage system.

a) Working:

A consortium of anaerobic bacteria has been formulated and adopted to work at temperature as low as 5 degree C. this is the component t which acts as inoculums to the bio digesters and converts organic matter to methane and carbon dioxide. The anaerobic processes in-activates the pathogens responsible for water borne diseases. Bio digester serves as reaction vessel for bio methanation and provides the anaerobic conditions & required temperature. The optimum temperature is maintained by microbial heat, insulation of reactor and solar heating.

b) Biological Process in Digesters:

Fecal matter is composed of carbohydrates, fats and proteins. A four step mechanism of decomposition is:

- (1) Hydrolysis: simple sugars, amino acids and fatty acids are break down into simpler compounds
- (2) Acidogenesis: formation of carbonic acid, alcohols, hydrogen and water
- (3) Acetogenesis: acetic acid, hydrogen and carbon dioxide is formed
- (4) Methanogenesis: conversion of all compounds to methane and carbon dioxide and neutral water (pH: 6-9)

The final product is left with no solids, only gases and water.

– Disadvantages:

Expert and skilled persons are required for design, construction, operation & maintenance depending upon the scale. Reuse of produced energy needs to be established. Sulphurous compounds leads to odour.

4) Pit Toilets

It is a dry system which collects human excrement in a large container or a pit or a simple slit trench with ventilation facility. They are basically used in rural areas or wilderness which lacks all the basic requirements. The waste pits in some cases may be large enough that the reduction in mass of the contained waste products by ongoing process of decomposition allows the pit to be more or less permanent. In other cases, when pit becomes too full, it may be emptied or hole may be covered by soil and rebuilt over a new pit.

The anaerobic decomposition takes place along with emissions of the gases and volume of sludge is reduced. There will also be die off bacteria and viruses during storage and as the water percolate through the soil so the pollution of groundwater occurs. Control of odour and insects can be achieved by a vented pit.

The period of emptying depends on the size of pit and its usage. it is desirable to store at least one year of sludge production.

5) Dehydrating Toilets

It is traditional indoor toilet in which urine is diverted to a collection tank or a soak pit under the toilet vault or outside the toilet and feces drop into one of the two vaults, below the toilet seat. When one vault is full, it is then sealed and another one is used. Materials like ash or soil or a mixture of sawdust/lime or soil/lime are added for defecation. The dry material assists the desiccation process and raises pH which aids in pathogen reduction.

6) Composting Toilet

It is a dry toilet that uses predominantly aerobic processing system that treats excreta with no water or low flush water via composting or managed aerobic decomposition. It can be used as an alternative to flush toilets in water scarce locations or where there is no provision of waste water treatment facility available or to capture the nutrients in human excreta. The decomposition of human sludge takes place in presence of air i.e. aerobic decomposition in an elevated latrine. They are used in many of the roadside facilities and public toilets.

Air can be introduced through an operating to pass through sludge and exit through the vent. Excess liquid is allowed to drain for collection or evaporation. The human excrement is normally mixed with sawdust, coconut coir or peat moss to support aerobic processing, absorb liquids and to reduce odor. The decomposition process is generally faster than anaerobic decomposition used in wet sewage treatment supply like septic tank.

E. Solid Waste Management

There are various measures by which solid waste can be collected or recycled or reused:

- (1) Collection of waste generated at source will not only reduce the burden on the Municipal Corporation but also it saves time and the cost employed on segregation. The authorities should install adequate collection bins at major routes and location and these bins should be emptied at frequent intervals to maintain a clean and hygienic environment.
- (2) Availability of adequate number of sweepers and waste disposal units should cover the entire event area. Containers or garbage bins should be placed in large open areas as approximately 4 per acres.
- (3) Reduce, reuse and recycle should be the main motto of any event to take place by preventing harmful biodegradable materials such as plastics, polythene from entering nearby water body and providing alternative options for processing.
- (4) Exploring creative and innovative ways in which solid waste can be used as a source of fuel in form of by products like biogas. Also an alternative option of eco packaging could be employed.
- (5) Waste disposal or landfill sites should be identified well in advance if required in case of urgency. The efficient facility to transport the waste from event site to dhalao and from dhalao to landfill or waste processing site should be ensured.
- (6) A maintenance schedule should be set up to monitoring the work efficiency of the sweepers and required workers.
- (7) The placement of containers and all the treatment units depends upon the size of event and the crowd gathered. In high traffic areas such as spectator stands, seating toilets, handwashing area, activity areas along walkways from food booth etc. the frequent and easy removal facility should be present for proper management.

F. Fire Services

Fire extinguishers should be installed at all places where there are maximum chances of frequent gatherings. Fire brigade should reach on time merely by fire alarms. In case of emergencies, a team should be dedicated to take control of the situation. The recycled water should be used for fire breakdown. The pantry, kitchens with LPG etc should be taken into special consideration. The combustible materials should be limited from the event area. The security personnel can ensure such strictness. The fire crackers, bits of cigarettes, smoking choolahs, hookas etc should be limited as far as possible.

G. Air Pollution Control

Air borne pollution is the important issue to be taken into consideration. The transfer or movement of crowd, vehicles etc leads to generation of dust. Even in high wind areas, there is dust all around. Fugitive dust particles adversely affect the workforce health and can interfere with the food and water quality.

Concern regarding smoking is also increasing day by day. It kills not only the smokers but it enhances the passive smoking effects. Smoking affects major body parts like heart, lungs, skin, bones, stomach, mouth, throat, reproduction and fertility as well. Public health can be accounted by implementing smoke free policies in major gathering areas. Benefits would include decreased cleaning costs, better public image, the reduction of second hand smoke drifting onto major public areas and reduce nuisance impact for general public as well.

Dust suppression could be possible as:

- (1) Applying gravel to a dirt road surface.
- (2) Sealing unpaved roads with pavement or impermeable materials.
- (3) Reducing exposed ground by maintaining vegetation.

H. Waste Water Control

All the liquid been generated from washing, cleaning or cooking activities should be disposed off efficiently in an approved manner. For a small scale function, a provision of pit can be used but for a large scale even proper treatment units should be employed. The organizers should have sewage pump out truck available on site at all time or on standby status for servicing the portable or other toilets.

In order to control insanitary conditions, sewage and waste water collection, treatment and disposal systems must be constructed and operated to prevent contamination of food or water supply. For insufficient sewerage system, the relevant health authorities should be contacted to determine the requirements for all type of sewage and waste disposal.

I. Noise Pollution Control

For public events especially outdoors there is a need to apply guidelines for the maximum noise level typically allowed at the boundary from the nearest private dwelling in the vicinity of the source of noise pollution. Allow 4-6 weeks time for an application to be processed so that there is time left for comments and objections by affected public and the relevant stockholders in the area. Information regarding time, location, duration of the event along with the sound equipment functionality should be provided by the organizing committee to the relevant authorities.

It may need to consult an acoustic consultant on the impact of loud noise from the event adjoining property and stock in rural area.

J. Mass Casualty Management

Mass casualties are characterized by the severity, diversity of injuries, availability of aid, time, etc. An effective mass management plan requires an establishment of baseline that is an assessment of current system's capacity against which planned changes can eventually be managed. A second key initiative should be taken to provide detailed understanding of hazards and risks. Emphasis should be paid on gathering retrospective data of previous similar events. This should not be limited to large scale disasters but also to the smaller ones like major traffic congestions, industrial incidents, chaos, building collapses and so on. It should not include a forward looking component assessing risk in future from the elements such as unsafe urban development or change in weather patterns.

This also necessitates the development of training modules and guidelines, monitoring and surveillance and early warning, stockpiling, collaboration with NGO's and corporate sectors as partners and thus finally developing a culture of community preparedness. It requires the paradigm change from the application of unlimited resources for the greatest good of each individual patient to the allocation of limited resources for the greatest good of the greatest number of casualties.

V. CONCLUSION

Kumbh Mela, the name itself projects the belief and vibrant traditions in India. With the religious sentiments, and adamant faith of thousands of people encasing the event, the Mela itself is a binder of different sects of people. The effects of Kumbh Mela had grown significantly over the past decades, its ability to draw the mass all around the world has grown significantly.

In the presence of proper planning, a successful event can be guaranteed. The Kumbh had implemented many modern techniques like CCTV, GPS, advanced water treatment units etc, their total efforts is no doubt incredible but there were few loop holes in the planning section. There was lack of awareness among the masses regarding the use of advanced technologies been implemented, inadequacy of skilled personals for operation & maintenance on time.

This paper has covers the overall effects on environment due to conductance of Maha Kumbh by data analysis and few technological advancements which may prove useful to other events over the world like Hajj, Rath Yatra, Durga Puja etc, other fairs and festivals, sporting events, conferences etc. The means of sustainable development can be achieved by fulfilling the basic needs of the masses.

A mass gathering like Kumbh Mela often brings together organization that has never worked together. It is very essential to establish excellent coordination and communication system between different organizing groups. A proper understanding, planning, timely risk assessment and proper resource utilization can make any event a successful one.

REFERENCES

- [1] Participation In Mass Gatherings Can Benefit Well Being: Longitudinal And Control Data From North Indian Hindu Pilgrimage Event: PLOS, October 2012.
- [2] "Millions Of Hindus Take To The Ganges At Maha Kumbh Mela" Reuters: Retrieved 15 January 2013
- [3] Official Site of Maha Kumbh Mela 2013: Powered by Omni-NET through UPLC
- [4] Sinhashta Khumbh Mela- Nashik is a pilgrim City, Maharashtra, Golden Research Thoughts, February 2014
- [5] Waste to energy- A case study of Eluru, Andhra Pradesh: 2010 by J. Sudhir Kumar, K. Venkata Subbaiah, Prasado Rao.P.V.V.
- [6] Green. W.G. May 2000: Mass Casualty Incident Management
- [7] Guidelines For Concerts, Events And Organized Gatherings, December 2009, Australia by Government of Western Australia
- [8] Mass gathering and Public health by Agis D. Tsouros and Panos A. Efstathiou: 2004
- [9] Kishore.S. July 2012: Toronto Host City: Mass Gathering Risks And Perception Applied Research And Innovation Center.
- [10] Communicable Disease Alert And Response For Mass Gatherings: Retrieved 04-11-2010
- [11] Harvard Global Health Institute 2012: The Kumbh Mela Annotated Bibliography Of Selected Resources Part 2
- [12] Indian Guidelines For Sector Reform And Successful Public-Private Relationships: January 2004
- [13] "Mass Gatherings And Public Health: 2007. Retrieved 04-11-2010
- [14] Guidelines For Management Of Public Health And Safety At Public Events: 2003, Department Of Health, Government Of South Australia
- [15] "Public Health Preparedness for Mass Gathering" Retrieved 04-11-2010
- [16] India, Urban Sanitation and Toilet Challenges, RTI press, September 2013.

This article is part of the supplement: [Thirteenth International Conference on Bioinformatics \(InCoB2014\): Computational Biology](#)

Research

Fragment based group QSAR and molecular dynamics mechanistic studies on arylthioindole derivatives targeting the α - β interfacial site of human tubulin

Chetna Tyagi¹, Ankita Gupta², Sukriti Goyal³, Jaspreet Kaur Dhanjal¹ and Abhinav Grover^{1*}

*Corresponding author: Abhinav Grover abhinavgr@gmail.com

Author Affiliations

¹School of Biotechnology, Jawaharlal Nehru University, New Delhi, India - 110067

²Department of Biotechnology, Delhi Technological University, New Delhi, India -110042

³Apaji Institute of Mathematics & Applied Computer Technology, Banasthali University, Tonk, Rajasthan, India - 304022

For all author emails, please [log on](#).

BMC Genomics 2014, **15**(Suppl 9):S3 doi:10.1186/1471-2164-15-S9-S3

The electronic version of this article is the complete one and can be found online at: <http://www.biomedcentral.com/1471-2164/15/S9/S3>

Published: 8 December 2014

© 2014 Tyagi et al.; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Formula display: ☒ **MathJax**

Abstract

Background

A number of microtubule disassembly blocking agents and inhibitors of tubulin polymerization have been elements of great interest in anti-cancer therapy, some of them even entering into the clinical trials. One such class of tubulin assembly inhibitors is of arylthioindole derivatives which results in effective microtubule disorganization responsible for cell apoptosis by interacting with the colchicine binding site of the β -unit of tubulin close to the interface with the α unit. We modelled the human tubulin β unit (chain D) protein and performed docking studies to elucidate the detailed binding mode of actions associated with their inhibition. The activity enhancing structural aspects were evaluated using a fragment-based Group QSAR (G-QSAR) model and was validated statistically to determine its robustness. A combinatorial library was generated keeping the arylthioindole moiety as the template and their activities were predicted.

Results

The G-QSAR model obtained was statistically significant with r^2 value of 0.85, cross validated correlation coefficient q^2 value of 0.71 and pred_r^2 (r^2 value for test set) value of 0.89. A high F test value of 65.76 suggests robustness of the model. Screening of the combinatorial library on

the basis of predicted activity values yielded two compounds HPI (predicted pIC50 = 6.042) and MSI (predicted pIC50 = 6.001) whose interactions with the D chain of modelled human tubulin protein were evaluated in detail. A toxicity evaluation resulted in MSI being less toxic in comparison to HPI.

Conclusions

The study provides an insight into the crucial structural requirements and the necessary chemical substitutions required for the arylthioindole moiety to exhibit enhanced inhibitory activity against human tubulin. The two reported compounds HPI and MSI showed promising anti cancer activities and thus can be considered as potent leads against cancer. The toxicity evaluation of these compounds suggests that MSI is a promising therapeutic candidate. This study provided another stepping stone in the direction of evaluating tubulin inhibition and microtubule disassembly degeneration as viable targets for development of novel therapeutics against cancer.

Keywords:

Cancer; arylthioindole; QSAR; tubulin; G-QSAR; inhibitor

Background

Tubulin inhibition has been considered a viable avenue for drug development in cancer management for a very long time [1-4]. It plays a major role in the formation of microtubule assembly. Microtubules are polar cytoskeletal filaments that either takes part in the formation of mitotic spindle and interphase networks or more complex formations like ciliary axoneme, centrioles and basal bodies [5-7]. Their disorientation, either through inhibition of tubulin polymerization or by blocking microtubule assembly, leads to metaphase arrest of cell division [8-10]. Colchicine, the most commonly used metaphase arrest agent, has a similar mode of action [11]. Apart from these, combretastatin A-4 and the *Catharanthus* alkaloids vincristine and vinblastine are also tubulin assembly inhibitors [12,13]. Combretastatin A-4 phosphate was found to block blood flow in capillaries supporting cancerous cells by completely disrupting the endothelial cell integrity leading to rapid tumor cell death [14,15]. These inhibitors lead to microtubule disorganization and renders cell apoptosis [16]. However, the use of these previously identified tubulin inhibitors were restricted on account of excessive toxicity, drug resistance and scarce bioavailability [17-20]. Thus, development of novel tubulin assembly inhibitors is the need of the hour.

A variety of chemical compounds are known to bind to the α - β interfacial site of tubulin. These chemical compounds can be grouped under following classes (a) group targeting the luminal site of α subunit i.e. taxol (b) group stabilizing microtubule polymerization i.e. vinblastine and (c) group targeting α - β interfacial site of tubulin that leads to cell apoptosis i.e. colchicine [21-23]. We are interested in the colchicine-like group of compounds for cancer management. The present study is based on a class of tubulin-inhibitors known as arylthioindoles that bind to the colchicine-binding site of β -tubulin close to the interface with α -tubulin. All the previous studies undertaken in this respect involved *Bos taurus* tubulin protein assembly comprising of chains A, B, C, D and E of which A and C belong to the α unit and B and D belong to the β unit [24,25]. Many previously known tubulin inhibitors consisted of the indole nucleus in the core structure and hence are touted to be one of the most potent compounds against tubulin

polymerization [17,26]. Arylthioindoles were also found to be potent inhibitors of the growth of MCF-7 human breast carcinoma cells [19].

Development of accurate and time effective drug discovery techniques is the need of the hour to propagate search for novel anti-tumorals. Exploiting one of the recent and innovative approaches known as fragment based group quantitative structure activity relationship (G-QSAR) [27], the relationship between different molecular fragments and their biological activity can be correlated and studied in detail giving site-specific clues for modification [28]. Such modifications in terms of substituents added or removed lead to activity enhancement. The knowledge of such modifications is based on various molecular descriptors calculated and used for G QSAR model construction. Various such studies have been reported and have proved to be very useful [29-31], many of them to discover cancer therapeutics [32]. These descriptors are calculated for various fragments defined by the user. The optimal subset of descriptors is chosen by any one of the variable selection methods which are most likely to describe all the physicochemical properties of the congeneric series required for their biological activity. Thus, it gives a better idea about which substitution site should be populated with which particular substituent for activity enhancement[33].

In this study, we search for tubulin inhibitors having a similar binding mode as that of colchicine at the α - β interfacial site. Arylthioindole moiety is known to be a potent anti-tubulin agent and has been studied very often for its anti-cancer properties but drug toxicity and less bioavailability were the problems encountered [34]. In order to exploit this avenue further, we created a robust, accurate and predictive G-QSAR model to enhance our understanding of arylthioindole derivatives as anti-cancer compounds in terms of structural requirements needed for drug development. Based on the G-QSAR model, we identified novel therapeutic compounds with improved tubulin assembly inhibition and potent anticancer activities. The compounds were validated for their interactive properties with the colchicine binding site of tubulin by docking analysis. The resultant top two compounds were also evaluated for their absorption, distribution, metabolism, excretion and toxicity (ADMET) properties.

Materials and methods

Compound dataset for model development

In this study, a congeneric series of 42 tubulin inhibitors belonging to the arylthioindole class of compounds [20,35] were selected for G-QSAR model development. Due to higher root-mean-square-deviation (RMSD) values, 6 compounds (6b, 15, 20b, 24, 28 and 41b in Additional file 1) were rejected and the model was built using 36 arylthioindole derivatives. The 2D structures were drawn using Marwin Sketch [36]. They were converted to 3D by *Vlife Engine platform* of VLifeMDS and later energy minimized using the force field batch minimization utility with default parameters[37]. These optimized compounds were finally used for G-QSAR model development. The template is used as a structural moiety common to all the compounds of the congeneric dataset with the substitution sites marked by unknown (dummy) atoms. In this case there were six substitution sites, depicted by R1 to R6 in the template.

Additional file 1. Table S1 - Structures and anti-cancer activity values of arylthioindole derivatives used in this study.

Format: DOCX Size: 1.8MB [Download file](#) [OPEN DATA](#)

Computation of Molecular Descriptors

After selecting the congeneric set of molecules and the template for model development, physicochemical descriptors were calculated using the *calculate descriptors* dialog of the G-QSAR module in VLife MDS. The software provides a large number of molecular descriptors belonging to the 2-dimensional and 3-dimensional classes. In case of Group QSAR, 2D descriptors are selected. All descriptors were chosen except dipole moment, electrostatic, semi-empirical and hydrophobicity as they are 3D descriptors and Information Theory-based descriptors. In G-QSAR, a variety of descriptors along with alignment independent descriptors for the fragments were calculated. Of the total 1027 descriptors calculated for all the six substitution sites, 298 were selected and the invariable descriptors were removed. Invariable descriptors are those which have same quantitative value for each data-point and thus should be discarded for their inefficacy in G-QSAR model development.

G-QSAR model development

For data selection using the *advanced data selection wizard*, the training and test set compounds were chosen after selecting the activity pIC₅₀ as dependent variable and all the calculated descriptors as independent variables. The test set including the compounds 10, 14, 18, 21, 22, 27b, 29, 30, 31, 35b (available in Additional file 1) was selected using *random selection method* with 80% compounds in the training set. After calculating the unicum statistics for the selected training and test set molecules, the *stepwise-forward variable selection* method along with PLS (Partial Least Square) [38] as the regression method for building the model was chosen through the *variable selection and model building wizard*. Keeping the cross-correlation limit set at 0.5, Ftest In at 4.0, Ftest Out at 3.0, term selection criteria being r^2 , variance cut-off at 0.1 with auto-scaling, the model was built.

Model validation

Many statistical parameters like n (number of compounds in regression), k (number of variables), degree of freedom, optimum component (number of optimum PLS components in the model), r^2 (squared correlation coefficient), F-test (Fischer's value), q^2 (cross-validated correlation coefficient), pred_r^2 (r^2 for external test set), Z score (randomisation test), $\text{best_ran_}q^2$ (highest r^2 value in the randomisation test) and $\text{best_ran_}r^2$ (highest r^2 value in the randomisation test) need to be taken into account to consider the model as a robust one. For a model to be statistically significant, the following conditions should be satisfied: r^2 , $q^2 > 0.6$ and $\text{pred}_r^2 > 0.5$. Since, F-test gives an idea of the chances of failure of the model, a value greater than 30 is considered to be good. On the other hand, low standard error values establish absolute quality of the model.

Internal and external validation

For internal validation using leave-one-out method, the cross-validated coefficient, q^2 is calculated using the given equation:

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2}$$

where y_i and $\hat{y}_i$ are the actual and predicted activities of the i th ($i = 1-26$ in Additional File 1) molecule in the training set, respectively, and y_{mean} is the average activity of all the molecules in the training set.

For external validation, the pred_r^2 value that gives an account of the statistical correlation between predicted and actual activities of the test set compounds was calculated as follows:

$$\text{pred}_r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2}$$

where y_i and $\hat{y}_i$ are the actual and predicted activities of the i th molecule ($i = 27-36$ in Additional File 1) in the test set, respectively, and y_{mean} is the average activity of all the molecules in the training set.

To avoid the risk of chance correlation, Y randomisation test was carried out by comparing the resultant linear model with those derived from random data set [39]. Various models were built on random datasets generated by rearranging the molecules in the training set so as to compare them with the obtained G-QSAR model on the basis of Z-score. A Z-score value is calculated by the following formula:

$$Z_{\text{score}} = (h - \mu) / \sigma$$

where h is the q^2 value calculated for the actual data set, μ is the average q^2 and σ is the standard deviation calculated for various models built on different random data sets.

Combinatorial library generation

A combinatorial library was generated using the *LeadGrow* module of VLife MDS using the template used before and a list of substituent chemical groups to be added to only three substitution sites R4, R5 and R6 based on the calculated descriptors. A library of 4257 compounds was created by substituting electronegative and electropositive atoms, alkyl groups, cyclic groups, aromatic rings and other bulky groups. Their biological activities were predicted using the G-QSAR model obtained.

Protein modelling and ligand docking

In order to investigate the extent of interactions involved between the most active compounds obtained from the combinatorial library, these compounds were docked to the homology modelled B-chain of tubulin protein using Schrodinger's Glide Module [40-42]. The homology model was obtained using Modeller 9.11 [43-46] taking *Bos taurus* tubulin B-chain [PDB ID: 1SA0] as the template [12]. The resultant homology model was validated using online PSVS suite [47] which comprises a number of software namely, DSSP [48,49], pdbStat (5.9), AutoAssign [50], RPF Analysis [51,52], PDB validation server, Verify3D (1.0) [53], ProsaII (2003) [54], PROCHECK (3.5.4) [55]. Along with these, Molprobit [56] programs and other software are also present.

The tubulin homology model was optimized for docking using Schrodinger's protein preparation wizard [57]. A total of 48 compounds were docked targeting the α - β interfacial site of tubulin heterodimer. Glide works by creating a cubic grid (20 Å side) around the user-specified critical residues and directing the approaching ligand at the specific site. Extra-precision (XP) docking was used to screen compounds with high binding affinity for tubulin. XP docking serves the purpose of correlating good poses with good scores and discarding the false positives [58]. Screening of compounds by docking for potential candidature as therapeutics has been a popular avenue in computational drug design [59-61]. The various interactions involved between the highly active compounds were evaluated using Ligplot [62].

Molecular Dynamics simulations of the modelled protein and docked complexes

To obtain energetically stable conformation of the modelled protein target and to get an insight into the stability of protein-ligand complexes, molecular dynamic simulations were carried out using Desmond Molecular Dynamics module [63-65] of Schrodinger Maestro by applying optimized Potentials for Liquid Simulations (OPLS) all-atom force field 2005. Prepared protein-ligand complexes were solvated with TIP4P water model in a triclinic periodic boundary box for MD simulations. To prevent the direct interaction of protein complex with its own periodic

image a boundary box is created and distance between protein complex and box wall is kept at 10 Å. Steepest descent method was used to minimize the energy of the prepared structures for a maximum of 5000 steps till a threshold of 25 kcal/mol/Å is achieved, which was then followed by Low-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newtonian minimizer until a convergence threshold of 1 kcal/mol/Å was met. Other parameters were kept as default for system equilibration. MD simulations were carried out for 10 ns at a constant temperature of 300 K, pressure 1 atm and at time step of 2 femtoseconds (fs). Long range electrostatic interactions were calculated using smooth particle mesh Ewald method [66] which was occurring during the MD simulations and coulombic short range interaction was calculated using a cut-off scheme, with a cut-off radius of 9 Å for calculation. The protein ligand complex was prepared for MD simulations using the above mentioned parameters. MD Simulation was then carried for a time period of 10 ns.

The root mean square deviation values (RMSD) for two top scoring ligands were calculated for the entire simulations trajectory with reference to their respective first frames. Radius of gyration (ROG) analyses was carried out for all the frames MD simulation of Ikb kinase beta and ligand complex.

Calculation of ADMET properties

The toxicity of final two compounds was evaluated using the Quikprop module [67] of Schrodinger suite which predicts various molecular properties and also provides ranges for comparing these properties with 95% of already available drugs. The analysis was followed by a toxicity analysis using an online webserver, admetSAR [68] which reads the smiles format and results in a number of ADMET values. AdmetSAR is a knowledge based tool comprising of ADMET related properties taken from large literatures which are further used to predict properties of unknown compounds.

Results and discussion

G-QSAR model developed for arylthioindole derivatives targeting α - β interfacial site of tubulin

We report a fragment based group QSAR model based on a congeneric series of arylthioindole moiety as the template targeting the α - β interfacial site of tubulin. Colchicine is also known to bind at the same site, though arylthioindoles were found to have much higher activity than colchicine against tubulin polymerization. Inhibition of tubulin polymerization can be a potent deterrent to cell division and hence can be seen as an avenue in developing cancer therapeutics. The tubulin D chain (β unit) structure was modelled based on the close homologous protein belonging to *Bos taurus*.

The G-QSAR model obtained can be represented by the following linear equation:

$$pIC_{50} = (0.536788 \times R4 - H \text{ Acceptor Count}) + (-0.412083 \times R6 - \chi^2) + (-0.538397 \times R5 - \text{slogp}) + (-0.365145 \times R6 - \text{NitrogensCount}) + (0.00256203 \times R6 - \text{Molecular Weight}) + 4.97922(1)$$

where pIC₅₀ is the negative logarithm of IC₅₀ activity values and was taken as the dependent variable in the model development. A total of 6 substitution sites were marked on the arylthioindole moiety which became the basis of fragmenting the derivatives for model development, of which 3 sites namely, R4, R5 and R6 were taken into account. All 2D molecular descriptors are calculated for each fragment generated. The five molecular descriptors selected on which the model is based are **H Acceptor Count** for the substitution site R4, **slogp** for the

site R5 and **chi2**, **NitrogensCount** and **Molecular Weight** for the site R6. The multiplied numerical terms associated with the descriptors are the respective coefficients and the last numerical term is the regression constant. Using the model generated, if the values of these descriptors are known for novel compounds, the biological activity in terms of pIC₅₀ can be determined. Out of these, only R4-H Acceptor Count and R6-Molecular Weight are positively contributing to the biological activity and the rest are negatively contributing. To determine the crucial structural features required for good activity, these descriptors are needed.

R4-H Acceptor Count: This descriptor signifies the number of hydrogen bond acceptor atom and corresponds to the R4 site which was originally substituted by either hydrogen atom or oxymethyl (-OCH₃) group in the series taken for model development. As it is a positive contributor, the presence of hydrogen bond acceptor atoms renders high biological activity to the molecule. The percentage of its contribution in enhancing the biological activity was 33.877 %.

R6-chi2: In theoretical terms, this descriptor stands for a retention index of second order which is derived directly from gradient retention times. The retention index of a chemical compound is the normalised retention time which is done to convert these into system independent constants in chromatographic analysis. In chromatography, retention time is the time required for a solute to migrate or elute from the column and this property depends upon the physical properties and behaviour of the molecule. The descriptor was calculated for the substitution site R6 present on the indole core and was mainly substituted by groups like hydrogen atom, halogens like Cl, I, Br, F and other groups like NH₂, NO₂, CH₃, OCH₃, OCH₂CH₃, OCH(CH₃)₃ etc. This descriptor contributed negatively to the biological activity to an extent of 23.018 %.

R5-slogp: This descriptor also contributed negatively to the biological activity up to an extent of 17.542% and corresponds to the substitution site R5. Slogp describes the value of log of octanol/water partition coefficient (including implicit hydrogens). It is an atomic property model and calculates logP value from the correct protonation state structure. The R5 site was mainly substituted with hydrogen atoms or oxymethyl (-OCH₃) groups.

R6-NitrogensCount: As the name suggests, this descriptor stands for the number of nitrogen atoms present in the compound. This physicochemical descriptor contributes negatively to the activity of arylthioindole derivatives by 15.952%. This suggests that a highly electronegative group, in this case, specifically nitrogen is deterrent for the molecule's activity if present at the R6 position. It should avoid a substitution involving nitrogen atoms.

R6-Molecular Weight: This descriptor signifies molecular weight of the compound and is positively contributing in the biological activities by 9.611%. This indicates the importance of substituents with higher molecular weight present at the R6 substitution site.

Model evaluation and validation of G-QSAR model developed for arylthioindole derived compounds

Unicolumn statistics for the chosen training and test set

For model evaluation and validation the complete dataset is divided into training set and test set. The developed model is validated by predicting the inhibitory activity (in terms of pIC₅₀) of the training set (known as internal validation) and the test set (known as external validation). The test set chosen can always be evaluated beforehand using the unicolumn statistics (Table 1). The unicolumn statistics can be interpreted in terms of the maximum and minimum of training and test set. The min of test set should be equal or more than the min of training set

and the max of the test set should be equal or less than the max of training set. This data is in absolute compliance with the conditions mentioned and shows that the test set is interpolative (derived within the min-max range of the training set). The relative difference of mean and point density distribution (along mean) of the two sets can be derived from the average and standard deviation. In this case, as the average value in the test set is slightly higher than the training set, the presence of relatively more active molecules as compared to the inactive ones is indicated. Also, a higher standard deviation for the training set indicates that training set constitutes widely distributed activity of the molecules as compared to the test set.

Table 1. Unicolumn statistics for the training and test set compounds.

Validation of the final G-QSAR model

The G-QSAR model is evaluated on the basis of certain statistical parameters for both internal and external validation. The number of compounds in the training set was specified by N which is 26. Considering the correlation coefficient, r^2 (0.8512), cross-validated correlation coefficient q^2 (0.7175), pred_r^2 (0.8968), low standard error value, r^2_{se} (0.1363), q^2_{se} (0.1878) and $\text{pred}_r^2_{se}$ (0.1127), the model can be stated to be a robust one. Along with this, a high F-test value (65.7699) implied that the model is 99.999 % statistically valid with less than 1 in 10000 chance of failure.

Also with this, the randomization test shows confidence of 100% (Alpha Rand $r^2=0.00000$) that the generated model is not random and hence is chosen as the G-QSAR model. Other important statistical parameters have been determined and Z-score is highlighted to emphasize its importance in QSAR model validation (Table 2). The values of selected descriptors for each compound in the dataset have been provided (Additional file 2). The Z score gives an idea about how far away is the observed value from the mean. A Zscore_r^2 of 6.48008, Zscore_q^2 of 4.73992 and $\text{Zscore}_{\text{pred}_r^2}$ of 1.70900 statistically validates the significance of the obtained G-QSAR model. In order to get a better idea of the validity of the model, p values were determined for each correlation coefficient using the corresponding Z scores. The P value_{r2} and P value_{q2} of less than 0.0001 tells that the hypothesis is statistically significant. The P value_{pred_r2} of 0.0875 is although, not quite statistically significant.

Table 2. The statistical parameters calculated for developed G-QSAR model.

Additional file 2. Table S2 - Calculated descriptor values and predicted activity of arylthioindole derivatives.

Format: DOCX Size: 14KB [Download file](#) [OPEN DATA](#)

The robustness of the model is better understood through the linear graphical representation between actual and predicted activities of the 36 compounds along with the contribution plot for each descriptor (Figure 1). The linear graphical representation shows the extent of variation between the actual and predicted activities of the congeneric set. The larger the distance of training and test set points from the regression line, more is the difference between the actual and the predicted activity values. The contribution of each descriptor specifies the properties that should be present in the drug lead for enhancing its inhibitory activity. Presence of descriptors with positive contribution increases its inhibitory activity while descriptors with negative contribution decrease the same. Radar plots for training and test sets are given (Figure 2). The radar graphs depict the difference in the actual and predicted activities for the training and the test sets separately by the extent of overlap between blue (actual activity) and

red (predicted activity) lines. The radar plot for training set represents a good r^2 value if the two lines show a good overlap while for the test set a good overlap represents high pred_r^2 value.



Figure 1. (a) Fitness Plot of the G-QSAR model obtained for actual and predicted activities. (b) Contribution Plot of the selected descriptors R4-H Acceptor Count, R6-chi2, R5-slogp, R6-NitrogensCount and R6-Mol.Wt.

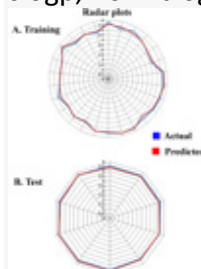


Figure 2. Radar graphs plotting the predicted and actual activities for (a) Training set and (b) Test set.

Activity prediction of combinatorial library generated using arylthioindole moiety as template

A combinatorial library was constructed resulting in 4257 compounds after substituting the R4, R5 and R6 sites with various chemical groups for which the descriptors were calculated. The three sites R4, R5 and R6 were populated by single atoms like N, C, O and halogens like Cl, F, Br and I, by alkyl groups and aromatic rings like phenoxy, phenyl, furan, pyrrole, pyridine, imidazole, 2-thiophene etc along with alkenes, acids, aromatic rings, aliphatic rings and other groups such as -O-CH₃, -O-C₂H₅, amide, cyanide, cyanate, isocyanate, -C=N, -N=C, azo, hydrazo, benz etc were also added. The biological activities of these compounds were predicted using the G-QSAR model. 48 arylthioindole derivatives were further selected to obtain mechanistic insights into their inhibitory properties based on high predicted activity scores and extrapolation values. Ideally, the compounds with extrapolation values being zero or close to zero are considered good for further analysis. The compound with the highest predicted activity score of 6.047, temp0592 was substituted with an electronegative H-bond acceptor NH₂ group at the R4 site, a hydrogen atom at the R5 site and a sulphur atom at the R6 site which has a molecular weight of 32.065. It is heavier than CH₂OH having a molecular weight of 31.021 which was substituted at the same site in temp0598 and has lower activity value of 6.042 as R6-Molecular Weight is a positively contributing descriptor. Similarly, temp0596 goes further down on activity value as R6 site is substituted by the ethyl group and hence having lesser molecular weight. The presence of H bond Acceptor at the R4 site enhances the inhibitory activity of arylthioindole derivatives as in temp0592, temp1656, temp2454 and many more. All these observations can be summed up as the presence of an electronegative H bond acceptor at the R4 site and a high molecular weight group at the R6 site enhances the activity of arylthioindoles. But an increased slogp value at the R5 site, increased retention index chi2 and nitrogen atom count at the R6 site decreases the biological activity. These observations establish the importance of such substituents at these sites to enhance the tubulin inhibitory

activity of arylthioindole derivatives. The predicted activity values with respect to the three substitution sites R4, R5 and R6 have been provided for 20 top scoring compounds (Table 3).

Table 3. Predicted activity values of the combinatorial library with respect to R4, R5 and R6 site substitutions.

Evaluation of screened compounds docked against α - β interfacial site of tubulin heterodimer

1. Validation of homology based model of human tubulin

The homology based protein structure of human tubulin was validated using PSVS validation suite. The conformational shifts in the protein model encountered after molecular dynamic simulations has been depicted (Figure 3A). After obtaining the MD simulated structure, Ramchandran plot analysis was carried out to find the percentage of amino acid residues falling in the sterically allowed and disallowed regions. According to PROCHECK analysis, 90.4 % residues fall under the most favoured regions and no residues in the disallowed regions. Of the 431 residues, 34 were glycine and 20 were proline and were analysed separately. A model having over 90% residues falling in the favoured region can be considered to be good one on account of the analysis of 118 structures of resolution 2.0 Å. Thus, the human tubulin model is a good one and can be considered for further analysis. The Ramchandran plot analysis using PROCHECK has been depicted (Figure 3B). According to Richardson's Lab molprobity analysis, 97 % of the residues fall under the favoured regions and 0.2 % under the disallowed regions. Just one outlier ASN347 was encountered. As Molprobity analyses Glycine, Proline and pre-proline separately owing to basic physicochemical differences in comparison to other amino acids, 4 different Ramchandran plots are given (Figure 3C). The average PROCHECK G factor value for phi-psi angles was -0.15 and for all dihedral angles was -0.23. The G score is the log odds score of observed distribution of torsion angles and covalent geometries and quantifies the goodness of the structure. The high the value of G score, better is that dihedral angle in the structure in terms of falling under the favourable regions in a Ramchandran plot. As the average G score is negative, high number of residues was seen having distortions in dihedral angles.

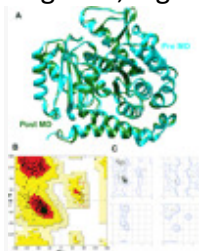


Figure 3. (a) Diagrammatic representation of structural differences between the pre MD and post MD simulated homology based model of human tubulin. (b) Ramchandran plot analysis from PSVS PROCHECK module. (c) Ramchandran plot analysis from Richardson lab's Molprobity module for glycine, proline and pre-proline separately and rest of the residues.

2. Screening of combinatorial library through docking and evaluation of the top scoring compounds

The 48 compounds having high predicted activity values chosen from the combinatorial library were docked against the tubulin assembly targeting the α - β interfacial site of the homology based protein structure of tubulin chain B. The top two compounds showing maximum affinity for the colchicine binding site were selected. Both the target-ligand complexes obtained were MD simulated and their interactions observed. The first compound with high affinity for tubulin, temp1662 named 5-(hydroxymethyl)-3- $\{[3-(hydroxymethyl) phenyl] sulfanyl\}$ -H-indol-2-

of (HPI)(Figure 4A) was involved in both hydrogen and hydrophobic interactions. The O atom of Pro243 formed a 2.72 Å long hydrogen atom with O1 atom of HPI and the rest others like Phe242, Gly244, Gln245, Leu246, Met323, Ala352, Val353, Cys354 and Thr351 were found to be involved in hydrophobic interactions (Figure 5A). After performing MD simulations in two slots of 10ns each, the stable average structure was obtained from 10 to 20ns. New interaction patterns could be seen post simulation with a 3.18 Å long H bond formed between the S atom of HPI and N atom of Gln245 which was involved in hydrophobic interaction before MD simulations. HPI seemed to have shifted from the earlier position owing to new hydrophobic interactions being formed including, Asn247, Pro243, Thr238 and Cys239. Rest all interactions were conserved with Leu246, gly244, Met323, Thr351, Ala352, Val353 and Cys354 (Figure 5B).

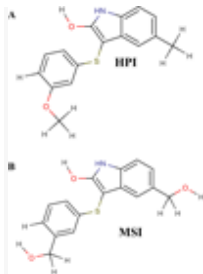


Figure 4. The structures of the two reported compounds (a) HPI and (b) MSI.

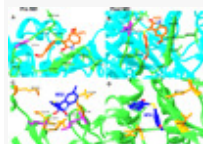


Figure 5. Diagrammatic representation of residues involved in various interactions with (a) HPI (shown in orange) forming hydrogen bond with Pro243 (purple) and residues involved in hydrophobic interactions (shown in green) (b) The post MD interactions of HPI showing one H bond with Gln245. (c) MSI (shown in blue) forming hydrogen bonds with Gly244 and Thr351 (purple) and residues involved in hydrophobic interactions are shown in orange. (d) The post MD interaction of MSI having no hydrogen bonds.

The second compound temp3789 named 3-[(3-methoxyphenyl) sulfanyl]-5-methyl-1H-indol-2-ol(MSI) (Figure 4B) was also found to be involved in both hydrogen bonds as well as hydrophobic interactions with the colchicine binding site in the modelled human tubulin protein. Leu246 and Thr351 were involved in forming hydrogen bonds of distances 2.97 and 3.29 Å respectively with MSI and Gly244, Gln245, Met323, Lys350, Ala352 and Val353 were involved in hydrophobic bonding. The hydrogen bonds were formed between Nitrogen atom of Lys246 with the O1 atom and Nitrogen atom of Thr351 with O2 atom of MSI (Figure 5C). After MD simulations of 10 ns, the interactions were seen to be reduced to hydrophobic interaction with Met323, Gln245 and Val353 (Figure 5D). The RMSD graphs for depicting the course of three MD simulations i.e. tubulin homology model (Figure 6A), ligand complexes with HPI (Figure 6B) and MSI (Figure 6C) have been given. When compared to the interactions involved between the most active arylthioindole derivative and tubulin, similar mode of action was discovered forming a 2.86Å long H-bond between N atom of Lys246 with O1 atom of the reference compound. Hydrophobic interactions were found to exist involving a few but same residues as those involved with HPI and MSI including Gly244, Gln245, Lys350, Ala352 and Val353 (Table 4). Thus, the two reported compounds HPI and MSI bind to the same colchicine binding site at the α - β interfacial cavity of the tubulin assembly and show good binding affinity (Figure 7). The predicted activities of HPI and MSI were 6.042 and 6.001 based on the G-QSAR

model and thus, are in absolute concurrence with the docking results in terms of high binding affinity with the tubulin assembly having similar mode of action as that of colchicine but with much higher biological activity than the most active compound of the congeneric series taken for this study.

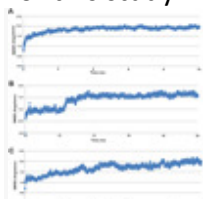


Figure 6. The graphical depiction of RMSD trajectory of (a) homology based tubulin model (from 0-10 ns) (b) target-HPI complex (from 10-20 ns) and (c) target-MSI complex (from 0-10 ns).

Table 4. Tubulin residues that are involved in hydrogen and hydrophobic interactions with HPI and MSI respectively in comparison with the reference compound.



Figure 7. (a) Depiction of the complete *Bos taurus* tubulin assembly comprising chains A, B, C, D and E of which A and C belong to the α chain (shown in orange); B and D belong to the β chain (shown in yellow). The modelled human tubulin β -chain has been compared (in red) and the ligand is shown in blue bound at the α - β interfacial colchicine binding site. (b) Depiction of the binding modes of HPI (in blue) and MSI (in green) compared to the most active arylthioindole derivative (in red) taken as the reference compound and important residues (in cyan).

ADMET analysis of the two top compounds, HPI & MSI

A total of 50 ADME properties were calculated using the Quikprop module. The first property #star depicts the number of properties that fall out of the similarity criteria of 95% of known drugs. The values for HPI and MSI were 0 and 1 respectively and signify that no molecular property of HPI falls out of the similarity criteria while only one criterion falls out for MSI. Another property CNS (central nervous system activity) ranging from -2 (inactive) to 2 (active) resulted in a value of -2 for HPI and zero for MSI. QPlogBB is the predicted brain/blood partition coefficient which had a value of -1.416 and -0.218 for HPI and MSI respectively. Descriptors including SASA (solvent accessible surface area), its hydrophobic component FOSA (saturated carbon and attached hydrogen) and hydrophilic component FISA (N, O and H on heteroatoms) were also predicted within the recommended ranges for both compounds. The compound HPI had a value of 81.27% and MSI had a value of 100% for percent human oral absorption descriptor. The predicted skin permeability factor QPlogkp resulted in a value of -3.342 and -1.635 for the two compounds HPI and MSI respectively. Another set of descriptors including QPlogPC16 (Free energy of solvation in hexadecane), QPlogPoct (Free energy of solvation in octanol), QPlogPw (Free energy of solvation in water) give the distribution of compounds in the body. QPlogPo/w (Predicted octanol/water partition coefficient) is the partition coefficient that

gives an idea of the hydrophobic nature of the chemical compounds and resulted in high values of 2.244 and 4.197 for MSI and HPI respectively suggesting their easy absorption through the lipid bilayer. It also finds out the number of various chemical groups present in the test compounds for e.g. amide, acid etc. The ionization potential and electron affinities (in electron Volts) for both the compounds were also obtained being 8.489 and 0.349 for HPI and 8.504 and 0.392 for MSI respectively. Lastly, it also evaluated the two compounds on the basis of Lipinski Rule-of-five and Jorgensen Rule-of-three violations which resulted in zero violations for both, HPI and MSI.

Certain other ADMET properties calculated by the online server admetSAR have also been evaluated. The absorption factors for HPI can be summarised in terms of blood brain barrier (BBB), human intestinal absorption (HIA) and CaCo2 permeability and resulted in a positive value (value above the prescribed threshold suggesting good permeability) with high probabilities. A compound with >30% HIA is absorbed easily. Similarly, CaCo2 permeability is taken as the *in vitro* model of human small intestinal mucosa as it provides a physical and biochemical barrier to ions and small molecules and hence must be tested for orally administered drugs. Under the metabolism category, HPI was found to be a non-substrate for CYP450 2C9 and 2D6, substrate for CYP450 3A4, inhibitor for 1A2 and 2C9 and non-inhibitor for 2C19 and 3A4 and thus, having a high cytochrome inhibitory promiscuity. Under the toxicity category, it was found to be non-carcinogenic but toxic for Fish, *Tetrahymena* and Honey Bee. Similarly, MSI was also evaluated which resulted in positive BBB and HIA permeability but no CaCo2 permeability. Under the metabolism category, it was found to be a non-substrate for CYP450 2C9, 2D6 and 3A4, an inhibitor of 1A2, 2C9 and 2C19 and non-inhibitor of CYP450 2D6 and 3A4 and thus, showed high cytochrome inhibitory promiscuity like HPI. Under the toxicity category, it was found to be non toxic for AMES test thus a non-carcinogen while having high toxicity for fish and honey bee but no toxicity for *Tetrahymena*. Thus, MSI can be selected for further evaluation as the drug candidate having low toxicity values as predicted.

Conclusions

In this study, we reported a novel fragment-based group QSAR approach exploiting the disintegration of tubulin assembly leading to interruption in cell division and eventually apoptosis as a potent avenue in cancer management. Tubulin heterodimers when assembled form the major building blocks of microtubules and hence have important functional role in maintaining cell structural integrity, motility and division in terms of spindle fibre formation. Inhibition of the tubulin formation is seen as a viable option for deterring the growth of cancerous cells. The arylthioindole moiety has been considered to be a potential anti-tubulin compound having similar mode of action as other agents involved in cell division arrest like colchicine and combretastatin. A G-QSAR model can provide significant results in terms of the crucial chemical fragments required for enhancing the activity of an already present template compound known to be inhibitory to our target of interest. A statistically robust G-QSAR model with the test set comprising 10 compounds out of a total of 36 compounds gave an insight into the contribution of various substitutions at the three sites R4, R5 and R6 for which 5 descriptors were calculated namely, R4-H AcceptorCount, R6-chi2, R5-slogp, R6-NitrogensCount, R6-Molecular Weight. A combinatorial library was created by varying substituents at these sites and their activity was predicted based on the model. All compounds having an electronegative

H-bond acceptor at the R4 site and a substituent with high molecular weight at the R6 site were found to have high activities (pIC₅₀) while the presence of nitrogen atom at R6 site resulted in decrease of biological activity of the arylthioindole derivatives. All highly active compounds were docked at the colchicine binding α - β interfacial site of the homology-based human tubulin chain B. After final analysis of the binding affinities and interactions involved, we reported two compounds, HPI and MSI with predicted activities of 6.042 and 6.001 respectively in terms of pIC₅₀. Out of the two, the second compound MSI promises to be the first choice for further evaluation as the drug candidate, being less toxic in comparison to HPI. The present study elucidated the crucial structural requirements to enhance the anti-tubulin activity of arylthioindole derivatives. The study provides a better understanding of the inhibitory roles of the identified compounds in microtubule disassembly and paves way for consideration of these compounds as potent anticancer leads.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CT and AGr designed the methods and experimental setup. CT and AG carried out the implementation of the various methods and were assisted by AG and SG. CT and AGr wrote the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

AG is thankful to Jawaharlal Nehru University for usage of all computational facilities. AG would like to thank University Grants Commission, India for the Faculty Recharge position.

Declaration

AG would like to acknowledge financial support from Department of Science and Technology, Government of India towards publication of this article.

This article has been published as part of *BMC Genomics* Volume 15 Supplement 9, 2014: Thirteenth International Conference on Bioinformatics (InCoB2014): Computational Biology. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/15/S9>.

References

1. Checchi PM, Nettles JH, Zhou J, Snyder JP, Joshi HC: **Microtubule-interacting drugs for cancer treatment.** *Trends Pharmacol Sci* 2003, **24**(7):361-365.
2. Nicolaou K, Vourloumis D, Li T, Pastor J, Winssinger N, He Y, Ninkovic S, Sarabia F, Vallberg H, Roschangar F: **Designed epothilones: Combinatorial synthesis, tubulin assembly properties, and cytotoxic action against taxol-resistant tumor cells.** *Angew Chem Int Edit* 1997, **36**(19):2097-2103.
3. Pettit GR, Singh SB, Niven ML, Hamel E, Schmidt JM: **Isolation, structure, and synthesis of combretastatins A-1 and B-1, potent new inhibitors of microtubule assembly, derived from Combretum caffrum.** *J Nat Prod* 1987, **50**(1):119-131.
4. Mooberry SL: **Microtubules as a target for anticancer drugs.** *New Frontiers and Treatment Paradigms for Metastatic Breast Cancer* 2011, **28**:7.

5. Lockwood AH: **Tubulin assembly protein: immunochemical and immunofluorescent studies on its function and distribution in microtubules and cultured cells.**
Cell 1978, **13**(4):613-627.
6. Nogales E: **Structural insights into microtubule function.**
Annu Rev Biochem 2000, **69**(1):277-302.
7. Hammond JW, Cai D, Verhey KJ: **Tubulin modifications and their cellular functions.**
Curr Opin Cell Biol 2008, **20**(1):71-76.
8. Pellegrini F, Budman DR: **Review: tubulin function, action of antitubulin drugs, and new drug development.**
Cancer Invest 2005, **23**(3):264-273.
9. Prinz H, Ishii Y, Hirano T, Stoiber T, Camacho Gomez JA, Schmidt P, Düssmann H, Burger AM, Prehn JH, Günther EG: **Novel benzylidene-9 (10 H)-anthracenones as highly active antimicrotubule agents. Synthesis, antiproliferative activity, and inhibition of tubulin polymerization.**
J Med Chem 2003, **46**(15):3382-3394.
10. Prakasham A, Saxena A, Luqman S, Chanda D, Kaur T, Gupta A, Yadav D, Chanotiya C, Shanker K, Khan F: **Synthesis and anticancer activity of 2-benzylidene indanones through inhibiting tubulin polymerization.**
Bioorgan Med Chem 2012, **20**(9):3049-3057.
11. Wilson L, Meza I: **The mechanism of action of colchicine. Colchicine binding properties of sea urchin sperm tail outer doublet tubulin.**
J Cell Biol 1973, **58**(3):709-719.
12. Ravelli RB, Gigant B, Curmi PA, Jourdain I, Lachkar S, Sobel A, Knossow M: **Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain.**
Nature 2004, **428**(6979):198-202.
13. Jordan MA, Thrower D, Wilson L: **Effects of vinblastine, podophyllotoxin and nocodazole on mitotic spindles. Implications for the role of microtubule dynamics in mitosis.**
J cell Sci 1992, **102**(3):401-416.
14. Lin CM, Ho HH, Pettit GR, Hamel E: **Antimitotic natural products combretastatin A-4 and combretastatin A-2: studies on the mechanism of their inhibition of the binding of colchicine to tubulin.**
Biochemistry 1989, **28**(17):6984-6991.
15. Shan YS, Zhang J, Liu Z, Wang M, Dong Y: **Developments of combretastatin A-4 derivatives as anticancer agents.**
Curr Med Chem 2011, **18**(4):523-538.
16. Risinger AL, Mooberry SL: **Microtubules as a target in cancer therapy.** In *Cytoskeleton and Human Disease*. Springer; 2012:203-221.
17. Beckers T, Mahboobi S: **Natural, semisynthetic and synthetic microtubule inhibitors for cancer therapy.**
Drugs Future 2003, **28**:767-785.
18. Mani S, Macapinlac M Jr, Goel S, Verdier-Pinard D, Fojo T, Rothenberg M, Colevas D: **The clinical development of new mitotic inhibitors that stabilize the microtubule.**
Anti-cancer Drug 2004, **15**(6):553-558.

19. De Martino G, La Regina G, Coluccia A, Edler MC, Barbera MC, Brancale A, Wilcox E, Hamel E, Artico M, Silvestri R: **Arylthioindoles, potent inhibitors of tubulin polymerization.**
J Med Chem 2004, **47**(25):6120-6123.
20. De Martino G, Edler MC, La Regina G, Coluccia A, Barbera MC, Barrow D, Nicholson RI, Chiosis G, Brancale A, Hamel E: **New arylthioindoles: potent inhibitors of tubulin polymerization. 2. Structure-activity relationships and molecular modeling studies.**
J Med Chem 2006, **49**(3):947-954.
21. Coluccia A, Sabbadin D, Brancale A: **Molecular modelling studies on arylthioindoles as potent inhibitors of tubulin polymerization.**
Eur J Med Chem 2011, **46**(8):3519-3525.
22. Lu Y, Chen J, Xiao M, Li W, Miller DD: **An overview of tubulin inhibitors that interact with the colchicine binding site.**
Pharm Res 2012, **29**(11):2943-2971.
23. Massarotti A, Coluccia A, Silvestri R, Sorba G, Brancale A: **The tubulin colchicine domain: a molecular modeling perspective.**
Chem Med Chem 2012, **7**(1):33-42.
24. Nogales E, Wolf SG, Downing KH: **Structure of the $\alpha\beta$ tubulin dimer by electron crystallography.**
Nature 1998, **391**(6663):199-203.
25. Liao Sy, Miao Tf, Chen Jc, Lu Hl, Zheng Kc: **Molecular modeling and design of arylthioindole derivatives as tubulin inhibitors.**
Chin J Chem Phy 2009, **22**(5):473-480.
26. Brancale A, Silvestri R: **Indole, a core nucleus for potent inhibitors of tubulin polymerization.**
Med Res Rev 2007, **27**(2):209-238.
27. Goyal M, Dhanjal JK, Goyal S, Tyagi C, Hamid R, Grover A: **Development of dual inhibitors against alzheimer's disease using fragment-based QSAR and molecular docking.**
BioMed Res Int 2014, **2014**:12.
28. Ajmani S, Jadhav K, Kulkarni SA: **Group-Based QSAR (G-QSAR): Mitigating interpretation challenges in QSAR.**
QSAR Comb Sci 2009, **28**(1):36-51.
29. Goodarzi M, da Cunha EF, Freitas MP, Ramalho TC: **QSAR and docking studies of novel antileishmanial diaryl sulfides and sulfonamides.**
Eur J Med Chem 2010, **45**(11):4879-4889.
30. Goyal S, Dhanjal JK, Tyagi C, Goyal M, Grover A: **Novel fragment-based QSAR modeling and combinatorial design of pyrazole-derived CRK3 inhibitors as potent antileishmanials.**
Chem Biol Drug Des 2014, **84**(1):54-62.
31. Goyal S, Grover S, Dhanjal JK, Tyagi C, Goyal M, Grover A: **Group-based QSAR and molecular dynamics mechanistic analysis revealing the mode of action of novel piperidinone derived protein-protein inhibitors of p53-MDM2.**
J Mol Graph Model 2014, **51**:64-72.

32. Tyagi C, Grover S, Dhanjal J, Goyal S, Goyal M, Grover A: **Mechanistic insights into mode of action of novel natural cathepsin L inhibitors.**
BMC Genomics 2013, **14**(Suppl 8):S10.
33. Ajmani S, Kulkarni SA: **Application of QQSAR for scaffold hopping and lead optimization in multitarget inhibitors.**
Mol Inform 2012, **31**(6-7):473-490.
34. Giansanti V, Piscitelli F, Camboni T, Prosperi E, La Regina G, Parks M, Silvestri R, Scovassi AI: **Arylthioindoles: Promising compounds against cancer cell proliferation.**
Oncol Lett 2010, **1**(1):109-112.
35. La Regina G, Edler MC, Brancale A, Kandil S, Coluccia A, Piscitelli F, Hamel E, De Martino G, Matesanz R, Díaz JF: **Arylthioindole inhibitors of tubulin polymerization. 3. Biological evaluation, structure-activity relationships and molecular modeling studies.**
J Med Chem 2007, **50**(12):2865-2874.
36. MarvinSketch[<http://www.chemaxon.com>] *webcite*
v. 5.12.1, ChemAxon; 2013.
37. VLifeMDS: Molecular Design Suite v. 4.3, VLife Sciences Technologies Pvt. Ltd., Pune, India; 2010.
38. Wold S, Sjöström M, Eriksson L: **PLS-regression: a basic tool of chemometrics.**
Chemometr Intell Lab 2001, **58**(2):109-130.
39. Rucker C, Rucker G, Meringer M: **y-Randomization and its variants in QSPR/QSAR.**
J Chem Inf Model 2007, **47**(6):2345-2357.
40. Schrodinger suite Schrodinger, LLC, New York, NY; 2009.
41. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK: **Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.**
J Med Chem 2004, **47**(7):1739-1749.
42. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL: **Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening.**
J Med Chem 2004, **47**(7):1750-1759.
43. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.**
J Mol Biol 1993, **234**(3):779-815.
44. Fiser A, Do RKG, Šali A: **Modeling of loops in protein structures.**
Protein Sci 2000, **9**(9):1753-1773.
45. Eswar N, Webb B, Marti-Renom MA, Madhusudhan M, Eramian D, Shen My, Pieper U, Sali A: **Comparative protein structure modeling using Modeller.**
Curr Protoc Bioinformatics 2006, 5.6.1-5.6.30.
46. Marti-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A: **Comparative protein structure modeling of genes and genomes.**
Annu Rev Biophys Bio 2000, **29**(1):291-325.
47. Bhattacharya A, Tejero R, Montelione GT: **Evaluating protein structures determined by structural genomics consortia.**
Proteins 2007, **66**(4):778-795.

48. Joosten RP, Te Beek TA, Krieger E, Hekkelman ML, Hooft RW, Schneider R, Sander C, Vriend G: **A series of PDB related databases for everyday needs.**
Nucleic Acids Res 2011, **39**(suppl 1):D411-D419.
49. Kabsch W, Sander C: **Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.**
Biopolymers 1983, **22**(12):2577-2637.
50. Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C-y, Powers R, Montelione GT: **Automated analysis of protein NMR assignments using methods from artificial intelligence.**
J Mol Biol 1997, **269**(4):592-610.
51. Huang YJ, Powers R, Montelione GT: **Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics.**
J Am Chem Soc 2005, **127**(6):1665-1674.
52. Huang YJ, Tejero R, Powers R, Montelione GT: **A topology-constrained distance network algorithm for protein structure determination from NOESY data.**
Proteins 2006, **62**(3):587-603.
53. Eisenberg D, Lüthy R, Bowie JU: **VERIFY3D: assessment of protein models with three-dimensional profiles.**
Method Enzymol 1997, **277**:396.
54. Sippl MJ: **Recognition of errors in three-dimensional structures of proteins.**
Proteins 1993, **17**(4):355-362.
55. Laskowski RA, MacArthur MW, Moss DS, Thornton JM: **PROCHECK: a program to check the stereochemical quality of protein structures.**
J Appl Crystallogr 1993, **26**(2):283-291.
56. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS: **MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.**
Nucleic Acids Res 2007, **35**(suppl 2):W375-W383.
57. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W: **Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.**
J Comput Aid Mol Des 2013, **27**(3):221-234.
58. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT: **Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes.**
J Med Chem 2006, **49**(21):6177-6196.
59. Goyal M, Grover S, Dhanjal JK, Goyal S, Tyagi C, Chacko S, Grover A: **Novel natural structure corrector of ApoE4 for checking Alzheimer's disease: benefits from high throughput screening and molecular dynamics simulations.**
Biomed Res Int 2013, **2013**:620793.
60. Dhanjal JK, Grover S, Sharma S, Singh A, Grover A: **Structural insights into mode of actions of novel natural Mycobacterium protein tyrosine phosphatase B inhibitors.**
BMC Genomics 2014, **15**(Suppl 1):S3.

61. Dhanjal JK, Goyal S, Sharma S, Hamid R, Grover A: **Mechanistic insights into mode of action of potent natural antagonists of BACE-1 for checking Alzheimer's plaque pathology.**
Biochem Biophys Res Commun 2014, **443**(3):1054-1059.
62. Wallace AC, Laskowski RA, Thornton JM: **LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.**
Protein Eng 1995, **8**(2):127-134.
63. Beckstein O, Fourrier A, Iorga BI: **Prediction of hydration free energies for the SAMPL4 diverse set of compounds using molecular dynamics simulations with the OPLS-AA force field.**
J Comput Aid Mol Des 2014, **28**(3):265-276.
64. Guo Z, Mohanty U, Noehre J, Sawyer TK, Sherman W, Krilov G: **Probing the alpha-helical structural stability of stapled p53 peptides: molecular dynamics simulations and analysis.**
Chem Biol Drug Des 2010, **75**(4):348-359.
65. Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD: **Scalable algorithms for molecular dynamics simulations on commodity clusters.** In *SC 2006 Conference, Proceedings of the ACM/IEEE: 2006.* IEEE; 2006:43-43.
66. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG: **A smooth particle mesh Ewald method.**
J Chem Phys 1995, **103**(19):8577-8593.
67. Quikprop v. 3.4, Schrodinger, LLC, New York, NY; 2011.
68. Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, Lee PW, Tang Y: **admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties.**
J Chem Inf Model 2012, **52**(11):3099-3105.

INTERACT-O-FINDER: A Tool for Prediction of DNA-Binding Proteins Using Sequence Features

Monika Samant · Minesh Jethva · Yasha Hasija

Accepted: 2 December 2014
© Springer Science+Business Media New York 2014

Abstract Protein–DNA interactions carry out many important regulatory functions in our body which are essential for growth and survival. The molecular level understanding of these interactions helps us to decipher the mechanism of regulation. Conventionally these interactions were identified using small scale experimental techniques and high throughput technologies. But these approaches being time consuming and expensive, arouse the need of a computational approach for the prediction of interactions. To this end, a machine learning approach for predicting and classifying DNA-binding proteins has been used. Weka and LibSVM platform have been used to develop an effective classifying model. The different classifiers were applied on the attributes generated from protein sequence and tried to develop an efficient model for classifying a protein as DNA-interacting or DNA-non-interacting. Among several classifying algorithms applied to generate models, best performance was achieved using LibSVM with 87.83 % accuracy. The tool named INTERACT-O-FINDER, based on the prediction model is available at <http://interacto.eurekanow.org/index.html>.

Keywords DNA · Attributes · Classifiers · Algorithms · LibSVM

Introduction

Proteins are the biomolecules which play pivotal role in major functional processes inside the body. They interact with other biomolecules and regulate many physiological processes critical for the body growth and survival. Protein–DNA interactions are the major interactions which play significant roles in gene regulation (András and Jeffrey 2006; Blouin et al. 2009). To decipher the mechanisms of gene regulation, understanding the molecular details of protein–DNA interactions is essential. The mechanisms underlying the DNA–protein interactions are not completely understood. From the present experimental studies, information about individual DNA–protein complexes have been deciphered which may not always apply to generalized cases. Also, the single experimental study cannot be used to govern general rule for protein interactions; these studies may be collectively used to infer the mechanism behind these interactions. This knowledge helps us to understand further underlying mechanisms of protein–DNA interactions, which has been a topic of significant interest for many years (Guy et al. 2009; Nagarajan et al. 2013; Bhardwaj et al. 2005; Jones 1999; Andrea et al. 2012).

DNA-binding proteins interact with DNA and play important role in DNA processing events such as replication, transcription and translation regulation i.e. maintenance and regulation of gene expression (Yan et al. 2006; Chuanxin et al. 2013; Zhou and Yan 2011). Many experimental techniques have been extensively used for the identification of proteins involved in DNA binding interactions. However, these techniques are costly and time consuming, thereby limiting genome wide predictions of DNA-binding interactions (Ahmad and Sarai 2004; Caragea et al. 2010; Luscombe et al. 2000; Ren et al. 2000; Tjong and Zhou 2007).

M. Samant · Y. Hasija (✉)
Department of Biotechnology, Delhi Technological University,
Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India
e-mail: yashahasija@gmail.com

M. Jethva
Synthetic Biology and Biofuels, International Center for Genetic
Engineering and Biotechnology, Aruna Asaf Ali Marg,
Delhi 110067, India

With the advances in high throughput sequencing, there has been an exponential increase in the number of macromolecular sequences, structures, gene expression measurements. These progresses have transformed the status of biology from a data-poor science to a data-rich science (Stawiski et al. 2003; Fang et al. 2008; Gao and Skolnick 2009). In spite of these advances, biology has remained a largely descriptive science and it requires costly and time consuming experimental procedures for data analysis (Kumar et al. 2007; Shao et al. 2009). Advances in machine learning provide improved methods for learning from highly unbalanced datasets and reliable methods for assessing the performance of the resulting models (Lin et al. 2011; Xiong et al. 2011; Sarai and Kono 2005; Liu and Bradley 2012; Langlois and Lu 2010; Si et al. 2011).

The most important issue in machine learning approach is selection of protein attributes and choice of classifying algorithm. Till now, models have been built for DNA-binding proteins prediction by either frequency or composition attributes individually from sequence features (Zhao et al. 2010; Yang et al. 2010; Chu et al. 2009; Chawla et al. 2004). Additionally other servers for predicting DNA interacting proteins include DISIS (Ofra et al. 2007), DNABindR (Yan et al. 2006), BindN (Wang and Brown 2006), DP-Bind (Hwang et al. 2007), DBS-Pred (Ahmad and Sarai 2004), MetaDBSite (Si et al. 2011) are based on protein sequences only. Therefore, a better model is required which exploits already present features and adds newly discovered features for the purpose of prediction of DNA-binding proteins. In this study, a prediction model has been proposed, which was developed on several physio-chemical features such as frequency and composition extracted from protein sequences. This study has used already present features and added several new features for development of a better model (Jones et al. 2003; Hall et al. 2009; Yang et al. 2010; Larrañaga et al. 2006; Tarca et al. 2007; Wang et al. 2003; Chang and Lin 2011).

The goal of this work is to develop a model to study the protein sequence characteristics and to classify the proteins as DNA interacting or DNA non-interacting proteins. DNA-binding proteins differ in their binding properties which is believed to be dependent upon the primary structure i.e. amino acid sequence. The method is based on characterizing the function of the protein by unique properties of their sequences. Using a combination of general protein features and specific physiochemical properties extracted from the sequences, a model was trained successfully to differentiate DNA-binding proteins from non-DNA binding ones. Models were generated using different classifiers such as Random Forest, Naive Bayes and J48 and SVM. It was found that model generated with SVM

classifier showed best performance among all the models with 87.83 % accuracy.

Materials and Methods

Collection of Instances

For Positive Set

DNA-binding protein sequences were collected from the Protein Data Bank (PDB) <http://www.pdb.org/>. The list of the IDs of DNA-binding proteins was obtained from NPIDB <http://npidb.belozersky.msu.ru/>. The sequence data was obtained from PDB by using IDs as an input to the batch retrieval tool of PDB. Sequence information of a total of 2715 DNA–protein complexes was obtained. All the sequences obtained from PDB were in the form of DNA–protein complexes which meant that the fasta files obtained for each DNA–protein complexes contained nucleic acid sequences along with the interacting protein. The nucleotide sequences were excluded from all the individual fasta files to avoid misclassification.

Additionally, sequences <15 amino acids or with long stretches of X were excluded from the dataset in order to remove redundancy, possible inclusion of uncharacterized peptides. For the exclusion of homologous sequences, <30 % similarity filter was applied. Homology of the data was removed with the help of CD-HIT i.e. protein sequence clustering program (Yang et al. 2010).

For Negative Set

The negative set was made by first downloading all the protein IDs from Protein Data Bank (PDB) <http://www.pdb.org/> and then subtracting the DNA-binding protein IDs from it. The IDs retrieved after the subtraction were used as negative set. The sequence files of negative set were obtained and refined as in the case of positive data set.

Construction of Training Sets

Attribute Generation and Building Dataset

Physiochemical features such as frequency and composition were calculated for all the protein sequences. Frequency of all the 20 amino acids is calculated for attribute generation. The ‘frequency’ is defined as the percentages of each of the 20 amino acids in the protein:

$$fA1 = \frac{NA}{N} \quad (1)$$

Where fA1 is the frequency of amino acid A1, NA is the number of each of the 20 amino acids in the protein and N is the total number of amino acid in the protein sequence i.e. length of the sequence.

For generation of “composition” set, several physiochemical features were selected and each amino acids is classified into three classes related to them (Table 1). The composition attribute was calculated as percentage for each group.

Attributes based on the frequency and compositions were generated for all the files. Three set of files were created. First file contained only amino acid frequencies, second file contained composition of the sequence such as degree of hydrophobicity, polarity, and solvent accessibility etc. and third file contain both these features. Seven training sets were created for all three categories with equal number of positive and negative entries. Since negative set entries were greater than the positive set entries, equal number of positive and negative entries are randomly entered into the training set to avoid bias. This file generation was carried out by an in-house developed shell script. The final training set contained total 988 entries with equal number of positive and negative set entries.

Balancing and Building the Random Training Dataset

The number of DNA-interacting proteins was much higher than non-DNA interacting proteins which causes imbalance in the training data and affects the performance of ML algorithms. Therefore, it was required to build balanced datasets. Seven balanced datasets, each consisting of 988 instances (494 non-PDIs and 494 PDIs) from original

dataset. The dataset were generated by randomly picking the instances from original datasets. The randomization helps to avoid the bias in the generated datasets.

The final training file was entered into Weka and Lib-SVM and learned with different classifiers such as Naive Bayes, Random forest, J48 and SVM. Then model was optimized by altering the gamma and cost parameters of the classifiers to obtain the more accurate results.

Development of User-interface

Using this model, an online tool is developed to predict DNA-binding capability of the proteins. The tool takes protein sequence as an input and then generates descriptors using in-house developed python script. The descriptors were taken as an input into the model which classifies the input into positive (DNA-binding) and negative (DNA non-binding) and gives an output. The web-server has a user friendly interface and was hosted on apache-tomcat based server and developed using HTML in the front-end, PHP and Javascript at the back-end for connecting the model with the user interface.

Results

In this work, a tool is developed for predicting DNA-interacting proteins. A model based on sequence features was generated by machine learning algorithms. The effect of physiochemical features i.e. frequency and composition were analysed on the model accuracy. Attributes based on frequency of each of the 20 amino acids and different compositions such as hydrophobicity, polarities etc. were

Table 1 Table describing the composition features divided into three groups

Physicochemical feature	Group 1	Group 2	Group 3
Hydrophobicity	Polar R K E D Q N	Neutral G A S T P H Y	Hydrophobicity C L V I M F W
Normalized Vander Waals volume	Vol. range (0–2.78) G A S T P D C	Vol. range (2.95–4.0) N V E Q I L	Vol. range (4.03–8.08) M H K F R Y W
Polarity	Polarity value (4.9–6.2) L I F W C M V Y Polarizability value (0–1.08) G A S D T	Polarity value (8.0–9.2) P A T G S Polarizability value (0.128–0.186) C P N V E Q I L	Polarity value (10.4–13.0) H Q R K N E D Polarizability value (0.219–0.409) K M H F R Y W
Charge	Positive K R	Neutral A N C Q G H I L M F P S T W Y V	Negative D E
Secondary structure	Helix E A L M Q K R H	Strand V I Y C W F T	Coil G N P S D
Solvent accessibility	Buried A L F C G I V W	Exposed R K Q E N D	Intermediate M P S T H Y

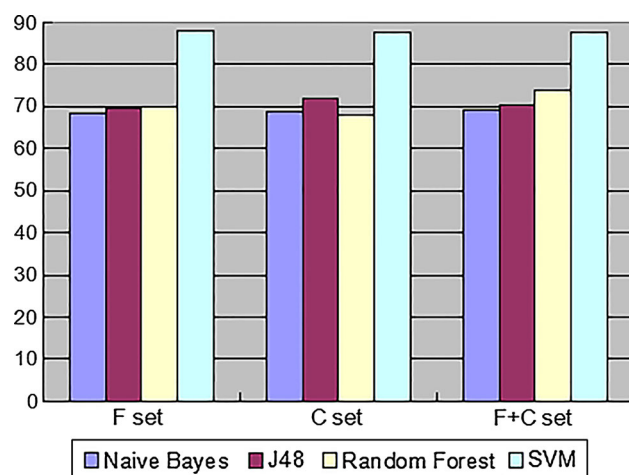


Fig. 1 This graph shows the result statistics of accuracy of models generated with different classifiers on various datasets. F set consists of only frequency attributes, C set consists of only composition attributes and F+C set contains both frequency and composition attributes

generated and models were developed with their combinations. Nine models were generated with four classifying algorithms applied on different datasets. Naive Bayes, J48, Random forest and SVM were the classifying algorithms adopted for generation of prediction model. Their performance was measured on the basis of accuracy of the models. Graph in Fig. 1 shows the result statistics of accuracy of models generated with different classifiers on various datasets. Model generated with SVM algorithm on LibSVM showed best accuracy of 87.83 %. The tool based on this prediction model is present on the web server named INTERACT-O-FINDER at <http://interacto.eureka.org/index.html>.

Discussion

The prediction of DNA interacting proteins from protein sequence is of great importance to understand the mechanism behind DNA–protein interactions. A lot of research efforts have gone into classifying proteins either as DNA interacting or DNA non-interacting. Several bioinformatics strategies have been explored to develop an efficient model for this classification. Machine learning tools have been applied using different classifiers and different kinds of features/attributes based on protein sequence, structure and binding affinities. Such approaches in machine learning were used to develop the most optimal model by exploring the contribution of different attributes to interaction.

Among the available servers, BindN and DP-Bind were found to have the highest accuracies. These servers are based on few sequence features such as pKa value,

hydrophobicity index and molecular mass of amino acids. These servers have been developed on SVM based classifiers. The datasets used in the earlier studies were downloaded from PDB. BindN takes only three sequence features and is developed on three machine learning algorithms. It holds an accuracy of 70 %, whereas DP-Bind, which is also based on sequence features and had been developed on three machine learning algorithms, has a better accuracy i.e. 77 %, than BindN.

It was observed that models generated by only frequency or composition attributes showed less accuracy than models developed on the combination of frequency and composition attributes. The results demonstrated that hybrid features contributed most to the prediction performance. Several classifying algorithms such as J48, Random forest, Naïve Bayes and SVM were used to develop a better and efficient model. Model generated with SVM algorithm on LibSVM showed best accuracy among all the classifiers with an accuracy of 87.83 %. This is also evident from the graph in Fig. 1 which shows that SVM outperforms all the other classifiers in all the cases (frequency, composition and combination of both). The predictions based on “C set” features slightly improved with J40 classifier as compared to based on “F set” features. But SVM gave best results when combination of both frequency and composition of the proteins.

Conclusion

To best of our knowledge, there are few prediction servers available. Out of which, BindN (Wang and Brown 2006) and DP-Bind (Hwang et al. 2007) have the best performance. Till now, most of the models have been built on sequence features. The performance achieved with our study is higher than the already present servers. Already present servers are either based on solely frequency or composition, whereas our model is based on combination of both frequency and composition attributes. It was also observed that SVM based prediction has the best accuracy among models developed from other classifiers.

In the present study, we presented a SVM based method for prediction of DNA-interacting proteins with the accuracy of 87.83 %. The method has been implemented to develop a web based tool, INTERACT-O-FINDER, which is available online at <http://interacto.eureka.org/index.html>.

Acknowledgement Authors thank Nitin Thukral for scientific discussions.

Conflict of interest The authors declare no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Ahmad S, Sarai A (2004) Moment-based prediction of DNA-binding proteins. *J Mol Biol* 341:65–71
- András S, Jeffrey S (2006) Efficient prediction of nucleic acid binding function from low-resolution protein structures. *J Mol Biol* 358:922–933
- Andrea S, Ondřej K, Filip Ž, Jakub T (2012) Prediction of DNA-binding proteins from relational features. *Proteome Sci* 10:66
- Bhardwaj N, Langlois RE, Zhao G, Lu H (2005) Kernel-based machine learning protocol for predicting DNA-binding proteins. *Nucleic Acids Res* 33:6486–6493
- Blouin S, Craggs TD, Lafontaine DA, Penedo JC (2009) Functional studies of DNA–protein interactions using FRET techniques. *Methods Mol Biol* 543:475–502
- Caragea C, Caragea D, Silvescu A, Honavar V (2010) Semi-supervised prediction of protein subcellular localization using abstraction augmented Markov models. *BMC Bioinform* 11:S6
- Chang CC and Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:27:1–27:27
- Chawla NV, Japkowicz N, Kotcz A (2004) Editorial: special issue on learning from imbalanced data sets. *SIGKDD Explor Newsl* 6:1–6
- Chu WY, Huang YF, Huang CC, Cheng YS, Huang CK, Oyang YJ (2009) ProteDNA: a sequence-based predictor of sequence-specific DNA-binding residues in transcription factors. *Nucleic Acids Res* 37:W396–W401
- Chuanxin Z, Jiayu G, Honglin L (2013) An improved sequence based prediction protocol for DNA-binding proteins using SVM and comprehensive feature analysis. *BMC Bioinform* 14:90
- Fang Y, Guo Y, Feng Y, Li M (2008) Predicting DNA-binding proteins: approached from Chou's pseudo amino acid composition and other specific sequence features. *Amino Acids* 34:103–109
- Gao M, Skolnick J (2009) A threading-based method for the prediction of DNA-binding proteins with application to the human genome. *PLoS Comput Biol* 5:e1000567
- Guy N, András S, Christina L, Nir BT (2009) Identification of DNA-binding proteins using structural, electrostatic and evolutionary features. *J Mol Biol* 387:1040–1053
- Huang Y, Niu B, Gao Y, Limin F, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680–682
- Hwang S, Gou Z, Kuznetsov IB (2007) DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins. *Bioinformatics* 23:634–636
- Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195–202
- Jones S, Shanahan HP, Berman HM, Thornton JM (2003) Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. *Nucleic Acids Res* 31:7189–7198
- Kumar M, Gromiha M, Raghava G (2007) Identification of DNA-binding proteins using support vector machines and evolutionary profiles. *BMC Bioinform* 8:463
- Langlois RE, Lu H (2010) Boosting the prediction and understanding of DNA-binding domains from sequence. *Nucleic Acids Res* 38:3149–3158
- Larrañaga Pedro, Calvo Borja, Santana Roberto, Bielza Concha et al (2006) Machine learning in bioinformatics. *Brief Bioinform* 7:86–112
- Lin WZ, Fang JA, Xiao X, Chou KC (2011) IDNA-prot: identification of DNA binding proteins using random forest with grey model. *PLoS ONE* 6:e24756
- Liu LA, Bradley P (2012) Atomistic modeling of protein–DNA interaction specificity: progress and applications. *Curr Opin Struct Biol* 22:397–405
- Luscombe NM, Austin SE, Berman HM, Thornton JM (2000) An overview of the structures of protein–DNA complexes. *Genome Biol* 1:1–37
- Mark Hall, Eib Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian Witten H (2009) The WEKA data mining software: an update. *SIGKDD Explor Newsl* 11:10–18
- Nagarajan R, Shander A, Gromiha M (2013) Novel approach for selecting best predictor for identifying binding sites in DNA binding proteins. *Nucleic Acids Res* 41:7606–7614
- Ofran Y, Mysore V, Rost B (2007) Prediction of DNA-binding residues from sequence. *Bioinformatics* 23:i347–i353
- Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E et al (2000) Genome-wide location and function of DNA binding proteins. *Science* 290:2306–2309
- Sarai A, Kono H (2005) Protein–DNA recognition patterns and predictions. *Annu Rev Biophys Biomol Struct* 34:379–398
- Shao X, Tian Y, Wu L, Wang Y, Jing L, Deng N (2009) Predicting DNA-and RNA-binding proteins from sequences with kernel methods. *J Theor Biol* 258:289–293
- Si J, Zhang Z, Lin B, Schroeder M, Huang B (2011) MetaDBSite: a meta approach to improve protein DNA-binding sites prediction. *BMC Syst Biol* 5:S7
- Stawiski EW, Gregoret LM, Mandel-Gutfreund Y (2003) Annotating nucleic acid-binding function based on protein structure. *J Mol Biol* 326:1065–1079
- Tarca AL, Carey VJ, Chen XW, Romero R, Draghici S (2007) Machine learning and its applications to biology. *PLoS Comput Biol* 3:e116
- Tjong H, Zhou HX (2007) DISPLAR: an accurate method for predicting DNA-binding sites on protein surfaces. *Nucleic Acids Res* 35:1465–1477
- Wang L, Brown SJ (2006) BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Res* 34:W243–W248
- Wang G, Dunbrack RL Jr (2003) PISCES: a protein sequence culling server. *Bioinformatics* 19:1589–1591
- Xiong Y, Liu J, Wei DQ (2011) An accurate feature-based method for identifying DNA-binding residues on protein surfaces. *Proteins* 79:509–517
- Yan C, Terribilini M, Wu F, Jernigan R, Dobbs D, Honavar V (2006) Predicting DNA-binding sites of proteins from amino acid sequence. *BMC Bioinform* 7:262
- Yang L, Xia JF, Gui J (2010) Prediction of protein–protein interactions from protein sequence using local descriptors. *Protein Pept Lett* 17:1085–1090
- Zhao H, Yang Y, Zhou Y (2010) Structure-based prediction of DNA-binding proteins by structural alignment and a volume-fraction corrected DFIRE-based energy function. *Bioinformatics* 26:1857–1863
- Zhou W, Yan H (2011) Prediction of DNA-binding protein based on statistical and geometric features and support vector machines. *Proteome Sci* 9(Suppl 1):S1. doi:10.1186/1477-5956-9-S1-S1

Methods for Improving Thermodynamic Performance of Vapour Compression Refrigeration Systems using Thirteen Ecofriendly Refrigerants in Primary Circuit and Tio2 Nano Particles Mixed with R718 used in Secondary Evaporator Circuit for Reducing Global Warming and Ozone Depletion

R.S Mishra

Department of Mechanical Engineering, Delhi Technological University, New Delhi, India

Article Info

Article history:

Received 20 October 2014

Received in revised form

30 October 2014

Accepted 30 November 2014

Available online 15 December 2014

Keywords

Performance Improvement Vapour Compression Refrigeration Systems, Energy and Exergy Analysis, First and Second Law Analysis, Nano Refrigerants

Abstract

The methods for improving first law and second law efficiency have been considered in this paper by using water as secondary coolant in evaporator with nano particles such as Al_2O_3 and TiO_2 mixed R718 refrigerant is investigated in this paper. Detailed energy and exergy analysis of multi-evaporators at different temperatures in the vapour compression refrigeration systems have been done in terms of performance parameter for R507a, R125, R134a, R290, R600, R600a, R410a, R407c, R404a and R152a refrigerants.

The numerical computations have been carried out for both systems. The use of nano particles improves the first law and second law performance significantly. The best performance is found using R152a and worst performance is observed using R410a. Due to flammable nature of R290, R600, R600a and R152a

The results were compared by using water in secondary circuit with nano refrigerants and without nano particles used and it was found that use of nano particles improves thermal performances.

The first law performance improvement in terms of COP and second law performance in terms of exergetic efficiency (rational efficiency) using TiO_2 is better than using Al_2O_3 with R718 refrigerant in the secondary evaporator circuit.

Nomenclature

COP	Coefficient of Performance (Non-Dimensional)
VCR	Vapour Compression Refrigeration
CFC	Chlorofluorocarbon
HCFC	Hydro Chlorofluorocarbon
Q	Rate of Heat Transfer (kW)
W	Work Rate (kW)
T	Temperature (K)
Δ	Efficiency Defect (Non-Dimensional)
ΔT_{sc}	Degree of Sub cooling
EP	Exergy Rate of Product (kW)
EV	Expansion Valve
h	Specific Enthalpy (kJ/kg)
P	Pressure (kPa)
IR	Irreversibility (kW)
E_x	Exergy Rate of Fluid (kW)
M	Mass Flow Rate (kg/s)
S	Specific Entropy (kJ/kgK)
EF	Exergy Rate of Fuel (kW)
EL	Exergy Loss Rate (kW)
η	Efficiency (Non-Dimensional)
r	Refrigerant, Space to be Cooled
ex	Exergetic
EP	Evaporator
e	Evaporator
ev	Expansion Valve
c	condenser

Corresponding Author,

E-mail address: professor_rsmishra@yahoo.co.in

All rights reserved: <http://www.ijari.org>

1. Introduction

Refrigeration is a technology which absorbs heat at low temperature and provides temperature below the surrounding by rejecting heat to the surrounding at higher temperature. Simple vapour compression system which consists of four major components compressor, expansion valve, condenser and evaporator in which total cooling load is carried at one temperature by single evaporator but in many applications like large hotels, food storage and food processing plants, food items are stored in different compartment and at different temperatures. Therefore there is need of multi evaporator vapour compression refrigeration system. The systems under vapour compression technology consume huge amount of electricity, this problem can be solved by improving performance of system. The use of nano particles improves the first law and second law performance significantly. The best performance is found using R152a and worst performance is observed using R410a. Due to flammable nature of R290, R600, R600a and R152a

2. Literature Review

Vapour compression refrigeration system based applications make use of refrigerants which are responsible for greenhouse gases, global warming and ozone layer depletion. Montreal protocol was signed on the issue of substances that are responsible for depleting Ozone layer and discovered how much consumption and production of ozone depletion substances took place during certain time period for both developed and developing countries.

Another protocol named as Kyoto aimed to control emission of green house gases in 1997. The relationship between ozone depletion potential and global warming potential is the major concern in the field of GRT (green refrigeration technology) so Kyoto proposed new refrigerants having lower value of ODP and GWP. Internationally a program being pursued to phase out refrigerants having high chlorine content for the sake of global environmental problems. Due to presence of high chlorine content, high global warming potential and ozone depletion potential after 90's CFC and HCFC refrigerants have been restricted. Thus, HFC refrigerants are used nowadays, showing much lower global warming potential value, but still high with respect to non-fluorine refrigerants. Lots of research work has been done for replacing "old" refrigerants with "new" refrigerants. Ahamed et al. [1] emphasized on use of hydrocarbons and mixture compressor shows much higher exergy destruction as compared to rest of components in the vapour compression refrigeration system and this exergy destruction can be minimized by using of nanofluid and nanolubricants in compressor. Anand [2] did detailed exergy analysis of 2 ton of refrigeration capacity window air conditioning test rig with R22 as working fluid and reached to the conclusions, that irreversibility in system components will be highest when the system is 100% charged and lowest when 25% charged and irreversibility in compressor is highest among system components.

Arora et al. [3] developed numerical model of actual vapour compression refrigeration system with liquid vapour heat exchanger and did energy and exergy analysis on the same in the specific temperature range of evaporator and condenser and concluded that R502 is the best refrigerant compared to R404A and R507A and compressor is the worst component and liquid vapour heat exchanger is best component of the system in case of exergy transfer.

Bolaji et al. [4] had done experimentally comparative analysis of R32, R152a and R134a refrigerants in vapour compression refrigerator and concluded that R32 shows lowest performance whereas R134a and R152a showing nearly same performance but best performance was obtained of system using R152a.

Chopra et al. [5] carried out analysis of energetic and exergetic based comparison Multiple Evaporators with compound compression and flash intercooler with individual and multiple throttle valves and find out the effect of flash intercooler on second law performances.

Han et al. [6] Under different working conditions experimental results revealed that there could be replacement of R407C in vapour compression refrigeration system having rotor compressor with mixture of R32/R125/R161 showing higher COP, less pressure ratio and slightly high discharge compressor temperature without any modification in the same system.

Mishra et al. [7-9] performed numerical analysis of vapour compression refrigeration system using R134a, R143a, R152a, R404A, R410A, R502 and R507A, and discussed the effect of evaporator temperature, degree of subcooling at condenser outlet, superheating of evaporator outlet, vapour liquid heat exchanger effectiveness and degree of condenser temperature on COP and exergetic efficiency. They reported that evaporator and condenser temperature have significant effect on both COP and

exergetic efficiency and also found that R134a has the better performance while R404a has poor performance in all respect.

Saravana kumar [10] compared the performance between R134a and R290/R600a mixture on a domestic refrigerator which is originally designed to work with R134a and found that R290/R600a hydrocarbon mixture showed higher COP and exergetic efficiency than R134a. In their analysis, highest irreversibility obtained in the compressor compare to condenser, expansion valve and evaporator.

Based on the literature it was observed that researchers have gone through detailed first law analysis in terms of coefficient of performance and second law analysis in term of exergetic efficiency of simple vapour compression refrigeration system with single evaporator. Researchers did not go through the irreversibility analysis (second law analysis) of followings

- (i) Simple VCR with nano particles used as secondary evaporator circuit in the water cooled evaporator
- (ii) Detailed analysis of vapour compression refrigeration systems using thirteen ecofriendly refrigerants with effect of nano particles for improving their first and second law performances

This paper mainly deals with effect of nano particles (TiO_2 and Al_2O_3) mixed with R718 refrigerants was used in the water cooled evaporator for improving thermal performance of vapour compression refrigeration systems for keeping evaporator size constant due to enhancing heat transfer coefficient in the evaporator

3. Performance Evaluation

The computation modeling of vapor compression refrigeration systems was carried out with the help of EES for first and second law analysis in terms of energetic analysis i.e. COP (First law analysis) and exergetic analysis in terms of exergetic efficiency, exergy destruction ratio (EDR). In this analysis we assumed negligible pressure losses and heat losses. The comparative performance for condenser temperature varying between 320K to 330K with increment of 2 and evaporator temperature is varying from 265K to 281 K with increment of 4. The energy and exergy change in vapour compression refrigeration cycle have been calculated for various eco friendly refrigerants such as R125, R507, R-134a, R404a, R410a, R407c, R-290 (propane), R600 (butane), R-600a (isobutene) for environmental temperature of 298K. and results are shown. The performance of vapour compression refrigeration system using ecofriendly refrigerants in the primary circuit and TiO_2 nano particles mixed in R718 is used in the secondary circuit of evaporators are shown in Table-(1). to Table-(3) respectively. and it was found that maximum First law efficiency in terms of COP and maximum second law efficiency in terms of exergetic efficiency using ecofriendly R152a and minimum first and second law performance using R410a. Due to flammable nature of R152a, and R290, R600 and R600a which can be used by considering safety measure gives better performance. The R407c, R134a and R404a also gives good performance for replacing R502, R11 and R12 and R22 which produces global warming and ozone depletion

Table: 1. Performance prediction of vapour compression refrigeration system using TiO₂ in R718 in the secondary circuit and Ecofriendly refrigerants in primary circuit for condenser temperature 48°C and evaporator temperature of -5°C

Refrigerant	COP	ETA_II	EDR
R404a	4.36	0.4831	1.694
R410a	2.14	0.2396	6.836
R134a	4.36	0.4880	1.659
R152a	5.169	0.5786	1.143
R507a	4.328	0.4844	1.685
R407c	4.736	0.5301	1.391
R290	4.826	0.5402	1.335
R600	3.605	0.4036	2.429
R600a	4.009	0.4480	1.969
R125	4.03	0.4511	1.949

Table: 2. Variation of performance parameters with condenser temperature in the vapour compression refrigeration system using R-134a in primary circuit and Water in secondary circuit

Condenser temp. (°C)	C.O.P.	η second	EDR
47	2.919	0.3013	0.6987
49	2.825	0.2859	0.7141
51	2.736	0.2713	0.7287
53	2.652	0.2575	0.7425
55	2.572	0.2442	0.7558
57	2.497	0.2316	0.7684

Table: 2. explaining the variation of condenser temperature with first and second law performance parameters and it was observed that increasing condenser temperature reduces first and second law performances and also increases exergy destruction ratio while Table-3 shows the variation of evaporator temperature with first and second law performance parameters. and it was found that first law performance in terms of coefficient of performance and second law efficiency increases and exergy destruction ratio of system decreases

Table: 3. Variation of performance parameters with evaporator temperature in the vapour compression refrigeration system using R-134a in primary circuit and Water in secondary circuit

Evaporator temp. (°C)	C.O.P.	η second	EDR
-8	2.481	0.3105	0.6895
-4	2.767	0.3000	0.7000
0	3.096	0.2854	0.7146
4	3.467	0.2649	0.7451
8	3.880	0.2369	0.7631
12	4.327	0.1998	0.8002

performance have been obtained using R-152a ecofriendly refrigerant and worst performances were found using R-

410a ecofriendly refrigerants. Due to flammable nature of R290, R600, R600a and R152a it is recommended that R407c and R134a is suitable for industrial and commercial applications.

4. Conclusions and Recommendations

In this paper, first law and second law analysis of vapour compression refrigeration systems using multiple evaporators and single compressor and single expansion valve with ecofriendly refrigerants in the system and R718 (water used in secondary circuit with and without nano particles mixed with water used as refrigerant) have been presented. The conclusions of the present analysis are summarized below:

1. The First law efficiency (COP) and Second law efficiency (Exergetic efficiency) of vapour compression refrigeration systems using R718 mixed with nano particles gives better performance is than without nano particles used in the secondary circuit of water cooled evaporator for above mentioned ecofriendly refrigerants.
2. The First law efficiency (COP) and Second law efficiency (Exergetic efficiency) of vapour compression refrigeration systems using R152a refrigerant is higher but is has flammable nature similar to hydrocarbons then safety measures to be taken while using R152a or hydrocarbons (R290, R600 and R600a)
3. The first law performance in terms of Coefficient of performance and second law performance in terms of exergetic efficiency improves using TiO₂ in the secondary evaporator circuit as compared to Al₂O₃ in the secondary circuit
4. COP and exergetic efficiency for R507a and R134a are nearly matching the same values. are better than that for R125.
5. For practical applications R-407c, R134a and R404a, R125 can be used recommended because it is easily available in the market has second law efficiency slightly lesser than R-152a which was not applicable for commercial applications due to flammable nature and R717 is also toxic nature..
6. The first law performance improvement in terms of COP and second law performance in terms of exergetic efficiency (rational efficiency) using TiO₂ is better than using Al₂O₃ with R718 refrigerant in the secondary evaporator circuit.

References

- [1] Ahamed, et.al, A review on exergy analysis of vapor compression refrigeration system. Int J Renewable and sustainable energy reviews, 2011, 15:1593-1600.
- [2] Anand et.al, Exergy analysis and experimental study of a vapour compression refrigeration cycle. Int J Therm Anal Calorim, 2012, 110:961-971.
- [3] Arora, et.al, Theoretical analysis of a vapour compression refrigeration system with R502, R404A and R507A. Int J Refrigeration, 2008, 31:998-1005.
- [4] B.O. Bolaji et.al, Comparative analysis of performance of three ozone-friends HFC refrigerants in a vapor compression refrigerator. Int J Sustainable Energy & Environment, 2011, 2, 61-64

-
- [5] Chopra et.al, Energetic and Exergetic Based Comparison Multiple Evaporators with Compound Compression and Flash Intercooler with Individual or Multiple Throttle Valves International Journal of Advance Research & Innovations (1), 2013, 73-81
- [6] Han, et.al, Cycle performance study on R32/ R125/ R161 as an alternative refrigerant to R407C. Int J Applied Thermal Engineering, 2007, 27:2559-2565
- [7] Mishra et.al, Methods for improving thermal performances of vapour compression Refrigeration system using eleven ecofriendly refrigerants”, ISTE conference on “Technological Universities and Institutions in New Knowledge Age: Future Perspectives and Action plan, ISTE, 149, 9, Delhi Technological University, 2013
- [8] Mishra, Irreversibility Analysis of Multi-Evaporators Vapour Compression Refrigeration Systems Using New and Refrigerants: R134a, R290, R600, R600a, R1234yf, R502, R404a and R152a and R12, R502” International Journal of Advance Research & Innovations International Journal of Advance Research & Innovations, 1, 2013, 180-193
- [9] Mishra, Thermodynamic performance evaluation of multiple evaporators , single compressor ,single expansion valve and liquid vapour heat exchanger in vapour compression refrigeration systems using thirteen ecofriendly refrigerants for reducing global warming and ozone depletion, International Journal of Advance Research & Innovations, 1, 2014, 163-171
- [10] Mishra, Thermodynamic performance evaluation of multiple evaporators , single compressor ,single expansion valve and liquid vapour heat exchanger in vapour compression refrigeration systems using thirteen ecofriendly refrigerants for reducing global warming and ozone depletion, International Research Journal of Sustainable Science & Engineering (Monthly Peer Review Journal , 2(3), 2014, 1-10
- [11] R. Saravanakumar, V. Selladurai, Exergy analysis of a domestic refrigerator using eco-friendly R290/ R600a refrigerant mixture as an alternative to R134a. Int J Therm Anal Calorim. 2013

Microstructured-Core Photonic Crystal Fiber with All Normal Dispersion for Supercontinuum Generation

Than Singh Saini, Ajeet Kumar*, Ravindra Kumar Sinha

TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Department of Applied Physics, Delhi Technological University, Delhi-110 042, INDIA

**Email:ajeetdph@gmail.com*

Abstract: Design of microstructured core photonic crystal fiber in fused silica with all-normal dispersion has been proposed for supercontinuum generation. Structural parameters have been optimized to achieve high nonlinearity and zero dispersion wavelength at 960 nm.

OCIS Codes: 190.0190 Nonlinear optics; 190.4370 Nonlinear optics, fibers

1. Introduction

Photonic crystal fibers (PCFs) or microstructured optical fibers (MOFs) are the exciting research topic among the scientists working in the field of optics and photonics during the last decade because of their unique properties that can not be achieved using conventional optical fibers [1-5]. Scientists are trying very hard to generate new frequency sources for applications in different fields of science and technology. Husakou et al. reported numerical modelling of supercontinuum generation (SCG) in photonic crystal fiber using femtosecond pulses and the role of soliton fission in the spectral broadening was discussed for the first time [6]. The dispersion profile of the propagating mode plays very important role in SCG because it determines the extents to which different spectral components of an ultrashort pulse propagated at different phase velocities in the photonic crystal fiber. PCF with all-normal dispersion is desirable for obtaining the intrinsic coherent property of super continuum generation [7].

In this paper, we present a new design of microstructured-core (MC) photonic crystal fiber (PCF) for nonlinear applications. The structural parameters have been tailored for obtaining all-normal dispersion with zero dispersion wavelength (ZDW) at 960 nm.

2. Design of MC PCF

We present a new photonic crystal fiber structure with three layers of air holes arrange in trigonal lattice pattern in silica glass. One air hole has been removed to construct a core. As shown in Fig. 1, to enhance nonlinearity, we have inserted updoped material nano-rods in core region. In the cladding region, center to center distance of holes is taken as constant and represented by Λ . The diameter of air holes in first and third rings have been considered identical and represented by d_1 , while the diameter of air holes in second ring is comparative larger and denoted by d_2 .

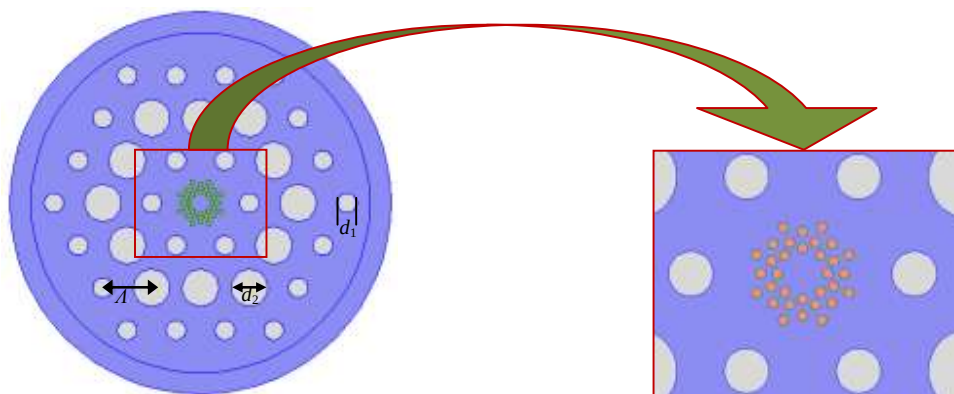


Fig. 1: Proposed microstructured core PCF structure.

3. Results and discussion

The diameters of air holes in second ring in cladding region have been optimized to obtain all-normal flat dispersion profile. It is clear from Fig. 2(a) that the dispersion can be control by adjusting d_2 . When the value of d_2 increases (keeping pitch of the holes (Λ) constant at value of 1.1 μm), the dispersion curve shifts toward the anomalous dispersion region. As illustrated in Fig. 2(b), on optimized parameters, we obtained all normal dispersion profile with zero dispersion wavelength (ZDW) at 960 nm. Proposed microstructured photonic crystal fiber offers very low negative dispersion with approximately $-3 \text{ ps}/(\text{nm} \cdot \text{Km})$ within the wavelength range from 900 – 1050 nm.

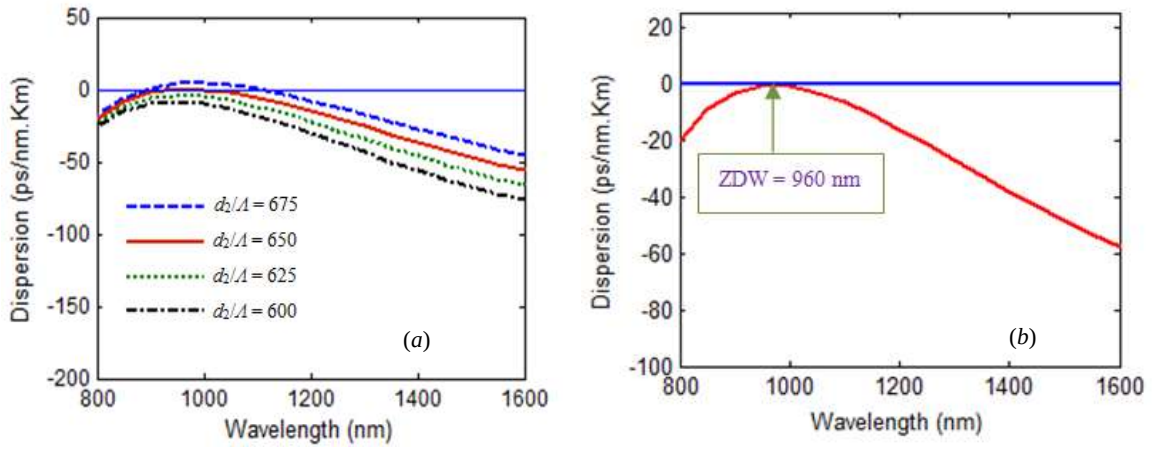


Fig. 2. (a) Influence of parameter d_2 on the dispersion of the proposed PCF design with $d_1/\Lambda = 0.4$ and $\Lambda = 1.1 \mu\text{m}$; **(b)** Dispersion profile of proposed microstructured PCF at optimized parameters with $d_1/\Lambda = 0.4$, $d_2/\Lambda = 0.645$ and $\Lambda = 1.1 \mu\text{m}$.

For the optimized parameters, (i.e. $d_1/\Lambda = d_3/\Lambda = 0.4$, $d_2/\Lambda = 0.645$ and $\Lambda = 1.1 \mu\text{m}$), the spectral variation of nonlinear coefficient and effective mode area have been shown in Fig.3(a). Our simulated results show that the proposed photonic crystal fiber structure offer nonlinear coefficient (γ) as high as $77 \text{ W}^{-1} \text{ Km}^{-1}$ with effective mode area, $A_{\text{eff}} = 2.23 \mu\text{m}^2$ at 960 nm. The electric field distribution in proposed microstructured core PCF has been illustrated in Fig. 3(b).

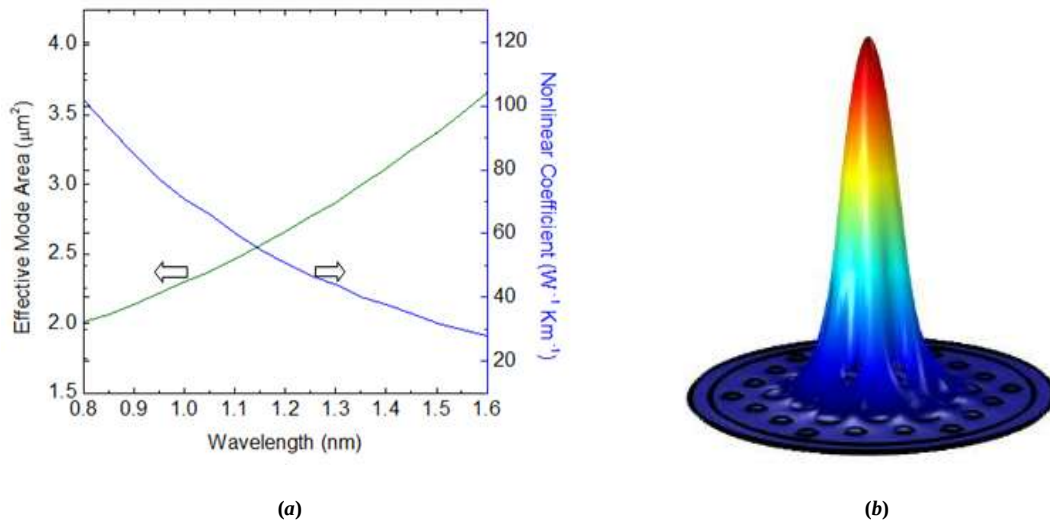


Fig. 3: (a) Spectral variation of nonlinear coefficient and effective mode area of LP_{01} mode with $d_1/\Lambda = d_3/\Lambda = 0.4$, $d_2/\Lambda = 0.645$ and $\Lambda = 1.1 \mu\text{m}$; **(b)** Electric field distribution of LP_{01} mode in proposed PCF at 960 nm with $d_1/\Lambda = d_3/\Lambda = 0.4$, $d_2/\Lambda = 0.645$ and $\Lambda = 1.1 \mu\text{m}$;

4. Conclusions

A new microstructured core photonic crystal fiber design in silica glass has been presented in this paper. The structure parameters have been tailored for all-normal and flat dispersion. A flat curve with dispersion value approximately $-3 \text{ ps}/(\text{nm.km})$ at the wavelength range from 900 – 1050 nm has been achieved. Proposed structure offers nonlinear coefficient as high as $77 \text{ W}^{-1} \text{ Km}^{-1}$ at 960 nm wavelength with $2.23 \mu\text{m}^2$ effective mode area. This new type of photonic crystal fiber can be a good candidate for generating efficient supercontinuum which is applicable for various nonlinear applications such as optical tomography, pump-probe spectroscopy, metrology or non-linear microscopy.

5. Acknowledgement

The authors gratefully acknowledge the (i) TUN-IND bilateral research project, Department of Science and Technology, Govt. of India; and (ii) Initiatives and support towards establishment of the “TIFAC-Center of Relevance and Excellence” in Fiber Optics and Optical Communication at Delhi Technological University (Formerly Delhi College of Engineering) Delhi, through the “Mission REACH” program of Technology Vision-2020 of the Government of India.

6. References

- [1] P. Russell, “Photonic Crystal Fibers,” *Science* **299**, 358-362 (2003).
- [2] R. K. Sinha and S. K. Varshney, “Dispersion properties of photonic crystal fibers”, *Micro. Opt. Technol. Lett.* **37**, 129-132 (2003).
- [3] K. Kishor, R. K. Sinha, A. D. Varshney, J. Singh, “Characterization of specially designed polarization maintaining photonic crystal fiber from far field radiation patterns”, *Opt. Commun.* **283**, 5007-5011 (2010).
- [4] K. Kishor, R. K. Sinha, A. D. Varshney, “Experimental verification of improved effective index method for endlessly single mode photonic crystal fiber” *Optics and Lasers in Engg.* **50**, 182 (2012).
- [5] T. S. Saini, A. Kumar, V. Rastogi, and R. K. Sinha, "Selectively filled large-mode-area photonic crystal fiber for high power applications", in *Proc. SPIE 8847, Photonic Fiber and Crystal Devices: Advances in Materials and Innovations in Device Applications VII*, (2013) 88471Q.
- [6] A. V. Husakou and J. Herrmann, “Supercontinuum generation of higher-order solitons by fission in photonic crystal fibers,” *Phys. Rev. Lett.* **87**, 203901(2001).
- [7] A. Hartung, M. Heidt, H. Bartelt, “Design of all-normal dispersion microstructured optical fibers for pulse-preserving supercontinuum generation,” *Optics Express* **19**, 7742 (2011).

Modified $f(R, T)$ gravity theory and scalar field cosmology

Vijay Singh · C.P. Singh

Received: 17 April 2014 / Accepted: 7 November 2014
© Springer Science+Business Media Dordrecht 2014

Abstract In this paper, we explore the behaviors of scalar field in modified $f(R, T)$ gravity theory within the framework of a flat Friedmann-Robertson-Walker cosmological model. The universe is assumed to be filled with two non-interacting matter sources, scalar field (normal or phantom) with scalar potential and matter contribution due to $f(R, T)$ action. We first explore a model where the potential is a constant, and the universe evolves as a de Sitter type. This model is compatible with phantom scalar field only which gives fine tuning with the recent observations. The model exhibits a wide variety of early time physical phenomena that eventually behaves like a cosmological constant at late times. The model shows transition from decelerated to accelerated expansion of the universe. We also explore a model where the scalar field potential and the scale factor evolve exponentially as a scalar field. This model is compatible with normal scalar field only and describes transition from inflationary to the decelerated phase at early times and quintessence to phantom phase at late times. We constraint our results with the recent observational data and find that some values of parameters are consistent with SNe Ia and $H(z)$ + SNe Ia data to describe accelerated expansion only whereas some one give decelerated and accelerated expansions with $H(z)$, WMAP7 and WMAP7 + BAO + $H(z)$ observational data.

Keywords Modified gravity · Scalar field · Phantom field

1 Introduction

The results from Ia supernova (Riess et al. 1998; Perlmutter et al. 1999; Suzuki et al. 2012), Cosmic Microwave Background (CMB) anisotropy from Planck (Ade et al. 2014), Wilkinson Microwave Anisotropy Probe (WMAP9) (Hinshaw et al. 2013), Large Scale Structures (Spergel et al. 2003), Baryon Acoustic Oscillations (BAO) (Anderson et al. 2013) etc. strongly indicate that we are living in an accelerating universe. These observations also deduce the scenario of a spatially flat, low density universe dominated by some kind of exotic matter, known as dark energy (DE) (Sahni 2004; Frieman et al. 2008; Caldwell and Kamionkowski 2008). The DE is supposed to have the dynamics to produce sufficient acceleration in late time of evolution of the universe. Several approaches have pursued to understand the fundamental nature of DE. The classical correction in Einstein's field equations, i.e., the addition of cosmological constant within General Relativity (GR) stands as the simplest candidate of DE (Padmanabhan 2003). The cosmological constant (Λ) is the most efficient way to generate accelerated expansion but it faces the fine-tuning problem in cosmology (Peebles and Ratra 2003). In order to overcome this serious problem many theorists have proposed a number of other probable candidates of DE such as quintessence (Martin 2008), phantom (Nojiri and Odintsov 2006), k-essence (Chiba et al. 2000), tachyons (Padmanabhan and Chaudhury 2002), Chaplygin gas (Bento et al. 2002) and so on but there is no direct detection of such exotic fluids. The absence of evidence for the existence of DE motivates to explore other possible ways to alleviate the most crucial problem of cosmic acceleration.

One way is the modification in the geometrical part of Einstein-Hilbert action (Magnano et al. 1987). It has been observed that the positive power of R in action supports

V. Singh · C.P. Singh (✉)
Department of Applied Mathematics, Delhi Technological
University (Formerly Delhi College of Engineering), Bawana
Road, Delhi 110 042, India
e-mail: cpsphd@rediffmail.com

V. Singh
e-mail: gtrcosmo@gmail.com

the inflationary epoch whereas the negative power serves as effective DE admitting late time acceleration (Nojiri and Odintsov 2003; Carloni et al. 2005; Carroll et al. 2005). The generalization of such modification causes the substitution of R with an arbitrary function of Ricci scalar, known as $f(R)$ theory of gravity. The $f(R)$ modified theory produces both cosmic inflation and mimic behavior of DE including present cosmic acceleration (De Felice and Tsujikawa 2005; Capozziello and Francaviglia 2008; Sotiriou and Faraoni 2010; Nojiri and Odintsov 2011). Amendola et al. (2007) have discussed the cosmologically viable conditions in $f(R)$ theory which describe the DE models. Some other modified theories such as Gauss-Bonnet $f(G)$ gravity (Nojiri and Odintsov 2005; Nojiri et al. 2006) and $f(T)$ theories (T is the torsion) (Ferraro and Fiorini 2007; Linder 2010), etc. have proposed to describe the DE models.

Harko et al. (2011) have proposed $f(R, T)$ modified theory of gravity, where the gravitational Lagrangian is modified by an arbitrary function of the Ricci scalar R and the trace of the stress-energy tensor T . The authors have intimated that the coupling between matter and geometry leads to the model which depends on a source term representing the variation of the matter-stress-energy tensor with respect to the metric. In addition, the new matter and time-dependent terms in the gravitational field equations play the role of effective cosmological constant. The reconstruction of cosmological models based on modified theory of gravity is one of the interesting issue in cosmology. Therefore, some authors have attempted to reconstruct viable $f(R, T)$ gravity models to resolve some of the open challenges in cosmology. In reconstruction method one assumes that the expansion history is known exactly and inverts the field equations to deduce what class of the modified theory give rise to a given FRW model. The existence of such solutions is particularly relevant because it represents all possible cosmological evolution. Houndjo (2012) has reconstructed $f(R, T) = f(R) + f(T)$ using an auxiliary scalar field with two known forms of scale factor and got transition from matter-dominated phase to an accelerated phase. Jamil et al. (2012) have reconstructed $f(R, T)$ gravity with dust fluid which reproduces the Λ -CDM model, Einstein's static Universe, de Sitter's Universe, phantom and non-phantom models. Houndjo and Piattella (2012) have reconstructed $f(R, T)$ dark matter and holographic DE models with dust which describe the same expansion history generated in GR. Recently, Singh and Singh (2014) have reconstructed power-law and de Sitter $f(R, T)$ gravity models with perfect fluid and found viable solutions.

Usually, the universe is modeled with a perfect fluid but the observations suggest that cosmic dynamics cannot be explained by standard matter (Weinberg 1989). The observational results lead to the search of some kind of matter fields which would generate sufficient negative pressure to drive late time cosmic acceleration. One of such

matter fields is the scalar field which goes slowly down to its potential such that the potential term dominates over the kinetic term and produces necessary negative pressure causing acceleration in late times (Caldwell et al. 1998; Sahni 2004).

Most of the studies have discussed with a minimally coupled scalar field representing quintessence where the equation of state (EoS) parameter ω is greater than -1 . The Hubble parameter measurement from the Hubble Space Telescope (HST) and luminosity measurements of type Ia supernovae restrict the EoS parameter of DE in the interval $-1.38 < \omega < -0.82$ (Melchiorri et al. 2003). Therefore, it is quite possible that the universe has a phantom phase at late times where ω is less than -1 (Caldwell 2000; Carroll et al. 2003; Singh et al. 2003). The measurements from Ia supernova (Huterer and Cooray 2005), Gold with HST and Supernova Legacy Survey (Alam et al. 2007), and the latest SNe with the WMAP5 and Sloan Digital Sky Survey (Zhao and Zhang 2010) are in good agreement to describe phantom era. More recently, the results from Planck Collaboration (Ade et al. 2014) and first Panoramic Survey Telescope and Rapid Response System (Scolnic et al. 2014) motivate to concentrate specially on the value of ω . The principle tool to find the most precise value of ω at present is the combination of the most mature, well-studied and robust probes of DE: SNe Ia, BAO and CMB anisotropies.

The purpose of this paper is to study the scalar field cosmological models in modified $f(R, T)$ gravity within the framework of a flat Friedmann-Robertson-Walker metric. We assume that the universe contains two matter sources, one is scalar field (normal or phantom) with scalar potential and another one is the matter contribution due to $f(R, T)$ gravity. We reconstruct a particular form of $f(R, T) = R + 2f(T)$ with two known forms of evolution of the universe, namely de Sitter model and power-law cosmology to explore the expansion history of the universe. In de Sitter model we consider a flat potential whereas in power-law model the scale factor and potential are assumed to be evolve exponentially with scalar field. The cosmic evolution of each model is traced out through EoS parameter of matter due to $f(R, T)$ gravity, effective EoS parameter for effective matter and deceleration parameter. We find out the exact value of the parameters in model with exponential potential using the constraints from $H(z)$, SNe Ia, $H(z) + \text{SNe Ia}$ (Gumjudpai 2013), WMAP7 and WMAP7 + BAO + $H(z)$ (Rani et al. 2014) observations for power-law cosmology.

The paper is organized as follows. In Sect. 2, we present a brief review of scalar field cosmology in modified $f(R, T)$ theory of gravity. Section 3 is divided into two Sects. 3.1 and 3.2. The reconstruction of $f(R, T)$ gravity for de Sitter model with constant potential and its physical interpretation are presented in Sect. 3.1. Section 3.2 is devoted for the reconstruction $f(R, T)$ with exponential scale factor and exponential scalar potential. We also discuss the consistency of

this model with the observations in this subsection. Finally, the findings of the models are summarized in Sect. 4.

2 Scalar field cosmology in $f(R, T)$ theory of gravity

Scalar fields have been presented in literature on gravity and cosmology for many decades. It is widely recognized that a minimally coupled scalar field in GR can be represented as a perfect fluid (see, e.g., Madsen 1985; 1988). From the conceptual point of view a scalar field and a perfect fluid are very different system; a fluid is obtained by microscopic quantities associated with its constituent particles, and the fluid laws can be obtained via a kinetic theory using microscopic model of fluid particles and of their interactions. A scalar field, instead, is more fundamental and does not derive an average; it is not made of point like-particles and does not need an average to be defined- it is always a continuum. As a consequence, scalar field stress-energy tensor involves first order derivatives of scalar field, while the perfect fluid stress energy tensor does not involve derivatives explicitly but can be described using only the energy density, pressure and four velocity field. Hence, the fluid description of a scalar field is not an identification, but only a convenient correspondence for formal purposes. In the early universe, scalar field drives super-fast expansion in inflationary scenario whereas it could be responsible to the present acceleration in various models of DE at late times (Copeland et al. 2006). Chimento et al. (1999) and Paul (1999) have studied some of the most intriguing aspects of our universe containing a self-interacting scalar field possessing an interaction potential as the only matter source. Recently, Harko et al. (2014) have presented several exact cosmological solutions with scalar field based on the mathematical analysis of the Ricci type equation in GR. Therefore, the theoretical and observational investigations of scalar field models are an essential task in cosmology.

In this paper, we consider a minimally coupled scalar field ϕ in a gravitational action self interacting through a potential $V(\phi)$ in $f(R, T)$ gravity which is given as

$$S = \frac{1}{2} \int [f(R, T) + 2\mathcal{L}_\phi] \sqrt{-g} d^4x, \quad (1)$$

where $\mathcal{L}_\phi$ corresponds to the matter Lagrangian of scalar field, $f(R, T)$ is an arbitrary function of the Ricci scalar curvature R , and the trace T of the energy-momentum tensor. We use the system of units $8\pi G = 1 = c$. The energy-momentum tensor of matter source is given as

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L}_\phi)}{\delta g^{\mu\nu}}, \quad (2)$$

and its trace is defined by $T = g^{\mu\nu}T_{\mu\nu}$. We consider that the term $\mathcal{L}_\phi$ accounts for the scalar field content of the universe

and depends on the metric tensor components $g_{\mu\nu}$, and not on its derivatives. Therefore, Eq. (2) simplifies to

$$T_{\mu\nu} = g_{\mu\nu}\mathcal{L}_\phi - 2\frac{\partial\mathcal{L}_\phi}{\partial g^{\mu\nu}}. \quad (3)$$

Variation of action (1) with respect to metric tensor results the field equations of $f(R, T)$ gravity in the form

$$\begin{aligned} f_R(R, T)R_{\mu\nu} - \frac{1}{2}f(R, T)g_{\mu\nu} + (g_{\mu\nu}\square - \nabla_\mu\nabla_\nu)f_R(R, T) \\ = T_{\mu\nu} - f_T(R, T)(T_{\mu\nu} + \ominus_{\mu\nu}), \end{aligned} \quad (4)$$

where f_R and f_T denote the derivatives of $f(R, T)$ with respect to R and T , respectively. As usual notations ∇_μ is covariant derivative, $\square \equiv \nabla_\mu\nabla^\mu$ is the d'Alembert operator and $\ominus_{\mu\nu}$ is defined by

$$\ominus_{\mu\nu} \equiv g^{\alpha\beta} \frac{\delta T_{\alpha\beta}}{\delta g^{\mu\nu}}, \quad (5)$$

which by use of (3) becomes

$$\ominus_{\mu\nu} = -2T_{\mu\nu} + g_{\mu\nu}\mathcal{L}_\phi - 2g^{\alpha\beta} \frac{\partial^2\mathcal{L}_\phi}{\partial g^{\mu\nu}\partial g^{\alpha\beta}}. \quad (6)$$

Since, the field equations of $f(R, T)$ theory depend on $\ominus_{\mu\nu}$, i.e., on the physical nature of the matter source. Therefore, a number of models corresponding to various forms of $f(R, T)$ may be generated depending on the nature of the matter source. Also, the equations of $f(R, T)$ gravity are much complicated even for FRW metric as compared to GR. Therefore, it is very difficult to reconstruct a general form of $f(R, T)$ or to solve the field equations, in general. Consequently, most of the works in $f(R, T)$ gravity are carried out by assuming a number of suitable forms of $f(R, T)$, such as $f(R, T) = R + 2f(T)$, $f(R, T) = \mu f_1(R) + \nu f_2(T)$, ($f_1(R)$ and $f_2(T)$ are arbitrary functions of R and T , and μ and ν are real constants), $f(R, T) = Rf(T)$, etc., (Harko et al. 2011; Jamil et al. 2012; Houndjo 2012; Houndjo and Piattella 2012; Houndjo et al. 2013; Chakraborty 2013). In this paper we consider the following simplest particular form (Harko et al. 2011):

$$f(R, T) = R + 2f(T), \quad (7)$$

i.e. the action is given by the same Einstein-Hilbert of GR plus a function of T . The term $2f(T)$ in the gravitational action modifies the gravitational interaction between matter and curvature. It is to be noted that Jamil et al. (2012) have reconstructed the same form of $f(R, T)$ described in (7) by treating R as a constant. However, in the present work, we consider R as a function of cosmic time t and try to find a general form of $f(T)$ from known scale factors.

Using (7) into (4), we get

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = T_{\mu\nu} - 2(T_{\mu\nu} + \ominus_{\mu\nu})f'(T) + f(T)g_{\mu\nu}, \quad (8)$$

which is considered as the field equations of $f(R, T)$ gravity. Here, a prime denotes derivative with respect to the argument. The assumption (7) is particularly interesting choice. In order to compare (8) with Einstein's, we find that the gravitational field equations (8) can be recast in such a form that the higher order corrections coming both from the geometry, and from matter-geometry coupling, provide an energy-momentum tensor of geometrical and matter origin, describing an effective source term on right hand side of (8).

In what follows, we present the field equations for a spatially homogeneous and isotropic model in the presence of scalar field to reconstruct $f(R, T)$ given in (7) with two known models.

3 Model and field equations

A spatially flat Friedmann-Robertson-Walker (FRW) line element is given by

$$ds^2 = dt^2 - a^2(t)[dr^2 + r^2(d\theta^2 + \sin^2\theta d\vartheta^2)], \quad (9)$$

where $a(t)$ is the scale factor.

We consider that the universe is filled with scalar field minimally coupled to gravity. Therefore, the energy-momentum tensor of a scalar field ϕ with self-interacting scalar field potential $V(\phi)$ has the form

$$T_{\mu\nu} = \epsilon\phi_{,\mu}\phi_{,\nu} - g_{\mu\nu}\left(\frac{\epsilon}{2}g^{\rho\sigma}\phi_{,\rho}\phi_{,\sigma} - V(\phi)\right), \quad (10)$$

where $\epsilon = \pm 1$ correspond to normal (quintessence) and phantom scalar fields, respectively. The trace of the energy-momentum tensor $T = g^{\mu\nu}T_{\mu\nu}$ in a homogeneous and isotropic model is given by

$$T = -\epsilon\dot{\phi}^2 + 4V(\phi), \quad (11)$$

where a dot denotes the derivative with respect to t .

The matter Lagrangian of the scalar field is given by

$$\mathcal{L}_\phi = -\left(\frac{1}{2}\epsilon\dot{\phi}^2 - V(\phi)\right). \quad (12)$$

Using (12) into (6) the tensor $\ominus_{\mu\nu}$ becomes

$$\ominus_{\mu\nu} = -2T_{\mu\nu} - g_{\mu\nu}\left(\frac{1}{2}\epsilon\dot{\phi}^2 - V(\phi)\right). \quad (13)$$

The field equations (8) for (9), (10) and (13) yield

$$3H^2 = \rho_\phi + \rho_f, \quad (14)$$

$$2\dot{H} + 3H^2 = -(p_\phi + p_f), \quad (15)$$

where $H = \dot{a}/a$ is the Hubble parameter. Here, ρ_ϕ and p_ϕ are the energy density and pressure of scalar field, respectively, which are given as

$$\rho_\phi = \frac{1}{2}\epsilon\dot{\phi}^2 + V(\phi), \quad (16)$$

$$p_\phi = \frac{1}{2}\epsilon\dot{\phi}^2 - V(\phi), \quad (17)$$

and ρ_f and p_f respectively are the energy density and pressure due to the modified gravity as described by

$$\rho_f = 2\epsilon\dot{\phi}^2 f'(T) + f(T), \quad (18)$$

$$p_f = -f(T). \quad (19)$$

We consider that the scalar field and matter contribution due to $f(R, T)$ are non-interacting and together represent an effective matter. The EoS parameter of scalar field and matter due to $f(R, T)$ are defined by $\omega_\phi = p_\phi/\rho_\phi$ and $\omega_f = p_f/\rho_f$, respectively.

The Klein-Gordon equation obeyed by the scalar field is given as

$$\epsilon\ddot{\phi} + 3\epsilon H\dot{\phi} + \frac{dV(\phi)}{d\phi} = 0. \quad (20)$$

Now, we reconstruct $f(R, T)$ function of the form considered in (7) for known history of evolution of the universe in the following models.

3.1 Model with constant potential

Let us explore a model where the scalar field potential is a constant, i.e., $V(\phi) = V_0$, and the universe behaves like a de Sitter model with an exponential expansion

$$a(t) = a_0 e^{H_0 t}, \quad (21)$$

where a_0 and H_0 are positive constants. This form of scale factor gives a constant value of Hubble parameter

$$H(t) = H_0. \quad (22)$$

Using (11), (16), (18), and (22) into (14), we obtain

$$2(4V_0 - T)f'(T) + f(T) + \frac{1}{2}(4V_0 - T) + V_0 - 3H_0^2 = 0, \quad (23)$$

which gives

$$f(T) = 3H_0^2 + V_0 - \frac{T}{2} + T_1\sqrt{2(T - 4V_0)}, \quad (24)$$

where T_1 is a constant of integration. We must have $T - 4V_0 > 0$ for a real solution and one may observe from (11) that $T - 4V_0 > 0$ for $\epsilon = -1$. Therefore, this model is compatible with phantom scalar field cosmology only.

Now, the $f(R, T)$ in (7) takes the form

$$f(R, T) = R + 2(3H_0^2 + V_0) - T + 2T_1\sqrt{2(T - 4V_0)}. \quad (25)$$

Let us discuss the behaviors of some physical quantities and cosmological parameters of the present model.

Using (22) into (20), we get

$$\phi(t) = \phi_1 - \frac{\phi_0 e^{-3H_0 t}}{3H_0}, \quad (26)$$

where ϕ_0 and ϕ_1 are constants of integration. We observe that the scalar field ϕ increases with time and becomes constant as $t \rightarrow \infty$. Therefore, the kinetic term $\dot{\phi}^2/2$ decreases with time.

The energy density and pressure of scalar field are respectively given by

$$\rho_\phi = V_0 - \frac{1}{2}\phi_0^2 e^{-6H_0 t}, \quad (27)$$

$$p_\phi = -\left(V_0 + \frac{1}{2}\phi_0^2 e^{-6H_0 t}\right). \quad (28)$$

For reality of any cosmological model where the energy density must be positive, i.e., $\rho_\phi > 0$, we must have $\phi_0 < \sqrt{2V_0}$. We observe that ρ_ϕ starts with a finite value $V_0 - (\phi_0^2/2)$ at $t = 0$, which shows that the model avoids singularity. It increases with time and approaches to a constant value V_0 as $t \rightarrow \infty$. We also observe that the pressure is always negative, which increases during the evolution and becomes constant ($-V_0$) at late times. In other words, we can say that the potential dominates over the kinetic energy at late times. The EoS parameter ω_ϕ starts from $\omega_\phi < -1$ and approaches to $\omega_\phi = -1$ as $t \rightarrow \infty$. Therefore, we conclude that the model describes the behavior of a phantom scalar field that eventually behaves like a cosmological constant at late times.

The energy density and pressure of matter due to $f(R, T)$ gravity are respectively given by

$$\rho_f = 3H_0^2 - V_0 + \frac{1}{2}\phi_0^2 e^{-6H_0 t}, \quad (29)$$

$$p_f = -(3H_0^2 - V_0) + \frac{1}{2}\phi_0^2 e^{-6H_0 t} - \sqrt{2\phi_0^2 T_1} e^{-3H_0 t}. \quad (30)$$

Again, we must have $V_0 \leq 3H_0^2$ for reality of the model. We observe that ρ_f starts from a finite constant value $3H_0^2 - V_0 + (\phi_0^2/2)$, which also asserts that the model is non-singular. It decreases with time and approaches to a positive constant value $3H_0^2 - V_0$ as $t \rightarrow \infty$. The pressure p_f remains always negative and increases with time.

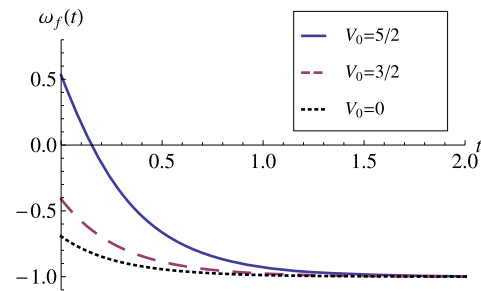


Fig. 1 ω_f versus t with $T_1 = -1$, $H_0 = 1$ and $\phi_0 = 0.5$

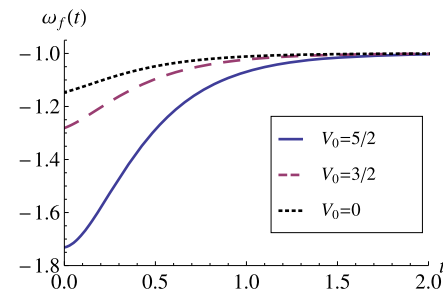


Fig. 2 ω_f versus t with $T_1 = 1$, $H_0 = 1$ and $\phi_0 = 0.5$

We may observe that the matter due to $f(R, T)$ gravity exhibits a wide variety of early time behaviors with different set of parameters which are consistent with the constraint $V_0 < 3H_0^2$. Figure 1 plots the graph of ω_f versus time t for a negative value of T_1 and different values of potential. From this figure we observe that ω_f changes its sign from positive to negative for large values of potential whereas it remains negative for small values of potential even though for zero potential. It means that the different set of parameters exhibits the behavior from ordinary matter to a normal scalar field (quintessence) and eventually behaving like a cosmological constant. Therefore, the matter due to $f(R, T)$ provides transition from decelerated to accelerated phase for some set of parameters. However, the universe accelerates throughout the evolution for some other set of parameters.

The behavior of ω_f for a positive value of T_1 with different sets of parameters under the constraint $V_0 < 3H_0^2$ is shown in Fig. 2. The figure shows that the matter due to $f(R, T)$ gravity always behaves like phantom matter ($\omega_f < -1$) during the evolution of the universe and cosmological constant at late times. Thus, the contribution of positive coefficient of T_1 accelerates the universe throughout the evolution.

Let us consider the case where $V_0 = 3H_0^2$. The EoS parameter has the form

$$\omega_f = 1 - \frac{2\sqrt{2}T_1}{\phi_0} e^{3H_0 t}, \quad (31)$$

which is free from scalar field potential but depends on ϕ_0 . It is clear that $\omega_f \leq 1$ for $T_1 \geq 0$ and $\omega_f > 1$ for $T_1 < 0$.

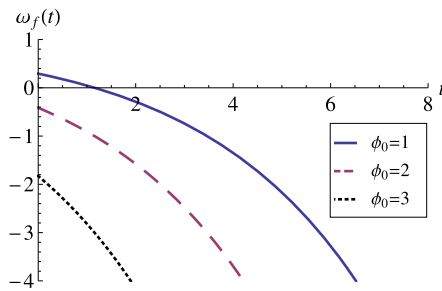


Fig. 3 ω_f versus t with $T_1 = 1$ and $H_0 = 0.1$

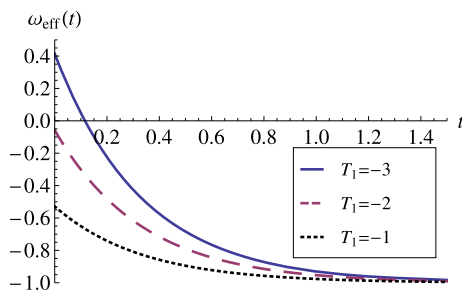


Fig. 4 ω_{eff} versus t with $H_0 = 1$, $\phi_0 = 1$ and with some negative values of T_1

The behavior of ω_f with a positive value of T_1 and for some different values of ϕ_0 is shown in Fig. 3. The EoS parameter shows a transition from a positive to negative value for small values of ϕ_0 whereas it gives negative value for sufficiently large values of ϕ_0 . Therefore, we observe that some sets of parameters give the behavior from ordinary matter to phantom scalar field crossing the phantom dividing line.

Note that instead of ω_f if we consider the effective EoS parameter ω_{eff} , then

$$w_{eff} = \frac{p_\phi + p_f}{\rho_\phi + \rho_f} = -1 - \frac{\sqrt{2\phi_0^2 T_1} e^{-3H_0 t}}{3H_0^2}, \quad (32)$$

which is also independent of scalar field potential. Figure 4 plots ω_{eff} versus time t for some negative values of T_1 with a particular set of H_0 and ϕ_0 . The effective EoS parameter ω_{eff} makes transition from some positive to negative values for some particular set of parameters. However, it remains negative ($-1 < \omega_{eff} < 0$) for some other set of parameters. Therefore, we conclude that this model describes transition from decelerated to accelerated phase for some set of parameters whereas it exhibits acceleration throughout the evolution for some other set of parameters and eventually behaves like a cosmological constant at late times.

The behavior of ω_{eff} for some positive values of T_1 and a particular set of values of H_0 and ϕ_0 is shown in Fig. 5. It is clear that the effective matter behaves as phantom matter during the evolution of the universe and cosmological

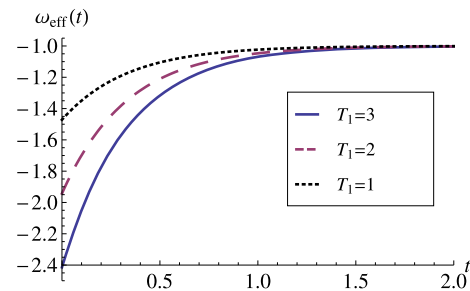


Fig. 5 ω_{eff} versus t with $H_0 = 1$, $\phi_0 = 1$ and with some positive values of T_1

constant at late times for this model. It is to be noted that $\omega_{eff} = -1$ for $T_1 = 0$.

3.2 Model with exponential potential

Models with an exponential scalar field potential arise naturally in alternative theories of gravity (Liddle and Scherrer 1998). Therefore, we assume the potential of the form (Burd and Barrow 1998; Coley and Goliath 2000)

$$V(\phi) = V_0 e^{-\beta\phi}, \quad (33)$$

where V_0 and β are non-negative constants. The parameter β has the dimension of inverse of mass as ϕ has the dimension of mass.

We further assume that the scale factor also evolves exponentially with scalar field (Sen and Banerjee 1997; Singh and Singh 2011) as

$$a(t) = a_0 e^{\alpha\phi}, \quad (34)$$

where a_0 and α are positive constants.

Using (33) and (34), Eq. (20) can be rewritten as

$$\frac{d}{dt}(\dot{\phi}^2 e^{6\alpha\phi}) = \frac{2V_0\beta}{(6\alpha - \beta)\epsilon} \frac{d}{dt}(e^{\phi(6\alpha - \beta)}), \quad (35)$$

which on integration gives

$$\dot{\phi}^2 = \frac{2V_0\beta e^{-\beta\phi}}{(6\alpha - \beta)\epsilon}, \quad (36)$$

where the integration constant is taken as zero for the sake of simplicity. Further, integration of (36) gives

$$\phi(t) = \frac{2}{\beta} \ln\left(\frac{\phi_0\beta}{2} \pm \frac{\beta\sqrt{A}}{2}t\right), \quad \beta \neq 0, \quad (37)$$

where $A = \frac{2V_0\beta}{(6\alpha - \beta)\epsilon}$ and ϕ_0 is a constant of integration. The real solution exists provided $A > 0$, i.e., $(6\alpha - \beta) > 0$ for $\epsilon = 1$ (normal scalar field) and $(6\alpha - \beta) < 0$ in case of $\epsilon = -1$ (phantom scalar field).

Consequently, the scale factor defined in (34) takes the form

$$a(t) = a_0 \left(\frac{\phi_0 \beta}{2} \pm \frac{\beta \sqrt{A}}{2} t \right)^{\frac{2\alpha}{\beta}}. \quad (38)$$

Let us assume that the universe originates from $a(0) = 0$, which implies $\phi_0 = 0$. Therefore, Eqs. (37) and (38) for an expanding universe respectively take the form

$$\phi(t) = \frac{2}{\beta} \ln \left(\frac{\beta \sqrt{A}}{2} t \right), \quad (39)$$

$$a(t) = a_0 \left(\frac{\beta \sqrt{A}}{2} t \right)^{\frac{2\alpha}{\beta}}. \quad (40)$$

Equation (39) shows that the scalar field $\phi(t)$ increases with time. The later expression gives power-law expansion of the universe. The Hubble parameter H and the deceleration parameter $q = -a\ddot{a}/\dot{a}^2$ respectively become

$$H = \frac{2\alpha}{\beta} \frac{1}{t}, \quad (41)$$

$$q = \frac{\beta}{2\alpha} - 1. \quad (42)$$

It is to be noted that $q > 0$ corresponds to a decelerated model and $q < 0$ to accelerated model. Equation (42) tells that the universe decelerates for $\beta/2\alpha > 1$, inflates marginally for $\beta/2\alpha = 1$ and accelerates for $\beta/2\alpha < 1$ in the present model.

The scalar potential $V(\phi)$ and the trace T respectively take the form

$$V(t) = \frac{2\epsilon(6\alpha - \beta)}{\beta^3} \frac{1}{t^2}, \quad (43)$$

$$T = \frac{12\epsilon(4\alpha - \beta)}{\beta^3} \frac{1}{t^2}. \quad (44)$$

Equation (43) shows that the scalar potential $V(t)$ decreases with time and tends to zero as $t \rightarrow \infty$.

The energy density and pressure of scalar field are respectively given by

$$\rho_\phi = \frac{12\epsilon\alpha}{\beta^3 t^2}, \quad (45)$$

$$p_\phi = \frac{4\epsilon(\beta - 3\alpha)}{\beta^3 t^2}. \quad (46)$$

As the energy density must be positive for reality of the model, i.e., $\rho_\phi > 0$, which is possible for $\epsilon = 1$. Therefore, the assumptions (33) and (34) are suitable to describe only normal scalar field model. The EoS parameter $\omega_\phi = p_\phi/\rho_\phi$ for scalar field becomes

$$\omega_\phi = \frac{2}{3} \left(\frac{\beta}{2\alpha} \right) - 1, \quad (47)$$

which is a constant. From (42) and (47), we have

$$q = \frac{1 + 3\omega_\phi}{2}, \quad (48)$$

which is the well known relation between deceleration parameter and EoS parameter in the standard flat model.

From (41) and (42), we get

$$H(t) = \frac{1}{(1+q)t}. \quad (49)$$

The scale factor in terms of redshift z is given by

$$a = \frac{a_0}{1+z}, \quad (50)$$

and therefore, the Hubble parameter in terms of z can be written as

$$H(z) = H_0(1+z)^{1+q}, \quad (51)$$

which shows that the evolution of the universe governs by the parameters H_0 and q .

Some authors have constrained on the parameters H_0 and q for the power-law cosmology. Kumar (2012) has constrained the parameters H_0 and q using 14 points of $H(z)$ data and 557 data points of SNe Ia data. Gumjudpai (2013) has constrained on the parameters H_0 and $1/1+q$ for WMAP7 and WMAP7 + BAO + $H(z)$ data sets. Recently, Rani et al. (2014) have also constrained these parameters for the latest 29 points of $H(z)$ data and 580 data points from Union 2.1 SNe Ia data. Since, in the present model we have two parameters, α and β , therefore, it is not possible to constraint separately. Using the values of q (Rani et al. 2014) and $1/1+q$ (Gumjudpai 2013), the values of $\beta/2\alpha$ and the corresponding evolution of the universe are given in Table 1. We observe that our model is best fitted with SNe Ia and $H(z)$ + SNe Ia for accelerated universe including error bars where as decelerated and accelerated universe are agreed with others observational data.

Using (16), (18), (39), (40), (43) and (44), Eq. (14) can be written as

$$2\beta T f'(T) + 3(4\alpha - \beta) f(T) - 3\alpha(\alpha\beta - 1)T = 0, \quad (52)$$

which is integrated to give

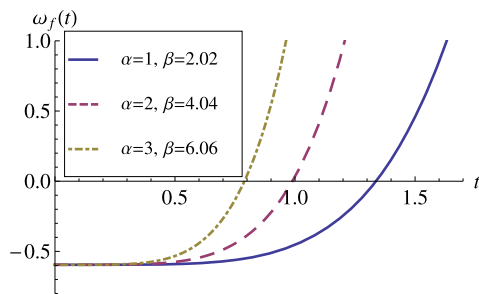
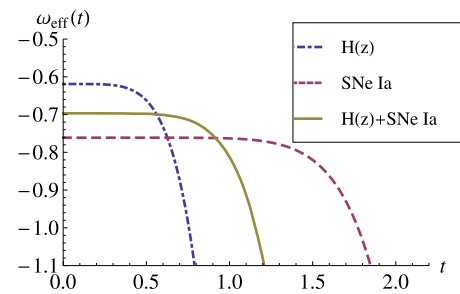
$$f(T) = \frac{3\alpha(\alpha\beta - 1)}{12\alpha - \beta} T + T_2 T^{-\frac{3(4\alpha - \beta)}{2\beta}}, \quad (53)$$

where T_2 is a constant of integration. The function $f(R, T)$ in (7) takes the form

$$f(R, T) = R + \frac{6\alpha(\alpha\beta - 1)}{12\alpha - \beta} T + 2T_2 T^{-\frac{3(4\alpha - \beta)}{2\beta}}. \quad (54)$$

Table 1 Values of $\beta/2\alpha$ and corresponding evolution of universe

Data	q	$\frac{1}{1+q}$	$\frac{\beta}{2\alpha}$	Nature of $\frac{\beta}{2\alpha}$ with error bars	Evolution of universe
$H(z)$	$-0.04^{+0.05}_{-0.05}$	–	$0.96^{+0.05}_{-0.05}$	> 1 (with +ve error) < 1 (with –ve error)	Decelerated expansion Accelerated expansion
SNe Ia	$-0.36^{+0.05}_{-0.05}$	–	$0.64^{+0.05}_{-0.05}$	< 1 (with +ve error) < 1 (with –ve error)	Accelerated expansion Accelerated expansion
$H(z) + \text{SNe Ia}$	$-0.21^{+0.04}_{-0.04}$	–	$0.79^{+0.04}_{-0.04}$	< 1 (with +ve error) < 1 (with –ve error)	Accelerated expansion Accelerated expansion
WMAP7	–	$0.99^{+0.04}_{-0.04}$	$1.01^{+0.04}_{-0.04}$	< 1 (with –ve error) > 1 (with +ve error)	Accelerated expansion Decelerated expansion
WMAP7 + BAO + $H(z)$	–	$0.99^{+0.02}_{-0.02}$	$1.01^{+0.02}_{-0.02}$	< 1 (with –ve error) > 1 (with +ve error)	Accelerated expansion Decelerated expansion

**Fig. 6** ω_f versus t for $T_2 = -1$ and some consistent values of α and β with WMAP7 and WMAP7 + BAO + $H(z)$ observations**Fig. 7** ω_f versus t for $T_2 = 1$, $\alpha = 1$, and some consistent values $\beta = 1.92$ ($H(z)$), $\beta = 1.28$ (SNe Ia) and $\beta = 1.58$ ($H(z) + \text{SNe Ia}$)

The energy density and pressure of the matter due to $f(R, T)$ gravity have the following expressions.

$$\rho_f = \frac{12\alpha(\alpha\beta - 1)}{\beta^3} \frac{1}{t^2}, \quad (55)$$

$$p_f = -\frac{36\alpha(4\alpha - \beta)(\alpha\beta - 1)}{\beta^3(12\alpha - \beta)} \frac{1}{t^2} - T_2 \left[\frac{12(4\alpha - \beta)}{\beta^3} \frac{1}{t^2} \right]^{-\frac{3(4\alpha - \beta)}{2\beta}}. \quad (56)$$

For $\rho_f \geq 0$, i.e., for real model we must have $(\alpha\beta - 1) \geq 0$.

We observe that the EoS parameter of matter due to $f(R, T)$ fluid is compatible to describe decelerated model for negative values of T_2 whereas positive values of T_2 are feasible to describe accelerating universe. Let us examine the behavior of EoS parameter ω_f under the constraints for which our model fits with the observations without including error bars. Figure 6 plots the graph between EoS parameter ω_f and t for decelerated ($T_2 > 0$) model for some values of α and β which satisfy the constraint from WMAP7 and WMAP7 + BAO + $H(z)$ given in Table 1. The figure shows that the EoS parameter starts from $\omega_f = -0.6$ irrespective of sets of parameters and approaches to the positive values,

which shows that the model transits from early inflationary phase to decelerated phase.

Figure 7 plots ω_f versus t for accelerated ($T_2 < 0$) model with $\alpha = 1$ and different values of β which satisfy the constraints from $H(z)$ ($\beta = 1.92$), SNe Ia ($\beta = 1.28$) and joint $H(z) + \text{SNe Ia}$ ($\beta = 1.58$). The graph clearly indicate that ω_f varies from $-1 < \omega_f < 0$ to $\omega_f < -1$. Hence, the model enters from quintessence to phantom phase in this case.

The effective energy density and pressure are respectively given by

$$\rho_{eff} = \frac{12\alpha^2}{\beta^2} \frac{1}{t^2}, \quad (57)$$

$$p_{eff} = -\frac{4(3\alpha - \beta)}{\beta^3} \frac{1}{t^2} - \frac{36\alpha(4\alpha - \beta)(\alpha\beta - 1)}{\beta^3(12\alpha - \beta)} \frac{1}{t^2} - T_2 \left[\frac{12(4\alpha - \beta)}{\beta^3} \frac{1}{t^2} \right]^{-\frac{3(4\alpha - \beta)}{2\beta}}. \quad (58)$$

One may observe that the effective EoS parameter exhibits the same behavior as ω_f .

Thus, the model successfully describes the transition from early inflationary phase to matter dominated era and it also exhibits late time acceleration of the universe providing transition from quintessence to phantom phase.

4 Conclusion

In this work, we have studied modified $f(R, T)$ gravity with scalar field in a flat FRW model. We have considered a particular form $f(R, T) = R + 2f(T)$, which leads to the field equations equivalent to the Einstein field equations with an effective energy momentum tensor containing a sum of normal (phantom) scalar field and the matter contribution due to $f(R, T)$ gravity. We have reconstructed this form of $f(R, T)$ with two well-known scale factors along with the constant and exponential scalar potentials, respectively. In the first model we have explored a model where the potential of the scalar field is constant, and the universe evolves as de Sitter universe with exponential expansion. In the second model, we have discussed a model where the scalar field potential and the scale factor evolve exponentially with the scalar field (normal or phantom), which lead to the power-law expansion of the universe. The behavior of each model has been examined through the deceleration parameter, the EoS parameters of matter contribution of $f(R, T)$ gravity (ω_f) and the effective EoS parameter (ω_{eff}) of matter due to $f(R, T)$ and scalar field. We have compared our models with the observational results of $H(z)$, SNe Ia, $H(z) + \text{SNe Ia}$, WMAP7 and $H(z) + \text{BAO} + \text{WMAP7}$. The summary of the results of both the models are as follows.

In the first model where we have considered a de Sitter expansion and a flat potential, we have found that the solutions are compatible with the phantom scalar field only. The model is free from big-bang singularity and exhibits a wide variety of early time behaviors of the universe with different sets of parameters. The model ultimately describes the behavior of cosmological constant at late times as expected. We have noted from Fig. 1 that ω_f changes its sign from positive to negative for large values of potential with negative value of T_1 whereas it remains negative for small values of potential even for zero potential. It means that the model exhibits transition from ordinary matter to a normal scalar field (quintessence) during early times and eventually behaves like a cosmological constant at late times. Therefore, the matter due to $f(R, T)$ describes evolution from decelerated to accelerated phase for some set of parameters. However, the universe accelerates throughout the evolution for some other set of parameters. In Fig. 2 we have observed that the matter due to $f(R, T)$ gravity always behaves like a phantom matter ($\omega_f < -1$) during the evolution of the universe and becomes cosmological constant at late times for positive values of T_1 . Therefore, the universe accelerates throughout the evolution for any set of parameters with positive values of T_1 . The effective matter also behaves in a similar manner as the matter due to $f(R, T)$ as shown in Figs. 4 and 5.

For the second model where we have assumed scalar potential and scale factor evolving exponentially with the

scalar field, we have found that the model is compatible with the normal scalar field only. The model leads to the power-law expansion of the universe which has a big-bang singularity at $t = 0$. The scalar field increases whereas the scalar potential decreases with time and tends to zero as $t \rightarrow \infty$. We have computed the value of $\beta/2\alpha$ with error bars using existing observational data from $H(z)$, SNe Ia, $H(z) + \text{SNe Ia}$, WMAP7 and WMAP7 + BAO + $H(z)$ available on power-law cosmology as given in Table 1. The behavior of the model is shown by EoS parameter for the parameters which satisfy these constraints. The model exhibits decelerated as well as accelerated universe. The decelerated model describes the early universe and shows transition from inflationary to decelerated phase as shown in Fig. 6. The model shows acceleration at late times and transits from quintessence to phantom phase (see Fig. 7). It is to be noted that these figures have been plotted excluding the error bars. The values of $\beta/2\alpha$ from SNe Ia and $H(z) + \text{SNe Ia}$ observations are well agreed for accelerated model including error bars whereas decelerated and accelerated nature of the universe have been observed with others observational data.

In concluding remark we can say that both the models are suitable to describe a wide variety of early and late time evolution of the universe in $f(R, T)$ theory, some of which give inflation, quintessence or a phantom phase in their respective EoS. One may expect more encouraging work in future on $f(R, T)$ theory to understand the mysterious dark side of our universe.

Acknowledgements The authors are thankful to the referees for their constructive comments and suggestions to improve the manuscript. The authors also express their thanks to Prof. Harko for his valuable suggestions. One of the author VS expresses his sincere thank to University Grants Commission (UGC), India, for providing Senior Research Fellowship (SRF).

References

- Ade, P.A.R., et al. (Planck Collaboration): *Astron. Astrophys.* **571**, A16 (2014). [arXiv:1303.5076](#) [astro-ph.CO]
- Alam, U., Sahni, V., Starobinsky, A.A.: *J. Cosmol. Astropart. Phys.* **0702**, 011 (2007). [arXiv:astro-ph/0612381](#)
- Amendola, L., et al.: *Phys. Rev. D* **75**, 083504 (2007). [arXiv:gr-qc/0612180](#)
- Anderson, L., et al.: *Mon. Not. R. Astron. Soc.* **427**, 3435 (2013). [arXiv:1203.6594](#) [astro-ph.CO]
- Bento, M.C., Bertolami, O., Sen, A.A.: *Phys. Rev. D* **66**, 043507 (2002)
- Burd, A., Barrow, J.D.: *Nucl. Phys. B* **308**, 929 (1998)
- Caldwell, R.R.: *Phys. Lett. B* **545**, 23 (2000). [arXiv:astro-ph/9908168](#)
- Caldwell, R.R., Kamionkowski, M.: *Annu. Rev. Nucl. Part. Sci.* **59**, 397 (2008). [arXiv:0903.0866](#) [astro-ph.CO]
- Caldwell, R.R., Dave, R., Steinhardt, P.J.: *Phys. Rev. Lett.* **80**, 1582 (1998)
- Capozziello, S., Francaviglia, M.: *Gen. Relativ. Gravit.* **40**, 357 (2008). [arXiv:0706.1146](#) [astro-ph]
- Carlton, S., Dunsby, P.K.S., Capozziello, S., Troisi, A.: *Class. Quantum Gravity* **22**, 4839 (2005). [arXiv:gr-qc/0410046](#)

- Carroll, S.M., Hoffman, M., Trodden, M.: Phys. Rev. D **68**, 023509 (2003). [arXiv:astro-ph/0211544](#)
- Carroll, S.M., et al.: Phys. Rev. D **71**, 063513 (2005). [arXiv:astro-ph/0410031](#)
- Chakraborty, S.: Gen. Relativ. Gravit. **45**, 2039 (2013). [arXiv:1212.3050](#) [physics.gen-ph]
- Chiba, T., Okabe, T., Yamaguchi, M.: Phys. Rev. D **62**, 023511 (2000)
- Chimento, L.P., Mendez, V., Zuccata, N.: Class. Quantum Gravity **16**, 3749 (1999)
- Coley, A., Goliath, M.: Class. Quantum Gravity **17**, 2557 (2000)
- Copeland, E.J., Sami, M., Tsujikawa, S.: Int. J. Mod. Phys. D **15**, 1753 (2006)
- De Felice, A., Tsujikawa, S.: Living Rev. Relativ. **13**, 3 (2005). [arXiv:1002.4928](#) [gr-qc]
- Ferraro, R., Fiorini, F.: Phys. Rev. D **75**, 084031 (2007)
- Frieman, J.A., Turner, M.S., Huterer, D.: Annu. Rev. Astron. Astrophys. **46**, 385 (2008). [arXiv:0803.0982](#) [astro-ph]
- Gumjudpai, B.: Mod. Phys. Lett. A **28**, 1350122 (2013)
- Harko, T., Lobo, F.S.N., Nojiri, S., Odintsov, S.D.: Phys. Rev. D **84**, 024020 (2011). [arXiv:1104.2669](#) [gr-qc]
- Harko, T., Lobo, F.S.N., Mak, M.K.: Eur. Phys. J. C **74**, 2784 (2014)
- Hinshaw, G., et al.: Astrophys. J. Suppl. **208**, 19 (2013). [arXiv:1212.5226](#) [astro-ph.CO]
- Houndjo, M.J.S.: Int. J. Mod. Phys. D **21**, 1250003 (2012). [arXiv:1107.3887](#) [astro-ph.CO]
- Houndjo, M.J.S., Piattella, O.F.: Int. J. Mod. Phys. D **2**, 1250024 (2012). [arXiv:1111.4275](#) [gr-qc]
- Houndjo, M.J.S., Batista, C.E.M., Campos, J.P., Piattella, O.F.: Can. J. Phys. **91**, 548 (2013). [arXiv:1203.6084](#) [gr-qc]
- Huterer, D., Cooray, A.: Phys. Rev. D **71**, 023506 (2005)
- Jamil, M., Momeni, D., Raza, M., Myrzakulov, R.: Eur. Phys. J. C **72**, 1999 (2012)
- Kumar, S.: Mon. Not. R. Astron. Soc. **422**, 2532 (2012). [arXiv:1109.6924](#) [gr-qc]
- Liddle, A.R., Scherrer, R.J.: Phys. Rev. D **59**, 023509 (1998)
- Linder, E.V.: Phys. Rev. D **81**, 127301 (2010)
- Madsen, M.S.: Astrophys. Space Sci. **113**, 205 (1985)
- Madsen, M.S.: Class. Quantum Gravity **5**, 627 (1988)
- Magnano, G., Ferraris, M., Francaviglia, M.: Gen. Relativ. Gravit. **19**, 465 (1987)
- Martin, J.: Mod. Phys. Lett. A **23**, 1252 (2008). [arXiv:0803.4076](#) [astro-ph]
- Melchiorri, A., Mersini, L., Odman, C.J., Trodden, M.: Phys. Rev. D **68**, 043509 (2003). [arXiv:astro-ph/0211522](#)
- Nojiri, S., Odintsov, S.D.: Phys. Rev. D **68**, 123512 (2003)
- Nojiri, S., Odintsov, S.D.: Phys. Lett. B **631**, 1 (2005). [arXiv:hep-th/0508049](#)
- Nojiri, S., Odintsov, S.D.: Phys. Lett. B **632**, 597 (2006). [arXiv:hep-th/0506212](#)
- Nojiri, S., Odintsov, S.D.: Phys. Rep. **505**, 59 (2011). [arXiv:1011.0544](#) [gr-qc]
- Nojiri, S., Odintsov, S.D., Sami, M.: Phys. Rev. D **74**, 046004 (2006). [arXiv:hep-th/0605039](#)
- Padmanabhan, T.: Phys. Rep. **380**, 235 (2003). [arXiv:hep-th/0212290](#)
- Padmanabhan, T., Chaudhury, T.R.: Phys. Rev. D **66**, 081301 (2002)
- Paul, B.C.: Pramana J. Phys. **53**, 833 (1999)
- Peebles, P.J.E., Ratra, B.: Rev. Mod. Phys. **75**, 559 (2003)
- Perlmutter, S., et al.: Astrophys. J. **517**, 565 (1999). [arXiv:astro-ph/9812133](#)
- Rani, S., Altaibayeva, A., Shahalam, M., Singh, J.K., Myrzakulov, R.: [arXiv:1404.6522](#) [gr-qc] (2014)
- Riess, A.G., et al.: Astrophys. J. **116**, 1009 (1998). [arXiv:astro-ph/9805201](#)
- Sahni, V.: Lect. Notes Phys. **653**, 141 (2004). [arXiv:astro-ph/0403324](#)
- Scolnic, D., et al.: Astrophys. J. (2014). [arXiv:1310.3824](#) [astro-ph.CO]
- Sen, S., Banerjee, N.: Astrophys. Space Sci. **254**, 133 (1997)
- Singh, C.P., Singh, V.: Mod. Phys. Lett. A **26**, 1495 (2011)
- Singh, C.P., Singh, V.: Gen. Relativ. Gravit. **46**, 1696 (2014)
- Singh, P., Sami, M., Dadhich, N.: Phys. Rev. D **68**, 023522 (2003). [arXiv:astro-ph/0305110](#)
- Sotiriou, T.P., Faraoni, V.: Rev. Mod. Phys. **82**, 451 (2010). [arXiv:0805.1726](#) [gr-qc]
- Spergel, D.N., et al.: Astrophys. J. Suppl. **148**, 175 (2003). [arXiv:astro-ph/0302209](#)
- Suzuki, N., et al.: Astrophys. J. **746**, 85 (2012). [arXiv:1105.3470](#) [astro-ph.CO]
- Weinberg, S.: Rev. Mod. Phys. **61**, 1 (1989)
- Zhao, G.-B., Zhang, X.-M.: Phys. Rev. D **81**, 043518 (2010). [arXiv:0908.1568](#) [astro-ph.CO]

Myconanosomes: Antibacterial, antifungal and immunomodulatory activities of silver nanoparticles synthesized from *Alternaria brassicae* (KF934409)

Faria Fatima¹, Neelam Pathak¹, Sarika Singh², Smita Rastogi Verma³ and *Preeti Bajpai¹

¹Department of Biosciences, Integral University, Lucknow-226026, India

²Toxicology Division, Central Drug Research Institute, Lucknow-226031, India

³Department of Biotechnology, Delhi Technological University, Delhi-110042, India

*Corresponding author email: preeti2874@gmail.com

Abstract

Silver nanoparticles (AgNPs) prepared from *Alternaria brassicae* may exhibit potential antimicrobial and immunomodulatory activity due to their inimitable character. The purpose of this study was to evaluate their antibacterial activity, antifungal and immunomodulatory efficacy on various pathogenic microorganisms and cell lines (*J774* and *THP1 α*). The characterization of these AgNPs was determined by using differential light scattering, UV-visible spectroscopy and transmission electron microscopic analyses. Mycologically produced AgNPs were found as spherical and irregular shaped measuring size range between 55.4 nm to 70.23 nm. The antimicrobicidal activity of these AgNPs against *Staphylococcus aureus*, *Trichoderma* sp. and *Fusarium semitectum* was evaluated by agar well diffusion method. Results showed that AgNPs inhibit the growth of various bacteria and fungi, which may be due to the disruption of cell membranes, leakage of cytoplasm and DNA degradation. Cytotoxicity analysis of AgNPs on cell lines revealed its dose dependent effect. Moreover, significant increase of intracellular reactive oxygen species was characterized in AgNPs treated cells after 4 hr of incubation. Thus, AgNPs may have a significant advantage over conventional antibiotics as microorganisms are acquiring resistance against the broad range of available antibiotics. In conclusion, study showed that these nanoparticles could be used as an antimicrobial agents as well as cost effective and nontoxic immunomodulatory delivery vehicle.

Keywords: *A. brassicae*; AgNP; antibacterial; antifungal; cytotoxicity; immunomodulatory

1. Introduction

Myco-nanotechnology is an emerging scientific discipline aimed towards studying the production of nanomaterials or nanostructures with desirable shapes and sizes by fungi. Potential appliances of myco-nanotechnology have fascinated scientist and researchers to contribute in providing incremental solutions through green chemistry approaches for targeted drug delivery. Thus, nano-sized particles have attracted worldwide attention due to their unusual photo-electrochemical, electronic chemical and optical properties.

AgNPs offer unique optical, catalytic and disinfectant properties gaining high scientific and commercial interest. They are used in broad range of different products ranging from acting as an antimicrobial agent, coatings of surgical instruments, contraceptive devices, wound dressing and prostheses [9] to the use in food container systems. Besides that, they are highly attractive for creation of advanced functional materials. To meet the wide scope of nanostructures, number of procedure like electrochemical methods, laser ablation method, microwave irradiation method, thermal decomposition may successfully produce pure well-defined nanoparticles. However, these procedures are capital-intensive, dangerous, energy use and often create health hazards due to the usage of toxic chemicals, which lead to production of harmful by-products [12] that makes their use inappropriate within biological systems.

The safety of nanomaterials marketed during past decades does not appear to be adequately addressed. There is a persistent requirement to develop new methods which should be safe, nontoxic, reliable, cost effective, and ecofriendly. The new methods should also be able to assess the cytotoxicological safety of current and future nanomaterials under different scenarios. Thus, synthesis of nanoparticles by using bio-organisms like bacteria, fungi, and plant extracts is compatible with the green chemistry principles [14]. Biosynthesis of metal nanoparticles is a kind of bottom up approach where the main reaction is the reduction/oxidation reaction. The enzymes from these microbes possess high redox potential which is usually responsible for reduction of metal compounds into their respective nanoparticulates (Figure 1) and synthesized nanomaterials would be intervened by biological systems. Recently, resistance to commercially available antimicrobial agents by pathogenic bacteria and fungi has been rising at an alarming rate and has become a severe problem [8].

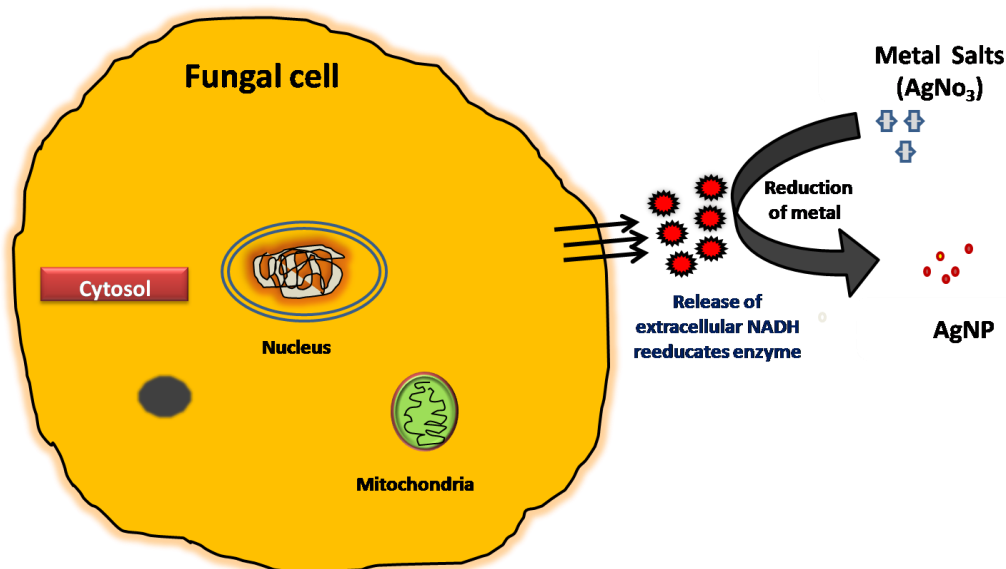


Fig.1 Schematic representation of reduction of metal ions by fungal secretory proteins

In the present study, a phosphate solubilizing fungi was preferred for nanostructure formulation due to their ability to produce large amount of secretory proteins [16]. Numbers of reports have showed the biological synthesis of metal nanoparticles but the potential of a phosphate solubilizing fungi, *A. brassicae* has not yet been demonstrated. The present investigation is an effort to synthesize AgNPs from *A. brassicae* (Accession No. KF934409). Moreover, antimicrobials of mycological origin have enormous therapeutic potential therefore, it is the need of time to develop new antimicrobial drugs for the treatment of infectious diseases.

2. Material and methods

2.1 Isolation, screening and characterization of *A. brassicae* (KF934409)

The fungus was isolated from rhizospheric region of Central Institute of Medicinal And Aromatic Plants (CIMAP), Lucknow, India and screened by its phosphate solubilizing ability on Pikovskaya's medium by using plate assay method [7]. The fungal colonies were subculture on fresh petriplates containing potato dextrose agar (PDA) media and the plates were incubated in inverted position for 72 h at $28 \pm 3^\circ \text{C}$ and examine for the halo zones around the colony. Further, the biochemical characterization of the fungus were tested on the basis of morphology of colony, conidial microscopic analysis, solubilization index (SI), cellulose hydrolysis test and starch hydrolysis test according to standard protocols [6].

The fungus was molecularly characterized by using 18S rRNA sequencing where genomic DNA was isolated and subjected to high-fidelity PCR using universal primers *i.e.* forward primer (5'-GGAAGTAAAAGTCGTAACAAGG-3') and reverse primer (5'-TCCTCCGCTTATTGATATGC-3') and analyzed on 1 % agarose gel. The PCR products were sequenced bi-directionally using the forward and reverse primers. Homology between this 18S rDNA sequence and the strains available at the public databases (Genbank, EMBL and DDBJ) was determined using BLASTN sequence match routines. The UPGMA (Unweighted Pair Group Mathematical Average) algorithm was used to perform hierarchical cluster analysis [13]. The sequences were aligned using CLUSTALW2 program and its phylogenetic and molecular evolutionary analysis were conducted. Sequences analyses were performed by alignment of the partial 18S rRNA gene sequences to those from the GenBank database, using the program BLAST (NCBI BLAST® homepage). The nucleotide sequences of 18S rRNA gene segments determined in this study have been deposited in GenBank database under accession number KF934409.

2.2 Myco-production of AgNPs and its characterization

The isolated fungus was grown in MGY (maltose, glucose, yeast potato broth which comprises of malt extract (0.5 %), glucose (1 %), yeast extract (0.3 %) and peptone (0.5 %)). The culture was incubated at 27° C and harvested after 5 days of growth by sieving followed by extensive washing with sterile double-distilled water. Initially 20 g of biomass (wet weight) was transferred to 100 ml deionized water for 48 hr at 27° C in an Erlenmeyer flask and agitated at 150 x rpm for release of secretory proteins. Silver nitrate (1 mM) was added to the Erlenmeyer flasks and the reaction was allowed to progress under dark conditions for production of AgNPs. Time-dependent formation of AgNPs was observed by using ultraviolet-visible spectrophotometer (Beckman DU-20 spectrophotometer). The scanning range was 350-650 nm for AgNPs at a scan speed of 420 nm/min. The data was recorded and analyzed using “UVWinlab” software.

2.2.1 Differential light scattering (DLS)

The suspensions of AgNPs was prepared in distilled water (dH₂O) by using a bath-sonicator (ULTRASONIK 57 X, 50/60 Hz, California, USA) prior to size measurements. Dynamic light

scattering size measurements were performed with the aid of a Malvern Zeta Sizer Nano ZS (Malvern Instruments, Worcestershire, UK) operating with version 5.03 of the systems Dispersion Technology Software (DTS Nano). The samples for DLS were equilibrated at 25° C for 3 min before each measurement. The refractive index (RI) of AgNP.dH₂O was 1.330.

2.2.2 Transmission electron microscopy (TEM)

The synthesized AgNPs were characterized by TEM analysis. The samples were prepared by placing a drop of synthesized nanoparticles over gold-coated negative grid followed by evaporation of the solvent [22]. TEM analysis was performed on Perkin-Elmer model which was operated at an accelerating voltage of 1000 kV.

2.3 Antibacterial efficacy

The antibacterial efficacy of synthesized AgNPs was determined by using the agar well diffusion assay method [4]. Pure cultures of five pathogenic bacteria namely, *B. subtilis*, *B. cereus*, *S. aureus*, *E. coli* and *E. aeruginosa* were procured from National Chemical Laboratory, Pune. Bacterial stock cultures were maintained at 4° C on nutrient agar media slants and subculture on nutrient broth media for antibacterial analysis of NPs. 5 mm wells in diameter were prepared and filled with AgNPs with a range of concentrations (10 µM, 50 µM, 100 µM, 150 µM). Each experiment was performed in triplicate and the average zone of inhibition, excluding well was recorded. 1 mM AgNO₃ were used as negative controls. The diameter of such zones of inhibition was measured for each organism and expressed in centimeter.

2.4 Antifungal efficacy

Disk diffusion method was used to evaluate the antifungal activity of Terconazole against *A. niger*, *Trichoderma sp.* and *F. semitectum* on PDA. The standard terconazole disks (Fu10; 10 µg/disk) were purchased from Hi-Media (Bangalore, India). To determine the combined effect, each standard paper disk was further imbued with 25 µl of freshly prepared AgNPs with a concentrations range of 10 µM, 50 µM, 100 µM, 200 µM. PDA plates were inoculated with a spore suspension (20 µl) of the test fungi. Standard antifungal terconazole disks were used as positive control. The terconazole disks were impregnated with AgNPs, placed onto the PDA medium and inoculated with tested fungi. The fungal cell filtrate, used for the synthesis of

AgNPs was used as negative control. The cultured plates were incubated at $28 \pm 4^\circ \text{C}$ for 7 days. The average inhibition zone, excluding well, for each case was measured.

2.5 Cell lines

Human macrophage cell line *THP1* α and mouse macrophage cell line *J774* was procured from the National Center of Cell Sciences, Pune, India and maintained at Animal Tissue Culture facility of Central Drug Research Institute (CDRI). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % Foetal bovine serum (FBS) and 1% antibiotic and antimycotic solution (50,000 units/L of penicillin and 50 mg/L of streptomycin) and 2 mM glutamine. Cultures were held in 75 cm culture flasks, at 37°C and 5 % CO_2 using standard cell culture methods.

2.5.1 Cytotoxicity assay

Cell viability was assessed by using the MTT [3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide] conversion assay [5]. Thus, 1×10^4 cells/ml were seeded on 96-well culture plates and incubated with increasing concentrations of nanoparticles (10 μM , 25 μM , 50 μM , 100 μM , 150 μM and 200 μM) for 24 hr at 37°C in CO_2 incubator. The MTT dye was added to each well and plate was incubated at 37°C for 4 hr. The absorbance of insoluble formazan salts was assessed at 550 nm using Powerwave XS (BIOTEK, USA) spectrophotometer [3]. Data produced were used to plot a dose-reaction curve and the concentration of these metal nanoparticles required to kill 50 % of cell population (IC_{50}) was determined.

$$\text{Cell viability (\%)} = \frac{\text{Mean OD}}{\text{Control OD}} \times 100$$

2.5.2 Intracellular reactive oxygen species estimation

Intracellular oxidative stress was calculated with the help of 2', 7'-dichlorofluorescein di-acetate (DCFH-DA), a well-accepted fluorescent marker for study of intracellular hydroperoxides [18]. The experiment was performed according to the protocol described [15] with slight modifications. Primarily, healthy confluent cells were harvested and seeded (1000 cells/well) into black bottomed 96 well plates (Nunc, Denmark) and allowed to adhere for a period of 24 hr

prior to exposure. For ROS quantification, both mouse and human macrophage cells were plated and distributed in triplicates. A working stock of 25 μ M DCFH-DA in phosphate buffered saline (PBS) was prepared and all test concentrations, unexposed negative controls and positive controls were prepared and exposed to the cells in this working stock. The negative control consisted of the working stock solely a 25 μ M DCFH-DA solution in PBS, the positive control consisted of 1 μ M hydrogen peroxide (H_2O_2) in 20 μ M DCFH-DA/PBS working stock solution and finally the test concentrations consisted of a continuous range of AgNPs. Lipopolysaccharide (1 μ g/ml) was used as a mitogen for stimulation of macrophages as well as comparison of phagocytic activity of stimulated and non-stimulated macrophages. The test concentration for AgNPs was 10 μ g/ml and the incubation period ranged from 2 hr to 6 hr. The rate of intracellular oxidative stress was monitored by measuring their fluorescence intensity via fluorometer (BIOTEK-FLX800'USA) emission at 520 nm (by 485 nm excitation).

2.6 Statistical analysis

All the experiments were conducted in triplicates and results were expressed as mean $\pm$ SD. One-way analysis of variance (ANOVA) with a Dunnett's test was performed for the multiple comparisons for normally distributed samples with homogenous variance. Statistically significant differences were set at $p < 0.05$.

3 Results

3.1 Screening and fungal characterization

The rhizospheric sample was collected, serially diluted and plated on Pikovskaya's media to check their phosphate solubilization efficiency. After 7 d of incubation, a fungal colony was observed, identified, as dark ash colour, looking like cotton having irregular edges and 1.2 cm in diameter. On Pikovskaya's medium, the transparency of the media is a primary indicator of phosphate solubilization to be visualized as halo zone formation around the colony. Thus among all, the fungal isolate FCK 20 was found as an efficient fungus on the basis of SI, pH and phosphatase enzyme production. This fungus was further characterized morphologically, biochemically and at molecular level. The SI of selected fungus was recorded with size of 2.9 cm in comparison with others strains. Further, colony morphology was analyzed on the Sabouraud agar medium and filamentous structural pattern was analyzed microscopically after

lactophenol staining. Moreover, this fungus gave a negative test for starch hydrolysis and cellulose degrading ability.

For molecular characterization, the genomic DNA was amplified using 18S rRNA specific primers (ITS 4 and ITS 5). An amplicon of 500 bp was observed on 1 % agarose gel. The DNA sequencing of the amplified product and its BLAST analysis confirmed the fungus as *A. brassicae* (KF934409). The phylogenetic analysis (Figure 2) revealed the fungus to belong to Pleosporaceae family and the genus was *Alternaria*.

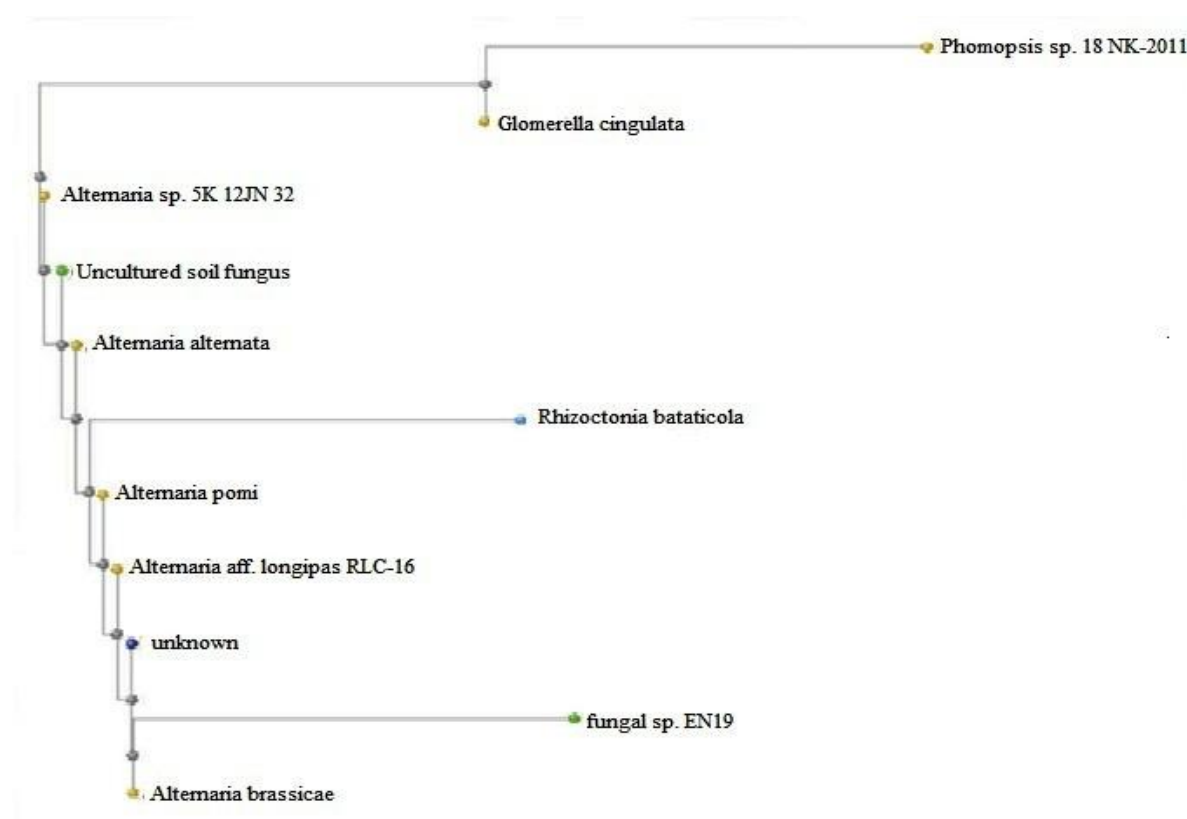


Fig. 2 Phylogenetic analysis of *A. brassicae* (KF934409)

3.2 Synthesis and size estimation of AgNPs

AgNPs were mycologically produced from *A. brassicae* (KF934409). 1 mM of silver nitrate was added to the fungal filtrate under magnetic stirring which was used as reducing and stabilizing agent for 1 mM of silver nitrate. Appearance of a brownish color in solution (Figure 3) reflected the formation of AgNPs in the reaction mixture which is due to the surface plasmon resonance,

exhibited by the AgNPs. The control sets did not demonstrate any color change under similar experimental conditions.

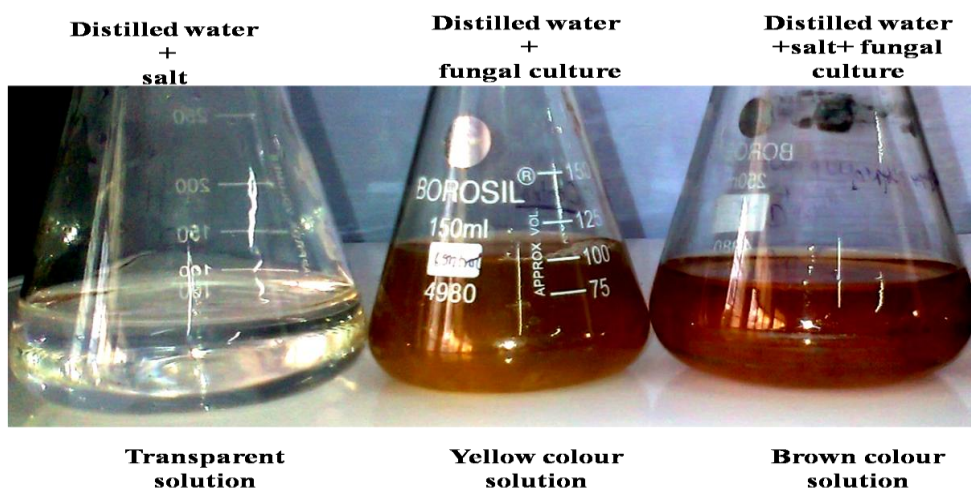


Fig.3 Nanoparticles formation on the basis of colour change due to surface plasmon resonance

3.2.1 Ultraviolet-visible spectrophotometric analysis

Ultraviolet-visible spectrophotometer showed no evidence of absorption in the range of 380-750 nm for the fungal extract. Whereas the AgNO_3 exposed fungal extract showed a distinct absorption at around 350 nm and 650 nm, with a peak at 370 nm for AgNPs [Figure 4 (a)].

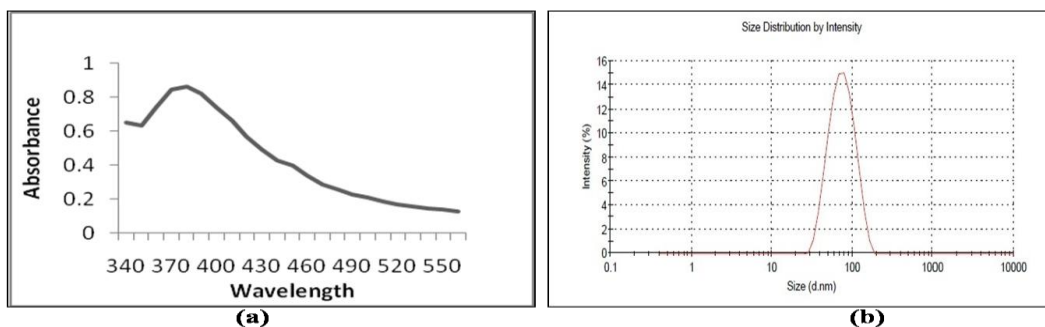


Fig.4 (a) UV-VIS spectrophotometry of AgNP nanoparticles and **4 (b)** DLS spectrum of Ag nanoparticles

3.2.2 Dynamic light analysis (DLS)

This technique enables the particle size determination by measuring the random changes in the intensity of light scattered from a suspension or solution. DLS spectra showed an intensity of

92.7 nm for AgNPs [Figure 4 (b)]. This size variation is due to oxidation of metal salts into their respective nanoparticles in the presence of enzymes.

3.2.3 Transmission electron microscopic (TEM) analyses

The morphology of AgNPs is spherical and the TEM micrographs suggest that particle diameters ranged from 48.4 to 65.23 nm (Figure 5). The dimensions of AgNPs were small enough and imaged as poly dispersed small and large spherical nanoparticles with variable diameter. This was further confirmed by the representative images recorded from the uniformly dispersed drop-coated film of the AgNPs on grid.

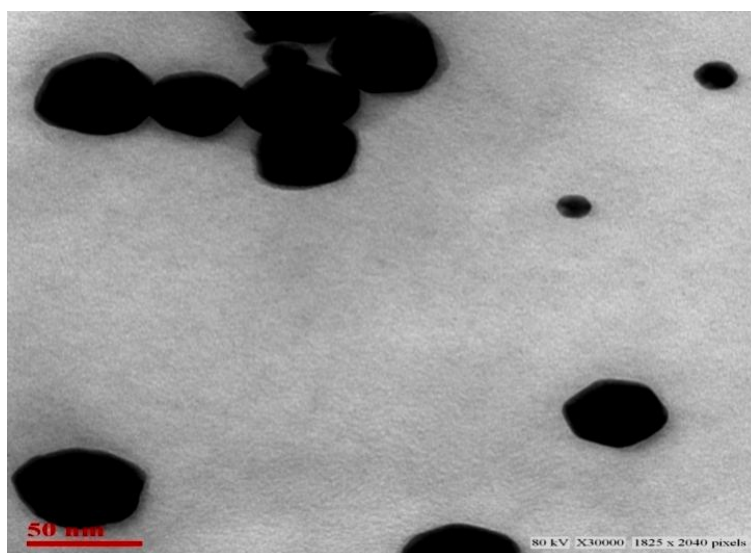


Fig.5 Transmission electron microscopic analysis of AgNP nanoparticles

3.3 Antibacterial activity of AgNPs

The bactericidal activity of AgNPs were studied using the pathogenic strains of bacteria as *S. aureus*, *E. coli*, *B. subtilis*, *B. cereus* and *E. aerogenes* using agar well diffusion method. After the incubation time, zone of inhibition (clear zones) were observed against all the tested microbes. The results recorded in centimetres for AgNPs are shown in Figure 6 (a) respectively. The effectiveness of AgNPs could be attributed to the fact that their larger surface area enabled them to have a better contact with the microbial cell wall. This is further supported by the observation that size dependent interaction of AgNPs with bacteria leads to its antibacterial activity [21]. The exact mechanism by which the reaction occurs is still largely unknown.

However, some literature proposes that Ag interact with the thiol groups of proteins, which is necessary for microbial respiration ability [10]. Ag might also interact with phosphorus containing compounds like DNA disturbing the replication process or preferably by their attack on the respiratory chain [10]. The direct interaction of Ag with cell membrane of bacteria has also been shown which consequently breaks the membrane. Earlier experimental evidences have also advocated the failure of replication ability by the DNA after treatment with silver ions [17]. The comparative histogram demonstrated that the best antibacterial efficacy of AgNPs was against *B. cereus* followed by *S. aureus*. The marked increase in antibacterial activity was demonstrated with increasing concentration of AgNPs [Figure 6 (a)]. In addition, the efficacy of AgNPs was also found to be enhanced in combination with the antibiotic rather than alone.

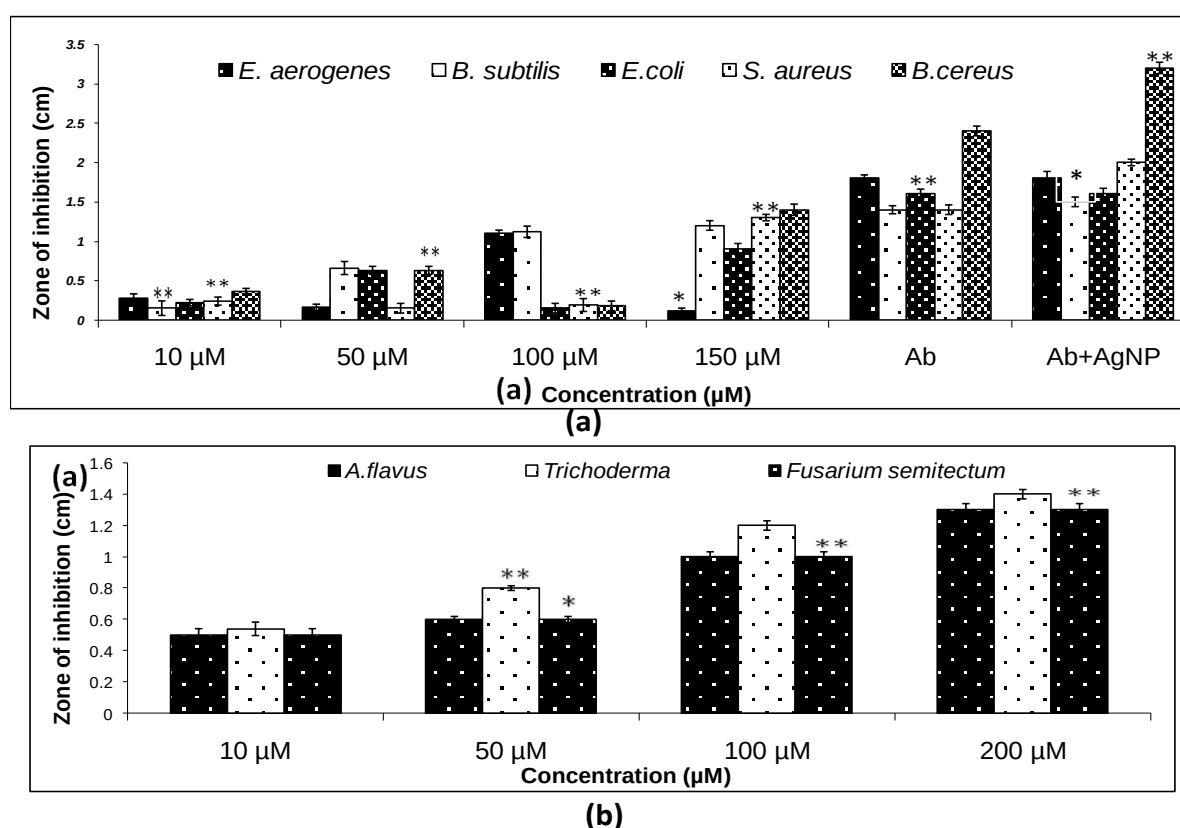


Fig. 6 Comparative analysis of (a) antibacterial activity and (b) antifungal activity of AgNP for different pathogenic bacteria and fungi. Results are presented in relative units compared with controls (Ab) which is not shown in the graph. Different signs (* and **) letters indicate significant differences ($p < 0.05$).

3.4 Antifungal activity of AgNPs

The colloidal AgNPs inhibited the growth of the fungus (*A. niger*, *Trichoderma sp.* and *F. semitectum*) which was seeded in the Muller Hinton agar plate and formed a zone of inhibition around the central cavity. The zone of inhibition with diameter of 1 cm was recorded in case of *A. niger*, 1.2 cm in *Trichoderma sp.* and 1.1 cm in case of *F. semitectum* [Figure 6 (b)]. The antifungal activity is due to the formation of insoluble compounds by inactivation of sulfhydryl groups in the fungal cell wall and disruption of membrane bound enzymes and lipids, which causes cell lysis [1]. We have observed that *Trichoderma sp.* and *F. semitectum* are susceptible to the lethal effects of the prepared silver due to the smaller size of AgNPs.

3.5 Cell lines

To determine the biomaterial capability of nanoparticles with no toxic effects, the cell viability and cytotoxicity assays were performed. Cell viability was determined by a standard MTT conversion assay.

3.5.1 *In vitro* cytotoxicity assay

Treatment of AgNPs to *J774* and *THP1 α* cell lines exhibited mild cytotoxicity in a dose dependant manner. At low doses (10 μ g) no cytotoxic effects were observed whereas at high doses of 100-150 μ g of AgNPs mild cytotoxicity was observed. Long-time exposure resulted in additional toxicity to the cells and reached to maximum dead cells after 24 hr of incubation, which might be due to over-accumulation of metal nanoparticles within the cell (Figure 7).

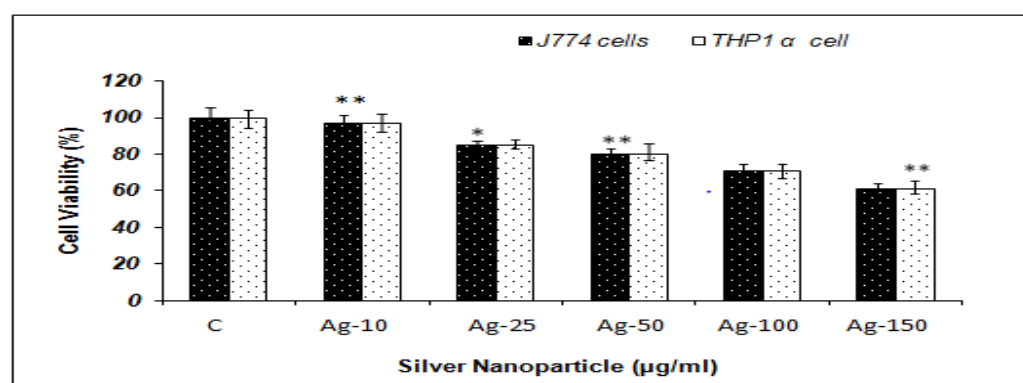


Fig.7 Dose-dependent effect of AgNP over cell viability using MTT assay on *J774* and *THP1 α* cells. Results are presented in relative units compared with controls. . Different signs (* and **) letters indicate significant differences ($p < 0.05$).

3.5.2 Intracellular reactive oxygen species estimation

The ROS estimation has shown that AgNP exposure to *J774* and *THP1* α cells caused time dependant increased production of ROS at respective concentration indicating macrophage stimulation. The maximum amount of ROS was produced uptill 4 hr of incubation, which declined at 6 hr (Figure 8). Here, the comparable ROS generation by AgNPs, from non-stimulated cells to mitogen-stimulated cells was observed which indicates the immunomodulatory potential of the synthesized AgNPs. This may be due to the small size of AgNPs as these could easily penetrate through the cell wall in an appreciable number [15].

The maximum free radical activity was obtained by the *J774* cells [Figure 8 (a)] and *THP1* α cells [Figure 8 (b)] after 4 h of incubation with AgNPs. Free radical generation and analogous encapsulation efficacy capability of AgNPs revealed the potential role of AgNPs as drug / vaccine delivery vehicle for macrophages as well as also indicate towards their immunomodulatory activity.

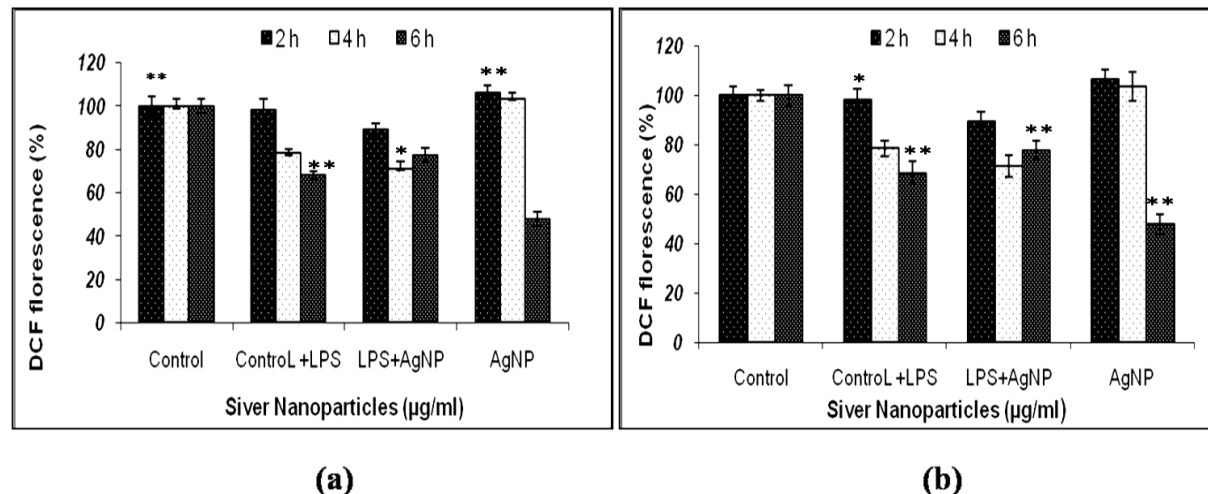


Fig. 8 ROS estimation in (a) *J774* cell lines and (b) *THP1* α cell lines after incubation with AgNPs at various time points (2 h, 4 h, 6 h). Results are presented in relative units compared with controls. . Different signs (* and **) letters indicate significant differences ($p < 0.05$).

4. Discussion

In the present study, a phosphate solubilizing fungal isolate, FCK 20 was isolated and subjected to its characterization. The sequence of 18S rRNA of the fungal strain FCK 20 was submitted to GenBank with an accession number of (KF934409). The homology search using BLAST

indicated a close genetic relatedness of the strain FCK 20 with the rRNA sequence of *A. brassicae* (18S: 100 % similarity with the reference sequence BankIt1680606 Seq5) in NCBI database. Such a higher identical value confirmed the strain FCK 20 to be *A. brassicae*.

Myconanotechnology has become an extensive field of study involving chemistry, physics, engineering, computing, electronics, energy, agriculture and biomedicine. In the realm of biomedicine, nanotechnology is widely touted as one of the next promising and important approaches to diagnose and treat various ailments [11]. Since physical and chemical methods of metal nanoparticle production are expensive and involve incorporation of toxic chemicals therefore their biological synthesis (bacterial, fungal and plant extract) would be preferred. This owes to their ease of availability, nontoxic nature and quicker synthesis, which prompted us to use this phosphate solubilizing fungus *A. brassicae* for the synthesis of AgNPs. In best of our knowledge this is the pioneer study showed the use of fungus for the production of nanoparticles. The confirmation of synthesis of AgNPs was based on surface plasmon resonance involving color alteration. Further, in the UV/VIS absorption, a strong peak located at 370 nm was observed. Average particle size calculated from DLS data was found to be 65.7 nm for AgNPs. Furthermore, the TEM image of the synthesized AgNPs has validated the formation of spherical, nanoparticles, which was 50 nm in diameter. This indicated the reduction of Ag^+ to elemental silver (Ag). The UV/VIS spectra of AgNPs was observed at 380 nm and showed their spherical shape. Additionally, the synthesized AgNPs were stable in solution over a time of three months at room temperature. Nano-silver is an effective and a fast-acting microbicide against a broad spectrum of pathogenic bacteria and fungi thus can be utilized in various processes in the medical field [9]. Our findings are in agreement to the previous reports and synthesized AgNPs have showed the considerable antibacterial activities along with the standard antibiotic (tetracycline) against *B. cereus*. The antibacterial activity of AgNPs is due to the permeability of the cell membrane [23] or formation of free radicals [20] or interaction of AgNPs with the thiol groups of many enzymes thus inactivating them. Beside this, the moderate antifungal activity of AgNPs against *Trichoderma sp.* and *F. semitectum* was due to formation of insoluble compounds by inactivation of sulfhydryl groups in the fungal cell wall and disruption of membrane bound enzymes and lipids, causing cell lysis [19].

The MTT assay determines the mitochondrial activity of the cells which, reduce the soluble, yellow MTT into an insoluble, purple formazon. The reduction of MTT to formazon indicates

the decrease in mitochondrial metabolism of the cells. Therefore, the absorption of formazon formed, directly co relates to the number of cells whose mitochondrial metabolism is intact even after the exposure of AgNPs. Dose response of AgNPs in *J774* cells and *THP1* α cells showed the decrease in the reduction of MTT to formazon with the increasing concentration and time of exposure. However, AgNPs were found to be significantly more toxic to *THP1* α cells as compared to *J774* cells, which could be attributed to the intrinsic anticancer property of AgNPs. This might be due to small particle size of AgNPs with enormous specific surface area, which facilitated further expression and dissolution of ions, leading to increased toxicity. AgNPs are highly reactive, exhibit oxidative potential, contains the ability to bind with biomolecules like proteins and DNA and consequently caused the disturbance in the functioning of bimolecules [2].

In both cell lines *J774* and *THP1* α a significant increase in the levels of ROS was observed in comparison to the unexposed controls. The maximum amount of ROS was produced after 4 hr of incubation with AgNPs, which declined at 6 h. ROS plays an important role in triggering cellular pathways that can lead to cellular death by either causing nuclear damage or by contributing in cell membrane rupture mechanisms. Exposure of AgNPs to cells caused increased production of ROS, suggested their potential use as an anti-microbicide, occupying its application in agriculture. Moreover, these nanoparticles can also act as immunomodulatory agent alone or in combination with established therapeutic immunomodulatory agents. As these NPs are simply engulfed by the macrophages, they also pose themselves as targeted drug / vaccine delivery vehicle to macrophages thereby a boom for development of a potent chemotherapeutic vehicle for diseases involving macrophages viz., leishmaniasis, tuberculosis etc. Hence, care has to be taken to utilize this marvel well and in a good, efficient and effective ways. Researchers should also understand its limitations and taking excessive care for its effect on environment and human health.

In future, these bio-synthesized nanomaterials (encapsulation) may lead to enhancement of agricultural productivity like slow release of phosphorus from fertilizers and its effective uptake by plants. Fungus *A. brassicae* contains phosphatase enzyme whose main function is to solubilize the insoluble form of phosphorus in to soluble one. Thus, the bio-nanoparticles prepared from this fungus may display slow release of encapsulated enzyme and hence may improve phosphate solubilization. These synthesized myco-nanoparticles may be mixed with

fertilizers to increase the uptake of phosphorus nutrient. However, in future, this strategy needs to be validated.

5. Conclusion

The present study showed the mycogenic production of AgNPs with particular emphasis on procedure of synthesis and its potential applications in human welfare as well as in other allied sectors. AgNPs were synthesized from fungus *A. brassicae* which has phosphate solubilizing ability. Findings showed that synthesized nanoparticles were spherical in shape and has a promising antimicrobial agent against both gram-positive / gram-negative bacteria and pathogenic fungi also. However, higher doses of AgNPs exhibited the cytotoxicity and ROS production in *J774* and *THP1 α* cells in time dependant manner. . Observations have also showed the bactericidal activity of synthesized nanoparticles thus implementing their role as antimicrobicidal agent. In conclusion findings showed that biosynthesized silver nanoparticles could be use in agriculture and also have antimicrobicidal activity which could be utilize for human welfare. However, further studies are required to fully characterize the mechanistic and toxicity aspects of synthesized nanoparticles.

Conflict of interest

The authors have no conflict of interest.

Acknowledgements

The authors are highly thankful to Vice Chancellor, Integral University for his support and encouragement. We sincerely thank Mr. Sharma ITRC, Lucknow for carrying out TEM analysis of nanoparticles. The grant of UGC-Maulana Azad National Fellowship is gratefully acknowledged.

References

1. B. Dorau, R. Arango, and F. Green, "Forest Products Society pp. Proceedings of the 2nd Wood-Frame Housing Durabili and Disaster Issues Conference," 133, Las Vegas, NV. 2004.

2. B. Reidy, A. Haase, A. Luch, and K. Dawson, "Mechanisms of AgNP release, transformation and toxicity: A critical review of current knowledge and recommendations for future studies and applications," *Lynch, Materials*, vol. 6, pp. 2295-2350, 2013.
3. C. Lam, J. James, and R. Mccluskey, "Pulmonary toxicity of single-wall carbon nanotubes in mice 7 and 90 days after intra-tracheal instillation," *Tox Sci*, vol. 77, pp. 126–34, 2004.
4. C. Perez, M. Pauli, and P. Bazerque, "An antibiotic assay by agar-well diffusion method," *Acta Biologiae et Medecine Experimentaalis*, vol. 15, pp.113-115, 1990.
5. F. Mosmann, "Methods: Rapid calorimetric assay for cellular growth and survival:. Application to proliferation and cytotoxicity assay," *J. Immunol*, vol. 65, pp.55–63, 1983.
6. G. Garrity, "Springer-Verlagpp. Bergey's Manual of Systematic Bacteriology," In *Bergey's Manual of Systematic Bacteriology* pp. 742 New York, 2001.
7. I. Pikovskaya, "Mobilization of Phosphate in Soil in Connection With Their Vital Activities of Some Microbial Species," *Microbiologiya*, vol. 17, pp. 362-370, 1948.
8. J. Wright, K. Lam, D. Hansen, and R. Burrell, "Efficacy of topical silver against fungal burn wound pathogens," *Am. J. Infest. Control*, vol. 27, pp. 344-350, 1999.
9. J.W. Alexander, "*History of the medical use of silver*. Surgical Infections," vol. 10, pp. 289-292, 2009.
10. K. Yliniemi, and M. Vahvaselka, "Antimicrobial activity of colloidal AgNPs prepared by sol-gel method," *Chem*, vol. 18, pp.199, 2008.
11. K. Hartman, L. Wilson, and M. Rosenblum, "Detecting treating cancer with nanotechnology," *Mol Diagn Ther*, vol. 12, pp. 1–14, 2008.
12. K. N. Thakkar, S. S. Mhatre, R. Y. Parikh, "Biological synthesis of metallic nanoparticles," *Nanomedicine: Nanotechnology, Biology, and Medicine*, pp. 1016, 2009.
13. L.A.S Dias, and A.C. Alfenas (Ed.), "Eletroforese de Isoenzimas e Proteínas Afinspp. Fundamentos e Aplicações em Plantas e Microrganismos, UFV, Viçosa *Análises multidimensionais*, pp. 405–473, 1998.
14. Li. Huang, J. Q. Sun, D. Lu, and Y. Su, "Biosynthesis of silver and gold nanoparticles by novel sundried *Cinnamomum camphora* leaf," *Nanotechnology*, vol. 18, pp. 11-15, 2007.

15. N. Swarnakar, K. Thanki, and S. Jain, "Effect of co-administration of CoQ10-loaded nanoparticles on the efficacy and cardiotoxicity of doxorubicin-loaded nanoparticles," *RSC Advances*, vol 3, pp. 14671, 2013.
16. P. Mohanpuria, N. K. Rana and S.K. Yadav, "Biosynthesis of nanoparticles : Technological concepts and future applications," *J. Nanopart*, vol. 10, pp. 507–517, 2008.
17. Q. Feng, G. Chen, F. Cui, T. Kim, and J. Kim, "A mechanistic study of the antibacterial effect of silver ions on *Escherichia coli* and *Staphylococcus aureus*," *J Biomed Mater Res*, vol. 52, pp. 662–668, 2000.
18. R. Cathcart, E. Schwieters, and B. N. Ames, "Detection of picomole levels of hydroperoxides using a fluorescent dichlorofluorescein assay," *Anal. Biochem*, vol. 134, pp. 111–116, 1983.
19. R. Liao, and J. Talbot, "Assessment of the effect of amphotericin B on the vitality of *Candida albicans*," *Antimicrobial agents and chemotherapy*, vol 43, pp. 1034-1041 1999.
20. S. Hussain, K. Hess, J. Gearhart, and K. Geiss, "In Vitro: Toxicity of Nanoparticles in BRL 3 A Rat Liver Cells," *J. Toxicol*, vol. 19, pp. 975–983, 2005.
21. S. Pal, Y. Tak, and J. Song, "Does the antibacterial activity of AgNPs depend on the shape of the Nanoparticle? A study of the gram-negative bacterium *E. coli*," *Appl. Environ. Microbiol*, vol. 73, pp. 1712-1720, 2007.
22. V. Germain, J. Li, D. Z. and M.P. Pileni, "Stacking faults in formation of silver nanodisks," *J. Phys Chem B*, vol. 107, pp. 8717–8720, 2003.
23. Y. Gavrieli, Y. Sherman, and S. Bensasson, "Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation," *J. cell Biol* vol. 119, pp. 493–501, 1992.

On the Applicability of Evolutionary Computation for Software Defect Prediction

Ruchika Malhotra

Department of Software Engineering
Delhi Technological University
Delhi, India
ruchikamalhotra2004@yahoo.com

Nakul Pritam

Department of Software Engineering
Delhi Technological University
Delhi, India
nakul.pritam@gmail.com

Yogesh Singh

Maharaja Sayajirao University of
Baroda
Gujrat, India
ys66@rediffmail.com

Abstract— Removal of defects is the key in ensuring long-term error free operation of a software system. Although improvements in the software testing process has resulted in better coverage, it is evident that some parts of a software system tend to be more defect prone than the other parts and identification of these parts can greatly benefit the software practitioners in order to deliver high quality maintainable software products. A defect prediction model is built by training a learner using the software metrics. These models can later be used to predict defective classes in a software system. Many studies have been conducted in the past for predicting defective classes in the early phases of the software development. However, the evolutionary computation techniques have not yet been explored for predicting defective classes. The nature of evolutionary computation techniques makes them better suited to the software engineering problems. In this study we explore the predictive ability of the evolutionary computation and hybridized evolutionary computation techniques for defect prediction. This work contributes to the literature by examining the effectiveness of the 15 evolutionary computation and hybridized evolutionary computation techniques to 5 datasets obtained from the Apache Software Foundation using the Defect Collection and Reporting System. The results are evaluated in terms of the values of accuracy. We further compare the evolutionary computation techniques using the Friedman ranking. The results suggest that the defect prediction models built using the evolutionary computation techniques perform well over all the datasets in terms of prediction accuracy.

Keywords— Defect Prediction; Search Based Software Engineering; Evolutionary Computation; Software Metrics

I. INTRODUCTION

The complexity and size of software is increasing due to which it is not unusual to produce software with defects. The identification of defective classes in the early phases of software development will help software developers focus the available testing resources on the defect prone parts of the software. This will help in producing high quality, low cost and maintainable software. There are various object-oriented metrics proposed in literature that capture various aspects of software design and can be effectively used in predicting defective classes. The models

built using these metrics can be used by software practitioners in early phases of software development for determining whether a class is defective or not.

Evolutionary computation has seen an explosion of interest in recent years in the various domains like medicine, automated designing, bioinformatics, etc. Harman et al. [1] has advocated that the predictive capability of these techniques must be evaluated in the software engineering domain, specifically defect prediction.

According to [2], there are many advantages of using evolutionary computation techniques including their ability to handle changing data, capability of performing where heuristics are not available or available heuristics are not satisfactory, the simplicity of the approach used by these algorithms to solve problems, etc.

Although there have been many studies that analyze the relationship between object-oriented metrics and defect proneness there is no study, to the best of the authors' knowledge, till date that analyses the predictive capability of evolutionary computing techniques for defect prediction. The aim of this study is to address the following issues,

- Analyze the predictive capability of evolutionary computing techniques for defect prediction.
- To validate the relationship between object-oriented metrics and defect proneness over 5 open source software systems.
- To determine which evolutionary computing technique performs the best.

In this paper we compare and assess the performance of evolutionary computing techniques for prediction of defective and non-defective classes. To conduct this study, we selected 15 evolutionary and hybrid evolutionary computing techniques.

The results are validated using 5 datasets collected from open-source software written using JAVA programming language. The validation method used in this study is 10-cross validation. The performance of the models built using evolutionary computing techniques is assessed using the accuracy measure. We also apply the statistical Friedman's test to evaluate which technique performs the best.

Thus, the main contributions of this study are, first, we compare and analyze the rarely used evolutionary computing techniques for software defect prediction. Second, we use 5 open source software systems developed using different methodologies in comparison to proprietary software.

The remainder of this paper is organized as, section 2 presents a summary of work done in this area, evolutionary computation and the configuration of a evolutionary techniques used in the study is discussed in section 3, section 4 provides insight to the data collection process including description of dependent and independent variables and performance measures used. The results are discussed in section 5 and 6. Discussion of the results along with the limitations of the study is given in section 7. Section 8 concludes this study.

II. RELATED WORK

In [3], the authors conducted a study over 8 medium sized systems. They concluded that Chidamber and Kemerer metrics [4] have an association with defect proneness.

In [5] it was concluded that only WMC and RFC have an association with defect proneness of a system. Three industrial software systems written in C++ were examined for the purpose of this study.

Emam et al. [6] discovered that size of a class (LOC) tends to have a confounding effect on almost all object-oriented metrics in regard of the defect prediction problem. In another study [7], they found out that export-coupling metric is the most significant predictor for defect proneness.

The authors of [8] concluded that all metrics but (Number of Children) NOC were effective in predicting defect proneness of a class. The study made use of a medium-sized system written in C++ and 49 metrics. The same study examined the impact of individual metrics and their combinations on defect proneness of a class. In [9] the same authors concluded that (Depth of Inheritance Tree) DIT is associated to defect proneness of a class inversely and NOC was deemed insignificant.

In [10], Yu et al. examined a part of a large network service management system written using Java to find correlation among metrics for predicting defect prone classes. A total of 8 metrics was used and 4 subsets with high correlation were found. It was concluded that (Coupling Between Objects) CBOin, (Response For a Class) RFCin and DIT insignificant predictors.

In [11, 12], the authors studied open source software systems for defect prediction. In [11], regression and machine learning methods were used over object-oriented metrics [12] to build models that predict defect prone classes. In both studies it was concluded that NOC is a weak predictor for at least 1 software system. In [12], it was further concluded that DIT is insignificant for 2 software systems.

Using the KC1 dataset, Zhou et al. [13] built defect prediction models that categorize defects by severity (high/low). Machine learning and regression techniques with Chidamber and Kemerer object-oriented metrics [4] were used in this study. The results

conclude that design level metrics are good predictors of defects having low severity compared to defects having high severity. The authors of [14] used Bayesian methods to associate product metrics to defect proneness over the same data set.

Kanmani et al. [15] used 2 neural network based models (probabilistic neural networks (PNN) and back-propagation neural networks (BPN)) to predict faults in a dataset generated from software modules written by students. They used object-oriented metrics in this study. The results suggested that PNN based models perform better among the two used in the study.

Aggarwal et al. [16] studied 12 software systems using 52 object-oriented metrics. It was discovered that import coupling and size metrics were related to defect proneness of classes while inheritance metrics found to be insignificant.

The authors of [17] examined the applicability of these techniques during its post release evolution. Three releases of Eclipse (eclipse.org) were used in this study and it was concluded that some metrics could predict defects at 3 levels of severity but the accuracy of prediction dropped from release to release.

Zhou et al. [18] re-examined the efficacy of complexity metrics in predicting class level defect proneness. Three versions of Eclipse (eclipse.org) were used to build models using 5 techniques namely, Naïve Bayes, Logistic Regression, Neural Network, ADTree, and KStar. It was concluded that LOC is the best predictor of defects followed by (Weighted Methods per Class) WMC. It was also found that when complexity metrics are used individually, their discriminating ability is moderate or nearly moderate.

Singh et al. [19] improved on the work done by [13]. They included the severity level 'medium' for the fact that defects having medium severity are more severe than defects having low severity and cannot be considered equal to them. They used machine learning techniques (Artificial Neural Network, Decision Tree) and discovered that CBO, WMC, LOC, and RFC are significant in predicting defects of all severity levels and the models for defects with low and medium severity levels have higher accuracies compared to models for defects with high severity.

In [20], the authors use AntMiner+, a derivative of the Ant Colony Optimization algorithm to predict defects in 3 public datasets. The results show that the AntMiner+ algorithm performs as well as machine learning algorithms.

The authors of [21] use the fault slip through metric to build defect prediction models using a variety of machine learning algorithms and 2 search based techniques, the Genetic Programming (GP) algorithm and the Artificial Immune Recognition System (AIRS). In their results, they found that GP has the highest recall along with the C4.5 Decision Tree algorithm and the Artificial Neural Network algorithm while AIRS reported the lowest recall among all selected algorithms.

In [22], the authors use a novel approach called the Evolutionary Decision Rules for Subgroup Discovery (EDRSD). They use 7 public datasets available on the PROMISE repository

to build defect prediction models. They found that the EDRSD technique is successful in identifying decision rules that characterize modules of defective nature. Rodríguez et al. [23] also uses the subgroup discovery approach for defect prediction and evaluates the prediction models on data available on the promise repository and the Eclipse project. The results show that subgroup discovery is effective in predicting defective modules in a software system.

Malhotra [24] analyses the applicability of 6 machine learning methods for defect prediction. The assessment is made on the results obtained on two public datasets. The results of this study show that decision trees are better than other machine learning and statistical techniques for defect prediction.

Only a few studies attempt to analyze the applicability of evolutionary computation in the domain of defect prediction in software. In this study we assess and evaluate the ability of evolutionary computation algorithms in regard of software defect prediction.

III. EVOLUTIONARY COMPUTATION

Evolutionary computation is a part of the computing technology wherein natural processes like navigation of a swarm

of birds, human reproduction, etc are modeled to develop algorithms that can learn and solve complex computing problems. For example, the Genetic Algorithm is inspired from the reproduction model of humans where every generation is formed by combining traits of chromosomes that are desired and simultaneously suppressing on the undesirable ones.

Evolutionary computing techniques bring along multiple advantages like superior ability to handle changing data, ability to perform in problems where heuristics are unavailable or unsatisfactory and simplicity of their approach [2].

The data analyzed by software engineering researchers is usually noisy and changes abruptly from system to system making it the perfect platform for evolutionary computing techniques. To make software engineering problems suitable for processing using evolutionary algorithms, performance measures like accuracy can be used as fitness functions to these algorithms [1].

In this study we assess the performance of 15 evolutionary computation algorithms. Table I provides a brief introduction to each algorithm and the configuration of the algorithm used for the purpose of this study.

TABLE I. DETAILS OF EVOLUTIONARY COMPUTATION TECHNIQUES

Abbreviation	Algorithm	Algorithm Description and Parameters of Test
CPSO	Constricted Particle Swarm Optimization [25]	It is a constricted particle swarm optimization technique. In this study we use 25 particles with a convergence radius of 0.1.
GANN	Artificial Neural Network with Genetic Algorithm [26]	An evolutionary artificial neural network what uses genetic algorithm to train the weights, select features, etc. GANN for this study uses 2 hidden layers comprised of 15 neurons. The algorithm works on 100 generations to report measures.
GAS-ADI	Genetic Algorithm Based Classifier [27]	The algorithm configuration used for this study performed 500 iterations with penalty=4.
GAS-I	Genetic Algorithm Based Classifier [28]	A Pittsburgh style classifier system based on GABIL. The algorithm configuration used in this study had fixed the number of iterations at 500 with penalty = 4.
GGA	Generational Genetic Algorithm [29]	The algorithm is configured to perform 10000 iterations over a population size of 51.
GFS-AB	GFS AdaBoost [30]	In this algorithm boosting is applied to fuzzy decision rules. The algorithm configuration has the number of labels and rules as 3 and 8 respectively.
GFS-LB	GFS LogitBoost [31]	A standard LogitBoost algorithm to learn fuzzy rules. The algorithm is configured to learn 3 labels and 25 rules.
GFS-MLP	GFS MaxLogitBoost [32]	A LogitBoost algorithm to learn fuzzy rules. It makes use of single winner inference. The configuration used in this study is, no. of labels = 3 and no. of rules = 8.
GFS SP	GFS SP [33]	Fuzzy Learning based on Genetic Programming Grammar Operators and Simulated Annealing. GFS SP is configured similar to GFS AdaBoost with the number of labels =3 and number of rules =8.
HIDER	Hierarchical Decision Rules [34]	An algorithm based on evolutionary computation that can learn in continuous and discrete domains. In this study, Hider uses a population size of 100 with 80 percent crossover, 50 percent mutation probability and a penalty factor of 1.
MPLCS	Memetic Pittsburgh Learning Classifier System [35]	MPLCS in this study is configured to perform 750 iterations with penalty = 4.
NNEP	Neural Network Evolutionary Programming [36]	It combines evolutionary learning with artificial neural networks. The algorithm is configured to use 4 hidden neurons and iterates over 200 generations.
REPSO	Real Encoding Particle Swarm Optimization [37]	It is a modified Particle Swarm Optimization algorithm that uses real encoding in place of binary encoding. This allows REPSO to be applied to both categorical and continuous data.
UCS	sUpervised Classifier System [38]	The number of explorations in UCS is fixed at 100000 with a population size of 6400.
XCS	X Classifier System [39]	A genetic algorithm implementation where fitness is determined by accuracy of prediction. XCS is configured similar to UCS, the population size is 6400 and 100000 iterations are performed using 2 point crossover method.

IV. RESEARCH METHODOLOGY

A. Dependent and Independent Variables

Defect proneness is the probability of a class having a positive outcome during defect detection [8]. In this study, defect proneness serves as the dependent variable. It is binary, i.e. a class is deemed defective if at least one defect is reported, and is defined at class level. To make predictions about a class being defective we use a set of 7 independent variables that is comprised of the Chidamber and Kemerer metric suite [4] and the Lines of Code metric (LOC). These metrics have been used by a large number of defect prediction studies in the past.

The data populated for any study can contain redundant and irrelevant features. An irrelevant feature does not provide the algorithm with meaningful information and a redundant feature only gives information that is available with some other feature in the dataset. To remove redundant and irrelevant features we performed correlation based feature sub selection on the dataset [40].

B. Empirical Data Collection

We use DCRS to generate the data used for this study. DCRS or Defect Collection and Reporting System is a Git repository mining tool that collects defect information from software systems maintained in Git repositories. DCRS also uses CKJM [41] to simultaneously calculate the metric values at class level for all classes present in the software system. To generate an output report, DCRS analyses the change logs of 2 versions of a software system that employ Git as a Version Control System (VCS). For instance, if version 1.0 (older version) and version 2.0 (newer version) of a software system are being analyzed; processing the change logs of version 2.0 generates the defect information and processing the classes in version 1.0 generates the metrics. However, DCRS only processes defects that are reported and fixed between version 1.0 and 2.0. Fig. 1 outlines the process flow of DCRS.

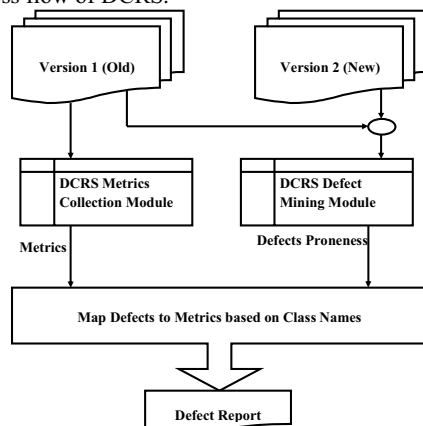


Fig. 1. Process flow of DCRS

Algorithmically, the following steps are executed by DCRS in order to generate a single report.

- Step 1: Filter out classes that are common to both versions of the software system.
- Step 2: Examine the older version of the two and populate object-oriented metrics for all classes.
- Step 3: Examine the change logs of the newer version of the software system to mine all defects that are fixed in the newer version. For each class, the number of defects fixed is counted.
- Step 4: Combine the outputs of step 2 and step 3 using class names as the common entity to generate the output report.

In this study we examine 5 datasets that are populated using DCRS.

1) Abdera (abdera.apache.org)

It is an implementation of the Atom Syndicate Format and the Atom Publishing protocol. This study uses version 1.0 for collection of metrics and version 1.1.2 for collection of defect data. The descriptive statistics for this dataset inform that the WMC lies between 0 and 347 with a mean value of 11.79. DIT and NOC lay between 0-4 and 0-26 respectively. CBO has a maximum value of 134 while RFC falls between 0-353. The LCOM has a mean value of 312.13 and lies between 0-59963. For LOC, the range is 0-187449 with a mean value of 498.96. The standard deviation is 23.079, 0.720, 1.806, 11.953, 29.359, 2672.269, and 7713.158 for WMC, DIT, NOC, CBO, RFC, LCOM and LOC respectively.

2) Betwixt (commons.apache.org/proper/commons-betwixt/)

It is a library that provides an XML introspection mechanism for mapping beans to XML. Using DCRS we examine version 0.7 for collection of object-oriented metrics and version 0.8 for procuring defect information. The standard deviation of independent variables for this dataset is 7.579, 0.532, 3.072, 4.753, 9.211, 174.402, and 164.605 for WMC, DIT, NOC, CBO, RFC, LCOM and LOC respectively. All independent variables have a minimum value of 0. The maximum values are WMC=63, DIT=4, NOC=57, CBO=58, RFC=64, LCOM=1953 and LOC=2760.

3) Collections (commons.apache.org)

Collections is an Java Collections Framework that includes data structures and prewritten functions to speed up development. Version 3.3of Collections is processed using DCRS to populate the independent variables (object-oriented metrics) and version 4.0 serves to provide defect information. The independent variables in this dataset have minimum value of 0. The maximum values for WMC, DIT, NOC, CBO, RFC, LCOM and LOC are 83, 5, 10, 59, 179, 3041 and 2139 respectively.

4) Commons IO (commons.apache.org)

Commons IO is a library that assists with IO functionality development. We examine version 2.3 and 2.4 of this software package using DCRS. Version 2.3 provides object-oriented

metrics while version 2.4 provides defect related information. The independent variables for this dataset have a minimum value 0 for DIT, NOC, CBO, and LCOM while WMC, RFC and LOC have a minimum value of 2. The maximum values are WMC=127, DIT=4, NOC=24, CBO=24, RFC=233, LCOM=8001, LOC=3168.

5) Apache Commons Net (*commons.apache.org*)

Apache Commons Net implements the client-side of many basic Internet protocols. Version 3.2 of Commons Net is used for populating metrics and version 3.3 provides defect information. Upon examining the dataset further it was observed that the minimum values for all independent variables in this dataset is 0 while the maximum values are 127, 4, 11, 26, 282, 7547, and 2378 for WMC, DIT, NOC, CBO, RFC, LCOM and LOC respectively. Table II provides the details of each dataset.

TABLE II. DETAILS OF DATASETS USED

Data Set	Lines of Code	No. Of Classes	No. Of Defective Classes	No. Of Non-Defective Classes	% Of Defective Classes
Abdera	298875	599	158	441	26.64
Betwixt	22349	356	70	286	19.67
Collections	83093	511	333	178	65.17
IO	29200	198	53	145	26.77
Net	45010	192	106	86	55.20

C. Performance Measures

Once the datasets are processed and models have been built, we use a carefully selected set of performance measures to evaluate their prediction capability. Below we have listed the performance measures put to use for model evaluation.

1) Recall

It is the ratio of the number of classes correctly identified as defective by the model to the actual number of defective classes in the system. It is also called True Positive Rate (TPR) or Sensitivity of the model.

2) Precision

It is the ratio of the number of correctly identified defective classes to the total number of classes identified as defective by the model.

3) Accuracy

It is the ratio of the number of correct predictions made by the model to the total number of classes processed by the model.

There are other performance measures like f-Measure but their stability while dealing with defect prediction problems has been questioned in [42].

V. RESULTS

In this section we discuss the results of this study. We assess and evaluate the performance of 15 evolutionary computation algorithms over 5 datasets populated by mining repositories of Apache Software Foundation. The data used in this study was populated using DCRS. A total of 7 metrics is used as

independent variables. We performed correlation based feature sub selection to remove redundant and irrelevant features from the dataset. Table III shows the metrics found relevant in all datasets after the application of CFS,

TABLE III. RELEVANT FEATURES

Dataset	Relevant Features
Abdera	WMC, RFC, LCOM, LOC
Betwixt	CBO, RFC
Collections	WMC, DIT, RFC, LCOM, LOC
IO	RFC, LCOM, LOC
Net	RFC, LCOM, LOC

All metrics except those found relevant are discarded and provided as input to algorithms for building models. The results are validated using 10-cross validation. In 10-cross validation, the dataset is divided into 10 parts. Out of 10, 9 parts are used to train the model while the remaining 1 part is used for evaluating the performance of the trained model. This is performed 10 times on different test partitions hence the name 10-cross validation.

The remainder of this section shows the results of the study on all five datasets.

A. Abdera

The dataset contains 26.64 percent defective classes. Table 4 provides the results of applying 15 selected evolutionary computation techniques for defect prediction on the Abdera dataset. The results show that the accuracy of defect prediction model is high. All algorithms have reported accuracy over 70%. As can be seen from Table IV, all the techniques have accuracy greater than 70 percent. The model built using MPLCS technique reported the best accuracy at 76.63 percent.

TABLE IV. ABDERA RESULTS

Algorithm	Recall	Precision	Accuracy
CPSO	3.80	23.08	71.29
GANN	18.35	53.70	74.29
GAS-I	15.19	57.14	74.62
GAS-ADI	19.62	64.58	75.96
GFS-AB	14.56	71.88	75.96
GFS LB	16.46	70.27	76.13
GFS MLB	8.23	68.42	74.79
GFS SP	21.52	60.71	75.63
GGA	23.42	61.67	75.96
HIDER	17.72	68.29	76.13
MPLCS	20.25	69.57	76.63
NNEP	18.35	65.91	75.96
REPSO	0.00	0.00	73.62
UCS	11.39	48.65	73.46
XCS	7.01	78.57	75.04

B. Betwixt

This dataset contains the least number of defective classes. It has only 19.67 percent defective classes. As can be seen from Table V the accuracy for all models is beyond 80% except GGA and CPSO, which reported an accuracy of 79.49% and 76.12% respectively. The best accuracy was reported by MPLCS at 83.43 percent.

TABLE V. BETWIXT RESULTS

Algorithm	Recall	Precision	Accuracy
CPSO	17.14	30.77	76.12
GANN	10.00	53.85	80.62
GAS-I	15.71	91.67	83.15
GAS-ADI	15.71	84.62	82.87
GFS-AB	11.43	88.89	82.30
GFS LB	14.29	90.91	82.87
GFS MLB	17.14	85.71	83.15
GFS SP	12.86	64.29	81.46
GGA	7.14	38.46	79.49
HIDER	10.00	53.85	80.62
MPLCS	15.71	100.00	83.43
NNEP	11.43	72.73	81.74
REPSO	0.00	0.00	80.06
UCS	15.71	84.62	82.87
XCS	12.86	90.00	82.58

C. Collections

Apache Commons Collection is the dataset has the highest percentage of defective classes (65.17%). The accuracy of models build for Collections has dropped in comparison to the previous two datasets. GAS-ADI reports the best accuracy of 65.17 percent. GGA reported the poorest accuracy at 59.30. Table VI provides the details of the results obtained on Apache Commons Collection.

TABLE VI. COLLECTIONS RESULTS

Algorithm	Recall	Precision	Accuracy
CPSO	82.88	65.71	60.67
GANN	98.20	64.88	64.19
GAS-I	96.70	64.79	63.60
GAS-ADI	100.0	65.17	65.17
GFS-AB	99.70	65.10	64.97
GFS LB	97.30	64.54	63.41
GFS MLB	99.70	65.10	64.97
GFS SP	95.50	64.50	62.82
GGA	89.19	63.33	59.30
HIDER	97.60	64.87	63.99
MPLCS	97.90	64.68	63.80
NNEP	96.40	64.59	63.21
REPSO	98.80	64.89	64.38
UCS	86.49	65.75	61.84
XCS	90.39	65.58	62.82

D. IO

The Apache Commons IO had a defect percentage of 26.77. Table VII provides the performance measures for the model built for this dataset. The best accuracy of 77.27 percent is reported by NNEP. The lowest accuracy is given by CPSO at 70.20 percent.

TABLE VII. IO RESULTS

Algorithm	Recall	Precision	Accuracy
CPSO	20.75	39.29	70.20
GANN	35.85	55.88	75.25
GAS-I	30.19	55.17	74.75
GAS-ADI	22.64	63.16	75.76
GFS-AB	11.32	50.00	73.23
GFS LB	9.43	45.45	72.73

GFS MLB	9.43	50.00	73.23
GFS SP	28.30	60.00	75.76
GGA	30.19	64.00	76.77
HIDER	28.30	57.69	75.25
MPLCS	15.09	57.14	74.24
NNEP	30.19	66.67	77.27
REPSO	0.00	0.00	72.73
UCS	15.09	53.33	73.74
XCS	3.85	66.67	74.11

E. Net

Table VIII shows the results obtained for Apache Commons Net. It is clear from the results that nearly all algorithms have excellent recall with REPSO having perfect recall. The precision, however, for REPSO was better than only one algorithm i.e. XCS. Since precision and recall for Apache Commons Net is more than the accuracy for all algorithms we can say that the percentage of correctly predicted defective classes for this dataset is excellent. Once again NNEP performed better than all other algorithms with an accuracy of 81.25.

TABLE VIII. NET RESULTS

Algorithm	Recall	Precision	Accuracy
CPSO	84.91	67.67	69.27
GANN	66.98	81.61	73.44
GAS-I	75.47	81.63	77.08
GAS-ADI	73.58	86.67	79.17
GFS-AB	79.25	82.35	79.17
GFS LB	76.42	88.04	81.25
GFS MLB	71.70	91.57	80.73
GFS SP	75.47	86.96	80.21
GGA	74.53	88.76	80.73
HIDER	74.53	83.16	77.60
MPLCS	74.53	86.81	79.69
NNEP	75.47	88.89	81.25
REPSO	100.0	61.63	65.63
UCS	68.87	81.11	73.96
XCS	92.45	59.39	60.94

VI. VALIDATION OF RESULTS

For the purpose of result validation we make use of the Friedman test. Friedman test is a non-parametric statistical test. This test is useful in determining the difference in performance of algorithms over a collection of data.

To begin, we first define the hypotheses for test.

- H_0 : there is no difference among performance of algorithms.
- H_1 : there is difference in performance of algorithms.

The next step is to set the significance level. The significance level for the purpose of this study is 0.05. The significance level of the Friedman test determines the probability of getting a Friedman statistic as good as the one returned by the test.

Upon setting the parameters, we provide the accuracy reported by all 15 algorithms over the 5 datasets as input to the Friedman test. Table IX presents the parameters of test for Friedman's test.

TABLE IX. FRIEDMAN TEST OUTPUT PARAMETERS

Parameter	Value
Chi-Square	27.193
DOF	14
p-Value	.018

The degrees of freedom (DOF) is equal to one less than the number of applied algorithms. In this study, we apply use 15 algorithms, which bring the degrees of freedom to 14.

Using the DOF parameter and the significance level, we look up the chi-2 tables to determine the critical value for the test. The critical value is the minimum value that the Friedman statistic should return to enable us to reject the H_0 hypotheses. Alternatively, we can check if the p-Value reported by the test is smaller than the significance level.

For 14 degrees of freedom at the significance level of 0.05 the critical value is 23.685. Since the calculated value of chi-square in the Friedman test is greater than the critical value determined from the table we reject the H_0 hypotheses and conclude that these is a difference among the performance of these algorithms. The difference is given by the ranks provided by the Friedman test to individual algorithms. Table X shows the ranking for all algorithms.

TABLE X. RESULTS OF FRIEDMAN TEST

Algorithm	Rank
GAS-ADI	4.50
MPLCS	4.60
GGA	5.40
GFS SP	5.60
GFS-LB	6.30
HIDER	6.90
GFS-AB	7.00
GAS-I	7.70
NNEP	7.90
GANN	8.40
GFS MLB	8.90
XCS	10.30
UCS	10.60
REPSO	11.50
CPSO	14.40

The best performing algorithm is given the lowest score for Rank and the algorithm with the highest Rank is the algorithm whose performance was the poorest. From Table 10 is can be seen that both PSO implementations are at the bottom (CPSO has rank=14.4 and REPSO has rank=11.5). GAS-ADI reported the lowest rank i.e. 4.50 making it the best performing technique in this study in terms of accuracy.

VII. DISCUSSION OF RESULTS

From the above results it is clear that the evolutionary computation techniques have the capability to predict software defects with high accuracy. Fig. 2 shows the accuracy of models across datasets. It is also clear that the performance of these algorithms in terms of recall is largely dependent on the percentage of defective classes in the input dataset. The datasets, which had a high number of defective classes, recorded excellent

recall in contrast to the datasets with less number of defective classes. Menzies et al. also observed the same [42].

When we look at precision, there was no direct relationship between precision and percentage of defective classes in software. This can be confirmed by evaluating the results for Betwixt and Net, both have showed excellent precision but recall for Betwixt was not good. On the other hand, Net had excellent recall.

The study examined 15 evolutionary and hybridized evolutionary techniques for their capability in predicting software defects. Our analysis suggests that GAS-ADI was the best performing algorithm when all datasets and performance measures are taken into consideration. The LCS methods, XCS and UCS, and the (Particle Swarm Optimization) PSO methods did not perform well.

A. Limitations of the Study

There exist some factors that might interfere with the correctness of the results in this study and limit its efficacy. While we have taken care to avoid any errors in the population, processing and evaluation of data during the course of this study it is imperative that we report issues that might have crept in. Malhotra et al. [43] found that the following 3 areas of concern are common to all empirical studies,

1) Construct validity

This refers to the accuracy at which the dependent and independent variables measure the proposed concept [43]. Since we make use of tools to measure object-oriented metrics and defects, this threat does not exist in our work to large extent.

2) Internal validity

Zhou et al. [13] define internal validity as “the degree to which conclusions can be drawn about the casual effect of independent variables on the dependent variables”. In this study, we only identify the relationship between object-oriented metrics and defect proneness. Since we are not concerned with the causal effect of object-oriented metrics on defect proneness this threat is not present in this study.

3) External validity

External validity refers to the extent the results of an empirical study are generalized. By including multiple datasets where each dataset has its own properties we can achieve generalization of results [43]. We have reduced this threat to large extent by evaluating results 5 open source software systems. Hence, the conclusions presented in this study are generalized. By performing replicated studies with varied data sets, we can further reduce this threat.

In totality, evolutionary computation techniques are definitely an area to explore further. Till date, no study has attempted to empirically validate their predictive capability for software defects. It is important that this domain be explored to further gain insight about the predictive accuracy of evolutionary

computing techniques and hybridized evolutionary computing techniques for predicting software defects using software systems of varying sizes.

In future, more studies of this nature should be performed to establish more evidence to support the results found in this research.

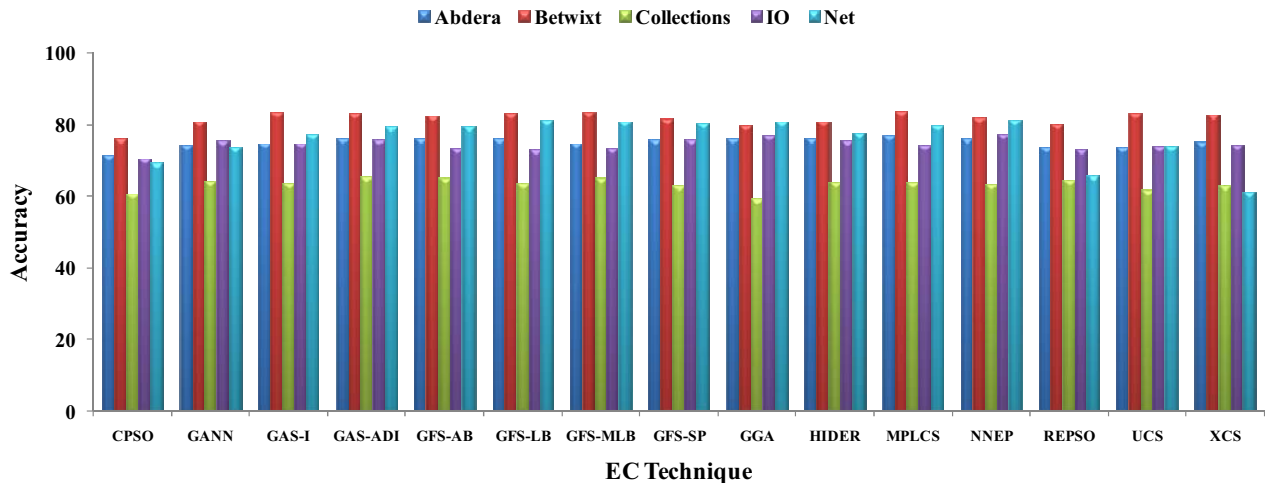


Fig. 2. Accuracy of models for all datasets

VIII. CONCLUSION AND FUTURE SCOPE

We examined 15 evolutionary and hybridized evolutionary techniques for their predictive capability in software defect prediction. To execute this study, we populated defect reports on 5 open source software systems using DCRS. The Apache Software Foundation maintains the selected software systems. We also evaluated the difference in performance of all 15 algorithms using the non-parametric Friedman test. Based on our findings we can conclude that evolutionary computing techniques are powerful methods for predicting software defects. The accuracy reported by all algorithms was high across all input datasets. Neural network techniques enhanced using evolutionary computation performed consistently well with NNEP performing slightly better than GANN. The best performing algorithm in this study was found to be GAS-ADI. This conclusion was made upon exhaustive comparison of its accuracies for all datasets with all other algorithms. The Friedman test also supports our finding and reports GAS-ADI with the least rank of 4.50. LCS and PSO based models did not perform as well as other algorithms and the same was reported by the Friedman test which ranked XCS, UCS, REPSO and CPSO at 10.30, 10.60, 11.50 and 14.40, respectively.

This work can be used as a starting point for future empirical research that examines the ability of evolutionary computing techniques to other software engineering problems. The industry can also benefit from the conclusions drawn in this study. By including defect prediction processes in their existing

development models the practitioners can better utilize their testing resources and simultaneously achieve better testing coverage. Evolutionary computing techniques and hybridized evolutionary computing techniques reported high accuracy for software defect prediction and the same models can be used by the industry. Some conclusions that can be applied to future research and industry are,

- The metrics RFC, LCOM and LOC were found to be significant attributes and can be used to predict defects in early stages of the software lifecycle, particularly during design.
- Although traditional machine learning based models built in earlier studies have shown promising results the accuracy reported by all defect prediction models was also high. Since theoretically, evolutionary computation techniques are better suited to software engineering problems [1] it is natural to test these models further on more complex data to further evaluate their ability to predict defects.

The results provided in this study are general since the 5 datasets showed healthy variation in terms of defective class percentage, total number classes as well as metric values

REFERENCES

- [1] M. Harman, "Why the Virtual Nature of Software makes it Ideal for Search Based Optimization", *Fundamental Approaches to Software Engineering*, 2010.
- [2] C. Grosan, and A. Abraham, "Hybrid Evolutionary Algorithms: Methodologies, Architectures, and Reviews", *Studies in Computational Intelligence*, vol. 75, pp. 1-17, 2011

- [3] V. Basili, L. Briand, and W. Melo, "A validation of object-oriented design metrics as quality indicators", *IEEE Transactions on Software Engineering*, vol. 22, No. 10, pp. 751-761, 1996
- [4] S. Chidamber, and C. Kemerer, "A metrics suite for object oriented design", *IEEE Transactions on Software Engineering*, vol. 20, no. 6, pp. 476-493, 1994
- [5] M.H. Tang, M.H. Kao, and M.H. Chan, "An empirical study on object-oriented metrics", in *Proceedings of Metrics*, pp. 242-249, 1999
- [6] K. El Emam, S. Benlarbi, N. Goel, and S. Rai, "The confounding effect of class size on the validity of object-oriented metrics", *IEEE Transactions on Software Engineering*, vol. 27, no. 7, pp. 630-650, 2001
- [7] K.E. Emam, W. Melo, and J.C. Machado, "The prediction of faulty classes using object-oriented design metrics", *The Journal of Systems and Software*, pp. 63-75, 2001
- [8] L. Briand, W. Daly, and J. Wust, "Exploring the relationships between design measures and software quality", *Journal of Systems and Software*, vol. 51, no. 3, pp. 245-273, 2000
- [9] L. Briand, J. Wust, and H. Lounis, "Replicated Case Studies for Investigating Quality Factors in Object-Oriented Designs", *Empirical Software Engineering*, vol. 6, no. 1, pp. 11-58, 2001
- [10] P. Yu, T. Systa, and H. Muller, "Predicting fault-proneness using OO metrics: An industrial case study", in *Proceedings of Sixth European Conference on Software Maintenance and Reengineering*, pp. 99-107, 2002
- [11] T. Gyimothy, R. Ferenc, and I. Siket, "Empirical validation of object-oriented metrics on open source software for fault prediction", *IEEE Transactions on Software Engineering*, vol. 31, no. 10, 897-910, 2005
- [12] H. Olague, L. Etzkorn, S. Gholston, and S. Quattlebaum, "Empirical validation of three software metrics suites to predict fault-proneness of object-oriented classes developed using highly iterative or agile software development processed", *IEEE Transactions on Software Engineering*, vol. 33, no. 8, pp. 402-419, 2004
- [13] Y. Zhou, and H. Leung, "Empirical analysis of object-oriented design metrics for predicting high severity faults", *IEEE Transactions on Software Engineering*, vol. 32, no. 10, pp. 771-784, 2006
- [14] G.J. Pai, and J.B. Dugan, "Empirical analysis of software fault content and fault proneness using Bayesian methods", *IEEE Transactions on Software Engineering*, vol. 33, no. 10, pp. 675-686, 2007
- [15] S. Kanmani, V.R. Uthariaraj, V. Sankaranarayanan, and P. Thambidurai, "Object oriented software fault prediction using neural networks", *Information and Software Technology*, vol. 49, no. 5, pp. 483-492, 2007
- [16] K.K. Aggarwal, Y. Singh, A. Kaur, and R. Malhotra, "Empirical Analysis for Investigating the Effect of Object-Oriented Metrics on Fault Proneness: A Replicated Case Study", *Software Process Improvement and Practice*, John Wiley & Sons, vol. 16, no. 1, pp. 39-62, 2009
- [17] R. Shatnawi, and W. Li, "The effectiveness of software metrics in identifying error prone classes in post-release software evolution process", *Journal of Systems and Software*, vol. 81, no. 11, pp. 1868-1882, 2008
- [18] Y. Zhou, B. Xu, and H. Leung, "On the ability of complexity metrics to predict fault-prone classes in object-oriented systems", *The Journal of Systems and Software*, 2010
- [19] Y. Singh, A. Kaur, and R. Malhotra, "Empirical validation of object-oriented metrics for predicting fault proneness models", *Software Quality Journal*, pp. 3-35, 2010
- [20] O. Vandecruys, D. Martens, B. Baesens, C. Mues, M. Backer, and R. Haesen, "Mining software repositories for comprehensible software fault prediction models", *Journal of Systems and Software*, 2008
- [21] W. Afzal, "Using Faults-Slip-Through Metric As A Predictor of Fault-Proneness", in *Asia Pacific Software Engineering Conference*, 2010
- [22] D. Rodriguez, R. Ruiz, J.C. Riquelme, and J.S. Aguilar-Ruiz, "Searching for rules to detect defective modules: A subgroup discovery approach", *Information Sciences*, 2012
- [23] D. Rodriguez, R. Ruiz, J. C. Riquelme, and R. Harrison, "A study of subgroup discovery approaches for defect prediction", *Information and Software Technology*, 2013
- [24] R. Malhotra, "Comparative Analysis of Statistical and Machine Learning Methods for Predicting Faulty Modules", *Applied Soft Computing*, 2014
- [25] T. Sousa, A. Silva, and A. Neves, "Particle Swarm based Data Mining Algorithms for classification tasks", *Parallel Computing*, vol. 30, pp. 767-783, 2004
- [26] X. Yao, "Evolving Artificial Neural Networks", in *Proceedings of the IEEE*, vol. 87, no. 9, 1999
- [27] J. Bacardit, and J.M. Garrell, "Evolving multiple discretizations with adaptive intervals for a Pittsburgh rule-based learning classifier system", in *Genetic and Evolutionary Computation Conference (GECCO'03)*. LNCS 2724, Springer 2003, Chicago (Illinois USA, 2003) pp. 1818-1831, 2003
- [28] J. Bacardit, and J.M. Garrell, "Bloat control and generalization pressure using the minimum description length principle for a Pittsburgh approach Learning Classifier System", *Lecture Notes in Computer Science*, 4399, pp. 59-79, 2007
- [29] J.R. Cano, F. Herrera, and M. Lozano, "Using Evolutionary Algorithms as Instance Selection for Data Reduction in KDD: An Experimental Study", *IEEE Transactions on Evolutionary Computation* 7:6, pp. 561-575, 2003
- [30] M.J. del Jesus, F. Hoffmann, L.J. Navascués, and L. Sánchez, "Induction of Fuzzy-Rule-Based Classifiers with Evolutionary Boosting Algorithms", *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 3, 2004
- [31] J. Otero, and L. Sánchez, "Induction of Descriptive Fuzzy Classifiers with the Logitboost Algorithm". In *Soft Computing*, 10:9, pp. 825-835, 2006
- [32] L. Sánchez, and J. Otero, "Boosting Fuzzy Rules in Classification Problems under Single-Winner Inference", *International Journal of Intelligent Systems*, 22:9, pp. 1021-1034, 2007
- [33] L. Sanchez, I. Couso, and J.A. Corrales, "Combining GP operators with SA search to evolve rule based classifiers", *Recent Advances in Genetic Fuzzy Systems*, vol. 136, no. 1-4, pp. 175-191, 2001
- [34] J.A. Aguilar-Ruiz, J.C. Riquelme, and M. Toro, "Evolutionary Learning of Hierarchical Decision Rules", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 33, no. 2, pp. 324-331, 2003
- [35] J. Bacardit, and N. Krasnogor, "Performance and Efficiency of Memetic Pittsburgh Learning Classifier Systems", *Evol Comput Fall*, vol. 17, no. 3, pp. 307-42, 2009
- [36] F.J. Marti'nez-Estudillo, C. Herva's-Marti'nez, P.A. Gutie'rrez, and A.C. Marti'nez-Estudillo, "Evolutionary product-unit neural networks classifiers", *Machine Learning for Signal Processing (MLSP 2006) / Life System Modeling, Simulation, and Bio-inspired Computing (LSMS 2007)*, vol. 72, no. 1-3, pp. 548-561, 2008
- [37] Y. Liu, Z. Qin, Z. Shi, and J. Chen, "Rule Discovery with Particle Swarm Optimization", in *Advanced Workshop on Content Computing (AWCC)*. LNCS 3309, Springer 2004, pp. 291-296, 2004
- [38] E. Bernad'o-Mansilla, and J.M. Garrell-Guiu, "Accuracy-Based Learning Classifier Systems: Models, Analysis and Applications to Classification Tasks", *Evolutionary Computation archive*, vol. 11, no. 3, pp. 209 - 238, 2003
- [39] S.W. Wilson, "Classifier Fitness Based on Accuracy". *Evolutionary Computation archive*, vol. 3, no. 2, pp. 149-175, 1995
- [40] M.A. Hall, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning", Working Paper 00/8, May 2000
- [41] ckjm — Chidamber and Kemerer Java Metrics — <http://www.spinellis.gr/sw/ckjm/>
- [42] T. Menzies, A. Dekhtyar, J. Distefano, and J. Greenwald, "Problems with precision: A response to 'comments on 'data mining static code attributes to learn defect predictors' " ", *IEEE Trans. Softw. Eng.*, vol. 33, no. 9, pp. 37-640, 2007
- [43] R. Malhotra, and M. Khanna, "Investigation of Relationship between Object-oriented Metrics and Change Proneness," *Int. J. Mach. Learn. & Cyber., Springer-Verlag*, vol. 4, no. 4, pp. 273-286, 2013

Performance analysis of ball end magnetorheological finishing using sintered magnetic abrasive based magnetorheological polishing fluid

Mahendra Singh Niranjana¹ Sunil Jha^{2*}

¹ Department of Mechanical Engineering, Delhi Technological University, New Delhi 110042, India. Email mahendraiitr2002@gmail.com

^{2*} Department of Mechanical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India, Email: suniljha@mech.iitd.ac.in , suniljha@gmail.com
Tel.: +91-11-2659-1125; Fax: +91-11-2658-2053.

Abstract

A scheme to finish surface of mild steel in nanometer level and obtain defect free surface/subsurface with sintered magnetic abrasives based magnetorheological polishing fluid (MRPF) on ball end magnetorheological finishing (BEMRF) tool is presented. Sintered magnetic abrasives were developed by uniformly mixing of 20 vol% carbonyl iron powder (CIP) CS grade and 25 vol% silicon carbide abrasives then sintered in tubular furnace at 1200°C using pressureless solid phase sintering method in argon atmosphere. The sintered product was crushed in ball mill to obtain sintered magnetic abrasives. Particle characterization was performed with scanning electron microscope and vibration sample magnetometer to characterize the morphology and magnetizability of sintered magnetic abrasives. MRPF was prepared with 55 vol% base fluid and 45 vol% sintered magnetic abrasives. Experiment was conducted on mild steel surface with given set of experimental conditions for 30 minutes on BEMRF tool and surface roughness before and after finishing were measured using Talysurf. The percent reduction in surface roughness ($\% \Delta R_a$) was calculated and compared with $\% \Delta R_a$ obtained by finishing the work-piece surface with unbonded magnetic abrasives based MRPF having same composition and found superior results. After comparison, a five level and three factors central composite design of experiments (DOE) was selected to develop the experimentation plan. Response surface methodology (RSM) was adopted to predict the profile and $\% \Delta R_a$ with given set of experimental conditions. An effective second order response surface model was developed with significant factors and their interactions. Physical interpretation of effect of machining parameters on $\% \Delta R_a$ was analyzed.

Keywords: Magnetorheological, Characterization, Sintering, MR finishing, DOE, RSM.

1. Introduction

Precision finishing of complex geometries which are difficult to machine by traditional machining system is always demanded in the present technological world. With the advent of new materials, the engineers are facing challenge for the finishing of these materials to meet their functional requirements. The conventional finishing processes alone are incapable to develop the desired surface characteristics on complex geometry and in exercising in-process control on finishing action. Advanced finishing processes like abrasive flow machining (AFM), magnetic float polishing (MFP), magnetic abrasive finishing (MAF) and magnetorheological finishing (MRF) have been evolved in the last few decades. AFM is used to finish complicated geometry by extruding abrasive laden polymeric medium through the work-piece but rheological properties of polymeric medium cannot be controlled by external means. Kordonski et al describe that magnetorheological (MR) finishing is most promising smart finishing methods applied for ultra

fine surface finishing. Golini et al has used this process for external finishing of optical lenses to the nanometer level in which forces can be controlled by external magnetic field. Shorey et al has considered this process as an advanced polishing technique to finish optics without propagating the subsurface damage layer. Kordonski et al described that the removal of subsurface damage by this process may be attributed to small normal stress applied on glass surface compared to large shear stress created by the interaction of tool's magnetic field and MR polishing fluid through the converging gap between the optic and tool's pole faces. Kim et al stated that the process can be used to finish variety of surfaces such as concave, convex, three dimensional and aspherical optical components. Seok et al has fabricated a curved surface on silicon based structure and surface characteristics were evaluated. Finite element method was used to study the magnetic field effect near the tool assembly on finished surface profile. Response surface methodology (RSM) has been used to predict the profile and roughness of curved

Performance analysis of ball end magnetorheological finishing using sintered magnetic abrasive based magnetorheological polishing fluid

work-piece surface at the given set of experimental parameters. Sadiq et al has used the process to finish external curved surfaces by giving rotation while MR polishing fluid was pushed up and down. The experimental study was conducted on aluminum and austenitic stainless steel to see the magnetic field effect on surface finish. Jung et al described that the process is not generally applied for finishing hard surface materials like Al_2O_3 -TiC hard disc slider. It may be due to wheel rotational speed which limits impressed abrasion energy and hence low material removal rate. These limitations were overcome by using sintered iron CNT compound abrasives based MRP fluid for MR finishing of hard materials. These abrasives increase the life time of consumables and material removal rate. Singh et al developed a new precision finishing process called ball end magnetorheological finishing (BEMRF) for nano level finishing of ferromagnetic/diamagnetic work-piece surface of 3D shapes using MRP fluid. The experiments were conducted on EN-31 and groove surface of copper to study the finishing time effect on surface roughness. The process has also been used to finish typical 3D work-piece surface obtained by milling process having different angle of projections like flat, 30° , 45° and curved surface. The experimental study has been done to see the effect of finishing passes on final surface finish. Finite element analysis has also been done to study the magnetic flux density distribution at tool tip surface with inserted typical 3D work-piece surface.

Literature survey revealed that the tool rotational speed limits the impressed abrasion energy on the work-piece surface and reduced material removal rate. The non magnetic abrasives which are held firmly by surrounding CIPs threw away at higher tool rotational speed and reduces shearing action. In the present work, the sintered magnetic abrasives based MRP fluid were planned to use on BEMRF tool to improve the material shearing action. The response surface model was developed for $\%\Delta R_a$ based on machining conditions using RSM. RSM was selected for identifying machining conditions influencing $\%\Delta R_a$ and the predictive model adequacy has been studied by analysis of variance.

2. Preparation of sample

2.1 Development of sintered magnetic abrasives

Sintered magnetic abrasives were developed by uniformly mixing 20 vol% CIP CS grade and 25 vol% silicon carbide (SiC) abrasives of 800 mesh sizes in ball mill. The mixed powder in alumina tray was allowed to sinter in tubular furnace in controlled argon atmosphere with temperature increment 4°C per minute reaches a

temperature 800°C and hold one hour at this temperature for complete degasification. Then temperature was increased with the same rate upto 1200°C and holds three hours at this temperature for getting phase transformation completely. After sintering, furnace cooling was done in argon atmosphere till the environment temperature 25°C achieved. The sintered products were then crushed in ball mill with tungsten carbide balls to obtain the sintered magnetic abrasives. Hence sintered magnetic abrasives composed of ferromagnetic phase and abrasive phase which are combined firmly.

2.2 Characterization of sintered magnetic abrasives

Particle characterization was performed on sintered magnetic abrasives to understand the MR effect and stability of synthesized MRP fluid. Sintered magnetic abrasives were studied with scanning electron microscope (SEM) and vibration sample magnetometer (VSM) to study morphology and magnetizability of sintered magnetic abrasives.

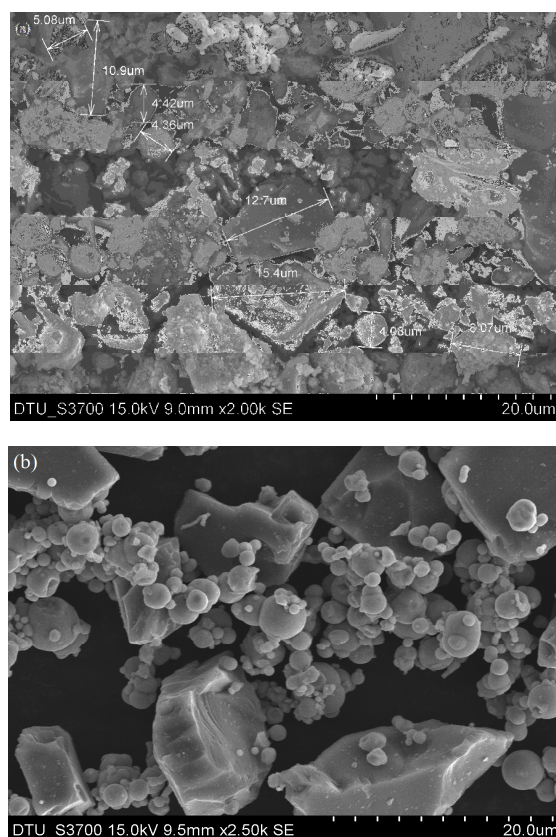


Figure 1 Scanning Electron Micrograph of (a) Sintered magnetic abrasives (b) Unbonded magnetic abrasives

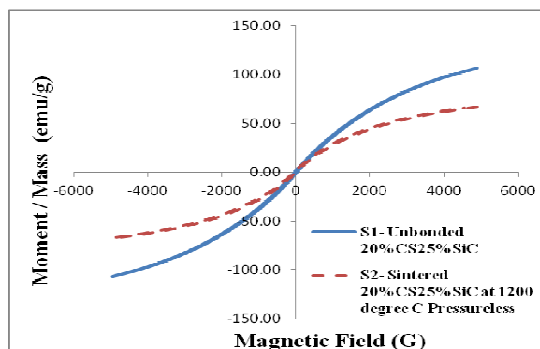


Figure 2 Vibration sample magnetometry data

2.3 Synthesis of magnetorheological polishing fluid

Unbonded magnetic abrasives based MRPF (Sample1) was synthesized with 20 vol% CIP CS grade, 25 vol% SiC abrasives and 55 vol% base medium. Another sintered magnetic abrasives based MRPF (sample2) was synthesized with composition of 45 vol% Sintered magnetic abrasives and 55 vol% viscoplastic base medium (20% by weight AP3 grease and 80% by weight heavy paraffin liquid).

2.4 Experiment

The experimental study has been done by synthesized MRPF (Sample 2) on mild steel work-piece surface at given machining parameters shown in table 1 with 50 mm/min work-piece feed rate for 30 minutes on BEMRF tool. Initial and final surface roughness were measured at the same point with data length of 4 mm and cut off evaluation length 0.25 mm with Taylor Hobson Talysurf and $\% \Delta R_a$ was calculated. The experiment was further conducted with unbonded magnetic abrasives based MRPF (Sample 1) on mild steel surface with same machining conditions then $\% \Delta R_a$ was calculated and given in table 2.

Table 1 Experimental parameters and conditions

Parameters	Conditions
Tool rotational speed (RPM)	334
Working Gap (mm)	0.75
DC power supply	4 A
Total finishing time (minute)	30
Abrasive- Silicon carbide (SiC)	800 mesh size

Table 2 Surface roughness before and after MR finishing

MRP fluid sample Number	Initial Roughness (μm)	Final Roughness (μm)	$\% \Delta R_a$
Sample 1 (Unbonded)	R_a 0.2163	R_a 0.1755	19%
	R_q 0.2753	R_q 0.2213	20%
	R_z 1.7240	R_z 1.5187	15%
Sample 2 (Sintered)	R_a 0.1908	R_a 0.1336	29.97%
	R_q 0.2451	R_q 0.1720	29.83%
	R_z 1.2138	R_z 0.8832	27.24%

The experimental plan was developed with Central composite design of experiment (DOE) and machining parameters such as magnetizing current, tool rotational speed and working gap were optimized for the response $\% \Delta R_a$. Design expert software was selected to optimize the machining parameters in BEMRF process. The selected machining parameters and their levels as given by Design expert software are shown in Table 3.

Table 3 Coded levels and corresponding actual values of machining parameters

Parameter	Unit	Levels				
		-1.682	-1	0	1	1.682
Magnetizing current (I)	A	2.3	3	4	5	5.7
Tool rotational speed (N)	RP M	332	400	500	600	668
Working gap (G)	mm	0.66	1	1.5	2	2.3

3. Result analysis and discussion

The morphology of magnetic abrasives has been studied by scanning electron microscope as shown in fig.1. The sharp edges of sintered magnetic abrasives are again exposed after crushing the sintered products in ball mill and shapes and sizes of different magnetic abrasive grains are shown in fig.1 (a). The CIP is

Performance analysis of ball end magnetorheological finishing using sintered magnetic abrasive based magnetorheological polishing fluid

diffused on the surface of abrasives and become magnetic abrasives. VSM was used to study the magnetic properties of sintered magnetic abrasives and unbonded magnetic abrasives with the same composition as shown in fig.2. Maximum magnetization was observed as 106.43 emu/g for unbonded magnetic abrasives and 67.0 emu/g for sintered magnetic abrasives at 4893.9 gauss magnetic field. The magnetic property is reduced due to phase transformation of CIP from ferromagnetic to paramagnetic (austenite region) when heated at high temperature and transforming into ferrite and coarse pearlite after furnace cooling. The SiC reacts with CIP and form iron silicides at the interfaces during solid state sintering process. After characterization, the experiments were conducted with sample1 and 2 on BEMRF tool and $\% \Delta R_a$ was calculated as shown in table 2.

The results show that the $\% \Delta R_a$ was found better by MR finishing surface of work-piece with sintered magnetic abrasives based MRPF (sample2) rather than unbonded magnetic abrasives based MRPF (sample1). The shearing action improvement of work-piece material is because of strong bonding between sintered magnetic abrasives and surrounding CIPs and/or sintered magnetic abrasives which impart more shear stress on the work-piece surface. After comparison, total 20 set of experiments were conducted randomly with the combinations of machining parameters given in table 4 and $\% \Delta R_a$ was calculated.

Table 4 Actual values and response $\% \Delta R_a$

Exp. No.	Actual Values					
	Current (I)	RPM (N)	Gap (G)	Initial R_a (μm)	Final R_a (μm)	$\% \Delta R_a$
1	3	400	1	0.2980	0.2192	26.44
2	5	400	1	0.2623	0.1574	39.99
3	3	600	1	0.2939	0.1771	39.74
4	5	600	1	0.2662	0.1728	35.08
5	3	400	2	0.2491	0.1914	23.16
6	5	400	2	0.2747	0.1870	31.93
7	3	600	2	0.2497	0.2222	11.01
8	5	600	2	0.2597	0.2188	15.75
9	2.3	500	1.5	0.2478	0.1953	21.18
10	5.7	500	1.5	0.2401	0.1647	31.40
11	4	332	1.5	0.2323	0.1881	19.03
12	4	668	1.5	0.2442	0.2073	15.11
13	4	500	0.66	0.2677	0.1327	50.43
14	4	500	2.3	0.2360	0.1863	21.06
15	4	500	1.5	0.2389	0.1677	29.80
16	4	500	1.5	0.2894	0.2070	28.47
17	4	500	1.5	0.2367	0.167	29.45
18	4	500	1.5	0.2586	0.1823	29.51
19	4	500	1.5	0.2389	0.1677	29.80
20	4	500	1.5	0.2725	0.2010	26.24

3.1 Response surface regression analysis

The response $\% \Delta R_a$ were input in the design expert software for further analysis and sequential model sum of squares were obtained for getting highest order polynomial. The quadratic terms with two factors interaction (2FI) and linear terms has highest significance because it has high F value and lowest p value suggesting its suitability. Lack of fit was calculated for each model. The analysis of variance (ANOVA) was applied to test the adequacy of model

and was used to investigate the machining parameters which significantly influence $\% \Delta R_a$.

Table 5 Analysis of variance (ANOVA) for $\% \Delta R_a$

Source	Sum of Square	DO F	Mean Square	F-Value	P-Value Prob>F	Remark
Model	1624.73	9	180.53	32.87	< 0.0001	significant
I	114.76	1	114.76	20.90	0.0010	
N	51.48	1	51.48	9.37	0.0120	
G	866.68	1	866.68	157.82	< 0.0001	
IN	61.83	1	61.83	11.26	0.0073	
IG	2.67	1	2.67	0.49	0.5017	
NG	168.54	1	168.54	30.69	0.0002	
I ²	4.33	1	4.33	0.79	0.3952	
N ²	209.19	1	209.19	38.09	0.0001	
G ²	112.54	1	112.54	20.49	0.0011	
Residual	54.92	10	5.49			
Lack of Fit	45.76	5	9.15	5	0.0510	Not significant
Pure Error	9.16	5	1.83			
Cor Total	1679.64	19				
R- squared = 0.9673			Adj R- squared = 0.9379			
Pred R- squared = 0.7619			Adeq precision = 22.968			

After dropping the insignificant terms such as IG and I² the model F value has been further improved to 44.79. It was found from ANOVA result that the working gap is the most significant machining parameter with significant factor of 167.97 which affect $\% \Delta R_a$. The contribution of magnetizing current and tool rotational speed is 22.24 and 9.98 respectively. To understand the MR finishing process in terms of $\% \Delta R_a$, the mathematical model was developed using multiple regression method and given by equation (1) in coded factors form, (2) in actual factors form.

The final equation in terms of coded factors is as follows-

$$\% \Delta R_a = 28.34 + 2.90A - 1.94B - 7.97C - 2.78AB - 4.59BC - 3.76B^2 + 2.85C^2$$

(1)

The final equation in terms of actual factors is as follows-

$$\% \Delta R_a = -142.35 + 16.78I - 0.605N - 4.22G - 0.028IN - 0.092NG - 3.76 \times 10^{-4} N^2 + 11.4G^2$$

(2)

3.2 Effect of working gap on percentage reduction in surface roughness

The working gap effect on $\% \Delta R_a$ at different magnetizing current and at tool rotational speed 500 rpm is shown in figure 3. From ANOVA table, it was observed that the working gap is most significant machining parameter with significant factor of 167.97. It was observed from fig.3 that $\% \Delta R_a$ decreases continuously as working gap increases. It is because the magnetic flux density as well as strength of finishing spot of MRPF decreases with increasing the working gap. Hence, sintered magnetic abrasives and/or surrounding crushed CIP were not able to hold the abrasives firmly during finishing and abrasive particles may not apply the required finishing force to shear off peaks from the work-piece surface.

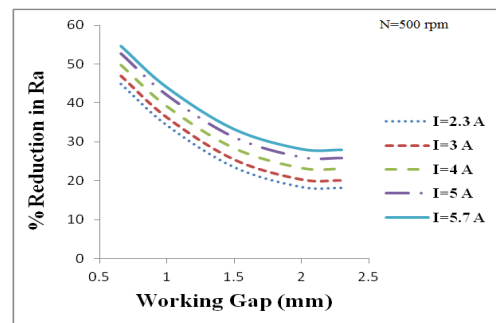


Figure 3 Effect of working gap on $\% \Delta R_a$

3.3 Effect of magnetizing current on percentage reduction in surface roughness

The magnetic current effect on $\% \Delta R_a$ at different working gap and at tool rotational speed 500 rpm is shown in figure 4. It was observed from ANOVA table that magnetizing current is more significant machining parameter with significant factor 22.24. It was observed from fig. 4 that $\% \Delta R_a$ increases with magnetizing current because magnetic flux density in finishing spot of MRPF increases with current and particles hold each other more firmly and results increasing finishing action.

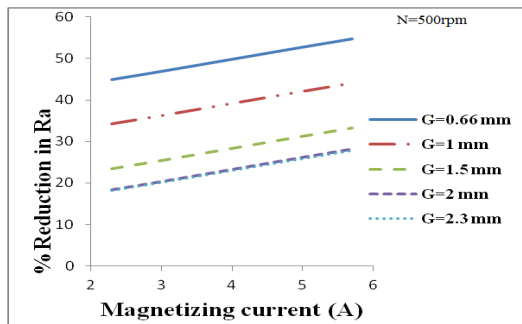


Figure 4 Effect of magnetic current on $\% \Delta R_a$

3.4 Effect of tool rotational speed on percentage reduction in surface roughness

The shear force acting on work-piece surface increases with rpm and hence $\% \Delta R_a$ increases then reaches an optimum value. When tool rotational speed is increased beyond optimum level, centrifugal force on particles increases then tool aging effect takes place which is responsible for decrease in $\% \Delta R_a$.

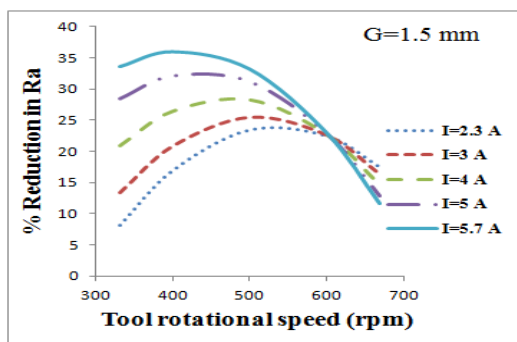


Figure 5 Effect of tool rotational speed on $\% \Delta R_a$

4. Conclusions

1. The sintered magnetic abrasives were developed at 1200°C using pressureless solid phase sintering method in control atmosphere of argon.
2. Morphology and characterization was carried out with scanning electron microscope and Vibration sample magnetometer.
3. The experiments were conducted on BEMRF tool with synthesized fluids (sample 1) and (sample 2) on mild steel work-piece surface and $\% \Delta R_a$ was calculated.
4. Experimental study shows that $\% \Delta R_a$ obtained by MR finishing the work-piece surface with sintered magnetic abrasives based MRPF was found better than $\% \Delta R_a$ obtained by MR finishing the work-piece surface with unbonded magnetic abrasives based MRPF at the same machining condition.

5. Experimental investigations and statistical analysis was carried out to study effect of current, tool rotational speed and working gap on $\% \Delta R_a$.

References

1. Kordonski, W.I. and Jacobs, S.D. (1996), Magnetorheological finishing, International Journal of Modern Physics B, Vol.10, pp 2857-2865.
2. Golini, D., Jacobs, S., Kordonski, W. and Dumas, P. (1997), Precision optics fabrication using Magnetorheological finishing, Advanced Materials for Optics and Precision Structures, CR67, pp 251-274.
3. Kordonski, W., Golini, D., Dumas, P. and Jacobs, S. (1998), Magnetorheological suspension-based finishing technology", proc. SPIE, Vol. 3326, pp 527-535.
4. Golini, D., Kordonski, W.I., Dumas, P. and Hogan, S. (1999), Magnetorheological finishing (MRF) in commercial precision optics manufacturing, proc. SPIE, Vol. 3782, pp 80-91.
5. Shorey, A.B., Jacobs, S.D., Kordonski, W.I. and Gans, R.F. (2001), Experiments and observations regarding mechanisms of glass removal in magnetorheological finishing, Applied Optics, Vol. 40/1, pp 20-33.
6. Kim, W-B, Lee, S.H., and Min, B.K. (2004), Surface finishing and evaluation of three-dimensional silicon microchannel using magnetorheological fluid, Journal of Manufacturing Science and Engineering, Vol.126, pp 772-778.
7. Seok, J., Kim, Y-J, Jang, K., Min, B-K and Lee, S.J., (2007), A study on the fabrication of curved surfaces using magnetorheological fluid finishing, International Journal of Machine Tools and Manufacture Vol. 47, pp 2077-2090.
8. Sadiq, A. and Shunmugam, M.S. (2009), Investigation into Magnetorheological abrasive honing (MRAH), International Journal of Machine tools and Manufacture, Vol.49, pp 554-560.
9. Jung, B., Jang, K., Min, B., Lee, S., and Seok, J. (2009), Magnetorheological finishing process for hard materials using sintered iron- CNT compound abrasives, International Journal of Machine tools and Manufacture, pp 1-11.
10. Singh, A.K., Jha, S. and Pandey, P.M. (2011), Design and development of nanofinishing process for 3D surfaces using ball end MR finishing tool, International Journal of Machine Tools & Manufacture, Vol. 51, pp 142-151.
11. Singh, A.K., Jha, S. and Pandey, P.M. (2012), Nanofinishing of a typical 3D ferromagnetic workpiece using ball end magnetorheological finishing process, International Journal of Machine Tools & Manufacture, Vol. 63, pp 21-31.

Performance Optimization of Four Stage Cascade Refrigeration Systems using Energy-Exergy Analysis in the R1234ze & R1234yf in High Temperature Circuit and Ecofriendly Refrigerants in Intermediate Circuits and Ethane in the Low Temperature Circuit for Food, Pharmaceutical, Chemical Industries

R.S. Mishra

Department of Mechanical Engineering, Delhi Technological University, New Delhi, India

Article Info

Article history:

Received 20 September 2014

Received in revised form

10 October 2014

Accepted 20 November 2014

Available online 15 December 2014

Keywords

Four stages Cascade Refrigeration,
Energy- Exergy Analysis,
First and Second Law Analysis,
Low Temperature Applications

Abstract

This paper mainly deals with the exergy analysis of Four stages cascade refrigeration systems used for low temperature applications using ecofriendly refrigerants. The effect of performance parameters (i.e. approaches, condenser temperature, and temperature variations in the evaporators) on the thermal performances in terms of second law efficiency of the system (exergetic efficiency) and exergy destruction ratio (EDR) and first law efficiency (i.e. overall coefficient of performance) have been optimized thermodynamically using of R1234yf and R1234ze in the high temperature circuits and mainly thirteen ecofriendly refrigerants in the intermediates circuits and ethane. It was observed that in the low temperature (between -80 °C to -88 °C) applications. It was observed that the best combination in terms of R1234ze-R134a-R410a-ethane gives better thermal performance than using R1234yf-R134a-R410a-ethane.

1. Introduction

Low-temperature refrigeration systems are typically required in the temperature range from -30°C to -100°C for applications in food, pharmaceutical, chemical, and other industries. At such low temperatures, single-stage compression systems with reciprocating compressors are generally not feasible due to high pressure ratios. A high pressure ratio implies high discharge and oil temperatures and low volumetric efficiencies and, hence, low COP values. Cascade refrigeration systems can be used to achieve low temperatures, where series of single-stage units are used that are thermally coupled through evaporator/condenser cascades. Each circuit has a different refrigerant suitable for that temperature. The high-temperature circuit uses high boiling point refrigerants such as R-1234ze, and R1234yf and Intermediate circuits thirteen ecofriendly refrigerants such as R-134a, R123, R125, HFC blends (R-507a), ammonia, propane, propylene, hydrocarbons, etc., whereas the low-temperature circuit uses low boiling refrigerants such as ethane, air, methane etc. In a vapour compression cascade refrigeration system where low pressure vapour in the evaporator stage is compressed and then recycled for condensation in the evaporator of a previous stage, the improvement of passing the low pressure refrigerant's vapour from an evaporator stage through a heat exchanger to heated vapour to ambient temperature, compressing said heated vapour, removing the compressor work by passing said compressed vapour through a evaporator, then cooling is by compressed vapour by passing it through heat exchanger in heat exchanger at a low pressure vapour, and condensing it in the evaporator of the next higher temperature cycle of the cascade system and

recycling the liquid to the low pressure evaporator stage. Most of the investigators even did not find out the effect of various approaches on the system performances on three and four stages cascade refrigeration systems and effect of ecofriendly refrigerants in the primary intermediate temperature circuit and secondary intermediate temperature circuit. This paper mainly deals with thermal performances in terms of exergy destruction ratio by thermodynamically analyzing the use of R1234yf and R1234ze in the high temperature circuits and mainly thirteen eco-friendly refrigerants in the intermediates circuits and Ethane in the low temperature applications.

2. Literature Review

Agnew et al.^[1] examined the performance of a three stage cascade refrigeration systems. Arora et al.^[2] analysis of a vapour compression refrigeration system with R502, R404A and R507. Bansal P.K.^[3] did thermodynamic analysis of carbon dioxide-ammonia (R744-R717) cascade refrigeration system. Bhattacharyya et al.^[4] studied the performance of a cascade refrigeration-heat pump system based on a model incorporating both internal and external irreversibility's. Cabell et al.^[5] Made a simplified steady-state modeling of a single stage vapors compression plant.. Calm^[6] reviewed the progression of refrigerants Ecir et al.^[7] used ten modeling techniques within data mining process for the prediction of thermo-physical properties of refrigerants. Jain et al.^[10] showed that carbon dioxide is a potential low temperature refrigerant. Kilicarslan^[12] experimental investigation and theoretical study of a different type of two-stage vapor compression cascade refrigeration system. Lee et al.^[13] studied thermodynamically a cascade refrigeration system that uses carbon dioxide and ammonia as refrigerants. Monte^[14] calculated thermodynamic properties of R407C Pearson^[16]

Corresponding Author,

E-mail address: professor_rsmishra@yahoo.co.in

All rights reserved: <http://www.ijari.org>

traced the development of the old carbon dioxide systems Samant [18] design and development of two stage cascade refrigeration system using CO₂ as LTC refrigerant and Propane as HTC refrigerant. Resat S. et al [17] exergy based thermo-economic optimization of subcooled and superheated vapour compression refrigeration cycle. Wang et al [18] potential benefits of compressor cooling. Syed M. [19] comprehensive investigation of numerical methods in simulating a steady-state vapor compression system. Xudong Wang, [20] Investigated potential benefits of compressor cooling Zubair [21] evaluated the performance of vapour Compression System and experimental investigation and theoretical study of a different type of two-stage vapor compression cascade refrigeration system. Experimental investigation and theoretical study of a different type of two-stage vapor compression cascade refrigeration system.

The above investigators did not studies the thermodynamic performances in terms of COP and exergetic efficiency and system exergy destruction ratio (EDR) for very low temperature application using ethane in the low temperature circuits and R1234ze and R1234yf in higher temperature circuit.

3. Description of four Stage Cascade Systems

The following assumptions have been taken for analyzing four stages cascade vapour compression system for low temperature applications. The cooling load is considered to be 35 “kW” Temperature of condenser is to be 70°C and temperature of ethane evaporator to be -88 °C, The temperature of high temperature evaporator to be 30 °C, temperature of secondary intermediate cascade evaporator is to be -50 °C, temperature of primary intermediate cascade evaporator is to be -10°C, temperature difference between cascade condenser and cascade evaporator is 10 °C in each stage.

4. Results and Discussions

Table-1(a) to 1(h) showing the first law efficiency of the four stage cascade refrigeration system using R1234ze and R1234yf in the high temperature circuit and R-410a in the secondary cascade intermediate temperature circuit which is a lower temperature circuit than primary intermediate cascade circuit and ethane in the lowest temperature circuit with different thirteen ecofriendly refrigerant in primary intermediate cascade circuit. It was observed that the use of R1234ze in higher temperature circuit gives better thermodynamic performance than using R1234yf in high temperature circuit. The best combination of four stage cascade is system is to be R1234ze-R123-R410a-ethane and R1234ze-R152a-R410a-ethane. The R123 refrigerants containing chlorine and R152, R600 1nd R290 are also flammable in nature can be considered by adopting safety measures. R717 is toxic in nature. Therefore R1234ze-R134a-R410a- Ethane is most suitable for practical application.

Table: 1(a). R1234ze in hot fluid circuit, R134a in secondary intermediate circuit, R404a in low temperature circuit and Different refrigerant in primary intermediate circuit

Primary Circuit Ecofriendly Refrigerant	COP _{overall}	EDR	Exergetic Efficiency
R290	0.4669	5.705	0.1491
R404A	0.4387	6.136	0.1401
R410A	0.4601	5.804	0.1470
R134A	0.4701	5.659	0.1502
R152A	0.4822	5.493	0.1540
R600	0.4799	5.523	0.1533
R600A	0.4698	5.663	0.1501
R407C	0.4385	6.14	0.1401
R507A	0.4428	6.070	0.1415
R113	0.4863	5.437	0.1553
R123	0.4873	5.424	0.1557
R717	0.4838	5.472	0.1545
R125	0.4231	6.399	0.1352

Table: 1(b). R1234ze in hot fluid circuit, R134a in secondary intermediate circuit, R404a in low temperature circuit and Different refrigerant in primary intermediate circuit

Primary Circuit Ecofriendly Refrigerant	COP _{Hot fluid circuit}	COP _{intermediate temperature circuit-I}	COP _{intermediate temperature circuit-II}	COP _{low temperature fluid circuit}
R290	4.375	2.412	3.829	2.307
R404A	4.375	2.098	3.829	2.307
R410A	4.375	2.332	3.829	2.307
R134A	4.375	2.541	3.829	2.307
R152A	4.375	2.604	3.829	2.307
R600	4.375	2.574	3.829	2.307
R600A	4.375	2.448	3.829	2.307
R407C	4.375	2.096	3.829	2.307
R507A	4.375	2.141	3.829	2.307
R113	4.375	2.659	3.829	2.307
R123	4.375	2.672	3.829	2.307
R717	4.375	2.625	3.829	2.307
R125	4.375	1.944	3.829	2.307

Table: 1(c). R1234yf in hot fluid circuit, R134a in secondary intermediate circuit, R404a in low temperature circuit and different refrigerant in primary intermediate circuit

Primary Circuit Ecofriendly Refrigerant	COP _{overall}	EDR	Exergetic Efficiency
R290	0.4543	5.891	0.1451
R404A	0.4271	6.330	0.1364
R410A	0.4478	5.991	0.1430
R134A	0.4574	5.844	0.1461
R152A	0.4691	5.674	0.1498
R600	0.4669	5.705	0.1491
R600A	0.4572	5.848	0.1460

R407C	0.4269	6.333	0.1364
R507A	0.4311	6.262	0.1377
R113	0.4731	5.618	0.1511
R123	0.474	5.604	0.1514
R717	0.4706	5.653	0.1503
R125	0.4121	6.598	0.1316

Table: 1(d). R1234yf in hot fluid circuit, R134a in secondary intermediate circuit, R404a in low temperature circuit and Different refrigerant in primary intermediate circuit

Primary Circuit Ecofriendly Refrigerant	COP _{Hot} fluid circuit	COP _{intermediate} temperature circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
R290	3.972	2.412	3.829	2.307
R404A	3.972	2.098	3.829	2.307
R410A	3.972	2.332	3.829	2.307
R134A	3.972	2.451	3.829	2.307
R152A	3.972	2.604	3.829	2.307
R600	3.972	2.574	3.829	2.307
R600A	3.972	2.448	3.829	2.307
R407C	3.972	2.096	3.829	2.307
R507A	3.972	2.141	3.829	2.307
R113	3.972	2.659	3.829	2.307
R123	3.972	2.672	3.829	2.307
R717	3.972	2.625	3.829	2.307
R125	3.972	1.944	3.829	2.307

Table: 1(e). R1234ze in hot fluid circuit, R134a in primary intermediate circuit, R404a in low temperature circuit, different refrigerant in primary intermediate circuit

Secondary intermediate circuit Ecofriendly Refrigerants	COP _{overall}	EDR	Exergetic Efficiency
R290	0.4701	5.659	0.1502
R404A	0.4622	5.773	0.1476
R410A	0.4701	5.659	0.1502
R134A	0.4703	5.657	0.1502
R152A	0.4720	5.633	0.1508
R600	0.4725	5.625	0.1509
R600A	0.4703	5.656	0.1502
R407C	0.4395	6.123	0.1404
R507A	0.4659	5.720	0.1488
R113	0.4742	0.5602	0.1515
R123	0.4737	0.5609	0.1513
R717	0.4650	5.733	0.1485
R125	0.4629	5.763	0.1479

Table: 1(f). R1234ze in hot fluid circuit, R134a in primary intermediate circuit, R404a in low temperature circuit, different refrigerant in primary intermediate circuit

Secondary Intermediate Circuit Ecofriendly Refrigerants	COP _H hot fluid circuit	COP _{intermediate} temperature circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
R290	4.375	2.451	3.828	2.307
R404A	4.375	2.451	3.625	2.307
R410A	4.375	2.451	3.829	2.307
R134A	4.375	2.451	3.822	2.307
R152A	4.375	2.451	3.878	2.307
R600	4.375	2.451	3.894	2.307
R600A	4.375	2.451	3.834	2.307
R407C	4.375	2.451	3.116	2.307
R507A	4.375	2.451	3.717	2.307
R113	4.375	2.451	3.941	2.307
R123	4.375	2.451	3.926	2.307
R717	4.375	2.451	3.695	2.307
R125	4.375	2.451	3.642	2.307

Table: 1(g). R1234yf in hot fluid circuit R134a in primary intermediate circuit, R404a in low temperature circuit and Different refrigerant in primary intermediate circuit

Secondary Intermediate Ecofriendly Refrigerant	COP _{overall}	EDR	Exergetic Efficiency
R290	0.4574	5.844	0.1461
R404A	0.4498	5.959	0.1437
R410A	0.4574	5.844	0.1461
R134A	0.4576	5.842	0.1462
R152A	0.4592	5.817	0.1467
R600	0.4598	5.809	0.1469
R600A	0.4577	5.841	0.1462
R407C	0.4279	6.316	0.1367
R507A	0.4533	5.906	0.1448
R113	0.4614	5.785	0.1474
R123	0.4609	5.792	0.1472
R717	0.4525	5.919	0.1445
R125	0.4505	5.949	0.1439

Table: 1(h). R1234yf in hot fluid circuit R134a in primary intermediate circuit, R404a in low temperature circuit and Different refrigerant in primary intermediate circuit

Secondary intermediate Ecofriendly Refrigerant	COP _H hot fluid circuit	COP _{intermediate} temperature circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
R290	3.973	2.451	3.828	2.307
R404A	3.973	2.451	3.625	2.307
R410A	3.973	2.451	3.829	2.307
R134A	3.973	2.451	3.832	2.307

R152A	3.973	2.451	3.878	2.307
R600	3.973	2.451	3.894	2.307
R600A	3.973	2.451	3.834	2.307
R407C	3.973	2.451	3.116	2.307
R507A	3.973	2.451	3.717	2.307
R113	3.973	2.451	3.941	2.307
R123	3.973	2.451	3.926	2.307
R717	3.973	2.451	3.695	2.307
R125	3.973	2.451	3.642	2.307

Table: 2(a) to 2(l) Sare showing the effect of approaches (differences in temperature between cascade condenser and cascade evaporators in the various intermediate temperature circuits and low temperature circuit) on the first law and second law performances on the four stage cascade refrigeration systems. It was observed that as approach is decreasing the system first and second law efficiency is increasing and exergy destruction ratio is decreases

Table: 2(a). R1234ze in High Temperature circuit Effect of Approach-1

Effect of Approach-1	COP _{overall}	EDR	Exergetic Efficiency
10	0.4701	5.659	0.1502
5	0.4953	5.321	0.1582
0	0.5207	5.013	0.1663

Table: 2(b). R1234ze in High temperature circuit Effect of Approach-1

Effect of Approach-1	COP _{Hot} fluid circuit	COP _{intermediate} temperature circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
10	4.375	2.451	3.829	2.307
5	4.375	2.783		2.307
0	4.375	3.175		2.307

Table: 2(c). R1234ze in High temperature circuit Effect of Approach-2

Effect of Approach-2	COP _{overall}	EDR	Exergetic Efficiency
10	0.4701	5.659	0.1502
5	0.4920	5.364	0.1571
0	0.5207	5.083	0.1644

Table: 2(d). R1234ze in High temperature circuit Effect of Approach-2

Effect of Approach-2	COP _{Hot} fluid circuit	COP _{intermediate} temperature circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
10	4.375	2.451	3.829	2.307
5	4.375	2.451	4.482	2.307
0	4.375	2.451	5.349	2.307

Table: 2(e). R1234ze in High temperature circuit Effect of Approach-3

Effect of Approach-3	COP _{overall}	EDR	Exergetic Efficiency
10	0.4701	5.659	0.1502
5	0.4920	5.303	0.1587
0	0.5247	4.966	0.1676

Table: 2(f). R1234ze in High temperature circuit Effect of Approach-3

Effect of Approach-3	COP _{Hot} fluid circuit	COP _{intermediate} temperature circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
10	4.375	2.451	3.829	2.307
5	4.375	2.451	3.829	2.623
0	4.375	2.451	3.829	3.013

Table: 2(g). R1234yf in High temperature circuit Effect of Approach-1

Effect of Approach-1	COP _{overall}	EDR	Exergetic Efficiency
10	0.4574	5.844	0.1461
5	0.4817	5.499	0.1539
0	0.5062	5.185	0.1617

Table: 2(h). R1234yf in High temperature circuit Effect of Approach-1

Effect of Approach-1	COP _{Hot} fluid circuit	COP _{intermediate} temperature circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
10	3.973	2.451	3.829	2.307
5	3.973	2.783	3.829	2.307
0	3.973	3.175	3.829	2.307

Table: 2(i). R1234yf in High temperature circuit Effect of Approach-2

Effect of Approach-2	COP _{overall}	EDR	Exergetic Efficiency
10	0.4574	5.844	0.1461
5	0.4817	5.499	0.1539
0	0.5062	5.185	0.1617

Table: 2(j). R1234yf in High temperature circuit Effect of Approach-2

Effect of Approach-2	COP _{Hot} fluid circuit	COP _{intermediate} temperature circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
10	3.973	2.451	3.829	2.307
5	3.973	2.783	3.829	2.307
0	3.973	3.175	3.829	2.307

Table: 2(k). R1234yf in High temperature circuit Effect of Approach-3

Effect of Approach-3	COP _{overall}	EDR	Exergetic Efficiency
10	0.4574	5.844	0.1461
5	0.4831	5.481	0.1543
0	0.5103	5.138	0.1629

Table: 2(f). R1234yf in High temperature circuit Effect of Approach-3

Effect of Approach-3	COP _H of fluid circuit	COP _{inte} mediate temperature e circuit-II	COP _{inter} mediate temperature circuit-II	COP _{low} temperature fluid circuit
10	3.973	2.451	3.829	2.307
5	3.973	2.451	3.829	2.623
0	3.973	2.451	3.829	3.013

Table: 3(a). Performance variation with condenser temperature of four stage cascade vapours compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tcond (°C)	COP _{overall}	EDR	Exergetic Efficiency
80	0.4204	6.447	0.1343
75	0.4459	6.02	0.1424
70	0.4701	5.659	0.1502
65	0.4933	5.347	0.1576
60	0.5157	5.07	0.1647
55	0.5378	4.822	0.1718
50	0.5595	4.595	0.1787
45	0.5812	4.387	0.1856
40	0.6029	4.193	0.1926

Table: 3(b). Performance variation with condenser temperature of four stage cascade vapours compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tcond (°C)	COP _{Hot} fluid circuit	COP _{inte} mediate temperature re circuit-II	COP _{inter} mediate temperature circuit-II	COP _{low} temperature fluid circuit
80	3.054	2.451	3.829	2.307
75	3.651	2.451	3.829	2.307
70	4.375	2.451	3.829	2.307
65	5.282	2.451	3.829	2.307
60	6.467	2.451	3.829	2.307
55	8.10	2.451	3.829	2.307
50	10.52	2.451	3.829	2.307
45	14.52	2.451	3.829	2.307
40	22.47	2.451	3.829	2.307

Table: 3(c). Performance variation with condenser temperature of four stage cascade vapours compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Tcond (°C)	COP _{overall}	System EDR	Exergetic Efficiency
80	0.3945	6.935	0.1260
75	0.4279	6.316	0.1367
70	0.4574	5.844	0.1461
65	0.4844	5.463	0.1547
60	0.5096	5.143	0.1628
55	0.5337	4.866	0.1705
50	0.5569	4.622	0.1779
45	0.5796	4.401	0.1851
40	0.6021	4.20	0.1923

Table: 3(d). Performance variation with condenser temperature of four stage cascade vapours compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Tcond (°C)	COP _{Hot} fluid circuit	COP _{inte} mediate temperature e circuit-II	COP _{inte} mediate temperature e circuit-II	COP _{low} temperature fluid circuit
80	2.573	2.451	3.829	2.307
75	3.216	2.451	3.829	2.307
70	3.973	2.451	3.829	2.307
65	4.904	2.451	3.829	2.307
60	6.106	2.451	3.829	2.307
55	7.748	2.451	3.829	2.307
50	10.17	2.451	3.829	2.307
45	14.14	2.451	3.829	2.307
0.1923	22.02	2.451	3.829	2.307

From Tables-3: it was observed that as condenser temperature decreases overall COP of system (First law efficiency of system) increases and System EDR is decreases and also second law efficiency (exergetic efficiency) increases. While First law efficiency of hot fluid circuit is increases. There will not be effect on other circuit first law efficiencies. The exergetic efficiency and Overall COP of system using R1234ze is higher than using R1234yf in the high temperature circuit while system EDR increases. As decreasing High temperature evaporator temperature, the first law efficiency (i.e. overall cop of system) and second law efficiency (i.e. exergetic efficiency of whole system) is is increasing up to maximum value at the evaporator temperature of 25°C and then decreases rapidly.

As decreasing High temperature evaporator temperature the system EDR first decreasing up to decreasing evaporator temperature and then further constant and then increasing and optimum becomes at evaporator temperature at 25°C in both cases using R1234yf and R1234ze in the high temperature circuit. However Cop of hot fluid circuit is decreases and COP of primary intermediate fluid circuit is increases.

Table: 4(a). Performance variation with evaporator temperature in the high temperature circuit of Four stage cascade vapour compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tevaporator High temp circuit (°C)	COP _{Overall}	System EDR	Exergetic Efficiency
50	0.4331	6.229	0.1383
45	0.4476	5.995	0.1430
40	0.4584	5.829	0.1464
35	0.4659	5.720	0.1488
30	0.4701	5.659	0.1502
25	0.4713	5.642	0.1505
20	0.4696	5.666	0.150
15	0.4652	5.729	0.1486
10	0.4583	5.831	0.1464
5	0.4489	5.974	0.1434
0	0.4375	6.158	0.1397

Table: 4(b). Performance variation with evaporator temperature in the high temperature circuit of four stage

cascade vapour compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tevaporator High temp circuit (°C)	COP _{Hot fluid circuit}	COP _{intermediate temperature circuit-II}	COP _{intermediate temperature circuit-II}	COP _{low temperature fluid circuit}
50	10.53	1.489	3.829	2.307
45	8.06	1.61	3.829	2.307
40	6.417	1.914	3.829	2.307
35	5.248	2.165	3.829	2.307
30	4.375	2.451	3.829	2.307
25	3.699	2.783	3.829	2.307
20	3.161	3.175	3.829	2.307
15	2.723	3.646	3.829	2.307
10	2.361	4.229	3.829	2.307
5	2.056	4.97	3.829	2.307
0	1.791	5.951	3.829	2.307

Table: 4(c). Performance variation with evaporator temperature in the high temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Tevaporator High temp circuit (°C)	COP _{Overall}	EDR	Exergetic Efficiency
50	0.4291	6.295	0.1371
45	0.4419	6.085	0.1412
40	0.4507	5.946	0.1440
35	0.4559	5.868	0.1456
30	0.4574	5.844	0.1461
25	0.4557	5.870	0.1456
20	0.4507	5.946	0.1440
15	0.4428	6.070	0.1414
10	0.4321	6.245	0.1380
5	0.4189	6.474	0.1338
0	0.4033	6.762	0.1288

Table: 4(d). Performance variation with evaporator temperature in the high temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Tevaporator High temp circuit (°C)	COP _{Hot fluid circuit}	COP _{intermediate temperature circuit-II}	COP _{intermediate temperature circuit-II}	COP _{low temperature fluid circuit}
50	9.804	1.489	3.829	2.307
45	7.462	1.690	3.829	2.307
40	5.906	1.914	3.829	2.307
35	4.799	2.165	3.829	2.307
30	3.973	2.451	3.829	2.307
25	3.333	2.783	3.829	2.307
20	2.825	3.175	3.829	2.307
15	2.413	3.646	3.829	2.307
10	2.072	4.229	3.829	2.307
5	1.786	4.970	3.829	2.307
0	1.544	5.951	3.829	2.307

As evaporator temperature decreases, the first law efficiency (overall COP) and second law efficiency (exergetic efficiency) of the system is increases and maximum efficiency is obtained at evaporator temperature of -10 °C and also cop of primary intermediate temperature circuit is decreases and secondary intermediate temperature circuit COP is increases.

Table: 5(a). Performance variation with cascade evaporator temperature in the primary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tevaporator (°C)	COP _{overall}	EDR	Exergetic Efficiency
20	0.4436	6.057	0.1417
15	0.4546	5.886	0.1452
10	0.4633	5.757	0.1480
5	0.4696	5.666	0.150
0	0.4738	5.607	0.1513
-5*	0.4758	5.579	0.1520
-10*	0.4759	5.579	0.1520
-15	0.4739	5.609	0.1514
-20	0.4701	5.659	0.1502
-25	0.4646	5.739	0.1484
-30	0.4573	5.485	0.1461
-35	0.4485	5.980	0.1433
-40	0.4383	6.143	0.140

Table: 5(b). Performance variation with cascade evaporator temperature in the primary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tevaporator (°C)	COP _{Hot fluid circuit}	COP _{intermediate temperature circuit-II}	COP _{intermediate temperature circuit-II}	COP _{low temperature fluid circuit}
20	4.375	10.33	1.460	2.307
15	4.375	7.951	1.629	2.307
10	4.375	6.367	1.818	2.307
5	4.375	5.239	2.032	2.307
0	4.375	4.396	2.278	2.307
-5*	4.375	3.744	2.565	2.307
-10*	4.375	3.224	2.905	2.307
-15	4.375	2.801	3.317	2.307
-20	4.375	2.451	3.829	2.307
-25	4.375	2.157	4.482	2.307
-30	4.375	1.907	5.349	2.307
-35	4.375	1.691	6.559	2.307
-40	4.375	1.505	8.368	2.307

Table: 5(c). Performance variation with cascade evaporator temperature in the primary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Tevaporator (°C)	COP _{overall}	EDR	Exergetic Efficiency
20	0.4319	6.249	0.1379
15	0.4425	6.075	0.1413

10	0.4509	5.944	0.1440
5	0.4570	5.851	0.1460
0	0.4610	5.791	0.1473
-5*	0.4630	5.762	0.1479
-10*	0.4630	5.762	0.1479
-15	0.4611	5.789	0.1473
-20	0.4574	5.844	0.1461
-25	0.4521	5.925	0.1444
-30	0.4451	6.034	0.1422
-35	0.4366	6.170	0.1395
-40	0.4267	6.336	0.1363

Table: 5(d). Performance variation with cascade evaporator temperature in the primary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Teva porat or (°C)	COP _{Hot} fluid circuit	COP _{inter} mediate temperatur e circuit-II	COP _{inte} rmediate temperatur e circuit-II	COP _{low} temperature fluid circuit
20	4.375	10.33	1.460	2. .307
15	4.375	7.951	1.629	2. .307
10	4.375	6.367	1.818	2. .307
5	4.375	5.239	2.032	2. .307
0	4.375	4.396	2.278	2. .307
-5*	4.375	3.744	2.565	2. .307
-10*	4.375	3.224	2.905	2. .307
-15	4.375	2.801	3.317	2. .307
-20	4.375	2.451	3.829	2. .307
-25	4.375	2.157	4.482	2. .307
-30	4.375	1.907	5.349	2. .307
-35	4.375	1.691	6.559	2. .307
-40	4.375	1.505	8.368	2. .307

Table: 6(a). Performance variation with cascade evaporator temperature in the secondary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tevapor ator (°C)	COP _{overall}	EDR	Exergetic Efficiency
-70	0.4631	5.76	0.1479
-65	0.4674	5.697	0.1493
-60	0.4701	5.66	0.1502
-55***	0.4710	5.647	0.1504
-50	0.4701	5.659	0.1502
-45	0.4675	5.696	0.1493
-40	0.4631	5.760	0.1479
-35	0.4569	5.851	0.1460
-30	0.4490	5.973	0.1434

Table: 6(b). Performance variation with cascade evaporator temperature in the secondary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tevap orator (°C)	COP _{Hot} fluid circuit	COP _{inte} rmediate temperatur e circuit-II	COP _{inter} mediate temperatur e circuit-II	COP _{low} temperature fluid circuit
-70	4.375	2.451	2.116	4.161
-65	4.375	2.451	2.426	3.508
-60	4.375	2.451	2.798	3.013
-55***	4.375	2.451	3.255	2.603
-50	4.375	2.451	3.829	2.307
-45	4.375	2.451	4.568	2.046
-40	4.375	2.451	5.556	1.825
-35	4.375	2.451	6.942	1.636
-30	4.375	2.451	9.022	1.42

As increasing evaporator temperature overall cop is increases while System EDR first decreases while second law efficiency is increases. The maximum efficiency was obtained at -55 oC and then decreases as increasing evaporator temperature. Secondary intermediate circuit COP is increases while low temperature circuit COP is decreases. Table-6(b): Performance variation with cascade evaporator temperature in the secondary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Table: 6(c). Performance variation with cascade evaporator temperature in the secondary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Tevaporator (°C)	COP _{overall}	EDR	Exergetic Efficiency
-70	0.4507	5.946	0.1440
-65	0.4549	5.883	0.1453
-60	0.4574	5.845	0.1461
-55***	0.4583	5.831	0.1464
-50	0.4574	5.844	0.1461
-45	0.4949	5.882	0.1453
-40	0.4507	5.946	0.1440
-35	0.4447	6.039	0.1421
-30	0.4370	6.163	0.1396

Table: 6(d). Performance variation with cascade evaporator temperature in the secondary intermediate temperature circuit temperature of four stage cascade vapour compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Teva porat or (°C)	COP _{Hot} fluid circuit	COP _{inter} mediate temperatur e circuit-II	COP _{inter} mediate temperatur e circuit-II	COP _{low} temperature fluid circuit
-70	0.3973	2.451	2.116	4.161
-65	0.3973	2.451	2.426	3.508
-60	0.3973	2.451	2.798	3.013
-55*	0.3973	2.451	3.255	2.603
-50	0.3973	2.451	3.829	2.307
-45	0.3973	2.451	4.568	2.046
-40	0.3973	2.451	5.556	1.825
-35	0.3973	2.451	6.942	1.636
-30	0.3973	2.451	9.022	1.42

Table: 7(a). Performance variation with low temperature evaporator temperature in the of four stage cascade vapour compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tevaporator (°C)	COP _{overall}	EDR	Exergetic Efficiency
-90	0.4701	5.659	0.1502
-85	0.5024	5.14	0.1629
-80	0.5358	4.676	0.1762
-75	0.5701	4.259	0.1902
-70	0.6056	3.884	0.2048
-65	0.6421	3.546	0.220
-60	0.6796	3.239	0.2359
-55	0.7181	2.962	0.2524

Table: 7(b). Performance variation with low temperature evaporator temperature in the of four stage cascade vapour compression refrigeration system using R1234ze in High temperature circuit and R404a in the low temperature circuit

Tevaporator (°C)	COP _{Hot} fluid circuit	COP _{intermediate} temperature e circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
-90	4.375	2.451	3.829	2.307
-85	4.375	2.451	3.829	2.697
-80	4.375	2.451	3.829	3.185
-75	4.375	2.451	3.829	3.812
-70	4.375	2.451	3.829	4.645
-65	4.375	2.451	3.829	5.804
-60	4.375	2.451	3.829	7.525
-55	4.375	2.451	3.829	10.35

Table: 7(c). Performance variation with low temperature evaporator temperature in the of our stage cascade vapour

compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Tevaporator (°C)	COP _{overall}	EDR	Exergetic Efficiency
-90	0.4574	5.844	0.1461
-85	0.4881	5.314	0.1584
-80	0.5207	4.840	0.1712
-75	0.5538	4.415	0.1847
-70	0.5878	4.032	0.1987
-65	0.6228	3.686	0.2134
-60	0.6587	3.373	0.2287
-55	0.6956	3.09	0.2445

Table: 7(d). Performance variation with low temperature evaporator temperature in the of four stage cascade vapour compression refrigeration system using R1234yf in High temperature circuit and R404a in the low temperature circuit

Tevaporator (°C)	COP _{Hot} fluid circuit	COP _{intermediate} temperature e circuit-II	COP _{intermediate} temperature circuit-II	COP _{low} temperature fluid circuit
-90	3.972	2.451	3.829	2.307
-85	3.972	2.451	3.829	2.697
-80	3.972	2.451	3.829	3.185
-75	3.972	2.451	3.829	3.812
-70	3.972	2.451	3.829	4.645
-65	3.972	2.451	3.829	5.804
-60	3.972	2.451	3.829	7.525
-55	3.972	2.451	3.829	10.35

Table-7(a) to 7(d) are showing the effect of low temperature evaporator on the system first and second law performances and system exergy destruction ratio and various circuit first law performances . as low temperature evaporator temperature is decreasing , the first law and second law efficiencies are increasing and exergy destruction ratio is decreasing and high temperature circuit first law performances in increasing.

References

- [1] B. Agnew, et.a., A finite time analysis of a cascade refrigeration system using alternative refrigerants, Applied Thermal Engineering, 24, 2004, 2557–2565
- [2] Akhilesh Arora et.al, Theoretical analysis of a vapour compression refrigeration system with R502, R404A and R507A, International Journal of Refrigeration, 31, 2008, 1–8
- [3] P. K. Bansal, Thermodynamic analysis of an R744–R717 cascade refrigeration system, international journal of refrigeration 31, 2008, 45-54.
- [4] Souvik Bhattacharyya, Exergy maximization of cascade refrigeration cycles and its numerical verification for a trans critical CO₂-C₃H₈ system, International Journal of Refrigeration, 30, 2007, 624-632
- [5] R. Cabello, Simplified steady-state modelling of a single stage vapour compression plant. Model development and validation, Applied Thermal Engineering, 25, 2005, 1740–1752
- [6] James M. Calm, The next generation of refrigerants – Historical review, considerations, and outlook, International Journal of Refrigeration, 31, 2008, 1123-1133.
- [7] Ecir Ug'ur Küçüksille, Data mining techniques for thermophysical properties of refrigerants, Energy Conversion and Management, 2008
- [8] J. Guilpart, Comparision of the performance of different ice slurry types depending on the application temperature, International Journal of Refrigeration, 29, 2006, 781-788
- [9] M. Herbert et.al, On the Performance of R22 and A HfC/HC Refrigerant Mixture over A Range of Charge Quantity In A Window Air conditioner, Acre-conf, 2009
- [10] P. K. Jain, Cascade Systems: Past, Present, and Future, 2007 ASHRAE Trans. 113(1), 245-252(DA-07-027)
- [11] Jonathan Winkler, Comprehensive investigation of numerical methods in simulating a steady-state vapor compression system, International Journal of Refrigeration, 31, 2008, 930-942
- [12] A. Kilicarslan, An experimental investigation of a different type vapor compression cascade refrigeration system, Applied Thermal Engineering, 24, 2004 2611–2626

- [13] F. De. Monte, Calculation of thermodynamic properties of R407C and R410A by the Martin–Hou equation of state — part II: technical interpretation, *International Journal of Refrigeration*, 25, 2002, 314–329
- [14] F. Monte, de Calculation of thermodynamic properties of R407C and R410A by the Martin–Hou equation of state - part I: theoretical development, *International Journal of Refrigeration*, 25, 2002, 306–313.
- [15] Mukhopadhyay, CO₂–C₃H₈ cascade refrigeration–heat pump system: Heat exchanger inventory optimization and its numerical verification, *international journal of refrigeration*, 31, 2008, 1207–1213
- [16] D. Peter, Refrigeration of the Future for Air Conditioning, *Air conditioning and Refrigeration Journal*, 8, 2008, 20-27
- [17] Reşat Selbaş, Önder Kızılkın, Arzu Şencan, Thermoeconomic optimization of subcooled and superheated vapor compression refrigeration cycle, *Energy*, 31, 2006, 2108–2128.
- [18] Samant Maj, Design and Development of two stage cascade refrigeration system, *Mechanical Engineering Department, IIT Delhi*, 2006
- [19] Tzong-Shinge.al, Thermodynamic analysis of optimal condensing temperature of cascade-condenser in CO₂/NH₃ cascade refrigeration systems, *International Journal of Refrigeration*, 29, 2006, 1100-1108
- [20] Xudong Wang, Yunho Hwang, Reinhard Radermacher, Investigation of potential benefits of compressor cooling, *Applied Thermal Engineering*, 28, 2008, 1791–1797
- [21] S. M. Zubair, Performance evaluation of Vapour Compression System, *International Journal of Refrigeration*, 22, 1999, 235-243
- [22] C. P. Arora, *Refrigeration and Airconditioning*, Second edition, New Delhi: Tata McGraw Hill, 2002

Secure Code Assignment to Alphabets Using Modified Ant Colony Optimization along with Compression

Hitesh Hasija

Information Technology
Delhi Technological University
Delhi, India
hitoo.hasija@gmail.com

Rahul Katarya

Information Technology
Delhi Technological University
Delhi, India
rahuldtu@gmail.com

Abstract— Assigning ASCII values to the plaintext alphabets during transmission of documents is very common, and as we know that the ASCII value of ‘a’ is less than that of ‘z’. But, what would happen if the document contains more ‘z’ as compared to ‘a’. In that case, achieving security through encryption - decryption and compression is very difficult. This paper proposes a technique of assigning codes to plain text alphabets, based on the frequency of occurrences of those alphabets by using ant colony optimization with roulette wheel selection algorithm, so that security in code generation could be achieved. Obtained codes are then applied over a particular encryption algorithm such that the cipher text generated should contain those alphabets only whose code values are near about similar to that of plain text alphabets, in order to achieve compression simultaneously. Getting all this within specified time constraint makes this paper more impressive.

Index Terms— ASCII, Ciphertext, ant colony optimization, relative occurrence matrix, encryption, decryption, compression, travelling salesman problem, roulette wheel selection algorithm

I. INTRODUCTION

Providing data security and compression during transmission of data over a network are two different tasks. Plain text alphabets during transmission are usually coded with ASCII values as guided in [1], and getting those alphabets back from ASCII values is a very easy task for intruders. Thus, assigning ASCII values is no more secure. Similarly, what happens if those alphabets whose ASCII values are large appearing more frequently in the document than those whose values are small? In that case, large numbers of bits are required for transmission of alphabets. Hence, it is clear that reassigning of the codes is necessary to achieve compression.

This paper derives a technique of assigning codes to alphabets based on their occurrences in the document.

This process of assigning codes is not similar to the one which could be achieved directly by assigning low value codes to frequently occurring alphabets and vice-a-versa. Rather, it is obtained by applying modified ant colony optimization, just like as an analogy to the travelling salesman problem. Ant colony parameters are decided by the user. Roulette wheel selection algorithm provides security to the obtained codes, so that an intruder could not be able to detect it. Compression is achieved by specifying an encryption - decryption algorithm as referenced by [2], [3], [4] and [5]. It assures by generating cipher text whose code values are near about equivalent to the code values of plain text entered by the user. Therefore, alphabets appearing more frequently would be assigned codes with low number of bits and the number of bits for generating cipher text codes will remain near about same to that of plain text.

This paper first provides background details in section (II), then modified Ant Colony Optimization technique is introduced in section (III), section (IV) describes about assigning codes to alphabets with encryption - decryption and it is the basic theme of the paper as well as provides an example to understand this paper, section (V) deals with the flowchart of complete paper and demonstration of the code, Section (VI) having conclusion and future work, Section (VII) consists of references.

II. BACKGROUND

Cryptography is the technique of generating the secret messages as suggested by [6] and [7]. The goals of cryptography are as follows:--

1. Authentication
2. Privacy
3. Integrity
4. Non repudiation
5. Access control.

Encryption is a subset of cryptography. It is the process of transforming the plaintext into cipher text. Decryption is the process of transforming the cipher text into plain text. The main purpose of cryptographic algorithm is to make it

as difficult as possible, so that decryption of cipher text should not be possible without the knowledge of key. As referenced by [8], It is categorized into two types:-

1. Symmetric key encryption
2. Asymmetric key encryption

In symmetric key encryption, the same key is used by both the sender and receiver before the process of encryption and decryption. So, the secret key plays an important role and its length (in bits) is responsible for its strength.

In asymmetric key encryption, there exists a pair of public key and private key. The Public key is used for encryption and the private key is used for decryption [9]. Private Key is known only to the sender, thus it is also known as secret key. But public keys can be stored in public databases for anyone to see [10].

This paper is based on symmetric Key encryption.

III. ANT COLONY OPTIMIZATION (ACO)

ACO is a meta-heuristic approach, as referenced by [11] meaning that it's a general framework that can be used to create a specific algorithm to solve a specific graph path problem. Ant colony optimization learns from the past learning experiences to determine the best trail for future iterations as referenced by [12].

A. Basic ACO Parameters

Ant colony optimization (ACO) is an artificial intelligence technique based on the pheromone-laying behavior of ants. It can be used to find solutions to exceedingly complex problems that seek the optimal path through a graph. In this case, the goal is to find the shortest path that visits each of alphabets in the form of nodes exactly once. The program uses 26 ants, where each ant represents a potential solution. ACO requires the specification of several parameters. The ants are initialized to random trails; after initialization, the best ant has a shortest trail length. The key idea of ACO is the use of simulated pheromones, which attract ants to better trails through the graph. The main processing loop alternates between updating the ant trails based on the current pheromone values and updating the pheromones based on the new ant trails. After the maximum number of times (1,000) through the main processing loop, the program found the best trail and its corresponding length.

B. ACO with graph based approach

For a real-world graph problem, you'd probably read a graph-adjacency and distance-between-node data from a text file into some sort of data structure. Here we simulate a graph by creating an array of arrays where the row indexes i and j represents the cost of moving from-

alphabet and the to-alphabet respectively. Notice that all alphabets (nodes) are connected somehow, distances are either symmetric or asymmetric and the distance from an alphabet to itself or the relative occurrence of an alphabet to itself is zero. The program uses Fisher-Yates shuffle algorithm to randomize the order of the alphabets in the trail. Pheromones are chemicals ants placed on their trails; they attract other ants. More ants will travel on the shorter trail to an alphabet and deposit more pheromones, than on longer trails. The pheromones slowly evaporate over time. The key to the ACO algorithm is the process that updates each ant and trail by constructing a new—we hope better—trail based on the pheromone and distance information.

C. Formulas of Ant Colony

$$p_{xy}^k = \frac{(\eta_{xy}^\beta)(\tau_{xy}^\alpha)}{\sum_{y \in allowed} (\eta_{xy}^\beta)(\tau_{xy}^\alpha)} \quad \text{-----}(1)$$

Where;

τ_{xy} is the amount of pheromone which is deposited for transition from alphabet x to y ,

α is a parameter used to control the influence of τ_{xy} ,

η_{xy} is the heuristic function for transition of state ' xy ', $(1/d_{xy})$, where d is the distance between alphabets ' xy ' (in the form of relative occurrence matrix) and

$\beta \geq 1$ is a parameter for controlling the influence of η_{xy} .

When all the ants have completed a solution, the trails are updated by

$$\tau_{xy} \leftarrow (1 - \rho)\tau_{xy} + \sum_k (\Delta \tau_{xy}^k) \quad \text{----} (2)$$

Where, τ_{xy} is again the amount of pheromone deposited for a state transition xy ,

ρ is the pheromone evaporation coefficient and

$(\Delta \tau_{xy}^k)$ is the amount of pheromone deposited by k_{th} ant, and it is calculated by the formula:-

$$(\Delta \tau_{xy}^k) = \begin{cases} Q/L_k & \text{if ant } k \text{ uses curve } xy \text{ in its tour} \\ 0 & \text{otherwise} \end{cases} \quad \text{-----} (3)$$

Where, L_k is the cost of the k_{th} ant's tour and Q is a constant.

Equation (1) is used to determine the probability of selecting a path based on the values of heuristic function as guided by the matrix and equation (2) is used to determine the updated values of pheromone with the help of equation (3).

D. Modified ACO Implementation

In this paper, ACO algorithm is incorporated with roulette wheel selection algorithm. In this approach an augmented array is used which holds cumulative probabilities. The size of the augmented array is one greater than the array of probabilities, and cell (0) is seeded with 0.0. Each cell in augmented array is the cumulative sum of the probabilities. After the augmented array has been constructed, a random number p between 0.0 and 1.0 is generated. Suppose $p = 0.538$ as shown. If that p value falls between the values at index 2 and index 3 in the augmented array, which means that, node 2, is selected as the next node. Each pheromone value is decreased, simulating evaporation, and increased, simulating the deposit of pheromones by ants on the trail. The decrease effect is produced by multiplying the current pheromone value by a factor less than 1.0 that depends on global parameter ρ . The larger ρ is, the greater the decrease in pheromone value. The increased effect is produced by adding a proportion of the current ant's total trail length, where the proportion is determined by global parameter Q . Larger values of Q increase the amount of pheromone added.

IV. ASSIGNING CODES TO ALPHABETS

The strength of a key is always an issue to think upon it. But, the question is how to decide the length of a key, because it should not be so long that it will increase the time complexity of our program, it should also not be so small that it could be broken by anyone. Hence, the key generation mechanism plays a vital role in secure data transmission.

This paper discovers a novel approach for the generation of key based on the document to be encrypted. It considers the frequency of occurrence of an alphabet into the document for the generation of code. For example, the frequency of occurrence of $a=10$, $b=20$, $c=30$ and $d=40$ in a particular document. It signifies that, "a" is occurring 10 times in the document, "b" is occurring 20 times and so on respectively. Based on these occurrences of an alphabet in the document, the relative occurrence matrix is designed by using the formula:-

$$RFO(i, j) = |FO(i) - FO(j)| \quad \text{----- (4)}$$

Where, $RFO(i, j)$ means the relative frequency of occurrence of the alphabets i and j ,

And,

$FO(i)$ means the frequency of occurrence of the alphabet i ,

The corresponding relative occurrence matrix for this example, according to equation (4) is as follows:-

Table (1) Relative Occurrence matrix

	A	B	C	D
A	0	10	20	30
B	10	0	10	20
C	20	10	0	10
D	30	20	10	0

Table (1) matrix is like a graph problem in which the alphabets would be acting as a node and the values inside the matrix specifies the values on the edges. The main aim of designing metrics is to determine the order of encrypting the alphabets because this paper focuses on assigning less number of bits to frequently occurring alphabets. So, we have to first determine, the descending order of alphabets with respect to their occurrences. We can determine that order, simply by calculating their occurrences in the document. But, the codes generated in this way would not provide any security to the algorithm, and for that we are using Ant colony optimization.

Table (1) matrix acts as an analogy to the travelling salesman problem. In the travelling salesman problem, our main aim is to determine the order in which the nodes or cities should be visited, such that all the cities should be visited only once and with an optimum trail length. Similarly, here also aim is to determine the order in which the alphabets should be assigned codes such that, frequently occurring alphabets should be assigned less number of bits and vice-a-versa. In this problem, just like travelling salesman problem, the edges contain values of the relative occurrences of alphabets. The order of alphabets obtained after applying ant colony optimization over this graph (specified in the form of a 2-D matrix), tells us the decreasing order of occurrences of the alphabets. But, this order is not equivalent to the decreasing order arrangement of alphabets based on their occurrences into the document (which could be obtained

directly). It depends on the roulette wheel selection algorithm used within it. The order in which the nodes should be visited, not only depends on the values of the edges but also on the four constants. They are:-

1. Alpha (Pheromone Influence factor)
2. Beta (Local node Influence factor)
3. Rho (Pheromone Evaporation coefficient factor)
4. Q (Pheromone deposit factor)

As the values of these constants are modified, the order in which the alphabets should be assigned codes also get modified. Thus, we can generate different codes based on the values of these four constants. It is clear that, code generation depends on the plaintext and also on the edges containing values, as specified in the relative occurrence matrix. The order contains first those alphabets with large number of occurrences, so that they should be assigned codes accordingly. As the first appearing alphabet with code 0, second appearing alphabet with code 1, third appearing alphabet with code 2, and so on. In this way, security as efficient code generation along with proper compression by assigning low code values to frequently occurring alphabets is achieved. In the example illustrated above, as it is clearly visible that the output of the ant colony optimization would be d, c, b, a. It means that, alphabet d is occurring more and hence it should be assigned code value 0, c should be assigned code value 1, and so on.

Let us consider for a while that, the output of ant colony optimization is like z, y, x, w, v and so on up to a. Thus, “z” should be assigned code 0, “y” should be assigned code 1 and so on “a” should be assigned code 25. Thus, the encryption algorithm would go like this:-

Z=0, Y=1, X=2, W=3, V=4, U=5, T=6, S=7, R=8, Q=9, P=10, O=11, N=12, M=13, L=14, K=15, J=16, I=17, H=18, G=19, F=20, E=21, D=22, C=23, B=24, A=25.

1. Plain text Input
P Q R S
2. Code Input
10 9 8 7
minimum content=7
3. Mod Content
3 2 1 0
4. Key
T U V W
5. Code Key
6 5 4 3
min=3
6. Mod Key
3 2 1 0

7. Right Shift with Mod Key
0 3 2 1
8. Adding min value to this shifted version
3 6 5 4
9. Final encrypt
W T U V
10. Adding min value array to the Mod Content array
6 8 6 4
11. Ciphertext
T R T V

First of all codes are assigned to the plaintext input entered by the user. Secondly, the minimum value is determined from that array. Thirdly, mod value is determined by taking mod of all those values with minimum content value. Fourthly, codes are assigned to the key. Fifthly, the minimum value is determined from that array. Sixthly, mod values are determined based on it. In the seventh step, the contents of the Mod Key array are shifted to right and min value is added to all those values. In the tenth step, again min value is added to all the values of Mod Content array. Finally, the cipher text is obtained from those obtainable values, by assigning proper alphabets to those codes values.

The corresponding decryption algorithm is like this:-

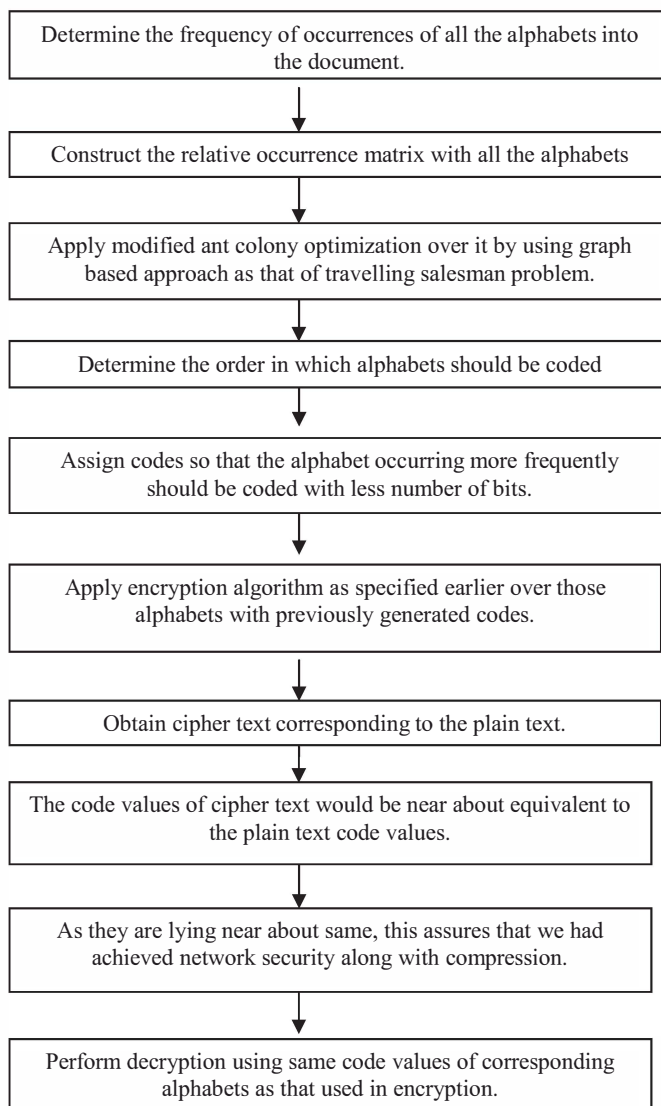
1. Ciphertext
T R T V
2. Code for cipher text
6 8 6 4
3. Code for Final Encrypt
3 6 5 4
4. Difference of upper two arrays
3 2 1 0
5. Adding minimum content value to the upper array
10 9 8 7
6. Plain text obtained
P Q R S

In the first step of decryption phase, cipher text is obtained. Secondly, codes are assigned to them. Thirdly, code for final encrypt as specified in the encryption process is obtained. Fourthly, differences of the two array values obtained in step second and third are determined and in the fifth step value of minimum content is added to that array. Finally, the codes are converted into corresponding alphabets and ultimately plaintext is obtained. From this example we can clearly understand the beauty of this algorithm that it converts our plain text “PQRS” into cipher text as “TRTV”. As it is justified earlier itself that, codes assigned to “PQRS and TRTV” are near about the same. So, the plaintext appearing more frequently would be converted into the cipher text whose codes are near about similar, means that letters appearing

more frequently would be assigned codes with low values or they would be assigned codes consisting of the lower number of bits as guided by [13],[14] and [15]. This is the biggest advantage of using this particular encryption technique over all the other approaches.

Another advantage of this paper is that it does not use simple ant colony optimization, rather a variation of it, in which roulette wheel selection algorithm is used for determining the next node to be traversed because it provides randomization.

V. FLOWCHART



```

file:///F:/programs for visual studio/races4/races4/bin/Debug/races4.EXE

Applying Modified Ant colony optimization based on
Frequency of occurrences of Alphabets

Number of alphabets in document = 26

Number of ants = 26
Maximum time limit= 1000

Alpha (pheromone influence factor) = 3
Beta (local node influence factor) = 2
Rho (pheromone evaporation coefficient factor) = 0.01
Q (pheromone deposit factor) = 2.00

Getting values for Relative occurrence matrix

Initializing ants to random trails

0: [ 25 21 20 15 . . . 9 11 22 12 ] len = 2130.0
1: [ 25 18 15 4 . . . 5 9 12 1 ] len = 1980.0
2: [ 25 13 6 23 . . . 17 5 8 21 ] len = 1900.0
3: [ 25 23 6 13 . . . 18 12 8 7 ] len = 1940.0

Best initial trail length: 1900.0

Initializing pheromones on trails

Entering UpdateAnts - UpdatePheromones loop

New best length of 1850.0 found at time 2
New best length of 1820.0 found at time 10
New best length of 1500.0 found at time 12
New best length of 1440.0 found at time 201
New best length of 1410.0 found at time 394
New best length of 1370.0 found at time 570
New best length of 1290.0 found at time 844

Time complete

Best trail found:
25 4 2 3 0 1 16 10 6 14 15 7 5 13 22 24 19 20 18 17 21
23 9 11 8 12

The order in which the nodes should be assigned codes is as follows:-
z e c d a b q k g o p h f n w y t u s r v
x j l i m

Length of best trail found: 1290.0

End of Ant Colony Optimization over relative occurrence matrix
  
```

FIGURE (1)

In the figure (1), numbers of ants taken are 26 and it could be varied. The time constraint is also there in order to determine the optimum path within the limit of 1000. Ants are initialized to the alphabet with highest frequency of occurrence as over here is 'z'. Then best trail is found and the order in which codes should be assigned is determined. As the order first contains value 25 means that z should be assigned first code value 0, then it contains value 4 means that e should be assigned code value 1, similarly c should be assigned 2, d with 3 and so on. This order is not exactly equivalent to the decreasing orderly arrangement of frequency of occurrence of alphabets, thus it can't be easily judged by the intruder, but then too it resembles highly.

```

file:///F:/programs for visual studio/races5/races5/bin/Debug/races5.EXE

The sequence of code values obtained for alphabets starting from 'A' and ending
to 'Z' are as follows
4      5      2      3      1      12      8      11      24      22
7      23      25      13      9      10      6      19      18      16
17     20     14     21     15     0

-----Encryption starts from here-----

Enter the number of characters you want to enter
4
Enter the characters to be converted into cipher text
p
o
g
k
Enter the four characters of your key
q
b
a
k

Input array is as follows
p      o      g      k
Code values for input array is as follows
10     9      8      7
The values for modcontent array are as follows
3      2      1      0
Input key is as follows
q      b      a      k
Code values for input key is as follows
6      5      4      7
The values for modcontent key array are as follows
2      1      0      3
Encrypt key after shifting is as follows
3      2      1      0
After adding minimum value to the encrypt key we get
7      6      5      4
Final encrypted key is
k      q      b      a
Ciphertext for code based values are
10     8      6      4
Final encrypted cipher text is as follows
p      g      q      a
-----Decryption starts from here-----
Difference array obtained from upper two arrays is as follows
3      2      1      0
Code based plain array obtained is as follows
10     9      8      7
Final Plain Text we get is
p      o      g      k

```

FIGURE (2)

In figure (2), after assigning codes to the alphabets, encryption is done over it with plain text taken over here as 'pogk' and key used as 'qbak' with generated cipher text to be transmitted as 'pgqa'. As the code of 'pogk' and 'pgqa' is almost equivalent to each other i.e. codes with less number of bits. Thus, the basic theme of this paper is justified, i.e. compression is achieved along with secure code generation.

VI. CONCLUSION AND FUTURE WORK

This paper incorporates a method for efficient code generation based on the frequency of occurrences of alphabets into the document. It uses relative occurrence matrix and ant colony optimization with suitable values of

constants for that. The key gets changed by changing either the values of constants or by changing the frequency of occurrences of alphabets. Frequently occurring alphabets should be assigned low code values or low number of bits. Then, encryption algorithm converts plain text into cipher text such that the cipher text consists of those alphabets whose code values are near about equivalent to that of plain text in order to satisfy both the requirements of compression as well as information security. Thus, it benefits the user by providing a strong key whose length is about 26 i.e. codes assigned to 26 alphabets. The strength of this key could be judged from this fact that, it is obtained after solving a graph problem with 26 nodes and with different values of edges. This code is not only dependent on the plaintext, but also on other parameters. Thus, it is very difficult to judge if intruder do not know about plaintext and ant colony optimization parameters both. One of the biggest advantages of this code generation technique is modified ant colony optimization, i.e. ant colony incorporated with roulette wheel selection algorithm, which provides a proper variation in encryption key generation.

In the future work, limitations of the proposed algorithm are proposed by using variable size of key length for encryption and decryption with the fixed input size.

VII. REFERENCES

- [1] Udepal Singh, Upasna Garg, "An Ascii value based text data encryption system";2013 International Journal of Scientific and Research Publications, Volume 3, Issue 11, ISSN 2250-3153
- [2] Akanksha Mathur;"An ASCII value based data encryption algorithm and its comparison with other symmetric data encryption algorithms";2012 International Journal on Computer Science and Engineering (IJCSSE)
- [3] Ajit Singh, Upasna Jauhari, "Data Security by preprocessing the text with Secret Hiding";2012 Advanced Computing:An International Journal, Volume 3, No. 3, 2012
- [4] Joyshree Nath, Asoke Nath, "Advanced Steganography Algorithm using Encrypted secret message", 2011 International Journal of Advanced Computer Science and Applications, Volume 2, No. 3
- [5] Majdi Al-qdah, Lin Yi Hui, , "Simple Encryption Decryption Application, International Journal of Computer Science and Security",Volume 1, Issue 1
- [6] Verma, Sharad Kumar, Ojha D.B., "An application of data encryption technique using random number generator";2012 International Journal of research studies in computing, Volume 1 Number 1
- [7] Gary C. Kessler "An Overview of Cryptography" , May 1998
- [8] Dr. Anwar Pasha Abdul Gafoor Deshmukh, Dr. Riyazuddin Qureshi, "Transparent data encryption-solution for Security

- of database contents”,2011 (IJACSA)International Journal of advanced computer and applications, Volume 2, No. 3
- [9] Gurjeevan Singh, Ashwani Kumar Singla, K.S. Sandha, “Throughput Analysis of Various Encryption Algorithms”, International Journal of Computer Science and Technology, Vol. 2, Issue 3, Septemver 2011.
- [10] Diaa Salama Abd Elminaam, Hatem Mohamed Abdual Kader, and Mohiy Mohamed Hadhoud, “Evaluating the Performance of Symmetric Encryption Algorithms”, International Journal of Network Security, Vol.10, No.3, PP.216–222, May 2010.
- [11] Marco Dorigo, Mauro Birattari, and Thomas Stützle;”Ant colony optimization”;University ‘e Libre de Bruxelles, Belgium;2006 IEEE
- [12] Feng-Hsu Wang;”On extracting recommendation knowledge for personalized web-based learning based on ant colony optimization with segmented-goal and meta-control strategies”; 2011 Elsevier Ltd
- [13] John A. Copeland;”Data Compression technique for PC Communication”;IEEE Journal on selected areas in communication, Volume 7 ,No. 2,February 1989
- [14] Wei-Wei Lough and M.P.Gough;”A fast adaptive Huffman coding algorithm”;IEEE Transactions on Communications, Volume 41,No. 4,April 1993
- [15] Ningde Xie, Guiqiang Dong, Tong Zhang,”Using lossless data compression in data storage systems: not for saving space”;IEEE Transactions on computers, Volume 60, No. 3,March 2011

Simultaneous Reconfiguration, Optimal Placement of DSTATCOM, and Photovoltaic Array in a Distribution System Based on Fuzzy-ACO Approach

Hajar Bagheri Tolabi, *Member, IEEE*, Mohd Hasan Ali, *Senior Member, IEEE*, and M. Rizwan, *Member, IEEE*

Abstract—In this paper, a combination of a fuzzy multiobjective approach and ant colony optimization (ACO) as a metaheuristic algorithm is used to solve the simultaneous reconfiguration and optimal allocation (size and location) of photovoltaic (PV) arrays as a distributed generation (DG) and distribution static compensator (DSTATCOM) as a distribution flexible ac transmission system (DFACT) device in a distribution system. The purpose of this research includes loss reduction, voltage profile (VP) improvement, and increase in the feeder load balancing (LB). The proposed method is validated using the IEEE 33-bus test system and a Tai-Power 11.4-kV distribution system as a real distribution network. The results proved that simultaneous reconfiguration and optimal allocation of PV array and DSTATCOM unit leads to significantly reduced losses, improved VP, and increased LB. Obtained results have been compared with the base value and found that simultaneous placement of PV and DSTATCOM along with reconfiguration is more beneficial than separate single-objective optimization. Also, the proposed fuzzy-ACO approach is more accurate as compared to ACO and other intelligent techniques like fuzzy-genetic algorithm (GA) and fuzzy-particle swarm optimization (PSO).

Index Terms—Ant colony optimization (ACO), distribution static compensator (DSTATCOM), distribution system, fuzzy sets, photovoltaic (PV), reconfiguration.

I. INTRODUCTION

DISTRIBUTION systems have two types of switches, i.e., 1) tie and 2) sectionalizing [1]. By changing the switches status between feeders, the structure of the distribution network will change and it is known as reconfiguration. The main objective of reconfiguration is to reduce losses, increase stability and reliability, improve voltage profile (VP), and relieve overload in the distribution network. The concept of reconfiguration of distribution network was first proposed by Merlin and Back [2] in 1975.

Manuscript received April 21, 2014; revised July 26, 2014, September 19, 2014, and October 15, 2014; accepted October 17, 2014. Date of publication November 20, 2014; date of current version December 12, 2014. Paper no. TSTE-00165-2014.

H. Bagheri Tolabi is with the Department of Electrical Engineering, Faculty of Engineering, Khorramabad Branch Islamic Azad University, Khorramabad 68135/536, Iran (e-mail: hajar.bagheri@hotmail.com).

M. H. Ali is with the Department of Electrical and Computer Engineering, University of Memphis, TN 38152 USA (e-mail: mhali@memphis.edu).

M. Rizwan is with the Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India (e-mail: rizwaniit@yahoo.co.in).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSTE.2014.2364230

A number of methods of reconfiguration in distribution networks have been proposed by researchers and also available in literature. Recently, Kavousi-Fard and Niknam [3] have proposed the reconfiguration problem with respect to the reliability using a self-adaptive modified optimization algorithm.

In restructured power systems, the use of distributed generation energy resources including photovoltaic (PV), fuel cells, small wind turbines, etc., is playing an important role because of various advantages. The advantage of distributed generation energy resources includes reduction in power loss, improvement in VP, and increase in the reliability of the network. To achieve the benefits of DG units, the selection of optimal locations and capacity is becoming the major problem. Various methods have been proposed by researchers in order to find the optimal placement and capacity of DG units. These methods are often based on artificial intelligence and heuristic algorithms. Kollu *et al.* [4] proposed a Harmony Search (HS) algorithm-based novel method to allocate DG units optimally in distribution system for power loss reduction and VP improvement.

Distribution flexible ac transmission system (DFACT) devices are used in distribution systems with different applications and controlling methods for improving the power quality indices. DSTATCOM, unified power flow controller (UPFC), and dynamic voltage restorer (DVR) are widely used DFACT devices. To find the optimal location and capacity of DFACT devices has a considerable impact in distribution systems. Some researchers have proposed various methods to find the optimal location and size of DFACT unit. Farhoodnea *et al.* [5] presented a firefly algorithm-based novel method to optimally place the DSTATCOM in distribution system.

In addition, several methods have been proposed for the simultaneous reconfiguration and optimal allocation of DG or DFACT unit in distribution system. However, simultaneous reconfiguration with both DG and DFACT is rarely available in the literature.

In this paper, the ant colony optimization (ACO) approach has been used for simultaneous multiobjective reconfiguration and optimal allocation of PV and DSTATCOM in a distribution network. The main objective of the work includes loss reduction, VP improvement, and equalizing the feeder load balancing (LB). To avoid the convergence problem, the input and output data are normalized in the same range based on fuzzy sets.

II. PROBLEM FORMULATION

The problem is formulated in the following steps.

A. Power Flow Equations

Power flows in a distribution system are computed by the following set of equations [6]:

$$\begin{aligned}
 P_{k+1} &= P_k - P_{\text{loss},k} - P_{Lk+1} \\
 &= P_k - \frac{R_k}{|V_k|^2} \left\{ P_k^2 + (Q_k + Y_k |V_k|^2)^2 \right\} - P_{Lk+1} \quad (1) \\
 Q_{k+1} &= Q_k - Q_{\text{loss},k} - Q_{Lk+1} \\
 &= Q_k - \frac{X_k}{|V_k|^2} \left\{ P_k^2 + (Q_k + Y_{k1} |V_k|^2)^2 \right\} - Y_{k1} |V_k|^2 \\
 &\quad - Y_{k2} |V_{k+1}|^2 - Q_{Lk+1} \\
 |V_{k+1}|^2 &= |V_k|^2 + \frac{R_k^2 + X_k^2}{|V_k|^2} (P_k^2 + Q_k^2) - 2(R_k P_k + X_k Q_k) \\
 &= |V_k|^2 + \frac{R_k^2 + X_k^2}{|V_k|^2} \left(P_k^2 + (Q_k^2 + Y_k |V_k|^2)^2 \right) \\
 &\quad - 2 \left(R_k P_k + X_k (Q_k + Y_k |V_k|^2) \right) \quad (2)
 \end{aligned}$$

where

- P_k real power flowing out of bus k ;
- Q_k reactive power flowing out of bus k ;
- $P_{\text{loss},k}$ real power loss at bus k ;
- $Q_{\text{loss},k}$ reactive power loss at bus k ;
- P_{Lk+1} real load power at bus at bus $k+1$;
- Q_{Lk+1} reactive load power at bus at bus $k+1$;
- R_k resistance of the line section between buses k and $k+1$;
- X_k reactance of the line section between buses k and $k+1$;
- Y_k shunt admittance at any bus k ;
- V_k voltage amplitude at bus k .

B. PV Array Equations

A PV array is a series-parallel combination of PV panels.

A PV system for residential, commercial, or industrial energy supply normally contains PV array, dc-dc converter, inverter, and a tracking system. An example of the electricity generated by PV system is shown in Fig. 1.

The active and reactive powers of PV unit can be calculated using the following equations [7]:

$$\begin{aligned}
 P_{\text{pv}} &= \left[\frac{V_k^2}{R_k} P_{\text{pv,loss}} - (P_k^2 + Q_k^2) \right. \\
 &\quad \left. - (Q_{\text{pv}}^2 - 2P_k P_{\text{pv}} - 2Q_k Q_{\text{pv}}) \left(\frac{G}{L} \right) \right]^{1/2} \quad (3)
 \end{aligned}$$

$$\begin{aligned}
 Q_{\text{pv}} &= \left[\frac{V_k^2}{R_k} P_{\text{pv,loss}} - (P_k^2 + Q_k^2) \right. \\
 &\quad \left. - (P_{\text{pv}}^2 - 2P_k P_{\text{pv}} - 2Q_k Q_{\text{pv}}) \left(\frac{G}{L} \right) \right]^{1/2} \quad (4)
 \end{aligned}$$

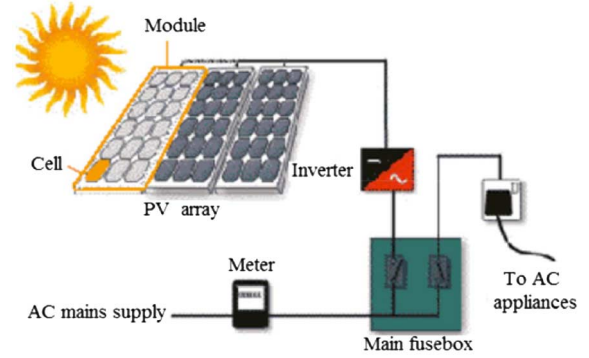


Fig. 1. Sample PV array.

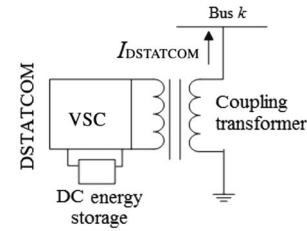


Fig. 2. DSTATCOM connected to bus k .

where

- P_{pv} real power supplied by PV;
- Q_{pv} reactive power supplied by PV;
- G distance from source to the PV location in km;
- L total length of the feeder from source to bus k in km.

C. DSTATCOM Equations

The DSTATCOM is one of the DFACT devices that basically consists of a coupling transformer with a leakage reactance, a GTO/IGBT voltage source converter (VSC), and a dc capacitor. Fig. 2 shows a bus in a distribution system equipped with a DSTATCOM.

The active and reactive powers of DSTATCOM can be calculated using the following equations [8]:

$$P_{\text{DSTATCOM}} = (V_k V_M / X_L) \sin \delta \quad (5)$$

$$Q_{\text{DSTATCOM}} = (V_k^2 / X_L) - (V_k V_M / X_L) \cos \delta \quad (6)$$

where V_k is the bus voltage, V_M is the DSTATCOM voltage, X_L is the line reactance, and δ is the phase angle displacement between V_k and V_M .

D. Fuzzy Multiobjective Function

The multiobjective function is a constrained optimization problem to find an optimal configuration for the distribution system, PV, and DSTATCOM allocation. This function consists of three goals: 1) reducing the loss; 2) increasing the LB; and 3) improving the VP.

The constraints of the problem can be presented as

- 1) $V_{k\min} \leq V'_k \leq V_{k\max}$
- 2) $|I'_{k,k+1}| \leq |I_{k,k+1\max}|$

- 3) $\sum_{k=1}^{n_s} P_{pv_k} + P_{DSTATCOM_k} \leq \sum_{k=1}^{n_s} (P_k + P_{loss,k})$
- 4) $|\det(A^1)| = 1$ (This constraint checks the radial structure of the network and feeding all available nodes. The details are given in the Appendix.)

where

V'_k	voltage at bus k after configuration;
$V_{k\max}$	maximum voltage at bus k ;
$V_{k\min}$	minimum voltage at bus k ;
$I'_{k,k+1}$	current in line section between buses k and $k+1$ after reconfiguration;
$I_{k,k+1\max}$	maximum current limit of line section between buses k and $k+1$;
n_s	total number of sections in the system;
A^1	reduced bus incidence matrix.

The first-term of the objective function reflects real power loss that is defined using

$$P_{loss} = \sum_{j=1}^{n_f} \sum_{k=1}^{n_s} R_k |I_k|^2 \quad (7)$$

where I_k is the current passing through line k and n_f is the total number of feeders.

The second term of the objective function reflects the improvement in the VP, which is shown by VP index in

$$VP = \sum_{j=1}^{n_f} \sum_{k \in lb} |V_k - V_{ref,k}| \quad (8)$$

where lb is the collection of the load buses and $V_{ref,k}$ is the nominal voltage at load bus k .

The third term of the objective function is considered for the LB index of the lines in the feeders, which is shown in

$$LB = \sum_{j=1}^{n_f} \sum_{k=1}^{n_s} \left(\frac{I_k}{\frac{1}{n_f} \sum_{k=1}^{n_s} I_k} \right)^2 \quad (9)$$

When there are different objectives to be satisfied simultaneously, a comparison is required to get the best solution. Since different terms of the multiobjective function are in various ranges, all the values are normalized in the same range to avoid the convergence problem [9]. Each variable has a membership function (μ) that determines the rank and effectiveness of its variable as follows:

$$\mu_{fij}(X) = \begin{cases} 1, & f_{ij}(X) \leq f_i^{\min} \\ \frac{f_i^{\max} - f_{ij}(X)}{f_i^{\max} - f_i^{\min}}, & f_i^{\min} < f_{ij}(X) < f_i^{\max} \\ 0, & f_i^{\max} \leq f_{ij}(X) \end{cases} \quad (10)$$

where $f_{ij} = \frac{T_{ij}}{T_{i0}}$, that is T_{ij} is the obtained value for the i th term of the objective function ($i = 1, 2, 3$) for the j th solution, T_{i0} is the base value for the i th term of the objective function before optimization, and $f_i^{\min}$ and $f_i^{\max}$ are the best and worst

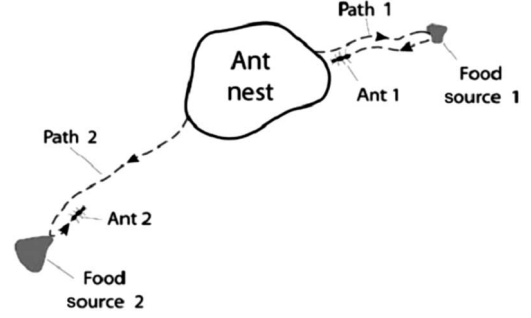


Fig. 3. Diagram of ants foraging for food.

solutions that observed in the single-objective optimization area for the i th term in the objective function, respectively.

In this paper, for the first term of objective function (p_{loss}), $f_1^{\min}$ and $f_1^{\max}$ are considered as 0.5 and 1, respectively, for the second term (VP), $f_2^{\min}$ and $f_2^{\max}$ are considered as 0.05 and 0.1, respectively, and for the third term (LB), $f_3^{\min}$ and $f_3^{\max}$ are considered as 0.1 and 0.5, respectively.

The three fuzzified objective functions mentioned above are integrated into a fuzzy multiobjective function through appropriate weighting factors. Equation (11) has described the fuzzy objective function ($F(X)$) for the j th solution as follows:

$$F_j(X) = \sum_{i=1}^3 w_i \cdot \mu_{fij}(X) \quad (11)$$

where w_i is the weighting factor for the i th term of the objective function. The weighting factors are considered $w_1 = w_2 = w_3 = 0.33$, in which the three objectives are assumed to have equal importance.

III. ACO

The ACO is a swarm intelligence-based technique, proposed by Dorigo *et al.* for the solution of combinatorial problems [10]. The ACO algorithm is originally inspired by the biological behavior of the ants and specifically their way of communication. This inspiration comes from the ability of real ants to find the short paths in their movement from and to their nests when searching for food source [11]. However, ants do not communicate with each other in a direct way but they exchange information through what is known as pheromones. A simple diagram of ant foraging behavior is shown in Fig. 3.

In this figure, two ants leave their nest in different directions at the same time for searching the food. As they move about, they deposit a pheromone trail that evaporates slowly and is detectable by other ants. If no pheromone exists initially outside the nest, the paths of the two ants are generally random [12], [13]. The evaporation of pheromone is also an important part of the foraging process that helps deemphasize older trails leading to exhausted food sources.

The ACO algorithms attempt to exploit the efficiency of ant foraging behavior by creating an abstract environment of possible paths, and simulating ants traveling along these paths. Each individual chooses a different path and comes up with different solutions in the search space region. When an ant passes through a path, it deposits the pheromone. This approach

helps ACO algorithms avoid getting trapped in local optima [11], [14].

A. ACO Procedure

The following procedure has been applied for the given problem [12]:

- 1) appropriate problem representation;
- 2) heuristic desirability (η) of edges;
- 3) construction of feasible solutions;
- 4) pheromone updating rule;
- 5) probabilistic transition rule.

A suitable heuristic desirability of traversing could be any subset evaluation function. The heuristic desirability of traversal and edge pheromone levels is combined to form the so-called probabilistic transition rule, denoting the probability of an ant at feature i choosing to travel to feature j at time t :

$$P_{i,j}^k(t) = \begin{cases} \frac{[\tau_{i,j}]^\alpha [\eta_{i,j}]^\beta}{\sum_{i \in T} [\tau_{i,j}]^\alpha [\eta_{i,j}]^\beta}, & \text{if } i, j \in T \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where k is the k th ant, $\alpha > 0$ and $\beta > 0$ are two parameters (the choice of α and β is determined experimentally), $\tau_{i,j}$ is the pheromone value for the ij th path of the ACO algorithm, $\eta_{i,j}$ is heuristic desirability for the ij th path of the ACO algorithm, and T is the total number of paths currently not visited by an ant. As suggested in (12), the transition probability is determined by the pheromone and heuristics of the trail, which corresponds to the individual cost of the ij th path. This path is said to be accepted when higher probability value is gained as a result of high quantity of pheromone or heuristic desirability [10]. After all ants have completed their tours, the pheromone level is updated by

$$\tau_{i,j}(t+1) = (1 - \rho) \cdot \tau_{i,j}(t) + \sum_{k=1}^m \Delta\tau_{i,j}^k(t) \quad (13)$$

where $\rho \in (0, 1)$ is the pheromone trail evaporation rate and m is the number of ants. The parameter ρ is used to avoid unlimited accumulation of the pheromone trails and enables the algorithm to forget previously done bad decisions. $\Delta\tau_{i,j}^k(t)$ is the amount of pheromone ant k deposits on the paths; it is defined as

$$\Delta\tau_{i,j}^k(t) = \begin{cases} 1/L^k(t), & \text{if path } ij \text{ is used by ant } k \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

where $L^k(t)$ is the length of the k th ant's tour. According to (14), the shorter the ant's tour is, the more pheromone is received by paths belonging to the tour [10].

The flowchart of the ACO procedure is described in the following steps:

- 1) reading the input data;
- 2) initializing the system;
- 3) creating the graphs for each ant, with its respective nodes and edges;

- 4) updating the list of feasible operation and probability values so that ants schedule their next operation until they reach food node; and
- 5) analysis of the best solution, updating pheromones, and checking the stopping criterion [11], [14], [15].

IV. PROPOSED FUZZY-ACO METHOD

In this section, the proposed fuzzy-ACO approach has been used for reconfiguration and optimal placement of DSTATCOM and PV array in distribution system.

An optimal feeder topology can be represented as

$$S^1 = [\text{tie switches}^1, \text{PV}^1, \text{DSTATVOM}^1].$$

In the S^1 vector, PV^1 is a matrix with two columns as follows:

$$\text{PV}^1 = [\text{size}_{\text{pv}}^1 \text{ location}_{\text{pv}}^1]$$

and DSTATCOM is a 1×2 matrix as follows:

$$\text{DSTATCOM}^1 = [\text{size}_{\text{DSTATCOM}}^1 \text{ location}_{\text{DSTATCOM}}^1]$$

where $\text{size}_{\text{pv}}^1$ and $\text{location}_{\text{pv}}^1$ are the initial proposed capacity and bus location for the PV unit, respectively, and $\text{size}_{\text{DSTATCOM}}^1$ and $\text{location}_{\text{DSTATCOM}}^1$ are the initial capacity and bus candidate to install the DSTATCOM unit, respectively.

By updating the ACO algorithm, the second, third, and the i th solution vectors are generated with the new proposed tie switches, PV, and DSTATCOM size and location as follows:

$$S^i = [\text{tie switches}^i, \text{PV}^i, \text{DSTATVOM}^i].$$

For each i th solution, a power flow program is carried out, the fuzzy membership values, and the fitness of objective function are evaluated and compared with the previous solution and the better solution has been selected and replaced. The proposed method is summarized in the following steps.

- Step 1) Read data of distribution system (bus, load, branch, sectionalizing, number of tie switches, etc.), initialize the ACO parameters and run the power flow program [6].
- Step 2) Generate a solution vector as “ S ” by offering the tie switches, PV, and DSTATCOM sizes and locations in the network without violating the constraints. Run the power flow program, evaluate the fuzzy membership value (μ) based on (10) for each objective function. Calculate the fitness of fuzzy objective function according to (11). Save the solution as the best solution.
- Step 3) Update the ACO pheromone value using (13) and (14). Generate a new solution by updating ACO, and compare the fitness value of the new solution with the best solution. If the new generated solution has a better fitness value than best solution, save it as the best solution.
- Step 4) Check the number of iterations, and if the number of iterations does not exceed from the specified value, repeat step 3).

TABLE I
SELECTED PARAMETERS FOR THE ACO

Pheromone trail weight (α)	0.72
Heuristic weight (β)	0.66
Evaporation rate (ρ)	0.075
Total ants (m)	40
Iteration	500

Step 5) Defuzzification of the best solution and print the result.

Step 6) Stop.

V. SIMULATION AND NUMERICAL RESULTS

Based on the proposed methodology, a program has been written in MATLAB software. In order to evaluate the performance of the proposed approach, the program has been applied on two test systems at the nominal load. In this paper, five different cases have been considered to analyze the effectiveness of the proposed approach.

Case I) The base system without reconfiguration, PV, and DSTATCOM units.

Case II) The base system only with reconfiguration.

Case III) The base system with reconfiguration and PV allocation.

Case IV) The base system with reconfiguration and DSTATCOM allocation.

Case V) The base system with simultaneous reconfiguration, PV, and DSTATCOM allocation.

The proposed approach is also applied on second test system at various load levels, i.e., 0.5 (light), 1.0 (nominal), and 1.6 (heavy) for cases I and V.

Using ACO, it is observed that by increasing the number of ants, the algorithm has shown a stronger global searching ability, but it requires more computation time. It is observed that the result is the best when $30 \leq m \leq 50$.

In case of smaller the value of α and β , the more possible ants may choose the paths that have already been chosen. If the value of α and β is too small, i.e., less than 0.5, the search may be fall in the local minima. If the value of α and β is large, i.e., more than 1, the influence of stimulating function will be increased; however, the convergent speed reduces. Based on the simulation results, it is observed that α and β lie in the range of 0.5–1 for the proposed approach.

Also, it is observed that, by increasing the evaporation rate (ρ), the pheromone evaporates with more speed, and therefore, the global searching capability increases, but the convergent speed of the algorithm decreases. The global searching capability of the ACO algorithm can be affected by decreasing the evaporation rate parameter. So, based on the simulation results, the outcome is the best if $0.02 \leq \rho \leq 0.1$. Accordingly, the selected ACO parameters for simulation are presented in Table I.

A. Test Systems

The test systems that are used in this study have some tie lines. It is considered the total number of tie switches are

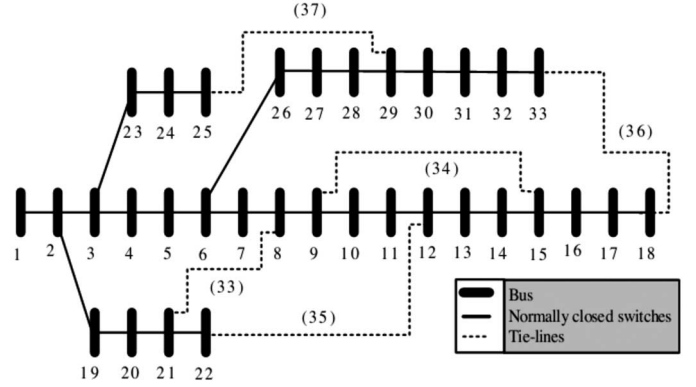


Fig. 4. 33-Bus test system with five tie lines.

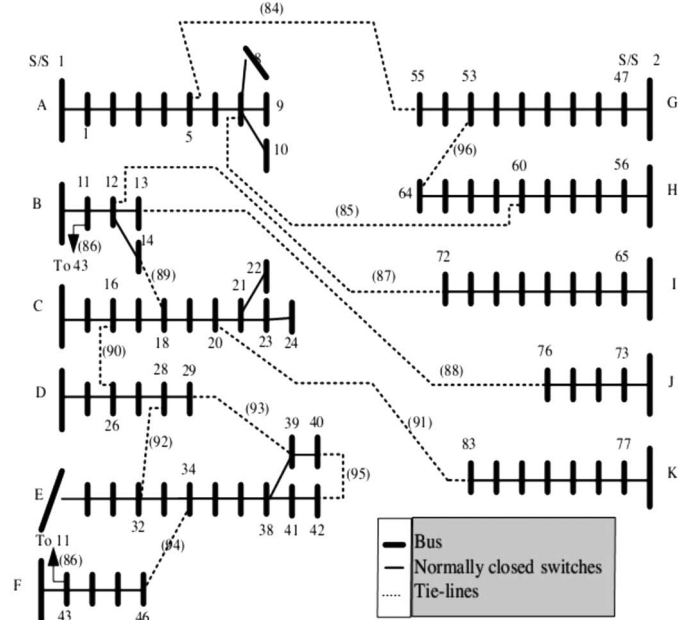


Fig. 5. TPC test system with 13 tie lines.

maintained after reconfiguration. In other words, whenever one of the tie switches is closed, one of the closed switches will be opened and the network remains in radial structure simultaneously feeding all loads. The following test systems are considered to evaluate the proposed method.

- 1) The first test system is a 33-bus distribution system with a total load of 3.7 MW and 2.3 MVAR having 5 tie and 32 sectionalizing switches [16]. The power flow is performed using $S_{base} = 100$ MVA and $V_{base} = 12.66$ kV. The single line diagram of this test system is presented in Fig. 4.
- 2) The second test system is a real distribution network from the Taiwan Power Company (TPC), which is shown in Fig. 5. This is practical 11.4-kV system equipped with 11 feeders, 13 tie switches, and 83 sectionalizing switches. The total system load, which is considered as balanced and constant, is 28.35 kW and 20.7 kVAR, respectively [17]. The power flow calculation is performed based on $S_{base} = 100$ MVA and $V_{base} = 11.4$ kV.

TABLE II
RESULTS OF 33-BUS TEST SYSTEM

Case	Tie switches	PV size (MVA) at bus	DSTATCOM size (MVA) at bus	P_{loss} (kW), VP, LB
I	33, 34, 35, 36, 37	–	–	202.5, 1.70, 67.71
II	31, 37, 7, 14, 9	–	–	143.69, 1.47, 58.42
III	28, 20, 17, 10, 14	1.387 at 9	–	93.30, 1.23, 49.64
IV	7, 9, 14, 32, 37	–	0.98 at 19	110.22, 1.08, 49.51
V	9, 14, 27, 33, 36	1.316 at 9	0.74 at 10	48.73, 0.82, 43.05

B. Obtained Results

The results of the proposed approach for 33-bus test system are presented in Table II for the five tested cases. It is observed from Table II that base case power loss in the system is 202.5 kW, which is reduced to 143.69, 93.30, 110.22, and 48.73 kW for cases II to V respectively. VP index is calculated as 1.7, 1.47, 1.23, 1.08, and 0.82 and LB index is obtained 67.71, 58.42, 49.64, 49.51, and 43.05 for cases I to V, respectively. Also, Table II shows the optimal location and size for PV and DSTATCOM units for the cases III to V.

By comparing the improvements obtained for cases II to V with the base case (case I), it is found that simultaneous reconfiguration with PV and DSTATCOM allocation (case V) has 75.93%, 36.42%, and 51.76% improvements in P_{loss} , LB, and VP indices, respectively.

Also, the results of ACO method (without fuzzy system) for 33-bus test system shows that the power loss is 184.12, 136.93, 1145.27, and 89.36 kW, VP index is 1.56, 1.48, 1.30, and 1.07, and LB index is 63.37, 56.79, 52.18, and 49.43 for cases II to V, respectively. Obtained results show that the performance of fuzzy-ACO approach is better as compared to the ACO (without fuzzy system) for 33-bus test system.

Similarly, the proposed approach is also applied on second test system, i.e., the TPC, and it is observed that all three indices of P_{loss} , LB, and VP have been improved for all cases II to V.

The results of the proposed method on TPC test system are presented in Table III for all five tested cases. It is observed from Table III that the base case power loss in the system is 531 kW, which is reduced to 383.15, 371.42, 368.09, and 293.55 kW for cases II to V, respectively. VP index is calculated as 2.5, 2.34, 2.03, 1.92, and 1.67 and LB index is obtained as 140.4, 111.38, 115.70, 102.65, and 98.47 for cases I to V, respectively. Also, Table III includes the optimal location and size for PV and DSTATCOM units for the cases III to V on TPC test system. The results of ACO approach (without fuzzy system) on TPC test system shows that the power loss is 427.55, 376.21, 392.80, and 352.04 kW, VP index is 2.31, 2.17, 2.10, and 1.93, and LB index is 132.04, 128.19, 121.53, and 117.99 for cases II to V, respectively. Obtained results show that the performance of fuzzy-ACO approach is better performance as compared to ACO (without fuzzy system) on TPC test system.

Obtained results are compared on both the systems under study and the following conclusions have been obtained.

TABLE III
RESULTS OF TPC TEST SYSTEM

Case	Tie switches	PV size (MVA) at bus	DSTATCOM size (MVA) at bus	P_{loss} (kW), VP, LB
I	84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96	–	–	531, 2.5, 140.4
II	7, 13, 34, 39, 41, 61, 84, 86, 87, 89, 90, 91, 92	–	–	383.15, 2.34, 111.38
III	7, 13, 34, 39, 42, 55, 72, 86, 89, 90, 91, 92, 96	2.761 at 25	–	371.42, 2.03, 115.70
IV	7, 13, 34, 39, 41, 61, 84, 86, 87, 89, 90, 91, 92	–	2 at 20	368.09, 1.92, 102.65
V	7, 13, 34, 39, 42, 55, 72, 86, 89, 90, 91, 92, 96	2.453 at 23	1.870 at 79	293.55, 1.67, 98.47

- 1) The performance of fuzzy-ACO method is better as compared to ACO (without fuzzy system).
- 2) As can be seen from Tables II and III, by applying the fuzzy-ACO method, the active power loss is reduced, VP is improved, and the LB of the feeders has been increased for cases II to V.
- 3) From Tables II and III, the results prove the superiority of case V (fuzzy-ACO approach) as compared to other cases. In other words, by investigating various cases involving reconfiguration, PV allocation, DSTATCOM allocation, and combination of them, it is found that reconfiguration simultaneous placement of PV and DSTATCOM units can be more beneficial than separate single-objective optimization. By comparing the improvements obtained by cases II to V as compared to the base case (case I) on the 33-bus test system, it is found that simultaneous reconfiguration with PV and DSTATCOM allocation (case V) has the most improvements in P_{loss} (75.93%), LB (36.42%), and VP (51.76%) indices. These results are repeated for the TPC test system with 44.71% improvement in P_{loss} , 29.86% in LB, and 33.2% in VP indices.
- 4) Since the obtained P_{loss} index for case III is better than the corresponding value for case IV in both test systems, it can be concluded that for power loss reduction, simultaneous reconfiguration and PV allocation is more effective than simultaneous reconfiguration along with DSTATCOM allocation. On the other hand, the calculated LB and VP indexes for case IV are better than the corresponding values for case III in both test systems. Therefore, in order to improve the LB or VP indexes of distribution system, simultaneous reconfiguration and DSTATCOM allocation can be more effective than reconfiguration along with PV allocation.
- 5) By comparing the improvements obtained by cases II to V from Tables II and III, it is confirmed that reconfiguration and simultaneous allocation of PV or DSTATCOM (cases III and IV) lead to more improvement in P_{loss} , LB, and VP indices as compared to single reconfiguration (case II).

TABLE IV
OBTAINED RESULTS BASED ON VARIOUS LOAD LEVELS
(TPC TEST SYSTEM)

Item		Load level		
		Light (0.5)	Nominal (1)	Heavy (1.6)
Case I	Tie switches	84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96	84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96	84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96
	P_{loss} (kW)	122.76	531	1510.34
	VP	1.7	2.5	2.74
	LB	108	140.4	185.02
Case V	Tie switches	7, 13, 34, 39, 42, 55, 72, 86, 89, 90, 91, 92, 96	7, 13, 34, 39, 42, 55, 72, 86, 89, 90, 91, 92, 96	7, 13, 34, 39, 41, 61, 84, 86, 87, 89, 90, 91, 92
	P_{loss} (kW)	103.61	293.55	1201
	VP	1.61	1.67	2.49
	LB	100.43	98.47	151.33

To assess the performance of the proposed approach at different load levels, the TPC test system is simulated at three load levels: 0.5 (light), 1.0 (nominal), and 1.6 (heavy) for cases I and V. The obtained results are presented in Table IV. It can be seen from Table IV that at light load (for case V), the base case power loss is 122.76 kW, which is reduced to 103.61 kW and the base case VP index is 1.7 that has been improved to 1.61. Also, at this load level, the LB index for the base case is 108, indicating that it is improved to 100.43. Similarly, it can be seen from Table IV that all three indices of P_{loss} , VP, and LB have been improved as compared to the base case at heavy load. The obtained results show that the performance of the proposed approach at different load levels is satisfactory.

Case V (at nominal load) is simulated using fuzzy-genetic algorithm (GA) and fuzzy-particle swarm optimization (PSO) and compared with the results obtained by fuzzy-ACO method. All parameters of each algorithm are tuned for achieving the best solution. The results of different methods on both test systems are presented in Table V for the case V).

From Table V, for the 33-bus test system P_{loss} is calculated as 73.26, 69.15, and 48.73 kW, VP is obtained 0.87, 0.84, and 0.82 and LB is 48.59, 45.67, and 43.05 using fuzzy-GA, fuzzy-PSO, and fuzzy-ACO approaches, respectively. Due to similar results obtained for the TPC test system (second part in Table V), it is confirmed that the performance of the fuzzy-ACO is better than that of the fuzzy-GA and fuzzy-PSO in all terms of loss reduction, improving VP, and equal LB for the both test systems. Also, it is observed that the performance of fuzzy-ACO approach is better than that of other tested approaches because of the lowest fitness equal to 0.147 in comparison with the fuzzy-PSO (0.164) and fuzzy-GA (0.181). Fig. 6 shows the percentage improvements in P_{loss} , LB, and VP indices using the fuzzy-GA, fuzzy-PSO, and fuzzy-ACO approaches for case V as compared to the base case (case I) at nominal load. It can be seen from Fig. 6 that among three terms P_{loss} , LB, and VP, the improvement in P_{loss} is more significant as compared to LB and VP for all three fuzzy-GA, fuzzy-PSO, and fuzzy-ACO approaches. This means that reconfiguration and simultaneous placement of PV and DSTATCOM units has a greater impact on loss reduction as compared to VP or LB improvement.

TABLE V
COMPARISON OF THE SIMULATION RESULTS FOR DIFFERENT METHODS

	Methods	Item	Case V
33 Bus test system	Fuzzy-GA	Tie switches	7, 9, 14, 32, 37
		P_{loss} (kW)	73.26
		VP	0.87
		LB	48.59
		PV size (MVA) at bus	1.284 at 17
	Fuzzy-PSO	DSTATCOM (MVA) at bus	0.69 at 9
		Tie switches	28, 20, 17, 10, 14
		P_{loss} (kW)	69.15
		VP	0.84
		LB	45.67
		PV size (MVA) at bus	1.309 at 7
		DSTATCOM (MVA) at bus	0.72 at 23
	Fuzzy-ACO	Tie switches	9, 14, 27, 33, 36
		P_{loss} (kW)	48.73
		VP	0.82
		LB	43.05
		PV size (MVA) at bus	1.316 at 9
		DSTATCOM (MVA) at bus	0.74 at 10
TPC test system	Fuzzy-GA	Tie switches	7, 13, 34, 39, 42, 55, 72, 86, 89, 90, 91, 92, 96
		P_{loss} (kW)	301.32
		VP	1.79
		LB	112.45
		PV size (MVA) at bus	2.380 at 37
	Fuzzy-PSO	DSTATCOM (MVA) at bus	1.706 at 62
		Tie switches	7, 13, 34, 39, 41, 61, 84, 86, 87, 89, 90, 91, 92
		P_{loss} (kW)	298.32
		VP	1.74
		LB	108.23
		PV size (MVA) at bus	2.416 at 21
		DSTATCOM (MVA) at bus	1.793 at 83
	Fuzzy-ACO	Tie switches	7, 13, 34, 39, 42, 55, 72, 86, 89, 90, 91, 92, 96
		P_{loss} (kW)	293.55
		VP	1.67
		LB	98.47
		PV size (MVA) at bus	2.453 at 23
		DSTATCOM (MVA) at bus	1.870 at 79

Accordingly, after the loss reduction, good improvement is seen in the VP index for all three methods on both test systems.

VI. CONCLUSION

In this paper, a fuzzy-ACO-based algorithm has been presented for simultaneous reconfiguration and allocation of PV and DSTATCOM units. The proposed approach is employed to mitigate the power loss, VP improvement, and equalizing the feeder LB in distribution system. To test the effectiveness of the proposed approach, five different cases have been tested on 33-bus test system and a real distribution network, i.e., Taiwan Power Company. Among the five cases, the multiobjective reconfiguration and simultaneous PV and DSTATCOM allocation case is found to be better than the others. For 33-bus system, the loss reduction is 75.93%, VP and LB improvements

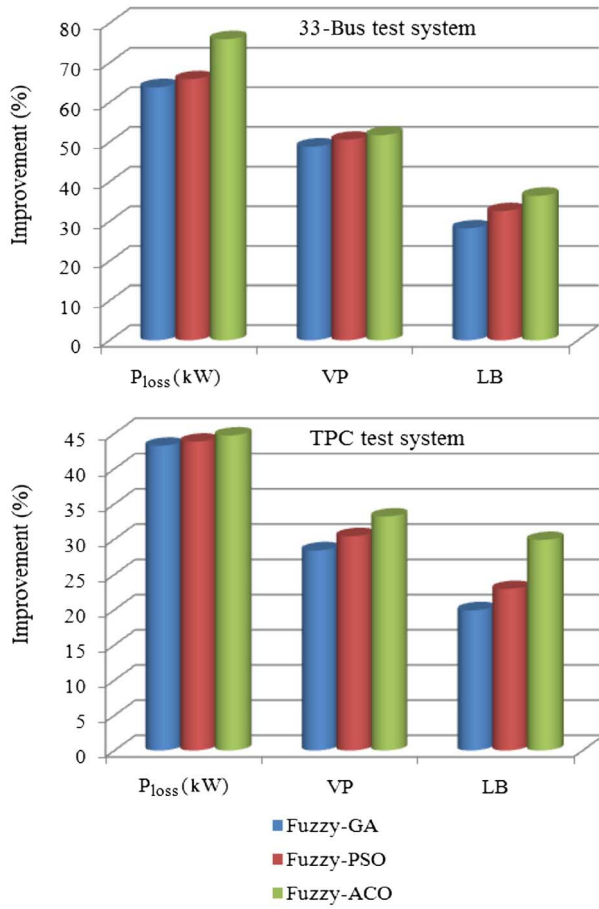


Fig. 6. Percentage improvements in P_{loss} , LB, and VP indexes using fuzzy-GA, fuzzy-PSO, and fuzzy-ACO methods.

are 51.76% and 36.42%, respectively, as compared to the base system. However, in case of Taiwan power company network, the loss reduction is 44.71%, and VP and LB improvements are 33.2% and 29.86, respectively. Obtained results are compared with the fuzzy-GA and fuzzy-PSO at nominal load and found to be better than the above-mentioned approaches because of the lowest fitness. Also, the performance of fuzzy-ACO is better as compared to the ACO method. Furthermore, the reconfiguration along with PV and DSTATCOM is more advantageous than the individual one. Finally, obtained results confirm the satisfactory operation of the proposed approach at different load levels.

APPENDIX

A method based on [18] is used in which a graph (i.e., distribution system with total number of n_f feeders) can be described by a bus incidence matrix (A). $A = [a_{bn}]$, where

$$a_{bn} = \begin{cases} 0, & \text{if branch } b \text{ is not connected to node } n \\ 1, & \text{if branch } b \text{ is directed away from node } n \\ -1, & \text{if branch } b \text{ is directed toward node } n. \end{cases}$$

A radial distribution feeder that feeds all of loads can be simulated by a tree graph, i.e., a connected graph (i.e., there is not any isolated load) without any cycle (i.e., the network has a radial structure). For each tree graph, satisfying the following

equation is necessary (it is not the sufficient condition): $q = p - 1$, where p is the total number of vertices (nodes) and q is the total number of edges (branches).

For each feeder, a bus incidence matrix is defined (A). For each matrix, one node must be selected as the reference. The column related to the reference nodes is deleted from A , and the remaining matrix is determined by A^1 for a tree graph, A^1 is a square matrix. The determinant of A^1 is then calculated. If $|\det(A^1)| = 1$ (it is the sufficient condition to having a tree structure), then the distribution system is radial and all loads are feeding, else, this means that the network is not radial or some loads are isolated from the network.

REFERENCES

- [1] H. B. Tolabi, M. Gandomkar, and M. Bahreyni Borujeni, "Reconfiguration and load balancing by software simulation in a real distribution network for loss reduction," *Can. J. Elect. Electron. Eng.*, vol. 2, no. 8, pp. 386–391, Aug. 2011.
- [2] A. Merlin and H. Back, "Search for a minimal-loss operating spanning tree configuration in an urban power distribution system," in *Proc. 5th Power Syst. Comput. Conf. (PSCC)*, Cambridge, U.K., 1975, pp. 1–18.
- [3] A. Kavousi-Fard and T. Niknam, "Multi-objective stochastic distribution feeder reconfiguration from the reliability point of view," *Energy*, vol. 64, pp. 342–354, Jan. 2014.
- [4] R. Kollu, S. R. Rayapudi, and V. L. N. Sadhu, "A novel method for optimal placement of distributed generation in distribution systems using HSDO," *Int. Trans. Elect. Energy Syst.*, vol. 24, pp. 547–561, Apr. 2014.
- [5] M. Farhoodnea, A. Mohamed, H. Shareef, and H. Zayandehroodi, "Optimum D-STATCOM placement using firefly algorithm for power quality enhancement," in *Proc. 7th Power Eng. Optim. Conf. (PEOCO)*, Langkawi, Malaysia, 2013, pp. 98–102.
- [6] R. S. Rao, S. V. L. Narasimham, M. R. Raju, and A. S. Rao, "Optimal network reconfiguration of large-scale distribution system using harmony search algorithm," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1080–1088, Aug. 2011.
- [7] R. Srinivasa Rao, K. Ravindra, K. Satish, and S. V. L. Narasimham, "Power loss minimization in distribution system using network reconfiguration in the presence of distributed generation," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 317–325, Feb. 2013.
- [8] T. Xiangqian, X. Keqing, S. Ming, and M. Xianhong, "Reactive power and unbalance compensation with DSTATCOM," in *Proc. Int. Conf. Elect. Mach. Syst. (ICEMS)*, Nanjing, China, 2005, vol. 2, pp. 1181–1184.
- [9] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, 1965.
- [10] M. Dorigo, V. Maniezzo, and A. Colomi, "The ant system: An auto-catalytic optimizing process," Politecnico di Milano, Italy, Tech. Rep. 91-016 Revised, 1991.
- [11] M. Dorigo, "Optimization, learning and natural algorithms," Ph.D. dissertation, Dipartimento di Elettronica, Politecnico di Milano, Milano, Italy, 1992.
- [12] V. Maniezzo, L. M. Gambardella, and F. DeLuigi, *Ant Colony Optimization: New Optimization Techniques in Engineering*, G. C. Onwubolu and B. V. Babu, Eds. New York, NY, USA: Springer, 2004, pp. 101–117.
- [13] M. Gandomkar and H. B. Tolabi, "Investigation of simulated annealing, ant-colony and genetic algorithms for distribution network expansion planning with distributed generation," in *Proc. 9th WSEAS Int. Conf. Instrum. Meas. Circuits Syst. (IMCAS'10)*, China, 2010, pp. 48–52.
- [14] M. Dorigo, V. Maniezzo, and A. Colomi, "The ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst. Man Cybern. B*, vol. 26, no. 1, pp. 29–41, Feb. 1996.
- [15] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 53–66, Apr. 1997.
- [16] M. E. Baran and F. Wu, "Network reconfiguration in distribution system for loss reduction and load balancing," *IEEE Trans. Power Del.*, vol. 4, no. 2, pp. 1401–1407, Apr. 1989.
- [17] L. W. Oliveira *et al.*, "Optimal reconfiguration and capacitor allocation in radial distribution systems for energy losses minimization," *Elect. Power Energy Syst.*, vol. 32, pp. 840–848, Oct. 2010.
- [18] J. Grainger and William D. Stevenson, Jr., "Power flow solutions," in *Power System Analysis*, 1st ed. New York, NY, USA: McGraw-Hill, 1994.



Hajar Bagheri Tolabi (S'12–M'13) received the B.Sc. degree in power engineering from Bu-Ali Sina University, Hamedan, Iran, in 2006, and M.Sc. degree in power engineering from Islamic Azad University, Saveh Branch, Saveh, Iran, in 2010.

Currently, she is a Lecturer with Islamic Azad University, Khorramabad Branch, Khorramabad, Iran. She has published/presented one book and more than 20 research papers in reputed international and national journals and conference proceedings. Her research interests include renewable energy, distribution systems, and optimization.



Mohd Hasan Ali (S'00–M'04–SM'08) received the Ph.D. degree in electrical and electronic engineering from Kitami Institute of Technology, Kitami, Japan, in 2004.

He was as a Lecturer with EEE Department, RUET, Rajshahi, Bangladesh, from 1995 to 2004, and later became an Assistant Professor in 2004. He was a Postdoctoral Research Fellow under the Japan Society for the Promotion of Science (JSPS) Program, Kitami Institute of Technology, from November 2004 to January 2007. He also worked as a Research Professor with the Electrical Engineering Department, Changwon National University, Changwon, South Korea, from February 2007 to December 2007. He was a Postdoctoral Research Fellow with the Electrical and Computer Engineering Department, Ryerson University, Toronto, ON, Canada, from January 2008 to June 2009. He also worked as a Faculty with the Electrical Engineering Department, University of South Carolina, Columbia, SC, USA, from July 2009 to August 2011. Currently, he is working as an Assistant Professor with the Electrical and Computer Engineering Department, University of Memphis, Memphis, TN, USA, where he leads the Electric Power and Energy Systems (EPES) Laboratory of the university. His research interests include advanced power systems, smart-grid and micro-grid systems, renewable energy systems, energy storage systems, and flexible ac transmission systems (FACTS). He has more than 125 publications including 2 books, 1 book chapter, 43 top-ranked journal papers, 62 peer-reviewed international conference papers, and 20 national conference papers.

Dr. Ali is the Chair of the PES of the IEEE Memphis Section.



M. Rizwan (M'14) received the Ph.D. degree in power engineering from Jamia Millia Islamia, New Delhi, India, in 2011.

Currently, he is an Assistant Professor with the Department of Electrical Engineering, Delhi Technological University, New Delhi, India. He has published/presented more than 45 research papers in reputed international and national journals and conference proceedings. He is also associated with many journals in different capacities. His research interests include renewable energy, particularly solar photovoltaic and soft computing techniques.

Dr. Rizwan has been awarded two sponsored research projects in the area of renewable energy systems and power quality.

Slow light enhanced slotted photonic crystal waveguide

Preeti Rani, Yogita Kalra and R. K. Sinha

TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Department of Applied Physics, Delhi Technological University (Formerly Delhi College Of Engineering, University of Delhi), Bawana Road, Delhi-110042, India
dr_rk_sinha@yahoo.com

Abstract: Slotted waveguide with slow light effect in photonic crystal composed of hexagonal arrangement of square air holes on SOI substrate has been proposed. The proposed structure has high group index and low Group Velocity Dispersion.

OCIS codes: (160.5298) Photonic crystals; (160.5293) Photonic band gap materials; 130.5296 Photonic crystal waveguides;

1. Introduction

In recent years, slow light with remarkable low group velocity has attracted wide attention in the research arena. Slow light effect leads to the increase in the light matter interaction which allows the designing of highly sensitive and miniaturized devices [1, 2]. Photonic crystal is one of the most suitable platforms for the realization of slow light in a waveguide configuration. The conventional silicon on insulator (SOI) based photonic crystal (PhC) waveguide [3-5] confines light into the high index guiding region. The slotted PhC waveguide [6, 7] has unique property to guide and confine light into the low index slot with enhanced field.

2. Structure design

In this paper, PhC based slotted waveguide on SOI substrate has been proposed. The PhC consists of hexagonal arrangement of square air holes in silicon slab on SOI substrate. The slot has been created by removing a row of air holes and by replacing it with an air slot of width 200nm as shown in figure 1. The width of air holes has been taken as 300nm. A 2D plane wave expansion (PWE) method with slab equivalent index method has been used to analyze the properties of the proposed structure. The 220nm thickness of the silicon slab has an effective index of 2.79. The dispersion curve for the proposed structure is shown in figure 2 where the dotted lines correspond to the two modes and dark straight line is the light line. These two modes lie within the band gap range of 0.2717- 0.3731, as calculated by PWE method. The flat band section of the even mode below the light line, where $\frac{d\omega}{dk}$ is small, has been utilized for the slow light effect in the proposed slotted PhC waveguide.

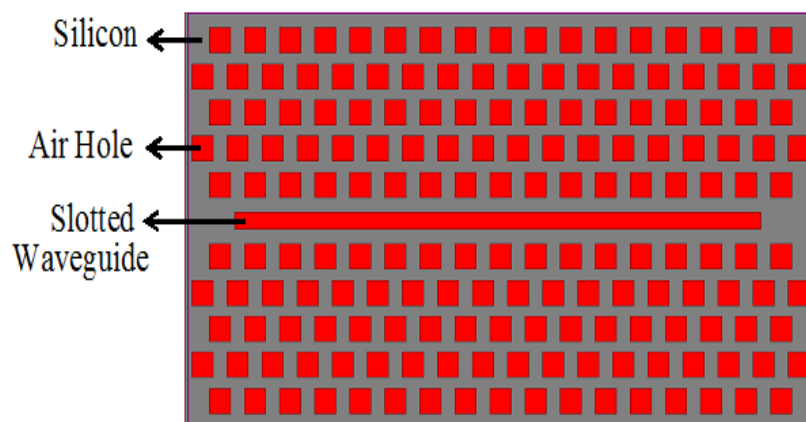


Figure 1 Schematic representation of slow light enhanced slotted waveguide structure

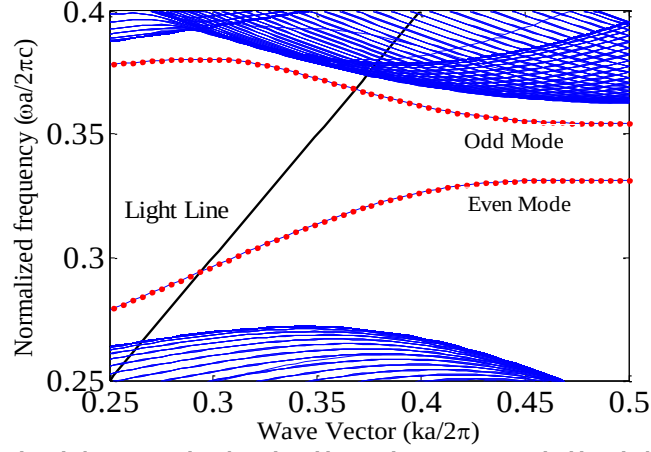


Figure 2 Dispersion relation of the slotted PhC waveguide, where dotted lines indicates an even and odd mode, blue lines indicates slab modes and dark straight line corresponds to the light line

3. Results and discussion

The group index n_g and GVD parameter (β) of the light waves has been calculated from the equations given below, where ω is the frequency, k is the wave propagation vector, c is the velocity of light in vacuum and v_g is the group velocity.

$$v_g = \frac{d\omega}{dk} \quad (1)$$

$$n_g = \frac{c}{v_g} \quad (2)$$

$$\beta = \left(\frac{d^2k}{d\omega^2} \right) = \frac{1}{c} \frac{dn_g}{d\omega} \quad (3)$$

For the proposed structure, group index and group velocity dispersion parameter has been calculated and plotted with respect to the normalized frequency as shown in figures 3 and 4 using above mentioned equations. From figures 3 and 4, we have concluded that for the desired frequency range in the flat band section of the dispersion curve the group index for the proposed structure is high and the group velocity dispersion is low. In the proposed design, the slotted waveguide satisfies the basic requirements of slow light i.e. low GVD and high n_g . Hence, we have concluded that the proposed design can be efficiently used for the enhancement of slow light effect which can further be used in the sensing applications.

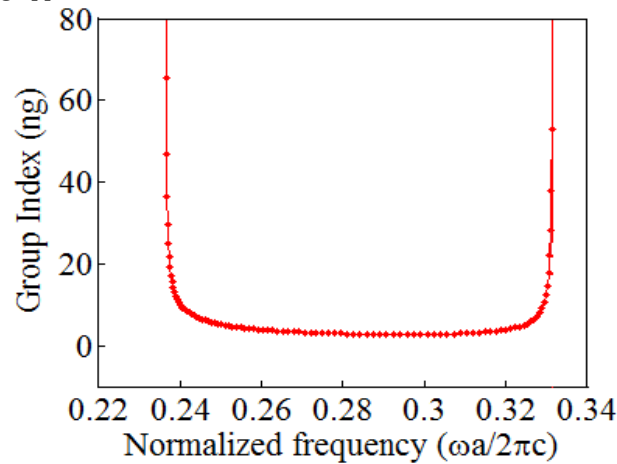


Figure 3 Group index versus normalized frequency for the slotted PhC waveguide

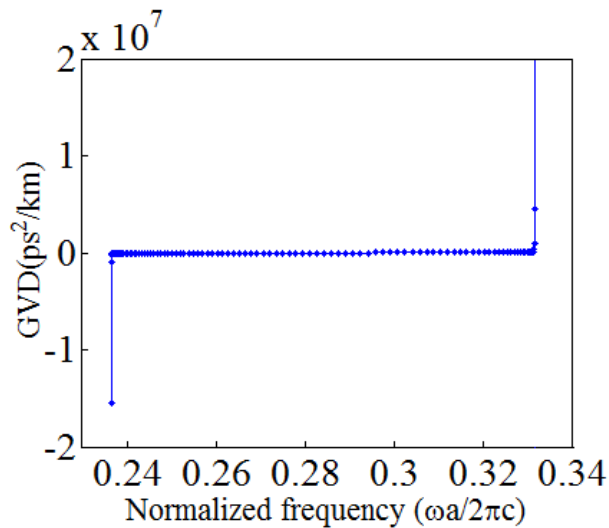


Figure 4 Group velocity dispersion versus normalized frequency

4. Conclusion

In this paper, slotted PhC waveguide based structure on SOI platform has been demonstrated. The proposed structure has high group index and low group velocity dispersion in the slow light region of the dispersion curve. The present results have provided basis for the realization of slow light based miniature optical devices.

5. Acknowledgement

Authors gratefully acknowledge the initiatives and support from TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication at Delhi College of Engineering, now DTU, Delhi under Mission REACH program of Technology Vision-2020, Government of India.

6. References

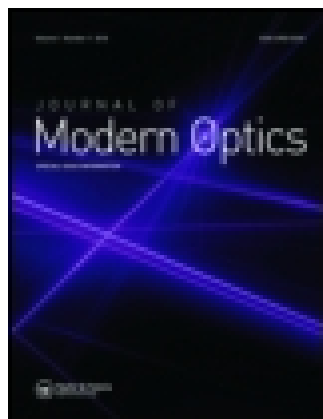
- [1] J. B. Khurgin and R. S. Tucker, "Slow Light: Science and Applications," Boca Raton, ed. (FL, USA: CRC Press, 2008).
- [2] J. -M. Brosi, "Slow-light photonic crystal devices for high speed optical signal processing," Ph.D. dissertation, Univ. Karlsruhe (TH), Karlsruhe, Germany, July 2008.
- [3] T. F. Krauss, "Why Do we need slow light?," Nature photonics, vol. 2, pp. 448-450, 2008.
- [4] T. Baba, "Slow light in photonic crystals," Nature photonics, vol. 2, no. 8, pp. 465-473, 2008.
- [5] Swati Rawal, R. K. Sinha and Richard M. De La Rue, "Slow light miniature devices with ultra-flattened dispersion in silicon-on-insulator photonic crystal," Optics Express, vol. 17, pp. 13315-13325, 2009.
- [6] Y. Zhao, Ya-Nan Zhang, Di Wu and Qi Wang, "Wideband slow light with large group index and low dispersion in slotted photonic crystal waveguide," Journal of Light wave Technology, vol. 30, no. 17, pp. 2812-2817, 2012.
- [7] Ya-Nan, Yong Zhao and Qi Wang, "Optimizing the slow light properties of slotted photonic crystal waveguide and its applications in a high sensitivity gas sensing system," Measurement science and Technology IOP publishing, vol. 24, pp. 105109, 2013.

This article was downloaded by: [Delhi Technological University]

On: 23 December 2014, At: 02:13

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Modern Optics

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tmop20>

Slow light generation in single-mode tellurite fibers

Than Singh Saini^a, Ajeet Kumar^a & Ravindra Kumar Sinha^a

^a TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Applied Physics Department, Delhi Technological University, Delhi, India

Published online: 08 Dec 2014.



[Click for updates](#)

To cite this article: Than Singh Saini, Ajeet Kumar & Ravindra Kumar Sinha (2014): Slow light generation in single-mode tellurite fibers, Journal of Modern Optics, DOI: [10.1080/09500340.2014.988767](https://doi.org/10.1080/09500340.2014.988767)

To link to this article: <http://dx.doi.org/10.1080/09500340.2014.988767>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Slow light generation in single-mode tellurite fibers

Than Singh Saini, Ajeet Kumar and Ravindra Kumar Sinha*

TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Applied Physics Department, Delhi Technological University, Delhi, India

(Received 19 July 2014; accepted 13 November 2014)

Theoretical investigations of stimulated Brillouin scattering-based tunable slow light in (i) *Er*-doped tellurite and (ii) undoped tellurite fibers are reported. Maximum allowable pump power for undistorted output pulse, Brillouin gain, time-delay, figure-of-merit, and time-delay slope efficiency of both the fibers has been obtained. We have found that (i) Brillouin gain up to ~91 dB and time delay of 140 ns can be achieved using 1100 mW pump power in 2 m *Er*-doped fiber and (ii) Brillouin gain up to ~86 dB and time delay of ~227 ns using 23 mW pump power in 100 m undoped tellurite fiber can be achieved. Simulated results indicate that the time delay in fibers can be tuned with the pump power to obtain tunable slow light features in these fibers. We feel that detailed theoretical investigations and simulations carried out in the study have potential impact in the design and development of slow light-based photonic devices.

Keywords: slow light; single mode-tellurite fibers; fiber nonlinear optics; stimulated Brillouin scattering

1. Introduction

The term “slow light” means the propagation of light pulse at a very low group velocity. It arises when the light pulse is substantially slowed down by the interaction with propagating medium. The interest on the topic of slow light is increasing day-by-day because of the many applications such as, data synchronization, optical memories, optical buffering, optical signal processing, microwave photonics, and precise interferometric instruments [1–3]. Slow light techniques give the promising solutions for broadband and tunable time delay for microwave and millimeter wave systems. On-chip all-optical signal processing photonic devices are possible by storing the light and enhancing the nonlinearity. When light interacts with the optically resonant medium, its speed and other propagation characteristics can be controlled [4]. In the literature, there are many methods to observe the phenomena of slow light which includes stimulated Brillouin scattering (SBS) and stimulated Raman scattering (SRS) in standard single-mode fiber [5–9], Raman-assisted optical parametric amplification [10], coherent population oscillation [11], and slow light relying on large dispersive structures [12]. Among all above-mentioned slow light-based technologies, SBS and SRS in optical fibers have attracted more interest due to its compatibility with optical fiber communication systems, room temperature operation, and tunable working wavelength feature. Earlier, silica-based fibers have been used for SBS-based slow light applications [13]. However, to obtain effective time delay and gain in silica fibers, the required length of

fiber is in the order of kilometers. For more practical applications, several efforts have been done to reduce the fiber length using high refractive index materials [14–19] and photonic crystal fibers [20]. It has been confirmed that the tellurite fibers are one of the most promising gain medium for Brillouin fiber lasers, amplifiers, and slow light generation [19]. Numerical studies of SBS in standard silica fibers for slow light applications have been done earlier in various research and review articles [21–23]. However, there is a need to explore features of tunable slow light in tellurite optical fibers. In the study, a detailed theoretical investigation of SBS-induced slow light in two different tellurite fibers has been carried out. Two nonlinear coupled SBS equations have been solved using the experimental data available in [18] for *Er*-doped tellurite fiber and in [19] for undoped tellurite fiber. The geometrical parameters for *Er*-doped tellurite fiber are as: the core diameter is 2.6 μm with index contrast of 1.6%, cutoff wavelength of 1.3 μm and fiber length of 2 m. While for undoped tellurite fiber, the core diameter is 2.2 μm , with index contrast of 2.2%, cutoff wavelength of 1.3 μm , and fiber length of 100 m. Our analytically calculated results on Brillouin gain and time delay match well with experimentally measured results, as mentioned in [18] and [19]. The effects of Brillouin gain, time delay, and figure of merit (FOM) of both the fibers have been calculated numerically. Our simulated results indicate that the time delay in fibers can be tuned with pump power and thereby, tunable slow light features can be obtained from these fibers. The pump power

*Corresponding author. Emails: rksinha@dce.edu, dr_rk_sinha@yahoo.com

can be tuned between 12.3 and 1100 mW for 2 m long *Er*-doped tellurite fiber, while between 1.5 and 23 mW for 100 m undoped tellurite fiber for getting desirable time delay.

2. Theory of SBS-based slow light

In the process of SBS, there is an interaction between a highly intense pump wave and a counter-propagating, typically weaker signal wave, which is detuned in frequency [24]. As a result, a slowly traveling intensity wave of the frequency equal to the difference between the frequencies of the pump and signal wave is generated which causes the traveling density variations (called acoustic wave). Due to the photo-elastic effect, the acoustic wave leads to a traveling grating of refractive index variation along the fiber. The traveling grating can strongly couple optical power between the pump and signal waves at phase matching condition, when the frequency difference between pump and signal power is equal to the Brillouin frequency shift (Ω_B).

Let ω_p is the frequency of continuous pump wave with amplitude A_p and ω_s is the frequency of counter-propagating signal wave with amplitude A_s . The signal frequency is much smaller than frequency of pump wave and $\omega_p - \omega_s = \Omega_B$.

The coupled nonlinear differential equations for pump and signal waves are given by the following equations.

$$\frac{dA_p}{dz} = -g(\omega_s)A_p^2A_s^2 - \alpha_p A_p \quad (1)$$

$$-\frac{dA_s}{dz} = g(\omega_s)A_p^2A_s^2 - \alpha_s A_s \quad (2)$$

assuming that the pump wave is undepleted, and ($\alpha_p \approx \alpha_s$) = α , the solution of above equations is given by [13,25].

$$A_s(0) = A_s(L)\exp[(g(\omega_s)L_{\text{eff}} - \alpha L)] \quad (3)$$

with,

$$L_{\text{eff}} = \alpha^{-1}(1 - \exp(-\alpha L)) \quad (4)$$

and

$$g(\omega_s) = \frac{Kg_0|A_p|^2}{1 - j2(\omega_p - \omega_s - \Omega_B)/\Gamma_B} \quad (5)$$

where L and L_{eff} are the real and effective length of the fiber respectively, α is the loss in the fiber, and K is the polarization factor and depends on the polarization properties of the fiber. The value of polarization factor is 1, if the polarization is maintained and 0.5, if polarization is not maintained. Although, some results [17,18,26]

showed that $K = 0.667$ is more appropriate, for the fiber which is low-birefringence and very high polarization beat length. In our simulation, $K = 0.667$ has been used. The real part of $g(\omega_s)$ gives the frequency-dependent amplitude gain of signal and the imaginary part of it gives the corresponding phase delay in optical pulse which is illustrated in Figure 1. Γ_B is the SBS line width.

For peak value of Brillouin gain or at $\omega_p - \omega_s = \Omega_B$, $g(\omega_s)$ becomes $Kg_0|A_p|^2$. where g_0 is the line center SBS gain coefficient and related to the inherent material gain coefficient as $g_0 = g_B/A_{\text{eff}}$; and g_B is the inherent Brillouin gain coefficient as calculated from the Brillouin gain bandwidth using the following relation [27].

$$g_B = \frac{2\pi n^7 p_{12}^2}{c\lambda_p^2 v_A \Delta v_B} \quad (6)$$

where n is the refractive index of material of fiber, p_{12} is the longitudinal elasto-optic coefficient, c is the velocity of light, λ_p is the pump wavelength, ρ is the material density, v_A is the acoustic velocity, and Δv_B is the Brillouin gain bandwidth.

The real and imaginary part of $g(\omega_s)$ are related by the following Kramers–Kronig relations [28].

$$I_m[g(\omega_s)] = \frac{2}{\pi} \int_0^\infty \frac{\omega' R[g(\omega')]}{\omega'^2 - \omega_s^2} d\omega' \quad (7)$$

The maximum allowable pump power P_{max} (the maximum pump power below which the output pulse is undistorted) related to the Brillouin gain coefficient g_B by the following relation [29].

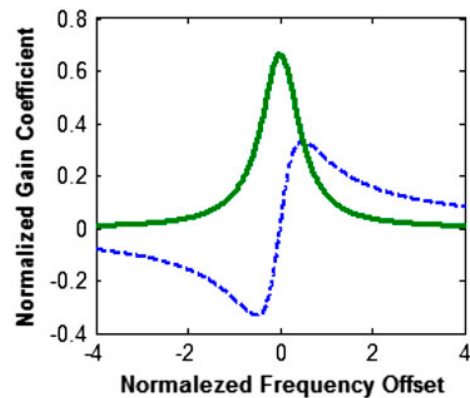


Figure 1. Illustration of real (solid line) and imaginary (dotted line) parts of the SBS gain coefficient, $g(\omega_s)$ as a function of normalized frequency offset $(\omega_p - \omega_s - \Omega_B)/\Gamma_B$ for a continuous pump wave. It is to be noted that the gain coefficient is normalized to $g_0|A_p|^2$. (The colour version of this figure is included in the online version of the journal.)

$$P_{\max} = 21 \frac{A_{\text{eff}}}{K g_B L_{\text{eff}}} \quad (8)$$

where A_{eff} is the effective cross-sectional area and is defined as $A_{\text{eff}} = \pi \omega_0^2$ ($\omega_0 = \frac{1}{e^2}$ intensity radius of the Gaussian distribution). Once pump power approached to $P_{\max}$, then the backscattered wave can be generated from the background noise in the fiber, which leads to serious output pulse distortion. Therefore, the value of pump power must be smaller than the $P_{\max}$. The minimum value of pump power which is required to initiate the SBS effect (when Brillouin gain is larger than the total loss in the fiber) is given as

$$P_{\min} = \frac{\alpha A_{\text{eff}} L}{K g_B L_{\text{eff}}} \quad (9)$$

where α is the loss coefficient of the fiber. The SBS-induced time delay per unit length and per unit pump power by slow light fiber devices can be expressed as [4].

$$\frac{\Delta t_d}{P_p L_{\text{eff}}} = \frac{g_o K}{\Gamma_B} \quad (10)$$

Another important parameter for slow light is time delay slop efficiency which is defined as below [19].

$$S_{dg} = \frac{\Delta t_d}{G_B} = \frac{1}{8.686 \pi \Delta \nu_B} \quad (11)$$

To insure the suitability of the fibers as a SBS slow light medium, it is very important to know their FOM. The FOM of the fiber is defined as [30].

$$\text{FOM} = \frac{G}{P_{\text{pump}} L_{\text{eff}} n} \quad (12)$$

where G is the Brillouin gain in dB. For simulating the above equations, we have developed MATLAB codes and simulated results have been verified with the experimental results published earlier in [18,19].

3. Results and discussion

3.1. Simulation of SBS slow light in Er-doped tellurite fiber

The pump and signal waves propagation are simulated over a 2-m long Er-doped tellurite fiber, which are based on the same values used in [18] for all simulation parameters; specifically, the pump power, $P_p = 360$ mW at 1542 nm, the Brillouin gain coefficient, $g_B = 1.47 \times 10^{-10}$ m/W, Brillouin shift, $\Omega_B = 7.882$ GHz, the mode effective area, $A_{\text{eff}} = 9.18 \mu\text{m}^2$, the fiber loss = 0.51 dB/m (i.e. $\alpha = 0.117/\text{m}$), and the polarization factor, $K = 0.667$. Using above-mentioned parameters,

the Brillouin gain of the fiber against the frequency difference between pump and signal ($\omega_p - \omega_s$) in 2-m long Er-doped tellurite fiber has been illustrated in Figure 2. It is clearly seen that the peak gain of ~29 dB is obtained in this fiber at frequency shift of 7.882 GHz.

Figure 3 illustrates the effect of pump power on Brillouin gain. It is clear from this figure that the Brillouin gain linearly increases with pump power. Our simulation result shows that for pump power of 360 mW, the value of Brillouin gain is 29.72 dB, which corresponds to 0.0825 dB/mW of the pump power. It is very close to the experimentally measured value [18], which validates that the formulae for Brillouin gain and MATLAB codes developed for its calculation are correct. Maximum 57.8 dB unsaturated gain has been achieved at pump power of 700 mW. As shown in Figure 3, the minimum pump power which is required to initialize the SBS effect is approximately 12.3 mW, as confirmed by solving the equation (9). The maximum pump power up to which the output pulse does not get affected by the distortion is about 1100 mW which is obtained from Equation (8). Pump power above 1100 mW leads to serious output pulse distortion. The maximum Brillouin gain which can be achieved at 1100 mW is about 91 dB for 2-m long fiber.

Figure 4 explains the dependence of the SBS-induced time delay per unit pump power on the real length of the fiber. For 2-m length of fiber, 127 ps/mW time delay has been obtained. Therefore, pump power with 360 mW introduces total time delay of 45.72 ns in 2-m length of Er-doped fiber.

Finally, the time delay in pulse with the pump power has been illustrated in Figure 5. The maximum time delay using 2 m of Er-doped fiber has been obtained up to 140 ns at 1100 mW pump power. The time delay slop

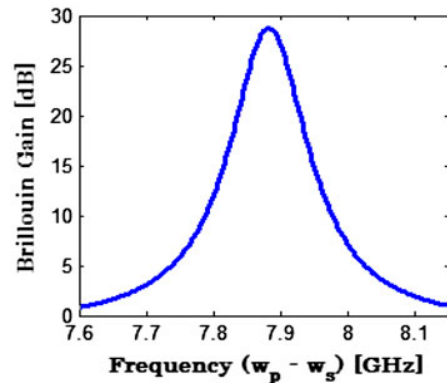


Figure 2. Brillouin gain of the fiber with frequency difference between pump and signal ($\omega_p - \omega_s$) in 2 m Er-doped tellurite fiber. (The colour version of this figure is included in the online version of the journal.)

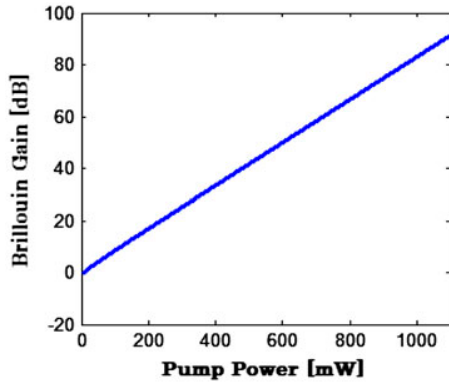


Figure 3. Effect of pump power on Brillouin gain. (The colour version of this figure is included in the online version of the journal.)

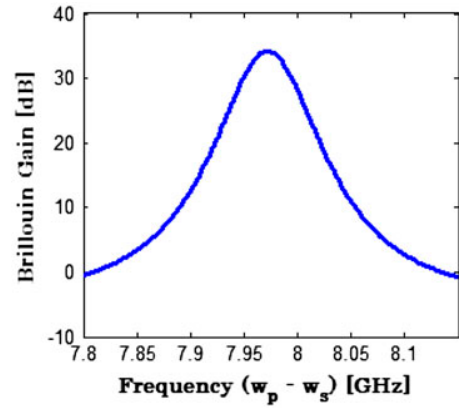


Figure 6. Brillouin gain of the fiber against the frequency difference between pump and signal ($\omega_p - \omega_s$) in 100 m long undoped tellurite fiber with 10 mW pump power. (The colour version of this figure is included in the online version of the journal.)

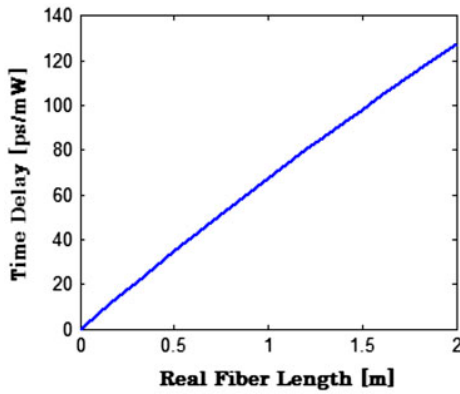


Figure 4. Time delay as a function of real fiber length. (The colour version of this figure is included in the online version of the journal.)

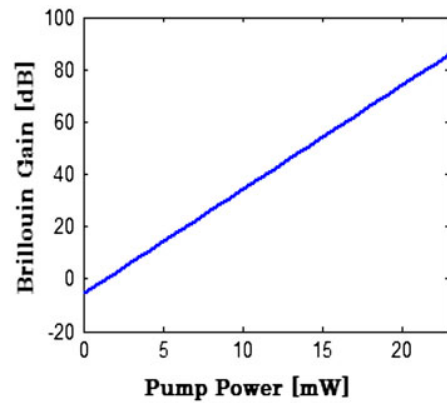


Figure 7. Brillouin gain as a function of pump power. (The colour version of this figure is included in the online version of the journal.)

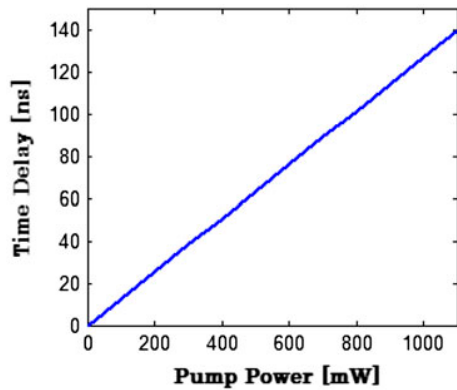


Figure 5. Time delay as a function of pump power. (The colour version of this figure is included in the online version of the journal.)

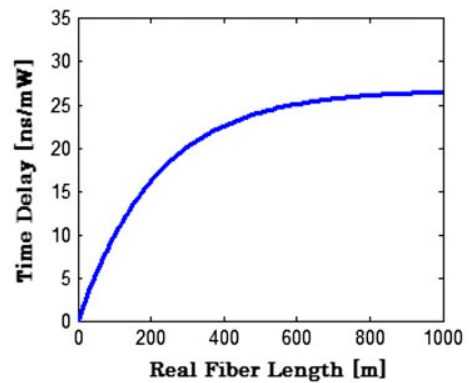


Figure 8. Time delay per unit power with real fiber length. (The colour version of this figure is included in the online version of the journal.)

efficiency for this fiber calculated from Equation (11) is 1.54 ns/dB, and the FOM of fiber is 22 which is calculated using the Equation (12).

3.2 Simulation of SBS slow light in undoped tellurite fiber

The second fiber which we have chosen is undoped tellurite fiber with the following parameters as in Ref. [19]. Pump power, $P_p = 10$ mW at 1550 nm, length of the fiber = 100 m, the Brillouin gain coefficient, $g_B = 1.6989 \times 10^{-10}$ m/W, Brillouin shift, $\Omega_B = 7.972$ GHz, the mode effective area, $A_{\text{eff}} = 6.9697 \mu\text{m}^2$, the fiber loss = 0.0558 dB/m (i.e. $\alpha = 0.01354/\text{m}$), and the polarization factor, $K = 0.667$. Using the above parameters, the Brillouin gain of the fiber against the frequency change between pump and signal ($\omega_p - \omega_s$) in undoped tellurite fiber has been illustrated in Figure 6. The maximum value of Brillouin gain of 34.2 dB has

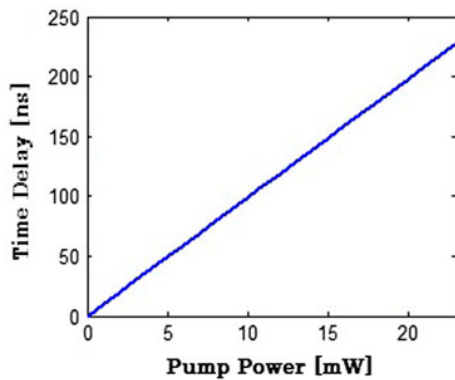


Figure 9. Total time delay with pump power. (The colour version of this figure is included in the online version of the journal.)

been obtained at 7.972 GHz frequency shift with pump power of 10 mW, which is well matched with the experimental result given in Ref. [19] for same fiber parameters.

Brillouin gain of the fiber linearly increases with pump power which is shown in Figure 7. Simulation results show that for pump power of 23 mW, the value of Brillouin gain is ~ 86 dB, which corresponds to 3.73 dB/mW of the pump power. The minimum pump power which is required to initialize the SBS effect is about 1.5 mW for this fiber. The maximum pump power for which output pulse does not affected by the distortion is 23.5 mW which is calculated by Equation (8). Based on the simulated results for doped and undoped tellurite fibers, we can conclude that the maximum allowable pump power is larger for Er-doped tellurite fiber in comparison to that of for undoped tellurite fiber. Therefore, for obtaining same time delay, we required short length of Er-doped tellurite fiber in comparison to undoped tellurite fiber.

The dependence of the SBS-induced time delay per unit pump power on the real length of the fiber has been shown in Figure 8. The time delay per unit power obtained by 100 m fiber is about 9.88 ns/mW. Therefore, pump power with 10 mW introduces total time delay of ~ 99 ns in 100 m length of undoped tellurite fiber. As shown in Figure 8, the maximum value of time delay per mW power for 1000 m long fiber is 26.5 ns/mW.

The total time delay with the pump power has been illustrated in Figure 9. The maximum time delay using 100 m undoped tellurite fiber has been obtained up to 227 ns at 23 mW pump power. The time delay slop efficiency for this fiber calculated from Equation (11) is 1.747 ns/dB, and the FOM has been calculated 30 using the Equation (12). All the features of both the fibers are tabulated in Table 1.

Table 1. Fibers characteristics.

Parameters	Er -doped tellurite fiber	Undoped tellurite fiber
Brillouin gain coefficient, g_B (m/W)	$1.47 \times 10^{-10\text{a}}$	$1.6989 \times 10^{-10\text{b}}$
Brillouin frequency shift, Ω_B (GHz)	7.882 ^a	7.972 ^b
Brillouin gain bandwidth, $\Delta\nu_B$ (MHz)	23 ^a	20.98 ^b
Fiber loss (dB/m)	0.51 ^a	0.0558 ^b
A_{eff} (μm^2)	9.18 ^a	6.9697 ^b
Time delay (ns/mW/m)	0.07142*	0.123*
Time delay slope efficiency (ns/dB)	1.54*	1.747*
FOM (dB/W m)	22*	30*
Maximum time delay at maximum allowable power (ns)	140* (for 2-m length)	227.3* (for 100-m length)
Maximum Brillouin gain at maximum allowable power (dB)	90.95* (for 2-m length)	85.9* (for 100-m length)

^aRef. [18].

^bRef. [19].

*Simulated.

4. Conclusion

In conclusion, we have simulated two different types of tellurite fibers for tunable slow light generation based on SBS. The fiber characteristics for slow light such as Brillouin threshold power, Brillouin gain, time delay, time delay slop efficiency, and FOMs of both types of fibers have been calculated. Our simulated results indicate that (i) Brillouin gain up to ~91 dB and maximum time delay of 140 ns can be achieved using 2-m long Er-doped tellurite fiber with maximum allowable pump power of 1100 mW and (ii) Brillouin gain up to ~86 dB and maximum time delay of 227 ns can be achieved from 100 m undoped tellurite fiber with 23 mW pump power. The results of this work indicate that the time delay in fibers can be tuned with the pump power, and thereby, tunable slow light features can be obtained from these fibers. These tellurite fibers are expected to have potential applications in realization of compact slow light devices.

Acknowledgements

The authors gratefully acknowledge the (i) initiatives and support toward establishment of the "TIFAC-center of Relevance of Excellence" in Fiber Optics and Optical Communication at Delhi Technological University (Formerly Delhi College of Engineering) Delhi, through the "Mission REACH" program of Technology Vision-2020 of the Government of India and (ii). TUN-IND bilateral research project, Department of Science and Technology, Govt. of India.

References

- [1] Mok, L.T.; Eggleton, B.J. *Nature (news & Views)* **2005**, *433*, 811–812.
- [2] Rawal, S.; Sinha, R.K.; De La Rue, R.M. *J. Nanophoton* **2012**, *6*, 063504-1–063504-12.
- [3] Santagiustina, M.; Eisenstein, G.; Thevenaz, L.; Capmany, J.; Mork, J.; Reithmaier, J.P.; Rossi, A.D.; Sales, S.; Yvind, K.; Combrie, S.; Bourderionnet, J. *IEEE Photon. Soc. Newsletter* **2012**, *26*, 5–12.
- [4] Song, K.Y.; Herraiez, M.G.; Thevenaz, L. *Opt. Exp.* **2005**, *13*, 82–88.
- [5] Sharping, J.; Okawachi, Y.; Gaeta, A. *Opt. Exp.* **2005**, *13*, 6092–6098.
- [6] Ren, L.Y.; Tomita, Y. *J. Opt. Soc. Am. B* **2009**, *26*, 1281–1288.
- [7] Wang, S.H.; Ren, L.Y.; Liu, Y.; Tomita, Y. *Opt. Exp.* **2008**, *16*, 8067–8076.
- [8] Dahan, D.; Eisenstien, G. *Opt. Exp.* **2005**, *13*, 6234–6248.
- [9] Ippen, E.P.; Stolen, R.H. *Appl. Phys. Lett.* **1972**, *21*, 539–540.
- [10] Shumakher, E.; Willinger, A.; Blit, R.; Dahan, D.; Eisenstein, G. *Opt. Exp.* **2006**, *14*, 8540–8545.
- [11] Bigelow, M.S.; Lepeshkin, N.N.; Boyd, R.W. *Science* **2003**, *301*, 200–202.
- [12] Baba, T. *Nature Photon.* **2008**, *2*, 465–473.
- [13] Schneider, T.; Hannover, D.; Junker, M. *J. Lightwave Technol.* **2006**, *24*, 295–304.
- [14] Abedin, K.S. *Opt. Exp.* **2005**, *13*, 10266–10271.
- [15] Florea, C.; Bashkansky, M.; Dutton, Z.; Sanghera, J.; Pureza, P.; I. Aggarwal, *Opt. Exp.* **2006**, *14*, 12063–12070.
- [16] Lee, J.H.; Tanemura, T.; Kikuchi, K.; Nagashima, T.; Hasegawa, T.; Ohara, S.; Sugimoto, N. *Opt. Lett.* **2005**, *30*, 1698–1700.
- [17] Abedin, K.S. *Opt. Exp.* **2006**, *14*, 11766–11772.
- [18] Abedin, K.S.; Lu, G.W.; Miyazaki, T. *Electron. Lett.* **2008**, *44*, 16–17.
- [19] Qin, G.; Sotobayashi, H.; Tsuchiya, M.; Mori, A.; Suzuki, T.; Ohishi, Y. *J. Lightwave Technol.* **2008**, *26*, 492–498.
- [20] Yang, S.; Chen, H.; Qiu, C.; Chen, M.; Chen, M.; Xie, S.; Li, J.; Chen, W. *Opt. Lett.* **2008**, *33*, 95–97.
- [21] Zhu, Z.; Gauthier, D.J.; Okawachi, Y.; Sharping, J.E.; Gaeta, A.L.; Boyd, R.W.; Willner, A.E. *J. Opt. Soc. Am. B* **2005**, *22*, 2378–2384.
- [22] Kalosha, V.P.; Chen, L.; Bao, X. *Opt. Exp.* **2006**, *14*, 12693–12703.
- [23] Hou, S.; Wang, Z.; Shang, Y.; Liu, Y.; Lei, J.; Xu, Y. *PIERS Proceedings*, Suzhou, Sept 12–16, 2011.
- [24] Agrawal, G.P. *Nonlinear Fiber Optics*, 5th ed. Academic; San Diego, CA, 2013.
- [25] Boyd, R.W. *Nonlinear Optics*, 3rd ed. Academic; Orlando, FL, 2008.
- [26] Van Deventer, M.O.; Boot, A.J. *J. Lightwave Technol.* **1994**, *12*, 585–590.
- [27] Cotter, D. *Electron. Lett.* **1982**, *18*, 495–496.
- [28] Zadok, A.; Eyal, A.; Tur, M. *Appl. Opt.* **2011**, *50*, E38–E49.
- [29] Smith, R.G. *Appl. Opt.* **1972**, *11*, 2489–2494.
- [30] Song, K.Y.; Abedin, K.S.; Hotate, K.; Herraiez, M.G.; Thevenaz, L. *Opt. Exp.* **2006**, *14*, 5860–5865.

Stimulated Brillouin Scattering Based Tunable Slow Light in Tellurite Photonic Crystal Fiber

Than Singh Saini, Ajeet Kumar, and Ravindra Kumar Sinha*

TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication,
Department of Applied Physics, Delhi Technological University, Delhi-110 042, INDIA
Email: dr_rk_sinha@yahoo.com

Abstract: Theoretical investigation of tunable slow-light based on stimulated Brillouin scattering in Tellurite photonic crystal fiber is reported. Time-delay upto 88 ns can be achieved using 1 meter long photonic crystal fiber pumped with 140 mW.

OCIS codes: (190.4370) Nonlinear optics, fibers; (290.5900) Scattering, stimulated Brillouin.

1. Introduction

Stimulated Brillouin scattering (SBS) is applicable in the development of slow-light devices, distributed fiber sensors and new laser sources [1-4]. Slow light techniques provide the promising solutions for broadband and tunable time delay for microwave and millimeter wave systems. On-chip all-optical signal processing photonic devices are possible by storing the light and enhancing the nonlinearity. When light interact with the optically resonant medium, its speed and other propagation characteristics can be controlled [5]. Small-core photonic crystal fiber (PCF) is an excellent medium to slow down the light with a moderate pump power. Large-delay slow light based on stimulated Brillouin scattering in a short length of small-core pure silica photonic crystal fiber has been realized by Yang *et al.*[6]. SBS of visible light in silica based small-core photonic crystal fiber pumped with 532 nm has been characterized by Woodward *et al.* [7]. Due to its larger refractive index contrast, PCFs can have tighter confinement of both light and the sound and hence enhance acousto-optic interactions.

In this paper, theoretical investigation of SBS induced slow light in Tellurite photonic crystal fiber has been carried out. The influences of PCF characteristics such as maximum pump power for undistorted output pulse, minimum input signal power (or Brillouin threshold) required to initiate the SBS effect, Brillouin gain and time delay experienced by the pulse on PCF geometry have been calculated numerically. Our simulated results indicate that the time delay experienced by pulse in proposed PCF structure can be tuned with pump power and thereby tunable slow light features can be obtained from proposed PCF.

2. Theory of SBS

Let ω_p is the frequency of continuous pump wave with amplitude A_p and ω_s is the frequency of counter-propagating signal wave with amplitude A_s . The signal frequency is much smaller than frequency of pump wave and $\omega_p - \omega_s = \Omega_B$. The coupled non-linear differential equations for pump and signal waves is given by Eq. (1 and 2)

$$\frac{dA_p}{dz} = -g(\omega_s)A_p^2A_s^2 - \alpha_p A_p \quad (1)$$

$$-\frac{dA_s}{dz} = g(\omega_s)A_p^2A_s^2 - \alpha_s A_s \quad (2)$$

Assuming that the pump wave is undepleted, and $(\alpha_p \approx \alpha_s) = \alpha$, the solution of these equations is given by [8]

$$A_s(0) = A_s(L) \exp [(g(\omega_s)L_{eff} - \alpha L)] \quad (3)$$

Where L and L_{eff} are the real and effective length of the fiber respectively.

The maximum allowable pump power P_{max} (the maximum pump power above which the output pulse get distorted), related to the Brillouin gain coefficient g_B by the following relation [9, 10]

$$P_{max} = 21 \frac{A_{eff}}{K g_B L_{eff}} \quad (4)$$

Where, A_{eff} is the effective cross-sectional area. Once pump power approached to P_{max} then the backscattered wave can be generated from the background noise in the fiber, which leads to serious output pulse distortion. Therefore, the value of pump power must be smaller than the P_{max} .

The minimum value of pump power which is required to initiate the SBS effect is given by

$$P_{min} = \frac{\alpha A_{eff} L}{K g_B L_{eff}} \quad (5)$$

The SBS-induced time delay per unit length can be expressed as [11]

$$\frac{\Delta t_d}{L_{eff}} = \frac{P_p g_0 K}{\Gamma_B} \quad (6)$$

3. Design of Proposed Photonic Crystal Fiber

The cross-sectional view of the proposed PCF design has been illustrated in Fig. 1(a). It consists of an array of air holes arranged in triangular lattice pattern in tellurite glass. The diameter of all air holes is considered identical and represented by 'd' while the centre to centre distance of air holes (i.e. pitch) is taken as 'A'. In our simulation the pitch of the PCF has been kept fixed at 1µm.

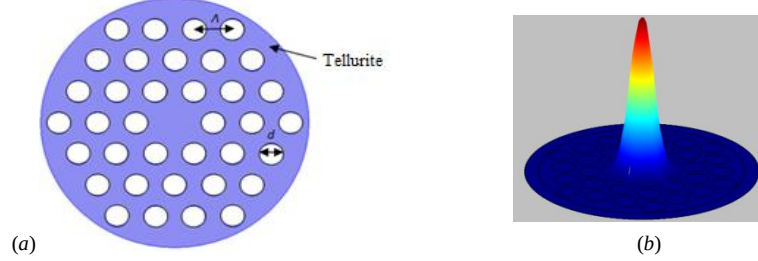


Fig. 1 (a) Transverse cross-section of proposed PCF structure; (b) Electric field distribution in PCF at 1.064 µm when $d/A = 0.8$.

4. Results and Discussion

The effective mode area plays very important role in enhancing nonlinear effects in PCF. In this work the effective mode area of PCF has been controlled by tuning the air filling fraction (which is defined as the ratio of the diameter of air holes to the pitch of the air holes) of the PCF. The influence of air filling fraction (i.e. d/A) on the effective mode area (A_{eff}) of propagating mode in PCF has been illustrated in figure 2(a). The effective mode area of propagating mode decreases on increasing d/A .

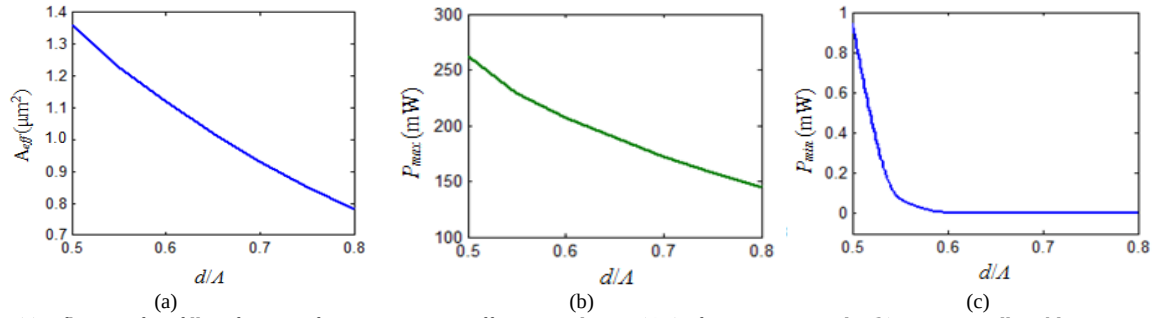


Fig. 2(a) Influence of air filling fraction of PCF structure on effective mode area (A_{eff}) of propagating mode; (b) Maximum allowable pump power as a function of air filling fraction of PCF structure; (c) Minimum required power to initiate SBS as a function of air filling fraction of PCF structure.

When the value of air filling fraction increases the core size of PCF decreases, this restricts on maximum allowable pump power (P_{max}) for undistorted output pulse from the PCF. As shown in Fig. 2 (b), the maximum allowable pump power decreases from 261 mW to 144 mW on increasing the d/A value ranging from 0.5 to 0.8. In most of the applications, reduction of threshold power for SBS is desirable. Simulation results show that the Brillouin threshold power of our proposed PCF is very less. As shown in Fig. 2(c), the threshold power or minimum required pump power (P_{min}) to initiate the SBS decreases on increasing the d/A value of the PCF structure. The Brillouin threshold power is approximately 0.9 mW for d/A value of 0.5 and it decreases rapidly on increasing the value of d/A .

The confinement loss is also a very important feature of photonic crystal fibers. As illustrated in Fig. 3(a) the maximum confinement loss of PCF is 0.33 dB/m at $d/\Lambda = 0.5$ and it decreases on increasing the value of d/Λ . This is due to the fact that on increasing the holes diameter d , the difference between core-cladding refractive index increases leading to higher confinement of light. The influences of pump power on time delay per unit length at different air filling fraction ranging from 0.5 to 0.8 have been illustrated in Fig. 3(b). Time delay in output pulse increases linearly with increasing pump power. Desirable time delay can be obtained by tuning the pump power for a given PCF. For example time delay can be tuned upto ~ 88 ns by pumping between 1mW to 140 mW for one meter long PCF having d/Λ equal to 0.8.

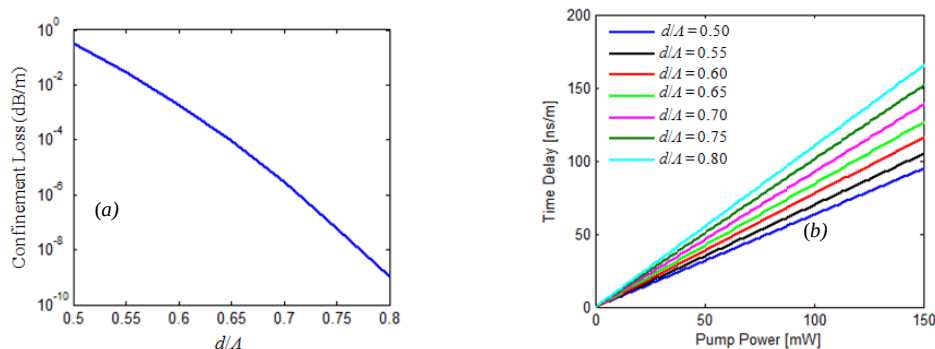


Fig. 3 (a) Confinement loss as a function of air filling fraction of PCF; (b) Time delay per unit length versus pump power at different values of air filling fraction of PCF structure.

5. Conclusion

A Tellurite photonic crystal fiber structure has been designed and analyzed for tunable slow light generation based on stimulated Brillouin scattering. The fiber characteristics for slow light such as Brillouin threshold power and time delay experienced by the pulse for proposed structure has been calculated. Our simulated results indicate that maximum time delay upto 88 ns can be achieved using 1 meter long photonic crystal fiber pumped with 140 mW. In addition to geometrical parameters of proposed photonic crystal fiber, the time delay can be tuned with the pump power and thereby tunable slow light features can be obtained from proposed photonic crystal fiber. The designed photonic crystal fiber is expected to have potential applications in realization of compact slow light devices.

6. Acknowledgement

The authors gratefully acknowledge the (i) Initiatives and support towards establishment of the “TIFAC-Center of Relevance and Excellence” in Fiber Optics and Optical Communication at Delhi Technological University (Formerly Delhi College of Engineering) Delhi, through the “Mission REACH” program of Technology Vision-2020 of the Government of India; and (ii) TUN-IND bilateral research project, Department of Science and Technology, Govt. of India.

7. References

- [1] J. T. Mok and B. J. Eggleton, "Expect more delays," *Nature* **433**, 811-812 (2005).
- [2] S. Rawal, R. K. Sinha, and R. M. De La Rue, "Slow light miniature devices with ultra-flattened dispersion in silicon-on-insulator photonic crystal," *Opt. Exp.* **17**, 13315-13325 (2009).
- [3] S. Rawal, R. K. Sinha and R. M. De La Rue, "Silicon-on-insulator photonic crystal miniature devices with slow light enhanced third-order nonlinearities," *J. Nanophoton.* **6**, 063504 (2012).
- [4] M. Santagiustina, G. Eisenstein, L. Thevenaz, J. Capmany, J. Mork, J.P. Reithmaier, A. D. Rossi, S. Sales, K. Yvind, S. Combrie and J. Bourderionnet, "Slow light devices and their applications to microwaves and photonics," *IEEE Photon. Soc. Newsletter*, 5-12 (2012).
- [5] K. Y. Song, M. Gonzalez Herraiz, and L. Thevenaz, "Observation of pulse delaying and advancement in optical fibers using stimulated Brillouin scattering," *Opt. Exp.* **13**, 82-88 (2005).
- [6] S. Yang, H. Chen, C. Qiu, M. Chen, M. Chen, S. Xie, J. Li and W. Chen, "Slow-light delay enhancement in small core pure silica photonic crystal fiber based on Brillouin scattering," *Opt. Lett.* **33**, 95-97 (2008).
- [7] R. I. Woodward, E. J. R. Kelleher, S. V. Popov and J. R. Taylor, "Stimulated Brillouin scattering of visible light in small-core photonic crystal fibers," *Opt. Lett.* **39**, 2330-2333 (2014).
- [8] R. W. Boyd, *Nonlinear Optics*, III ed. Orlando: (Academic, 2008).
- [9] E. P. Ippen and R. H. Stolen, "Stimulated Brillouin scattering in optical fibers," *Appl. Phys. Lett.* **21**, 539-541 (1972).
- [10] R. G. Smith, "Optical power handling capacity of low loss optical fibers as determined by stimulated Raman and Brillouin scattering," *Appl. Opt.* **11**, 2489-2494 (1972).
- [11] A. Zadok, A. Eyal and M. Tur, "Stimulated Brillouin scattering slow light in optical fibers," *Appl. Opt.* **50**, E38-E49 (2011).



ISSN NO. 2320-5407

Journal homepage: <http://www.journalijar.com>

INTERNATIONAL JOURNAL
OF ADVANCED RESEARCH

RESEARCH ARTICLE

Study of Flexural, Tensile, Impact properties and Morphology of Potato Starch/Polypropylene blends.

A.P. Gupta and Aftab Alam*

Department of Applied Chemistry and Polymer Technology, Delhi Technological University, Delhi 11042 (India)

Manuscript Info

Manuscript History:

Received: 15 September 2014

Final Accepted: 25 October 2014

Published Online: November 2014

Key words:

Thermoplastic Starch, polypropylene powder, PP-g-MA, Tensile Strength, Flexural Strength, Impact Strength, SEM, FTIR, MFI

*Corresponding Author e-mail: aftab1238@gmail.com

A.P. Gupta

Abstract

The effect of potato starch contents on mechanical and morphological properties of polypropylene starch blends have been investigated in this work. The maleic anhydride grafted PP (PP-g-MA) was used as coupling agent to improve the adhesion between starch and polymer matrix in melt blending as interfacial strength between reinforcement and polymeric matrix has direct effect on the properties of blend. For this purpose the content of starch varied from 15 to 30% and coupling agent (i.e. PP-g-MA) was taken 10%. The blends were compounded on co-rotated twin screw extruder. Then, mechanical properties such as tensile, flexural and impact were evaluated. It was found that the tensile and impact strength remained almost constant, while flexural strength of the blend improved by 45%, and percentage strain decreased with increasing starch content in the blend. The scanning electron micrograph (SEM) supports the finding as it shows uniform distribution of starch particle in the matrix. The chemical structure was analyzed by FTIR. The Melt flow Index (MFI) of blends also supported the uniform distribution of starch within the polymeric matrix.

Copy Right, IJAR, 2014,. All rights reserved

Introduction

Polypropylene (PP) was the first synthetic stereo-regular polymer to achieve industrial importance. Being lightest and highly crystalline, polypropylene exhibits high stiffness, hardness and tensile strength. Hence, due to high strength to weight ratio, it is widely used in automobile, electrical and household equipment, furniture, and packaging applications¹. The PP is derived from non-renewable petroleum resources and it is non degradable in nature, which may cause environmental disaster without proper disposal². Hence several government and non-government communities are now more sensitive to the impact of discarded plastic on the environment, including deleterious effects on wildlife and on the aesthetic qualities of cities and forests. The biodegradable polymers opened the way for new consideration of waste management strategies since these materials are designed to degrade under environmental conditions or in municipal and industrial biological waste treatment facilities.

The blend of natural material like starch and synthetic polymers offers a practical approach to this environmental threat. The starch may contribute towards the development of such biodegradable polymers because its properties are quite similar to the petrochemical based polymers³⁻⁵. Many attempts have used renewable and sustainable materials to prepare biodegradable PP blends to reduce its environment impact⁶. Starch is an important material of choice as approximately 75% of all organic material on earth is present in the form of polysaccharides. The main issue around its use, have been that it is hydrophilic in nature, brittleness in the absence of suitable plasticizers and deterioration of mechanical properties upon exposure to environmental conditions⁷⁻¹⁰. Since, PP and Starch did not react chemically, the maleic anhydride grafted PP (PP-g-MA) used as coupling agent to improve the compatibility between polymer matrix and starch by formation of hydrogen bonding¹¹⁻¹⁴.

Hence, authors selected potato starch, PP and PP-g-MA to prepare biodegradable blend of PP and starch. The content of coupling agent PP-g-MA has been maintained to 10%. In this present work 15 to 30% potato starch

treated with glycerol, have been extruded with co-rotated twin screw extruder. The granule so obtained was evaluated for Tensile, Flexural, and Impact strength. The fractured surfaces of the blend were investigated by scanning electron microscope (SEM) providing information for the adhesion of starch/polypropylene.

1. EXPERIMENTAL

2.1 Materials:

Polypropylene (Homopolymer) powder (Melt Flow Index: 35 g/10 min. at 230°C/2.16 Kg load, Grade PP H350FG) was procured from M/s Reliance Industries. The native potato starch (M.W. 162.14) was procured from Loba Chemie, Mumbai. The glycerol was obtained from M/s Fisher Scientific (99.5% pure). The maleic anhydride grafted PP (PP-g-MA) grade Epolene G-3003, was procured from Eastman Chemical Products Co. USA.

2.2 Preparation of PP/Starch Blends:

The potato starch was dried to 1-2% moisture content using a laboratory oven at 70 °C for six hours and then stored in sealed plastic bags to avoid moisture. The dried starch after proper cooling was mixed with 25% glycerol. The stirring continued till the uniform suspension was obtained. The prepared suspension was left for overnight at ambient conditions of 30°C and 60% RH. After allowing swelling action for 24 hours suspension was rotated in high speed mixer at 3000 rpm for approximately 5 min. so that it gets converted into powder form by heat and shear. The thermoplastic starch powder (TPS) so obtained was mixed with polypropylene powder with the experimental formulation listed in Table No.1.

Both PP in powder form and Starch (TPS powder) were mixed homogeneously in high speed mixer for 2 min. and blends were prepared by melt blending with the help of co-rotating twin screw extruder. The temperature profile of twin screw extruder was 170°C to 190°C and melt temperature was 175°C, screw rpm kept at 275 and melt pressure was maintained around 25MPa during extrusion.

2.3 CHARACTERIZATION:

2.3.1 Fourier Transmission Infrared Spectroscopy:

Fourier Transmission Infrared Spectroscopy (FTIR) spectra of injection molded sample have been recorded by employing M/s Agilent Technologies model carry 630. The samples have been scanned in frequency region 4000 to 600 cm⁻¹ with resolution 4 cm⁻¹.

2.3.2 Tensile Properties:

The tensile properties for the blends have been conducted using Instron universal testing machine model 3369 according to ASTM D 638. The specimen was prepared by injection molding and dimensions were according to ASTM standard. The crosshead speed of 50mm/min at 23°C was taken for testing of specimen. The specimen samples were conditioned at 23°C and 50% RH for 24 hours. Five specimens were used to obtain the average value of the tensile properties.

2.3.3 Flexural Properties:

The flexural properties for the blend have been conducted by using Instron universal testing machine model 3369 according to ASTM D 790. The specimen was prepared by injection molding and dimensions were according to ASTM standard. The crosshead speed was 5mm/min was taken for testing of specimen. The samples were conditioned at 23°C and 50% RH for 24 hours. Five samples were used to obtain the average value of the flexural properties.

2.3.4 Impact Strength:

The Izod Impact strength for blends have been conducted by using Ceast Spa impact tester model Resil Impactor according to ASTM D 256. The specimen was pre-conditioned at 23°C and 50% RH for 40 hours. The energy of the hammer was 2.75 J.

2.3.5 Melt Flow Index:

The Melt Flow Index for the blends have been conducted according to ASTM D 1238. The test conducted at 230°C at 2.16 kg load. Five tests were done to obtain the average value of the melt flow.

2.3.6 Morphology Test:

The scanning electron microscopy (SEM) Hitachi model S3700N, was used to evaluate the sample microstructure. The samples have been first dried in an oven to remove the moisture and then sputter coated with a thin layer of gold to avoid electrical charging.

2. RESULTS AND DISCUSSION:

SAMPLE	Polypropylene (Homopolymer Grade H350FG)	Starch (TPS)	PP-g-MA (Epolene G-3003)
PP	100 %	0 %	0 %
PPST-A	75 %	15 %	10 %
PPST-B	70 %	20 %	10 %
PPST-C	65 %	25 %	10 %
PPST-D	60 %	30 %	10 %

3.

4. Table No.1: Formulation of PP/Starch Blends

3.1 Fourier Transmission Infrared Spectroscopy(FTIR):

Figure No.1 depicts FTIR spectrum of PP and its blends with different compositions of starch. FTIR spectroscopy was employed to prove the compatibilization among PP-g-MA and starch during the melt blending process on polypropylene matrix. Several strong peaks were observed in $3100\text{--}3600\text{ cm}^{-1}$ (O-H stretching from starch), 1749 cm^{-1} (C=O stretching from starch), 1653 cm^{-1} (intramolecular hydrogen bonding), 1255 cm^{-1} (-O-C(O)- and C-OH stretching), 1151 cm^{-1} (C-O-C stretching), 1025 cm^{-1} (C-O stretch vibration), which are characteristic peaks of starch^{11,15}. The peak observed near to 1740 cm^{-1} evidencing strongly that anhydride group is present in PP backbone¹¹. The CH_3 bands at 2950 cm^{-1} , 2868 cm^{-1} and 1375 cm^{-1} , the CH_2 bands at 2917 cm^{-1} , 2837 cm^{-1} and 1457 cm^{-1} , and three isotactic peaks at 1167 cm^{-1} , 997 cm^{-1} and 973 cm^{-1} confirmed the presence of the polypropylene component in all the blends¹⁶.

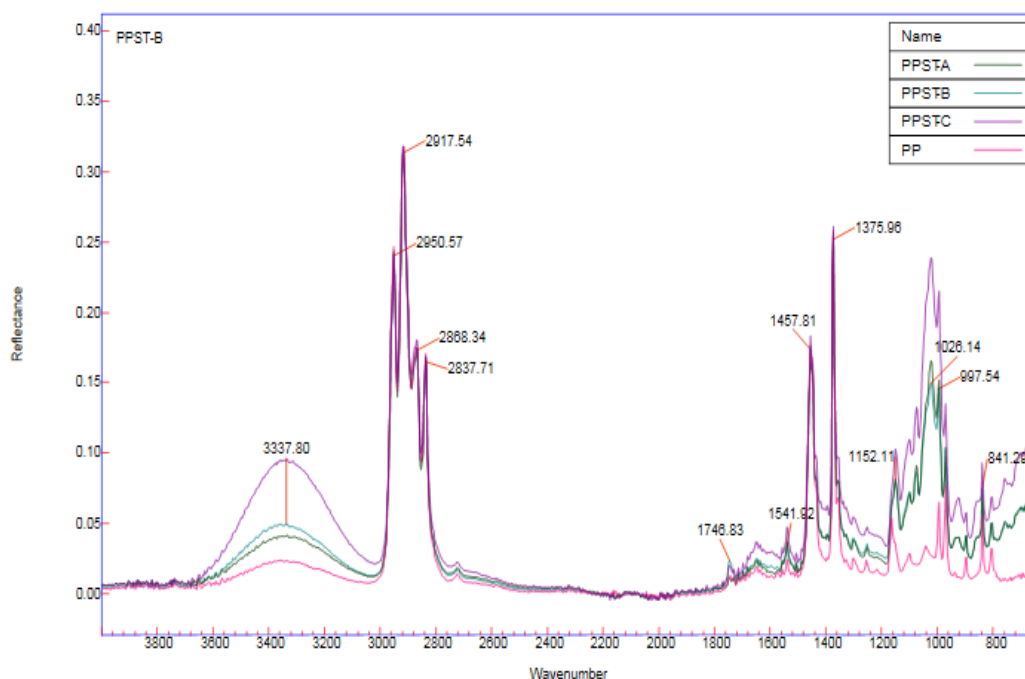


Figure No.1: FTIR Spectrum of Samples PP, PPST-A, PPST-B and PPST-C.

3.2 Mechanical Properties:

As the interfacial bonding is greatly improved because of the presence of PP-g-MA and plasticization of starch, a considerable enhancement in the mechanical properties of the PP matrix is expected, which could be verified by the test results. The addition of starch improved the tensile strength and impact strength remained almost same provides a scope of development. The flexural strength has increased by approximately 45% (Figure no. 2). The %age strain has decreased by 80%. Because % strain is related to the ductility of a material, the presence of PP-g-MA sharply increases the stiffness (flexural strength), thus decreases ductility of the blends.

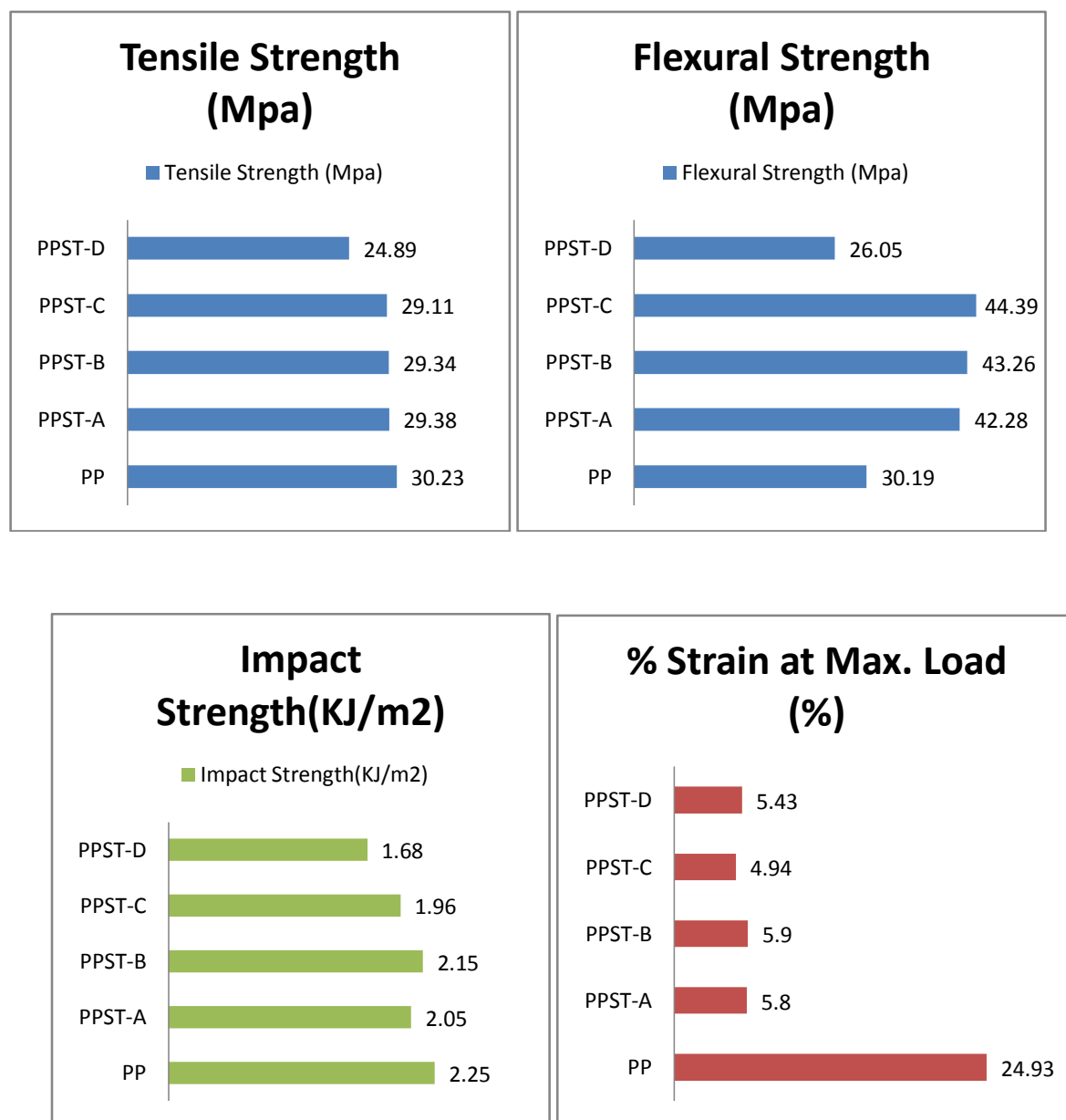


Figure no. 2: The tensile, flexural, impact and % strain of PP/Starch blend.

3.3 Melt Flow Index (MFI):

The melt flow index (MFI) is used as a means of measuring of uniformity of the flow rate of the thermoplastic material. It is the most accepted and widely used practical method to relate the molecular weight and viscosity of the polymer in industry. The MFI value of different compositions of PP and starch indicates the uniform distribution of starch in polypropylene matrix. The PPST-D composition having 30% loading of starch in polypropylene matrix have low value, which exhibits poor dispersion/compatibilization of starch in blend (Figure No. 3), and same has also been reflected by other tests.

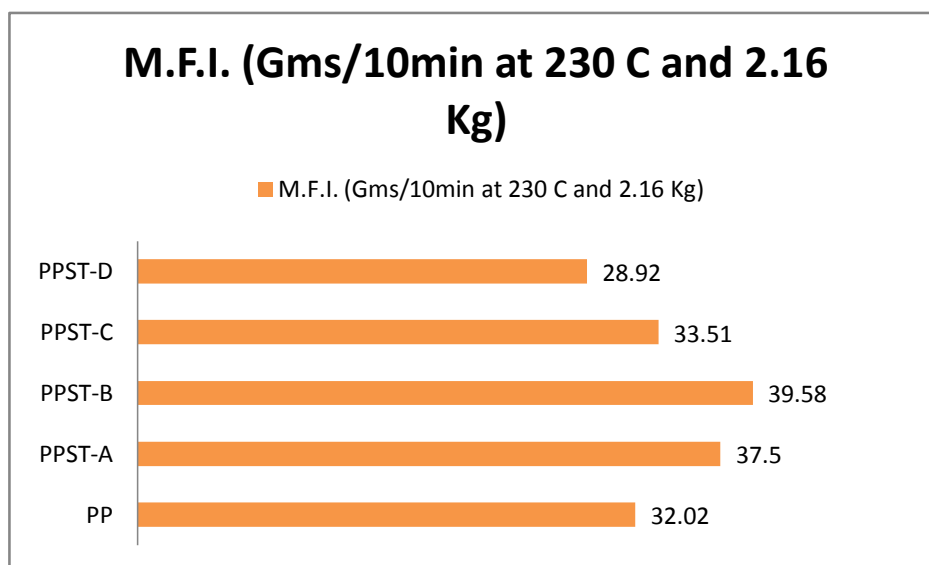
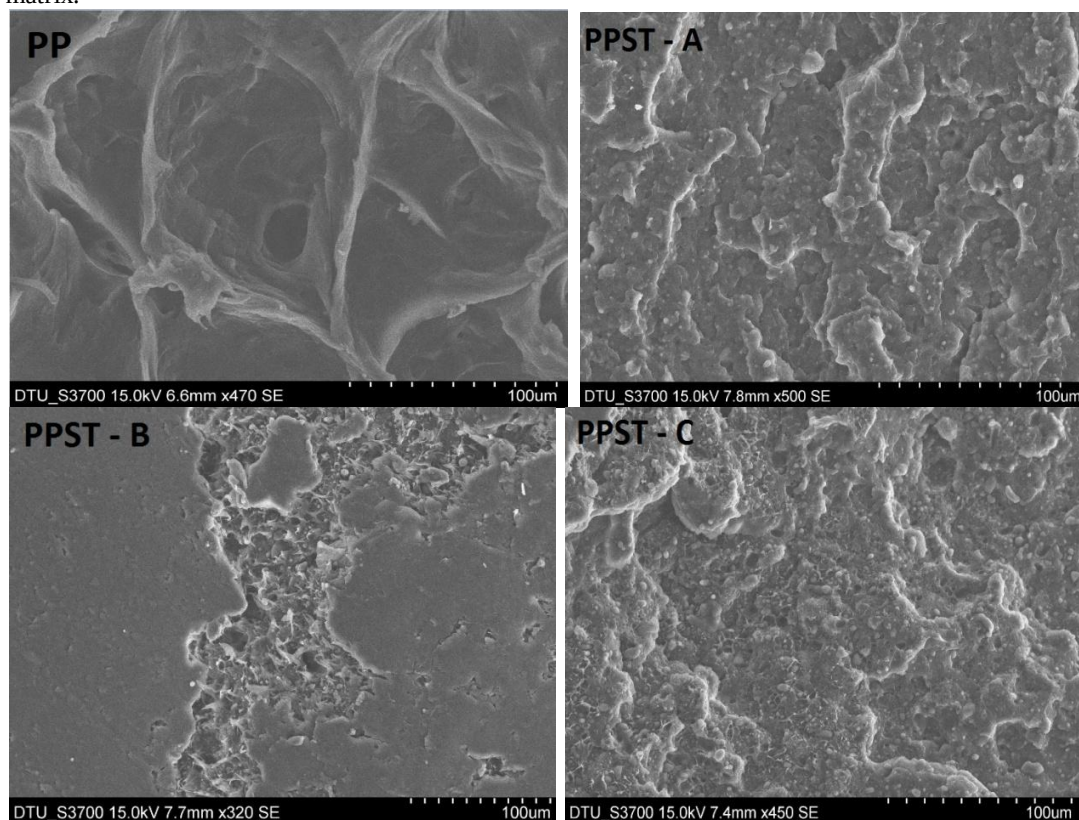


Figure No 3: Melt Flow Index of PP/Starch blends.

3.4 Morphology Analysis:

The SEM micrographs of PP/Starch blend of various compositions are presented in Figure No 4. These micrographs revealed that thermoplasticized starch granules are being continuously and uniformly mixed throughout the PP matrix and form a homogeneous phase. The SEM micrograph ensured that presence of polar group enhances adhesion between two phases (hydroxyl group of starch and polar group of PP). Therefore, the PP-g-MA played a vital role in reducing the interfacial energy and promoting the interfacial adhesion between the potato starch and PP matrix.



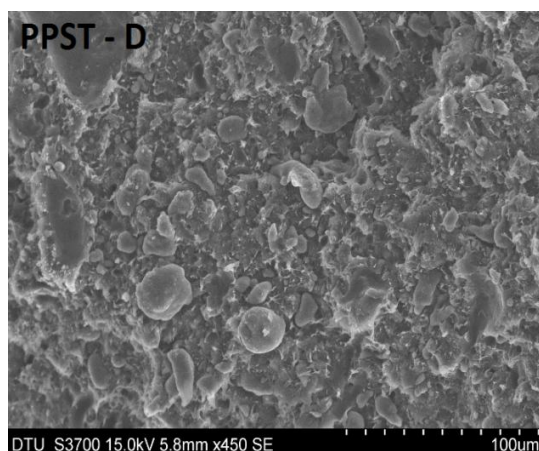


Figure No.4: SEM images of the fracture surfaces of PP/Starch blends at magnification with a scale of 100 um.

Conclusion:

PP/Starch blending using PP-g-MA as the coupling agent has significantly improved mechanical properties and improved the adhesions between starch and polypropylene matrix. The adhesion of 25% starch in PP, increased the flexural strength by 45%, while tensile strength and impact strength remain same is a significant discovery. The FTIR, SEM and Melt flow Index of the blends also indicated improvement in interfacial adhesion between starch and polymer. This discovery may prove a milestone towards developing bio-based polymeric materials in future.

Reference:

1. Karger-Kocsis, J.; Polypropylene as AZ reference; Chapman & Hall; London 1999.
2. Harding, K.G.; Dennis J.S.; Von Blotnitz, H.; Harrison, S.T.I., Environmental analysis of plastic production processes: Comparing petroleum-based polypropylene and polyethylene with biologically-based poly-B-hydroxybutric acid using life cycle analysis. *J. Biotechnol.* 2007, 13, 57
3. Huang J.C., Shetty A.S., and Wang S.W.; Biodegradable Plastics: A review *Adv. Polym. Tech.*, 1990, 10, 23-30
4. Dbi Y. and Fukuda K.; Biodegradable Plastics and Polymers. Elsevier Science B.V., 1994, 601-608.
5. Potts J.E.; Environmentally Degradable Plastics. In: Kirk-Othmer Encyclopedia of Chemical Technology, 3rd ed., Supp. Vol, John-Wiley, New York, 1981, 638-668
6. Mohanty, A.K.; Misra, M.; Drzal, L.T.; Sustainable bicomposites from renewable resources: Opportunities and challenges in the green material world. *J. Polym Environ.* 2002, 10, 19
7. Gupta, A.P.; and Verma, D.K.; Carboxymethyl guar gum- silver nanocomposite: green synthesis, characterization and an optical sensor for ammonia detection, *Advance in natural sciences: nanoscience and nanotechnology*, IOP, Publishing, Vietnam Academy of science and technology, 2014, 5, 035018 (8pp).
8. Griffin G.J.L.; Starch polymer blends, *Polym. Degrad. Stab.* 1994, 45, 241
9. Sathya Kalambur and Syed S.H. Rizvi "Overview of starch based plastic blends" *Journal of Plastics Films and Sheeting* 2006, 22, 1, 39-58.
10. Ramis X; Cadenato, A.; Safla, J.M.; Morancho, J.M.; Valles, A.; CONTAT, I.; Ribes, A. Thermal degradation of polypropylene/Starch-based materials with enhanced biodegradability. *Polym. Degrad. Stab.* 2004, 86, 483
11. Gupta A.P.; Vijay K., Sharma M., Formulation and Characterization of Biodegradable Packaging Film Derived from Potato Starch & LDPE Grafted with Maleic Anhydride-LDPE composition. *J. Polym. Environ.* 2010, 18, 484-491
12. Willet J.L.; Mechanical Properties of LDPE/granular starch composite. *J. Appl. Polym. Sci.*, 1994, 54, 1485-1495
13. Kim M.; Evaluation of Degradability of Hydroxyl propylated potato starch/polyethylene blend films. *Carbo. Polym.*, 2003, 54, 173-181.
14. Lim S.T., and Jane J.L.; Effect of Starch Granule size on Physical Properties of Starch filled polyethylene film. *Biotech.* 1992, 8, 51-57.
15. Wang, S.J., Yu, J.G., Yu, J.L.; Preparation and Characterization of compatible thermoplastic starch/polyethylene blend. *Polym. Degrad. Stab.* 2005, 87, 395
16. Devasahayam, S., Sahajwalla, V., and Sng M.; Investigation into Failure in Mining Wire Ropes- Effect of Crystallinity. *J of Org. Polym. Materials.* 2013, 3, 34-40



Test Suite Optimization using Mutated Artificial Bee Colony

Ruchika Malhotra¹ and Manju Khari²

¹ Delhi Technological University,
Department of Software Engineering, Delhi, India
ruchikamalhotra2004@yahoo.com

² Ambedkar Institute of Advanced Communication Technology and Research,
Department of Computer Science and Engineering, Delhi, India
Email: manjukhari@aiactr.ac.in

Abstract— Software test suite optimization is one of the most important issue in software testing as testing consumes a lot of time in executing redundant test cases. In this paper, we have proposed and implemented a new approach for test suite optimization, namely, Mutated Artificial Bee Colony. Artificial Bee colony algorithm combines local search carried out by employed and onlooker bees with global search managed by scouts and gives optimal results. To further improve the global search capabilities of ABC algorithm, Mutation function of Genetic Algorithm is permuted with Onlooker bee and Scout bee, giving four different approaches for test suite optimization that are, Simple ABC, ABC with mutation at onlooker bee, ABC with mutation at Scout bee and ABC with mutation at both onlooker bee and scout bee. All approaches will be compared on the basis of runtime and number of iterations. Based on the experimental results, it has been verified and validates the proposed algorithm. The proposed algorithm would be beneficial for the software testers for selecting a minimal test suite for testing.

Index Terms— Test suite optimization, Artificial Bee Algorithm, Genetic Algorithm, Mutated Artificial Bee Colony, Mutation function.

I. INTRODUCTION

Testing is an integral part in software development process. It is broadly deployed in every phase of the software development life cycle. Software testing process typically consumes at least 50% of the total cost involved in software development. Testing basically comprises of test case generation and optimization of the generated test cases. Executing all the generated test cases is expensive in terms of cost and time because of the existence of redundant test cases. The solution to this issue is executing an optimal test suite covering each functionality module atleast once. Tests optimization is highly dependent on the testing team's efficiency and is not suitable for large programs; on the other hand, automated test optimization mitigates the risk of test-maintenance failure, improves efficiency and hence reduces cost and time consumption. Automated test data optimization is an important problem. Meta heuristic search techniques offer much promise in response to optimization problems. Meta heuristic search techniques are high-level frameworks,

which utilize heuristics to seek solutions for combinatorial problems at a reasonable computational cost [1, 2, 3, 4, 5, 6, 8, 10]. Many existing meta heuristic optimization approaches have been applied for efficient test case selection and optimization such as Hill Climbing (HC), Simulated Annealing (SA), Genetic Algorithm (GA), Ant Colony Optimization (ACO), Memetic Algorithm (MA), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC)[12,14,15,17], Shuffled Frog Leaping (SFL) and Tabu Search (TS).

In this paper we have proposed a new algorithm, Mutated Artificial Bee Colony (MABC), which is inspired by the biological behaviour of honey bees and Darwin's theory of natural selection for test suite optimization. The proposed algorithm works on the principal of ABC with the application of mutation function of Genetic Algorithm for optimization. This paper is divided in seven sections. Second section covers the review of literature and all the previous work performed in the area of heuristics and test suite optimization. In the third section, key research concepts are mentioned. Section four includes proposed algorithm MABC for optimization of test cases. In section five, a case study is implemented on the principle of MABC. In section six, empirical results and their analysis has been described to validate the results. At the end, we conclude in section seven that MABC algorithm is the best for optimization of the test suite.

II. RELATED WORK

An extensive research work has been carried out in the field of test suite optimization. Different researchers have used different algorithms while optimizing test data/suites. In 1995, Harmen worked on GA and concluded that GA is better than Random Testing [9]. In 2004, McMinn worked on GA, HC and SA and concluded that GA performs the best [1]. In 2005, Elbeltagia et al. worked on GA, MA, ACO and PSO algorithms finally concluded that PSO algorithm performs the best [3]. In 2005, Girgis worked on GA and concluded that GA is better than Random Testing. In 2006, Kanmani et al. worked on GA, HC and SA then concluded Evolutionary algorithms are better in local search [8]. In 2006, Hassan et al. worked on GA and PSO and concluded that PSO algorithm performs the best [11]. In 2007, Karaboga et al. worked on GA, ABC and PSO algorithm and concluded that ABC performs the best [13]. In 2007, Tapkan et al. worked on GA, ABC and SA and concluded that ABC performs the best [7]. In 2008, Garg worked on GA, MA and concluded that MA performs the best [4]. In 2009, Karaboga et al. worked on GA, ABC and PSO and concluded that ABC performs the best [12]. In 2009, Jeya Mala et al. worked on GA and ABC and concluded that parallel ABC performs the best [6]. In 2009, Singh worked on GA, ABC, and PSO and concluded that ABC performs the best [9]. In 2010, Jeya Mala et al. worked on GA, ABC and concluded that ABC performs the best [5]. In 2011, Maragathavalli et al. GA, HC, SA, ACO and concluded Genetic is better in local search. In 2011, [10] Malhotra et al. worked on GA and concluded that GA is better than Path Testing [2]. In 2012, Karabulut et al. works on discrete ABC algorithm for TSP [13]. Shanthia et al. proposed a collaborative ABC clustering using SPNN [14]. In 2013, Cuevas et al. introduces a new algorithm based on ABC optimization which is proposed to reduce the number of search locations in Matching (BM) motion estimation [15].

As far as Literature survey is concerned, Karabulut et al. has described the discrete artificial bee colony (DABC) algorithm for TSP. He uses the time window in order to reduce the total travel cost of the problem. The algorithm is used in both scheduling and logistic problems. It relies on the destruction and construction phases of iterated greedy algorithm to generate neighboring food sources in a framework of ABC algorithm. It also considers classical 1-opt local search algorithm to further enhance the solution quality [13]. Subotic has constrained the optimization problem with more onlooker bees and the candidate solution in onlooker bee phase is formed using three solutions [16]. Karaboga et al. has explained the usage of ABC on symbolic regression which is very important practical problem. Symbolic regression is a process of obtaining a mathematical model using given finite sampling of values of independent variables and associated values of dependent variables [12]. Alam et al. has explained the ABC algorithm with improved explorations for Numerical Function Optimization. It describes a novel algorithm that modifies both the selection and perturbation operations of the basic ABC algorithm in an explorative way. ABC-IX employs a probabilistic, explorative selection scheme based on simulated annealing which can accept both better and worse candidate solutions. It also maintains a self-adaptive perturbation rate, separately for each candidate solution, to promote more explorations [17]. Shanthia et al. has explained the collaborative ABC clustering using SPNN. It mainly deals with methods for clustering of continuous signals. In this process, centers of classes are determined by neural network with process input. In conclusion, C-ABC has great explorative search features and better convergence [14]. Abu-Mouti et al. presents an overview of the literature employing the ABC algorithm in their solution approach. Also provide a concise survey of newest heuristic methods [18]. Qiao et al. introduces ABC to the optimization of the multi-objective on construction project time-cost-quality and

compares the results of ABC with the results of GA or PSO. He proves effectiveness of ABC [19]. Cuevas et al. proposes a new algorithm based on ABC optimization which reduces the number of search locations in BM motion estimation. Also, BM motion estimation plays a very important role in video conferencing [15]. Kiran et al. presents a hybridization of PDO and ABC approaches based on recombination procedure. The information flow between particle swarm and bee colony helps increase global and local search abilities of the hybrid approach which is referred to hybrid approach on PSO and ABC [20]. At the end, Gao et al. has applied Gaussian ABC algorithm approach to Loney's Solenoid Benchmark Problem [21]. Table I shows the various meta heuristic algorithms [3, 4] used for optimization on few theoretical parameters like ease of use, scalability, global/local search ability etc.

TABLE I: COMPARISON OF META HEURISTIC ALGORITHM

Algorithm	Advantages	Disadvantages	Ease of use	Local/Global optimum	Scalable
GA [1,2,3,4,8, 10]	1. Do not break easily even if the inputs change slightly 2. Answer gets better with time.	1. Gives slow results 2. Weak local search ability	Very easy to understand and it practically does not demand the knowledge of mathematics.	Global	Yes
ABC [5,6, 12,15,17]	Good quality solutions of the difficult optimization problems can be discovered.	Not yet discovered	Easy	Global	Yes
HC [3,8]	1. Simple 2. Gives fast results	1.Leads to local optimal 2.Highly dependent on starting solution.	Easy	Local	No
SA [3,8]	1. Less dependent on the starting solution 2. Less restrictive	If cooling is too rapid search, not enough of the search space will be explored, hence stuck in local optima.	Medium	Global	Yes
MA [3, 4]	Strong local search ability	Gives slow results	Difficult	Global	Yes
ACO [1,3,8]	1.Retains memory of entire colony instead of previous generation only 2.Less affected by poor initial solutions	Convergence is guaranteed, but time to convergence uncertain	Easy	Global	Yes
PSO [1,3,8,11]	1.Easy to perform 2.Few parameters to adjust 3.Efficient in global search	1.Slow convergence in refined search stage 2. Weak local search ability	Medium	Global	Yes

III. KEY RESEARCH CONCEPTS

A. Cyclomatic Complexity

Cyclomatic complexity is a metric used to measure the complexity of software. It was developed by McCabe in 1976 and is used to indicate the complexity of a program. It directly measures the number of linearly independent paths through a program's source code. Cyclomatic complexity is computed using the control flow graph of the program. From the given CFG (control flow graph), DD (decision to decision) graph is made. DD graph is path of execution between to decisions [2, 3,8].

B. Genetic Algorithm

GA is an example of evolutionary algorithm [1, 2, 3, 10]. The idea is based on the randomly initialized population of individuals. Each individual is a potential candidate of the problem. In GA, Fitness function is used to evaluate the adequacy and quality of individual. It includes three basic steps: selection, crossover, mutation.

C. Artificial Bee Colony

ABC works on foraging process of bees and is a population based algorithm. In ABC, food source represents the possible solution of problem [5, 6, 12, 14, 15, 17]. It includes three types of bees which are: Employee bee, Onlooker bee and Scout bee. ABC is based on the randomly initialized population of the solution. On the basis of fitness value, the best optimal solution is selected.

IV. PROPOSED ALGORITHM- MUTATED ARTIFICIAL BEE COLONY

In this section, the proposed algorithm **Mutated Artificial Bee Colony** is presented. We optimize software test data using a hybrid of ABC and GA. In this algorithm, Mutation function (sub part of GA) is applied at both onlooker bee and scout bee in ABC. Initially a random population of test cases is generated which is equal to or greater than the cyclomatic complexity of the software under test. Objective values and fitness values of all the test cases are calculated and the test cases covering different paths of the software are memorized. The algorithm then runs till all the independent paths are covered or number of iterations is less than MABC. Test cases then enter the employed bee phase in which neighbour for each test case is found and compared to the original test case. If the new test case has objective value greater than the old test case, it replaces the original test case and is memorized, else trial of the test case is incremented. Probabilities of all the test cases are then calculated. In the onlooker bee phase a random number is generated for each test case. If the probability of the test case is greater than the random number generated, then neighbour of the test case is evaluated, else, mutation function is applied to the test case giving an entirely new test case. Again, the better test case is memorized and trial is incremented if no better test case is found. If trial of any test case exceeds the max trial limit (set by the programmer), the test case enters the scout bee phase and mutation operation is applied to the test case. Now, if the mutated test case covers any new path then it is memorized and trial is set to zero. Else, random test case is generated to replace the old test case. Figure 1 depicts the proposed MABC algorithm. We applied the following algorithm for our test case optimization problem. r is a random number in the range $[0, 1]$.

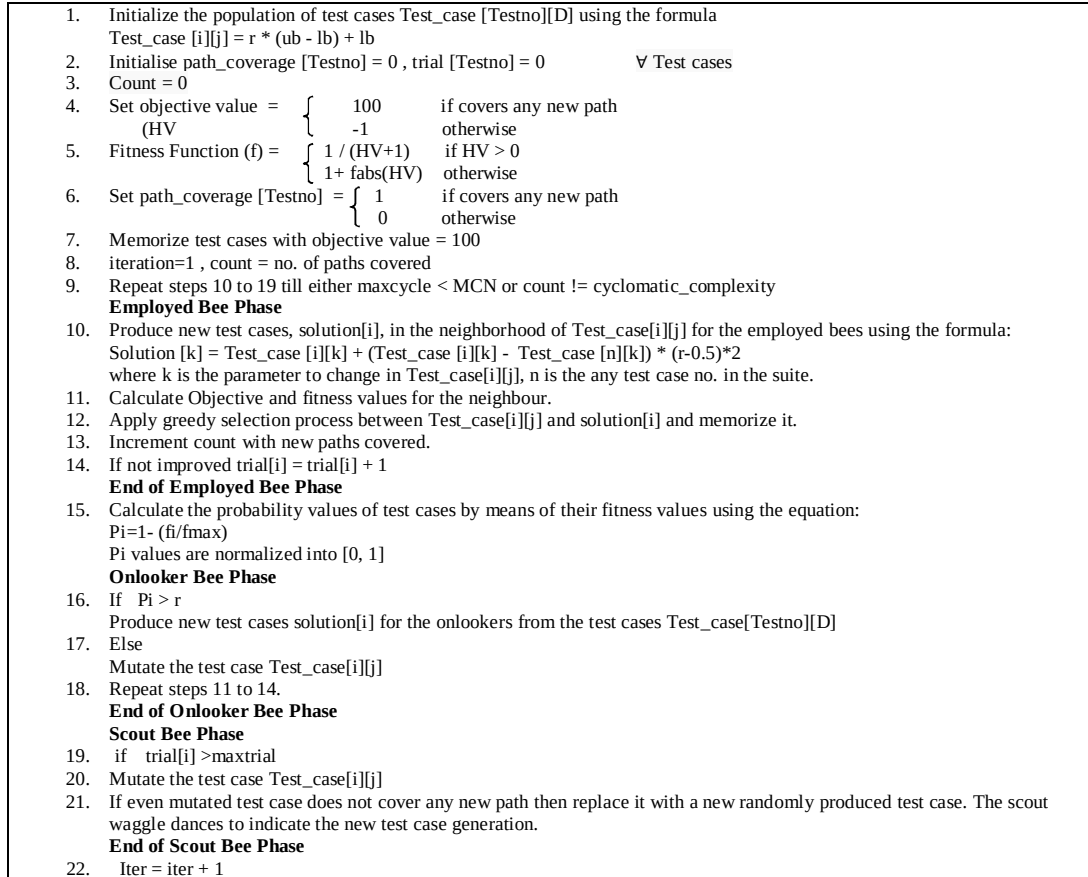


Figure 1: Proposed Algorithm of Mutated Artificial Bee Colony

V. MABC – IMPLEMENTATION

CASE STUDY: COMPARISON OF THREE NUMBERS

In this section, we present an implemented sample program to compare three numbers. Program take three inputs a, b and c. The program code is written in C++ language and uses 48 Lines of code (LOC). There are seven independent paths in the given problem. We have an initial set of randomly chosen test cases.

The table II presents generated No of parameters, Population size i.e. 10, Max cycle i.e. 20, Max trial i.e. 5. Initial Population of test cases is generated using the formula $\text{Test_case}[i][j] = r * (\text{ub} - \text{lb}) + \text{lb}$. Here, upper bound (ub) and lower bound (lb) for the parameters a, b and c are 10 and -10 respectively. The population hence generated and calculated objective values and fitness values for all the test cases produced are shown in Table II. The initial population generated covers four independent paths and hence the value of variable count becomes four. The paths covered along with the respected test case are memorized and shown in Table III.

TABLE II: INITIAL POPULATION, OBJECTIVE VALUE, AND FITNESS VALUE

Testcase	Parameter			Objective Value (HV)	Fitness Value (fi)
	a	b	c		
0	-9	-5	7	100	0.0099
1	0	-4	9	100	0.0099
2	4	7	4	100	0.0099
3	-3	-2	-5	-1	2
4	5	9	-9	-1	2
5	1	1	8	-1	2
6	0	-2	4	-1	2
7	2	-4	-7	100	0.0099
8	9	2	5	-1	2
9	-5	-9	9	-1	2

TABLE III: PATHS COVERED AFTER INITIAL POPULATION

Path _No	Test_case		
	a	b	c
1	2	-4	-7
2	0	-4	9
3	-		
4	4	7	4
5	-9	-5	7
6	-		
7	-		

As $\text{count} < 7$ and max cycle limit is not exceeded, first iteration starts with the employed bee phase calculating the neighbours for all the test cases using the formula, $\text{solution}[k] = \text{Test_case}[i][k] + (\text{Test_case}[i][k] - \text{Test_case}[n][k]) * (r-0.5)*2$, where k is the parameter to change in $\text{Test_case}[i][j]$ and n is the any test case number in the test suite. Table IV shows the corresponding neighbours for all the test cases and depicts if the neighbour is improved or not and shows the probabilities for the test cases are calculated using the formula i.e. $P_i = 1 - (f_i/f_{\max})$.

In the onlooker bee phase, appropriate function to be performed for each test case is selected depending upon its probability value given in Table V. Table marks the function performed by a test case and also shows the changed test case.

The improved test case number 2 i.e. [4 7 4] is calculated as shown below:

$$K = 3 * 0.6 = 1.8 = 1$$

$$N = 10 * 0.81 = 8$$

$$\text{Sol}[k] = 7 + (7 - 2) * (0.2 - 0.5) * 2 = 4$$

$$\text{Neighbour} = [4 \quad 4 \quad 4]$$

TABLE IV: EMPLOYED BEE PHASE FOR ITERATION 1 WITH ITS PROBABILITY

Test case	Neighbours			Improved Test cases?	Trial[i]	Probability
	a	b	c			
0	-9	-5	7	×	1	0.99
1	0	-4	9	×	1	0.99
2	4	7	4	×	1	0.99
3	-3	-2	-5	×	1	0
4	5	9	-9	×	1	0
5	1	1	8	×	1	0
6	0	-2	4	×	1	0
7	2	-4	-7	×	1	0.99
8	9	2	5	×	1	0
9	-5	-9	9	×	1	0

TABLE V: ONLOOKER BEE PHASE FOR ITERATION 1

Test case	Neighbour	Mutation	Changed Test case			Improved Test case?	Trial[i]
			a	b	c		
0	✓		-9	-7	7	×	2
1	✓		0	-4	9	×	2
2	✓		4	4	4		0
3		✓	-7	2	10	×	2
4		✓	1	10	-8	×	2
5		✓	0	-1	9	×	2
6		✓	4	-6	6	×	2
7	✓		2	-4	-5	×	2
8		✓	1	3	1	×	2
9		✓	-7	-10	10	×	2

The new neighbour explored an uncovered path i.e. path 6 and hence memorized by the application of algorithm. The value of count is also incremented to 5. Since trial for any of the test case does not exceed the preset limit, therefore, scout bee phase is not executed. The test suite after first iteration is as shown in Table VI. The whole algorithm is iterated second time until the termination criterion is not met. Table VII shows the employed bee phase for iteration 2. Test case number 2 i.e. [4 2 4] covers a new path number 3 with 'a' equals to 'c' and greater than 'b' and value of count is also incremented to 6. The onlooker bee phase for iteration 2 is shown in Table VIII. Test case [-3 -2 -5] got improved by mutation function covering path 7 and is shown in Table IX.

TABLE VI: PATHS COVERED AFTER ITERATION 1

Path_No	Test_case		
	a	b	c
1	2	-4	-7
2	0	-4	9
3	-		
4	4	7	4
5	-9	-5	7
6	4	4	4
7	-		

Since -13 exceeded the lower bound (i.e. -10), it is changed to -10 and the new test case becomes [-3 -10 -10] which is an improved one. The value of variable count now becomes 7 and is equal to the cyclomatic complexity, which is our termination criterion; hence, the algorithm is terminated here. Table X shows the optimised test suite.

TABLE VII: EMPLOYED BEE PHASE FOR ITERATION 2

Testcase	Neighbour			Improved Testcase?	Trial[i]
	a	b	c		
0	-9	-7	7	×	3
1	0	-4	7	×	3
2	4	2	4	✓	0
3	1	-2	-5	×	3
4	5	7	-9	×	3
5	2	1	8	×	3
6	-1	-2	4	×	3
7	2	-1	-7	×	3
8	6	2	5	×	3
9	-5	-9	5	×	3

TABLE VIII: ONLOOKER BEE PHASE FOR ITERATION 2

Testcase	Neighbour	Mutation	Changed Test case			Improved Test case?	Trial[i]
			a	b	c		
0	✓		-10	-7	7	×	4
1	✓		0	-4	10	×	4
2	✓		4	6	4	×	1
3		✓	3	-10	-10		0
4		✓	7	8	9	×	4
5		✓	3	5	10	×	4
6		✓	8	2	5	×	4
7	✓		2	-4	-5	×	4
8		✓	-9	5	-5	×	4
9		✓	-4	9	1	×	4

TABLE IX: MUTATION FUNCTION

Parameter	Binary Equivalent	Changed Bit	Changed Parameter
-3	10011	00011	3
-2	10010	11010	-10
-5	10101	11101	-13

TABLE X: OPTIMIZED TEST SUITE

Path_No	Test case		
	a	b	c
1	2	-4	-7
2	0	-4	9
3	4	2	4
4	4	7	4
5	-9	-5	7
6	4	4	4
7	3	-10	-10

VI. EMPIRICAL RESULTS AND ANALYSIS

In order to validate the efficiency of the MABC algorithm, we use 10 C++ programs on four algorithms and compared them on the basis of their mean time and mean number of iterations. Mean time is obtained after calculating the mean of the runtime obtained by 10 runs of a program and mean number of iterations is obtained after calculating the mean of the number of iterations obtained by 10 runs of a program.

- Simple ABC : ABC -1
- ABC with Mutation function at Onlooker Phase : ABC -2
- ABC with Mutation function at Scout Phase : ABC -3
- ABC with Mutation function at both Onlooker Phase and Scout Phase : ABC -4

Table XI gives the mean of time required, obtained after running a program for 10 times to increase the robustness of the algorithm. The results for all the four algorithms is listed and compared. It could be seen that MABC shows results in minimum time out of all the algorithms considered in 80% of the cases.

TABLE XI: MEAN TIME REQUIRED FOR DIFFERENT ALGORITHMS

S. No	Program Name	ABC -1	ABC - 2	ABC - 3	ABC - 4
1.	To find the greatest of three numbers.	0.352	0.483	0.323	0.36
2.	To determine the nature of roots of a quadratic equation.	0.184	0.246	0.164	0.192
3.	To test if a number is even or odd	0.033	0.036	0.044	0.032
4.	Triangle classification Problem	0.44	1.02	0.52	0.413
5.	Leap year	0.455	0.845	0.523	0.431
6.	Prime Number	0.05	0.1	0.07	0.05
7.	Armstrong Number	1.92	3.73	2.67	1.87
8.	Palindrome	0.11	0.19	0.16	0.06
9.	Calculator	0.22	0.31	0.28	0.16
10.	To display the season with month as an input	0.38	0.49	0.42	0.22

The bar graph shown in Figure 2 depicts the time required for implementing four algorithms on all programs. Table XII depicts the mean number of iterations obtained for each algorithm after running a program for 10 times to increase the robustness of the algorithm. ABC – 4 outperforms as compared to the other algorithms in terms of number of iterations. It gives an optimized test suite in minimum number of iterations. Bar graph shown in Figure 3 depicts the algorithm versus mean number of iterations required for various programs using all the four algorithms. Table XIII shows the number of input parameters for each program and number of optimized test cases given as an output by MABC algorithm.

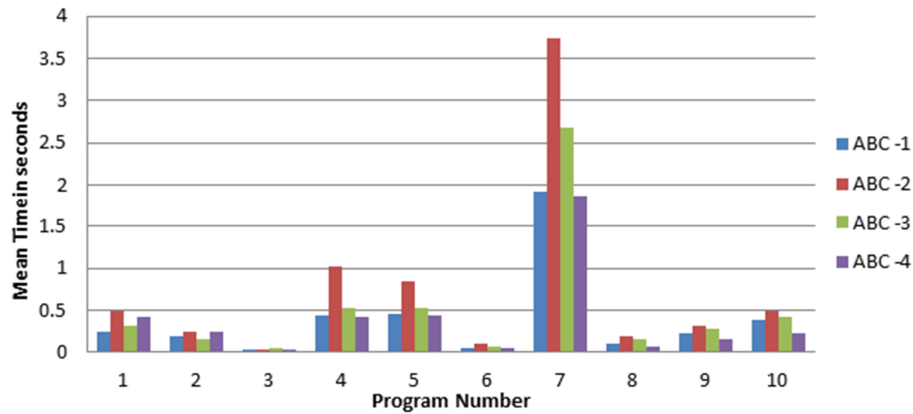


Figure 2: Algorithms against time required

TABLE XII: MEAN NUMBER OF ITERATIONS REQUIRED FOR DIFFERENT ALGORITHMS

S. No	Program Name	ABC -1	ABC - 2	ABC - 3	ABC - 4
1.	To find the greatest of three numbers.	7.1	7.5	8.8	6.9
2.	To determine the nature of roots of a quadratic equation.	6.9	6.5	6.8	5.6
3.	To test if a number is even or odd	1	1	1	1
4.	Triangle classification Problem	16	20	16.8	11
5.	Leap year	23	24	24.2	21
6.	Prime Number	2	2.7	2.1	1
7.	Armstrong Number	83.3	92.4	89.2	55.1
8.	Palindrome	3	6	5	1
9.	Calculator	4.4	5.3	4.9	2.6
10.	To display the season with m2onth as an input.	7.4	8.3	7.6	3.7

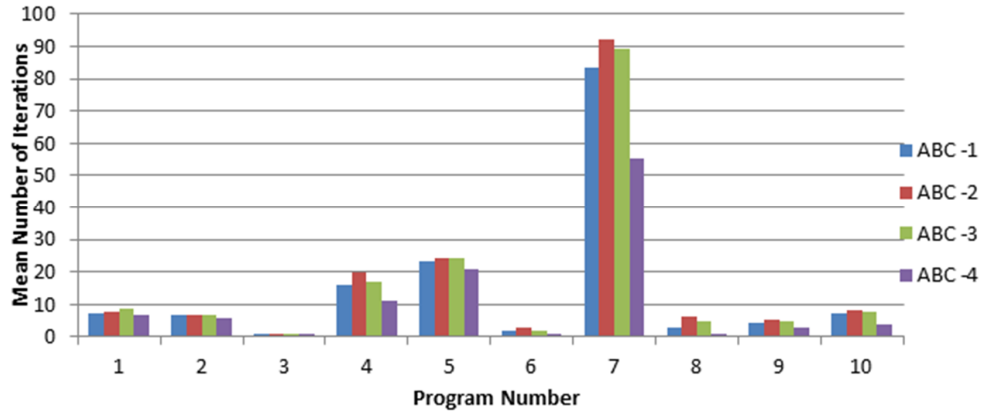


Figure 3: Algorithms against number of iteration

TABLE XIII: NUMBER OF INPUT PARAMETERS AND NUMBER OF OPTIMIZED TEST CASES

S. No	Program Name	Input Parameters	Number of Optimised Test cases
1.	To find the greatest of three numbers.	3	7
2.	To determine the nature of roots of a quadratic equation.	1	3
3.	To test if a number is even or odd	1	2
4.	Triangle classification Problem	3	4
5.	Leap year	1	3
6.	Prime Number	1	2
7.	Armstrong Number	1	2
8.	Palindrome	1	2
9.	Calculator	3	5
10.	To display the season with month as an input.	1	5

In this paper a new algorithm was proposed by combining ABC and mutation function of GA. The performance and accuracy of the proposed algorithm were examined on the basis of two parameters they are number of iterations and on the basis of runtime. The results show that ABC with mutation at both onlooker bee and scout bee (MABC) performs the best. ABC is recently proposed population based algorithm and it has the advantages of finding the best global optimal solution while optimization. The performance of ABC is good in terms of local and global optimization due to the selection schemes employed and the neighbour production mechanism used for test data generation. MABC algorithm performs better than rest modified ABC algorithm and efficiently proven in this paper to solve the optimal test data. This algorithm is used for test data optimization with the functionality of global as well as local search. This algorithm will be helpful for researchers in the field of software testing for optimization of automated test data generation.

VII. CONCLUSIONS

In this paper, we have compared various meta heuristic algorithms used for optimization on few theoretical parameters like ease of use, scalability, global/local search ability etc. We, then demonstrated the application of ABC in software test suite optimization and presented different types of ABC, those are – ABC - 1, ABC - 2, ABC - 3 and ABC - 4, which can be applied for test suite optimization. The proposed algorithm is the MABC - the version of ABC on which mutation operation of GA is applied at both onlooker and scout bee phases (ABC - 4). We verified and validated the MABC algorithm by running it on 10 C++ programs and comparing the results with those obtained from ABC - 1, ABC - 2 and ABC - 3. It is shown that MABC algorithm gives results in lesser time and with less number of iterations and negligible increase in the space complexity and enhances the local and global search ability of the ABC algorithm.

In future we plan to validate the MABC technique on large sized software in order to obtain well formed and generalizable results. The quality of the final best algorithm solution might be improved further by using an efficient and exploitative search.

REFERENCES

- [1] P. McMinn, "Search-based software test data generation: a survey," *Software Testing Verification Reliability*, 2004.
- [2] P. Maragathavalli, M. Anusha, P. Geethamali and S. Priyadharsini, "Automated Test Data Generation for modified Condition/ Decision coverage using Genetic Algorithm," *International Journal of Engineering Science and Technology (IJEST)*, vol. 3, 2011.
- [3] E. Elbeltagi, T. Hegazy and D. Grierson, "Comparison among five evolutionary-based optimization algorithms," Elsevier Ltd, pp. 43–53, 2005.
- [4] P. Garg, "A Comparison between Memetic algorithm and Genetic algorithm for the cryptanalysis of Simplified Data Encryption Standard algorithm," IEEE Computer Society, 2008.
- [5] D. Jeya Mala, V. Mohan and M. Kamalpriya, "Automated software test optimization framework – an artificial bee colony optimization-based approach," *IET Software*, vol. 4, pp. 334–348, 2010.
- [6] D. Jeya Mala and V. Mohan, "ABC Tester - Artificial Bee Colony Based Software Test Suite Optimization Approach," *International Journal of Software Engineering (IJSE)*, vol. 2, 2009.
- [7] N. Karaboga, "A new design method based on artificial bee colony algorithm for digital IIR filters," Elsevier, pp. 328–348, 2009.
- [8] S. Kanmani and P. Maragathavalli, "Search-Based Software Test data generation using Evolutionary Testing Techniques," *International Journal of Software Engineering (IJSE)*, vol. 1, 2006.
- [9] A. Singh, "An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem," Elsevier, pp. 625–631, 2009.
- [10] R. Malhotra and M. Garg, "An Adequacy Based Test Data Generation Technique Using Genetic Algorithms," *Journal of Information Processing Systems*, vol. 7, 2011.
- [11] R. Hassan, B. Cohanin, O. D. Weck, "A comparison of Particle Swarm Optimization and the Genetic Algorithm," Springer, 2006.
- [12] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Applied Mathematics and Computation*, Elsevier, vol. 214, pp. 108–132, 2009.
- [13] K. Karabulut and M. F. Tasgetiren, "A Discrete Artificial Bee Colony Algorithm for the Team Orienteering Problem with Time Windows," *IEEE Symposium on Computational Intelligence in Production and Logistics Systems (CIPLS)*, 2013.
- [14] D. Shanthia and R. Amalraj, "Collaborative Artificial Bee Colony Optimization Clustering Using SPNN," *International Conference on Communication Technology and System Design 2011*, Elsevier, vol. 30, pp. 989–996, 2012.
- [15] E. Cuevas, D. Zaldívar, M. P. Cisneros, H. Sossa and V. Osuna, "Block matching algorithm for motion estimation based on Artificial Bee Colony (ABC)," *Applied Soft Computing*, Swarm intelligence in image and video processing, Elsevier, vol. 13, pp. 3047–3059, 2012.
- [16] M. Subotic, M. Tuba and N. Stanarevic, "Parallelization of the artificial bee colony (ABC) algorithm," *Recent Advances In Neural Networks, Fuzzy systems & evolutionary computing*, ACM, pp. 191–196, 2010.
- [17] M. S. Alam, Md. M. Islam and K. Murase, "Artificial Bee Colony Algorithm with improved Exploration for Numerical Function Optimization," *Intelligent Data Engineering and Automated Learning - IDEAL 2012*, Lecture Notes in Computer Science, Springer, vol. 7435, pp. 1–8, 2012.
- [18] F. S. Abu-Mouti and M. E. El-Hawary, "Overview of Artificial Bee Colony (ABC) Algorithm and Its Applications," *Systems Conference (SysCon)*, IEEE International, pp. 1–6, 2012.
- [19] C. Qiao and H. Wang, "The Further Research on the Application of ABC to the Optimization and Control of Project," *Engineering Management Research*, Published by Canadian Centre of Science and Education, vol. 1, 2012.
- [20] M. S. Kiran and M. Gündüz, "A recombination-based hybridization of particle swarm optimization and artificial bee colony algorithm for continuous optimization problems," *Applied Soft Computing*, Elsevier, vol. 13, pp. 2188–2203, 2013.
- [21] W. Gao, X. China, S. Liu, L. Huang, "A novel Artificial Bee Colony Algorithm based on Modified Search Equation and Orthogonal Learning," *Cybernetics*, IEEE Transactions, vol. 43, pp. 1011–1024, 2013.



Text Categorization – A Review

Rajni Jindal^a, Shweta Taneja^{b,*}

^a Professor, Dept. of Computer Engineering, Delhi Technological University, Formerly Delhi College of Engineering (DCE), Bawana Road, Delhi-42.
^b Research Scholar, Dept. of Computer Engineering, Delhi Technological University, Formerly Delhi College of Engineering (DCE), Delhi-42.

Abstract

With the growth of internet, the amount of digital information is growing exponentially day by day. This information may be structured or unstructured in nature. So, a need to convert unstructured text into structured text and to infer knowledge was felt. As a result of this, the field of text mining emerged. Text documents may be in the form of online news articles, emails, scientific documents or reports. Text Classification or Categorization is the process of assigning of free occurring text documents into some predefined categories based on the nature and characteristics of the document. A number of statistical and machine learning techniques or classifiers are applied to text categorization. To summarize the work, in this paper, we give a survey of state of art in text categorization, different classifiers available and applications of text categorization.

Keywords: Text Mining, Text Categorization, Classifiers

1. Introduction

The term text mining is the process of extraction of useful or interesting knowledge from unstructured or natural language text data [1]. It is an interdisciplinary field that is it uses methods and techniques from other disciplines like machine learning, information retrieval, natural language processing, knowledge management etc. The text documents may be in the form of news stories, emails, scientific research papers, business documents etc. Text categorization or text classification is a task to assign a document to a predefined class or category. That is, given a collection of text documents and a set of topics or categories, text classification is a process of assigning correct category to each document. There are a variety of applications of text categorization like document sorting, text indexing, Web page categorization, text filtering which includes email spam filtering etc.

There are a variety of methods available for performing text categorization. These include decision trees, K Nearest Neighbor method (K.N.N.), Support Vector Machines (S.V.M), Genetic Algorithms, Neural Networks, Bayesian Classifiers etc. In this paper, we have given a survey of all these techniques available. We have tried our best to present the latest work going on in the field of text classifiers.

The paper is organized as follows: Section 2 explains the concept of text categorization giving in detail the

process. In Section 3, we focus on different classifiers available like KNN method, SVM method, Genetic Algorithms and Naïve Bayes method showing the work done in separate tables. Section 5 gives the work done in case of hybrid classifiers, that is which is formed by combining different Classifiers. Finally Conclusion is given in the next section.

2. Text Categorization

The aim of text categorization is to approximate a category assignment function $F : D \times C \rightarrow \{0, 1\}$, where D is the set of all possible documents and C is the set of predefined categories [6]. If a document d belongs to category c , then the value of F is 1, otherwise it is 0. In text categorization, depending on the type of application and value of function F , there is single label text categorization in which document belongs to exactly one category whereas in multi label, a document may belong to more than one category. Another case is binary categorization in which the number of categories is two. It is the simplest and most popular technique for text categorization.

The process of text categorization involves following steps: firstly raw unstructured text document is taken, it undergoes pre processing. In pre processing, stop words [2] are removed, stemming[3] is done, HTML tags are removed etc. The task of pre processing is to convert an

unstructured document to a structured form. After that, feature selection and feature representation is done. Features are the words, multi words, frequently occurring phrases present in the document. The features are selected and are used to represent a document. Normally, a document is represented as a vector of features. Different methods are used for feature selection and representation like LSI (Latent Semantic Indexing) [4], TF-IDF (term frequency – inverse document frequency) [5] etc. After this step, a suitable machine learning algorithm is used for training a text classifier. But before using the classifier, its predictive accuracy is checked using a test set. If it is acceptable, then only the classifier is used to classify new documents. The process of text categorization is shown in fig 1.

3. Types of Classifiers

There are different types of statistical and machine learning classifiers available. These include decision trees; Regression based methods, K nearest neighbor, Bayes method, Support Vector Machines, Genetic Algorithms, Neural Networks etc. Different authors have contributed in this field. As a lot of work is going on in individual classifiers, so we have shown the contribution made by different authors in a tabular way. We have made separate tables (tables 1 to 4) for various classifiers. The work is shown in a chronological order with respect to time. The parameters taken are dataset used, approach followed with results and accuracy.

3.1 K N N Classifier

The K .N.N is a famous and simple algorithm for text classification. To classify a new document, the algorithm searches for the K nearest neighbours of the document. Depending on the classes or categories of them, it makes a decision . This method is simple to implement but is slow in performance as it needs to compare the test document with all training documents. The table 1 given shows the survey of state of art in K Nearest Neighbor method.

3.2 SVM Classifier

The S.V.M method is a more powerful method for text categorization [17]. In this method, we construct a hyper plane that separates positive examples from negative examples. It can work for both linearly separable problems as well as for non-linearly separable problems. Another peculiar property of S.V.M. method is its ability to learn which is independent of the dimensionality of feature space. In text mining, we deal with documents having high dimensionality of feature space. So, it performs best with text documents. The table 2 given shows the survey of state of art in S.V.M method.

3.3 Naïve Bayes Classifier

The Naive Bayes Classifier is a statistical Classifier method. It is based on Bayesian theorem which predicts class membership probabilities. It is an easy method and gives good results in some cases. But it has a disadvantage that it assumes class conditional independence. And in real world , dependency exists between variables or words in text. This becomes difficult to handle. The table 3 given shows the survey of state of art in Naïve Bayes method.

3.4 Genetic Algorithms

Genetic Algorithms are based on the theory of survival of the fittest. It is an optimization method which is used for its simplicity. Nowadays this method is used for the improvement of text categorization process. The table 4 given shows the survey of state of art in Genetic Algorithms method.

3.5 Combination of multiple Classifiers

In the area of text categorization, authors have proposed a hybrid approach that is they have combined various classifiers taking their strong points and discarding the weaknesses. Following table 5 shows the work done by different authors after combining the classifiers.

4. Comparison between Classifiers

From our survey on different types of text classifiers, we have analyzed that S.V.M (Support Vector Machines) is the most popular, efficient and best learning algorithm for text categorization. Next is K.N.N (K Nearest Neighbor) algorithm as it is also effective and simple in classification of textual documents. Although it has some drawbacks like it is termed as a lazy learning algorithm as it spends a lot of time in classifying unknown documents. A lot of work is going on to improve the performance of this algorithm. Other Classifiers are Naïve Bayes method, Genetic Algorithms and Neural Networks. There are some authors that take a hybrid or combination of different Classifiers like S.V.M and K.N.N or K.N.N and Naïve Bayes. The hybrid Classifiers prove to be better than individual Classifiers. The survey done can be shown in the following figure 2 in terms of a pie chart as shown below.

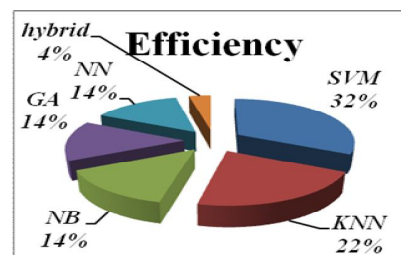
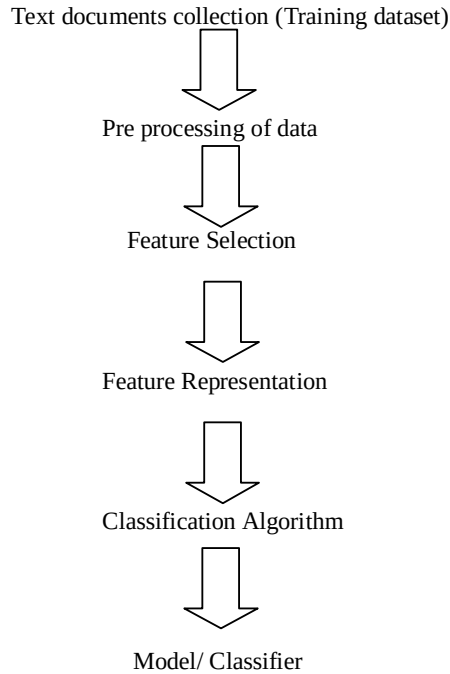


Figure 2: Figure showing the efficiency of different text classifiers

Training phase



Testing Phase

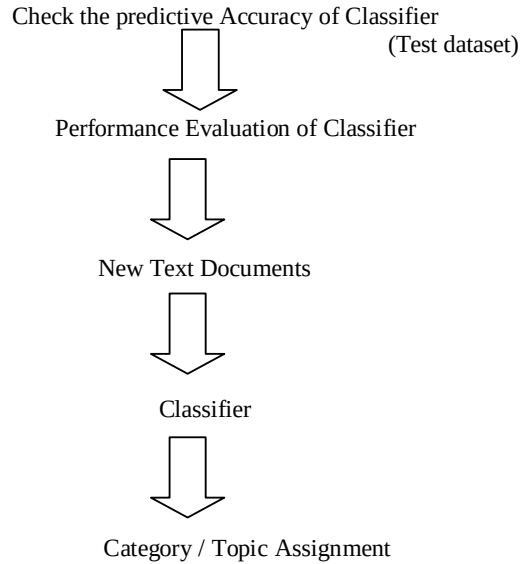


Fig 1: Architecture of text classifiers

Table 1: Table showing survey of the work done in K N N method.

S. No	Authors	Details of Publication	Dataset Used	Approach Used	Result and Accuracy
1	Yiming Yang, Jan Pederson [7]	1997	Reuters corpus and Ohsumed Collection	They focussed on dimensionality reduction. They evaluated five methods: term selection based on document frequency, information gain, mutual information and chi square test.	Information gain and chi-square methods are most effective.
2.	Songbo Tan [8]	Journal of Expert Systems with Applications, Elsevier, 2005.	Two datasets: Reuters 21578 and TDT2	They proposed a neighbor weighted KNN algorithm for unbalanced datasets. It assigns a big weight to neighbors from small classes and little weight to neighbors of big classes.	The proposed algorithm achieves good performance improvement on imbalanced corpora.
3.	Zhi Hong Deng and Shi Wei Tang [9]	Springer Conference 2005	Newsgroups 18828 and Ohscal	They proposed a non VSM KNN algorithm for text classification based on correlations between categories and features.	They evaluated the algorithm on two datasets and compared it against traditional KNN. Their algorithm performs better.
4.	Hatem A. Fayed and Amir F. Atiya [10]	IEEE Transactions on Neural Networks, 2009	Real life datasets from UCI repository.	They have proposed a new condensing approach to remove patterns which cause burden . They have defined a chain of nearest neighbors from alternating classes and then set a cut off for the patterns in the training set.	Their approach is proved to be a simple and fast condensing algorithm and maintains the same level of classification accuracy as the traditional KNN method.
5.	Pedro J. Garcia Laencina et. Al [11]	Journal of neurocomputing, Elsevier, 2009.	Two complete datasets and two incomplete datasets from UCI repository	They propose a novel KNN imputation procedure for missing data using a feature weighed distance metric.	The proposed algorithm is efficient and robust.
6.	Jun Toyoma et. Al [12]	Journal of Pattern Recognition, Elsevier, 2010	Japanese phoneme dataset	They proposed a probably correct approach in which the correct set of k nearest neighbors is obtained in high probability to reduce the search time.	The proposed algorithm has advantage that in it, the searching time is reduced.

Table 1: Table showing survey of the work done in K N N method.

7.	Eric Bax [13]	IEEE Transactions on information theory, 2012.	----	They present a method to calculate probably approximately correct error bounds for k nearest neighbor classifiers.	It presents a bound on out of sample (examples not used for training a classifier) error rate for k nearest neighbor classifier.
8.	Shichao Zhang [14]	The journal of systems and software, Elsevier, 2012.	Real datasets from UCI repository	They proposed a novel KNN imputation method to impute missing data: GKNN-grey KNN. It is based on calculating gray distance between missing data and training data.	They have implemented the proposed algorithm and have shown that it is more efficient than existing methods.
9.	Shengyi Jiang et. Al [15]	Journal of Expert Systems with Applications, Elsevier, 2012	Three datasets: Reuters 21578, Fudan University corpus and Ling Spam email spam filtering corpus.	They have improved KNN algorithm by using one pass clustering algorithm in it.	They have tested on three datasets and proved their algorithm to be better.
10.	Gleb Beliakov, Gang Li [16]	Journal of Pattern Recognition Letters, Elsevier, 2012.	----	They have improved the speed and stability of KNN method by replacing the sort operation with calculating the order statistics.	Their approach is proved to be superior.

Table 2: Table showing survey of the work done in S.V.M method.

S. No	Authors	Details of Publication	Dataset Used	Approach Used	Result and Accuracy
1.	Thorsten Joachims [17]	ACM Digital library, 1998.	Two datasets: Reuters 21578 and Ohsumed corpus	He compared SVM method with Naïve Bayes , Rocchio algorithm, KNN and Decision tree.	SVM is the best classifier with 66% accuracy.
2.	Chung Hong Lee, Hsin Chang Yang [18]	----	Random set of documents consisting of five classes.	They used a hybrid approach for measuring semantic relatedness among text. They developed several text classifiers using SVM.	SVM Classifier with a Gaussian Kernel is the best having greater than 90% recall ratio.
3.	M. Arun Kumar, M. Gopal [19]	Journal of Pattern Recognition Letters, Elsevier , 2010.	Three datasets: Web KB, 20 Newsgroups, Industry Sector	They performed empirical comparison of standard O.A.A and O.A.O together with three improvements made to these standard approaches.	For small scale text categorization tasks, one against all (O.A.A) will have better performance than other methods.
4.	Li Zhixing , Xiong Zhongyong Zhang, Liu, Li Kuan [20]	Journal of Pattern Recognition Letters, Elsevier , 2011	Three datasets: Reuters 21578, 20 Newsgroups and Tan Corp	They proposed a concise semantic analysis technique for text categorization tasks.	Their proposed method reaches a comparable performance with SVM classifier in text categorization tasks.
5.	Shweta Mayor and Bhasker Pant[21]	International Journal of Engineering Science and Technology, 2012	50 news instances from newspapers or sites	They have implemented SVM for calculating term frequency. They also have used LIBSVM tool.	The accuracy rate obtained was 66.67%.

Table 3: Table showing survey of the work done in Naive Bayes method.

S. No	Authors	Details of Publication	Dataset Used	Approach Used	Result and Accuracy
1.	Sang Bum et al. [22]	IEEE Transactions on Knowledge and Data Engineering, 2006	Reuters -21578, 20 Newsgroups	They propose two empirical heuristics to overcome a problem in naïve bayes method.	
2.	Yi Hsing Chang, Hsiu Yi Huang [23]	IEEE Conference, 2008	319 documents from Electronic Thesis and Dissertations	Proposed an automatic document classifier system based on ontology and Naïve Bayes Classifier.	The average effectiveness of document classification in 11 categories is about 89%.
3.	Daniele Soria et al. [24]	Journal of Knowledge-Based Systems, Elsevier, 2011.	breast cancer data and on three UCI datasets	They have proposed a new method which does not restrict the variables to have values which are normally distributed.	Their algorithm performed well in all four datasets.
4.	Wei Zhang, Feng Gao [25]	Journal of Procedia Engineering 2011.	-----	They have proposed an auxillary feature selection method.	The proposed method improves the performance of naïve bayes classifier.
5.	Appavu alias Balamurugan et al. [26]	Journal of Knowledge-Based Systems 2011.	Various databases	They have proposed a novel algorithm which extends the traditional naïve bayes method. They have used a partial matching method.	The proposed algorithm performs better than traditional algorithm.

Table 4: Table showing survey of the work done in Genetic Algorithm method.

S. No	Authors	Details of Publication	Dataset Used	Approach Used	Result and Accuracy
1	John Atkinson, Abutridy, Chris Mellish, Stuart Aitken [27]	IEEE Intelligent Systems, 2004.	-----	Their approach had two steps: first is preprocessing step that produces training data and initial population of genetic algorithm after information extraction from documents. It will generate rules. The second step is genetic algorithm based knowledge discovery in which hypothesis are produced.	They implemented a prototype of model in Prolog and achieved good accuracy which is assessed by experts.
2.	S.M.Khalessizadeh et al. [28]	World Academy of Science, Engineering and Technology, 2006	Text in Persian language	They have proposed a new algorithm using genetic algorithm based on concept standard deviation is explored.	The proposed algorithm has a good performance.
3.	Zhu Zhen fang, Liu Pei yu and Lu Ran [29]	IEEE Symposium, 2008	-----	They introduced simulated annealing mechanism of genetic algorithm to solve the problem of text categorization.	The proposed method had good accuracy and recall rate.
4.	Adriana Pietramala et al. [30]	Springer Conference, 2008	Reuters-21578, Ohsumed Collection	They have presented a genetic algorithm called Olex-GA, based on rule based text classifiers.	The proposed method is equally competitive with other classifiers.
5.	Harun Uğuz [31]	Journal of Knowledge-Based Systems , 2011	Reuters 21578 and Classic 3 datasets	They have used a two stage feature selection and feature extraction method using genetic algorithms.	The proposed method achieves good effectiveness.

Table 5: Table showing survey of the work done in hybrid method.

S. No	Authors	Details of Publication	Dataset Used	Approach Used	Result and Accuracy
1.	Y.H.Li and A.K.Jain [32]	The Computer Journal, 1998.	Yahoo newsgroups dataset having 7 classes.	They took a combination of four classifiers: Naïve Bayes, Nearest Neighbour, decision tree and subspace method .They used three approaches: simple voting, dynamic classifier selection and adaptive classifier combination.	Naïve Bayes and subspace methods are better having an accuracy of 83%. The approach of adaptive combination is the best.
2.	Yiming Yang and Xin Liu [33]	ACM Digital Library, 1999.	Reuters -21578	They combined five classifiers: SVM, KNN ,Neural Networks , LLSF(linear least squares fit) and Naïve Bayes and compared the results.	SVM, KNN and LLSF are better if data is less and all methods perform equally if data is large.
3.	Akinori Fujino et al. [34]	Journal of Information Processing & Management, 2007.	They took four test collections.	They have focused on a hybrid of classifiers using probabilistic generative and discriminative approaches. They have used naïve bayes method.	Their proposed method achieved good classification accuracy.
4.	Chung Hong Lee and Hsin Chang Yang [35]	Expert Systems with Applications, Elsevier 2009.	They took 6000 news documents.	They combined SVM (supervised technique) and SOM(self organizing maps- unsupervised technique) for multilingual text categorization.	The proposed hybrid method provided good accuracy.
5.	Chin Heng Wan et al [36]	Journal of Expert Systems with Applications, 2012.	Benchmark text datasets.	They have combined KNN with Support Vector Machines.	Their method has achieved good classification accuracy.

in Emails, Word sense disambiguation, Authorship attribution, Opinion Mining and Analysis, Biomedical text mining to many more

5. Applications of Text Classification

The application areas of text categorization include a wide variety of domains. These range from text indexing, text retrieval, extraction of metadata, document sorting and text filtering [6], Web page Categorization, Spam Filtering

6. Conclusion

Our review in this paper focuses on the existing literature present in the field of text classifiers. From our analysis, we have found that no single classifier is the best or can be stated as standard model in text categorization. All methods perform in a different way depending on the type of data as well as the type of application or domain. But still from our analysis, we can conclude that SVM and KNN methods perform better than the other existing methods. SVM method can manage high dimensional input space and KNN method is simple and an effective classifier.

References

- [1] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco, 2000.
- [2] H. Lili. and H. Lizhu. , Automatic identification of stopwords in Chinese text classification. In proceedings of the IEEE international conference on Computer Science and Software Engineering, 2008, p. 718 – 722.
- [3] M.F.Porter 1980. An algorithm for suffix stripping, *Program*, 14 (3), 1980, p. 130-137.
- [4] S. Deerwester, S.T. Dumais, T. K. Landauer, G.W. Furnas and R. Harshman , 1990. Indexing by Latent Semantic Analysis. *Journal of American Society of Information Science*, 41(6), 1980, p. 391-407.
- [5] K.S. Jones, A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, vol. 28, no. 1, 1972, p. 11-21.
- [6] R. Feldman, J. Sanger, *The Text Mining Handbook*, Cambridge Press.
- [7] Y. Yang, J. Pederson, A Comparative study on Feature Selection in Text Categorization, 1997.
- [8] S. Tan, Neighbor – weighted K- nearest neighbor for unbalanced text corpus, *Journal of Expert Systems with Applications*, Elsevier, 2005, p. 667-671.
- [9] Z. H. Deng and S. W. Tang, A Non – VSM Algorithm for Text Classification, *LNAI 3584*, p. 339-346, Springer.
- [10] H. A. Fayed and A. F. Atiya , A Novel Template Reduction Approach for the K- Nearest Neighbor Method, *IEEE Transactions on Neural Networks*, vol. 20, no. 5, 2009.
- [11] P. J. G. Laencina et. al, K nearest neighbours with mutual information for simultaneous classification and missing data imputation, *Journal of neurocomputing*, Elsevier, 2009, p. 1483-1493..
- [12] J. Toyoma et.al., Probably correct k- nearest neighbor search in high dimensions, *Journal of Pattern Recognition*, Elsevier, 2010, p. 1361-1372.
- [13] E. Bax , Validation of K- Nearest Neighbor Classifiers, *IEEE Transactions on information theory*, vol. 58, no. 5, 2012.
- [14] S. Zhang, Nearest Neighbor selection for iteratively KNN imputation, *The Journal of systems and software*, Elsevier, 2010, p. 2541-2552.
- [15] S. Jiang et. al, An improved K-nearest neighbor algorithm for text categorization, *Journal of Expert Systems with Applications*, Elsevier, 2012, p. 1503-1509.
- [16] G. Beliakov, G. Li, Improving the speed and stability of the K-nearest neighbors method, *Journal of Pattern Recognition Letters*, Elsevier, 2012, p. 1296-1301.
- [17] T. Joachims, *Text Categorization with Support Vector Machines: Learning with many Relevant Features*, ACM Digital library, 1998.
- [18] C. H. Lee, H. C. Yang, A Classifier based Text Mining Approach for Evaluating Semantic Relatedness Using Support Vector Machines, in proceedings of IEEE International Conference, 2005.
- [19] M. Arun Kumar, M. Gopal, A comparison study on multiple binary-class SVM methods for unilabel text categorization, *Journal of Pattern Recognition Letters*, Elsevier, 2010, p. 1437-1444.
- [20] L. Zhixing , X. Z. Zhang, Liu, L. Kuan, Fast text categorization using concise semantic analysis, *Journal of Pattern Recognition Letters*, Elsevier, 2011, p. 441-448.
- [21] S. Mayor, B. Pant, Document Classification using Support Vector Machine, *International Journal of Engineering Science and Technology*, vol. 4, 2012.
- [22] S. B. Kim et al., Some Effective Techniques for Naïve Bayes Text Classification, *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, 2006.
- [23] Y. H. Chang, H. Y. Huang, An Automatic Document Classifier system based on Naïve Bayes Classifier and Ontology , in proceedings of IEEE International Conference on Machine Learning and Cybernetics, 2008.
- [24] D. Soria et al., A ‘non-parametric’ version of the naive Bayes classifier, *Journal of Knowledge-Based Systems*, vol. 24, Issue 6, Elsevier, 2011, p. 775–784.
- [25] W. Zhang, F. Gao, An Improvement to Naive Bayes for Text Classification, *Journal of Procedia Engineering*, vol.15, 2011, p. 2160–2164.
- [26] Balamurugan et al., NB+: An improved Naïve Bayesian algorithm, *Journal of Knowledge-Based Systems*, vol.24, 2011, p. 563–569.
- [27] J. Atkinson, Abutridy, C. Mellish, S. Aitken, Combining Information Extraction with Genetic Algorithms for Text Mining, *IEEE Intelligent Systems*, 2004.
- [28] S.M.Khalessizadeh et al., Genetic Mining: Using Genetic Algorithm for topic based on concept distribution, *World Academy of Science, Engineering and Technology*, 2006.
- [29] Z. Z. fang, L. Pei yu and L. Ran, Research of text classification technology based on genetic annealing algorithm, *IEEE International Symposium on Computational Intelligence and Design*, 2008.
- [30] A. Pietramala et al., A Genetic Algorithm for Text Classification Rule Induction, in proceedings of ECML PKDD, LNAI 5212, Springer Conference, 2008
- [31] H. Uğuz, A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm, *Journal of Knowledge-Based Systems*, vol.24, 2011, p. 1024–1032.
- [32] Y.H.Li and A.K.Jain, Classification of Text Documents, *The Computer Journal*, vol. 41, no.8, 1998.
- [33] Y. Yang and X. Liu, A re-examination of text categorization methods, in proceedings of SIGIR, ACM Digital Library, 1999.
- [34] A. Fujino et al., A hybrid generative/discriminative approach to text classification with additional information, *Journal of Information Processing & Management*, vol.43, 2007, p. 379–392.
- [35] C. H. Lee and H. C. Yang , Construction of supervised and unsupervised learning systems for multilingual text categorization, *Expert Systems with Applications*, Elsevier, 2009, p. 2400-2410.
- [36] C. H. Wan et al., A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine, *Journal of Expert Systems with Applications*, vol.39, 2012, p. 11880–11888.
- [37] W. Lam , M. Ruiz, P. Srinivasan, Automatic Text Categorization and Its Application to Text Retrieval, *IEEE Transactions on Knowledge and Data Engineering*, vol. 11 ,no. 6, 1999, p. 865-879.

Thermodynamic Irreversibilities Analysis of Industrial Ice Plant

R. S. Mishra^{*}, S. S. Kachhwaha

Department of Mechanical Engineering, Delhi Technological University, New Delhi, India

Article Info

Article history:

Received 20 September 2014

Received in revised form

10 October 2014

Accepted 20 November 2014

Available online 15 December 2014

Keywords

Energy and Exergy Analysis,
Efficiency Defect,
Vapour Compression Refrigeration
System,
Ice Plant

Abstract

The behavior of single-stage vapour compression-refrigeration cycle, using NH₃, has been investigated by the exergy method. The condenser and the evaporator's saturation temperatures were varied from 295 K to 305 K and from 249 K to 239 K respectively. The effects of temperature changes in the condenser and evaporator on the plant's irreversibility rate were determined. The greater the temperature difference between either (i) the condenser and the environment, or (ii) the evaporator and the cold room, the higher the irreversibility rate. Reduction in the irreversibility rate of the condenser gives approximately 2.58 times greater reduction in the irreversibility rate for the whole plant, whereas reduction in the evaporator's irreversibility rate gives a 2.32 times greater mean reduction in the irreversibility rate of the whole plant. Because the changes in the temperatures in the condenser and the evaporator contribute significantly to the plant's overall irreversibility, there is considerable scope for optimizing the condition imposed upon the condenser and evaporator.

Nomenclature

A	Area (m ²)
c	Constant
E_Q^D	Exergy rate, (kW)
h	Specific enthalpy, (kJ/kg)
h_f	Enthalpy of saturated liquid refrigerant, (kJ/kg)
I	Irreversibility rate, (kW)
m_R	Mass flow rate of refrigerant, (kg/s)
P	Pressure, (bar)
Q_0	Heat transfer rate, (kW/m ² K)
r	Compression ratio
s	Specific entropy, (kJ/kg K)
s_f	Entropy of saturated liquid, (kJ/kg K)
T	Temperature, (K)
W	Work input, (kW)
X	Dryness fraction

Greek Symbols (all dimensionless)

η_m	Mechanical efficiency
η_e	Electric motor efficiency
$\sigma_{k,I}$	Coefficient of structural bonds
η_R	Rational efficiency
δ_k	Efficiency defect

Abbreviations

COMP	Compressor
COND	Condenser
COP	Coefficient of performance
CSB	Coefficient of structural bond
EVAP	Evaporator
t	total

1. Introduction

Thermodynamic processes in refrigeration system releases large amount of heat to the environment.

Corresponding Author,

E-mail address: professor_rsmishra@yahoo.co.in

All rights reserved: <http://www.ijari.org>

Further, heat transfer between the system and the surrounding environment takes place at a finite temperature difference, which is a major source of irreversibility for the cycle and also responsible for the system performance degradation. The losses in the cycle need to be evaluated considering respective individual thermodynamic processes by applying first and second laws. Energy analysis (First Law) is still the most commonly used method in the analysis of thermal systems which only concern with the conservation of energy, and gives no information on how, where, and how much the system performance is degraded. On the other hand, second law is used to describe the quality of energy of materials. The first law optimization results in maximizing the coefficient of performance (COP) thus providing maximum heat removal from minimum power input; while the second law optimization is used for maximizing the exergy efficiency and minimizing entropy generation within the system, hence providing maximum cooling for the smallest distraction of available energy (exergy).

The exergy method, known as the second law analysis calculates the exergy loss caused by irreversibility which is an important thermodynamic property that measures the useful work that can be produced by a substance or the amount of work needed to complete a process. Thus exergy analysis is powerful tool in the design, optimization, and performance evaluation of energy systems.

The principles and methodologies of exergy analysis are well established [1-4]. An exergy analysis is usually aimed to determine the maximum performance of the system and identify the sites of exergy destruction. Exergy analysis of a complex system can be performed by analyzing the components of the system separately. Identifying the

main sites of exergy destruction shows the direction for potential improvements. An important objective of exergy analysis for system that consume work such as refrigeration, liquefaction of gases, and distillation of water is to find the minimum work required for a certain desired result [5].

There are several studies on the exergy analysis of refrigeration system [6, 7]. Bejan [7] showed that the exergetic efficiency decreases as the refrigeration temperature decreases. He offered two simple models to explain this trend. In his model, thermodynamic imperfections are explained largely by the heat transfer irreversibility. The behaviour of two stage compound-cycle with flash intercooling, using refrigerant R22 has been investigated by Nikolaidis and Probert [8] using exergy method. A computational model based on the exergy analysis is presented by Yumrutas et. al [9] for the investigation of the effects of the evaporating and condensing temperatures on the pressure losses, exergy losses, second law of efficiency, and the COP of a vapour compression cycle.

In the present research work, exergy analysis is performed on the operating data of vapour compression refrigeration cycle of an industrial ice plant. The expressions for the exergy losses (lost works) for the individual processes of the cycle as well as the coefficient of performance (COP) and second law efficiency for the entire cycle has been obtained. Effect of variation of condensing and evaporating temperatures on exergy losses, second law efficiency and COP has been investigated. The concept of structure coefficient (coefficient of structure bond) is used to explain the relation between irreversibilities of system & its component.

2. Plant Description and Thermodynamic Formulation

As shown schematically in Figure 1,

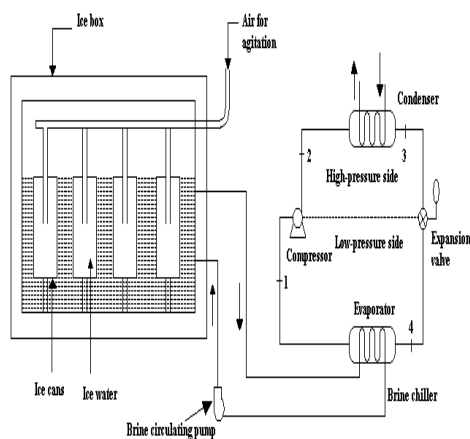


Fig. 1a. Schematic Diagram of an Industrial Ice Plant

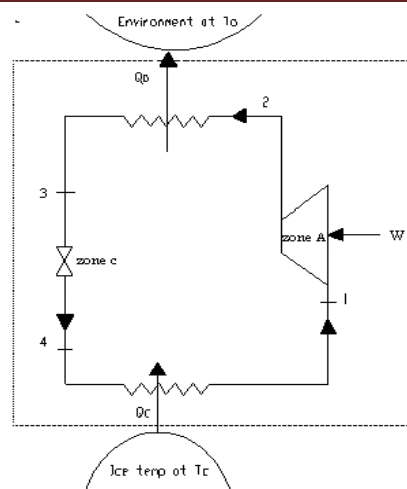


Fig. 1a. Schematic Diagram of an Industrial Ice Plant

The main components of an industrial ice plant are compressor, expansion valve, condenser and evaporator (brine chiller connected with ice box). Ice tank is made of mild steel plates and having heavy insulation of granulated corks around 300 mm in thickness. The tank is filled with recirculated brine in which cans filled with water are immersed. The tank is provided with piping for blowing air into ice cans to create agitation. In the present industrial ice plant, single stage vapour compression system has been used for ice production. Reverse osmosis water is being used in the plant for production of ice cubes. The corresponding temperature versus entropy diagram for this system is given in Figure 2.

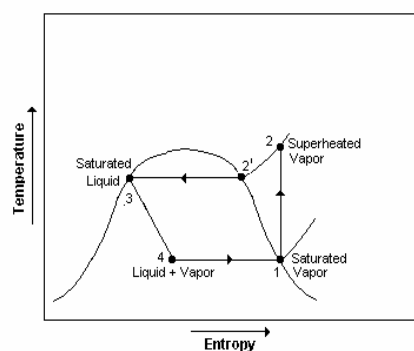


Fig. 2. T-s Diagram of Single Stage Vapour Compression Cycle

At point 1, the circulating refrigerant enters the compressor as a saturated vapour. From point 1 to point 2, the refrigerant vapour is isentropically compressed and exits the compressor as a superheated vapour. From point 2 to point 2', the superheated vapour travels through part of the condenser which removes the superheat by cooling the vapour. Between point 2' and point 3, the vapour travels through the remainder of the condenser and is condensed into a saturated liquid. The condensation process occurs essentially at constant pressure. Between points 3 and 4, the saturated liquid refrigerant passes through the expansion valve and undergoes an abrupt decrease of pressure. This

process results in the adiabatic flash evaporation and causes auto-refrigeration of a portion of the liquid refrigerant (typically, less than half of the liquid flashes). The adiabatic flash evaporation process is isenthalpic (i.e., occurs at constant enthalpy) in nature. Between points 4 and 1, the cold and partially vaporized refrigerant travels through the coil or tubes in the evaporator where it is totally vaporized by the warm water (from the space being refrigerated) that circulates across the coil or tubes in the evaporator. The evaporator operates at essentially constant pressure.

The resulting saturated refrigerant vapor returns to the compressor inlet at point 1 to complete the thermodynamic cycle. The various data obtained from the 20 Ton capacity industrial ice plant are as shown in table-1

Table: 1. Data for Single Stage Vapour Compression Industrial Ice Plant

Observation	Data
Refrigerant	NH ₃
Cooling load	20 T
Saturated vapour evaporating temperature	-24 °C
Saturated vapour condensing temperature	22 °C
Ambient temperature	17 °C
Temperature of ice	-5 °C
Brine water temperature	-10 °C
Reciprocating compressor efficiencies	
Isentropic efficiency	0.73
Mechanical efficiency	0.80
Electrical efficiency	0.90

It should be noted that the above discussion is based on the ideal vapor-compression refrigeration cycle which does not take into account the frictional pressure drop and heat loss in the system, slight internal irreversibility during the compression of the refrigerant vapour, or non-ideal gas behaviour (if any).

3. Irreversibility Analysis of Ice Plant

To facilitate the thermodynamic analysis, simplified schematic of the present single stage vapour compression industrial plant with control volume is shown in Figure 3, where Q_0 is the heat rejection rate from the condenser, Q_c is the heat input rate (cooling load) to the evaporator and W is the work input required to the compressor. For the thermodynamic analysis of the industrial ice plant working on vapour compression system (shown in Figure 3), the principles of mass conservation, first and second laws of thermodynamics are applied to each component of the system. Each component can be treated as control volume with inlet and outlet streams, heat transfer and work interactions. For the system, the mass conservation is governed by following equation:

$$\sum \dot{m}_i - \sum \dot{m}_o = 0 \quad (1)$$

Where m is the mass flow rate and suffix i and o represents inlet and outlet of the component.

The first law of thermodynamics yields the energy balance of each component of the system as follows:

$$\sum (\dot{m}h)_i - \sum (\dot{m}h)_o + [\sum \dot{Q}_i - \sum \dot{Q}_o] + W = 0 \quad (2)$$

The overall energy balance of the system requires that the sum of the evaporator, condenser and compressor must

be zero. The vapour compression system is assumed to be in steady state condition and further if the pump work for brine solution circulation and the environmental heat losses are neglected, the energy balance for the entire system can be written as

$$Q_0 = Q_c + W \quad (3)$$

The energy balance equations of various components of a vapour compression system are given below:

Mass flow rate of circulated refrigerant can be calculated as

$$\dot{m}_R = Q_c / (h_1 - h_4) \quad (4)$$

For the compressor (by convention, the work done by compressor is presumed to be negative);

$$W_c = - \dot{m}_R (h_2 - h_1) / \eta_m \eta_e \quad (5)$$

Where η_m and η_e are mechanical and electric motor efficiencies respectively

Heat transfer rate in the condenser (zone B) is given by

$$Q_0 = \dot{m}_R (h_3 - h_2) \quad (6)$$

Coefficient of performance of refrigeration is

$$COP = Q_c / W \quad (7)$$

For state point 4 dryness fraction and specific entropy can be represented as:

$$x_4 = (h_3 - h_f) / (h_1 - h_f) \quad (8)$$

Where h_f is the enthalpy of saturated liquid refrigerant at evaporator pressure

Specific entropy is:

$$s_4 = s_f + x_4 (s_1 - s_f) \quad (9)$$

Where s_f is the entropy of saturated liquid refrigerant at evaporator pressure.

Second law analysis is a relatively new concept, which has been used for understanding the irreversible nature of real thermal processes and defining the maximum available energy. The second law analysis is based on the concept of exergy, which can be defined as a measure of work potential or quality of different forms of available energy relative to the environmental conditions. In other words, exergy can be defined as the maximum theoretical work derivable by the interaction for an energy resource with the environment.

Exergy analysis is applied to a system describes losses both in the components of the system and for the system as a whole. With the help of exergy analysis the magnitude of these losses or irreversibility's and their order of importance can be understood respectively. With the use of irreversibility, which is a measure of process imperfection, the optimum operating conditions can be easily determined. It is possible to say that exergy analysis throw an insight to indicate the possibilities of thermodynamic improvement for the process under consideration. The formulation for exergy analysis is described below:

The difference of the flow availability of a stream and that of the same stream at its restricted dead state is called flow exergy (ϵ) and by ignoring chemical exergy terms, flow exergy (ϵ) is given by

$$\epsilon = (h - T_0 s) + V^2/2 + gZ - (h_0 - T_0 s_0) \quad (10)$$

Ignoring the potential and kinematic energy terms, Eq. (10) becomes

$$\epsilon = (h - T_0 s) - (h_0 - T_0 s_0) \quad (11)$$

The exergy balance equation is given by

$$E_w = \sum E_Q + \sum (\dot{m}\epsilon)_i - \sum (\dot{m}\epsilon)_o + T_0 S_{gen} \quad (12)$$

In equation (12) the term $T_0 S_{\text{gen}}$ is defined as the irreversibility (I) and can be written as:

$$I = T_0 S_{\text{gen}} \quad (13)$$

The above exergy analysis formulation has been performed on each component of the vapour compression system shown in Figure 3, and corresponding irreversibility of each component is calculated. Using this formulation described below:

By carrying out an exergy-rate balance for the compressor, the irreversibility rate (zone A):

$$I_A = W + E_1 - E_2 \quad (14)$$

The exergy-rate balance for the evaporator (zone D):

$$I_D = E_4 - E_1 - E_Q^D \quad (15)$$

Where, for the evaporator

$$E_Q^D = Q_C (T_0 - T_C) / T_C \quad (16)$$

The exergy rate balance for the condenser, (zone B) is given by

$$I_B = E_2 - E_3 \quad (17)$$

The exergy-rate balance in the throttling valve, (zone C) is given by

$$I_C = E_3 - E_4 \quad (18)$$

Efficiency defect (δ_k) of k^{th} component of the system may be expressed as fractions of input which are lost through irreversibility

$$\delta_k = I_k / W \quad (19)$$

Where I_k is the irreversibility rate of the k^{th} component of the system under consideration

The total irreversibilities of the system components is expressed as

$$I_t = I_A + I_B + I_C + I_D \quad (20)$$

The relative irreversibility of the k^{th} component of plant is

$$= I_k / I_t \quad (21)$$

Structural coefficients are used in the study of the system structure, optimization of plant components and product pricing in multi-product plants. The change of local irreversibility rates and exergy fluxes in relation to the overall plant's irreversibility rate is effectively expressed by the coefficient of structural bonds (CSB) which is defined by

$$\sigma_{k,i} = [\partial I_t / \partial x_i] / [\partial I_k / \partial x_i] \quad (22)$$

Where x_i is the i^{th} parameter of the system which produces the changes in k^{th} component. The effect of a change in x_i on the system would be to alter the rate of exergy input while leaving the output constant. This acceptance confirms to the usual practice of specifying a plant in terms of its output rather than its input. From the exergy balance of the system

$$E_{\text{IN}} = E_{\text{OUT}} + I_t \quad (23)$$

But $E_{\text{OUT}} = \text{constant}$, thus

$$\Delta E_{\text{IN}} = \Delta I_t \quad (24)$$

As seen from equation (24) changes in the irreversibility of the system are equivalent to changes in the exergy input. In general, the ratio of the rates of exergy output to exergy input is less than unity. This ratio denotes the degree of thermodynamic perfection of the process and is called the rational efficiency (η_R).

$$\eta_R = E_{\text{OUT}} / E_{\text{IN}} \quad (25)$$

Plant rational efficiency

$$\eta_{R, \text{plant}} = 1 - \sum \delta_k \quad (26)$$

4. Result and Discussions

Based on the operating plant data shown in Table 1, enthalpy, entropy and exergy at different state points are determined using REFPROP (Table-2)

Table: 2. Thermodynamic properties at state points

State	Temperature (K)	Pressure (bar)	Enthalpy (kJ/kg)	Entropy (kJ/kg K)	Exergy (kJ)
1	249	1.57	1.575	6.4407	88.35
2	410.98	9.09	1.827	6.4407	105.85
2'	295	9.09	1.624	5.8266	102.74
3	295	9.09	0.445	1.8301	101.47
4	249	1.57	0.445	1.9055	100.09

Table: 3. Efficiency defects expressed as percentage for various values of condenser temperatures for a given evaporator temperature = 249 K

Component	305	303	301	299	297	295
Evaporator	15.25	15.95	16.71	17.53	18.42	19.38
Condenser	9.55	8.7	7.75	6.69	5.50	4.15
Compressor	41.39	41.58	41.78	41.98	42.19	42.4
Throttling valve	5.45	5.27	5.08	4.9	4.72	4.54
Sum	71.64	71.5	71.32	71.1	70.82	70.48
η_R	28.36	28.5	28.68	28.9	29.18	29.52

Table: 4. Efficiency defects expressed as percentage for various values of evaporator temperatures for a given condenser temperature = 295 K

EVAPORATOR TEMPERATURE (K)						
Component	249	247	245	243	241	239
Evaporator	19.38	20.34	21.18	21.92	22.57	23.14
Condenser	4.15	3.92	3.71	3.51	3.33	3.16
Compressor	42.4	42.18	41.96	41.74	41.52	41.3
Throttle valve	4.54	4.69	4.84	4.99	5.14	5.29
Sum	70.48	71.13	71.69	72.17	72.56	72.89
η_R	29.52	28.87	28.31	27.83	27.44	27.11

In the present study the condenser and evaporator temperatures are varied between 295 K to 305 K, and 249 K to 239 K, respectively with an interval of 2 K. Table 3 and 4 represents the efficiency defects in percentages for various values of condenser and evaporator temperatures respectively.

As shown in Table 3, when the condenser temperature is varied from 295 K to 305 K, the total efficiency defect in the system increases from 70.48 % to 71.64 %. Whereas the plant's rational efficiency decreases from 29.52 % to 28.36

%. The order of efficiency defects (maximum to minimum) in system components are compressor, evaporator, condenser and throttle valve at 295 K. The efficiency defects in condenser and throttle valve increases with condenser temperature whereas for evaporator and compressor, efficiency defects decreases with increase in condenser temperature. As shown in Table 4, when the evaporator temperature is varied from 249 K to 239 K, the total efficiency defect in the system increases from 70.48 % to 72.89 % whereas the plant's rational efficiency decreases from 29.52 % to 27.11 %. The order of efficiency defects (maximum to minimum) in system components are compressor, evaporator, throttle valve and condenser for all evaporator temperature. The efficiency defects in evaporator and throttle valve increases with decrease in evaporator temperature whereas for condenser and compressor, efficiency defects decreases with decrease in evaporator temperature.

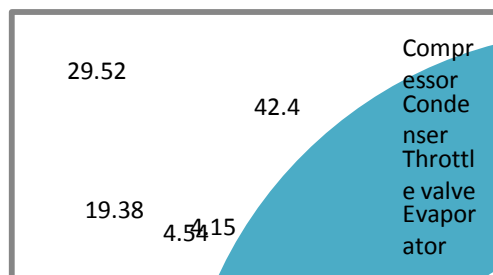


Fig. 3. Control volume representation of the single stage vapour compression system of industrial ice plant

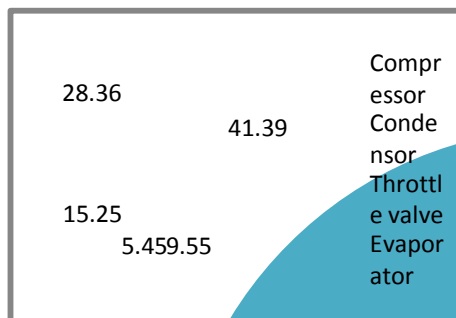


Fig. 4. Pie chart of efficiency defects for a condenser temperature of 295 K

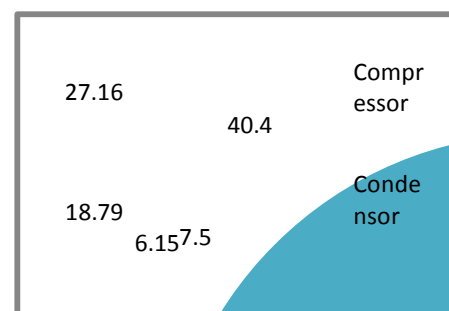


Fig. 5. Pie chart of efficiency defects for $T_{\text{evap}} = 239$ and $T_{\text{cond}} = 305$ K Slope of the line = 2.58

The three pie charts of Figures 4-6 each show percentages of the input that are lost through irreversibility's in four zones of the plant. There is a direct relationship between the plant's component irreversibility's and their effects on the plant's rational efficiency. The fraction representing the proportion of the input lost through irreversibility's in the sub regions are denoted (usually as percentages) by the appropriate efficiency defect (δ_k). The result presented in Figures 4 and 5 are obtained for condenser saturation temperature from 295 and 305 K respectively, assuming a constant ambient-temperature of 293 K. We see that for $T_{\text{cond}} = 295$ K, the plant's rational efficiency is 29.52 %.

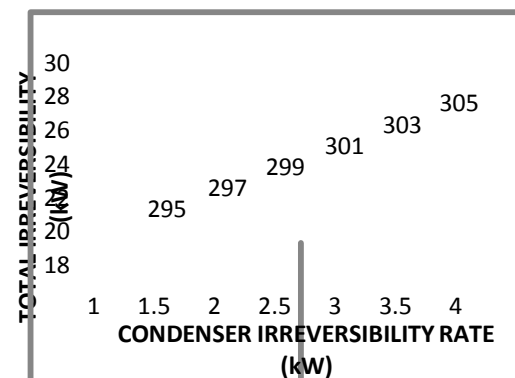


Table 5. Irreversibility rates (kW) for the stated value of condenser temperature (K)

(T _{cond}); [T _{evap} = 249]						
Component	305	303	301	299	297	295
Compressor	15.98	15.35	14.72	14.10	13.49	12.88
Throttle valve	2.1	1.94	1.79	1.64	1.51	1.37
Evaporator	5.89	5.89	5.89	5.89	5.89	5.89
Condenser	3.68	3.21	2.73	2.24	1.75	1.26
Total	27.67	26.40	25.14	23.89	22.65	21.42
Temp (K)	305	303	301	299	297	295

Data given in Figure 6, are determined at the extreme conditions [$T_{\text{evap}} = 239$ K and $T_{\text{cond}} = 305$ K]. At this condition the total efficiency defect in the system is 72.84 % where as the plant's rational efficiency is 27.16 %. The effect of variation of condenser and evaporator temperatures on irreversibility rates of system components are shown in Table 5 and 6 respectively. When the condenser temperature is varied from 295 K to 305 K, the total irreversibility rate in the system increases from 21.42 kW to 27.67 kW. The order of irreversibility rates (maximum to minimum) in system components are compressor, evaporator, throttle valve and condenser at 295 K. The irreversibility rates in compressor, condenser and throttle valve increases with increase in condenser temperature whereas for evaporator, it remains constant.

Table: 6. Irreversibility Rates (kW) for the Stated Value of Evaporator Temperature T_{evap} ; for Condenser Temperature $T_{\text{cond}} = 249\text{K}$

Component	249	247	245	243	241	239
Compressor	12.88	13.61	14.35	15.12	15.91	16.71
Throttle valve	1.37	1.51	1.65	1.8	1.96	2.13
Evaporator	5.89	6.56	7.24	7.94	8.64	9.36
Condenser	1.26	1.26	1.26	1.26	1.26	1.26
Total	21.42	22.95	24.53	26.14	27.8	29.5
Component	249	247	245	243	241	239

As shown in Table 6, when the evaporator temperature is varied from 249 K to 239 K, the total irreversibility rate in the system increases from 21.42 kW to 29.5 kW. The order of irreversibility rates (maximum to minimum) in system components are compressor, evaporator, throttle valve and condenser. The irreversibility rates in compressor, evaporator and throttle valve increases with decrease in evaporator temperature whereas for condenser, it remains constant. Figure 7 shows the variation of total irreversibility rate with condenser's irreversibility rate for different condenser temperature. As the condenser temperature increases, the condenser's irreversibility rate increases. For a unit change in condenser's irreversibility, the total irreversibility of the plant increases by 2.58. Figure 8 shows the variation of total irreversibility rate with evaporator's irreversibility rate for different evaporator temperature. As the evaporator temperature decreases, the evaporator's irreversibility rate increases

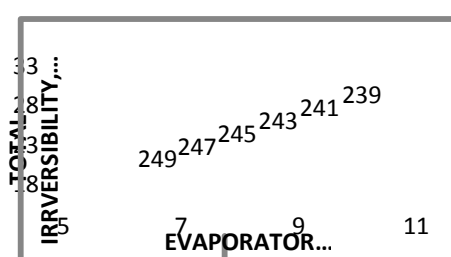


Fig: 8. Plot for Determining the CSB of the Condenser T_{evap} with as a Variable and $T_{\text{cond}} = 249\text{K}$

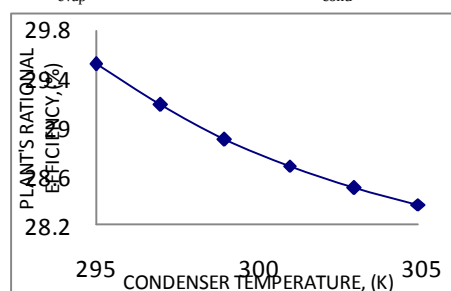


Fig: 9. Plant's Rational Efficiency Dependence upon the Condenser Temperature

For a unit change in evaporator's irreversibility, the total irreversibility of the plant increases by 2.32

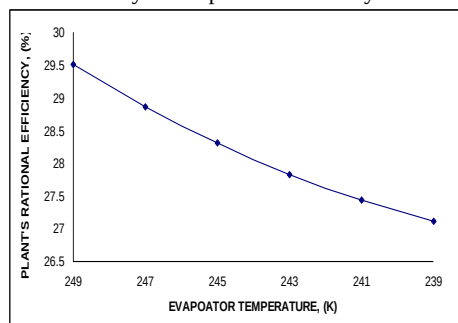


Fig: 10. Plant's Rational Efficiency Dependence upon the Evaporator Temperature

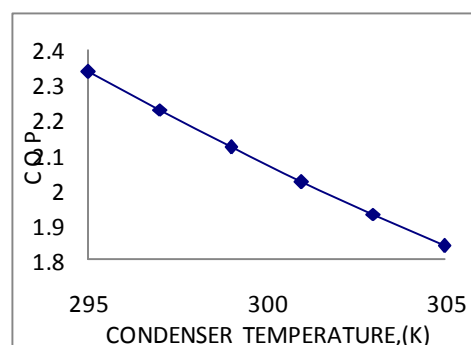


Fig: 11. Plot between the COP vs. Condenser Temperature

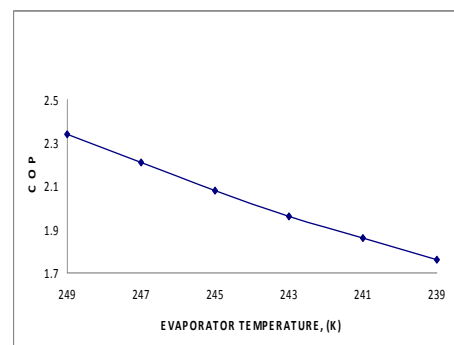


Fig: 12. Plot between the COP vs. Evaporator Temperature

The results presented in Figures 9 to 12 are obtained for different atmosphere conditions and Table 7 and 8 shows the variation of plant's rational efficiency with different temperature conditions.

Table: 7(a). Irreversibility's and Plant's Rational Efficiencies for the Stated Values of T_{cond} ; $T_{\text{evap}} = 249\text{K}$

(T_{cond}) (K)	Total plant's irreversibility rate (I_t) (kW)	Condenser's irreversibility rate (I_c) (kW)	(η) (%)
295	21.42	1.26	29.52
297	22.65	1.75	29.18
299	23.89	2.24	28.9
301	25.14	2.73	28.68
303	26.40	3.21	28.5
305	27.67	3.68	28.36

Table: 7(b). Irreversibility's and Plant's Rational Efficiencies for the Stated Values of T_{cond} ; $T_{\text{evap}} = 249$ K

(T_{cond}) (K)	Total plant's irreversibility rate (I_t) (kW)	Condenser's irreversibility rate(I_c)(kW)	(COP)
295	21.42	1.26	2.34
297	22.65	1.75	2.23
299	23.89	2.24	2.12
301	25.14	2.73	2.02
303	26.40	3.21	1.93
305	27.67	3.68	1.84

Table: 8a. Irreversibility's and Plant's Rational Efficiencies for the Stated Values of T_{evap} for given Condenser Temperature ($T_{\text{cond}} = 295$ K

Evap Temp (T_{evap}) (K)	Total plant's irreversibility rate (I_t) (kW)	Evap's irreversibility rate(I_e)(kW)	(COP)
249	21.42	5.89	2.34
247	22.95	6.56	2.21
245	24.53	7.24	2.08
243	26.14	7.94	1.96
241	27.8	8.64	1.86
239	29.5	9.36	1.76

Table: 8b. Irreversibility's and Plant's Rational Efficiencies for the Stated Values of T_{evap} for $T_{\text{cond}} = 295$ (K)

Evap temp (T_{evap}) (K)	Plant's rational efficiency (η) (%)	Evap. Temp (T_{evap}) (K)	Total plant's irreversibility rate (I_t) (kW)
249	29.52	249	21.42
247	28.87	247	22.95
245	28.31	245	24.53
243	27.83	243	26.14
241	27.44	241	27.8
239	27.11	239	29.5

References

- [1] Moran M J, Availability Analysis: A guide to efficient energy use (Prentice Hall & Englewood Cliffs, NJ) 1982
- [2] Szargut J, et al, Energy analysis of thermal, chemical, and metallurgical processes (Hemisphere publishing corporation, Springer-Verlag, NJ) 1988
- [3] T. J. Kotas, The exergy method of thermal plant analysis, 2nd edn, (Krieger publishing company, USA) 1995
- [4] Bejan et al, Thermal Design and Optimization, John Wiley & Sons, 1996
- [5] M Kanoglu, Exergy analysis of the multistage cascade refrigeration cycle used for natural gas liquefaction, Int J Energy Res. 2002
- [6] A Bejan, Theory of heat transfer irreversible refrigeration plants, Int J Heat mass transfer, 32(9), 1989, 1631-39.
- [7] G Wall, Optimization of refrigeration machinery, Int J Ref, 14, 1999, 336-340

The plant's rational efficiency and the COP decreases with increase in condenser temperature. Also with decrease in evaporator temperature with condenser temperature kept constant plant's rational efficiency decreases. The result indicates that available part of input energy (exergy) continuously decreases with increase in condenser temperature and decrease in evaporator temperature. From the pie chart (figure 3) we can see the percentages of irreversibility rate of different component of the system. By comparing different pie chart it is clear that contribution of compressor, evaporator, and condenser in total irreversibility is much larger than contribution of throttling device and also as the temperature of condenser rises, the irreversibility of each component rises. It can also be found out from the Figure 7 that there is direct linear relationship between total irreversibility and condenser irreversibility as the condenser temperature rises keeping the evaporator temperature constant. We have also found the effect of rational efficiency on condenser temperature keeping evaporator temperature constant. And it is being found out from the results that rational efficiency of plant decreases with increase in condenser temperature.

5. Conclusions

The analysis of the single stage vapour compression refrigeration plant's performance by the exergy method demonstrates how effective this method is for analyzing behavior. Employing the concepts of efficiency defect and rational efficiency has enabled the proportions of input lost through irreversibility's, in various plant sub regions, to be evaluated easily.

Using the technique of the coefficient of structural bond has demonstrated that a change in any component variable in a plant component significantly influences the other plant component in the plant as a whole, and a reduction of irreversibility's rate in a plant component give a greater reduction in the irreversibility rate of plant as whole. The greater the value of ΔT , the greater the irreversibility. Because ΔT_{cond} and ΔT_{evap} affect the plant's rational efficiency, they need to be optimized each particularly heat transfer area chosen for the two heat exchanger.

- [8] C. Nikolaidis, D. Probert, Exergy method analysis of a two stage vapour-compression refrigeration-plants performance, Applied energy, 60, 1998, 241-256
- [9] R Yumrutas, M. Kunduz, M Kanoglu, Exergy analysis of vapour compression refrigeration systems, Exergy, an Int J, 2, 2002, 266-272.

Total Quality Management in Technical Educational System

R. S. Mishra*, Ashok Kumar Madan

Department of Mechanical Engineering, Delhi Technological University, New Delhi, India

Article Info

Article history:

Received 25 September 2014

Received in revised form

15 October 2014

Accepted 25 November 2014

Available online 15 December 2014

Keywords

Quality management,
Technical education,
Institutional improvement

Abstract

The Government of India is utilizing huge amount on the improvement in quality of technical education in the country. The various modeling techniques are available in the literature to evaluate the quality in technical education system are described. The best criteria for judging for best institution in the university system have also been suggested in this paper.

1. Introduction

Education is an important aspect for the development of economy and to be self sustained. It predicts the growth and living standard. Technical education enhances further economic growth and provides livelihood to the needed. Quality education is an absolute necessity at every level of society. Parents are interested to enroll their children in the best college/institute in order to provide them quality education at reasonable costs. The main stakeholders of any TES are faculty, students, management and the infrastructure, which are responsible for efficient functioning of a TES. There is an utmost need to provide good quality technical education. In order to achieve it there is a need to develop certain procedures which would help in assessing the standard and quality of technical education so that its quality can be improved and monitored from time to time.

In this work we have done research on various aspects of technical education by using different techniques. The basis of the research is provided by the literature review done.

- “Quality” means those characteristics of processes that satisfy the needs of customers (students) and thus ensure their satisfaction. The purpose of such higher quality is to ensure a greater satisfaction of students.
- “Quality” means lack of deficiencies – lack of such errors which result in exploitation, non satisfaction of faculty/staff and students.

In education, quality means good academic culture, excellent academic results, progressive and adaptive management, clean administration and prominent profile of outgoing students. It involves the expectations and perceptions of students, faculty, supporting staff, administrators, parents of the students, government, industry (recruiters) and society etc. They interact with the system in different ways and their objectives may be different. So the implementation of a quality improvement program necessitates the identification of various factors in an educational set-up and determination of their criticality.

Corresponding Author,

E-mail address: professor_rsmishra@yahoo.co.in

All rights reserved: <http://www.ijari.org>

The stakeholders are classified into three groups: input, transformation and output. Students and parents are included in input stakeholders, the faculty is the transformation stakeholders and the corporations and society are the output stakeholders. The main objective of a TES is the development of methodologies for improving the quality of education and establishment of a new system of its own

2. Technical Education Present-Past Scenario

Education depends on several mechanisms. A technical education system (TES) or process consists of three different stages, such as input, process and the output with a feedback mechanism which makes it a closed loop shown in fig-1.

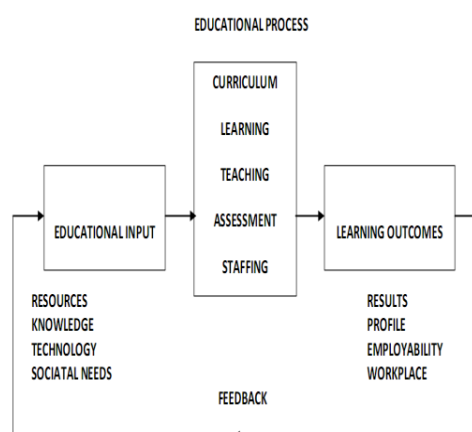


Fig: 1. Educational process^[1]

Quality is a key factor for determining the success and survival. It is an important factor in manufacturing and service industries and has been extended to other fields across the world. Quality of a product is very important for an enterprise, it acts as a lifeline for the industries and hence they employ the best possible means to ensure quality of their product. The International Standard specifies

requirements for a quality management system where an institute:

- i. Needs to motivate its ability to consistently provide knowledge to students that meets the international standards and are able to provide good quality service to the nation and society to grow.
- ii. Aims to quality satisfaction through effective application of the system, including processes for continual improvement of the system and the assurance of conformity to students and top authority.

3. Application of Total Quality Management in Technical Educational System

Quality is an attitude best defined not by the head of the institution/university but by those the system serves i.e. students. Various definitions of quality are:

- Deming defines quality as meeting and exceeding the customer's needs and expectations-and then continuing to improve.
- Juran defines it as fitness for use-does it fit the customer?

The American Society for Quality Control states that: "quality is the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs".

4. Parameters for total Quality Management in Technical Educational System

The important parameters in quality in technical education are as follow:

- Training on state-of-the-art technology.
- Comprehensive learning resources.
- Opportunities for campus training & placement.
- Close supervision of students' work.
- Expertise in subjects and well-organized lectures.
- Good communication skill of academic staff.
- Well-equipped laboratories with modern facilities.
- Design of course structure based on job requirements.
- Encouragement for sports, games and cultural activities.
- Cleanliness, orderliness, systematic and methodical.
- Available regularly for students' consultation.
- Effective classroom management.
- Recognition of the students.
- Adaptability to modern techniques.

These parameters have been considered further in analysis of quality in technical education system.

The major thrust areas to improve the quality of technical educational institutions are

- (i) Students Quality
- (ii) Faculty Quality
- (iii) Management Quality
- (iv) Infrastructure in an Institution
- (v) Accountability

The analogy between manufacturing system and educational system are shown in fig-2.

5. Literature Review

The literature is replete with various works bordering on university admission, student performance, and related problem. In 1954, the University of New Zealand Council for Educational Research investigated the relationship between academic standards of students on entrance and their first year university work. The study found that the median correlation found among the many sets of variables representing general school performance and general university performance was indicated by a *tau* coefficient of 0.36 for the first year students undertaking their studies on a full time basis.

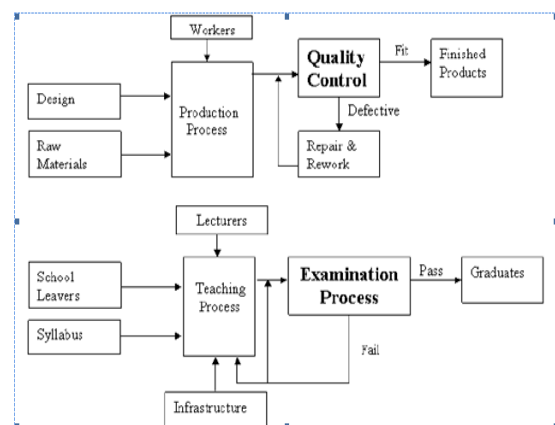


Fig. 2. Analogy between Manufacturing System and Educational System^[1]

In 1975, Bakare (2) summarized the factors and variables affecting student's performance into the intellectual and non-intellectual factors, emphasizing that the intellectual abilities were the best measure (Bakare 1975). He categorized causes of poor academic performance into four major classes:

- (i) Causes resident in society
- (ii) Causes resident in school
- (iii) Causes resident in the family
- (iv) Causes resident in the student. Studies such as (Lage and Tregelia, 1996)

(4) and (Dyran, 1977) (3) looked at a more general aspects of success while Anderson et al., 1994 (5) studied the effect of factors such as gender, student age, and students' high school scores in mathematics, English, and economics, on the level of university attainment. According to their study, students who received better scores in high school also performed better in university. Another aspect discovered was that men had better grades than women and choose to drop from school less often. Johnes J. (6) analysed the teaching efficiency of the teachers using Data envelopment analysis (DEA) technique. The results suggested that efficiencies derived from DEAs performed at an aggregate level include both institution and individual components, and are therefore misleading. Temponi C. (2005) (7) analyze the main elements of continuous improvement (CI) in higher education and the concerns of academia's stakeholders in the implementation of such an approach. Thakkar J. et al (2006) (8) used a quality function deployment (QFD) which prioritizes technical requirements and correlates them with various customers students requirements for the present Indian context. As an extension

to the basic model of QFD house of quality (HOQ), the scope for futuristic improvements is explored through a four-phased QFD process. Challenges involved in the implementation of TQM are investigated using an approach of force field analysis. They recognized the need for continuous improvement, cultural change and effective use of financial resources to improve the value addition at each level. They developed an understanding of the issues to be addressed at each phase of TQM implementation. Mahapatra S. S. and Khan M. S. (2007) (9) gave a measuring instrument known as EduQUAL for evaluation of quality in Technical Education System (TES). They carried out a Factor analysis on responses obtained through cross-sectional questionnaire survey on various items to validate dimensionality of the instrument and it is found that 28 items loaded above 0.5. Neural network models have been proposed to assess the degree of satisfaction of various stakeholders in TES. Mahapatra S. S. and Khan M. S. (10) designed a measuring instrument known as EduQUAL and an integrative approach using neural networks for evaluating service quality is proposed. The dimensionality of EduQUAL is validated by factor analysis followed by vari-max rotation. Four neural network models based on back-propagation algorithm are employed to predict quality in education for different stakeholders. The study demonstrated that the P-E gap model is found to be the best model for all the stakeholders. Sensitivity analysis of the best model for each stakeholder was carried out to appraise the robustness of the model. Finally, areas of improvement were suggested to the administrators of the institutions. Cristea L. and Gogoncea D. (11) applied fuzzy approach in quality management of higher education. They concluded that the adoption of the fuzzy formalism is a possible solution to the standardization in the domain of quality, in which to the usual terms are still given various meanings by the managers of various firms, on markets that visibly oscillate between globalization and regionalization.

The literature review shows the importance of quality management in technical education. It also shows the need and importance of quality management. Hence we can conclude that there is an immense scope of research in the field as very nominal amount of work has been done in this area of research.

The literature review concludes that there is an utmost need of quality management in technical education. It shows the extensive work carried in the field of research. It also shows the need and importance of quality management in technical education. The literature review also states the various techniques used for quality management in technical education.

Various scientists have done a lot of research in the field of technical education. The research is done for prediction of quality of technical education, quality assessment and quality management in technical education. Various techniques have been employed for this purpose but no technique is found to be very accurate. Techniques like ANN, Fuzzy Logic, SPC, Fuzzy AHP and ISM have not been used for quality management in technical education. Hence, there is need to apply these techniques and find whether these techniques can be used for quality management in technical education or not.

6. Factors Affecting Total Quality in Technical Educational Systems

Some of the factors which motivated me to undertake the subject research work are as under:

- Due to globalization we need more and more technical manpower of high quality in the industries, research centers and institutions.
- The number of the technical institutions is increasing exponentially under the govt. efforts of liberalization and reform of higher technical education to make India the largest English speaking technical manpower of the world.
- Industry is becoming more competitive locally as well as globally and for that we need higher number of technical manpower at all levels. For this improvement in the quality of TE is must and is a need of hour.
- These institutes are merely producing technical graduates which do not meet the expectations of the industries.
- The rapid growth in the knowledge and technology makes the technical curriculum obsolete quickly. Therefore, technical institutions, industries and research centers need to maintain close contacts, inter-academia partnership and higher quality standards in technical education.
- Under the present scenario and for the reasons already stated, a lot of sub-standard technical institutional have already come up and many are in the pipeline to get established.
- Further there is large amount of technical students and professionals across national borders making it imperative and concerned agencies not to be restricted to national considerations only. Therefore, high quality standards have to cater for global processes of internationalism in TE.
- However, at present very little amount of research work seems to have taken place in this field.

The main Objectives of the present research study are

1. To develop quality management models for quality management and assessing in technical education which includes self assessment models, placement model, performance model, determination of attributes for a technical education system.
2. To develop strategies for improving the productive behavior and quality in technical education.
3. To develop competitive indices in terms of ranking of technical institutions in India.
4. To show the validity of application of the soft computing and artificial intelligence techniques in the field of quality management in technical education.

7. Research Methodology

Quality control in technical or higher education is a major issue from past decades. Research has been going on in this field for years and different types of techniques or methods have been employed by the researchers in order to achieve desired results.

Some of the techniques used are:

- Artificial Neural network

- Fuzzy logic
- Analytical Hierarchy Process (AHP)
- Statistical Process Control (SPC)
- Interpretive Structural Modeling (ISM)

The above techniques are used very effectively and have also proven to give good results and hence helped various institutions to control the quality of technical education all over the world.

In this work we have used various methods for improving and analyzing the quality of technical education. The techniques have been discussed below.

7.1. Artificial Neural Network

Inspired by the structure of the brain, a neural network consists of a set of highly interconnected entities, called Processing Elements (PE) or units. Each unit is designed to mimic its biological counterpart, the neuron. Each accepts a weighted set of inputs and responds with an output. Neural networks address problems that are often difficult for traditional computers to solve, such as speech and pattern recognition, weather forecasts, sales forecasts, scheduling of buses, power loading forecasts, early cancer detection, etc. A neural network is a more general method of regression analysis. Some of the advantages of the network over conventional regression include the following:

- There is no need to specify a function to which the data are to be fitted. The function is an outcome of the process of creating a network.
- The network is able to capture almost arbitrarily nonlinear relationships.
- With Bayesian methods, it is possible to estimate the uncertainty of extrapolation.

7.2. Fuzzy Logic

Logic started as the study of language in arguments and persuasion, and it may be used to judge the correctness of a chain of reasoning, in a mathematical proof for example. In two valued logic a proposition is either true or false, but not both. The "truth" or "falsity" which is assigned to a statement is its truth value. In fuzzy logic a proposition may be true or false or have an intermediate truth-value, such as may be true. The sentence the level is high is an example of such a proposition in a fuzzy controller. It may be convenient to restrict the possible truth values to a discrete domain, say (0, .5, and 1) for false, may be true and true in that case we are dealing with multi valued logic. In practice a finer subdivision of the unit interval may be more appropriate. New product (read courses) development (NPD) is closely linked to an institute's competitiveness. Managing NPD is complex and requires consideration of customer (read student) requirements, technical issues, and competing courses and curriculums. The more closely the course fits the students' expectations, the greater the likelihood of successful course and curriculum development. Quality function deployment (QFD) is a well-known tool for identifying customer needs and translating customer requirements into a technical response. QFD translates customer requirements into technical specifications appropriate for each stage of product development and production. QFD considers customer requirements by examining development space as well as product differentiation, position, and characteristics.

Moreover, QFD can enable businesses to integrate R&D, manufacturing, and management when drafting a marketing policy. QFD is based on the construction and analysis by the house of quality (HOQ), which documents the transformation of customer needs into technical specifications.

The competitive evaluation of HOQ ranks each customer requirement to combine the data of each competing product. Corporations can then employ the combined data for product differentiation and positioning. Importance ratings represent the relative importance of each customer requirement, although assigning ratings to customer requirements is sometimes made difficult by issues of objectivity and significance. Previous investigations have ranked the importance of customer requirements by focus group opinion, expert opinion, and analytical hierarchy process (AHP) analysis. Nevertheless, ranking techniques used in the past may be subjective, complex, controversial, and time-consuming. This study integrates fuzzy logic to rank each customer requirement item and calculate evaluating data in order to analyze product features and conduct product positioning more simply, accurately, objectively, and scientifically.

7.3. Fuzzy AHP

Many fuzzy AHP methods have been proposed by various authors. These methods are systematic and useful approaches to the alternative selection and gives justification to the problem by using the concepts of fuzzy set theory and hierarchical structure analysis. Decision makers have experienced that it is more confident and easy to give interval judgments than fixed value judgments. This is due to the fuzzy nature of the comparison process. The Fuzzy-AHP methodology extends Saaty's AHP by combining it with the fuzzy set theory. In the Fuzzy-AHP, fuzzy ratio scales are used to indicate the relative strength of the factors in the corresponding criteria. Therefore, a fuzzy judgment matrix can be constructed. The final scores of alternatives are also represented by fuzzy numbers. The optimum alternative is obtained by ranking the fuzzy numbers using special algebra operators. The next three steps can summarize the procedure of applying Fuzzy-AHP:

- Construct a hierarchical structure for the problem to be solved.
- Establish the fuzzy judgment matrix and a fuzzy weight vector.
- Rank all alternatives and select the optimal one.

In this methodology, all elements in the judgment matrix and weight vectors are represented by triangular fuzzy numbers. Using fuzzy numbers to indicate the relative contribution or impact of each alternative on a criterion, a fuzzy judgment vector is then obtained for each criterion. The fuzzy judgment matrix A is built with all the fuzzy judgment vectors. The weight vector W is used to represent the decision maker's opinion of the relative importance of each criterion during the decision process. A fuzzy number $\tilde{x}$ expresses the meaning 'about x'. Each membership function is defined by three parameters of the symmetric triangular fuzzy number, (l, m, r), left, middle and right points of the range over which the function is defined. Fuzzy membership function and the definition of a fuzzy number are shown in Fig. 1. When the decision-maker

faces a complex and uncertain problem and expresses his/her comparison judgments as uncertain ratios, such as 'about two times more important', 'between two and four times less important', etc., the standard AHP steps, and specially, eigen value prioritization approach, cannot be considered as straightforward procedures. Indeed, the assessment of local priorities, based on pair wise comparisons needs some prioritization method to be applied.

7.4. Statistical Process Control (SPC)

SPC stands for Statistical Process Control, which the statistical technology and methodology are applied to monitor the quality of product during manufacturing process in real time, it could distinguish and pick up the abnormal deviation of quality from the normal deviation scientifically and precisely, so that a certain kind of early warning could be given when an abnormal is found in manufacturing process, and a certain measures would be taken by relevant persons, e.g. try to find the causes, try to eliminate the abnormal, and try to restore a stable process, etc. it's necessary for a enterprise to achieve the targets of quality control and improving. Walter Shewhart, the founder of SPC, had presented some famous comments on application of SPC:

There are 2 factors appeared in deviation for all manufacturing processes, one is a stable factor, the normal deviation, which is caused by process itself and another is an interrupted factor, the abnormal deviation, which its causes could be found out."

The abnormal deviation can be found out and eliminated by some effective methods, but the normal deviation will never be disappeared unless the basic manufacturing process is altered."

The 3σ SPC control charts could be used to distinguish the abnormal deviation from normal deviation."

SPC is not only a tool for identifying trends or changes, but also provides relevant information whether a process is in control or not. The data obtained from questionnaire are plotted about the mean, range, or proportion. Using the appropriate control chart, one can make diagnosis about the process. If a sample exceeds a control limit, there is strong possibility that an assignable cause exists for this variation, such as major differences attributed to the implementation of a new initiative. If a sample does not exceed a limit, then sample-to-sample variation may just be due to common cause variation. By allowing variations to be examined in a logical manner, control charts can provide engineering educators with the information needed to make a systematic change. SPC has been widely used in quality management for enterprises; it has been proved effective which relevant means of SPC are applied to monitor production process. To promote the teaching quality is one of the main objectives for colleges and universities in current period. Teaching means process.

7.5. Interpretive Structural Modeling (ISM)

Interpretive Structural Modeling is an interactive learning process. The method is interpretive in that the group's judgment decides whether and how items are related, it is structural in that, on the basis of the relationship, an overall structure is extracted from the complex set of items, and it is modeling in that the specific

relationships and overall structure are portrayed in a diagraph model. ISM methodology helps to impose order and direction of relationships among elements of a system (Sage, 1977). It provides us a means by which order can be imposed on the complexity of such variables (Mandal and Deshmukh, 1994; Jharkharia and Shankar, 2005). However, the direct and indirect relationships between the factors describe the situation far more accurately than the individual factor taken in isolation. Therefore, ISM develops insights into collective understandings of these relationships. The application of ISM helps to reassess perceived priorities and improve their understanding of the linkages among key concerns.

8. Results and Discussions

The research and methodology which gives a description of all the techniques used for research work. It also consists all the case studies done for quality management in technical education. A case study was done on comparison of different colleges under one university. The comparison has been done on the basis of various parameters like ISO certification of college, department and labs, Accreditation of college, department, laboratory, UG courses and PG courses, research activities in college and department. UG and PG courses offered both in the department, passing percentage of the students both at UG and PG level in each branch for the college and department, number of PHD holders(faculty) in the college and department, student placement, research scheme offered books published, research paper publication, etc. With regards to the ranking of the engineering institutes in a university system following parameters of acceptance have been considered. If the ranking is $> 80\%$: it is center of excellence. If the ranking is $60\% - 79\%$: the center/institute is of grade A.

If the ranking is $50\% - 59\%$: the center/institute is of grade B. If the ranking is $40\% - 49\%$: the center/institute is of grade C. If the ranking is $< 40\%$: the center/institute is of grade D and it needs to be improved to enhance the quality of the engineering education. The following observations are.

The research resulted in successful implementation of all the above mentioned techniques in the field of technical education. This gives us a new breakthrough in this field of research and also shows the efficient results that were obtained using the above mentioned techniques. The use of the techniques namely, ANN, Fuzzy Logic and MATLAB, SPC, Fuzzy AHP and ISM have greatly helped in assessing and controlling the quality of technical education. Results from each technique give us unique results and hence it is very helpful. The research depicts the need for good quality education and also shows its importance. The techniques used in the research are the latest techniques and have not been applied extensively in this field. The results obtained were found to be very effective and satisfactory. An accuracy of over 73% was achieved by the application of Artificial Neural Network technique. This shows the efficiency of the ANN methodology as a prediction tool for selection criterion for candidates seeking admission into a university. One limitation of this model stems from the fact that not all the relevant performance influencing factors are obtainable from the pre-admission record forms filled by the

students. A model incorporating the use of results from a carefully designed oral interview administered to the students may likely be an improvement over the present model. Also the extension this research to non-engineering departments is recommended.

The Fuzzy Logic technique has been applied for prediction of the various factors of technical education system. A non traditional approach has been proposed to infer statistical and Fuzzy rules from quantitative database. Fuzzy Logic tool of MATLAB software was used for the analysis work. Each factor was assigned with several fuzzy sets. Using fuzzy set concepts, fuzzy rules were inferred then Mat Lab Fuzzy logic tool box is used for generating rules. This approach suggests that for large data base decision can be taken more effectively than traditional methodology with less mental fatigue. This method is just one of the many methods used to generate rules in an adaptive system. Using MAT LAB technique gives us the precise information about where the problem exists and how much input is desired to improve the quality. Quality of teaching is a major focus for colleges and universities. The management and monitoring of teaching process should be enhanced. SPC is applied to analyze the result of second semester examination of an institute in NCR region, of all the streams held in MAY/JUNE-2009. The technique proves to be effective and the SPC control chart shows the problems occurring in the streams. Fuzzy AHP is an effective MCDM technique and can be used effectively for assessing quality in technical education. Since we are aware of the fact that in today's world, decisions are made in increasingly complex environments. Fuzzy decision making theory can be used for this purpose. This research concludes that Fuzzy AHP in an effective MCDM technique and can be applied to the education sector for assessing quality in technical education. We can conclude that of the four Main Attributes, Faculty Quality is the most important, followed by Management Inputs, then Students Quality and in the end Infrastructure. In terms of the Sub Attributes we can conclude some of the very important Sub Attributes are (global weight >0.04) as follows: Good Communication Skills (GCS), Curriculum Design (CD), Qualification of Faculty (Qua), Timely Assessment of Faculty & Students (TA F& S), Teaching & Industrial Experience (T&I Ex), Training & Placement (T&P), Background & Merit of Entering Students (B&M ES), Well Equipped Labs and Classrooms (WE L&C), Attitude Towards Learning (ATL). The objective of the Interpretive Structural Modeling (ISM) is to develop a hierarchy of enablers that would help in management of quality in a technical education system. These enablers are important input for upgrading the standard of technical education and to compete with the fast changing technology. Quality of Technical education can improve in effective manner if all the variables are

improved in the given hierarchy. The driver dependence diagram helps to classify various enablers of effective Technical education system. There are no variables in the autonomous cluster, which indicates no variable can be considered as disconnected from the other variables and therefore there is an imperative need to pay an attention to all the identified enablers for improving the quality of Technical education. In the next cluster we have Effective classroom management, Available regularly for students' consultation, Recognition of the students, Close supervision of students' work, Opportunities for campus training & placement, Quality of Technical Education which have low driving power and high dependency that means they are dependent on the driving enablers. There are no linkage enablers which has a strong driving power as well as strong dependence. Thus, it can be inferred that among all the 13 enablers chosen in this study, no enabler is unstable. The driver power dependence diagram indicates that enablers such as Design of course structure based on job requirements, Well-equipped laboratories with modern facilities, Adaptability to modern techniques, Expertise in subjects and well-organized lectures, Training on state-of-the-art technology, Good communication skill of academic staff, Comprehensive learning resources are at the bottom of the model having strong driving power and low dependency. These enablers will help to achieve its desired objective and are classified as independent enablers or drivers.

9. Conclusions and Recommendations

Most of the techniques used in this research work are the latest and indicate their effectiveness in achieving the desired goal in the quality in technical education. The following conclusions have been drawn.

- This research work shows the importance of above methodologies in the field of quality management in technical education.
- This gives us a new breakthrough in this field of research and also shows the efficient results that were obtained using the above mentioned techniques.
- The use of the techniques namely, ANN, Fuzzy Logic and MATLAB, SPC, Fuzzy AHP and ISM have greatly helped in assessing and controlling the quality of technical education.
- The research is limited in this area and there is a scope of future work that can be done in this field of research. Techniques like nanotechnology, genetic algorithm, etc. can also be applied for quality management in technical education. So in future application of many other techniques can also be checked in this area of research.

References

- [1] P. Venkataram, A. Giridharan, Quality Assurance and Assessment in Technical Education System: A Web Based Approach, ICEE 2007
- [2] C. C. Bakare, Some Psychological Correlates of Academic Success and Failure, African Journal of Educational Research, 1975
- [3] K. Dynan, C. Rouse, The Under Utilization Of Women In Economics: A Study, 1997
- [4] M. Lage, M. Tregelia, The Impact Of Integrating Scholarship On Women Into Introductory Economics: Evidence From One Institution, Journal Of Economic Education, 1996, 27:26-36.
- [5] G. Anderson, D. Benjamin, M. Fuss, The Determinant Of Success In University, 1994
- [6] Jill Johnes, Measuring teaching efficiency in higher education: An application of data envelopment analysis toeconomics graduates from UK

- Universities 1993, European Journal of Operational Research, 174, 2006, 443–456
- [7] C. Temponi, Continuous improvement framework: implications for academia, Emerald Group, Quality Assurance in Education, 13(1), 2005
- [8] Jitesh Thakkar, Anil Shastree, Anil Shastree, Total quality management (TQM) in self-financed technical institutions, Emerald Group. Quality Assurance in Education, 14(1), 2006
- [9] S. S. Mahapatra, M. S. Khan, A framework for analysing quality in education settings, European Journal of Engineering Education, 32(2), 2007, 205–217
- [10] S. S. Mahapatra, M. S. Khan, A neural network approach for assessing quality in technical education: an empirical study, International Journal of Productivity and Quality Management 2007, 2(3)
- [11] L. Cristea, Quality Management or Management of Quality in Higher Education, International Conference Man in the Knowledge Based Organization, Sibiu, Romania, 2006, 72-75
- [12] Roediger Voss, Thorsten Gruber, The desired teaching qualities of lecturers in higher education: a means end analysis, Quality Assurance in Education, 14(3), 2006, 217-242
- [13] S. S. Mahapatra, M. S. Khan, A framework for analysing quality in education settings, European Journal of Engineering Education, 32(2), 2007, 205–217
- [14] Yau Tsai, Sue Bevertson, Top-down management: an effective tool in higher education?, International Journal of Educational Management, 1(1), 2007, 6
- [15] H. Ahmad, A. Francis, M. Zairi, Business process reengineering: critical success factors in higher education, Emerald Group, Business Process Management Journal, 13(3), 2007 451-469
- [16] S. S. Mahapatra, M. S. Khan, A neural network approach for assessing quality in technical education: an empirical study, International Journal of Productivity and Quality Management 2007, 2(3)
- [17] Dr. Subrata Das, Dr. Anindya Ghosh, Quality assurance in textile education
- [18] Gitachari Srikanthan, John F. Dalrymple, A conceptual overview of a holistic model for quality in higher education, Emerald Group, International Journal of Educational Management, 21(3), 2007
- [19] Yousif Bahzad, Zahir Irani, Developing a Quality Assurance Model for Small Military Institutions, European and Mediterranean Conference on Information Systems 2008, 25-26 2008.
- [20] C. M. Gheorghe, D. Constantinescu, M. Covrig, The Determination of Factors Underlying Decision Making Process on Quality Improvement in Technical Education, IEEE 2008
- [21] Ziren Wang, Ronghua Liang, Discuss on Applying SPC to Quality Management in University Education”, The 9th International Conference for Young Computer Scientists, IEEE, 2008
- [22] Prof. dr. eng. Lidia Cristea, Lecturer dr. eng. Dan Gogoncea, Fuzzy Approach in Quality Management of Higher Education, The 9th International Conference for Young Computer Scientists, IEEE, 2008
- [23] C. M. Bhatia, Smita Bhatia, An Integrated Approach for Restructuring Higher Technical Education
- [24] Hafiz Muhammad Inamullah, M. Naseeruddin, Ishtiaq Hussain, The Development of Technical Education In Pakistan, International Business & Economics Research, 2009, 8(1)
- [25] C. C. Bakare, Some Psychological Correlates of Academic Success and Failure, African Journal of Educational Research. 2, 1975
- [26] M. Lage, M. Tregelia, The Impact of Integrating Scholarship on Women into Introductory Economics: Evidence from One Institution, Journal of Economic Education. 27, 1996, 26-36
- [27] K. Dynan, C. Rouse, The Under Utilization Of Women In Economics: A Study, 1997
- [28] G. Anderson, D. Benjamin, M. Fuss, The Determinant Of Success In University, 1994
- [29] Muhammad Z Mamun, TQM of the non-govt. universities in Bangladesh, ICMIT 2000
- [30] L. Zadeh, Fuzzy sets, Information Control, 8, 1965, 338–353
- [31] P. J. M. Van Laarhoven, W. Pedrycz, A fuzzy extension of Saaty’s priority theory. Fuzzy Sets and Systems 11, 1983, 229–241.
- [32] C. Kahraman, Z. Ulukan, E. Tolga, A fuzzy weighted evaluation method using objective and subjective measures. In: Proceedings of the International ICSC Symposium on Engineering of Intelligent Systems (EIS’98), 1, University Laguna Tenerife, Spain, 1998, 57–63
- [33] H. Deng, Multicriteria analysis with fuzzy pairwise comparison. International Journal of Approximate Reasoning, 21(3), 1999, 215–231
- [34] M. Lee, H. Pham, X. Zhang, A methodology for priority setting with application to software development process. European Journal of Operational Research 118, 1999, 375–389.
- [35] C.-H. Cheng, K.-L. Yang, C.-L. Hwang, Evaluating attack helicopters by AHP based on linguistic variable weight. European Journal of Operational Research, 116 (2), 1999, 423–435
- [36] K.-J. Zhu, Y. Jing, D.-Y. Chang, A discussion of extent analysis method and applications of fuzzy AHP. European Journal of Operational Research 116, 1999, 450–456
- [37] Cengiz Kahraman, Ufuk Cebeci, Da Ruan, Multi-attribute comparison of catering service companies using fuzzy AHP: The case of Turkey, Int. J. Production Economics 87, 2004, 171–184
- [38] Da-Yong Chang, Applications of the extent analysis method on fuzzy AHP, European Journal of Operational Research 95, 1996, 649-655
- [39] C.-H. Cheng, Evaluating naval tactical missile systems by fuzzy AHP based on the grade value of membership function. European Journal of Operational Research 96 (2), 1997, 343–350
- [40] M. Weck, F. Klocke, H. Schell, E. R. Uenauver, Evaluating alternative production cycles using the

- extended fuzzy AHP method. European Journal of Operational Research 100 (2), 1997, 351–366
- [41] D.-Y. Chang, Extent Analysis and Synthetic Decision, Optimization Techniques and Applications, 1, World Scientific, Singapore, 1992, 352
- [42] D.-Y. Chang, Applications of the extent analysis method on fuzzy AHP. European Journal of Operational Research 95, 1996, 649–655
- [43] A. K. Madan, Mishra R. S. Forecasting the academic results of students using ANN: a case study of an engineering college. ISTE, 2(4), 2010
- [44] A. K. Madan, R. S. Mishra, Prediction of quality in technical education using Fuzzy Logic. BVCOE-JMET, 4(1), 2010
- [45] A. K. Madan, R. S. Mishra, A Fuzzy AHP approach for Assessing Quality in technical education. BVCOE-JMET, 4(1), 2011
- [46] A. K. Madan, R. S. Mishra, Model for evaluation of routing and machine flexibility. BVCOE-JMET, 2(1), 2012

Use of Hydrocarbons in Low Temperature Circuit in Terms of First Law and Second Law Efficiency of Four Stage Cascade Refrigeration of Semen Preservation

R.S. Mishra

Department of Mechanical Engineering, Delhi Technological University, New Delhi, India

Article Info

Article history:

Received 20 October 2014

Received in revised form

30 October 2014

Accepted 20 November 2014

Available online 15 December 2014

Keywords

Four stage cascade Refrigeration system,
Energy –Exergy Analysis,
Low Temperature cascade System,
Irreversibility Prediction

Abstract

This paper mainly deals with the utility of hydrocarbons, R404a and other ecofriendly refrigerants in the low temperature circuit for semen preservation. The energy-exergy analysis have been carried out to find out the feasibility of above ecofriendly refrigerants in the Four stage cascade refrigeration system using R1234ze in high temperature circuit and R134a in primary intermediate circuit and R410a in secondary intermediate circuit. It was observed that the use of hydrocarbon is feasible for improving thermal performance of the above system by considering safety measures. The best second law performance was observed using R600a and low performance was observed using ethylene in the low temperature evaporator circuit without safety measure the system using R404a in low temperature evaporator circuit gives better performance using ethylene in low temperature circuit.

1. Introduction

Low-temperature refrigeration systems are typically required in the temperature range from -100°C to -140°C for applications in food, pharmaceutical, chemical, and other industries. At such low temperatures, single-stage compression systems with reciprocating compressors are generally not feasible due to high pressure ratios. A high pressure ratio implies high discharge and oil temperatures and low volumetric efficiencies and, hence, low COP values. Cascade refrigeration systems can be used to achieve low temperatures, where series of single-stage units are used that are thermally coupled through evaporator/condenser cascades. Each circuit has a different refrigerant suitable for that temperature. The high-temperature circuit uses high boiling point refrigerants such as R-1234ze, and R1234yf and Intermediate circuits thirteen ecofriendly refrigerants such as R-134a, R123, R125, HFC blends (R-507a), ammonia, propane, propylene, hydrocarbons, etc., whereas the low-temperature circuit uses low boiling refrigerants such as ethane, air, methane etc. In a vapour compression cascade refrigeration system where low pressure vapour in the evaporator stage is compressed and then recycled for condensation in the evaporator of a previous stage, the improvement of passing the low pressure refrigerant's vapour from an evaporator stage through a heat exchanger to heated vapour to ambient temperature, compressing said heated vapour, removing the compressor work by passing said compressed vapour through a evaporator, then cooling is by compressed vapour by passing it through heat exchanger in heat exchanger at a low pressure vapour, and condensing it in the evaporator of the next higher temperature cycle of the cascade system and recycling the liquid to the low pressure evaporator stage.

Most of the investigators even did not find out the effect of various approaches on the system performances on three and four stages cascade refrigeration systems and

effect of ecofriendly refrigerants in the primary intermediate temperature circuit and secondary intermediate temperature circuit. This paper mainly deals with thermal performances in terms of exergy destruction ratio by thermodynamically analyzing the use of R1234yf and R1234ze in the high temperature circuits and mainly thirteen eco-friendly refrigerants in the intermediates circuits and Ethane in the low temperature applications.

2. Literature Review

Agnew et al.^[1] explained the performance of a three stage cascade refrigeration systems for low temperature applications. Bansal P.K.^[3] carried out thermodynamic analysis of carbon dioxide–ammonia (R744–R717) cascade refrigeration system. Bhattacharyya et al.^[4] studied a carbon dioxide–propane (R744–R290) optimum cascade evaporating system to define an evaporating temperature of R744 for application in heating circuits. Lee et al.^[5] analyzed a carbon dioxide–ammonia (R744–R717) cascade system thermodynamically to determine the optimum condensing temperature of R744 in the low-temperature circuit. Nikolaidis et al.^[6] studied analytically that change in evaporator and condenser temperatures of two stage vapour compression refrigeration plant using R22 add considerable effect on plant irreversibility. They suggested that there is need for optimizing the conditions imposed upon the condenser and evaporator. Reddy et al.^[7] performed numerical analysis of vapour compression refrigeration system using R134a, R143a, R152a, R404A, R410A, R502 and R507A, and discussed the effect of evaporator temperature, degree of subcooling at condenser outlet, superheating of evaporator outlet, vapour liquid heat exchanger effectiveness and degree of condenser temperature on COP and exergetic efficiency. They reported that evaporator and condenser temperature have significant effect on both COP and exergetic efficiency and also found that R134a has the better performance while R407C has poor performance in all respect. Xuan et al.^[8] presented in this manuscript about the replacement of R502 by mixture

Corresponding Author,

E-mail address: professor_rsmishra@yahoo.co.in

All rights reserved: <http://www.ijari.org>

of HFC-161 in vapour compression refrigeration system and conducted experimental study it was found that mixture of HFC-161 gives same and higher performance than R404A at lower and higher evaporative temperature respectively on the vapour compression refrigeration system designed for R404A.

In spite large number of research have been done by several investigators on single stage vapour compression systems and few investigators on double and double stage systems, there are several problems are to be solved by using more than two stages cascade refrigeration systems for low temperature applications below -90°C for semen preservation. These above investigators did not studies the thermodynamic performances in terms of COP and exergetic efficiency and system exergy destruction ratio (EDR) for very low temperature application using hydrocarbon and other eleven refrigerants used in the low temperature circuits and R1234ze in higher temperature circuit for semen preservation and other applications

3. Results and Discussions

The performance of four stages cascade refrigeration systems using twelve ecofriendly refrigerants of -100°C in the lower temperature evaporator circuit and R1234ze in HTC (Higher temperature circuit) and R134a in first intermediate temperature circuit and R410a is shown in the table-1. It was observed that R245fa gives better first law efficiency in terms of overall COP of the system. However second law efficiency of R600a is better in spite by using R245fa and second law efficiency of R507a is lowest and system EDR is highest. EDR of system using R404a is also nearly approaching to EDR using R507a gives more irreversibility occurred in the system components while using these refrigerants. Although circuit first law efficiency of LTC using R114 is highest as shown in Table-1(b) respectively

Table: 1(a). Four Stage Cascade Refrigeration Systems For Low Temperature Applications Using R1234ze in hot fluid circuit-R134a-R410a-Varying Ecofriendly Refrigerants in low temp circuit ($T_{\text{eva}} = (-100^{\circ}\text{C}, T_{\text{cond}} = -70^{\circ}\text{C}, T_{\text{cascade-eva}} = 30^{\circ}\text{C}, T_{\text{cascade-eva}} = (-20^{\circ}\text{C}, T_{\text{cascade-eva}} = -70^{\circ}\text{C})$) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

LTC Eco Friendly Refri.	COP _{overall}	EDR	Exergetic Efficiency
R290	0.4023	3.201	0.2380
R404A	0.3963	7.143	0.1228
R143A	0.4012	3.957	0.2019
R236fa	0.4032	7.984	0.1113
R142b	0.4002	3.763	0.2099
R600	0.4041	3.208	0.2376
R600A	0.4037	2.879	0.2578
R507A	0.4017	7.179	0.1223
R114	0.4024	9.584	0.09448
R227ea	0.4011	14.37	0.06506

R245fa	0.4032	3.86	0.2058
R125	0.4020	4.42	0.1845

Table: 1(b). Four Stage Cascade Refrigeration Systems For Low Temperature Applications Using R1234ze in hot fluid circuit-R134a-R410a-Varying Ecofriendly Refrigerants in low temp circuit ($T_{\text{eva}} = (-100^{\circ}\text{C}, T_{\text{cond}} = -70^{\circ}\text{C}, T_{\text{cascade-eva}} = 30^{\circ}\text{C}, T_{\text{cascade-eva}} = (-20^{\circ}\text{C}, T_{\text{cascade-eva}} = -70^{\circ}\text{C})$) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

LTC Eco Friendly Refri.	COP _H of fluid circuit	COP _{int} intermediate temperature circuit-II	COP _{in} intermediate temperature circuit-II	COP _{low} temperature fluid circuit
R290	4.375	2.279	2.116	2.948
R404A	4.375	2.279	2.116	2.829
R143A	4.375	2.279	2.116	2.927
R236fa	4.375	2.279	2.116	2.968
R142b	4.375	2.279	2.116	2.907
R600	4.375	2.279	2.116	2.987
R600A	4.375	2.279	2.116	2.979
R507A	4.375	2.279	2.116	2.936
R114	4.375	2.279	2.116	3.014
R227ea	4.375	2.279	2.116	2.924

Similarly the variation of system performance with hydrocarbon and R404a and Ethylene in the LTC (lower temperature evaporator circuit evaporator temperature) varying from -160°C to -130°C and R1234ze in HTC (Higher temperature circuit) and R134a in first intermediate temperature circuit and R410a in second intermediate temperature circuit is also shown in Table 2(a) to 2(e) and circuit performance s in terms of first law efficiency of each temperature circuit is also shown in Table 3(a) to 3(e) respectively. It was observed that using R1234ze in high temperature circuit gives better thermal performance than using R1234yf in HTC. It was also observed that interchanging Refrigerants in the intermediate temperature circuits do not affect system overall second law performances also there is a slightly variation in the system exergy destruction ratio. The second law efficiency in terms of exergetic efficiency of R600a is better than R290 and R40a and lowest performance obtained using ethylene in the LTC. It was also observed that One can used R600 hydrocarbon below 136°C only but theoretically the performance is also shown in Table 2(e) and 3(e) respectively.

Table: 2(a). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a in second cascade intermediate circuit and Ecofriendly R600a refrigerant in low temp circuit for ($T_{\text{cascade-eva1}} = 0^{\circ}\text{C}, T_{\text{cond}} = 70^{\circ}\text{C}, T_{\text{cascade-eva2}} = -60^{\circ}\text{C}, T_{\text{cascade-eva3}} = -120^{\circ}\text{C},$) APPROACH1=APPROACH2=APPROACH3= 10°C and

Compressor-Efficiency1= Compressor-Efficiency2=
Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA_Temp	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1436	3.255	0.2350
-155	0.1582	3.145	0.2413
-150	0.1730	3.063	0.2461
-145	0.1879	3.006	0.2496
-140	0.2029	2.972	0.2518
-135	0.2179	2.959	0.2526
-130	0.2328	2.963	0.2523

Table: 2(b). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R290 refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

Eva_Temp	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1385	3.411	0.2267
-155	0.1539	3.259	0.2348
-150	0.1696	3.144	0.2413
-145	0.1853	3.063	0.2461
-140	0.2110	3.011	0.2494
-135	0.2165	2.983	0.2511
-130	0.2319	2.978	0.2514

Table: 2(c). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R404a refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

Eva_Temp	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1369	3.461	0.2242
-155	0.1672	3.203	0.2379
-150	0.1520	3.313	0.2318
-145	0.1826	3.124	0.2425
-140	0.1979	3.073	0.2455
-135	0.2132	3.046	0.2472
-130	0.2283	3.040	0.2475

Table: 2(d). Variation of low temperature evaporator temperature with thermal performance of four stage cascade

refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly ethylene refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

Eva_Temp	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1366	3.472	0.2236
-155	0.1524	3.302	0.2325
-150	0.1684	3.175	0.2395
-145	0.1843	3.085	0.2448
-140	0.2002	3.026	0.2484
-135	0.2160	2.994	0.2504
-130	0.2315	2.985	0.2509

Table: 2(e). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R600 refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

Eva_Temp	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1418	3.308	0.2321
-155	0.1568	3.182	0.2391
-150	0.1719	3.088	0.2446
-145	0.1872	3.022	0.2486
-140	0.2024	2.982	0.2511
-135	0.2176	2.964	0.2523
-130	0.2326	2.966	0.2521

Table: 3(a). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R600a refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA_Temp	COP _{HTC}	COP _{ICT 1}	COP _{ICT 2}	COP _{LTC}
-160	1.797	1.784	1.212	1.255
-155	1.797	1.784	1.212	1.784
-150	1.797	1.784	1.212	1.889

-145	1.797	1.784	1.212	2.349
-140	1.797	1.784	1.212	2.966
-135	1.797	1.784	1.212	3.834
-130	1.797	1.784	1.212	5.138

Table: 3(b). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R290 refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.797	1.784	1.212	1.170
-155	1.797	1.784	1.212	1.447
-150	1.797	1.784	1.212	1.80
-145	1.797	1.784	1.212	2.259
-140	1.797	1.784	1.212	2.875
-135	1.797	1.784	1.212	3.742
-130	1.797	1.784	1.212	5.046

Table: 3(c). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R404a refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.797	1.784	1.212	1.145
-155	1.797	1.784	1.212	1.741
-150	1.797	1.784	1.212	1.409
-145	1.797	1.784	1.212	2.168
-140	1.797	1.784	1.212	2.737
-135	1.797	1.784	1.212	3.524
-130	1.797	1.784	1.212	4.685

Table: 3(d). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly ethylene refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C

and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.797	1.784	1.212	1.14
-155	1.797	1.784	1.212	1.417
-150	1.797	1.784	1.212	1.769
-145	1.797	1.784	1.212	2.226
-140	1.797	1.784	1.212	2.84
-135	1.797	1.784	1.212	3.703
-130	1.797	1.784	1.212	5.001

Table: 3(d). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R404a refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.797	1.784	1.212	1.225
-155	1.797	1.784	1.212	1.505
-150	1.797	1.784	1.212	1.861
-145	1.797	1.784	1.212	2.322
-140	1.797	1.784	1.212	2.942
-135	1.797	1.784	1.212	3.813
-130	1.797	1.784	1.212	5.121

Similarly system performance in terms of first law efficiency (overall system COP) and system exergy Destruction ratio (EDR) and second law efficiency in terms of exergetic efficiency (rational efficiency) using hydrocarbons in the lower temperature circuit and R1234yf in High temperature circuit and R134a in first intermediate temperature circuit and R410a in second intermediate evaporator temperature circuit is also shown in Table 4 and circuit first law performances in Table 5 respectively. It has lower performance than R1234ze in HTC. Following observations using hydrocarbons and R404a and ethylene in lower temperature evaporator circuit in the four stages cascade refrigeration systems have been drawn.

Use of hydrocarbons in the lower temperature circuit gives better first law performance in terms of overall system coefficient of performance and second law system performance in terms of exergetic efficiency. By using Hydrocarbons in lower temperature evaporators circuit safety precautions must be taken because these refrigerants also having flammable properties. R600a hydrocarbon gives best first law and second law performance than using R404a in LTC. Ecofriendly refrigerant R1234yf using high temperature circuit gives lower performances than R1234ze but it has lower global warming potential (approximately

Four) than R1234ze (GWP is six). This is a advantage using R1234yf in the high temperature circuit.

Table: 4(a). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234yf in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R600a refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1345	3.541	0.2202
-155	0.1481	3.426	0.2259
-150	0.1619	3.342	0.2303
-145	0.1757	3.286	0.2333
-140	0.1895	3.253	0.2351
-135	0.2033	3.242	0.2358
-130	0.2171	3.250	0.2353

Table: 4(b). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234yf in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R290 refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1298	3.706	0.2125
-155	0.1442	3.547	0.2199
-150	0.1587	3.439	0.2258
-145	0.1733	3.346	0.2301
-140	0.1878	3.293	0.2329
-135	0.2021	3.267	0.2343
-130	0.2163	3.266	0.2344

Table: 4(c). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234yf in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R404a refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{overall}	EDR	Exergetic Efficiency
---------------------	------------------------	-----	----------------------

-160	0.1284	3.759	0.2101
-155	0.1424	3.605	0.2171
-150	0.1565	3.491	0.2227
-145	0.1707	3.411	0.2267
-140	0.1849	3.360	0.294
-135	0.1990	3.334	0.2307
-130	0.2130	3.332	0.2309

Table: 4(d). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234yf in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly ethylene refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1280	3.771	0.2096
-155	0.1427	3.593	0.2177
-150	0.1576	3.461	0.2242
-145	0.1724	3.369	0.2289
-140	0.1870	3.309	0.2321
-135	0.2016	3.279	0.2337
-130	0.2159	3.273	0.2340

Table: 4(e). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234yf in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R600 refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} = 70^\circ\text{C}$, $T_{\text{cascade_eva2}} = -60^\circ\text{C}$, $T_{\text{cascade_eva3}} = -120^\circ\text{C}$,) APPROACH1=APPROACH2=APPROACH3= 10°C and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA _{Temp}	COP _{overall}	EDR	Exergetic Efficiency
-160	0.1329	3.596	0.2176
-155	0.1468	3.466	0.2239
-150	0.1609	3.369	0.2289
-145	0.1750	3.303	0.2324
-140	0.1891	3.263	0.2346
-135	0.2031	3.247	0.2354
-130	0.2161	3.253	0.2351

Table: 5(a). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234yf in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R600a refrigerant in low temp circuit for ($T_{\text{Cascade_eva1}} = 0^\circ\text{C}$, $T_{\text{cond}} =$

70 °C, $T_{\text{cascade-eva2}} = -60$ °C, $T_{\text{cascade-eva3}} = -120$ °C,)
 APPROACH1=APPROACH2=APPROACH3= 10 (°C) and
 Compressor-Efficiency1= Compressor-Efficiency2=
 Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA_Temp	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.544	1.784	1.212	1.255
-155	1.544	1.784	1.212	1.535
-150	1.544	1.784	1.212	1.889
-145	1.544	1.784	1.212	2.349
-140	1.544	1.784	1.212	2.966
-135	1.544	1.784	1.212	3.834
-130	1.544	1.784	1.212	5.138

Table: 5(b). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R290 refrigerant in low temp circuit for ($T_{\text{cascade-eva1}} = 0$ °C, $T_{\text{cond}} = 70$ °C, $T_{\text{cascade-eva2}} = -60$ °C, $T_{\text{cascade-eva3}} = -120$ °C,) APPROACH1=APPROACH2=APPROACH3= 10 (°C) and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA_Temp	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.544	1.784	1.212	1.170
-155	1.544	1.784	1.212	1.447
-150	1.544	1.784	1.212	1.80
-145	1.544	1.784	1.212	2.259
-140	1.544	1.784	1.212	2.875
-135	1.544	1.784	1.212	3.742
-130	1.544	1.784	1.212	5.046

Table: 5(c) Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R404a refrigerant in low temp circuit for ($T_{\text{cascade-eva1}} = 0$ °C, $T_{\text{cond}} = 70$ °C, $T_{\text{cascade-eva2}} = -60$ °C, $T_{\text{cascade-eva3}} = -120$ °C,) APPROACH1=APPROACH2=APPROACH3= 10 (°C) and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=83.8340%

EVA_Temp	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.544	1.784	1.212	1.145
-155	1.544	1.784	1.212	1.409
-150	1.544	1.784	1.212	1.784
-145	1.544	1.784	1.212	2.168

-140	1.544	1.784	1.212	2.737
-135	1.544	1.784	1.212	3.524
-130	1.544	1.784	1.212	5.685

Table: 5(d). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly Ethylene refrigerant in low temp circuit for ($T_{\text{cascade-eva1}} = 0$ °C, $T_{\text{cond}} = 70$ °C, $T_{\text{cascade-eva2}} = -60$ °C, $T_{\text{cascade-eva3}} = -120$ °C,) APPROACH1=APPROACH2=APPROACH3= 10 (°C) and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA_Temp	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.544	1.784	1.212	1.140
-155	1.544	1.784	1.212	1.417
-150	1.544	1.784	1.212	1.769
-145	1.544	1.784	1.212	2.226
-140	1.544	1.784	1.212	2.840
-135	1.544	1.784	1.212	3.703
-130	1.544	1.784	1.212	5.001

Table: 5(e). Variation of low temperature evaporator temperature with thermal performance of four stage cascade refrigeration system using R1234ze in hot fluid circuit and R134a in first cascade intermediate circuit and R410a- in second cascade intermediate circuit and Ecofriendly R600 refrigerant in low temp circuit for ($T_{\text{cascade-eva1}} = 0$ °C, $T_{\text{cond}} = 70$ °C, $T_{\text{cascade-eva2}} = -60$ °C, $T_{\text{cascade-eva3}} = -120$ °C,) APPROACH1=APPROACH2=APPROACH3= 10 (°C) and Compressor-Efficiency1= Compressor-Efficiency2= Compressor-Efficiency3= Compressor-Efficiency4=80%

EVA_Temp	COP _{HTC}	COP _{ICT- i}	COP _{ICT- ii}	COP _{LTC}
-160	1.544	1.784	1.212	1.255
-155	1.544	1.784	1.212	1.535
-150	1.544	1.784	1.212	1.889
-145	1.544	1.784	1.212	2.349
-140	1.544	1.784	1.212	2.966
-135	1.544	1.784	1.212	3.834
-130	1.544	1.784	1.212	5.138

4. Conclusions

The following conclusions were drawn while analyzing four stages cascade refrigeration systems.

- The use of R1234ze has Global warming potential potential (GWP) = 6 gives better thermal performance than R1234yf of 4GWP in the higher temperature circuit of four stage cascade refrigeration system.
- There is not much performance improvements using R134a as compared with R410a in the first intermediate temperature circuit.

(iii) The second law efficiency using R600a in the low temperature evaporator circuit gives better performance

than using R290 and R404a refrigerant in lower temperature circuit (LTC).

References

- [1] Alberto Dopazo et al., Theoretical analysis of a CO₂-NH₃ cascade refrigeration system for cooling applications at low temperatures, *Int. Journal of Applied thermal engineering*, 29, 2009, 1577-1583
- [2] Bansal. P.K et.al, Analysis of an R744-R717 cascade refrigeration system. *Int J Refrigeration*, 31, 2008, 45-54.
- [3] Bhattacharyya S at.al, Optimization of a CO₂-C₃H₈ cascade system for refrigeration and heating. *International Journal of Refrigeration*, 28, 2005, 1284-1292.
- [4] Lee TS, Liu CH, Chen TW, Thermodynamic Analysis of Optimal Condensing Temperature of Cascade-Condenser In CO₂/NH₃ Cascade Refrigeration Systems, *International Journal Of Refrigeration*, 29, 2006, 1100-1108
- [5] Nikolaidis, et.al., Exergy method analysis of a two-stage vapour-compression refrigeration-plants performance. *Int J Applied Thermal Engineering*, 60, 1998, 241-256.
- [6] Reddy V.S. et.al, N. L Panwar, S.C Kaushik-Exergy analysis of a vapour compression refrigeration system with R134a, R143a, R152a, R404A, R407C, R410A, R502 and R507A. *Clean Techn Environ Policy*, 14, 2012, 47-53
- [7] Xuan et.al, Guangming Experimental study on HFC-161 mixture as an alternative refrigerant to R502, *Int J Refrigeration*. In press, 2014