I. R. IRAN

ISSN: 1728-1431
e-ISSN: 1735-9244

International Journal of Engineering

Journal Homepage: www.ije.ir

MERC

TRANSACTIONS A: Basics

Volume 33, Number 07, July 2020

Materials and Energy Research Center

In the Name of God

INTERNATIONAL JOURNAL OF ENGINEERING
Transactions A: Basics

DISCLAIMER

The publication of papers in International Journal of Engineering does not imply that the editorial board, reviewers or publisher accept, approve or endorse the data and conclusions of authors.

International **J**ournal of **E**ngineering *Transactions A: Basics* (ISSN 1728-1431) (EISSN 1735-9244)
International **J**ournal of **E**ngineering *Transactions B: Applications* (ISSN 1728-144X) (EISSN 1735-9244)
International **J**ournal of **E**ngineering *Transactions C: Aspects* (ISSN 2423-7167) (EISSN 1735-9244)
Web Sites: www.ije.ir & www.ijeir.info E-mails: ije.editor8@gmail.com, Tel: (+9821) 88771578, Fax: (+9821) 88773352
Materials and Energy Research Center (MERC)

ISSN: 1728-1431, e-ISSN: 1735-9244

International Journal of Engineering, Volume 33, Number 7, July 2020

CONTENTS:

Civil Engineering

H. O. Abbas	Laboratory Study on Reinforced Expansive Soil with Granular Pile Anchors	1167-1172
S. Ghasemi; G. Ghodrati Amiri, M. Mohamadi Dehcheshmeh	Structural Damage Assessment via Model Updating Using Augmented Grey Wolf Optimization Algorithm	1173-1182
K. Shaiksha Vali, B. S. Murugan, S. K. Reddy, E. Noroozinejad Farsangi	Eco-friendly Hybrid Concrete Using Pozzolanic Binder and Glass Fibers	1183-1191

Electrical & Computer Engineering

S. M. Mirrezaei	Bandwidth and Delay Optimization by Integration of Software Trust Estimator with Multi-user Cloud Resource Competence	1192-1200
M. Hassanpour, H. Malek	Learning Document Image Features with Squeeze Net Convolutional Neural Network	1201-1207
D. Bidekani Bakhtiarvand, S. Farzi	An Ensemble Click Model for Web Document Ranking	1208-1213
A. Aliabadian, M. R. Zahabi, M. Mobini	Optimal Singular Value Decomposition Based Pre-coding for Secret Key Extraction from Correlated Orthogonal Frequency Division Multiplexing Sub-channels	1214-1222
M. Grailoo, B. Alizadeh	AIOSC: Analytical Integer Word-length Optimization Based on System Characteristics for Recursive Fixed-Point Linear Time Invariant Systems	1223-1230
A. R. Fereydouni, A. Charmin, H. Vahdaty, H. Nasir Aghdam	Channel Estimation and Carrier Frequency Offset Compensation in Orthogonal Frequency Division Multiplexing System Using Adaptive Filters in Wavelet Transform Domain	1231-1239
Z. Shaeiri, S. J. Kazemitabar	Fast Unsupervised Automobile Insurance Fraud Detection Based on Spectral Ranking of Anomalies	1240-1248

V. Kehri, R. N. Awale	A Comparative Analysis of Wavelet-Based FEMG Signal Denoising with Threshold Functions and Facial Expression Classification Using SVM and LSSVM	1249-1256
M. Abbasi, M. Rafiee	Efficient Parallelization of a Genetic Algorithm Solution on the Traveling Salesman Problem with Multi-core and Many-core Systems	1257-1265
R. Mortazavi	Repeated Record Ordering for Constrained Size Clustering	1266-1273

Industrial Engineering

H. Beiki, S. M. Seyedhosseini; V. R. Ghezavati, S. M. Seyedaliakbar	A Location-Routing Model for Assessment of the Injured People and Relief Distribution under Uncertainty	1274-1284
M. A. Fattahzadeh, A. Saghaei	A Statistical Method for Sequential Images–based Process Monitoring	1285-1292
M. Khedmati, F. Seifi, M. J. Azizi	Time Series Forecasting of Bitcoin Price Based on Autoregressive Integrated Moving Average and Machine Learning Approaches	1293-1303
Z. Saeidi-Mobarakeh, R. Tavakkoli-Moghaddam, M. Navabakhsh, H. Amoozad-Khalili	A Bi-level Meta-Heuristic Approach for a Hazardous Waste Management Problem	1304-1310
H. Hernadewita, B. I. Saleh	Identifying Tools and Methods for Risk Identification and Assessment in Construction Supply Chain	1311-1320
A. Salmi, L. Bousshine, K. Lahlou	A New Model of Equivalent Modulus Derived from Repeated Load CBR Test	1321-1330

Material Engineering

M. A. Aswad, S. H. Awad, A. H. Kaayem	Study on Iraqi Bauxite Ceramic Reinforced Aluminum Metal Matrix Composite Synthesized by Stir Casting	1331-1339

Mechanical Engineering

T. A. Hafiz	Life Prediction of Carbon Fiber Reinforced Polymers using Time Temperature Shift Factor	1340-1346

A. Fekari, N. Javani, S. Jafarmadar	Experimental Study of Management Systems for Emission Reduction in Ignition Engines	1347-1353
S. M. Moghimi, M. Abbasi, M. Khaki Jamei, D. D. Ganji	Effect of Non-uniform Magnetic Field on Non-newtonian Fluid Separation in a Diffuser	1354-1363
D. Changbin, L. Yongping, W. Yongqiao	Meshing Error of Elliptic Cylinder Gear Based on Tooth Contact Analysis	1364-1374
M. Akbari, G. Shafabakhsh, M. R. Ahadi	Single-vehicle Run-off-road Crash Prediction Model Associated with Pavement Characteristics	1375-1386
K. Divsalar, R. Shafaghat, M. Farhadi, R. Alamian	Numerical Simulation of Hydrodynamic Properties of Alex Type Gliders	1387-1396
A. Vyas, J. Menghani, H. Natu	Metallurgical and Mechanical Properties of Laser Cladded AlFeCuCrCoNi-WC10 High Entropy Alloy Coating	1397-1402
Z. H. Wu, Y. Q. Xu, K. A. Liu, Z. Y. Chen	Numerical Analysis of Grease Film Characteristics in Tapered Roller Bearing Subject to Shaft Deflection	
Mining Engineering		1403-1412
A. Hajarian, M. Osanloo	A New Developed Model to Determine Waste Dump Site Selection in Open Pit Mines: An Approach to Minimize Haul Road Construction Cost	1413-1422
M. Dvoynikov, A. Kunshin, P. Blinov, V. Morozov	Development of Mathematical Model for Controlling Drilling Parameters with Screw Downhole Motor	1423-1430
K. R. Argimbaev, D. N. Ligotsky, K. V. Mironova, E. V. Loginov	Investigations on Material Composition of Iron-containing Tails of Enrichment of Combined Mining and Processing in Kursk Magnetic Anomaly of Russia	1431-1439

IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1167-1172

International Journal of Engineering

Journal Homepage: www.ije.ir

Laboratory Study on Reinforced Expansive Soil with Granular Pile Anchors

H. O. Abbas*

Department of Civil Engineering, University of Diyala, Diyala, Iraq

PAPER INFO

Paper history:
Received 06 February 2020
Received in revised form 03 March 2020
Accepted 07 March 2020

Keywords:
Sand Pile
Expansive Soil
Swelling
Granular Pile Anchor

ABSTRACT

Granular Pile Anchor (GPA) considers one of the solution foundation techniques, designed to mitigate the lifting of the sole resulting from expansive soils. This study work is to investigate the uplift movement response of GPA in expansive soil and evaluation performance in this soil. The effects of several parameters, such as length (L) of GPA and diameter (D), the thickness (H) of expansive clay layer and the existence sandy soil layer are investigated. The results evidenced the effectiveness and ability of GPA to reduce the lifting movement of the expansive soil and presented that the lifting movement can be decreased with rising length to some extent and the GPA diameter. The lifting movement of GPA-Foundation System is controlled by 3 separate variables, these are L/H and L/D ratios and diameter. The lifting movement can be decreased by up to (47%) if GPA is embedded in layer of expansive soil at L/H = 1, and by 83% if GPA is in expansive soil and extensive sandy soil is embedded at L/H = 1.4 with the similar GPA diameter and foundation.

***doi**: 10.5829/ije.2020.33.07a.01*

1. INTRODUCTION

Until the end 1930s, geotechnical engineers did not recognize the damage associated with buildings on extensive soils. In 1938, the US Bureau of Reclamation made the first recorded observation of ground lifting. documented evidence of the problems related with expansive soils is worldwide. According to Jones and Holtz [1], the damages in light weight buildings and roads caused by expansive soils attained to $2.2 billion in USA only. At bottom tip of pile, pore water pressure is increased due to vibration source [2]. In South Africa over (R100 million) is spent on affecting remedial works on buildings on expansive soils [3]. There are many methods that can be utilized to reduce the damage effects from expansive soils. This includes replacement of soil, chemical, and physical treatment and the use of special techniques. Nine essential additives in addition to three mixed additives with different ratios have been used and implemented during the installation of helical pile and some of these additives gave good treatment for problem of expansive soil. The use of these methods is retained over a long period of time [4, 5]. However, many of them have certain limitations and can be very expensive [6]. Stone columns or granular piles are a known technique for soil improvement, which can reduce the build-up and increase the load-bearing capacity of soft clay beds [7, 8]. GPA foundation resisted swelling with increase diameter and length of pile as a result from friction around the perimeter of pile [9]. The results of numerical study depicted that the influence of GPA foundation to be a valuable in solving problems of swelling in expansive soil [10]. This study is an endeavor to better understand the GPA performance and behavior in expansive soils to decrease resistance and lift the pullout load. The following parameters are investigated: GPA footing system performance under swelling and the GPA adequacy and validity as a reliable solution to the problems of expansive.

Various parameters are examined that are taken into account in GPA design, such as the length of GPA system (L), the diameter (D), expansive soil layer depth (H) and the depth in sandy layer (stable region).

2. EXPERIMENTAL PROGRAM

2. 1. Materials Used Three materials are used in this study; these can be described as the following:

*Corresponding Author Email: *temimi71@yahoo.com* (H. O. Abbas)

Please cite this article as: H. O. Abbas, A laboratory Study on Reinforced Expansive Soil with Granular Pile Anchors, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1167-1172

1. A bed of foundation is represented by expansive soil.
2. Sand is used in granular pile material and stable zone.
3. Steel as a matter for shallow foundation, rod of anchor and anchor plate.

2. 2. The Properties of Expansive Soil The expansive clay soil utilized in present research was produced artificially by admixing Iraqi bentonite from the city of Al-Anbar / Bushayrah Valley, 35 km south of the base of Al-Waleed military, at a depth of 3.5 m from natural ground level, with natural soil from Al-Khalis region with ratio of (1:1). Table 1 presented the routine tests results of expansive soil.

2. 3. Sand Properties The material, which is used as stable zone and a granular pile, is poorly graded dense clean sand obtained from the local markets. Table 2 presented the laboratory tests results.

2. 4. Granular Pile Anchor System The anchor parts of granular piles consists from steel rod and plate of circular shape. To perform the anchor system, the steel rod penetrate the granular pile and connected with model footing in upper end by a bolt, while, the other end (lower end) is connected with anchor plate by a bolt also. The diameters of anchor plate are chosen in the same diameters of GPA models. The dimensions of the anchors (rod and plate) are shown in Table 3 and in Figure 1.

2. 5. The Model Footing Steel circular plate with (20 cm) diameter and (3 mm) thickness is utilized as a shallow footing model. A hole of (30mm) diameter is fixed at the center of footing in order to connect with steel rod of GPA by bolt, as shown in Figure 2.

TABLE 1. Summary of properties of expansive soil used

Test description	Soil Property	Value
Atterberg Limits (ASTM D-4318)	Liquid limit(L.L),%	91
	Plastic Limit(P.L),%	38
	Plasticity Index(P.I),%	53
Specific Gravity (ASTM D-854)	Specific Gravity (Gs)	2.75
Grain size analysis (ASTM D-422)	Gravel,%	2
	Sand,%	43
	Silt and Clay %	55
	UnifiedSoil Classification System(USCS)	CH
Consolidation (ASTM D-3084) Method (A)	Swelling Potential, %	15
	Swelling Pressure, kPa.	210
Standard Compaction Test(ASTM D-1557)	Max. Dry Unit Weight,(kN/m^3)	13.5
	Optimum Moisture Content(O.M.C),%	16

TABLE 2. Summary of the properties of sandy soils

Test Description	Property	Value
Grain Size Analysis (ASTM D-422)	D10(mm)	0.10
	D30(mm)	0.17
	D60(mm)	0.22
	Coefficient of Uniformity (Cu)	2.20
	Coefficient of Curvature (Cz)	1.31
	Unified Soil Classification System (USCS)	SP
Specificgravity (ASTM D-854)	Specific Gravity (Gs)	2.69
Maximum Unit Weight(ASTM D-4253)	Max. Unit Weight , kN/m³	16.7
Minimum Unit Weight(ASTM D-4254)	Min. Unit Weight, kN/m³	13.3
Chosen	Experimental Relative Densities (Dry) ,%	80
Calculated	Experimental Unit Weight (Dry), kN/m³	16

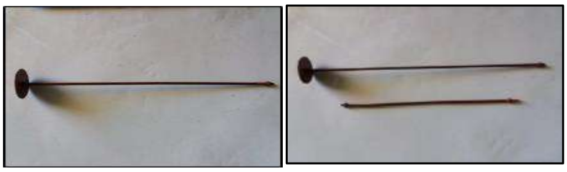


Figure 1. Plates of anchor system

TABLE 3. Dimensions of anchors (rod and plate)

	Length (mm)			Diameter (mm)
Anchor rod	250	300	350	30
	Diameter (mm)		Thickness (mm)	
Anchor plate	25	50	3	

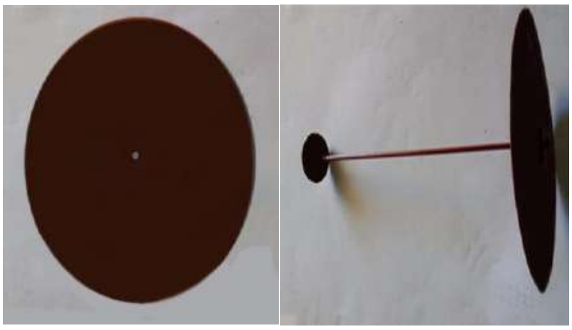


Figure 2. Plates of model footing and anchor system

2. 6. The Test Tank (Container Model) and Testing Frame Shallow footing model, granular pile anchor model, sandy soil and the expansive clay soil below placed in a cylindrical steel tank to simulate the real case in the field as well as possible. The test tank is made of (4 mm) thick steel plate with interior dimensions of (31 cm) in diameter and height of (55cm). The upper distance of 10cm of height of container is lifted in order to perform the saturation process of soil within the test tank.

2. 7. The Models of Granular Pile Anchor The test program is performed on single GPA with various lengths and diameters. The diameter of granular pile anchor (D) is varied as 2.5 and 5 cm. For each diameter, the length of granular pile anchor (L) varies from 250, 300, and 350 mm; these lengths of GPA are taken as a function of L/H where (H) indicates that expansive soil bed depth (H = 25 cm), because, the ratio of L/H is became as 1, 1.2 and 1.4. Consequently, the range of the L/D ratio of the granular pile anchors varied from 5 to 14. In total, six GPA models in addition to unreinforced model were used in the testing program as shown in Figure 3.

2. 8. Granular Pile Anchor Installation The following procedure is used in order to install the GPA in expansive soil bed and sand layer:

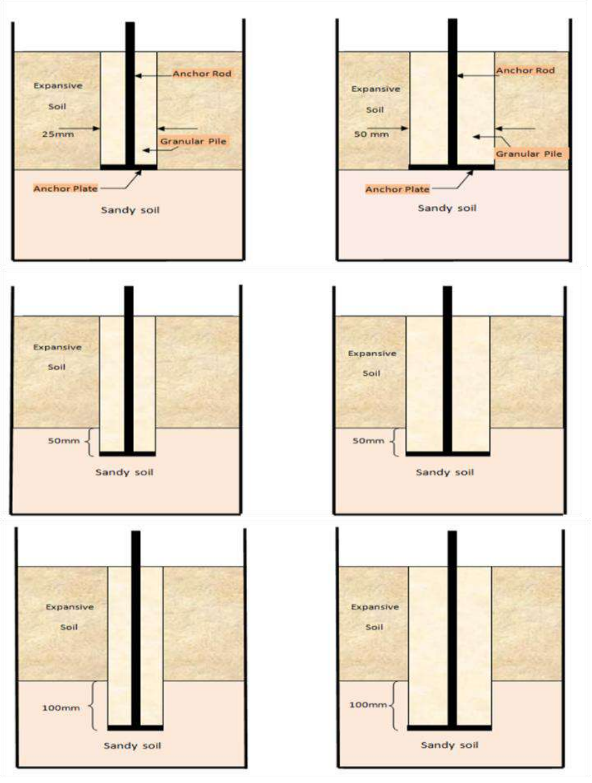


Figure 3. Details of cross sections of (GPA) models used in this study

1. After preparing and compacting the expansive soil bed, the nylon cover is removed and the top surface is leveled.
2. The hole is carefully made in the middle of the expansive soil bed surface and sand layer by gradually bringing a steel pipe with the specific diameter to the desired depth. Verticality of steel pipe is controlled during test.
3. The unit of the anchor rod with the lower anchor plate with the specific depth and diameter is inserted perpendicularly into the hole. At the same time, the hole is gradually filled with poor sand and gently compacted utilizing a steel stuffing rod with the desired relative density (80%). Finally, a GPA with dry unit weight of 16 kN/m³ is formed at the specific depth and specific diameter.

2. 9. Testing Procedure The swelling test is performed on a bed of expansive soil, which is not reinforced and reinforced with GPA. After the preparation of an expansive soil and sand layers, the test configuration steps are followed to perform the unreinforced and reinforced tests with GPA expansive soil and sand beds as:
1. The model foundation is placed in the middle of the soil bed itself for the case of unreinforced expansive soil and linked with a bolt to the steel rod anchor.
2. The 0.01 mm accuracy dial gauge with is positioned on the surface of footing and attached to the frame of loading with specific instruments in order to record the reading of swelling during the swelling process.
3. From the container top, water is added gradually to the soil bed.
4. After (30-60) days, the saturation is approximately completed. All readings are recorded during this period.
5. At the test end, the moisture content of the samples of soil taken at different depths of the soil bed is verified to certify the saturation degree. The degree of saturation must be reached (100%).

3. TEST RESULTS AND DISCUSSION

Many factors are investigated such as GPA size, L/D, Ls/H and L/H ratios. The results are analyzed, discussed and displayed in simplified manner.

3. 1. Uplift Movement of Unreinforced and Reinforced with GPA A first reference test was performed on an unreinforced, expansive soil bed under the model foot to determine the degree of improvement that was achieved after the introduction of the GPA. One model test is performed on unreinforced expansive soil to obtain the final uplift movement. Figure 4 shows the time-uplift movement relationship curves. It can be seen that, the relationship is not linear and the uplift movement of expansive soil bed continuously increases with time up to maximum values of 40 mm at time of 60 days. At this step, the saturation of the expansive soil is adjusted and the test is completed. Also, the typical relations are noticed of uplift movement-time of six reinforced models of GPA at various cases. In general, the uplift movement does not appear linear and rises continuously over time until reached equilibrium after 30 days. The rate of uplift movement suddenly reduced after about 20 days for reinforced soil is due to cylinderical sand layer around rod anchor which allows to water seep through expansive layer and increase swelling at the begginig period of test. When soil is saturated the swelling begins to decrease. At this stage, the swelling is stopped and the saturation is completed. The uplift movement of GPA reduces with GPA extended to sandy layer. This indicates the efficiency of (GPA) in decreasing the uplift movement. This is consistent with literature [9-16].

3. 2. Effect of GPA Size on the Uplift Movement Results Figure 5 shows the variation of the maximum uplift movement with diameters and lengths of GPA. The results reflect the effect of GPA on the uplift movement response of GPA-Foundation System and capability of GPA in reduce the uplift movement of expansive soil bed. It can be clearly observed that, for a given diameter of GPA the uplift movement reduces with augmented embedment length of GPA. This reduction of uplift movement can be attributed to the effect of anchor system afforded in the granular pile anchor, which made it tension-resistance member, and friction or shear resistance rallied around cylindrical pile soil volume interface, which resist forces resulting from swelling pressure of expansive soil bed. This behavior in agreement with data reported in literature [9-16].

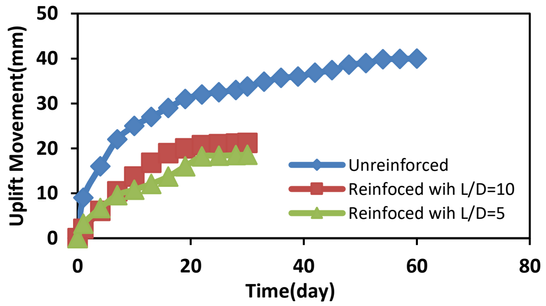


Figure 4. Uplift movement–time relationship for Unreinforced and reinforced of length to diameter ratios L/D=10 and L/D=5

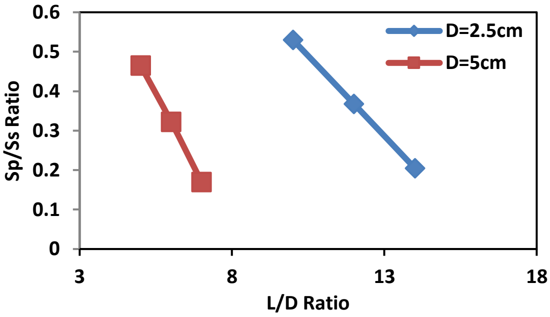


Figure 5. Variation of the normalized ratio (Sp/Ss) with (L/D ratio) of (GPA) for two diameters

3. 3. Effect of L/D Ratio of GPA on the Uplift Movement Results Figure 5 shows the relation between the proportion of maximum uplift movement with and without GPA reinforcement, Sp/Ss and L/D ratio of GPA, where Sp denoted as maximum uplift movement of foundation with reinforcement GPA. It can be seen that the max. Uplift movement of footing with GPA reinforcement reduces with rising of L/D ratio of GPA for a given diameter and expansive soil bed thickness (H), this is due to augmenting in its length. This performance may be assigned to the augment the surface area of GPA and its weight that causes increasing in pullout resistance along the GPA-soil interface against the swelling pressure. This performance agrees with the results of laboratory, field, and numerical results reported in literature [9-16].

3. 4. Effect of L/H Ratio on the Uplift Movement Results Figure 6 shows the relationship between the Sp/Ss ratio and L/H ratio for various cases of diameters. The figure reflects the L/H effect on the maximum uplift movement of footing with GPA reinforcement. Generally, the uplift movement reduces with the increase in the ratio L/H because of anchoring effect of GPA. A significant reduce was noticed at L/H=1.4, i.e., the GPA extended halfway down the sandy soil. This performance may be explained that as the swelling pressure effect of expansive soil reduces with rising L/H ratio, pullout resistance of GPA with length becomes equal to 1.4 depth of the expansive layer, which means GPA contributes to the reduction of the uplift movement.

3. 5. Effect of Ls/H Ratio on the Uplift Movement Results It is obvious from Figure 7; the maximum uplift movement reduces with increasing extended length in sand layer. The part of pile depth extended in sand soil is affected frequently on reducing uplift movement. This figure depicts the dimensionless ratio Sp/Ss plotted anti Ls/H ratio, a depth of extended of pile in sand layer to the depth of expansive soil. A proposed relationship was noticed within a limited number of model tests achieved for specified soil. Extrapolating the results gives the ratio Sp/Ss=0; this mean no uplift movement at this depth. In field, the required fixed depth at which no movement in pile may be concluded from this relation. The factor of safety is high in this case because all models tests were not included the applied load, which certainly reduces uplift movement.

3. 6. Degree of Improvement The results of the uplift movement of both the reinforced and the unreinforced expansive soil with GPA are combined and compared according to Figures 5 and 6 to evaluate the effectiveness and ability of GPA in decreasing the uplift movement. As previously stated, the unreinforced untreated expansive soil has reached a last uplift movement 40mm for 60 days. Since no technology was provided in the expansive floor to stop the lifting movement, the expansive floor swelled completely. However, in the case of reinforced treated expansive soil bed with GPA, the uplift movement is reduced considerably. It can be concluded from the results that there are three main variables that control the uplift movement performance of GPA which were L/H, L/D and Ls/H ratios. The uplift movement of GPA was influenced by me one or two or all of these variables, a decrease in uplift movement and the improvement degree rise with rising L/D, Ls/H and L/H ratios. The proportion of decrease in uplift movement and the improvement degree can be articulated as a percentage of maximum uplift movement without using GPA as in this equation:

$$\text{Degree of Improvement (\%)} = \frac{Ss-Sp}{Ss} x100 \ldots\ldots (4.1)$$

where Ss and Sp represent the maximum uplift movement of the footing without and with GPA reinforcement. It should be observed that a slight decrease in uplift movement was noticed at L/H=1, L/D=10 and D=2.5cm of 47% as a degree of improvement, whereas greater decrease in uplift movement was noted in L/H=1.4, L/D=14 and D=2.5cm of 80 % as an improvement degree. This reflects an individual's ability and efficiency (GPA) to reduce the uplift movement when extended in stable layer. The degree of improvement in the reduction of the uplift movement is summarized in Table 4.

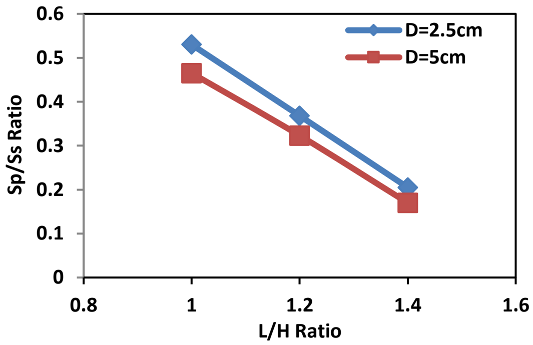


Figure 6. Variation of the normalized ratio (Sp/Ss) with (L/H ratio) of (GPA) for two diameters

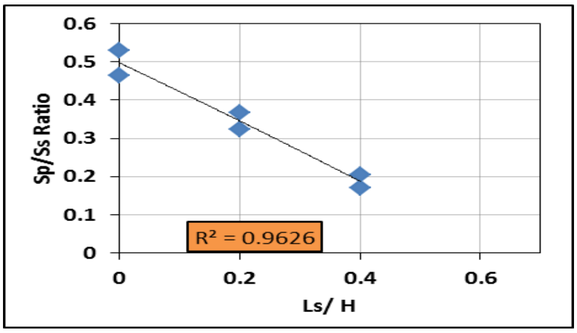


Figure 7. Dimensionless relationship of ratio Sp/Ss with ratio Ls/H of GPA

TABLE 4. Degree of improvement for all cases

L/D Ratio	Improvement Degree %
10	47
12	63
14	80
5	54
6	68
7	83

4. CONCLUSIONS

Conclusions of this study are summarized as follows:

1. The installation of GPA in expansive soil effectively decreases the values of uplift movement for the different groupings of the GPA, diameter (D) and length (L), the amount of uplift movement decreases with augmenting length and diameter.
2. Three main parameters affecting and controlling movement of (GPA). These are the ratios of length to diameter (L/D), extended length in sand layer to depth of swelling soil (Ls/H) and the length to the thickness of expansive soil (L/H).
3. The maximum reduction of approximately (47%) in the lifting movement is noted when (GPA) is integrated at (L=H) and reaches (83%) at (L = 1.4H). This means that (GPA) is suitable choice for structure constucted on expansive soil.
4. The time required to increase rate of saturation of expansive soil is clearly reduces when installing (GPA) in expansive soil and sandy soil.
5. A dimensionless relationship may be used to determine the safe depth in sand layer to provide a sufficient anchorage. Future studies are required to establish formula that gives values of movement in any embedment depth.

5. REFERENCES

1. Jones, D. E. and Holtz, W. G. “Expansive Soils—the Hidden Disaster”, ***Journal of Civil Engineering, ASCE***, Vol. 43, No. 8, (1973), 49-51.

2. Basha, A.M. and Elsiragy, M., N.," Effect of Sheet Pile Driving on Geotechnical Behavior of Adjacent Building in Sand: Numerical Study", ***Civil Engineering Journal***, Vol. 5, No. 8, (2019), 1726-1737. doi.org/10.28991/cej-2019-03091366.

3. Williams, A.A. & Pellissier, J.P. “New Options for Foundations on Heaving Clay, Geotechnics in the African Environment”, Proceedings of the 10th Regional Conference for Africa on Soil Mechanics and Foundation Engineering, Vol. 1, (1993), 243-247. doi.org/10.1016/0148-9062 (93)90574-w.

4. Al-Busoda and Abbase, "Mitigation of Expansive Soil Problems by Using Helical Piles with Additives" ***Journal of Geotechnical Engineering***, Vol. 2, No. 3, (2015), 30-40.

5. Herrero, O.R. “Special Issue on Construction on Expansive Soils”, ***Journal of Construction Facilities, ASCE***, ***Journal of Performance and Construction Facilities***, Vol. 25, No. 1, (2011), 2-4. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000179

6. Dafalla, M.A., Shamrani, M.A. “Expansive Soil Properties in a Semi-Arid Region”, ***Research Journal of Environmental and Earth Sciences***, Vol. 4, No. 11, (2012), 930-938.

7. Greenwood, D.A. “Mechanical Improvement of Soils below Ground Surface”, Conference on Ground Engineering, Institution of Civil Engineers, (1970), 11-22.

8. Hughes, J.M., Withers, N.J. “Reinforcing of Soft Cohesive Soils with Stone Columns”, ***International Journal of Rock Mechanics and Mining Sciences and Geomechanics***, Vol. 11, No.11, A234 (1974). doi.org/10.1016/0148-9062 (74)90643-3.

9. Sharma, R.K., " A Numerical Study of Granular Pile Anchors Subjected to Uplift Forces in Expansive Soils Using PLAXIS 3D " (2018), ***Indian Geotech***. doi.org/10.1007/s40098-018-0333-3**.**

10. Sfoog, E.H., Siang, A.J.L., Albadri, W.M., Naji. N., Yi, S.S. and Guntor, N.A.A., " (Finite Element Modelling of Innovative Shallow Raft Foundation with Granular Pile Anchor System for Expansive Clays" The second Global Congress on Construction, Material and Structural Engineering IOP Conf. Series: Materials Science and Engineering, 713, (2020).

11. Aljorany, A., Ibrahim, S.F. and Al-Adeley, A., “Heave Behavior of Granular Pile Anchor-Foundation System (GPA-Foundation System) in Expansive Soil”, ***Journal of Engineering***, Vol. 20, No. 4, (2014).

12. Ismail, M.A. & Shahin, M. “Finite Element Analysis of Granular Pile Anchors as A Foundation Option for Reactive Soils”, International Conference on Advances in Geotechnical Engineering, Perth, Australia, (2011), 1047-1052.

13. Krishna, P.H., Murty, V.R. & Vakula, J. “A Filed Study on Reduction of Flooring Panels Resting on Expansive Soils Using Granular Anchor Piles and Cushions”, ***The International Journal of Engineering and Science,*** Vol. 2, No. 3, (2013), 111-115.

14. Phanikumar, B.R., Sharma R.S., Srirama, R.A, Madhav, M.R. “Granular Pile-Anchor Foundation (GPAF) System for Improving Engineering Properties of Expansive Clay Beds***”, Geotechnical Testing Journal, ASTM***, Vol. 27, No. 3, (2004), 279-287. doi.org/10.1520/gtj11387.

15. Phanikumar, B.R., Rao, A.S., Suresh, K..., “Field Behavior of Granular Pile Anchors in Expansive Soils”, ***Ground Improvement Journal, Proceeding of Institution of Civil Engineering***, Vol. 4, (2008), 199-206. doi.org/10.1680/grim.2008.161.4.199 17.

16. Rao, A.S., Phanikumar, B.R., Babu, R.D. & Suresh, K..,“Pullout Behavior of Granular Pile Anchors in Expansive Clay Beds In-situ”, ***Journal of Geotechnical and Geoenvironmental Engineering, ASCE***, Vol. 133, No. 5, (2007), 531-538, doi.org/10.1061/ (ASCE) 1090-0241(2007)133:5(531).

Persian Abstract

چکیده

لنگر شمع گرانول (GPA) یکی از تکنیک های پایه حل می باشد ، طراحی شده برای کاهش برخواستن از جا کنده شدن و برداشتن تنها حاصل از خاکهای گسترده است. هدف از این مطالعه بررسی واکنش حرکتی رو به بالا (GPA) در عملکرد گسترده خاک و ارزیابی آن در این خاک است. تأثیر پارامترهای مختلفی از جمله طول (L) از (GPA) و قطر (D) ، ضخامت (H) لایه رس وجود لایه خاک شنی بررسی شده است. نتایج به اثبات رساندن اثربخشی و توانایی (GPA) در کاهش حرکت بالابر خاک توسعه پرداخته را نشان می دهد که حرکت بالابر را می توان با افزایش طول تا حدی و قطر (GPA) کاهش داد. حرکت بلند کردن سیستم (GPA-Foundation system) توسط (۳) متغیر جداگانه کنترل می شود ، اینها [(L/H) و (L/D)] و قطر هستند. اگر (GPA) در لایه ای از خاک وسیع در (L/H=1) تعبیه شده باشد ، و اگر (GPA) در خاک گسترده و شنی گسترده باشد ، حرکت بالابر را تا ٤٧٪ کاهش می یابد. خاک در قطر و پایه مشابه (GPA) مشابه (L/H=1.4) تعبیه شده است.

International Journal of Engineering

Journal Homepage: www.ije.ir

Structural Damage Assessment via Model Updating Using Augmented Grey Wolf Optimization Algorithm

S. Ghasemi*[a], G. Ghodrati Amiri[b], M. Mohamadi Dehcheshmeh[a]

[a] *Civil Engineering Department, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran*

[b] *Centre of Excellence for Fundamental Studies in Structural Engineering, School of Civil Engineering, Iran University of Science & Technology, Tehran, Iran*

PAPER INFO

Paper history:
Received 31 October 2019
Received in revised form 05 April 2020
Accepted 11 June 2020

Keywords:
Augmented Grey Wolf Optimization Algorithm
Damage Detection
Flexibility Matrix
Modal Data
Static Strain Energy

ABSTRACT

Some civil engineering-based infrastructures are planned for the structural health monitoring (SHM) system based on their importance. Identifiction and detecting damage automatically at the right time are one of the major objectives this system faces. One of the methods to meet this objective is model updating whit use of optimization algorithms in structures.This paper is aimed to evaluate the location and severity of the damage combining two being-updated parameters of the flexibility matrix and the static strain energy of the structure using augmented grey wolf optimization (AGWO) and only with extracting the data of damaged structure, by applying 5 percent noise. The error between simulated and estimated results in average of ten runs and each damage scenario was less than 3 percent which proves the proper performance of this method in detection of the all damages of the 37-member three-dimensional frame and the 33-member two-dimensional truss. Moreover, they indicate that AGWO can provide a reliable tool to accurately identify the damage in compare with the particle swarm optimizer (PSO) and grey wolf optimizer (GWO).

***doi**: 10.5829/ije.2020.33.07a.02*

1. INTRODUCTION

After long term utilization, the infrastructures should be evaluated in terms of safety and sustainability. Over time, a structure may lose its desired performance due to the factors such as earthquakes, floods, storms, etc. This may even leads to its collapse.

With the advent of advanced technologies including sensor networks, information and signal processing and managing systems [1–3], The SHM process has been able to enhance safety, sustainability, the development of infrastructure, measuring and management of cost of exploitation over time. The use of monitoring provides with the required information in building smart structures; such as, equipment needed to measure or record data before the structure gets damaged more.

The Prognosis of damage by using traditional methods of local inspection or testing due to an increase in the number and dimensions of structures and their deterioration is not feasible because inspection of such structures is time consuming, costly and along with human error. Therefore, to control the remaining useful life of large and more complex structures, new methods based on changes in the vibration properties of structures have been developed; that are commonly referred to as damage detection methods [4]. The basic idea is that the modal data of the structure, such as frequency and mode shapes are influenced by the physical properties of the structure, so changes in the physical properties of the structure lead to change in its modal properties. Consequently, by comparing the modal characteristics of the structure before and after the damage, the location and severity of damage to the structure can be detected [5].

First time Holland [6] investigated the problem of detecting damage based on natural frequency with genetic algorithm. One method that has attracted many researchers today is the numerical update model method. Detection of damage without the need for undamaged structural data is one of the advantages of this method.

*Corresponding Author Email: *ghasemi@std.iaushk.ac.ir* (S. Ghasemi)

Please cite this article as: S. Ghasemi, G. Ghodrati Amiri, M. Mohamadi Dehcheshmeh, Structural Damage Assessment via Model Updating Using Augmented Grey Wolf Optimization Algorithm, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1173-1182

Defining the objective function and determining the being-updated parameters are among the most important factors affecting the success of these method. In updating the numerical model using the inverse problem, the difference between the simulated and estimated results is minimized with the help of the optimizer algorithm.

A comprehensive review by Friswell and Mottershead [7] has been conducted on various methods of updating the model. Hajela and Soeiro [8] examined two optimazation methods on a 15-member -dimensional truss, and obtained acceptable results. Boulkaibet et al. [9], used the Monte Carlo combining simulation, were able to provide a more precise method and introduce probabilities in a model update process for more sophisticated systems. Other researchers utilizing the sensitivity of the frequency response functions detected the severity of the damage [10, 11]. Ghodrati et al. [12, 13] used the flexibility matrix parameters and the modal residual force and came up with strong and stable results.

Flexibility matrix and static strain energy are two being-updated parametes used in this method which are introduced in this paper.

Yan and Golinval [14] also use covariance-based sub-space detection techniques to identify modal parameters. The stiffness matrix variations to detect damage is used as it is significantly altered due to major damage to the structure, but if the damage is small, this method is not very effective. Dynamic data and flexibility can be elicited out of dynamic experiments and structure frequency response measurements, respectively. One of the methods to detect vibration-based damage is to use statistical analysis [15, 16]. Tomaszewska [17] investigated the influence of statistical errors on damage detection methods and concentrated on the flexibility and mode shape curvature approaches as methodologies that use both natural frequencies and mode shapes. Damage detection from mode shape data requires measurements in many locations of a structure. Therefore, damage detection methods based on flexibility were utilized by researchers [18–20]. Li et al. [21] presented a generalized -flexibility matrix for a definite reduction of natural frequencies of higher modes. The flexibility matrix by applying a unit force to values of degrees of freedom (DOFs) can be used as a modal displacement to calculate the strain variation of members. Accordingly, an efficient method was used to detect multiple damage to the truss system using strain-based flexibility index (SCBFI) [22] and another flexibility-based damage probability index (FBDPI) [23], simulation results showed high performance. Zare Hosseinzadeh et al. [24] employed an effective method based on the calculation of static displacement by a matrix of flexibility. The efficiency of the proposed method was verified by an experimental study of a five-story structure with shear frame. Kaveh and Zolghadr [25] used the object function of flexibility matrix and modal strain energy (MSE) method as a conducting tool in order to direct a beam and portal frame's damage detection process.

Shi et al. [26] proposed a damage detection-method using differences in the MSE for the simple two-story plain structure. The results were partially successful in quantification of the structural damage in spite of errors. Modal-strain-energy-based methods have generally shown promise for locating damage [27–31]. However, while it has numerous advantages over other methods; recent research has shown that its application to three-dimensional frame-type structures is limited [32]. Seyedpoor and Yazdanpanah [33] found in a study on a static strain energe-based damage index (SSEBI) that this method is more reliable under similar conditions than modal strain energe-based damage index (MSEBI). Cha and Buyukozturk [34] discovered a new method for detecting damage in three-dimensional steel structures using the hybrid multiobjective genetic. Their method well detects the small damages when there is no noise. Li et al. [35] developed an Improved Modal Strain Energy (IMSE) method for detecting damage in offshore platform structures and compared it with Stubbs index method. Their comparative studies showed that the IMSE index outperformed the Stubbs index and exhibited stronger robustness.

In this paper, detection of damage considered in five sections of the introduction, overview of the AGWO, structural damage detection approach based on taking advantage of the mentioned being-updated parameters with the strategy of choosing the best performance, numerical examples. Finally, the summary is outlined in conclusions. Figure 1 illustrates the flowchart of research methodology of present work.

2. AUGMENTED GREY WOLF OPTIMIZER

Algorithm AGWO modifies the global algorithm's grey wolf optimization (GWO) by focusing on search parameter (A). This algorithm simulates the group behaviour of gray wolves in hunting, who have a leader called α. And secondary wolves with the name β, which help α in decision making (See Figure 2). Here α means estimated results to solve the problem in the research.

The hunting process is divided into four below categories.

2. 1. Searching for Prey

The exploration of the prey location can be achieved by the divergence of search agents, which can be achieved when $|A| > 1$, the main parameter responsible for exploration and exploitation is parameter A which mainly depends on parameter a as given in Equation (1).

$$\vec{a} = 2 - cos(rand) \times t/Max\_iter \tag{1}$$

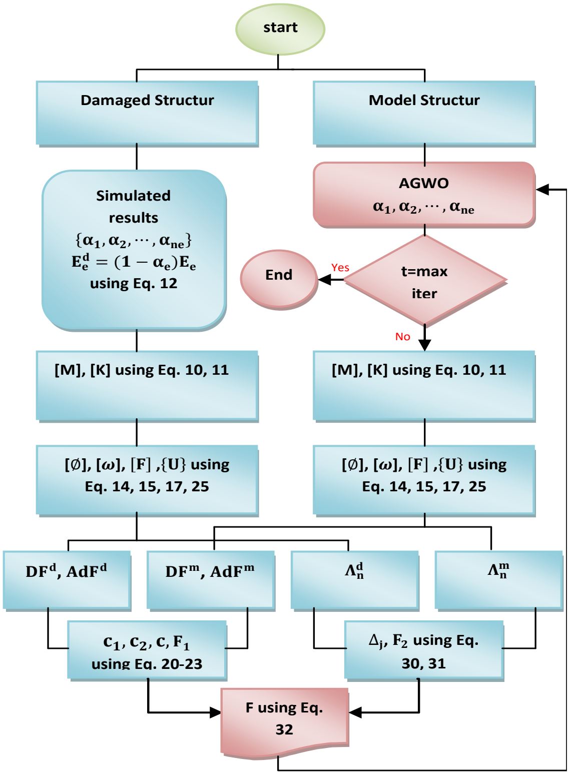

start
Damaged Structur
Model Structur
Simulated results
$\{\alpha_1, \alpha_2, \cdots, \alpha_{ne}\}$
$E_e^d = (1 - \alpha_e)E_e$
using Eq. 12
AGWO
$\alpha_1, \alpha_2, \cdots, \alpha_{ne}$
End
Yes
t=max iter
No
[M], [K] using Eq. 10, 11
[M], [K] using Eq. 10, 11
[∅], [ω], [F] ,{U} using Eq. 14, 15, 17, 25
[∅], [ω], [F] ,{U} using Eq. 14, 15, 17, 25
DF^d, AdF^d
DF^m, AdF^m
Λ_n^d
Λ_n^m
c_1, c_2, c, F_1 using Eq. 20-23
Δ_j, F_2 using Eq. 30, 31
F using Eq. 32

Figure 1. Flowchart of research methodology

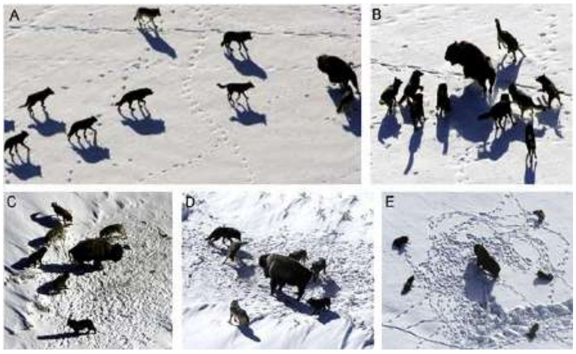


Figure 2. Hunting process [36]

$$\vec{A} = 2 - \vec{a}.\vec{r_1} - \vec{a} \tag{2}$$

$$\vec{C} = 2.\vec{r_2} \tag{3}$$

where r_1 and r_2 are uniformly distributed random vectors between 0 and 1 and parameter$a$changes nonlinearly and randomly from 2 to 1 with iteration (t) increased until it reaches maximum iteration.

2. 2. Encircling the Prey

The mathematical model of encircling the prey is expressed as follows:

$$\vec{D} = \left|\vec{C}.\vec{X}_{pi} - \vec{X}_i\right| \tag{4}$$

$$\vec{X}_{i+1} = \vec{X}_{pi} - \vec{A}.\vec{D} \tag{5}$$

where $\vec{X}$ is the position vector of grey wolf, $\vec{X}_p$ is the position vector of the prey.

2. 3. Hunting

In the proposed AGWO algorithm, the hunting will depend only on α and β as given in Equations (6)-(8).

$$\vec{D}_\alpha = \left|\vec{C}_1.\vec{X}_{\alpha i} - \vec{X}_i\right|, \vec{D}_\beta = \left|\vec{C}_2.\vec{X}_{\beta i} - \vec{X}_i\right| \tag{6}$$

$$\vec{X}_1 = \vec{X}_{\alpha i} - \vec{A}_{1.}\vec{D}_\alpha, \vec{X}_2 = \vec{X}_{\beta i} - \vec{A}_{2.}\vec{D}_\beta \tag{7}$$

$$\vec{X}_{1+i} = \vec{X}_1 + \vec{X}_2/2 \tag{8}$$

2. 4. Attacking the Prey

The exploitation of (attacking) the prey can be achieved by the convergence of search agents, which is investigated when |A| < 1 [37].

3. PROPOSED METHOD

The free vibration equation of a structural in an undamped state is written as follows:

$$[M]\{\ddot{X}\} + [K]\{X\}=0 \tag{9}$$

where [M] and [K] are the matrices of the mass and stiffness of the structure, respectively. These matrices can be obtained from the direct stiffness method for the number of elements (ne). Also $\{\ddot{X}\}$,{X}, $\left[M^e\right]$ and $\left[K^e\right]$ are acceleration, displacement vectors, the matrices of the mass and stiffness of each element, respectively.

$$[M] = \sum_{e=1}^{ne}\left[M^e\right] \tag{10}$$

$$[k] = \sum_{e=1}^{ne}\left[k^e\right] \tag{11}$$

Damage to the structure reduces the stiffness of the damaged element, which is a function of the modulus of elasticity. Thus, by reducing the modulus of elasticity of the elements using Equation (12), the actual damag to the structure is simulated.

$$E_e^d = (1 - \alpha_e)E_e \tag{12}$$

where E_e^d the modulus of elasticity of the damaged element, α_e the amount of damage to the element (a number between zero and one), where the zero indicates that there is no damage, the one indicates a damage of

100% of the element, and E_e is the modulus of elasticity of the element in the undamaged state.

Modal parameters are obtained by the solution to this equation:

$$[K-\omega_i^2 M][\phi_i]=0, i=1,2,\dots,n \tag{13}$$

According to the Equation (13), the mode shapes and the square of the natural frequencies of the structure can be obtained for n of DOFs, respectively:

$$[\omega]=\begin{bmatrix}\omega_{11}^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \omega_{nn}^2\end{bmatrix} \tag{14}$$

$$[\phi]=\begin{bmatrix}\phi_{11} & \cdots & \phi_{1n}\\ \vdots & \ddots & \vdots\\ \phi_{n1} & \cdots & \phi_{nn}\end{bmatrix} \tag{15}$$

In case of noise, its effect in this section is applied to the damaged structure using the following equation:

$$[\omega_p]=[\omega]\times(1+N\times rand) \tag{16}$$

where ω_p is the output, Nrepresents the noise level which is 5% in this paper, and the $rand$ vector [-1,1] is the random variable distributed by the software.

Thus, by using Equations (14) and (15) the flexibility matrix can be written as follows:

$$[F]_{n\times n}=[\phi]_{n\times nm}[\omega]_{nm\times nm}^{-1}[\phi]_{n\times nm}^T \tag{17}$$

in which, nm is the number of the modes used. Now, the diagonal and anti-diagonal elements of the Equation (17) are used, respectively:

$$DF=\{F_{1,1},F_{2,2},\dots F_{n,n}\}^T \tag{18}$$

$$AdF=\{F_{1,n},F_{2,n-1},\dots F_{n,1}\}^T \tag{19}$$

Based on Equations (18) and (19), four vectors are defined in order to determine C. DF^d and $\mathrm{AdF}^{\mathrm{d}}$are the two vectors defined for the damaged structure, and DF^{m} and AdF^m are those of the model structure.

$$c_1=\frac{\left|DF^{d^T}.DF^m\right|^2}{\left(DF^{d^T}.DF^d\right)\left(DF^{m^T}.DF^m\right)} \tag{20}$$

$$c_2=\frac{\left|AdF^{d^T}.AdF^m\right|^2}{\left(AdF^{d^T}.AdF^d\right)\left(AdF^{m^T}.AdF^m\right)} \tag{21}$$

$$C=(c_1\times c_2)^2 \tag{22}$$

Then, $F_1$is obtained:

$$F_1=\left(cos^{-1}(C)\times\frac{180}{\pi}\right)^{1/2} \tag{23}$$

The static strain energy of the structure can be simulated by applying the following equation for the nm mode used:

$$[K]\times\{U\}-\{P\}=0 \tag{24}$$

"If a static force, like the vector {P}, is applied to the free DOFs of the structure, the static displacements of these DOFs can be calculated by" [24]:

$$\{U\}=[K]^{-1}\times\{P\}=[F]_{n\times n}\times\{P\}_{n\times 1} \tag{25}$$

where $[K]^{-1}$ is the flexibility matrix, $\{U\}$the vector of static nodal displacement and {P} "A unique static load is as follows applied to the structure with n DOFs" [24]:

$$\{P\}=[1\quad 1\quad 1\quad \cdots\quad 1]^T \tag{26}$$

Then, using the Equation (25), the static strain energy of each element can be calculated as follows [33]:

$$\Lambda_e=\frac{1}{2}\left(u_e{}^T\times K^e\times u_e\right), e=1,2,\dots,ne \tag{27}$$

where u_e is the vector of static nodal displacement of each element, and Λ_e is the static strain energy of e-th element. For the function convergence, the static strain energy of the structure (Λ) gets normalized:

$$\Lambda_{norm}=\sqrt{\sum_{e=1}^{n_e}(\Lambda_e)^2} \tag{28}$$

$$\Lambda_{n,e}=\Lambda_e/\Lambda_{norm}, e=1,2,\dots,ne \tag{29}$$

where Λ_{norm} is the static strain energy norm of the structure, and $\Lambda_{\mathrm{n,e}}$ is the unit static strain energy of e-th element. Thus, by defining two vectors of Λ_n^d and $\Lambda_{\mathrm{n}}^{\mathrm{m}}$ for the damaged and model structures, F_2 is obtained as follows:

$$\Delta_j=\left|log(\Lambda_n^d)_j-log(\Lambda_n^m)_j\right|, j=1,2,\dots,ne \tag{30}$$

$$F_2=(max(\Delta_1,\Delta_2,\dots,\Delta_{ne}))^2 \tag{31}$$

Therefore, the objective function is defined as follows:

$$f(\alpha_1,\alpha_2,\dots,\alpha_{ne})=min(F_1,F_2) \tag{32}$$

4. ANALYSIS AND RESULT

In this section the applicability of the presented method is demonstrated by studying two-dimensional truss and three-dimensional frame structures under different damage patterns. Moreover, applying the minimum modes number, noise and various scenarios, the efficiency of the object function and accuracy algorithm AGWO in comparison with to GWO and PSO are examined. It should be declared that all analyses have been made in the workspace of MATLAB software.

4. 1. Two-Dimensional Truss

The finite element model of this structure consists of 33 elements as illustrated in Figure 3. Damage scenarios are given in Table 1, and its material properties are as follows: modules of elasticity $\mathrm{E}=2\times10^6\ \mathrm{kg/cm^2}$, mass density

$\rho = 7.85\, gr/cm$ 3 and area A $= 36.2\text{cm}^2$, respectively. Also, parameters of optimization algorithm as follows: maximum number of iterations=1000, number of population of wolves=100, upper bound=1, lower bound=0.

The results of damage detection of the two-dimensional truss data with 0% noise, and 5% noise for the first, second and thrid scenarios are presented in Figures 4, 5 and 6, respectively.

Convergence curves for the AGWO in the third damage scenario of the two-dimensional truss data with: 0% noise and 5% noise are illustrated in Figure 7.

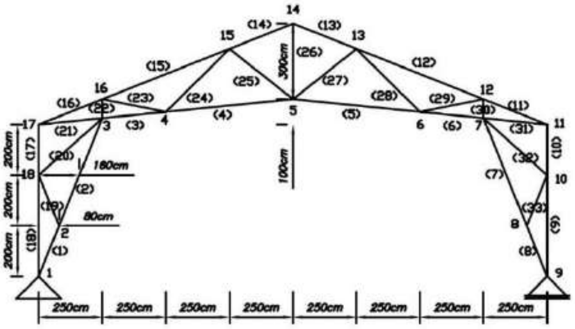


Figure 3. Two-dimensional truss

TABLE 1. Different damage scenarios for the two-dimensional truss

Damage scenario 1		Damage scenario 2		Damage scenario 3	
Element number	**Damage (%)**	**Element number**	**Damage (%)**	**Element number**	**Damage (%)**
2	10	9	5	1	20
		27	15	26	10
				33	25

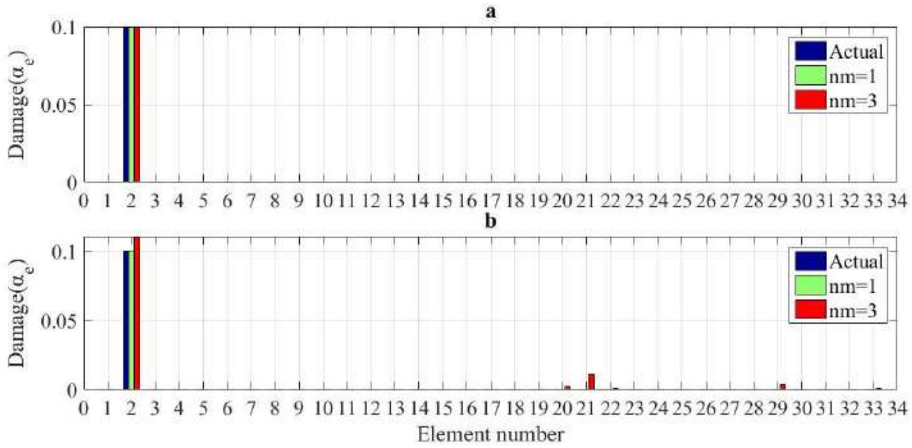


Figure 4. The results of damage detection in the first scenario of the two-dimensional truss data with: (a) 0% noise, (b) 5% noise

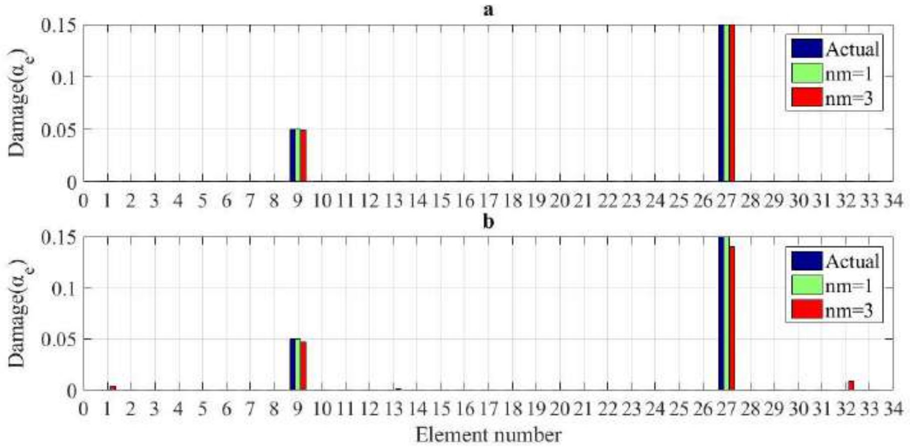


Figure 5. The results of damage detection in the second scenario of the two-dimensional truss data with: (a) 0% noise, (b) 5% noise

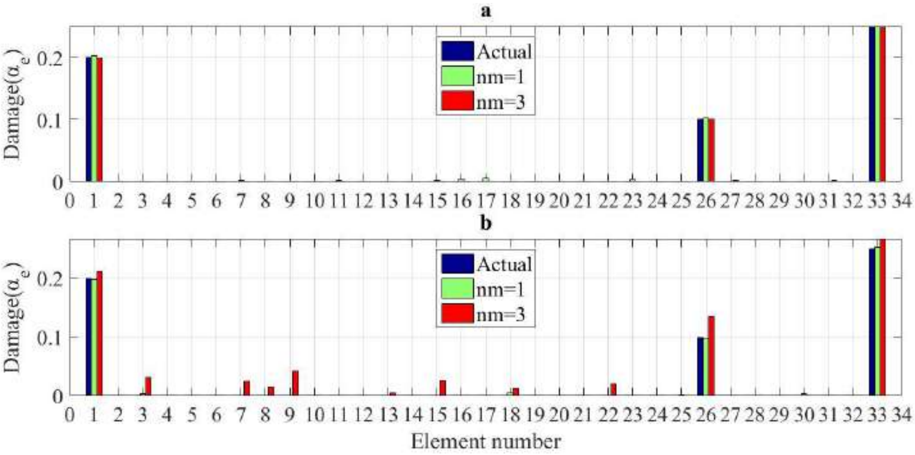


Figure 6. The results of damage detection in the third scenario of the two-dimensional truss data with: (a) 0% noise, (b) 5% noise

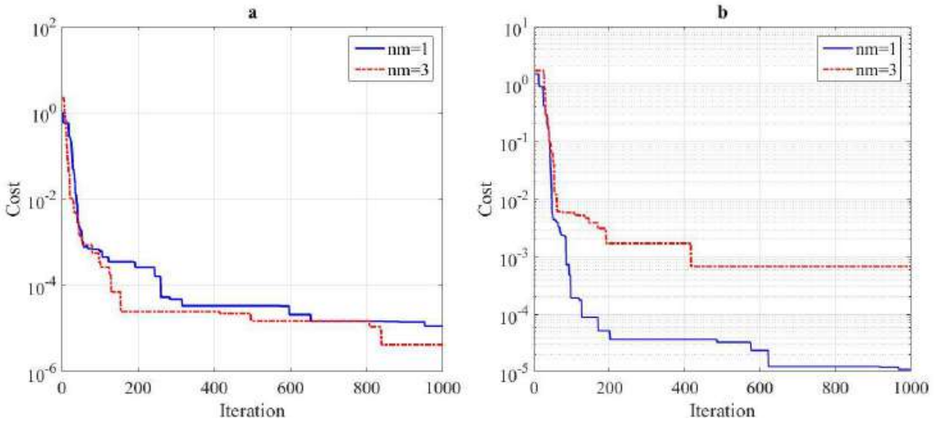


Figure 7. Convergence curves for the AGWO in the third damage scenario of the two-dimensional truss data with: (a) 0% noise, (b) 5% noise

4. 2. Three-Dimensional Frame The three dimensional frame model of this structure consists of 37 elements and 28 nodes which have six DOFs each as illustrated in Figure 8. Damaged scenarios are given in Table 2. For this structure, modules of elasticity are $\mathrm{E} = 2 \times 10^{6}\ \mathrm{kg/cm^2}$, mass density, ρ= *7.85 g/cm³* , moment of horizontal inertia$I_h = 4162cm^4$, moment of vertical inertia $\mathrm{I_v} = 4162\mathrm{cm}^4$, shear modulus $\mathrm{G} = 793000\ \mathrm{kg/cm^2}$, torsional constant $\mathrm{J} = 2081\mathrm{cm}^4$, area $\mathrm{A} = 64\mathrm{cm}^2$, the horizontal length $\mathrm{L_H} = 500\mathrm{cm}$, the vertical length $\mathrm{L_V} = 320\mathrm{cm}$. Also, parameters of optimization algorithm as follows: maximum number of iterations=1000, number of population of wolves=200, upper bound=1, lower bound=0.

The results of damage detection of the three-dimensional frame and the first seven modes for the first, second and thrid scenarios are presented in Figures 9, 10 and 11, respectively. Because of the random nature of the heuristic optimization algorithms, the average results of 10 damage detection independent runs of studied optimization algorithms investigated and shown in Figure 12.

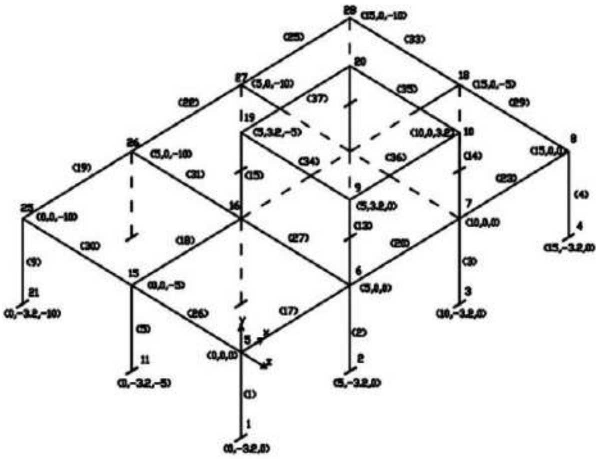


Figure 8. Three-dimensional frame

TABLE 2. Different damage scenarios for the three-dimensional frame

Damage scenario 1		Damage scenario 2		Damage scenario 3	
Element number	**Damage (%)**	**Element number**	**Damage (%)**	**Element number**	**Damage (%)**
10	10	7	5	1	10
		24	15	15	20
				35	25

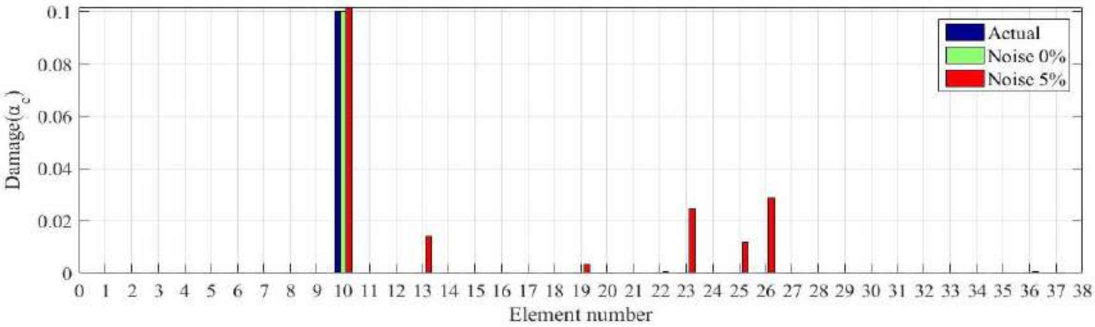


Figure 9. The results of damage detection in the first scenario of the three-dimensional frame for the first seven modes

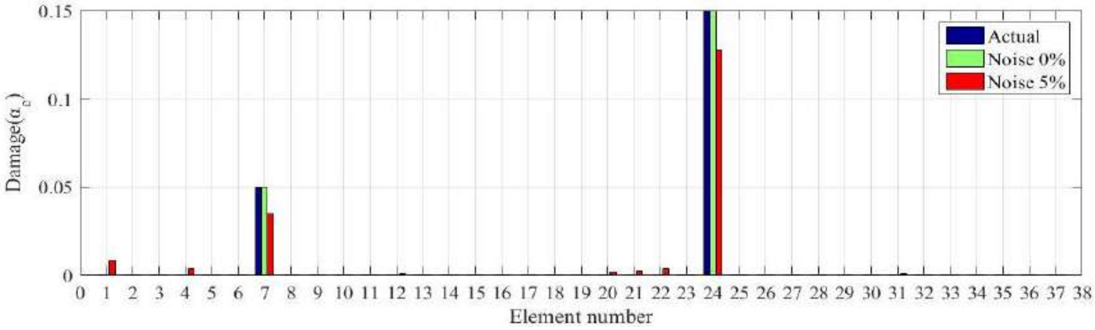


Figure 10. The results of damage detection in the second scenario of the three-dimensional frame for the first seven modes

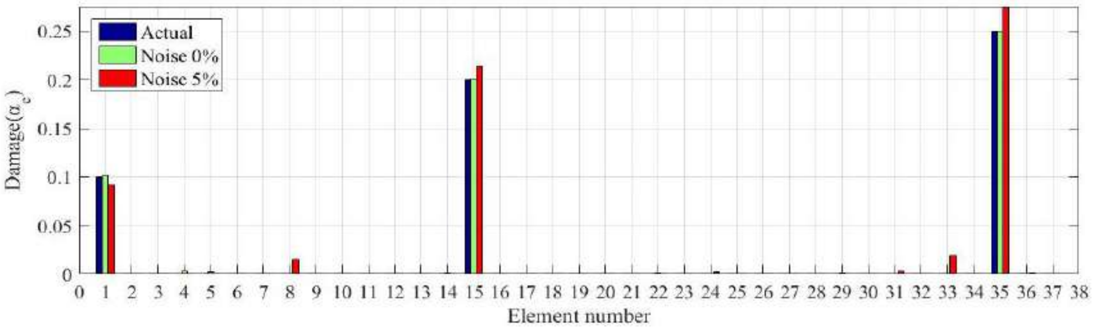


Figure 11. The results of damage detection in the third scenario of the three-dimensional frame for the first seven modes

To compare the reliability and efficiency of optimization algorithms, the best, worst and, the standard deviation (SD) of the results among the 10 independent runs are presented in Table 3. As shown in Figure 13 from the left to right, the convergence curves for the third damage scenario of the three-dimensional frame data with 0% noise and 5% noise are illustrated for the AGWO, GWO and PSO, respectively.

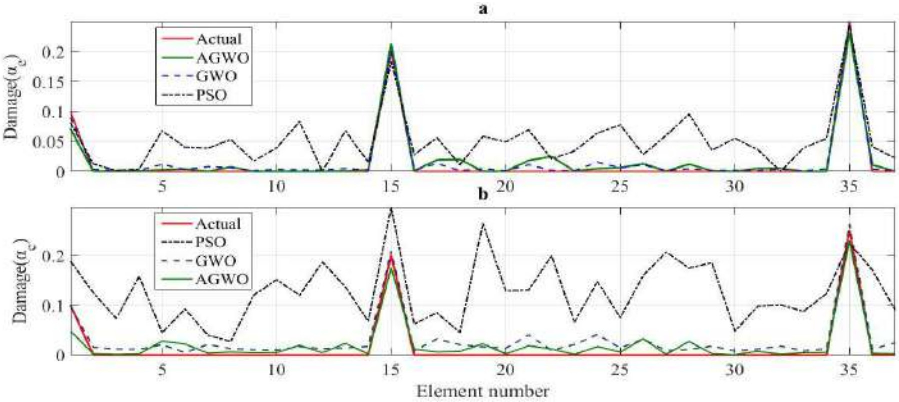


Figure 12. The average results damage detection of ten independent runs for the PSO, GWO and AGWO in the third damage scenario of the three-dimensional frame data with: (a) 0% noise, (b) 5% noise

TABLE 3. The best, worst, average and the standard deviation of the results among at ten independent runs for the optimization algorithms in the third damage scenario of the three-dimensional frame

Algorithm	AGWO $(f)_{min}$		GWO $(f)_{min}$		PSO $(f)_{min}$	
Noise	**0 (%)**	**5(%)**	**0(%)**	**5(%)**	**0(%)**	**5(%)**
Run Number 1	0.00000	0.00066	0.00000	0.00015	0	0.00610
Run Number 2	0.00000	0.00017	0.00031	0.00016	0.000109	0.00432
Run Number 3	0.00000	0.00076	0.00000	0.00011	0.001945	0.00772
Run Number 4	0.00000	0.00009	0.00000	0.00013	0.002431	0.00593
Run Number 5	0.00568	0.00276	0.00000	0.00031	0	0.00028
Run Number 6	0.00067	0.00021	0.00000	0.00025	0.004679	0.00311
Run Number 7	0.00000	0.00062	0.00000	0.00016	0.004165	0.00210
Run Number 8	0.00000	0.00064	0.00000	0.00008	0	0.00476
Run Number 9	0.00000	0.00323	0.00000	0.00026	0.000639	0.00217
Run Number 10	0.00058	0.00010	0.00000	0.00027	0	0.00054
Maximum	0.00568	0.00323	0.00031	0.00031	0.00468	0.00772
Minimum	0.00000	0.00009	0.00000	0.00008	0	0.00028
Average	0.00069	0.00092	0.00003	0.00019	0.00140	0.00370
Sd	0.00177	0.00113	0.00010	0.00008	0.00182	0.00248

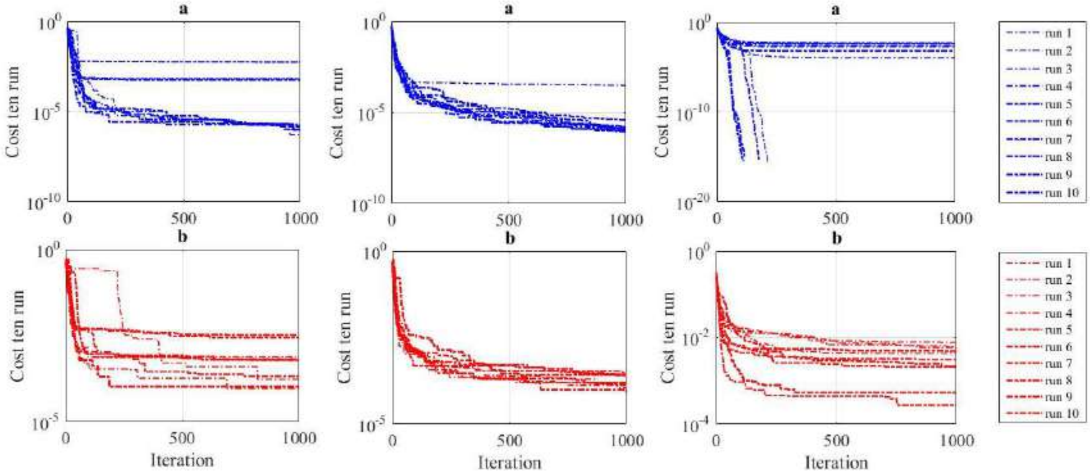


Figure 13. Convergence curves shown left to right respectively for the AGWO, GWO and PSO in the third damage scenario of the three-dimensional frame data with: (a) 0% noise, (b) 5% noise

5. CONCLUSION

In this study, an updating-based-model strategy is presented in which by combining two being-updated parameters of the flexibility matrix and the static strain energy of the structure along with the use of optimization, structural damage assessment is achieved.

Despite the limitation in process of damage assessment in two-dimensional and three-dimensional structures with high DOFs along with applying multiple damages in different parts of the structure, using noise. It is assumed that by using structural static strain energy advantages to improve the performance of the proposed objective function and reduce the weakness of small and general damage detection in flexibility-matrix-based methods and using the first few modes of the structure, damages are evaluated very precisely.

Moreover, by comparing different studies in section 4 including average results of the 10 runs, statistical results and convergence with other evolutionary optimization algorithms of PSO and GWO, the stability of the AGWO algorithm is evaluated.

The error between simulated and estimated results in average of ten runs and each damage scenario was less than 3 percent which proves the proper performance of this method in detecting the damage of the 37-member frame and the 33-member truss. Investigation on the experimental model, combining other being-updated parameters and using other new heuristic and multi objrctive algorithms in the method is recommended.

6. ACKNOWLEDGEMENTS

Hereby, we thank several people who helped us to complete this article: Farzaneh Haratyan, Masoud Shahi Juneghani and Marzieh Keshavarzi.

7. REFERENCES

1. Hamidian, D., Salajegheh, J., Salajegheh, E., “Damage Detection of Irregular Plates and Regular Dams by Wavelet Transform Combined Adoptive Neuro Fuzzy Inference System,” ***Civil Engineering Journal,*** Vol. 4-2 (2018), 305-319. doi: 10.28991/cej-030993
2. Saljoughi, A.S., Mehrvarz, M., and Mirvaziri, H., “Attacks and intrusion detection in cloud computing using neural networks and particle swarm optimization algorithms, ***Emerging Science Journal***, Vol. 1, No. 4, (2017), 179-191. doi: 10.28991/ijse-01120
3. Kouhdaragh, M., “Experimental Investigation of Damage Detection in Beam Using Dynamic Excitation System”, ***Civil Engineering Journal***, Vol. 3, No. 10 (2017), 920-928. doi: 10.28991/cej-030925
4. El-Wazery, M.S., Hassan, A., and Hassan, S., “Health Monitoring of Welded Steel Pipes by Vibration Analysis”, ***International Journal of Engineering, Transactions C: Aspects***, Vol. 28, No. 12, (2015), 1782-1789. doi: 10.5829/idosi.ije.2015.28.12c.11
5. Khoshnoudian, F., and Esfandiari, A., “Structural damage diagnosis using modal data”, ***Scientia Iranica***, Vol. 18, No. 4, (2011), 853-860. doi: 10.1016/j.scient.2011.07.012
6. Holland, J. H., “Adaptation in natural and artificial systems”,University of Michigan Press, (1975).
7. Friswell, M., and Mottershead, J.E., Finite element model updating in structural dynamics (Vol. 38). Springer Science & Business Media, (2013).
8. Hajela, P., and Soeiro, F.J., “Recent developments in damage detection based on system identification methods”, Structural Optimization, Vol. 2, No. 1, (1990), 1-10. doi: 10.1007/BF01743515
9. Boulkaibet, I., Mthembu, L., Marwala, T., Friswell, M. I., and Adhikari, S., “Finite element model updating using the shadow hybrid Monte Carlo technique”, ***Mechanical Systems and Signal Processing***, Vol. 52, (2015), 115-132. doi: org/10.1016/j.ymssp.2014.06.005
10. Esfandiari, A., Bakhtiari-Nejad, F., Rahai, A., and Sanayei, M., “Structural model updating using frequency response function and quasi-linear sensitivity equation”, ***Journal of Sound and Vibration***, Vol. 326, No. 3-5, (2009), 557-573. doi: 10.1016/j.jsv.2009.07.001
11. Shadan, F., Khoshnoudian, F., and Esfandiari, A., “A frequency response-based structural damage identification using model updating method”, ***Structural Control and Health Monitoring***, Vol. 23, No. 2, (2016), 286-302. doi: 10.1002/stc.1768
12. Ghodrati Amiri, G., Zare Hosseinzadeh, A., and Seyed Razzaghi, S. A., “Generalized flexibility-based model updating approach via democratic particle swarm optimization algorithm for structural damage prognosis”, ***Iran University of Science & Technology***, Vol. 5, No. 4, (2015), 445-464. doi: ijoce.iust.ac.ir/article-1-227-en.html
13. Ghodrati Amiri, G., Jafarian Abyaneh, M., and Zare Hosseinzadeh, A., “Model Updating-Based Approach for Damage Prognosis in Frames via Modal Residual Force”, ***International Journal of Civil and Environmental Engineering***, Vol. 10, No. 8, (2016), 1005-1011. doi: 10.5281/zenodo.1125801
14. Yan, A., and Golinval, J.C., “Structural damage localization by combining flexibility and stiffness methods”, ***Engineering Structures***, Vol. 27, No. 12, (2005), 1752-1761. doi: 10.1016/j.engstruct.2005.04.017
15. Catbas, F.N., Brown, D.L., and Aktan, A.E., “Use of modal flexibility for damage detection and condition assessment: case studies and demonstrations on large structures”, ***Journal of Structural Engineering***, Vol. 132, No. 11, (2006), 1699-1712. doi: 10.1061/(ASCE)0733 -9445(2006)132:11(1699)
16. Sung, S.H., Koo, K.Y., and Jung, H.J., “Modal flexibility-based damage detection of cantilever beam-type structures using baseline modification”, ***Journal of Sound and Vibration***, Vol. 333, No. 18, (2014), 4123-4138. doi: 10.1016/j.jsv.2014.04.056
17. Tomaszewska, A., “Influence of statistical errors on damage detection based on structural flexibility and mode shape curvature”, ***Computers & Structures***, Vol. 88, No. 3-4, (2010), 154-164. doi: 10.1016/j.compstruc.2009.08.017
18. Miguel, L.F.F., Miguel, L.F.F., Riera, J.D., and Menezes, R.C.R.D., “Damage detection in truss structures using a flexibility based approach with noise influence consideration”, ***Structural Engineering and Mechanics***, Vol. 27, No. 5, (2007), 625-638. doi: 10.12989/sem.2007.27.5.625
19. Yan, Y. J., Cheng, L., Wu, Z. Y., and Yam, L. H., “Development in vibration-based structural damage detection technique.” ***Mechanical Systems and Signal Processing***, Vol. 21, No. 5 (2007), 2198-2211. doi: 10.1016/j.ymssp.2006.10.002
20. Nobahari, M., and Seyedpoor, S.M.,“An efficient method for structural damage localization based on the concepts of flexibility matrix and strain energy of a structure”, ***Structural Engineering and Mechanics***, Vol. 46, No. 2, (2013), 231-244. doi:

10.12989/sem.2013.46.2.231

21. Li, J., Wu, B., Zeng, Q.C. and Lim, C.W., "A generalized flexibility matrix based approach for structural damage detection, ***Journal of Sound and Vibration***, Vol. 329, No. 22, (2010), 4583-4587. doi: 10.1016/j.jsv.2010.05.024

22. Montazer, M., and Seyedpoor, S.M., "A new flexibility based damage index for damage detection of truss structures", ***Shock and Vibration***, 2014, (2014). doi: 10.1155/2014/460692

23. Seyedpoor, S.M., and Montazer, M. "A damage identification method for truss structures using a flexibility-based damage probability index and differential evolution algorithm", ***Inverse Problems in Science and Engineering***, Vol. 24, No. 8, (2016), 1303-1322. doi: 10.1080/17415977.2015.1101761

24. Zare Hosseinzadeh, A., Ghodrati Amiri, G., and Koo, K.Y., "Optimization-based method for structural damage localization and quantification by means of static displacements computed by flexibility matrix", ***Engineering Optimization***, Vol. 48, No. 4, (2016), 543-561. doi: 10.1080/0305215X.2015.1017476

25. Kaveh, A., and Zolghadr, A., "Guided modal strain energy-based approach for structural damage identification using tug-of-war optimization algorithm", ***Journal of Computing in Civil Engineering***, Vol. 31, No. 4, (2017), 04017016. doi: 10.1061/(ASCE)CP.1943-5487.0000665

26. Shi, Z.Y., Law, S.S., and Zhang, L.M., "Structural damage detection from modal strain energy change", ***Journal of Engineering Mechanics***, Vol. 126, No. 12, (2000), 1216-1223. doi: 10.1061/(ASCE)0733-9399(2000)126:12(1216)

27. Wang, S., Zhang, J., Liu, J. and Liu, F.,"Comparative study of modal strain energy based damage localization methods for three-dimensional structure", In the 20th International Offshore and Polar Engineering Conference, 20-25 June, Beijing, China, (2010).

28. Seyedpoor, S.M., "A two stage method for structural damage detection using a modal strain energy based index and particle swarm optimization", ***International Journal of Non-Linear Mechanics***, Vol. 47, No. 1, (2012), 1-8. doi: 10.1016/j.ijnonlinmec.2011.07.011

29. Yan, W.J., Huang, T.L. and Ren, W.X., "Damage detection method based on element modal strain energy sensitivity", ***Advances in Structural Engineering***, Vol. 13, No .6, (2010), 1075-1088. doi: 10.1260/1369-4332.13.6.1075

30. Yan, W.J., Ren, W.X. and Huang, T.L., "Statistic structural damage detection based on the closed- form of element modal strain energy sensitivity", ***Mechanical Systems and Signal Processing***, Vol. 28, (2012), 183-194. doi: 10.1016/j.ymssp.2011.04.011

31. Entezami, A., and Shariatmadar, H., "Damage detection in structural systems by improved sensitivity of modal strain energy and Tikhonov regularization method", ***International Journal of Dynamics and Control***, Vol. 2, No. 4, (2014), 509-520. doi: 10.1007/s40435-014-0071-z

32. Li, H., Yang, H. and Hu, S.L.J., "Modal strain energy decomposition method for damage localization in 3D frame structures", ***Journal of Engineering Mechanics***, Vol. 132, N. 9, (2006), 941-951. doi: 10.1061/(ASCE)0733-9399(2006)132:9(941)

33. Seyedpoor, S.M., and Yazdanpanah, O., "An efficient indicator for structural damage localization using the change of strain energy based on static noisy data", ***Applied Mathematical Modelling***, Vol. 38, No. 9-10, (2014), 2661-2672. doi: 10.1016/j.apm.2013.10.072

34. Cha, Y.J., and Buyukozturk, O., "Structural damage detection using modal strain energy and hybrid multiobjective optimization", ***Computer-Aided Civil and Infrastructure Engineering***, Vol. 30, No. 5, (2015), 347-358. doi: 10.1111/mice.12122

35. Li, Y., Wang, S., Zhang, M. and Zheng, C. "An improved modal strain energy method for damage detection in offshore platform structures", ***Journal of Marine Science and Application***, Vol. 15, No. 2, (2016), 182-192. doi: 10.1007/s11804-016-1350-1.

36. Mirjalili, S., Mirjalili, S.M. and Lewis, A., "Grey wolf optimizer", ***Advances in Engineering Software***, Vol. 69, (2014), 46-61. doi: 10.1016/j.advengsoft.2013.12.007

37. Qais, M.H., Hasanien, H.M. and Alghuwainem, S., "Augmented grey wolf optimizer for grid-connected PMSG-based wind energy conversion systems", ***Applied Soft Computing***, Vol. 69, (2018), 504-515. doi: 10.1016/j.asoc.2018.05.006

Persian Abstract

چکیده

برخی از زیرساخت‌های عمرانی بر اساس اهمیت آن‌ها برای سیستم نظارت بر سلامت سازه (SHM) برنامه‌ریزی می‌شوند. شناسایی و تشخیص خودکار آسیب در زمان مناسب یکی از اهدف اصلی است که این سیستم با آن روبرو است. یکی از روش‌های برآورد این هدف، بروزرسانی مدل با استفاده از الگوریتم‌های بهینه‌سازی در سازه‌ها است. این مقاله با به‌کارگیری ازترکیب دو پارامتر به روز شونده ماتریس نرمی و انرژی کرنشی استاتیک با استفاده از(AGWO) و تنها بااستخراج از داده‌های سازه‌ی آسیب دیده به همراه ۵ درصد نویز، به ارزیابی موقعیت و شدت خسارت می‌پردازد. خطای بین نتایج شبیه‌سازی شده و تخمین زده شده در میانگین ده اجرا و هر سناریوی آسیب کمتر از ۳ درصد بوده که عملکرد مناسب این روش را در تشخیص کلیه‌ی آسیب‌های ساختاری قاب سه بعدی ۳۷ عضوی و خرپای دو بعدی ۳۳ عضوی اثبات می‌کند. علاوه بر این، آن‌ها نشان می‌دهند که (AGWO) می‌تواند یک ابزار قابل اعتماد برای شناسایی دقیق آسیب در مقایسه با بهینه‌ساز ازدحام ذرات (PSO) و بهینه‌ساز گرگ خاکستری (GWO) ارائه دهد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Eco-friendly Hybrid Concrete Using Pozzolanic Binder and Glass Fibers

K. Shaiksha Vali*[a], B. S. Murugan[a], S. K. Reddy[a], E. Noroozinejad Farsangi[b]

[a] *School of Civil Engineering, Vellore Institute of Technology, Vellore, India*
[b] *Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Kerman, Iran*

PAPER INFO

Paper history:
Received 14 March 2020
Received in revised form 04 May 2020
Accepted 03 June 2020

Keywords:
Environment-friendly
Glass Fiber Concrete
Highly Cement Replaced Concrete
Silica Fume
Super-plasticizer
Ground Granulated Blast Furnace Slag

ABSTRACT

Hybrid Concrete focused on development of buildings, highways, and other structures of civil engineering. In the current study, various mix combinations have been prepared and tested with different percentages of super-plasticizer at different levels of water reduction for obtaining the optimum mix. Further, study on different properties of hybrid concrete and replacement of ordinary portland cement (OPC) with ground granulated blast furnace slag (GGBFS), silica fume (SF) and glass fibers (GF) for obtaining highly cement replaced concrete (HCRC) and glass fiber concrete (GFC). The concrete performance was evaluated based on slump cone test, compressive strength test, split tensile strength test, flexural strength test, water absorption test and ultrasonic pulse velocity test. It was observed from the results that, the best performance of HCRC achieved at 50% GGBFS and 3% silica fume replacement. Further, in the case of GFC, 0.2% of glass fibers showed high performance in terms of split tensile and flexural strength at all ages. The optimized concrete mixtures like HCRC and GFC performed better than the control concrete (CC).

***doi**: 10.5829/ije.2020.33.07a.03*

1. INTRODUCTION

In today's world, the usage of concrete was increased enormously in different construction activities. One of the most important ingredients in the production of concrete was ordinary Portland cement. The high production of concrete involves more manufacturing and utilization of cement. The manufacturing process of cement leads to a huge release of CO_2 which results in environmental problems [1]. Because of this, many investigations have been carried out to find substitutes for cement which are cost-effective and environment-friendly. From the available literature, the substitutes to cement with different industrial by-products like fly ash, GGBFS, silica fume, metakaolin, rice husk ash, etc., had shown improved concrete properties [2–12]. Among various substitutes, GGBFS by-product gives good binding which results in improved artificial aggregates and concrete properties [13–15].

The utilization of various industrial by-products in concrete became popular because of their pozzolanic nature which improves the effective packing of mortar matrix with aggregate results in a solid concrete mix with very fine pore structure [16–18]. For the production of hybrid concrete, one part where the focus required was the mix design of concrete where the correct dosage of super-plasticizer was fixed based on water reduction percentage to improve concrete properties with the maximum replacement of cement content. Moreover, in this study, so many trials were conducted and tested to fix the exact dosage of super-plasticizer with an optimum percentage of GGBFS, SF, and GF. Through this study, the inclusion of GGBFS, SF, and GF to produce HCRC and GFC has been reported.

2. RESEARCH SIGNIFICANCE

The Portland cement which was an important ingredient of ordinary concrete plays the main role in obtaining the strength properties of concrete. But nowadays cement manufacturing leads to a huge release of CO_2 which results in environmental problems. Due to this, the investigation on different properties was carried out to

*Corresponding Author Email: *kolimishaiksha.vali2015@vit.ac.in* (K. Shaiksha Vali)

Please cite this article as: K. Shaiksha Vali, B. S. Murugan, S. K. Reddy, E. Noroozinejad Farsangi, Eco-friendly Hybrid Concrete Using Pozzolanic Binder and Glass Fibers, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1183-1191

substitute maximum cement content with GGBFS, SF with addition of GF to make concrete more effective and economical.

3. EXPERIMENTAL PROGRAM

3. 1. Materials

For the manufacturing of concrete specimens, different materials were used as follows and the physical and chemical characteristics of them are summarized in Table 1.

3. 1. 1. Cement

Ordinary Portland cement (OPC) of 53 grade was utilized in the entire study which was conforming to the BIS: 12269-2013 [19].

3. 1. 2. Industrial By-products

Industrial by-products like GGBFS and SF were used as partial replacement of OPC in this study. Both GGBFS and SF materials have been supplied by Aastra chemicals Chennai, which satisfies the requirements recommended by ASTM C1240-14 and ASTM C1240-15.

TABLE 1. Chemical and physical characteristics of different materials used in this study

Observations	OPC	GGBFS	SF
Chemical Characteristics			
SiO_2	22.3	35	99.88
Fe_2O_3	3	0.95	0.040
Al_2O_3	6.93	17.7	0.043
CaO	63.5	41	0.001
MgO	2.54	11.3	-
TiO_2	-	-	0.001
Na_2O	-	0.2	0.003
K_2O	-	-	0.001
$Ca(OH)_2$	-	-	-
MnO_2	-	2.7	-
SO_3	1.72	-	-
$CaCO_3$	-	10	-
P_2O_5	-	0.65	-
Glass content	-	92	-
Physical Characteristics			
Specific gravity	3.12	2.85	2.63
Appearance (powder)	Grey	Off-white	White
Specific surface area (m^2/kg)	290	409	819
Loss on ignition	0.84	0.26	0.015
pH Value	6.3	-	6.90
Moisture (%)	-	0.10	0.058

3. 1. 3. Aggregates

Crushed stone confirming to graded ordinary aggregates of size not more than 20mm as coarse aggregate and locally available natural river sand was used as fine aggregate which confirming to grading Zone II of BIS: 383-1989 [20]. The sieve analysis of natural aggregate and sand are given in Table 2.

3. 1. 4. Water

Potable tap water was used for the preparation and curing of concrete which confirming to BIS: 456-2000 [21].

3. 1. 5. Chemical Admixture

Commercially available Master Gelenium SKY 8233 super-plasticizer (SP) of specific gravity 1.08 has been used to improve workability, mechanical, and durability properties, which was high range water reducing admixture supplied by BASF, Chennai.

3. 1. 6. Alkali Resistant Glass Fibers

Alkali Resistant glass fibers were added in the production of concrete at different percentages and characteristics of fibers were given in Table 3. It was a lightweight and high tensile material, which was evaluated as per ASTM C1579 [22].

TABLE 2. Sieve analysis of natural aggregates and sand

Size of aggregate (mm)	Percentage of aggregates produced
Natural aggregate	
20	3.2
16	4.1
12.5	34.4
10	44.1
4.75	14.2
Sand	
4.75	4.8
2.36	12.8
1.18	49.6
600	11.6
300	16.4
150	4
Pan	0.8

TABLE 3. Characteristics of alkali-resistant glass fibers

Characteristics	
Specific gravity	2.68
Density	2.7
Elastic modulus (Gpa)	72
Tensile strength (Mpa)	1700
Fiber filament diameter (microns)	14
Length (mm)	12

3. 2. Methodology

3. 2. 1. Mix Proportions

In the present study, M30 grade concrete mix was designed based on the specifications BIS: 10262-2009 [23], BIS: 456-2000 [21], and SP: 29 [24]. Various trials have been performed to fix the correct dosage of super-plasticizer with respect to water reduction and optimum level of cement replacement by GGBFS, silica fume with the addition of glass fibers to attain desired target strength. The various mix combinations were given in Tables 4-7.

3. 2. 2. Samples Preparation

In this part, the materials were mixed properly in a tilting type mixer machine until the concrete attained uniform consistency. Thoroughly mixed concrete as shown in Figure 1(a) was compacted into the required molds in three equal layers (casting) as shown in Figure 1(b) and de-molded after 24 hours, followed by curing in water for 7 and 28 days as shown in Figure 1(c) and then tested at room temperature. The cube specimens of size 100 x 100 x 100 mm were used to conduct the compressive strength, water absorption test, and ultrasonic pulse velocity test. Similarly, the cylindrical specimens of size 200 x 100 mm were used to test split tensile strength and beam specimens of size 500 x 100 x 100 mm were used to test flexural strength.

4. RESULTS AND DISCUSSIONS

Initially to fix the optimum replacement of OPC by GGBFS, SF with the addition of GF various trials to be conducted as follows. Further, CC, HCRC, and GFC were tested with different properties, mix proportions of different concrete were given in Table 8.

TABLE 4. Mixture compositions with super-plasticizer content with water reduction and 7days Compressive strength

Mix ID	Mix Combinations	OPC (kg/m^3)	Water (kg/m^3)	Sand (kg/m^3)	Natural aggregate (kg/m^3)	SP (kg/m^3)	7 Days Strength (MPa)
C0	0%S.P Control	438	197.2	640	1077	-	27
C1	0.1%S.P(16%WR)	369	166	690	1161	0.369	٢٢,٧
C2	0.2%S.P(16%WR)	369	166	690	1161	0.738	24
C3	0.3%S.P(16%WR)	369	166	690	1161	1.107	24.8
C4	0.4%S.P(16%WR)	369	166	690	1161	1.476	23.1
C5	0.5%S.P(16%WR)	369	166	690	1161	1.845	22.9
C6	0.1%S.P(19%WR)	356	160	701	1179	0.356	22.7
C7	0.2%S.P(19%WR)	356	160	701	1179	0.712	27
C8	0.3%S.P(19%WR)	356	160	701	1179	1.068	27.4
C9	0.4%S.P(19%WR)	356	160	701	1179	1.424	26.9
C10	0.5%S.P(19%WR)	356	160	701	1179	1.78	25.9
C11	0.1%S.P(21%WR)	346	156	707	1191	0.346	22.1
C12	0.2%S.P(21%WR)	346	156	707	1191	0.692	23.7
C13	0.3%S.P(21%WR)	346	156	707	1191	1.038	21.3
C14	0.4%S.P(21%WR)	346	156	707	1191	1.384	26.9
C15	0.5%S.P(21%WR)	346	156	707	1191	1.73	27.6
C16	0.6%S.P(21%WR)	346	156	707	1191	2.076	26.6
C17	0.7%S.P(21%WR)	346	156	707	1191	2.422	23.5

TABLE 5. Mixture compositions with GGBFS as cement replacement with 7days compressive strength

Mix ID	Mix Combinations	OPC (kg/m^3)	GGBFS (kg/m^3)	Water (kg/m^3)	Sand (kg/m^3)	Natural aggregate (kg/m^3)	SP (kg/m^3)	7 Days Strength (MPa)
CG1	0.4%S.P+90C+10G(21%WR)	311.4	34.6	156	707	1191	1.384	26.3
CG2	0.4%S.P+80C+20G(21%WR)	276.8	69.2	156	707	1191	1.384	25.1
CG3	0.4%S.P+70C+30G(21%WR)	242.2	103.8	156	707	1191	1.384	25.6
CG4	0.4%S.P+60C+40G(21%WR)	207.6	138.4	156	707	1191	1.384	29.1
CG5	0.4%S.P+50C+50G(21%WR)	173	173	156	707	1191	1.384	28.8
CG6	0.4%S.P+40C+60G(21%WR)	138.4	207.6	156	707	1191	1.384	24.8
CG7	0.5%S.P+90C+10G(21%WR)	311.4	34.6	156	707	1191	1.73	27.9
CG8	0.5%S.P+80C+20G(21%WR)	276.8	69.2	156	707	1191	1.73	28.6
CG9	0.5%S.P+70C+30G(21%WR)	242.2	103.8	156	707	1191	1.73	28.5
CG10	0.5%S.P+60C+40G(21%WR)	207.6	138.4	156	707	1191	1.73	28.9
CG11	0.5%S.P+50C+50G(21%WR)	173	173	156	707	1191	1.73	29.9
CG12	0.5%S.P+40C+60G(21%WR)	138.4	207.6	156	707	1191	1.73	24.9

TABLE 6. Final Optimum Mixture compositions with GGBFS and silica fume as cement replacement

Mix ID	Mix Combinations	OPC (kg/m^3)	GGBFS (kg/m^3)	SF (kg/m^3)	Water (kg/m^3)	Sand (kg/m^3)	Natural aggregate (kg/m^3)	SP (kg/m^3)	7 Days Strength (MPa)
CGS1	0.5%S.P+49C+50G+1SF	171.27	173	1.73	156	707	1191	1.73	29.8
CGS2	0.5%S.P+48C+50G+2SF	169.54	173	3.46	156	707	1191	1.73	30.2
CGS3	0.5%S.P+47C+50G+3SF	167.81	173	5.19	156	707	1191	1.73	31.2
CGS4	0.5%S.P+46C+50G+4SF	166.08	173	6.92	156	707	1191	1.73	29.1
CGS5	0.5%S.P+45C+50G+5SF	164.35	173	8.65	156	707	1191	1.73	27.9

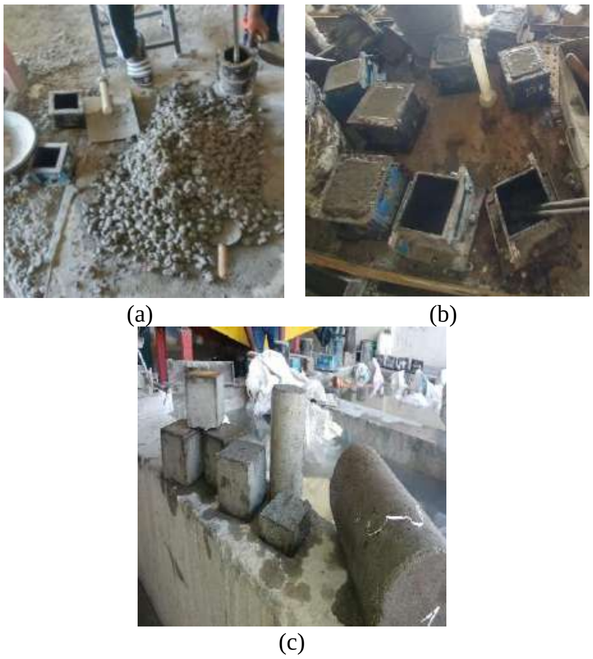
TABLE 7. Final Optimum Mixture compositions with GGBFS and silica fume as cement replacement with glass fibers

Mix ID	Mix Combinations	OPC (kg/m^3)	GGBFS (kg/m^3)	SF (kg/m^3)	Water (kg/m^3)	Sand (kg/m^3)	Natural aggregate (kg/m^3)	SP (kg/m^3)	GF (kg/m^3)	7 Days Strength (MPa)
CGSF1	0.5%S.P+47C+50G+3SF+0.1%GF	167.81	173	5.19	156	707	1191	1.73	2.4	28.9
CGSF2	0.5%S.P+47C+50G+3SF+0.2%GF	167.81	173	5.19	156	707	1191	1.73	4.8	29.7
CGSF3	0.5%S.P+47C+50G+3SF+0.3%GF	167.81	173	5.19	156	707	1191	1.73	7.2	27.3

4. 1. Optimizing the Dosage of SP and Replacements with Different Materials

4. 1. 1. Optimizing Dosage of SP with Respect to Water Reduction

The water-reducing admixture SP was considered in the mix design to reduce the cement content in concrete. To get the optimum mix, various mix combinations with different percentages of super-plasticizer with respect to different water reduction percentages were tested with 7 days compressive strength and given in Table 4. Among 18 mix combinations (C0 to C17), the highest compressive strength was obtained for C14 and C15 mix of 26.9 and 27.6 MPa which was higher than the control concrete (C0) as 27 MPa. Similarly, the lowest compressive strength of 21.3 MPa was obtained for C13 mix which was lower than the control concrete (C0) as presented in Table 4.



(a) (b) (c)

Figure 1. (a) Fresh concrete, (b) Casting specimens, (c) Curing specimens

4. 1. 2. Optimizing OPC Replacement with GGBFS The optimum mixes C14 and C15 were taken and cement replaced with GGBFS at different percentages from 10 to 60% (CG1 to CG12) as given in Table 5. The highest 7 days compressive strength of 29.9 MPa was achieved for CG11 mix which was 8.3% more than C15 control mix. As the cement replacement by GGBFS at all the levels the compressive strength was obtained higher. From the results, it was observed that 50% replacement with GGBFS was optimum which obtained the desired strength.

4. 1. 3. Optimizing OPC Replacement with GGBFS and SF The optimum mix CG11 was selected from Table 6 and OPC has been replaced with SF from 1 to 5% (CGS1 to CGS5). The highest 7 days compressive strength of 31.2 MPa was achieved at 3% SF (CGS3) which was 4.3% more than CG11 mix.

The finer size of SF with high pozzolanic nature was responsible to achieve good strength. Hence, it was observed that using 50% GGBFS and 3% SF replacement will attain the desired target strength.

4. 1. 4. Optimizing OPC Replacement with GGBFS and SF with the Addition of GF The optimum mix CGS3 was taken and added GF at 0.1 to 0.3% (CGSF1 to CGSF3) by the total volume of concrete is given in Table 7. The highest 7 days compressive strength was noted 29.7 MPa at 0.2% glass fibers which were higher than the CC.

4. 2. Compressive Strength The 7, 28 days compressive strengths of different types of concrete were tested by a universal testing machine as shown in Figure 2(a) and the results were given in Table 9. It was noticed that the HCRC mix showed higher compressive strength than CC, because of high CaO and less Al_2O_3 content which results in from a pozzolanic reaction. The compressive strength of HCRC and GFC was slightly higher around 2% than CC but with the addition of GF, 1% declined in compressive strength was observed. Compressive strength values were decreased for OPC replacement beyond the optimum percentage because of escaping out of excess lime that leads to a decline in pore bonding strength [25–28].

4. 3. Split Tensile Strength The split tensile test was performed as per BIS: 516-1959 [29] in a universal testing machine as shown in Figure 2 (b) and the strength value of different mixes were shown in Table 9. From the results, it was noticed that the positive influence of glass fibers on split tensile strength. The highest split tensile strength of 4.73 MPa was observed for the GFC mix and lowest for CC mix of 4.12 MPa. With respect to CC mix tensile strength was increased by about 11% for HCRC mix and 14.8% for GFC mix at 28 days. With addition of 0.2% GF in HCRC mix 3.5 % more split tensile strength was achieved for the GFC mix. By replacing OPC with GGBFS, SF, the interfacial transition zone (ITZ) becomes solid which results in the enhancement of tensile strength [12, 26]. The GF used in the present study has 6-12 mm length which increases the resistance of concrete against splitting. A similar performance of GF at an optimum dosage has been noted in earlier studies [30].

TABLE 8. Final optimum Mixture compositions with GGBFS, SF with GF

Mix ID	Mix Combinations	OPC (kg/m^3)	GGBFS (kg/m^3)	SF (kg/m^3)	Water (kg/m^3)	Sand (kg/m^3)	Natural aggregate (kg/m^3)	SP (kg/m^3)	GF (kg/m^3)
*CC	0%S.P+100C	438	-	-	197.2	640	1077	-	-
HCRC	0.5%S.P+47C+50G+3SF	167.81	173	5.19	156	707	1191	1.73	-
GFC	0.5%S.P+47C+50G+3SF+0.2GF	167.81	173	5.19	156	707	1191	1.73	4.8

(a) (b)

Figure 2. (a) Compressive strength, (b) Split tensile strength

4. 4. Flexural Strength

Flexural strength test was conducted as per BIS: 516-1959 [29] in the flexural testing setup as shown in Figure 3 and the values of various mixes were presented in Table 9. It was observed from the results that the highest flexural strength of 6.03 MPa for GFC mixes and lowest of 4.91 MPa for CC mix at 28 days. With respect to CC mix, flexural strength was increased by about 18.3% for the HCRC mix and 22.8 % for the GFC mix. With addition of 0.2% GF in HCRC mix, 3.8 % more flexural strength was attained for the GFC mix. From the above results, the utilization of GGBFS, SF, and GF enhanced the strength for all mixes, in comparison with the CC mix.

4. 5. Water Absorption

The water absorption test was conducted as per ASTM C642-2013 [31], by oven dry process after 7, 28 days of specimen curing as shown in Figure 4. The effect of GGBFS, silica fume with glass fibers on the water absorption presented in Figure 5. The highest water absorption was observed for CC mix as 2.4% and lowest for GFC mix as 1.95%. From the results, it was noticed that the HCRC and GFC mixes show 11.7 and 18.7% lesser water absorption than CC mix. Similarly, the GFC mix shows 8% lesser water absorption value than the HCRC mix. The above test results show that lower water absorption values have higher compressive strengths. The lower water absorption may occur because of higher pozzolanic effect by GGBFS and SF which results in a decrease in pore structure to produce denser concrete [4, 5].

Figure 3. Flexural strength

Figure 4. Water absorption test

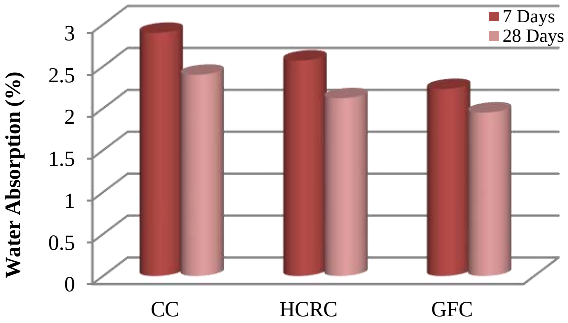


Figure 5. Water absorption values

TABLE 9. Mechanical properties of Final optimum mixtures

Mix Type	Compressive Strength (MPa)		Split Tensile Strength (MPa)		Flexural Strength (MPa)	
	7 Days	28 Days	7 Days	28 Days	7 Days	28 Days
CC	27.0	38.9	3.26	4.12	3.67	4.91
HCRC	31.2	39.7	3.18	4.57	4.19	5.81
GFC	29.7	39.3	3.28	4.73	4.52	6.03

4. 6. Ultrasonic Pulse Velocity Ultrasonic pulse velocity (UPV) test was an indicator to check the homogeneity of concrete in the form of porosity and permeability as per BIS: 13311(1) – 1992 [32] as shown in Figure 6. A higher UPV value was generally related to a solid structure of concrete, in which all the results show excellent quality. The UPV values for different mixes at 7, 28 days was represented in Figure 7. The UPV values for HCRC and GFC mixes were higher than CC mix at 7 and 28 days which exhibit an excellent quality of concrete. From the results, it was observed that GGBFS, SF have lesser specific gravity than OPC which helps the concrete to form dense structure results in enhancement of characteristics of concrete.

4. 7. Cost Analysis In this study, the different concrete mixes were produced with the replacement of OPC by GGBFS, SF with addition of GF. So, this section aims to study the entire cost obtained in the production of CC, HCRC, and GFC mixes. The mixes were compared with each other with the available market prices of various materials used in the production of concrete. The cost of various mixes presented in Table 10 was calculated for one meter cube based on the quantity of materials as per the final mix design. Based on the results, the CC mix is costlier than HCRC and FRC mixes. The highest cost savings in the production of CC to HCRC mix is 26% and followed by CC to FRC mix is 18.3%. Therefore, concrete production with HCRC and GFC mix will have ecological and economical benefits in practice.

Figure 6. UPV test

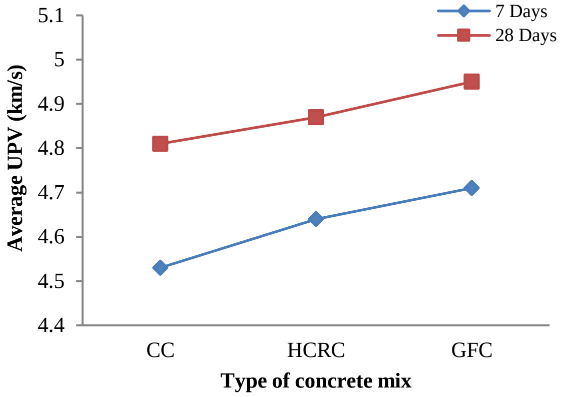


Figure 7. Average ultrasonic pulse velocity values

TABLE 10. Materials and cost per meter cube of concrete for different mixes

Mixture ID		CC		HCRC		GFC	
Materials	Cost per kg (US $)	Materials and cost per m^3		Materials and cost per m^3		Materials and cost per m^3	
		Materials (kg)	Cost (US $)	Materials (kg)	Cost (US $)	Materials (kg)	Cost (US $)
OPC	0.10	438	43.8	167.81	16.78	167.81	16.78
GGBFS	0.029	-	-	173	5.02	173	5.02
SF	0.11	-	-	5.19	0.57	5.19	0.57
Sand	0.0066	640	4.23	707	4.67	707	4.67
Natural Aggregate	0.013	1077	14.0	1191	15.48	1191	15.48
SP	1.98	-	-	1.73	3.42	1.73	3.42
GF	0.99	-	-	-	-	4.8	4.75
Water	0.0013	197.2	0.26	156	0.20	156	0.20
Cost of concrete per m^3 (US $)			62.29		46.14		50.89

5. CONCLUSIONS

Based on the experimental investigations carried out on different mixes, the following conclusions were drawn.

1. So many trials were conducted to fix the correct dosage of super-plasticizer with respect to water reduction percentage before utilizing in the mass concrete applications.

2. To achieve the desired target strength, an optimum of 0.5% of SP with 21% water reduction was used in the entire study.
3. By utilizing the combination of GGBFS, SF, and GF had improved the particle filling and pore structure which tends to the enhancement of all the concrete properties.
4. The higher test results were observed with the mix containing 50% GGBFS, 3% SF and 0.2% GF. Because of the higher specific surface area of materials have high pozzolanic action which results in C-S-H gel which helps in improving the concrete properties.
5. The cost to produce HCRC mix reduces to 26% when compared with CC mix. Similarly, the cost to produce FRC mix reduces to 18.3% when compared with CC mix.
6. Utilizing the combination of GGBFS and SF at high percentages as a substitute for OPC produces ecological and sustainable concrete.

6. REFERENCES

1. Shahba, S., Ghasemi, M., and Marandi, S. M., " Effects of Partial Substitution of Styrene-butadiene-styrene with Granulated Blast-furnace Slag on the Strength Properties of Porous Asphalt", ***International Journal of Engineering, Transactions A: Basics***, Vol. 30, No. 1, (2017), 40–47. doi:10.5829/idosi.ije.2017.30.01a.06
2. Martin, A., Pastor, J.Y., Palomo, A. and Jiménez, A.F., "Mechanical behaviour at high temperature of alkali-activated aluminosilicates (geopolymers)", ***Construction and Building Materials***, Vol. 93, (2015), 1188–1196. doi:10.1016/j.conbuildmat.2015.04.044
3. Shaiksha Vali, K. and Bala Murugan, S., "Effect of different binders on cold-bonded artificial lightweight aggregate properties", ***Advances in Concrete Construction***, Vol. 9, No. 2, (2020), 183–193. 183–193, doi:10.12989/acc.2020.9.2.183
4. Li, H., Zhang, M.H. and Ou, J. P., "Flexural fatigue performance of concrete containing nano-particles for pavement", ***International Journal of Fatigue***, Vol. 29, No. 7, (2007), 1292–1301. https://doi.org/10.1016/j.ijfatigue.2006.10.004
5. Shaiksha Vali, K. and Bala Murugan, S., "Utilization of cementitious materials with cold-bonded artificial aggregate in concrete", ***International Journal of Engineering and Advanced Technology***, Vol. 9, No. 1, (2019), 385–388. doi:10.35940/ijeat.A9376.109119
6. Huang, X., Ranade, R., Zhang, Q., Ni, W. and Li, V.C., "Mechanical and thermal properties of green lightweight engineered cementitious composites", ***Construction and Building Materials***, Vol. 48, (2013), 954–960. doi:10.1016/j.conbuildmat.2013.07.104
7. Morsy, M.S., Al-Salloum, Y.A., Abbas, H. and Alsayed, S.H., "Behavior of blended cement mortars containing nano-metakaolin at elevated temperatures", ***Construction and Building Materials***, Vol. 35, (2012), 900–905. doi:10.1016/j.conbuildmat.2012.04.099
8. Khalaj, G. and Nazari, A., "Modeling split tensile strength of high strength self compacting concrete incorporating randomly oriented steel fibers and SiO 2 nanoparticles", ***Composites Part B: Engineering***, Vol. 43, No. 4, (2012), 1887–1892. doi:10.1016/j.compositesb.2012.01.068
9. Shaiksha Vali, K., Bala Murugan, S., and Murugan, B., "Impact of Nano SiO_2 on the Properties of Cold-bonded Artificial Aggregates with Various Binders Impact of Nano SIO_2 on the Properties of Cold-bonded Artificial Aggregates with Various Binders", ***International Journal of Technology***, Vol. 10, No. 5, (2019), 897–907. doi:10.14716/ijtech.v10i5.2590
10. Shi, X., Xie, N., Fortune, K. and Gong, J., "Durability of steel reinforced concrete in chloride environments: An overview," ***Construction and Building Materials***, Vol. 30, (2012), 125–138. https://doi.org/10.1016/j.conbuildmat.2011.12.038
11. Said, A.M., Zeidan, M.S., Bassuoni, M.T. and Tian, Y., "Properties of concrete incorporating nano-silica", ***Construction and Building Materials***, Vol. 36, (2012), 838–844. doi:10.1016/j.conbuildmat.2012.06.044
12. Rath, B., Deo, S., and Ramtekkar, G., "Durable Glass Fiber Reinforced Concrete with Supplimentary Cementitious Materials", ***International Journal of Engineering, Transactions A: Basics***, Vol. 30, No. 7, (2017), 964–971. doi:10.5829/ije.2017.30.07a.05
13. Turu'allo, G., "Using ggbs for Partial Cement Replacement in Concrete: Effects of Water-binder Ratio and ggbs Level on Activation Energy", ***International Journal of Technology***, Vol. 5, (2015), 327–336. doi:10.14716/ijtech.v6i5.1916
14. Choucha, S., Benyahia, A., Ghrici, M. and Mansour, M.S., "Correlation between compressive strength and other properties of engineered cementitious composites with high-volume natural pozzolana", ***Asian Journal of Civil Engineering***, Vol. 19, No. 5, (2018), 639–646. doi:10.1007/s42107-018-0050-3
15. Shaiksha Vali, K. and Bala Murugan, S., "Properties of glass fiber reinforced cold-bonded artificial lightweight aggregates with different binders", ***Revista Română de Materiale / Romanian Journal of Materials***, Vol. 50, No. 1, (2020), 40–50. https://www.researchgate.net/publication/339935578
16. Erdoğan, T., Admixtures for concrete, Middle East Technical University Press., (1997).
17. Mehta, P. and Monteiro, P., Concrete microstructure, properties and materials, New York: McGraw-Hill, (2017).
18. Sharma, S.K., Kumarb, P. and Roya, A.K., "Comparison of Permeability and Drying Shrinkage of Self Compacting Concrete Admixed with Wollastonite Micro Fiber and Fly Ash", ***International Journal of Engineering, Transactions B: Applications***, Vol. 30, No. 11, (2017), 1681–1690. doi:10.5829/ije.2017.30.11b.08
19. BIS 12269., 'Ordinary portland cement 53 grade specification', New Delhi, India, (2013).
20. BIS 383., 'Specification for Coarse and Fine Aggregates from Natural Sources for Concrete', New Delhi, India, (2016).
21. BIS 456., 'Plain and Reinforced Concrete - Code of Practice is an Indian Standard code of practice for general structural use of plain and reinforced concrete', New Delhi, India, (2000).
22. ASTM International C1579-13., 'Standard Test Method for Evaluating Plastic Shrinkage Cracking of Restrained Fiber Reinforced Concrete (Using a Steel Form Insert)', (2013).
23. BIS 10262., 'Concrete mix proportioning – guidelines', New Delhi, India, (2009).
24. SP 29., 'Specification for concrete mix design', Bureau of Indian standards, New Delhi, India.
25. Shaiksha Vali, K. and Bala Murugan, S., "Influence of industrial by-products in artificial lightweight aggregate concrete: An Environmental Benefit Approach", ***Ecology, Environment and Conservation***, Vol. 26, (2020), S233–S241. http://www.envirobiotechjournals.com/EEC/FebSupplIssue2020/EEC-33.pdf
26. Nazari, A. and Riahi, S., "Microstructural, thermal, physical and mechanical behavior of the self compacting concrete containing SiO2 nanoparticles", ***Materials Science and Engineering A***, Vol. 527, No. 29–30, (2010), 7663–7672. doi:10.1016/j.msea.2010.08.095

27. Jindal, B.B., Singhal, D., Sharma, S.K. and Ashish, D. K., “Improving compressive strength of low calcium fly ash geopolymer concrete with alccofine”, ***Advances in Concrete Construction***, Vol. 5, No. 1, (2017), 17–29. https://doi.org/10.12989/acc.2017.19.2.017

28. Zhang, M. H. and Li, H., “Pore structure and chloride permeability of concrete containing nano-particles for pavement”, ***Construction and Building Materials***, Vol. 25, No. 2, (2011), 608–616. doi:10.1016/j.conbuildmat.2010.07.032

29. BIS 516., ‘Methods of Tests for Strength of Concrete’, New Delhi, India, (1959).

30. Das, C.S., Dey, T., Dandapat, R., Mukharjee, B.B. and Kumar, J., “Performance evaluation of polypropylene fibre reinforced recycled aggregate concrete”, ***Construction and Building Materials***, Vol. 189, (2018), 649–659. doi:10.1016/j.conbuildmat.2018.09.036

31. ASTM-C642-13., “Standard Test Method for Density, Absorption, and Voids in Hardened Concrete, (2013).

32. BIS 13311(part 1)., ‘Specification for Non Destructive Testing for Concrete’, New Delhi, India, (1992).

Persian Abstract

چکیده

تمرکز استفاده از بتن‌های ترکیبی در ساخت و توسعه ساختمان‌ها، بزرگراه‌ها و سایر سازه‌های مهندسی عمران می‌باشد. در مطالعه حاضر، طرح مخلوط‌های متفاوتی از بتن‌های ترکیبی با درصدهای مختلف فوق روان‌کننده، مقادیر مختلف درصد آب جهت تعیین طرح بهینه در آزمایشگاه مورد بررسی قرار گرفته است. همچنین جایگزینی سیمان پرتلند معمولی با الیاف شیشه، دوده سیلیسی و سرباره کوره آهن‌گدازی به منظور حصول نتیجه بهینه و دریافت بتن با سیمان جایگزین شده نیز مورد بررسی قرار گرفت. در نهایت عملکرد بتن تولید شده توسط تست اسلامپ، تست مقاومت فشاری، تست مقاومت کششی، تست مقاومت خمشی، تست جذب آب و تست پالس اولتراسونیک مورد ارزیابی قرار گرفت. نتایج حاصله بیانگر عملکرد بهینه بتن ترکیبی در صورت استفاده از ۵۰٪ سرباره کوره آهن‌گدازی و ۳٪ دوده سیلیسی بوده است. همچنین در صورت استفاده از ۰/۲ ٪ الیاف شیشه، عملکرد کششی و خمشی بتن به طور قابل توجهی در تمامی سنین بهبود پیدا کرده است. در نهایت نشان داده شد که عملکرد بتن ترکیبی پیشنهادی به مراتب نسبت به بتن مرجع بهتر بوده است.

International Journal of Engineering

Journal Homepage: www.ije.ir

Bandwidth and Delay Optimization by Integration of Software Trust Estimator with Multi-user Cloud Resource Competence

S. M. Mirrezaei*

Faulty of Electrical and Robotic Engineering, Shahrood University of Technology, Shahrood, Iran

PAPER INFO

Paper history:
Received 27 March 2020
Received in revised form 02 June 2020
Accepted 08 June 2020

Keywords:
Bandwidth
Behavior Trust Estimator
Cloud Resources
Competence Optimization
Software-as-a-Service
User Feedback

ABSTRACT

Trust establishment is one of the significant resources to enhance the scalability and reliability of resources on cloud environments. To establish a novel trust model on SaaS (Software as a Service) cloud resources and to optimize the resource utilization of multiple user requests, an integrated software trust estimator with multi-user resource competence (IST-MRC) optimization mechanism is proposed in this paper. IST-MRC optimization mechanism combines its trustworthy properties with optimal resource allocation, on the requisition of the software apps, without any traffic occurrence in the cloud environment. Initially, a behavior trust estimator is developed in the IST-MRC mechanism to measure the trust value of the software service zone. The trust value is estimated, based on Software Availability Rate, Hit Rate, and User Feedback regarding the specific software apps. Next, the resources are optimized to multiple users using competence optimization. The competence optimization in the IST-MRC mechanism computes the processor speed, bandwidth, and latency to handle the varied traffic conditions on multiple user requests. Experiments are conducted to measure and evaluate factors, such as Successful Request Handles, Resource Utilization Efficiency, Latency Time, and Trust Success Ratio on the multiple users.

***doi**: 10.5829/ije.2020.33.07a.04*

1. INTRODUCTION

The ever-increasing and broadening limelight, tendered to the cloud computing environment, presents a natural means to expand the potential of content delivery networks to sustain the progress of the content linked to the text, image, and the like. The dynamic request redirection (DRR-CCMN) of cloud-centric media network [1] considers the drawback of redirecting the users' requests to multiple destination virtual machines (VMs) optimally and most favorably to lessen the cost of computation. The request arrival process is also geared up in such a way that it can effectively switch on between the two models - the normal and flash crowd. Nevertheless, DRR-CCMN has been substantiated to be unproductive when it is used between multiple dispatchers, with diverse traffic circumstances.

With umpteen intricacies mixed up with the distinct traffic conditions, one of the essential requirements for the accomplishment of the clouds, lies in the efficient organization of the task of the cloud virtual machines. The existing cloud schedulers not only overlook the overall cloud infrastructure but also ignore the entire infrastructural properties and, in doing so, end up with various security-related issues, hitches in the privacy of the data, and so on. A cloud scheduler, using the OpenStack (CS-OpenStack) prototype [2], presents a novel cloud scheduler, which takes into consideration the requirements of the users as well as the infrastructural properties.

In reference [3], the task of counting the third party auditor (TPA) is initiated to grant genuineness to cloud data storage. The TPA, in turn, verifies the data integrity of the dynamic cloud data storage that resolutely avoids any addition of a client to achieve the economies of scale for the cloud computing environment.

This ensures more protection during the processing of the multiple auditing tasks; however, simultaneously, there are apprehensions of losing control over their data. On account of Internet usage has reached a new high,

*Corresponding Author's Email: *sm.mirrezaei@shahroodut.ac.ir* (S. M. Mirrezaei) cloud computing environment affords dynamic scalability. Despite enjoying the privileges given by this new epoch, there are apprehensions of losing grip over the data. So a decentralized information accountability framework [4] is introduced to keep track of the users' data usage in the cloud computing environment via an object-centered approach. Not withstanding the inclusion of proficient and resourceful auditing mechanisms, the component of trust remains unattended.

In trustworthy clouds (TClouds) [5], the focus is on constructing trust models that include different levels of transparency and the establishment of trust. It has been ascertained to be beneficial and advantageous to both the cloud providers and the users; yet, here too, the procedure of evaluating or measuring the trust values continues to be left unaddressed.

In literature [6], a cloud-based information repository is used, which significantly concentrates on the notable aspects of the cloud computing environment, which amalgamate the users' health data and make an accessible analysis of the data, using a trusted third party. As a result, a trust framework [7] is evolved to obtain a broad set of parameters to, not only ascertain the trust, but also includes the cost factor that enables it to establish and negotiate. It includes a trust model to prop up the policy integration, using a high-powered approach.

Internet appliances compass electronic commerce, interactions between the users through electronic mail, and so on; nevertheless, there is still a higher level of unease regarding the trustworthiness of these types of devices. To boost the level of security, in literature [8], a trust evaluation model has been devised based on D-S evidence theory with sliding windows, using the cloud computing environment. It is endowed with opportunities to effectively identify malicious users and offer reliable information in making correct decisions regarding the system. However, the collision behavior remains unanswered.

In general, the framework applies to most of the trust-based models, shared on the software apps, where multiple users evaluate the software trust model utilizing the trust success ratio. Based on techniques as mentioned above and methods, I propose an integrated software trust estimator with a multi-user resource competence (IST-MRC) optimization mechanism to provide a novel trust model in software as a service (SaaS). Trust is the estimation of the cloud software environmental values, and the software trust model of IST-MRC mechanism evaluates the trust value of the resources, based on the observed behavior of the cloud computing environment. With this, the trust management mechanism in the IST-MRC reduces the complexity level of cloud service provisioning. As a result, the resource utilization in the IST-MRC mechanism provides entirely distributed resources to its multiple users, without any traffic occurrence. IST-MRC optimization mechanism combines its trustworthy properties with optimal resource allocation, on the requisition of the software apps, without any traffic occurrence in the cloud environment. This framework, besides using the behavior-based trust model, utilizes competence optimization to enhance its resource utilization efficiency. Extensive experiments exhibit that the optimization mechanism is equipped with a superior level of successful requests on the users with less Latency Time.

The main contributions of this paper are the introduction of:

- The proposed model uses an integrated software trust estimator with a multi-user resource competence (IST-MRC) optimization mechanism and leads to overcome certain drawbacks of the existing trust models in the cloud computing environment.
- IST-MRC mechanism uses the competence optimization method to utilize the higher processing speed of RAM with higher bandwidth for several specific software apps.
- The proposed optimization mechanism estimates the degree of resource utilization efficiency, based on the plea placed by the cloud user. Therefore, it evaluates the trust success ratio in a resourceful manner.
- To maximize the resource utilization efficiency on the multiple users, the processor speed, bandwidth, and latency to handle the wide-ranging traffic conditions on the multiple user requests are deliberated. According to the results obtained my approach is apt for multiple users with its wide-ranging traffic conditions.

The remaining sections are systematized as follows. In Section 2, an overview of the related works on the trust-based model in a cloud computing environment is deliberated. Section 3 analyzes the general outline of problem descriptions and lays on solutions accordingly. Section 4 elucidates my proposed trust-based model employing multiple users. Section 5 interprets the experimental results pertained to the validation of the performance of the proposed method. Section 6 puts across the concluding observations.

2. RELATED WORKS

Despite attaining tremendous insights, the primary concern being security, cloud computing is still at its early stage. In literature [9], security, trust, as well as the privacy of a cloud computing environment, has been addressed. An elaborate analysis has been executed on these three imperative factors; however, their deployment in the real-time environment continues to be a significant concern, and the time factor has not been scrutinized in a broader sense as well. In reference [10], a novel framework has been designed to support the application of knowledge discovery in a cloud computing environment, and it ably predicts its application execution time, using Rough Sets Theory; nevertheless, to a greater extent, the uncertainty still exists. Though the cloud computing environment is more suitable for representing trust, in the presence of dynamic updates, it is unwise to include trust. Strategic provisioning of cloud computing services [11] has been initiated to prefer amongst the paramount services. Lack of standard specification and low ratings are some of the issues, which persist. In reference [12], a novel, unfair rating filtering method is set up to furnish reputation and accordingly to opt for the most excellent services.

With the mounting utilization of the Internet, and its applications as well as the resources being delivered to the cloud users, on-demand, data security and privacy concerns have to be given prime attention to increase the users' responsiveness. In literature [13], a comprehensive comparative analysis was made to address the issues related to data security and privacy protections in the cloud computing environment. To curtail the consumption of energy in the cloud data center, virtual machine allocation algorithm [14] is put forward, using efficient resource allocation and particle swarm optimization (PSO) method. Even though the complexities mixed up with the computation are addressed, only, certain resources like CPU and disk have been incorporated. To address the challenges involved in the security aspect of the cloud computing environment, cloud authentication [15] is to be enhanced to provide a resourceful mechanism against a phishing attack in the initial stages.

The cloud computing environment is a framework that enables on-demand network access amongst its various users. One of the major features to be well-thought-out related to the quality of service is reliability, which has to be dealt with competently. An efficient trust assessment [16] is introduced before the sharing of the infrastructure, and it results in scalable and reliable service for its legitimate users.

One of the highest concerns related to cloud data storage is its inefficient data verification procedure and its reliability; it is a holdup in endorsing trust in the cloud computing environment together with SaaS [17-20].

In reference [21], cloud broker and its need in the related cloud environment are discussed. Also, cloud brokering frameworks of interconnected cloud environments are reviewed. Its authors presented a taxonomy of cloud brokering techniques and comparative analysis of cloud brokering techniques.

Xie et al. [22] proposed a trust model for a container cloud environment, which uses direct trust, recommendation trust, and cooperative trust to calculate the comprehensive trust degree in three trust ways. The results of the simulation experiments showed that the model can effectively solve the trusted problem in the container-based cloud.

Another article in this field is an attempt to address the issue of privacy of big data distributed in the field of cloud computing [23]. One of the most essential data privacy issues is authentication and protection of proprietary data. In the article above, this privacy issue is analyzed by Petri net modeling.

Nowadays, due to the advances in mobile and wireless communication, mobile devices are widely used in daily life. Meantime, in mobile devices, there exist diverse applications that are developed to satisfy the various requirements of mobile users. To relieve this problem, driven by edge computing, the central units (CUs) in fifth-generation wireless systems (5G) could be enhanced into edge nodes (ENs) for processing [24].

Cloud computing, the long-standing dream of computing as a tool, can transform a large part of the information technology industry and make the software even more attractive as a service and shape the way IT hardware is designed and purchased. The authors of [25] provided an architectural framework and principles for energy-efficient cloud computing environments. They call the resource allocation algorithms the load-power-aware in this architecture. The algorithm uses a heuristic to dynamically improve energy efficiency in the data center while guaranteeing the QoS.

3. OVERVIEW

3. 1. Problem Statement

The proposed in this paper uses the IST-MRC optimization mechanism, and it is set up to overcome certain drawbacks of the existing trust models in the cloud computing environment. Also, it is apt for multiple users with its wide-ranging traffic conditions. Before going into the facets of the proposed software trust model, it is requisite to define the scope of the study.

Cloud infrastructure is an archetype that primarily concentrates on data sharing, information, or resources over a massive network. The foremost notion behind cloud infrastructure is to emphasize computing with the storage available at any time and at anywhere. With the increasing use of internet-based cloud infrastructure, the availability of the two most vital elements - trust and security - becomes indispensable. An appropriate trust model provides optimal utilization of the resources, under different traffic conditions.

3. 2. Overview of the Entire Software Trust Model

In this section, the software trust model is briefly addresed. The software service provisioning IST-MRC is shown in Figure 1. In general, the framework applies to most of the trust-based models, shared on the software apps, where multiple users evaluate the software trust model utilizing the trust success ratio. Trust is the estimation of the cloud software environmental values, and the software trust model of IST-MRC mechanism evaluates the trust value of the resources, based on the observed behavior of the cloud computing environment. The evaluation of the cloud trust, ultimately, checks whether proper legal materials of the software are provided to the cloud server. The availability of the legal documents, therefore, ensures and establishes mutual trust between the resource providers and the users. With this, the trust management mechanism in the IST-MRC reduces the complexity level of cloud service provisioning. As a result, the resource utilization in the IST-MRC mechanism provides entirely distributed resources to its multiple users, without any traffic occurrence. The software service provisioning, through behavior trust estimator, is represented in Figure 1. Subsequently, the cloud end-users place the request of the needed software apps. The provisioning of the software services, initially, checks the trust value. The trust value is checked, using the IST-MRC mechanism by computing the Software Availability Rate, Hit Rate, and User feedback. The Software Availability Rate, initially, checks whether the licensed software (i.e. the legal documents) is accessible in the cloud zone. Followed by this, the Hit Rate measures the success ratio of the usability between the end-users. To end with, user feedback is also analyzed, using the IST-MRC mechanism to identify the quality of the software service, provided during the transaction.

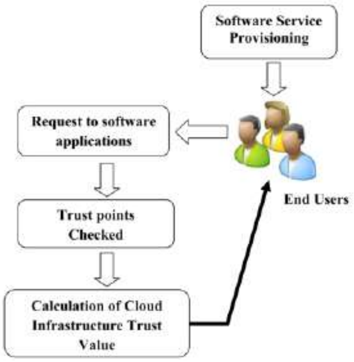

Software Service Provisioning
Request to software applications
End Users
Trust points Checked
Calculation of Cloud Infrastructure Trust Value

Figure 1. Software service provisioning in the IST-MRC mechanism

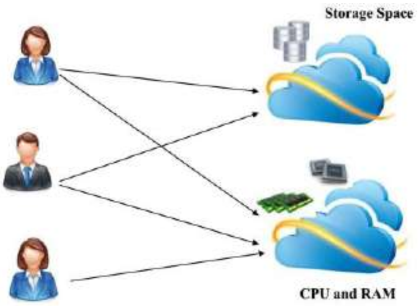


Figure 2. Multiple user requests to cloud resources

The multiple users request the resources for efficient utilization of the software apps for their systems. IST-MRC mechanism uses the competence optimization method to utilize the higher processing speed of RAM with higher bandwidth for several specific software apps. The multiple user requests on cloud infrastructure are depicted in Figure 2. As illustrated in this figure, the resources are made available for the various users, (i.e., User 1, User 2, User 3), depending on the requisition made by the cloud users of the cloud infrastructure. User 1 and User 2 in Figure 2 request for CPU processor and storage space, whereas User 3 requires only a CPU processor for processing the software apps. The whole requests are fulfilled not only with minimal latency time but also without any traffic occurrences.

4. SOFTWARE TRUST MODEL INTEGRATED WITH MULTI-USER RESOURCE COMPETENCE OPTIMIZATION

In this section, the optimization mechanism is described to ascertain a novel trust model and to optimize the resource utilization of the multiple user requests. The proposed optimization mechanism estimates the degree of resource utilization efficiency, based on the plea placed by the cloud user. Therefore, it evaluates the trust success ratio in a resourceful manner. The overall framework of the IST-MRC optimization mechanism is depicted in Figure 3. As illustrated in this figure, the multiple users entreat the cloud server for the software application. The cloud server builds up a trust model in the IST-MRC mechanism using a behavior trust estimator. The behavior of the software apps is perceived, and the trust values are assessed to make out its service provisioning efficiency. In conjunction with this, the trust value is evaluated with a higher scalability ratio, and then the integration work is carried out using the IST-MRC

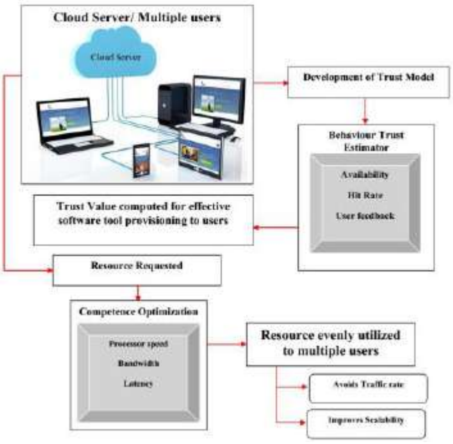

Cloud Server/ Multiple users
Cloud Server
Development of Trust Model
Behaviour Trust Estimator
Availability
Hit Rate
User feedback
Trust Value computed for effective software tool provisioning to users
Resource Requested
Competence Optimization
Processor speed
Bandwidth
Latency
Resource evenly utilized to multiple users
Avoids Traffic rate
Improves Scalability

Figure 3. Overall structural diagram of the IST-MRC optimization mechanism

mechanism. The users (i.e. multi-users) in the cloud zone are not provisioned with the valuable resources in an attempt to run the software apps. Consequently, the multiple users implore the cloud zone for the resources to run the application with a higher reliability rate. The IST-MRC mechanism uses the competence optimization method to handle the entire users appeal under different traffic conditions.

4. 1. Behavior Trust Estimator The user behavior trust estimator in the IST-MRC mechanism measures the performance factor of the Availability Rate (A_i), Hit Rate (H_i), and User Feedback Rate (UF_i). The Availability Rate of the software apps in the cloud zone refers to the ratio of the number of times the specific software apps are requested to the total number of times the software has been queried for the operation (i.e., transaction). The availability rate of the application in the IST-MRC mechanism is computed as,

$$A_i = \frac{Number\ of\ times\ the\ software\ requested}{Total\ number\ of\ times\ the\ software\ required} \qquad (1)$$

Assume '5' to be the count on the software requested in the cloud zone and '10' to be the number of times the software queried for the transaction.

$A_i = \frac{5}{10} = 0.5$

The availability rate or availability percentage for the specific software apps is subsequently evaluated to be '0.5'. The software hit ratio in the IST-MRC mechanism refers to the ratio of the number of successful operations in the cloud to the total number of operations placed over the specific software.

$$H_i = \frac{Number\ of\ successful\ operations}{Total\ number\ of\ requests\ placed\ over\ the\ operation} \qquad (2)$$

To evaluate the hit rate, let us assume that '5' successful operations are carried out in adherence to the '5' successful requests placed over the cloud zone.

$H_i = \frac{5}{5} = 1$

The IST-MRC mechanism achieves a higher hit ratio, without any delay when providing the software services to the end-users. In line with this, the higher hit ratio trust value offers better user feedback to the cloud system. The user feedback UF_i in the IST-MRC mechanism collects the feedback and identifies the trust value, based on the feedback obtained from the repository database. The user feedback ensures the scalability ratio of the cloud zone.

The feedback is dispersed over a range of 0 to 1 in the IST-MRC mechanism, where 1 represents the most trustworthy software collection, and 0 represents the non-trustworthy software collection. The feedback value for the above-computed hit ratio, UF_i is '1', and the feedback varies vigorously, based on the hit ratio attained during the operation. The algorithmic description of Behavior Trust Estimator is described as,

```
Behavior Trust Estimator
Input: Assume users {$U_1, U_2, U_3 ..., U_n$} for the software apps service requisition
Output: Trust Value computed with Higher Scalability Value
Begin
  Step 1: Initially measure the Availability Rate $A_i$ [shown in (1)]
  Step 2: Measure the Software Hit Rate $H_i$ [shown in (2)]
  Step 3: Final step checks the User Feedback, using (2) Measure Rate
     Fuzzy points '0' or '1' is computed in $UF_i$
  Step 4: Sum up the Availability Rate, Hit Rate, and User Feedback values
     Compute $BTE = A_i + H_i + UF_i$
End
```

The algorithm clearly describes the evaluation of behavior trust estimator value through step by step process. The behavior trust estimator is measured as the overall sum of the Availability Rate, Hit Rate, and User Feedback Rate. The behavior trust is formularized as,

$$Behavior\ Trust\ Estimator\ (BTE) = A_i + H_i + UF_i \qquad (3)$$

BTE value through the above-illustrated points is expressed as,

BTE = 0.5+1+1= 2.5

The BTE value attained from the assumed points is about '2.5'; finally, the behavior trust estimator achieves a higher scalability rate, using the IST-MRC mechanism, when compared with the existing systems.

4. 2. Competence Optimization The IST-MRC mechanism allocates the resources to the multi-user requests, under different traffic conditions. The resources, such as CPU processor of diverse types, RAM, and storage unit, are allocated to the multiple users, depending on their needs for the efficient processing of the software applications. To avoid any collision during the multi-user requests, the cloud provides the resources with different traffic conditions, using the competence optimization procedure.

The competence optimization computes three important measures, namely the bandwidth, processor speed, and latency, to avoid traffic occurrences throughout requests from multiple users. The main objective of competence optimization in the IST-MRC mechanism is not only to measure the availability of the resources but also to evaluate the required Request Ratio. In given that, it allocates the resources to all of the users without any delay. The competence optimization (CRO) of the resources, using the IST-MRC mechanism is computed as:

$$CRO = \{U_1(P,\ RAM,\ Storage\ space),\ U_2\ (P,\ RAM,\ Storage\ space),\ U_3(P,\ RAM)\} \qquad (4)$$

From Equation (4), the cloud allocates the resources following the request made by the Users 1, 2, and 3 for the efficient processing of the software application services. The *'P'* in Equation (4) denotes the processor requests; 'RAM' indicates RAM space needed for the software installation, and *'Storage Space'* is intended for storing the processed results. By reducing the complexity level, every one of the requests is undoubtedly fulfilled without any collision in the IST-MRC mechanism. Besides, the traffic under diverse conditions is handled by computing the speed of the processor and the bandwidth rate, on which the cloud server places the request. By avoiding collisions, efficient Traffic Handling (TH) is obtained using (5).

$$TH = [2*(Pspeed+RAMspeed)]+B/L \tag{5}$$

The optimized resources achieve efficient traffic handling and it is allocated to the multiple users, based on the processor speed, *'Pspeed'* and RAM speed, *'RAMspeed'*. Further, the IST-MRC mechanism sums up the ratio of Bandwidth Rate, *'B'* to the Delay Time, *'L'* to easily tackle all the requests under different traffic conditions. The competence optimization uses the cloud resources to run the software applications in support of the multiple user systems, and it, therefore, attains a higher reliability rate.

5. RESULT AND DISCUSSION

5. 1. Experimental Setup and Analysis

The proposed mechanism, the IST-MRC, has been implemented in JAVA with Cloudsim Simulator, and certain statistical results have been obtained to validate it. A particular toolkit is chosen as a simulation platform for easy evaluation of the experimental parameters. Cloudsim holds the cloud structure with its multiple software tools. The machine is simulated with the data center, which comprises a variety of processing speeds, CPU, and RAM. Virtual machines are used for the experimental work in Cloudsim. Cloud computing has emerged as the key technology to deliver consistent, secure, and scalable computational services. The parameters used to obtain the results for simulations are listed in Table 1. TheIST-MRC mechanism is compared against the existing dynamic request redirection (DRRCCMN) [1] in the cloud-centric media network and cloud scheduler using OpenStack (CSOpenStack) [2] prototype. The experiments are conducted with 100 users and 300 requests, as well as with an average BTE of 30. The experiments are used to measure and evaluate factors, such as Successful Request Handles, Resource Utilization Efficiency, Latency Time, and Trust Success Ratio on the multiple users.

TABLE 1. Parameters Setting of CloudSim

Entity Type	Parameters	Value
Tasks (cloudlet)	Length of Tasks	1000-20000
	Total number of Tasks	100-1000
	Priority of the Tasks	High, Medium and Low
Virtual Machine	Total number of VMs	50
	MIPS	500-2000
	VM Memory (RAM)	256-2048
	Bandwidth	500-1000
	Cloudlet Scheduler	Space shared and Time shared
	Number of PEs requirement	1-4
Data Center	Number of Data Center	10
	Number of Host	2-6
	VM Scheduler	Space shared and Time shared

5. 2. Performance Evaluation

a. Scenario 1: Successful Request Handles on Multiple Users

To better perceive the effectiveness of the proposed IST-MRC mechanism, substantial experimental results are tabulated in Table 2. The IST-MRC mechanisms compared against the existing dynamic request redirection (DRR-CCMN) [1] in a cloud centric media network and cloud scheduler using OpenStack (CS-OpenStack) [2] prototype. For experimental purposes, Java with Cloudsim simulator is used to experiment with the factors and analyze the measures with the help of the graph values.

Results are accessible for the mixed number of requests, being placed. Request handles on the multiple users measure the ratio of the number of times the software is requested to the total number of requests being made as in (1). The higher, the requests being handled, the more successful the method is. The results disclosed confirm that the successful request handles on the multiple users increase in conjunction with the increase in the number of requests being placed.

TABLE 2. Tabulation for Request Handles on Multiple Users

No. of Requests Placed (n)	Request Handles on Multiple Users		
	IST-MRC Mechanism	DRR-CCMN	CSOpenStack
20	14	9	7
40	32	18	15
60	52	40	32
80	48	36	30
100	88	72	60
120	105	90	80
140	125	100	85

For experimental purposes, the process is repeated until the 120 requests are fulfilled. Figure 4 illustrates the application handles on multiple users, based on the number of the demands being placed. The IST-MRC mechanism performs relatively well when compared to the two additional methods, DRR-CCMN [1] and CS-OpenStack [2]. The proposed approach has superior modifications, using a trust-based estimator, which eventually examines Availability Rate, Hit Rate, and User Feedback Rate and thereby increases the request handles on the multiple users quickly by 14 – 43 % when compared to the DRR-CCMN [1]. Moreover, with the severance of the trustworthy and non-trustworthy software collection, dynamically, based on the Hit Ratio, the request handles an increase on multiple users by 23-53 % when compared to CS-OpenStack.

b. Scenario 2: Resource Utilization Efficiency on Multiple Users

To maximize the resource utilization efficiency on multiple users, the processor speed, bandwidth, and latency to handle the wide-ranging traffic conditions on the multiple user requests are deliberated. In the experimental setup, the number of requests positioned ranges from 20 to 140. As illustrated in Figure 5, the IST-MRC mechanism measures resource utilization efficiency, which is measured in terms of percentage (%). The resource utilization efficiency, using the IST-MRC mechanism, offers equivalent values with the state-of-the-art methods. Besides, the resultant resource utilization provides the summation of the processor speed, bandwidth, and latency, and it is obtained using the equation given below.

$$RU = \sum_{i=1}^{n} P_i\,, RAM, Storage_i \qquad (6)$$

The targeting results of resource utilization efficiency, using the IST-MRC mechanism, compared with the two state-of-the-art methods, [1] and [2], are presented in Figure 5 for visual comparison, based on the number of the requests placed. The proposed approach differs from DRR-CCMN [1] and CSOpenStack [2] since we have incorporated competence optimization method that competently allocates the resources to the multiple resources. It can be evaluated under different traffic conditions by improving resource utilization efficiency. This value is 7–14% when compared to DRR-CCMN. Besides, the diverse traffic conditions, based on the processing speed and RAM speed, further enable the resource utilization efficiency by 2–18 % when compared with the CS-OpenStack [2].

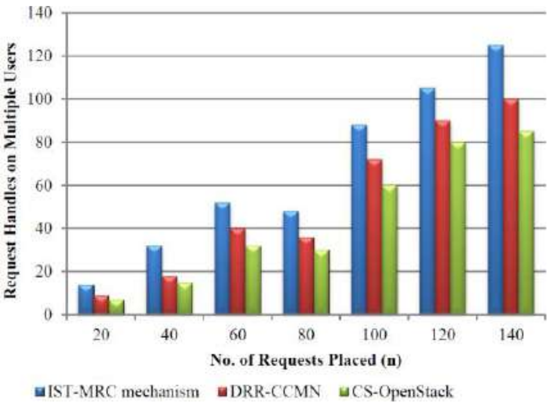


Figure 4. Distribution of request handles on multiple users in the experiments (n = 300 requests placed)

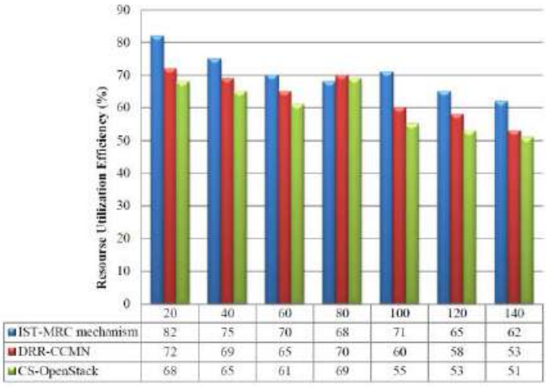


Figure 5. Distribution of resource utilization efficiency on multiple users in the experiments (n = 300 requests placed)

c. Scenario 3: Latency Time on Multiple Users

In Table 3, we show an analysis of Latency Time of IST-MRC mechanism concerning the behavior trust estimator value (BTE), ranging between 2.5 and 13.5, and measure the time taken to place a request between multiple users and the cloud server. It is calculated in terms of milliseconds (ms).

Figure 6 shows the Latency Time of the IST-MRC mechanism, DRR-CCMN [1], and CS-OpenStack [2] versus the increasing number of BTEs, from BTE = 2.5 to BTE = 13.5. The Latency Time improvement returned over DRR-CCMN, and CS-OpenStack decreases gradually as the number of BTE gets increased. For example, if BTE = 9, the growth of the IST-MRC mechanism, compared to DRR-CCMN, is 2.02 percent and when compared to CS-OpenStack, it is 4.92 percent. At the same time, if BTE = 12, the improvements are in the region of 1.64 and 3.56 percent, compared to DRR-CCMN and OpenStack, respectively. The trust value and resources are optimized separately in the IST-MRC mechanism. The proposed IST-MRC mechanism improves the Latency Time by 2 -7 % when compared to DRR-CCMN, and by 3–9 % when compared to CS-OpenStack.

d. Scenario 4: Latency Time on Trust Success Ratio

Figure 7 illustrates the Trust Success Ratio of the different users, User 1 to User 7, taken for experimental purposes. As shown in this figure, the percentage of Trust Success Ratio in the IST-MRC mechanism is higher than the other two methods. The reason is the highest number of requests being handled, using the proposed IST-MRC mechanism. In addition, the feedback application, obtained from the repository, improves the Trust Success Ratio by 3 – 12 % when compared to DRR-CCMN, and 7 - 20 % when compared to CS-OpenStack.

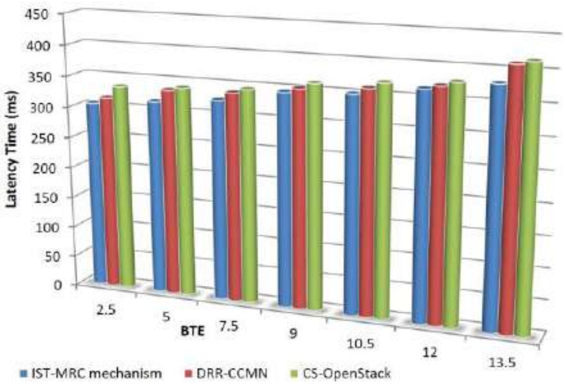


Figure 6. Distribution of Latency Time on multiple users in the experiments (BTE = 30 behavior trust estimators)

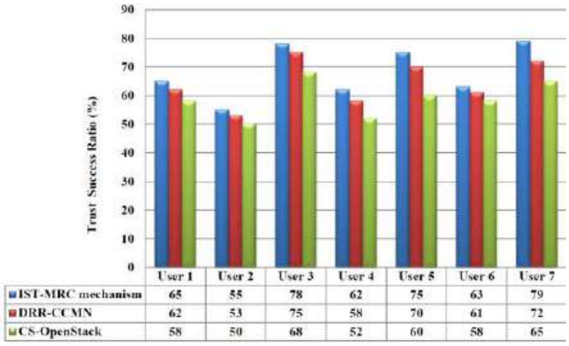


Figure 7. Distribution of Trust Success Ratio on multiple users in the experiments

TABLE 3. Tabulation of Latency Time

BTE	Latency Time (ms)		
	IST-MRC Mechanism	DRR-CCMN	CSOpenStack
2.5	305	318	335
5	315	325	340
7.5	325	338	345
9	345	349	352
10.5	350	353	360
12	365	370	375
13.5	380	395	410

6. CONCLUSION

In this paper, an integrated software trust estimator in alliance with the multi-user resource competence (IST-MRC) optimization mechanism was proposed in the cloud computing environment. The IST-MRC mechanism utilizes a behavior trust estimator to capably measure the trust value of the software service zone by applying Software Availability Rate, Hit Rate, and User Feedback according to the specific software apps. With the trust value computed for effective software provisioning to the multiple users, competence optimization is performed using processor speed, bandwidth, and latency to handle the varied traffic conditions on the multiple user requests. Finally, with the help of the feedback attained, the trustworthy and untrustworthy software collections are identified, and the trust value, based on the feedback, is provided. Experimental results reveal that the proposed IST-MRC mechanism not only leads to conspicuous improvement over Trust Success Ratio and Resource Utilization Efficiency but also outperforms the other methods - DRR-CCMN and CS-OpenStack- in Latency Time and successful requests handled on multiple users.

7. REFERENCES

1. Tang J., Peng Tay W., and Wen Y., "Dynamic request redirection and elastic service scaling in cloud-centric media networks." ***IEEE Transactions on Multimedia***, Vol. 16 (2014), 1434-1445. DOI: 10.1109/TMM.2014.2308726.

2. Abbadi I. M. and Ruan A., "Towards trustworthy resource scheduling in clouds." ***IEEE Transactions on Information Forensics and Security,*** Vol. 8, (2013), 973-984. DOI: 10.1109/TIFS.2013.2248726.

3. Wang Q., Wang C., Ren K., Lou W. and Li J., "Enabling public auditability and data dynamics for storage security in cloud computing." ***IEEE Transactions on Parallel and Distributed Systems,*** Vol. 22, (2010), 847-859. DOI: 10.1109/TPDS.2010.183.

4. Sundareswaran S., Squicciarini A., and Lin D., "Ensuring distributed accountability for data sharing in the cloud." ***IEEE Transactions on Dependable and Secure Computing,*** Vol. 9, (2012), 556-568, DOI: 10.1109/TDSC.2012.26.

5. Abbadi I, M. and Martin A., "Trust in the Cloud." Information security technical report, Vol. 16, (2011), 108-114.

6. Pandey S., Voorsluys W., Niu S., Khandoker A., and Buyya R., "An autonomic cloud environment for hosting ECG data analysis services." ***Future Generation Computer Systems,*** Vol. 28, (2012), 147-154, DOI: 10.1016/j.future.2011.04.022.

7. Takabi H., Joshi J. B., and Ahn G. "Security and privacy challenges in cloud computing environments." ***IEEE Security & Privacy,*** Vol. 8, (2010), 24-31, DOI: 10.1109/MSP.2010.186.

8. Wu X., Zhang R., Zeng B., and Zhou S., "A trust evaluation model for cloud computing." ***Procedia Computer Science*** Vol. 17 (2013), 1170-1177, DOI: 10.1016/j.procs.2013.05.149.

9. Sun D., Chang G., Sun L., and Wang X. "Surveying and analyzing security, privacy and trust issues in cloud computing environments." ***Procedia Engineering,*** Vol. 15, (2011), 2852-2856, DOI: 10.1016/j.proeng.2011.08.537.

10. Gao K., Wang Q., and Xi L. “Reduct algorithm based execution times prediction in knowledge discovery cloud computing environment.” ***The International Arab Journal of Information Technology,*** Vol. 11, (2014), 268-275, DOI: 10.5296/npa.v7i2.7797.

11. Whaiduzzaman M., Nazmul Haque M., Chowdhury M. R. K., and Gani A. “A study on strategic provisioning of cloud computing services.” ***The Scientific World Journal*** Vol. 14 (2014), DOI: 10.1155/2014/894362.

12. Wu Q., Zhang X., Zhang M., Lou Y., Zheng R., and Wei W. “Reputation revision method for selecting cloud services based on prior knowledge and a market mechanism.” ***The Scientific World Journal,*** Vol. 4, (2014), DOI: 10.1155/2014/617087.

13. Sun Y., Zhang J., Xiong Y., and Zhu G. “Data security and privacy in cloud computing.” ***International Journal of Distributed Sensor Networks,*** Vol. 10, (2014), 19-27, DOI: 10.1155/2014/190903.

14. Xiong A. and Xu C. “Energy efficient multi resource allocation of virtual machine based on PSO in cloud data center.” ***Mathematical Problems in Engineering,*** Vol. 2, (2014), DOI: 10.1155/2014/816518.

15. Ghazizadeh E., Zamani M., Ab Manan J., and Alizadeh M. “Trusted computing strengthens cloud authentication.” ***The Scientific World Journal,*** Vol. 4, (2014), DOI: 10.1155/2014/260187.

16. Varghese A. B., Hemalatha T, Sasidharan S., and Jophin S. “Trust Assessment Policy Manager in Cloud Computing-Cloud Service Provider's Perspective.” ***International Journal on Recent Trends in Engineering & Technology,*** Vol. 10, (2014), 46, DOI: 10.5772/intechopen.76338.

17. Arianyan E., Taheri H., and Khoshdel V. “Novel fuzzy multi objective DVFS aware consolidation heuristics for energy and SLA efficient resource management in cloud data centers**.”** ***Journal of Network and Computer Applications,*** Vol. 78, (2017), 43-61, DOI: 10.1016/j.jnca.2016.09.016.

18. Wang X., Chen X., Yuen C., Wu W., Zhang M., and Zhan C. “Delay-cost tradeoff for virtual machine migration in cloud data centers.” ***Journal of Network and Computer Applications,*** Vol. 78, (2017), 62-72, DOI: 10.1016/j.jnca.2016.11.003.

19. Kumar G. and Kumar Rai M. “An energy efficient and optimized load balanced localization method using CDS with one-hop neighbourhood and genetic algorithm in WSNs.” ***Journal of Network and Computer Applications,*** Vol. 78, (2017), 73-82, DOI: 10.1016/j.jnca.2016.11.013.

20. Lin Y., Lai Y., Teng H., Liao C., and Kao Y. “Scalable multicasting with multiple shared trees in software defined networking.” ***Journal of Network and Computer Applications,*** Vol. 78, (2017), 125-133, DOI: 10.1016/j.jnca.2016.11.014.

21. Chauhan S. S., Pilli E. S., Joshi R., Singh G., and Govil M. “Brokering in interconnected cloud computing environments: A survey.” ***Journal of Parallel and Distributed Computing,*** Vol. 133, (2019), 193-209, DOI: 10.1016/j.jpdc.2018.08.001.

22. Xie X., Tianwei Y., Xiao Z. and Xiaochun C. "Research on trust model in container-based cloud service." ***Computers, Materials and Continua,*** Vol. 56, No. 2, (2018), 273-283, DOI: 10.3970/cmc.2018.03587.

23. Asadi F., and Hamidi H. "An architecture for security and protection of big data." ***International Journal of Engineering,*** Vol. 30, No. 10, (2017), 1479-1486, DOI: 10.5829/ije.2017.30.10a.08.

24. Xu X., Liu X., Xu Z., Wang C., Wan S. and Yang X. "Joint Optimization of Resource Utilization and Load Balance with Privacy Preservation for Edge Services in 5G Networks." ***Mobile Networks and Applications*** (2019), 1-12, DOI: 10.1007/s11036-019-01448-8.

25. Speily, B., Omid R., and Rezai H. "Energy aware resource management of cloud data centers." ***International Journal of Engineering,*** Vol. 30, No. 11, (2017), 1730-1739, DOI: 10.5829/ije.2017.30.11b.14.

Persian Abstract

چکیده

اعتماد سازی یکی از موارد مهم برای افزایش قابلیت مقیاس پذیری و قابلیت اطمینان منابع در محیط ابری است. در این مقاله به منظور ایجاد یک مدل مطمئن جدید بر روی منابع ابری SaaS (نرم افزار به عنوان سرویس دهنده) ، بهینه سازی استفاده از منابع برای درخواست های چند کاربره وتخمین اطمینان نرم افزاری به صورت یکپارچه با مکانیزم بهینه سازی منابع چند کاربره (IST-MRC) پیشنهاد شده است. مکانیسم بهینه سازی IST-MRC، بدون هیچگونه ترافیکی در محیط ابری، ویژگی های قابل اعتماد خود را در صورت نیاز برنامه های نرم افزاری، با تخصیص منابع بهینه ترکیب می کند. در ابتدا، تخمین اطمینان رفتاری در مکانیسم IST-MRC برای ارزیابی اطمینان سرویسهای نرم افزاری ایجاد می‌شود. ارزیابی اعتماد بر اساس نرخ در دسترس بودن نرم افزار، میزان نرخ ضربه و بازخورد کاربر در مورد برنامه‌های نرم افزاری خاص صورت می‌پذیرد. در مرحله بعد، منابع با استفاده از بهینه‌سازی شایستگی برای چندین کاربر بهینه می‌شوند. بهینه سازی شایستگی در مکانیسم IST-MRC سرعت، پهنای باند و تأخیر پردازنده را محاسبه می‌کند تا بتواند درخواست‌های کاربرهای مختلف را درشرایط مختلف ترافیکی مدیریت کند. آزمایش‌های مختلفی برای اندازه‌گیری و ارزیابی پارامترهایی از قبیل مدیریت درخواست موفق، بهره وری در استفاده از منابع، مدت زمان تاخیر و نسبت موفق اعتماد در چند کاربر انجام شده است.

International Journal of Engineering

Journal Homepage: www.ije.ir

Learning Document Image Features With SqueezeNet Convolutional Neural Network

M. Hassanpour*, H. Malek

Department of Computer Science Engineering, Shahid Beheshti University, Tehran, Iran

PAPER INFO

Paper history:
Received 10 November 2017
Received in revised form 23 December 2017
Accepted 04 Januray 2018

Keywords:
Squeezenet
Convolutional Neural Network
Document Image Classification

ABSTRACT

The classification of various document image classes is considered an important step towards building a modern digital library or office automation system. Convolutional Neural Network (CNN) classifiers trained with backpropagation are considered to be the current state of the art model for this task. However, there are two major drawbacks for these classifiers: the huge computational power demand for training, and their very large number of weights. Previous successful attempts at learning document image features have been based on training very large CNNs. SqueezeNet is a CNN architecture that achieves accuracies comparable to other state of the art CNNs while containing up to 50 times less weights, but never before experimented on document image classification tasks. In this research we have taken a novel approach towards learning these document image features by training on a very small CNN network such as SqueezeNet. We show that an ImageNet pretrained SqueezeNet achieves an accuracy of approximately 75 percent over 10 classes on the Tobacco-3482 dataset, which is comparable to other state of the art CNN. We then visualize saliency maps of the gradient of our trained SqueezeNet's output to input, which shows that the network is able to learn meaningful features that are useful for document classification. Previous works in this field have made no emphasis on visualizing the learned document features. The importance of features such as the existence of handwritten text, document titles, text alignment and tabular structures in the extracted saliency maps, proves that the network does not overfit to redundant representations of the rather small Tobacco-3482 dataset, which contains only 3482 document images over 10 classes.

doi: *10.5829/ije.2020.33.07a.05*

1. INTRODUCTION

Archive offices usually contain a large corpus of various paper documents, and maintaining these documents can be a challenging issue. Although the trivial solution might be to convert these documents into digitally scanned document images and store them on disk with similar documents residing in the same folders on the file system, implementing this classification process through human labor can be extremely time consuming and frustrating. One good solution to this problem is to use of deep learning and Convolutional neural networks (CNN) for document image classification [1] since these networks have recently created a breakthrough in the field of image classification.

A CNN classifier trained with back propagation can be very powerful at learning rather complex visual concepts, but its often very large number of weights and depth means that it requires lots of computational power to converge on the learning task, and a considerable amount of memory for storing the weights. The large number of weights may not seem a problem when deploying on powerful machines with large memory, but can become a serious drawback when deploying on embedded systems that use much smaller memories. While one approach to this problem is to use techniques such as network pruning, quantization and huffman coding to reduce the model's size for inference [2], the strategy proposed by SqueezeNet is to reduce the number of convolution weights by squeezing convolutional feature maps using 1x1 filters while maintaining accuracy as high as possible by stacking 1x1 and 3x3 feature maps [3]. By doing so, the baseline SqueezeNet network reduces its size to less than 1 million weights while maintaining AlexNet [4] level accuracy on the ImageNet [5] dataset.

To justify the use of SqueezeNet for document image classification, we shall first make three important

*Corresponding Author Email: *mo.hassanpour@mail.sbu.ac.ir* (M. Hassanpour)

Please cite this article as: M. Hassanpour, H. Malek, Learning Document Image Features with SqueezeNet Convolutional Neural Network, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1201-1207

notes here. First, features extracted from document images are indeed robust to compression [6], therefore the SqueezeNet architecture may be applied to squeeze feature maps harmlessly. Second, features learned from ImageNet are transferrable to the document image domain and SqueezeNet performs well on the ImageNet dataset [6]. Third, state of the art CNN architectures experimented by Afzal et al. [7] (networks such as Resnet [8], GoogLeNet [9], AlexNet [4] and VGG-19 [10]), all achieve similar accuracy rates on the Tobacco-3482 dataset. Considering the previous three notes mentioned, along with the experience that SqueezeNet achieves AlexNet level accuracy on ImageNet, reaches us to the hypothesis that it is possible for SqueezeNet to also achieve state of the art level accuracy on document image classification, while using much less weights. We will be further experimenting this hypothesis for the first time in this work.

Document image datasets share important structural similarities with generic image datasets such as ImageNet, while also having fundamental differences. The similarity between domains can easily be proven by showing that pretraining a CNN on ImageNet and then training on a document image classification dataset results in higher accuracy rates compared to random initialization of the weights [6, 7]. One important result of the differences between the two image domains is that augmentation techniques are not applicable to document images. While techniques such as image translation, rotation and scaling are successfully used to expand a dataset and reduce chances of overfitting, they cannot be applied to document images as these translations often disturb original document image features. We experimented this by training a Spatial Transformer Network (STN) which tries to learn the optimal linear augmentation on input images [11], but the final accuracy of the system degraded significantly.

The outline of the paper is as follows: in the second section we investigate related works done on the subject of document image classification. In the third section we first discuss a number of important strategies of the SqueezeNet architecture, and then describe the complete training procedure along with the choice of hyperparameters. In the fourth section, evaluation of the models classification accuracy is reported and saliency maps for the network input are analyzed. In the last section, the paper is concluded and possible future strategies accuracy are noted to improve.

2. RELATED WORKS

Over the years, various improvements and optimizations have been made to improve accuracy rates on the document image classification problem, most of which rely on CNN feature extractors. Here we will discuss a number of these efforts.

Kumar et al. [12] used a codebook of Speeded Up Robust Features (SURF) descriptors along with a random forest classifier with Support Vector Machine (SVM) to classify document images. This approach achieved a final accuracy of %43.27 on Tobacco-3482 [12], a dataset which was also first introduced in this paper and later converted into being one of the de-facto standards for benchmarking document image classification models.

Kang et al. [13] introduced one of the first attempts to train a document image classifier with CNN. They implemented a CNN with two convolution layers, max pooling and fully connected layers each, used the ReLU (Rectified Linear Unit) activation function along with dropout [14] to enhance the training process and then trained their network on two separate datasets: Tobacco-3482 and NIST tax-form. This method achieved better accuracy rates compared to previous state of the art approaches such as the hidden tree Markov model [15] and random forest classifier with SURF descriptors [12]. The main shortcoming of their method was that the designed CNN was too simple and therefore had a limited learning capacity. This problem was later overcome by proposing much deeper and complex CNNs such as Alexnet [4].

In [6], Harley et al. used an ensemble of five AlexNet networks with Principal Component Analysis (PCA) to present a new state of the art for document image classification. Their work made three main contributions to the field. First they showed that features extracted from document images were robust to compression. Second they showed that using an ensemble of networks does not greatly improve the classification results, and therefore it is unnecessary to enforce region specific feature learning for the task, under the consumption that enough training data is provided. Third, they showed that features extracted from other image classification tasks can be well transferred to the document image classification task. Pretraining a network on ImageNet and then training on the document image classification task improved the final accuracy rate. They also introduced the large scale RVL-CDIP dataset which contains 400 thousand document images over 16 classes.

In 2017, Afzal et al. performed an exhaustive investigation on various CNN architectures for document image classification [7]. The results showed that using architectures more complex than AlexNet for the task does not result in a noticeable increase of accuracy rate on the Tobacco-3482 and RVL-CDIP datasets. They also reduced the error rate on the Tobacco-3482 dataset by more than half by pretraining on the very large RVL-CDIP dataset. A drawback of their methods was that the trained networks still contained too many weights.

Tensmeyer and Martinez applied an unsupervised clustering based approach to cluster visually similar noise images [16]. They employed a 3-stage scalable clustering approach which first clusters a subset of the data, then these clusters are further split to create purer subclusters, and at last a classifier is trained on top to recreate the subclusters. Their method showed promising results on five various document datasets.

3. METHODOLOGY

This section will go through the architectural details of SqueezeNet, along with the detailed procedure of training this network on the Tobacco-3482 dataset. The relatively small number of weights in SqueezeNet simplifies tasks such as implementing it on embedded systems and downloading its weights over the Internet. It also considerably speeds up the training process.

3. 1. SqueezeNet Architecture

This architecture was proposed by Iandula et al. in 2016 [3]. The SqueezeNet network itself consists of building blocks named Fire modules, as shown in Figure 1. Each Fire module is basically a Squeeze Layer followed by an Expand layer, where the Squeeze layer is simply a layer of 1x1 convolution maps, and the expand layer is a combination of 1x1 and 3x3 maps. The number of feature maps in the Squeeze layer is made less than or equal to the number of expand layer feature maps, therefore performing some kind of compression on the extracted feature maps while also reducing the number of network weights. These Fire modules are then eventually stacked together to build the microarchitecture of the SqueezeNet model, as can be seen in Figure 2. An important hyperparameter of the Fire module is the Squeeze ratio, the number of Squeeze layer feature maps divided by the number of expand layer feature maps. Increasing this ratio up to $\frac{1}{2}$ generally increases the networks accuracy rates at the cost of increasing the network's size.

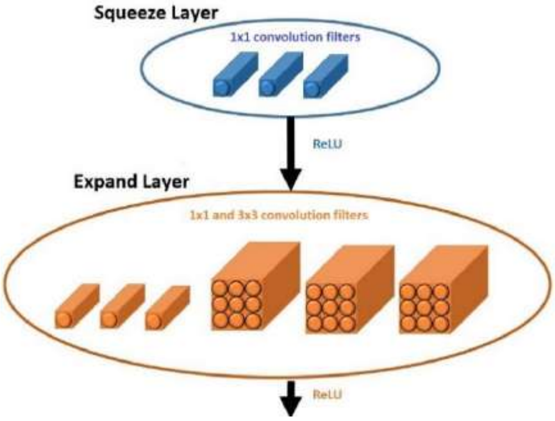


Figure 1. Overall structure of a Fire module in SqueezeNet model [3]

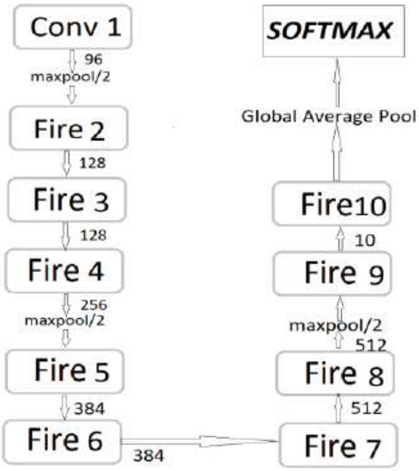


Figure 2. The macro architecture of a baseline SqueezeNet model [3]

SqueezeNet takes on three main strategies to improve the performance of traditional CNN networks. First, the majority of filters used in the network are 1x1 instead of 3x3; this greatly reduces the number of network weights. Second, it decreases the number of input channels to 3x3 filters. This approach also greatly reduces the number of network weights. Third, downsampling is performed later in the network on larger activation maps. The idea proposed is that a direct relationship exists between the size of activation maps of which downsampling is performed upon, and final classification accuracy results.

Another important strategy which greatly reduces the number of network weights, is the removal of the fully connected dense layers often used at the end of the network. This layer is replaced with a convolutional layer in which the number of output channels is equal to the number of data classes, and followed by a dropout layer and softmax activation function.

3. 2. SqueezeNet for Document Image Classification

According to classification results reported by [7], most deep CNN architectures achieve similar scores on the document image classification task both on small and large scale document image datasets separately. This gives us an intuitive understanding as of how much network complexity is correlated with classification accuracy on document image classification. It seems that raising the network's complexity higher than AlexNet level, does not have a significant effect on final classification accuracy.

As Harley et al. [6] showed, features extracted from document images are robust to compression. Therefore it is possible to effectively train SqueezeNet on this task, a network that continuously uses feature compression on every Squeeze layer to reduce the overall network size, while maintaining accuracy. This automatically makes SqueezeNet the superior choice for document image classification, as it achieves accuracy rates comparable to AlexNet while using as much as 50 times less weight [3].

3. 3. Training Procedure To evaluate the performance of SqueezeNet on document image classification, we trained this model on the Tobacco-3482 dataset which contains 3482 high resolution document images over 10 classes. A number of sample documents from this dataset can be seen in Figure 3. In each class, 80 images were used for training, 20 for validation and the rest for testing model performance.

All of the dataset's grayscale images were repeated over three channels, resized to 224x224 and mean subtracted. After performing experiments, we found a minibatch size of 64 and learning rate of 10^{-4} to be the best option for training. The optimizer we used for learning network weights was Adam [17], with the hyperparameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. This optimizer has proven to perform well when training CNN networks. All network weights were pretrained on ImageNet, and training was performed over 150 epochs where each training epoch took approximately 6 seconds on a Tesla K80 GPU. Due to the small dataset size, training was performed five times with five different train/validate/test splits and the accuracies achieved from these splits were averaged to get the final accuracy.

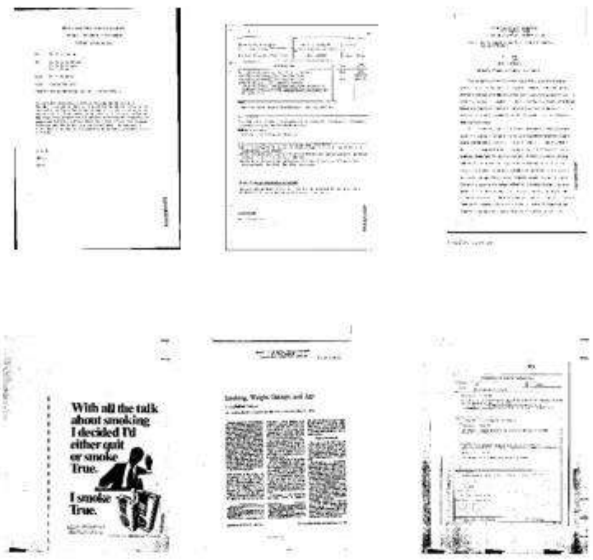


Figure 3. Sample documents from Tobacco-3482. The sample classes from top left to bottom right are memo, resume, note, advertisement, scientific and form

4. Evaluation

4. 1. Network Accuracy Our results show that the SqueezeNet model performs well compared to other CNN models, and achieves an accuracy of %74.5 as can be seen in Table 1, which is only 1 percent less than the accuracy achieved by AlexNet. It is worth mentioning that the original SqueezeNet paper by Iandola et al. [3] introduced two strategies for further improving SqueezeNet classification rates at the cost of adding more network weights. The first approach is to increase the network squeeze ratio from $\frac{1}{4}$ up to $\frac{1}{2}$. This will make the network perform less compression on feature maps which in turn results in less data loss due to compression and higher accuracy rates on the data.

It is very likely that this little tweak will also boost accuracy rates for the document image classification task (we have not experimented this due to the unavailability of weights pretrained with ImageNet and the hardware limitations we had for training on ImageNet). The second approach is to add skip connections to each Fire module for increasing learning capacity. Due to the small size of the Tobacco-3482 dataset, these connections will likely harm accuracy results, as it was also shown by Afzal et al. that the classification rate achieved by a Resnet-50 network on Tobacco-3482 with no document pretraining, stands far behind other CNN models that do not contain these connections [7]. The only explanation for this phenomenon is that networks containing skip connections require larger amounts of training data to converge on a supervised image classification problem.

4. 2. Saliency Map Visualization To show the effectiveness of SqueezeNet at learning document image features, we used saliency maps to visualize gradients of the network's output layer with respect to its input, as proposed posed by Simonyan et al. [18]. Simply put, we are trying to compute the gradient $\frac{\boldsymbol{\delta}\ \mathbf{output}}{\boldsymbol{\delta}\ \mathbf{input}}$, where the output is the network's softmax layer and the input is the input image we feed to the network.

TABLE 1. Comparison of classification results on Tobacco3482 with ImageNet pretraining between SqueezeNet using a Squeeze ratio of ¼ as experimented by us and other CNN architectures as experimented by Afzal et al. [7]

Network	Accuracy (%)	Num. Parameters
Resnet-50	67.93	25.6 M
GoogLeNet	72.98	4 M
SqueezeNet	74.40	0.8 M
AlexNet	75.73	62.3
VGG-16	77.52	138

Visualizing this gradient for each input image results in a saliency map which shows how the softmax output changes with respect to changes in the input. Brighter regions in this saliency map for each image indicate parts of the input that create higher activations on the output softmax neurons. This will let us know which features the network is paying more attention to, and whether it is learning a meaningful substructure or simply just overfitting on outlier features specific to the training data. This can become an issue especially when the dataset being trained on is small in size.

A number of the resulting saliency maps can be seen in Figure 4. Our visualization shows that the network is paying attention to a number of important features in documents, which shall be mentioned below.

- Document headers such as titles,
- The alignment of text paragraphs. Different document classes use different alignment methods for text paragraphs,
- Tables in documents. Particular document classes such as forms can be classified from other classes using this particular feature,
- Handwriting on the document is also a crucial feature. Notes and Letters containing handwritten signatures could be classified from other classes through this feature.

The visualized maps can also help us understand which document features are more important with respect to each document class.

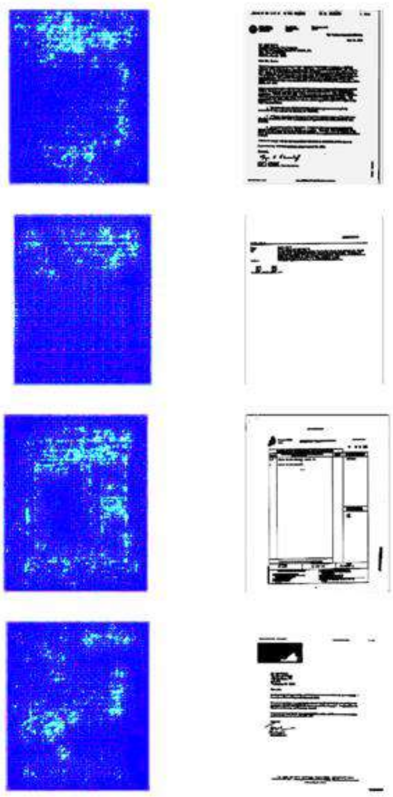


Figure 4. Four samples of extracted saliency maps (left column) and sample documents from the associated class (right column). The classes from top to bottom are Letter, Email, Form and Letter

5. DISCUSSION

Most attempts in this field have been focused on the use of Convolutional Neural Networks for document image classification. Although most of these networks are very large and expensive to train, the SqueezeNet CNN is able to achieve state of the art level accuracy in document image classification with only 800 thousand weights, and the relatively small size of this network makes it suitable for deployment on cheaper embedded devices. Although, one drawback of CNN networks in document image classification is that they are not able to exploit the sequential structure of a document image and the correlation between its elements (these are in fact important features in this context because a document image is somewhat sequential in nature [20], i.e the header, body and footer of a document image are very likely related to each other). The inability mentioned above is because convolutional architectures are not able to encode the position of features, and feature maps (even different regions of a single feature map) are computed independently so correlations cannot be exploited. Due to these shortcomings, future work in this field could possibly involve using recurrent architectures to exploit these attributes. In addition, image enhancement and binarization techniques can be used to enhance document images for a better classification result [19, 20].

A more recent learning framework such as Contrastive Predictive Coding (CPC) [21] may also be employed in the future to learn document image representations in an unsupervised manner which requires much less labelled data compared to supervised methods. CPC learns representations by predicting the future in latent space using autoregresstive models. A probabilistic contrastive loss is used to induce this latent space, and negative sampling makes the model's training procedure tractable. The advantages of this method compared to CNN is that its future prediction in latent space could be able to exploit the correlation between various parts of a document image, and the accuracy achieved by this method on ImageNet is comparable to fully supervised methods, despite using 2 to 5 times less training labels. Still, implementing this method on small embedded hardware remains a challenge, while this is not the case for SqueezeNet CNN.

6. CONCLUSION

In this work we studied previous works done on document image classification, then proposed that SqueezeNet is a suitable CNN architecture for this task. This architecture was then trained on the Tobacco-3482 dataset. The accuracy achieved by a baseline SqueezeNet with only 800 thousand weights, was comparable to other state of the art CNN architectures with weights in the order of tens of millions. We then visualized our network's saliency maps and investigated document features which were learned by the network.

7. REFERENCES

1. Vincent, N. and Ogier, J.-M., "Shall deep learning be the mandatory future of document analysis problems?", ***Pattern Recognition***, Vol. 86, (2019), 281-289. https://doi.org/10.1016/j.patcog.2018.09.010
2. Han, S., Mao, H. and Dally, W.J., "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding", arXiv preprint arXiv:1510.00149, (2015).
3. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J. and Keutzer, K., "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size", arXiv preprint arXiv:1602.07360, (2016).
4. Krizhevsky, A., Sutskever, I. and Hinton, G.E., "Imagenet classification with deep convolutional neural networks", in Advances in neural information processing systems., 1097-1105.
5. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L., "Imagenet: A large-scale hierarchical image database", in 2009 IEEE conference on computer vision and pattern recognition, Ieee., 248-255. DOI: 10.1109/CVPR.2009.5206848
6. Harley, A.W., Ufkes, A. and Derpanis, K.G., "Evaluation of deep convolutional nets for document image classification and retrieval", in 2015 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE. , 991-995. DOI: 10.1109/ICDAR.2015.7333910
7. Afzal, M.Z., Kölsch, A., Ahmed, S. and Liwicki, M., "Cutting the error by half: Investigation of very deep cnn and advanced training strategies for document image classification", in 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), IEEE. Vol. 1, 883-888. DOI: 10.1109/ICDAR.2017.149
8. He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition", in Proceedings of the IEEE conference on computer vision and pattern recognition., 770-778.
9. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., "Going deeper with convolutions", in Proceedings of the IEEE conference on computer vision and pattern recognition., 1-9.
10. Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, (2014).
11. Jaderberg, M., Simonyan, K. and Zisserman, A., "Spatial transformer networks", in Advances in neural information processing systems., 2017-2025.
12. Kumar, J., Ye, P. and Doermann, D., "Structural similarity for document image classification and retrieval", ***Pattern Recognition Letters***, Vol. 43, No., (2014), 119-126. https://doi.org/10.1016/j.patrec.2013.10.030
13. Kang, L., Kumar, J., Ye, P., Li, Y. and Doermann, D., "Convolutional neural networks for document image classification", in 2014 22nd International Conference on Pattern Recognition, IEEE., 3168-3172. DOI: 10.1109/ICPR.2014.546
14. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., "Dropout: A simple way to prevent neural networks from overfitting", ***The Journal of Machine Learning Research***, Vol. 15, No. 1, (2014), 1929-1958. DOI: 10.5555/2627435.2670313
15. Diligenti, M., Frasconi, P. and Gori, M., "Hidden tree markov models for document image classification", ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, Vol. 25, No. 4, (2003), 519-523. DOI: 10.1109/TPAMI.2003.1190578
16. Tensmeyer, C. and Martinez, T., "Confirm–clustering of noisy form images using robust matching", ***Pattern Recognition***, Vol. 87, (2019), 1-16. https://doi.org/10.1016/j.patcog.2018.10.004
17. Kingma, D.P. and Ba, J., "Adam: A method for stochastic optimization", arXiv preprint arXiv:1412.6980, (2014).
18. Simonyan, K., Vedaldi, A. and Zisserman, A., "Deep inside convolutional networks: Visualising image classification models and saliency maps", arXiv preprint arXiv:1312.6034, (2013).
19. He, S. and Schomaker, L., "Deepotsu: Document enhancement and binarization using iterative deep learning", ***Pattern Recognition***, Vol. 91, (2019), 379-390. https://doi.org/10.1016/j.patcog.2019.01.025
20. Guo, J., He, C. and Wang, Y., "Fourth order indirect diffusion coupled with shock filter and source for text binarization", ***Signal Processing***, Vol. 171, (2020), 107478. https://doi.org/10.1016/j.sigpro.2020.107478
21. Oord, A.v.d., Li, Y. and Vinyals, O., "Representation learning with contrastive predictive coding", ***arXiv preprint*** arXiv:1807.03748, (2018).

Persian Abstract

چکیده

توانایی دسته‌بندی کردن اسناد اسکن شده از روی تصویر، قابلیتی است که می‌توان از آن در کتاب‌خانه‌های دیجیتال یا سیستم‌های اتوماسیون اداری به خوبی بهره برد. در همین راستا، شبکه‌های عصبی پیچشی آموزش داده شده با الگوریتم پس انتشار به عنوان روشی امروزی و قدرتمند برای دسته‌بندی تصاویر شناخته می‌شوند. امّا چنین شبکه‌هایی در حال حاضر دارای دو اشکال هستند: هزینه محاسباتی آموزش دادن آن‌ها بسیار سنگین است و معمولاً حافظه بسیار زیادی را به جهت داشتن تعداد بسیار زیادی پارامتر اشغال می‌کنند. با این وجود موفقیت‌های به دست آمده در مساله دسته‌بندی اسناد اسکن شده عموماً از طریق آموزش دادن شبکه‌های پیچشی بسیار بزرگ حاصل شده‌اند. شبکه پیچشی عصبی SqueezeNet، شبکه‌ای به نسبت کوچک امّا قدرتمند است که قادر است با وجود داشتن تعداد پنجاه برابر پارامتر کمتر نسبت به شبکه‌ای قدرتمند مانند AlexNet، در مساله دسته‌بندی تصاویر ImageNet، دقّتی هم‌اندازه با آن را کسب نماید، امّا تاکنون عملکرد آن در مساله دسته‌بندی اسناد تصویری ارزیابی نشده است. به همین جهت ما در این تحقیق تصمیم گرفته‌ایم تا عملکرد SqueezeNet را در دسته‌بندی اسناد اسکن شده مورد بررسی قرار دهیم. ما نشان می‌دهیم که یک شبکه SqueezeNet پیش‌آموزش یافته از روی مجموعه داده ImageNet دقتی تقریباً معادل با ۷۵ درصد را بر روی مجموعه داده Tobacco–۳۴۸۲ متشکل از ۱۰کلاس به دست می‌آورد، که دقتی قابل قیاس با سایر شبکه‌ای عصبی پیچشی می‌باشد. سپس گرادیان خروجی شبکه نسبت به تصاویر ورودی را با استفاده از نقشه برجستگی مورد بررسی قرار می‌دهیم و نشان می‌دهیم که شبکه ویژگی‌های سودمند و معنی‌داری را از روی تصاویر آموزش دیده است. این در حالیست که در تحقیقات گذشته تلاشی در راستای ظاهرسازی و یا تفسیر ویژگی‌های آموزش دیده به وسیله شبکه به چشم نمی‌خورد. ما با تحلیل این نقشه‌های برجستگی نشان می‌دهیم که شبکه SqueezeNet با وجود آموزش دیدن بر روی مجموعه داده‌ای به نسبت کوچک، ویژگی‌هایی مانند امضا، عنوان، جدول و نوع خاص هم‌ترازی متن را به خوبی تشخیص می‌دهد و از آن‌ها در راستای دسته‌بندی و تفکیک اسناد بهره می‌برد.

IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1208-1213

International Journal of Engineering

Journal Homepage: www.ije.ir

An Ensemble Click Model for Web Document Ranking

D. Bidekani Bakhtiarvand*, S. Farzi

Department of Artificial Intelligence, Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran

PAPER INFO

Paper history:
Received 09 February 2020
Received in revised form 31 March 2020
Accepted 30 May 2020

Keywords:
Click Modeling
Document Re-ranking
Modeling Users' Behavior
Search Engine Result Page Enhancement

ABSTRACT

Annually, web search engine providers spend a lot of money on re-ranking documents in search engine result pages (SERP). Click models provide advantageous information for re-ranking documents in SERPs through modeling interactions among users and search engines. Here, three modules are employed to predict users' clicks on SERPs simultaneously, the first module tries to predict users' click behaviors using *Probabilistic Graphical Models*, the second module is a *Time-series Deep Neural Click Model* which predicts users' clicks on documents and finally, the third module is a similarity-based measure which creates a graph of document-query relations and uses *SimRank Algorithm* to predict the similarity. After running these three simultaneous processes, three click probability values are fed to an MLP classifier as inputs. The MLP classifier learns to decide on top of the three preceding modules, then it predicts a probability value which shows how probable a document is to be clicked by a user. The proposed system is evaluated on the Yandex dataset as a standard click log dataset. The results demonstrate the superiority of our model over the well-known click models in terms of perplexity.

doi: *10.5829/ije.2020.33.07a.06*

1. INTRODUCTION

Nowadays, web search engine providers invest huge amounts of money for ranking documents in SERPs according to users' satisfaction criteria [1]. Most users are willing to see search engine result pages (SERPs) which are arranged such that the more relevant documents appear in the higher ranks. Some of them never look at the second result page. Therefore, if the service provider does not want to see the users leaving the service, it should provide them with the most relevant results in the first page. Since sometimes a search engine is unable to arrange a SERP with desired ranking using the traditional methods, e.g. Page Rank, an innovative idea is to employ the collected knowledge from users' behaviors, e.g. how they interact with SERPs, when they stop clicking on documents, and amount of time they spend on documents. In other words, search engines can use the implicit feedback of users to improve the ranking of their documents. Typically, users click on the documents that they think are more in line with their information needs. Hence, a service provider can find relevant documents more efficiently before arranging a SERP by taking the users' footprints into account.

In recent years, researchers have been encouraged to work on models based on users' behavior such as clicks and mouse movements in order to enrich search engines' qualities [2–4] or effective advertising [5]. Since the click is the most frequent user behavior in web search, most researchers consider click models to provide some useful information for ranking documents. In online advertising markets, knowing about the click possibilities on items leads to changing the priority of displayed items.

In this paper, an Ensemble Click Model, hereafter named *ECM*, is introduced based on combining sophisticated well-known click models, i.e. a Probabilistic Graphical Model based (PGM) click model named User Browsing Model (UBM) [6] and a deep

*Corresponding Author Email: *danial.bidekani@email.kntu.ac.ir* (D. Bidekani Bakhtiarvand)

Please cite this article as: D. Bidekani Bakhtiarvand, S. Farzi, An Ensemble Click Model for Web Document Ranking, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1208-1213

neural network click model named NCM [2] and a structural similarity measure using SimRank [7]. The proposed system relies on the fact that each click model explores separate parts of the hypothesis space, with respect to its assumptions about users' click behaviors. Consequently, an Ensemble click model can scan a larger area of hypothesis space. Our proposed system has employed a Multi Layer Perceptron (MLP) classifier in order to predict the final decision based on UBM and NCM models outputs as Learning models, and SimRank as similarity measure.

In order to evaluate the ECM, a variety of experiments have been performed over a standard dataset called Yandex relevance dataset. The Yandex relevance dataset contains 30,717,251 unique queries and 117,093,258 documents. We followed two scenarios to evaluate the ECM in terms of effectiveness. The purpose of the first scenario is to analyze and set suitable parameters for the ECM and the second experiment attempts to compare the ECM with other well-known click models i.e. UBM, DBN (stands for Dynamic Bayesian Network) and NCM. The results imply that ECM has superiority over the well-known click models in terms of perplexity as quality metric.

The rest of the paper is organized as follows, Section 2 reviews all realted works, the proposed system is defined in Section 3. It is subdivided to four parts which are explained respectively. The experiments have been drawn and explained in Section 4 in which two research questions are discussed, and the conclusion is presented in Section 5. References are all listed in Section 6, respectively.

2. RELATED WORKS

PGM-based models have probabilistic backgrounds and try to model users' behaviors as a sequence of events. These events include attractiveness, examination, satisfaction, etc [8]. UBM and DBN are the state-of-the-art PGM-based click models [2]. The user browsing model is a well-known click model introduced by Dupret and Piwowarski [6]. The user browsing model considers the distance between last clicked document and current document examination to predict user's behavior. DBN is an extension of the Cascade Model [9] which considers a parameter for users satisfaction. The experiments show that the UBM outperforms DBN [3].

To the best of our knowledge, the first model which solved the problem of click modeling by the Neural Network approach was the Neural Click Model defined by Brisov et al. [2]. They defined several representations and also used LSTM and RNN models to train their proposed model. They reported better quality of their model over all PGM-based models and also defined another Neural Network based model using Encoder-Decoder architecture which was a little better than their previous model [10].

Researchers have applied the Convolutional Neural Network in the click modeling problem [11, 12]. There have been accomplished more specific researches on click modeling, specifically in mobile search [13] and sponsored search [14].

3. PROPOSED METHOD

We proposed an Ensemble Click Model which takes the advantages of both PGM-based and Neural network-based click models and a structural similarity-based algorithm called SimRank [7]. In particular, we paid attention to the idea of Ensemble Learning which deals to the concept that if a group of base learners attempts to learn the same problem, they can do a better classification by aggregating their viewpoints [15]. Therefore, base learners feed their decision output (a probability of how a document is likely to be clicked) to the combiner on one hand and SimRank predicts the similarity on the other hand, then the final decision will compute through either polling methods, e.g. majority voting, averaging and borda count, or applying a new classifier constructing an Ensemble Model. In contrast to most Ensemble Methods which use different datasets for every base learner, here, a single training and testing datasets have been employed. It provides us an outstanding achievement that the Ensemble Model may explore a larger subspace of the hypothesis space under the base learners' assumptions. The overall model architecture is depicted in Figure 1. The proposed model should pass four phases in order to train. According to Figure 1, in the undersampling phase, 50000 search sessions are selected from Yandex relevance dataset. In the Encoding phase, data is prepared in a way that it is suitable to base learners. Then in the base learners' training phase, each base learner will learn a hypothesis through its assumptions. Before executing last phase, the training set will be tested by each base learners. Then they return a probability of clicking on documents. These probabilities are fed to an MLP classifier as its attributes.

3. 1. Undersampling Phase

In 2011, Yandex published a dataset of its search engine which contained users' clicks history. Because of the computation limitation and unavailability of powerful resources, we had to undersample a set of sessions from Yandex relevance dataset by uniform random sampling. Yandex dataset contains ten documents for each query, however we consider the first six documents in this paper. Table 1 shows the general information of the undersampled dataset. Thus, in the first place, it needs to be shown that the undersampled dataset is a good representative of the whole Yandex dataset. In this regard, Table 2 represents

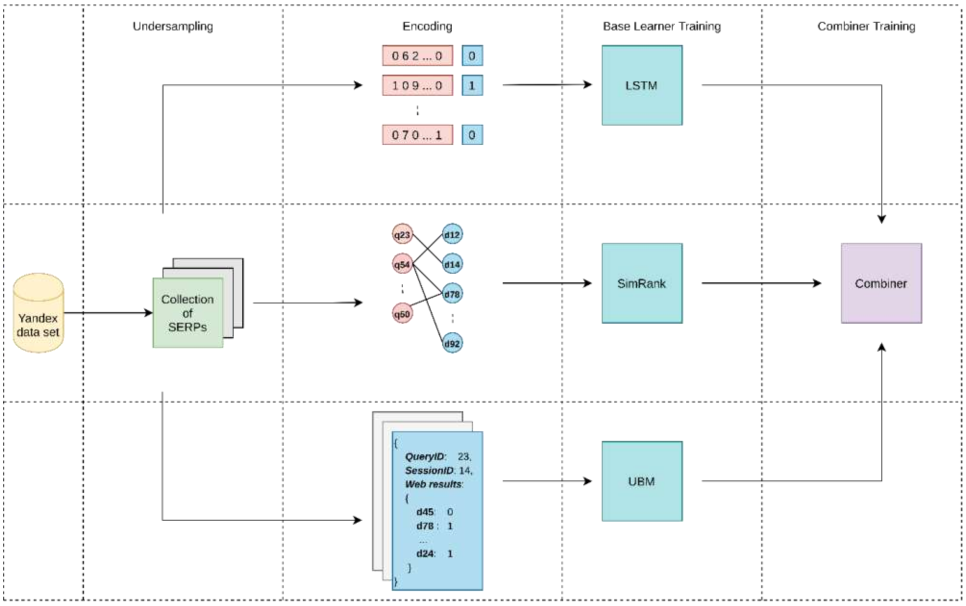


Figure 1. Ensemble click model training architecture

the click frequency ordered by documents ranking. It shows the normal behavior of users intuitively, because of the numerous clicks at the top ranks is the most seen users' behavior and the number of clicks has a descending order which illustrates that the probability of clicking on a document reduces by increasing its rank [16]. In other words undersampled dataset shows users' intention to click on the top documents in SERP which is the normal

behavior of search engine users. Figure 2 shows the users' intent on the sampled dataset. Furthermore, Table 3 shows the number of clicks in each query session of undersampled dataset. From the table, it can be understood that the session which contains less clicks on the results pages are in the majority.

3. 2. Encoding Phase Before each click model learns its hypothesis space, it is required to preprocess dataset in a suitable form according to the intended model. As observed in Figure 1 the encoding phase contains NCM transformation, UBM transformation and SimRank transformation.

Before each click model learns its hypothesis space, it is required to preprocess dataset in a suitable form.

The first representation should be in a vector form to feed in the NCM model. In order to transform dataset to the vector forms, a vector of 2^{SERP} is considered for every <query, document, rank> triplet. Here, the given SERP size is six, so every <query, document, rank> triplet creates a vector of 2^6, because in a six document SERP, there exist 2^6 different click patterns. A click pattern is a binary vector that its value shows the click or skip on the documents. To transform dataset to the vector forms, a vector of 2^6 is created for <query, document, rank> triplet and the search is began for SERPs which include the same <query, document, rank> triplet. Every SERP has a click pattern, it turns the click pattern binary value to an integer, and at the end it adds one unit to the corrosponding vector index. To make a more informative vector, the user interaction will be added at the end of the vectors, e.g. if the user clicks on the previous document, 1 will be appended to the vector representation otherwise 0. To create query

TABLE 1. Sampled dataset information

Item	Value
Search session size	50000
Query session size	166149
Train set query sessions size	124611
Test set query sessions size	41538
Unique query of train set query sessions size	62702
Unique query of test set query sessions size	22383

TABLE 2. Clicks and skips frequency in dataset

Rank	#Clicks	#Skips
1	76693	89456
2	32434	133715
3	21345	144804
4	16104	150045
5	12582	153567
6	10228	155921
Sum	169386	827508

representation, it is only needed to aggregate all vectors which have the same query [2].

This representation of UBM model includes query ID, session ID and the user history on each SERP whether it is a click or a skip.

In order to apply SimRank algorithm as a similarity measure of the ECM, data should be transformed into a graph form. This transformation creates a bipartite graph which contains two different node types; query node and document node. An edge is generated whenever at least there exist a click on a query and a document. The represented bipartite graph is weighted by the Click Through Rate (CTR). CTR for a query document pair is defined as the number of clicks when *q* as query and *d* as document appear together and receive a click event over the times that *q* and *d* appears together despite the event type (click or skip).

3. 3. Base Learner Training Phase

3. 3. 1. NCM Model

One of the base learners has the LSTM structure. It is shown that LSTM is an effective model to learn the sequences [17]. Since the click modeling is a sequential problem, it is a good intuition to learn a model based on the LSTM structure [2] called *NCM*. It considers query vector at first, then it predicts whether the user will click on the first document or not. Following the user's interactions with the first document, the model predicts on the second document, the considering query representation at first, and first document representation as next. This process continues until the last document in SERP.

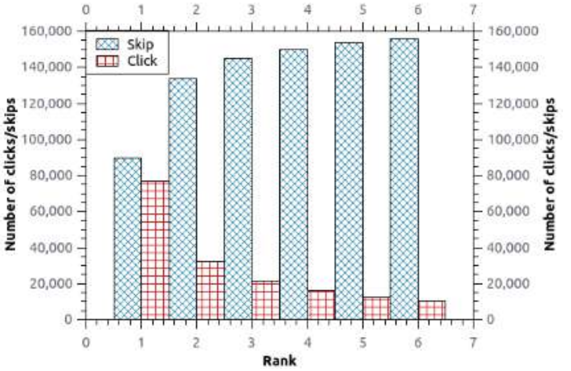


Figure 2. Click and Skip frequency @Rank

TABLE 3. Clicks per session frequency in the dataset

#Session	#Clicks
0	55498
1	77058
2	18804
3	8152
4	3883
5	1792
6	962
Sum	166149

3. 3. 2. UBM Model

The second learner is the UBM model which is a well-known PGM-based click model. UBM is an extension of the PBM model that considers the last clicked document in its assumption. It depends on the current document and the last clicked document ranks. This model defines two parameters, *examination parameter* which depends on the document rank and on the previously clicked document rank, and *Attractiveness parameter* which depends on a query and a document. Both examination and attractiveness parameters are learned by the Expectation Maximization (EM) algorithm [6].

3. 3. 3. Similarity-based Model

The third part of the ECM is different from the others. The similarity-based model tries to find the similarity between a query node and a document. It considers the SimRank algorithm as a similarity measure to predict a link between a query and a document. The idea behind SimRank is that the two objects are similar if they are referenced by similar objects. SimRank formula is as follows:

$$sim(q,d) = \frac{c}{|N(q)|.|N(d)|} \sum_{q \in N(q)} \sum_{d \in N(d)} sim(q,d) \qquad (1)$$

where *sim(q, d)* denotes the similarity between a document and a query which is initialized by the CTR measure. An object has the most similarity to itself, therefore *sim(q, q)* = 1 and *sim(d, d)* = 1.

3. 4. Combiner Training Phase

After the third phase, the base learner training phase, it is necessary to design a model which takes the base learners' output and learns to decide based on the base learners' opinions. MLP is a supervised learning algorithm which can learn a hypothesis spase based on gradient descent algorithm. As shown in Figure 1, MLP acts as a combiner learning in the presented model to aggregate the results of UBM, NCM and SimRank. The combiner takes the output from base learners as its input. Base learners output are similar in the context of domain. The domain is limited to a value in [0, 1]. After training the combiner model, when the model is required to predict on a <query, documen, rank> triplet, first the model transforms the input in the three different representations type as we discussed in Section 3.2, then each representation is fed to base learners as input, and three outputs comes from the base learners training phase. Next, these outputs are fed to the MLP classifier as input and in the end of the process, combiner decides on the base learners decision and predicts that the user finally clicks on the document or not.

4. EXPERIMENTS

In this section, we have provided two experiments to evaluate the model quality. Then we discuss the validity of the results and the superiority of ECM over the well-known click models.

In order to measure the quality of each click model, we used perplexity metric [6] to compare the accuracy of each click models. Perplexity is the best known metric in the context of click modeling. The perplexity measure for the model M on a set of sessions S in the rank r is calculated by Formula (2). The total perplexity is calculated by averaging over all ranks [2].

$$p_r(Model) = 2^{-\frac{1}{|S|}\sum_{s\in S}(c_r^{(s)}\log_2 q_r^{(s)}+(1-c_r^{(s)})\log_2(1-q_r^{(s)}))} \qquad (2)$$

Before addressing the results, it is a good idea to have an overall view to see how experiments are done. Because one of the three base learners is based on a Deep Neural Network, the results will be somewhat different for each execution. To make certain that the results make sense, we executed the NCM model over the training dataset 10 times. The results of 10 runs are included in Table 4. Because the execution of NCM model is required to execute the combiner model, we executed the combiner model 10 times as the same way to be sure of the results. The results for the combiner model executions are shown in Table 5.

4. 1. Experiment 1

After running the MLP classifier as the combiner model with different neural network structures, we came up with a two-layer neural network containing 20×5 neurons. The results for the combiner are depicted in Table 5. As it is shown in the table, the standard deviation is not noticeable which conforms the validity of the results.

TABLE 4. Results of 10 times execution of NCM

Run	@1	@2	@3	@4	@5	@6
1	1.789	1.519	1.423	1.384	1.289	1.249
2	1.811	1.545	1.462	1.335	1.299	1.242
3	1.800	1.539	1.390	1.329	1.272	1.236
4	1.719	1.560	1.468	1.360	1.290	1.208
5	1.729	1.514	1.368	1.320	1.298	1.231
6	1.834	1.512	1.405	1.317	1.286	1.237
7	1.688	1.537	1.426	1.329	1.298	1.231
8	1.703	1.510	1.384	1.355	1.298	1.245
9	1.703	1.542	1.448	1.323	1.285	1.244
10	1.673	1.493	1.399	1.313	1.272	1.246
AVG	1.745	1.527	1.419	1.338	1.290	1.239
STD DEV	0.057	0.020	0.033	0.022	0.010	0.011

TABLE 5. Results of 10 times execution of the Combiner

Run	@1	@2	@3	@4	@5	@6
1	1.578	1.052	1.027	1.011	1.008	1.004
2	1.495	1.035	1.025	1.013	1.007	1.004
3	1.421	1.026	1.031	1.012	1.008	1.003
4	1.519	1.057	1.031	1.014	1.009	1.005
5	1.505	1.053	1.029	1.012	1.007	1.004
6	1.420	1.034	1.022	1.011	1.006	1.004
7	1.420	1.036	1.025	1.010	1.005	1.003
8	1.501	1.031	1.032	1.014	1.007	1.006
9	1.544	1.051	1.030	1.013	1.008	1.003
10	1.530	1.036	1.026	1.011	1.006	1.003
AVG	1.493	1.041	1.028	1.012	1.007	1.004
STD DEV	0.055	0.011	0.003	0.001	0.001	0.000

4. 2. Experiment 2

After finding the best structure for the MLP classifier as the combiner of the proposed model, the model perplexity was measured and the results are depicted in Figure 3.

In order to compare the ECM with previous models, it is a good idea to draw the average perplexity of NCM, ECM and UBM models altogether. As it can be understood from Figure 3, the ECM has succeeded in all ranks and average perplexity. To ensure that these results are meaningful, the t-test experiment has been taken and the ECM has superiority over UBM (p-vlaue = 0.00010249401468) and NCM (p-value = 3.01806922588773E-05).

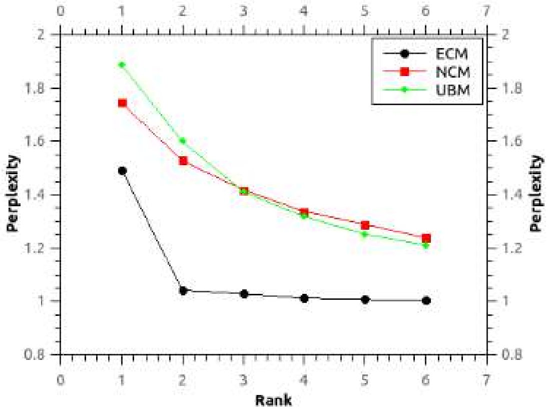


Figure 3. Comparison of the Click Models in terms of perplexity

5. CONCLUSION

In this paper, a new click model called ECM was introduced and it was shown that it has better performance than the well-known click models. The proposed model consists of a PGM-based click model called UBM and a Neural network based click model called NCM. An MLP classier is employed to decide based on UBM and NCM click models output. The superiority of ECM has been shown with experiments based on perplexity measure.

6. REFERENCES

1. Ghose, A., Ipeirotis, P. G., and Li, B., "Examining the impact of ranking on consumer behavior and search engine revenue", ***Management Science***, Vol. 60, No. 7, (2014), 1632–1654. https://doi.org/10.1287/mnsc.2013.1828
2. Borisov, A., Markov, I., De Rijke, M. and Serdyukov, P., "A neural click model for web search", In Proceedings of the 25th International Conference on World Wide Web, (2016), 531–541. https://doi.org/10.1145/2872427.2883033
3. Grotov, A., Chuklin, A., Markov, I., Stout, L., Xumara, F. and de Rijke, M., "A comparative study of click models for web search", In International Conference of the Cross-Language Evaluation Forum for European Languages, (2015), 78–90. https://doi.org/10.1007/978-3-319-24027-5_7
4. Liu, Z., Mao, J., Wang, C., Ai, Q., Liu, Y. and Nie, J.Y., "Enhancing click models with mouse movement information", ***Information Retrieval Journal***, Vol. 20, No. 1, (2017), 53–80. https://doi.org/10.1007/s10791-016-9292-4
5. Aouad, A., Feldman, J., Segev, D. and Zhang, D., "Click-based MNL: Algorithmic frameworks for modeling click data in assortment optimization", Available at SSRN 3340620, (2019). http://dx.doi.org/10.2139/ssrn.3340620
6. Dupret, G. E. and Piwowarski, B., "A user browsing model to predict search engine click data from past observations", In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, (2008), 331–338. https://doi.org/10.1145/1390334.1390392
7. Jeh, G. and Widom, J., "SimRank: a measure of structural-context similarity", In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, (2002), 538–543. https://doi.org/10.1145/775047.775126
8. Guo, F., Liu, C., and Wang, Y. M., "Efficient multiple-click models in web search", In Proceedings of the second acm international conference on web search and data mining, (2009), 124–131. https://doi.org/10.1145/1498759.1498818
9. Chapelle, O. and Zhang, Y., "A dynamic bayesian network click model for web search ranking", In Proceedings of the 18th international conference on World wide web, (2009), 1–10. https://doi.org/10.1145/1526709.1526711
10. Borisov, A., Wardenaar, M., Markov, I. and de Rijke, M., "A click sequence model for web search", In the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, (2018), 45–54. https://doi.org/10.1145/3209978.3210004
11. Liu, Q., Yu, F., Wu, S. and Wang, L., "A convolutional click prediction model", In Proceedings of the 24th ACM international on conference on information and knowledge management, (2015), 1743–1746. https://doi.org/10.1145/2806416.2806603
12. Ni, Z., Ma, X., Sun, X. and Bian, L., "A Click Prediction Model Based on Residual Unit with Inception Module", In Pacific Rim International Conference on Artificial Intelligence, (2019), 393–403. https://doi.org/10.1007/978-3-030-29911-8_30
13. Zheng, Y., Mao, J., Liu, Y., Luo, C., Zhang, M. and Ma, S., "Constructing Click Model for Mobile Search with Viewport Time", ***ACM Transactions on Information Systems (TOIS)***, Vol. 37, No. 4, (2019), 1–34. https://doi.org/10.1145/3360486
14. Zhang, Y., Dai, H., Xu, C., Feng, J., Wang, T., Bian, J., Wang, B. and Liu, T.Y., "Sequential click prediction for sponsored search with recurrent neural networks", In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, (2014). https://arxiv.org/abs/1404.5772
15. Dietterich, T. G., "Ensemble methods in machine learning", In International workshop on multiple classifier systems (pp. 1-15). Springer, Berlin, Heidelberg, (2000), 1–15. https://doi.org/10.1007/3-540-45014-9_1
16. Zhang, Y., Chen, W., Wang, D. and Yang, Q., "User-click modeling for understanding and predicting search-behavior", In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, (2011), 1388–1396. https://doi.org/10.1145/2020408.2020613
17. Sak, H., Senior, A. W., and Beaufays, F., "Long short-term memory recurrent neural network architectures for large scale acoustic modeling", In 15th Annual Conference of the International Speech Communication Association, Singapore, (2014), 338-342. https://www.isca-speech.org/archive/interspeech_2014/i14_0338.html

Persian Abstract

چکیده

به طور سالانه، شرکت‌های فراهم‌کننده‌ی سرویس موتور جست‌وجو، برای رتبه‌بندی مجدد اسناد در صفحه‌های جست‌وجو، مبالغ زیادی هزینه می‌کنند. کلیک‌مدل‌ها از طریق تعاملات کاربران با موتورهای جست‌وجو اطلاعات سودمندی برای رتبه‌بندی مجدد اسناد در صفحه‌های نتایج جست‌وجو فراهم می‌کنند. در این مقاله، به منظور پیش‌بینی کلیک کاربران بر روی صفحه‌های نتایج جست‌وجو، از سه ماژول به طور همزمان استفاده شده است، نخستین ماژول سعی در پیش‌بینی کلیک‌های کاربران، با استفاده از مدل‌های گرافی احتمالی دارد، ماژول دوم بر اساس شبکه‌های عصبی عمیق مختص سری‌های زمانی کلیک کاربران روی اسناد را پیش‌بینی می‌کند و در آخر، ماژول سوم که از یک معیار شباهت به نام SimRank بر روی گرافی از روابط کلیک–سند استفاده می‌کند. پس از اجرای همزمان این سه ماژول، سه مقدار احتمالاتی به دست آمده به عنوان ورودی‌های یک شبکه‌ی عصبی پرسپترون چندلایه استفاده می‌شود. شبکه‌ی عصبی پرسپترون چندلایه آموزش می‌بیند که روی تصمیم سه ماژول پایه، تصمیم بگیرد، سپس یک مقدار احتمالاتی به عنوان احتمال کلیک‌خوردن یک سند توسط کاربر پیش‌بینی کند. مدل ارائه‌شده با استفاده از مجموعه‌داده‌ی موتور جست‌وجوی Yandex ارزیابی شده است و نتایج حاکی از برتری این مدل نسبت به مدل‌های شناخته‌شده‌ی گذشته دارد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Optimal Singular Value Decomposition Based Pre-coding for Secret Key Extraction from Correlated Orthogonal Frequency Division Multiplexing Sub-channels

A. Aliabadian, M. R. Zahabi*, M. Mobini

Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran

PAPER INFO

Paper history:
Received 07 November 2019
Received in revised form 02 June 2020
Accepted 18 June 2020

Keywords:
Orthogonal Frequency Division Multiplexing
Physical Layer Security
Secret Key
Singular Value Decomposition based Channel Decorrelation

ABSTRACT

Secret key extraction is a crucial issue in physical layer security and a less complex and, at the same time, a more robust scheme for the next generation of 5G and beyond. Unlike previous works on this topic, in which Orthogonal Frequency Division Multiplexing (OFDM) sub-channels were considered to be independent, the effect of correlation between sub-channels on the secret key rate is addressed in this paper. As an assumption, a realistic model for dependency among sub-channels is considered. Benchmarked by simulation, the result shows that the key exchange rate may decline by up to 72% due to the correlation of sub-channels. A new approach for efficient key extraction is used in this study. To do this, a Singular Value Decomposition based (SVD-based) pre-coding is utilized to alleviate the sub-channels correlation and the channel noise. The low computational complexity of our proposed approach makes it a promising candidate for developing secure and high-speed networks. Results obtained through simulation indicate that applying pre-coding on the measured correlated data resulted in a minimum gain of 9 dB. In addition, the result also depicts the advantage of SVD versus other pre-coding techniques, namely PCA, DCT, and WT.

doi: *10.5829/ije.2020.33.07a.07*

1. INTRODUCTION

Efficient secret key generation of physical layer and authentication schemes based on wireless channels are developing issues in physical layer security, especially in Orthogonal Frequency Division Multiplexing (OFDM)-based communication systems. Therefore, the channel parameter has received a lot of attention in literature as the fundamental part of key construction techniques [1-3].

Depending on the environmental conditions, the Key Generation Rate (KGR) is defined as a parameter that specifies the rate of secret key bits generated per second. Similarly, Key Disagreement Rate (KDR) is defined as the distinction rate of the key bits generated by Alice and Bob as the two ends of the communication link.

In recent studies, OFDM structures have been utilized for key generation with long sequences and increasing the rate of key generation by obtaining a key for each sub-channel in a coherence time. Since KGR and KDR oppose each other, a reasonable tradeoff should be set. This setting is configured between them by considering the demands of the system and the user interface. In literature, the key generation method is separated into four stages including the channel probing, quantization, information reconciliation and privacy amplification [4-6].

In the channel probing stage, the transmitter and the receiver use the static period of channel parameters in a coherence time interval. The extracted parameters from the channel are channel impulse response, channel frequency response, received signal strength, and channel phase.

In this stage, due to several reasons such as the channel noise, random displacements, multipath and scattering, the values measured by both ends of the communication link are not equal so some pre-processing should be done [7]. Unequal measurements of the

*Corresponding Author Email: *zahabi@nit.ac.ir* (M. R. Zahabi)

Please cite this article as: A. Aliabadian, M. R. Zahabi, M. Mobini, Optimal Singular Value Decomposition Based Pre-coding for Secret Key Extraction from Correlated Orthogonal Frequency Division Multiplexing Sub-channels, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1214-1222

channel lead to disagreement among the keys, while a high KDR might lead to the deficiency of the key generation process [8, 9].

The quantization scheme is utilized to optimize the operation of randomness, KGR, and KDR by adjusting the level of quantization and the threshold limit [10, 11].

Another part of the information is also sent through public channels during the information reconciliation stage, which can be heard by Eve as a wire-tapper. This can potentially threaten the security of the key sequence. Privacy amplification is finally used for removing the revealed information from the agreed key sequence by legitimate users (Alice and Bob).

In the key generation, based on channel reciprocity, the secret key is made from one or more channel parameters such as channel phase, Received Signal Strength (RSS), and Channel State Information (CSI) [5, 12].

In the previous works on this topic, it is typically assumed that the sub-channels do not correlate for the sake of simplicity. In practice, there is a correlation between the sub-channels that suppresse the assumption of randomness. Thus, determining the secret key rate by considering the correlation between sub-channels and maintaining the randomness as well as increasing the KGR is essential in physical layer security. However, in this paper, the correlation between the sub-channels and its effect on the secret key rate is studied.

There is no theoretical model for the mutual correlation between the measured values due to the lack of closed form for it. Thus, the correlation can be improved by interpolation or filter configuration, which is commonly done through different experiments [8, 13]. Due to the considerable impact of noise in a slow fading channel, a Low Pass Filter (LPF*)* is needed to eliminate the high-frequency components of the noise and to enhance the correlation [14].

In [9], efficient signal pre-coding is addressed and it is demonstrated that Principal Component Analysis (PCA)-based pre-coding achieves a higher KGR than Discrete Cosine Transform (DCT) and Wavelet Transform (WT). In other words, the channel correlation can be eliminated by a signal pre-processing procedure such as PCA [15], DCT [16, 17], and WT [18, 19].

The singular value decomposition (SVD) is a factorization of a real or complex matrix by which the original matrix is expressed by/forms three matrix, namely USV*. One of the most widely used functions of SVD is noise elimination and reduction of measured correlations.

In this paper, inspired by [9], an SVD-based channel decorrelation is proposed which is more accurate than the PCA-based method, with a significant superiority from the computational complexity point of view.

The main contributions that distinguish our work from others in literature are as follows:

- In previous works, authors did not consider the correlation among sub-channels, for the sake of simplicity. However, in our study, the effect of correlation among sub-channels is evaluated by applying a new realistic model.
- An optimal SVD-based pre-coding scheme is presnted which has lower computational complexity than other works. Moreover, the Mutual Information (MI) calculation and KGR improvement are provided.
- Our proposed SVD-based method is numerically compared with some other approaches, especially with the PCA-based method as an appropriate benchmark to determine the method which is better for key extraction; to the best of our knowledge, the numerical aspect has not been yet considered in any previous literature. As a result of SVD-based pre-coding, one will be able to obtain an optimal key generation.

The remainder of this paper is organized as follows: In Section 2, the communication and adversary models are presented. The correlation of the sub-channels and its effect on the secret key rate is also derived in this section. In Section 3, our proposed SVD-based pre-coding is addressed. The comparison among SVD and other approaches are also done, and advantages of the proposed approach are further investigated. In Section 4, simulation results are expressed. Finally, Section 5 deals with conclusions.

2. OFDM SYSTEM MODEL WITH CORRELATED SUB-CHANNLE

Figure 1 shows an OFDM-based system model including three nodes in which Alice and Bob are known as legitimate users and Eve is known as the adversary or the interceptor [20]. In this model, the adversary can only initiate a passive attack. In other words, Eve can tap into the connection between legitimate users and search for the secret key based on her deductions. Therefore, she cannot affect the information between Alice and Bob [21, 22]. Alice and Bob use a half-duplex communication system. Thus, they cannot simultaneously send and receive a signal. Mathematically, Alice and Bob attempt to estimate the channel by sending signals according to the following formulas:

$$\mathrm{h_A = h_{BA} + z_A} \qquad , \qquad \mathrm{h_B = h_{AB} + z_B} \tag{1}$$

in which $\mathrm{z_A}$ and $\mathrm{z_B}$ refer to the channel noises at the locations of Alice and Bob respectively. $\mathrm{h_{BA}}$ and $\mathrm{h_{AB}}$ are the legitimate channel vectors in frequency domain.

The measured values of the channel such as CSI, RSS, and channel phase are collected by Alice and Bob. They probe the channel by sending successive time signals in each period. Due to channel reciprocity in coherence time, a high correlation exists between the measured values of Alice and Bob, and these measured

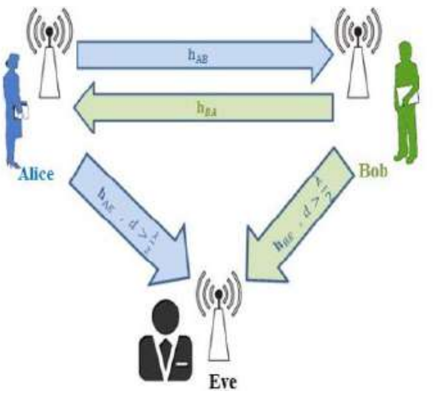


Figure 1. The communication model

values of the channel will change into a vector of bits after quantization. However, the quantized values of the channel slightly differ from each other due to the existence of noise. Because of the channel reciprocity between $\mathbf{h}_{\mathrm{BA}}$ and $\mathbf{h}_{\mathrm{AB}}$ which are the channel vectors between Alice and Bob in the frequency domain, they can be written as:

$$\begin{aligned} &\mathrm{h_{BA}} = \mathrm{h_{AB}} = \mathrm{h} = [\mathrm{h_1}, \mathrm{h_2}, \ldots \mathrm{h_N}], \\ &\mathrm{z_A} = [\mathrm{z_{A1}},\ \mathrm{z_{A2}}, \ldots \mathrm{z_{AN}}] \quad , \quad \mathrm{z_B} = [\mathrm{z_{B1}},\ \mathrm{z_{B2}}, \ldots \mathrm{z_{BN}}]. \end{aligned} \tag{2}$$

In Figure 1, the terms $\mathbf{h}_{\mathrm{AE}}$ and $\mathbf{h}_{\mathrm{BE}}$ show information about Eve from the channel and it is assumed that Eve is positioned at a distance greater than one half-wavelength from Alice and Bob. Therefore, $\mathbf{h}_{\mathrm{E}}$ has no correlation with $\mathbf{h}_{\mathrm{A}}$ and $\mathbf{h}_{\mathrm{B}}$. Here, it is assumed that the sub-channels h_i, $i = 1,2,3, \ldots$ N have $CN(0, \sigma^2_{h_i})$ distribution and also $\boldsymbol{z}_{Bi}$ and $\boldsymbol{z}_{Ai}$ are independent noises with distributions $\boldsymbol{z}_{Ai} \sim CN(0, \sigma^2_{Z_{Ai}})$, and $\boldsymbol{z}_{Bi} \sim CN(0, \sigma^2_{\boldsymbol{z}_{Bi}})$ which are also independent of h_i. Therefore, the estimated values of $\mathbf{h}_{\mathrm{A}}$ and $\mathbf{h}_{\mathrm{B}}$, and their distributions can be readily written as:

$$\mathrm{h_A} = \mathrm{h_{BA}} + \mathrm{z_A} \quad , \quad \mathrm{h_B} = \mathrm{h_{AB}} + \mathrm{z_B} \tag{3}$$

$$\mathrm{h_A} \sim \mathrm{CN}(0, \mathrm{R_A}) \quad , \quad \mathrm{h_B} \sim \mathrm{CN}(0, \mathrm{R_B}) \tag{4}$$

$$(\mathrm{h_A}, \mathrm{h_B}) \sim \mathrm{CN}(0, \mathrm{R_{AB}}) \ , \ \mathrm{R_{AB}} = \begin{bmatrix} \mathrm{R_A} & \mathrm{R_C} \\ \mathrm{R_D} & \mathrm{R_B} \end{bmatrix} \tag{5}$$

where the covariance matrices $\boldsymbol{R}_A$ and $\boldsymbol{R}_B$ are both diagonal. In literature, for the sake of simplicity, it is assumed that the sub-channels have no correlation and the secret key rate in OFDM systems based on the sub-channel state information is calculated and analyzed. To increase the key generation rate, the corresponding keys k_i are generated for each independent sub-channel h_i from N sub-channels, and the long key sequence K is then generated as:

$$K = k_1 k_2 k_3 \ldots k_N \tag{6}$$

It should be noted that there is a correlation between the sub-channels which leads to the lower randomness in the practical scenarios. Thus, calculating the secret key rate by considering the amount of correlation among sub-channels and maintaining randomness as well as increasing the KGR would be crucial. The relation of the secret key rate for $\mathbf{h}_A$ and $\mathbf{h}_B$ is given as [14]:

$$\begin{aligned} I_m &= I(\mathbf{h}_{\mathrm{A}}, \mathbf{h}_{\mathrm{B}} | \mathbf{h}_{\mathrm{AE}}) = I(\mathbf{h}_{\mathrm{A}}, \mathbf{h}_{\mathrm{B}} | \mathbf{h}_{\mathrm{BE}}) = I(\mathbf{h}_A, \mathbf{h}_B) \\ &= H(\mathbf{h}_A) + H(\mathbf{h}_B) - H(\mathbf{h}_A, \mathbf{h}_B) \end{aligned} \tag{7}$$

in which H(.) refers to entropy function. Therefore, by considering the correlation among sub-channels in the OFDM system, the relation for the covariance between the two sub-channels h_{iA} and h_{jA} can be written as:

$$\begin{aligned} &COV(h_{iA}, h_{jA}) = COV\left[(h_i + n_{Ai}), (h_j + n_{Aj})\right] = \\ &E\left[(h_i + n_{Ai})(h_j + n_{Aj})^*\right] \\ &= E\left[h_i h_j^* + h_i n_{Aj}^* + h_j^* n_{Ai} + n_{Ai} n_{Aj}^*\right] \\ &= \begin{cases} E(h_i h_j^*) & i \neq j \\ E(|h_i|^2) + E(|n_i|^2) & i = j \end{cases}, \end{aligned} \tag{8}$$

where the covariance matrices R_A and R_B are diagonal matrices defined by:

$$\mathbf{R}_A = \begin{pmatrix} \sigma^2_{h_1} + \sigma^2_{n_{A_1}} & E(h_1 h_2^*) & \cdots & E(h_1 h_N^*) \\ E(h_2 h_1^*) & \sigma^2_{h_2} + \sigma^2_{n_{A_2}} & \ldots & \\ \vdots & & \ddots & \vdots \\ E(h_N h_1^*) & \ldots & & \sigma^2_{h_N} + \sigma^2_{n_{A_N}} \end{pmatrix} \tag{9}$$

$$\mathbf{R}_B = \begin{pmatrix} \sigma^2_{h_1} + \sigma^2_{n_{B_1}} & E(h_1 h_2^*) & \cdots & E(h_1 h_N^*) \\ E(h_2 h_1^*) & \sigma^2_{h_2} + \sigma^2_{n_{B_2}} & \ldots & \\ \vdots & & \ddots & \vdots \\ E(h_N h_1^*) & \ldots & & \sigma^2_{h_N} + \sigma^2_{n_{B_N}} \end{pmatrix} \tag{10}$$

Also $\mathbf{R}_C$ and $\mathbf{R}_D$ are defined by:

$$\mathbf{R}_C = \mathbf{R}_D = \begin{pmatrix} \sigma^2_{h_1} & E(h_1 h_2^*) & \cdots & E(h_1 h_N^*) \\ E(h_2 h_1^*) & \sigma^2_{h_2} & \ldots & \\ \vdots & & \ddots & \vdots \\ E(h_N h_1^*) & \ldots & & \sigma^2_{h_N} \end{pmatrix}. \tag{11}$$

Therefore, the formula for the secret key rate can be derived as follows:

$$\begin{aligned} &I(\mathbf{h}_A; \mathbf{h}_B) = log_2(|\pi e \mathbf{R}_A|) + log_2(|\pi e \mathbf{R}_B|) - \\ &log_2(|\pi e \mathbf{R}_{AB}|) = log_2\left(\frac{|\mathbf{R}_A| . |\mathbf{R}_B|}{|\mathbf{R}_A| . |\mathbf{R}_A - \mathbf{R}_D \mathbf{R}_A{}^{-1} \mathbf{R}_C|}\right). \end{aligned} \tag{12}$$

3. THE PROPOSED SVD-BASED PRE-CODING METHOD

In this section, our proposed SVD-based pre-coding method is presented. According to Equation (12), to achieve a more efficient secret key and preventing the similarity of key sequences, the correlation between sub-channels should be considered. Using Equation (12) in which the frequency correlation of the sub-channels is considered, a more accurate model for the KGR can be presented. It is worthwhile to point out that the correlation between the measured sub-channel coefficients of data can be eliminated by a signal pre-processing procedure such as the PCA [15], DCT [16, 17] and WT [18, 19]. In [9], a new pre-coding method is addressed and it is also demonstrated that PCA-based pre-coding achieves a higher KGR than both the DCT and the WT. In this paper, the SVD is applied on the covariance matrix of the channel $\mathbf{h}_A$. The correlation matrix is defined as:

$$\mathbf{R}_{\mathrm{A}} = E\{\mathbf{h}_A\mathbf{h}_A{}^H\} = \mathbf{U}_A(\mathbf{\Lambda}_A + \sigma_n^2\mathbf{I}_N)\mathbf{U}_A^H \tag{13}$$

where $\mathbf{U}_A = \left[u_A^{(1)}, u_A^{(2)}, \dots\ u_A^{(N)}\right]$ is the transform matrix and $\mathbf{\Lambda}_A$ is a diagonal matrix with sorted eigenvalues $(\lambda_1 \geq \lambda_2\ \dots \geq \lambda_i \geq\ \lambda_N)$. Our main goal is to obtain the optimal unitary matrix $\mathbf{V}^*$ through maximizing the MI which can be obtained using:

$$\mathbf{V}^* = Arg\ max_{\mathrm{v}}\ \tilde{\boldsymbol{I}} \qquad s.t\ \ \mathbf{V}^{\mathrm{H}}\mathbf{V} = \mathbf{I}_{\mathrm{M}},\ \ \mathbf{M} \leq \mathbf{N}. \tag{14}$$

Ultimately, $\mathbf{V}_A^*$, $\tilde{\mathbf{R}}_A$, $\tilde{\mathbf{R}}_B$, $\tilde{\mathbf{R}}_C$ and $\tilde{\mathbf{R}}_D$ are written as:

$$\mathbf{V}_A^* = \left[v_A^{*(1)}, v_A^{*(2)}, \dots\ v_A^{*(M)}\right], \tag{15}$$

$$\tilde{\mathbf{R}}_A = \mathbf{V}_A^*(\mathbf{\Lambda}_A + \sigma_n^2\mathbf{I}_M)\mathbf{V}_A^{*\mathrm{H}}, \tag{16}$$

$$\tilde{\mathbf{R}}_B = \mathbf{V}_B^*(\mathbf{\Lambda}_B + \sigma_n^2\mathbf{I}_M)\mathbf{V}_B^{*\mathrm{H}}, \tag{17}$$

$$\tilde{\mathbf{R}}_C = \mathbf{V}_C^*(\mathbf{\Lambda}_C + \sigma_n^2\mathbf{I}_M)\mathbf{V}_C^{*\mathrm{H}}, \tag{18}$$

$$\tilde{\mathbf{R}}_D = \mathbf{V}_D^*(\mathbf{\Lambda}_D + \sigma_n^2\mathbf{I}_M)\mathbf{V}_D^{*\mathrm{H}}. \tag{19}$$

The optimal transform matrix $\mathbf{V}_A^*$ is provided by the eigenvectors of the channel covariance matrix corresponding to the *M* maximum eigenvalues; the proof can be found in [9].

As explained above, it is demonstrated that the optimal V_A^* consists of the maximum *M* eigenvectors of the covariance matrix. This means that the proposed method mathematically obtains the optimal solution from the secret key rate aspect. Bob's pre-coding vector can be obtained similarly. In addition, it can be concluded from Equation (12) that:

$$\tilde{I}(\mathbf{h}_A; \mathbf{h}_B) = log_2\left(\frac{|\tilde{\mathbf{R}}_A|.|\tilde{\mathbf{R}}_B|}{|\tilde{\mathbf{R}}_A|.|\tilde{\mathbf{R}}_A - \tilde{\mathbf{R}}_D\tilde{\mathbf{R}}_A{}^{-1}\tilde{\mathbf{R}}_C|}\right). \tag{20}$$

The block diagram for the proposed system is shown in Figure 2.

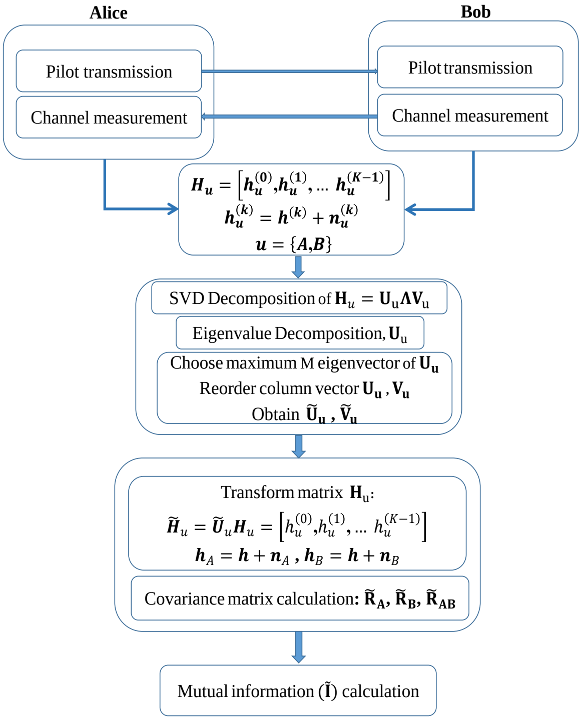

Alice
Bob
Pilot transmission
Channel measurement
Pilot transmission
Channel measurement
$\boldsymbol{H}_u = \left[\boldsymbol{h}_u^{(0)}, \boldsymbol{h}_u^{(1)}, \dots\ \boldsymbol{h}_u^{(K-1)}\right]$
$\boldsymbol{h}_u^{(k)} = \boldsymbol{h}^{(k)} + \boldsymbol{n}_u^{(k)}$
$\boldsymbol{u} = \{\boldsymbol{A}, \boldsymbol{B}\}$
SVD Decomposition of $\mathbf{H}_u = \mathbf{U}_{\mathrm{u}}\mathbf{\Lambda}\mathbf{V}_{\mathrm{u}}$
Eigenvalue Decomposition, $\mathbf{U}_{\mathrm{u}}$
Choose maximum M eigenvector of $\mathbf{U}_{\mathbf{u}}$
Reorder column vector $\mathbf{U}_{\mathbf{u}}$, $\mathbf{V}_{\mathbf{u}}$
Obtain $\tilde{\mathbf{U}}_{\mathbf{u}}$, $\tilde{\mathbf{V}}_{\mathbf{u}}$
Transform matrix $\mathbf{H}_{\mathrm{u}}$:
$\tilde{\boldsymbol{H}}_u = \tilde{\boldsymbol{U}}_u\boldsymbol{H}_u = \left[h_u^{(0)}, h_u^{(1)}, \dots\ h_u^{(K-1)}\right]$
$\boldsymbol{h}_A = \boldsymbol{h} + \boldsymbol{n}_A$, $\boldsymbol{h}_B = \boldsymbol{h} + \boldsymbol{n}_B$
Covariance matrix calculation: $\tilde{\mathbf{R}}_{\mathbf{A}}$, $\tilde{\mathbf{R}}_{\mathbf{B}}$, $\tilde{\mathbf{R}}_{\mathbf{AB}}$
Mutual information ($\tilde{\mathbf{I}}$) calculation

Figure 1. The block diagram for the proposed system

3. 1. Comparison between SVD and PCA

In this subsection, a comparison between SVD and PCA is presented. The SVD decomposes a diagonalizable matrix into special matrices that are straightforward to handle and to analyze; however, the PCA method maps some data linearly into different properties that are not correlated with each other. Meanwhile, PCA can use various algorithms to implement the principal component analysis such as Eigenvalue Decomposition of the covariance matrix. It is faster than SVD, but less accurate. Another algorithm is Alternating Least Squares (ALS) algorithm, which uses an iterative method with a random seed. It is designed to handle missing values [23].

The PCA-based approach is used to break down $\boldsymbol{R}_A$ into $\boldsymbol{U}_A\boldsymbol{\Lambda}_A\boldsymbol{V}_A^H$, only if $\boldsymbol{R}_A$ is a square matrix. Hereby, one result can be expanded for all matrices using SVD, which indicates the numerical advantages of this approach.
The matrices $\boldsymbol{R}_A\boldsymbol{R}_A^T$ and $\boldsymbol{R}_A^T\boldsymbol{R}_A$ are symmetric, square, positive semi-definite with positive eigenvalues, and both have the same rank. Since they are symmetric, it is inferred that its eigenvectors must be orthonormal. This is an essential property for symmetric matrices [24]. Therefore, using SVD, $\boldsymbol{R}_A$does not require to be square. The eigenvectors for $\boldsymbol{R}_A\boldsymbol{R}_A^T$ are shown as u_i and $\boldsymbol{R}_A^T\boldsymbol{R}_A$ shown as v_i, and these sets of eigenvectors are called $\boldsymbol{U}_A$ and $\boldsymbol{V}_A$ *singular vectors* of $\boldsymbol{R}_A$. The square roots of these eigenvalues are called *singular values*.

Comparing to Eigen-decomposition, SVD works on non-square matrices. $\boldsymbol{U}_A$ and $\boldsymbol{V}_A$ are invertible for any matrix in SVD and they are orthonormal. Besides, singular values are numerically more stable than eigenvalues [25]. Although the PCA-based method mathematically has an optimal performance for using in channel decorrelation and key extraction systems, the numerical aspect has not been considered yet in similar research efforts.

Using SVD has some numerical advantages over the PCA. A suitable approach for calculating PCA on a computer is needed to directly implement the eigenvalue decomposition of $\boldsymbol{R}_A\boldsymbol{R}_A^T$.

It results in some computational complexity that could be decreased by adopting SVD. Let $\tilde{\sigma}_\mathrm{i}$ be the output of an algorithm calculating singular values for $\mathrm{R_A}$, and σ_i be the appropriate singular value, it can be shown that:

$$|\tilde{\sigma}_i - \sigma_i| = O(\varepsilon\|\mathbf{R}_A\|)\,, \tag{21}$$

where $\|\mathbf{R}_A\|$ is a measure of the size of $\mathbf{R}_A$. On the other hand, for an algorithm that calculates the eigenvalues λ_i of $\mathbf{R}_A{}^\mathrm{T}\mathbf{R}_A$, it can be shown as:

$$|\tilde{\lambda}_i - \lambda_i| = O(\varepsilon\|\mathbf{R}_A{}^\mathrm{T}\mathbf{R}_A\|) = O(\varepsilon\|\mathbf{R}_A\|^2) \tag{22}$$

which in terms of the singular values of $\mathbf{R}_A$, it can be written as:

$$|\tilde{\sigma}_i - \sigma_i| = O\left(\varepsilon\frac{\|\mathbf{R}_A\|^2}{\sigma_i}\right). \tag{23}$$

Ultimately, when small singular values, e.g. $\sigma_i \ll \|\mathbf{R}_A\|$ are needed, using $\mathbf{R}_A$ will result in more accurate output than using $\mathbf{R}_A{}^\mathrm{T}\mathbf{R}_A$.

4. SIMULATION RESULTS

In this section, we simulate some examples to evaluate the performance of our proposed method. The Monte Carlo simulation is adopted using MATLAB to model both the SVD and PCA algorithms. In all simulation examples, we assume that the OFDM symbols undergo Rayleigh fading. Moreover, channel realizations are based on statistical distributions presented in [20]. In our simulation studies, some curves are presented for both dependent and independent of frequency domains to assess the effect of the SVD. Some verifications are further done to demonstrate the optimality of the proposed SVD-based pre-coding. Meanwhile, our proposed method is compared with a PCA-based pre-coding method in both the MI and computational complexity aspects. As a case study, we focus on the de-correlation algorithms for channel measurement with frequency correlation in the OFDM model. Channel measurements are generated by the channel responses of all the sub-channels of an OFDM symbol.

4. 1. Example 1: Considering Correlated Sub-Channels in OFDM Model

In this simulation example, we evaluate the secret key rate for both uncorrelated and correlated sub-channels. We assume that the variance of each sub-channel is random and normalized. The variance of sub-channels utilized in this example is listed in Table 1 for sub-channels.

Figure 3 shows the MI in terms of the SNR in outdoor conditions for constant T_m and independent sub-channels. As seen, the secret key rate is increased as the SNR becomes greater. Also, the secret key rate increases as the number of sub-channels increases. For N=8, 16, 32, the secret key rate is increased to 16.19, 27.42, 55.3 bits, respectively, where the target SNR is considered to be 8 dB. This increment in the secret key rate is due to the increase in the number of random sources as the greater number of sub-channels is utilized.

Figure 4 depicts the secret key rate for an OFDM-based system with correlated sub-channels with the above assumptions. As obviously observed, the rate in the dependent scenario is less than the independent scenario because of the correlation between sub-channels. As a result, where the target SNR is 8 dB, the secret key rate is obtained as Independent = 5.139, 8.814, 10.97, 15.17 bits, which are less than the values illustrated in Figure 3 for independent scenarios, for the same number of sub-channels.

Figure 5 illustrates the comparison of the secret key rate for both dependent and independent scenarios for an OFDM symbol including 32 sub-channels. As a result, the values of secret key rate where the target SNR is considered to be 8 dB are obtained as 15.17, 55.3 bits for correlated and independent sub-channels, respectively.

4. 2. Example 2: Evaluation of the Performance of Proposed SVD-Based Pre-coding

In this sub-section, an OFDM model including 32 sub-carriers is utilized to do the simulations. Figure 6 shows the covariance matrix of $\boldsymbol{h}_{iA}, \boldsymbol{h}_{jA}$. Each element in this pattern indicates the amount of correlation between the two sub-channels. As an up-down sorting, a significant correlation is illustrated in brown, red and yellow colors, respectively. The first *M*

TABLE 1. Variance vectors for different number of sub-channels

Number of sub-channels	**Variance**
$N = 4$	$\delta_h^2 = [1.1819 \quad 1.0935 \quad 0.3597 \quad 0.820]$
$N = 8$	$\delta_h^2 = [1.1819 \quad 1.0935 \quad 0.7653 \quad 1.2042 \quad 2.4845 \quad 1.299 \quad 0.3597 \quad 0.820]$
$N = 16$	$\delta_h^2 = \begin{bmatrix} 1.1194 & 0.6190 & 0.3706 & 0.7079 & 1.0966 & 0.9597 & 1.0638 & 0.2884 \\ 0.4019 & 0.8906 & 1.4591 & 0.5037 & 1.4586 & 0.8723 & 1.2445 & 0.6251 \end{bmatrix}$
$N = 32$	$\delta_h^2 = \begin{bmatrix} 1.1997 & 0.9788 & 0.9412 & 0.7271 & 1.5351 & 2.0229 & 1.3886 & 1.4335 \\ 1.1822 & 0.9545 & 0.4923 & 0.9560 & 2.5430 & 0.5750 & 0.9384 & 0.3889 \\ 1.0682 & 0.3771 & 0.9178 & 0.5328 & 0.8034 & 0.4233 & 0.3446 & 0.2924 \\ 0.4517 & 0.3455 & 0.3942 & 0.7221 & 0.5543 & 0.6546 & 2.3146 & 0.7294 \end{bmatrix}$

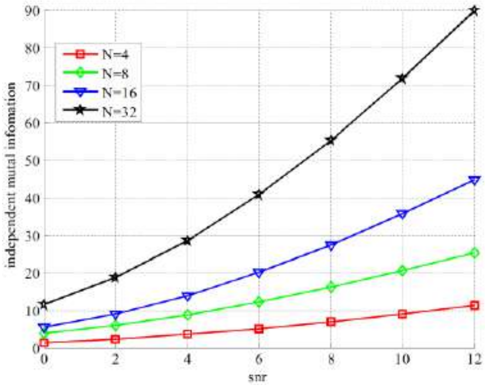


Figure 3. Secret key rate for independent sub-channels

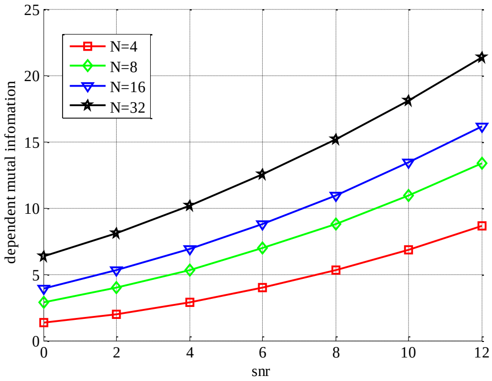


Figure 4. Secret key rate for correlated Sub-channels

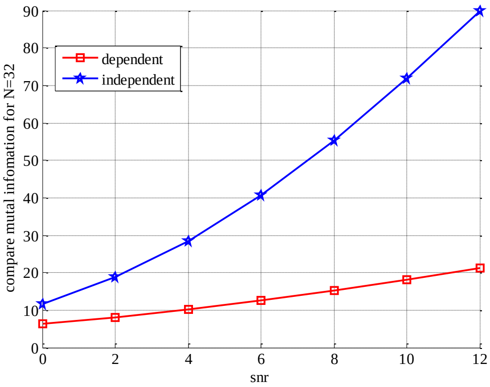


Figure 5. The comparison of the secret key rate between independent and dependent sub-channels for N = 32 and constant $\mathrm{T_m}$

rows of *V* represent the signal subspace and the other rows describe the noise subspace which may be separated. The most important feature of the SVD is that the matrix *U*, which transforms the matrix of correlated sub-channels into a matrix of uncorrelated coefficients is unitary.

Figure 7 illustrates a 3D-view of the covariance matrix values. It should be noted that in our proposed approach, we try to select the elements of the matrix that are effective in the key generation process by discarding the destructive elements.

According to the PCA approach, the covariance matrix of the ideal channel $\mathbf{h}_A$ can be decomposed as $\mathbf{R}_A = \mathbf{U}_A\mathbf{\Lambda}_A\mathbf{U}_A^{\mathrm{H}}$, where $\mathbf{U}_A = \left[u_A^{(1)}, u_A^{(2)}, \dots u_A^{(N)}\right]$ is the transform matrix and $\mathbf{\Lambda}_A$is a diagonal matrix with some sorted eigenvalues.

After transformation, matrix $\mathbf{H}_u$ is transformed into $\mathbf{Y}_u = \left[y_u^{(0)}, y_u^{(1)}, \dots y_u^{(K-1)}\right]$ by expanding the CSI estimates with their projections on $\mathbf{Y}_u = \mathbf{U}^{\mathbf{H}}\mathbf{H}_u$. Because only a part of $\mathbf{U}$ is used for key generation, column vectors can be re-ordered and a part be choosen as a $M \times K$ matrix $\mathbf{Y}_u$ as $\mathbf{V} = [\mathbf{V}^{(1)}, \dots, \mathbf{V}^{(M)}]$. This is an $N \times K$ matrix and satisfies $\mathbf{V}^{\mathrm{H}}\mathbf{V} = \mathbf{I}_M$, $M \leq N$. The output $M \times K$ matrix $\widetilde{\mathbf{Y}}_u$ is obtained by $\widetilde{\mathbf{Y}}_u = \mathbf{V}^{\mathrm{H}}\mathbf{H}_u$, where $\widetilde{\mathbf{Y}}_u$ is the result of the pre-coding procedure. Figure 7 shows the values of the new covariance matrix as long as we consider the equivalent channel. Figures 8 and 9 show the values of the covariance matrix of $\mathrm{R_B}$ after applying SVD and PCA pre-coding. As seen, the proposed algorithm will eliminate the peaks of the covariance matrix in Figure 9 and the smoother values are well selected.

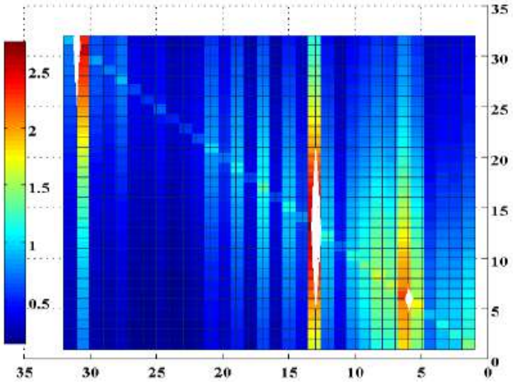


Figure 6. Covariance matrix of original channel. Each element indicates the correlation between two sub-channels $\boldsymbol{h}_{iA}$,$\boldsymbol{h}_{jA}$

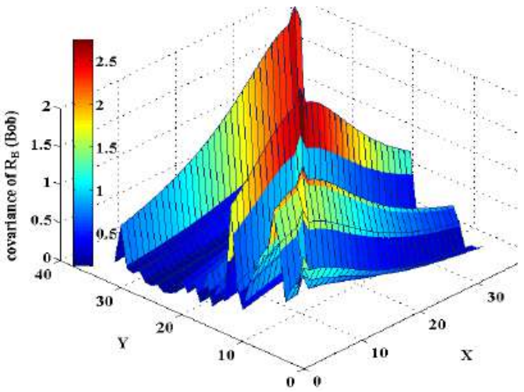


Figure 7. The covariance matrix of the original channel

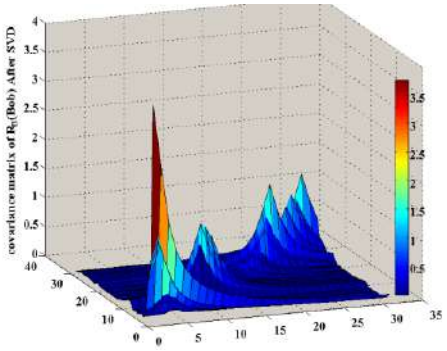


Figure 8. The covariance matrix of $\mathbf{R}_B$ (Bob) after SVD

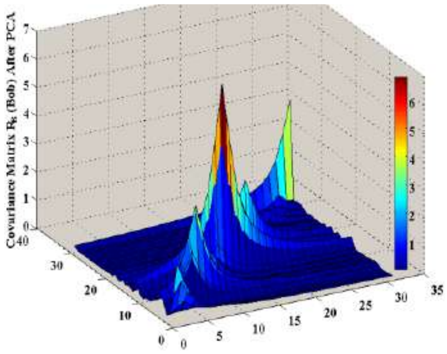


Figure 9. The covariance matrix of of $\mathbf{R}_B$ (Bob) after PCA

Figure 10 shows a comparison among the effect of the SVD, PCA, SVD and Wavelet matrix transforms on the MI. As expected, the proposed SVD-based method performs better than the other conventional methods from the MI aspect. In our simulations, the conventional OFDM-based key generation rate is first numerically plotted by substituting the covariance matrix $\mathbf{R}_A$ in Equation (13). Then, the transformed channel $\mathbf{R}_A$ is applied to Equation (12). As seen, where an OFDM symbol is utilized including *32* correlated sub-channels, 16 bits are otained in terms of the key generation rate at a target SNR of *8 dB*. It can be observed that using each of the above mentioned transform matrices will lead to improvements in terms of MI.

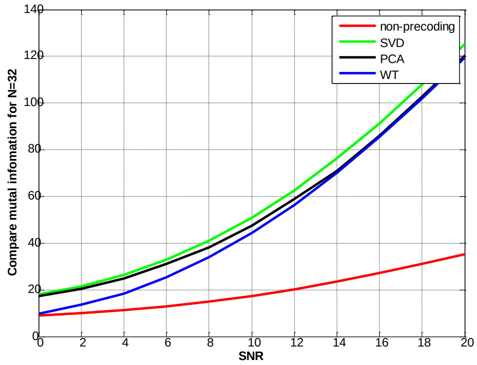


Figure 10. The effect of SVD-based pre-coding on the MI and comparison with DCT, PCA, and WT

4. 3. Example 3: On the Optimality of the SVD-Based Pre-coding In this simulation, we aim to obtain a contact zone between the SVD and PCA methods. As a result of SVD based pre-coding, we can achieve this optimal zone in which the performance of both methods are equal and complex. Ultimately, the proposed SVD-based pre-coding has a promising computational complexity. The comparison of MI performance is further performed in Figures 11 and 12 according to the variable M defined in Equation (14), for SNR=20 dB and SNR=10 dB respectively according to the results obtained from *32* correlated sub-channels, the SVD-based method outperforms the PCA-based method for both of the two scenarios.

Figures 12 and 13 show that the proposed SVD-based scheme has a promising superiority in terms of computational complexity due to adopting decomposition property of the SVD and eliminating of the non-related sub-routines of the PCA procedure. Assessment of the results for a case study including 32 correlated sub-channels verifies that the SVD based method achieves the optimal performance while consuming less processing time which is a crucial parameter for high-speed applications in 5G communications and other low-delay demanded secure applications.

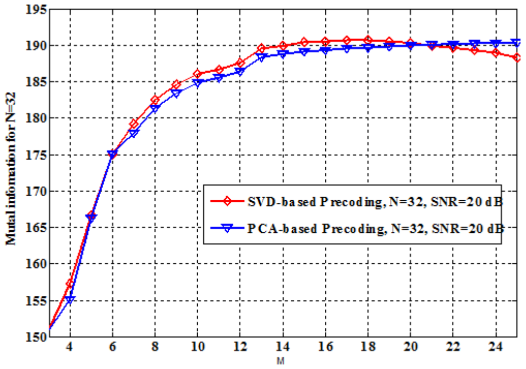


Figure 11. A comparison between the MI performance of the SVD and PCA methods for SNR=20db

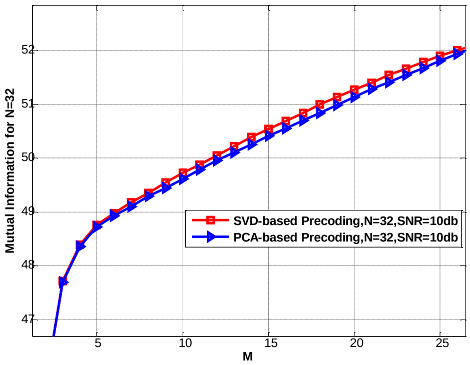


Figure 12. A comparison between the MI performance of the SVD and PCA methods for SNR=10db

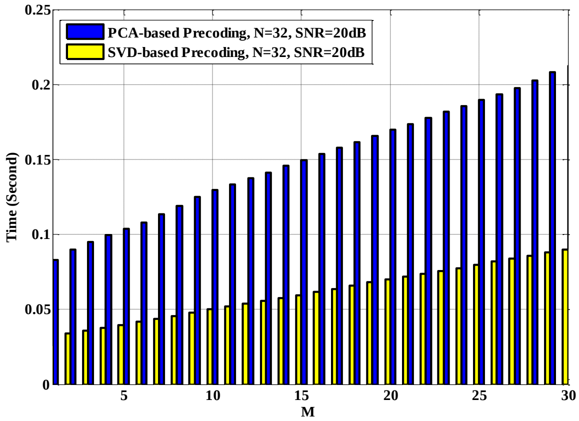


Figure 13. A comparison between the computational complexity of the SVD and PCA-based methods

5. CONCLUSION

In this paper, an SVD-based method was addressed for key extraction in OFDM-based communication systems. It was demonstrated that applying our proposed method has considerable advantages over the PCA, WT and DCT-based methods. Having considered a practical assumption, we investigated the effect of correlation among sub-channels on the secret key rate. Moreover, it was numerically shown that if we deal with temporal correlation alongside frequency correlation, KGR decreases and a more accurate value for secret rate is obtained. Simulation results demonstrated that the computational complexity of our proposed SVD-based pre-coding has promising superiorities as a better MI is obtained.

6. REFERENCES

1. Melki, R., Noura, H.N., Mansour, M.M. and Chehab, A., "A survey on ofdm physical layer security", ***Physical Communication***, Vol. 32, No., (2019), 1-30. https://doi.org/10.1016/j.phycom.2018.10.008
2. Chen, Y., Wen, H., Wu, J., Song, H., Xu, A., Jiang, Y., Zhang, T. and Wang, Z., "Clustering based physical-layer authentication in edge computing systems with asymmetric resources", ***Sensors***, Vol. 19, No. 8, (2019), 1926. https://doi.org/10.3390/s19081926
3. Jiang, Y., Zou, Y., Guo, H., Zhu, J. and Gu, J., "Power allocation for intelligent interference exploitation aided physical-layer security in ofdm-based heterogeneous cellular networks", ***IEEE Transactions on Vehicular Technology***, Vol. 69, No. 3, (2020), 3021-3033. doi: 10.1109/TVT.2020.2966637
4. Zhan, F. and Yao, N., "Efficient key generation leveraging wireless channel reciprocity and discrete cosine transform", ***KSII Transactions on Internet and Information Systems***, Vol. 11, No. 5, (2017), 2701-2722. https://doi.org/10.3837/tiis.2017.05.022
5. Zhang, J., Duong, T.Q., Marshall, A. and Woods, R., "Key generation from wireless channels: A review", ***IEEE Access***, Vol. 4, (2016), 614-626. doi: 10.1109/ACCESS.2016.2521718
6. Cheng, L., Zhou, L., Seet, B.-C., Li, W., Ma, D. and Wei, J., "Efficient physical-layer secret key generation and authentication schemes based on wireless channel-phase", ***Mobile Information Systems***, Vol. 2017, (2017), 1-13. https://doi.org/10.1155/2017/7393526
7. Genkin, D., Pachmanov, L., Pipman, I., Tromer, E. and Yarom, Y., "Ecdsa key extraction from mobile devices via nonintrusive physical side channels", In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. (2016), 1626-1638. https://doi.org/10.1145/2976749.2978353
8. Liu, H., Yang, J., Wang, Y. and Chen, Y., "Collaborative secret key extraction leveraging received signal strength in mobile wireless networks", In Proceedings IEEE INFOCOM, IEEE, (2012), 927-935. doi: 10.1109/INFCOM.2012.6195843
9. Li, G., Hu, A., Zhang, J., Peng, L., Sun, C. and Cao, D., "High-agreement uncorrelated secret key generation based on principal component analysis preprocessing", ***IEEE Transactions on Communications***, Vol. 66, No. 7, (2018), 3022-3034. doi: 10.1109/TCOMM.2018.2814607
10. Bloch, M., Thangaraj, A., McLaughlin, S.W. and Merolla, J.-M., "Ldpc-based secret key agreement over the gaussian wiretap channel", In IEEE International Symposium on Information Theory, IEEE, (2006), 1179-1183. doi: 10.1109/ISIT.2006.261991
11. Peng, L., Li, G. and Hu, A., "Channel reciprocity improvement of secret key generation with loop-back transmissions", In IEEE 17th International Conference on Communication Technology (ICCT), IEEE, (2017), 193-198. doi: 10.1109/ICCT.2017.8359629
12. Shehadeh, Y.E.H. and Hogrefe, D., "An optimal guard-intervals based mechanism for key generation from multipath wireless channels", In 4th IFIP International Conference on New Technologies, Mobility and Security, IEEE, (2011), 1-5. doi: 10.1109/NTMS.2011.5720584
13. Liu, H., Wang, Y., Yang, J. and Chen, Y., "Fast and practical secret key extraction by exploiting channel response", In Proceedings IEEE INFOCOM, IEEE, (2013), 3048-3056. doi: 10.1109/INFCOM.2013.6567117
14. Zhang, J., Marshall, A., Woods, R. and Duong, T.Q., "Efficient key generation by exploiting randomness from channel responses of individual ofdm subcarriers", ***IEEE Transactions on Communications***, Vol. 64, No. 6, (2016), 2578-2588. doi: 10.1109/TCOMM.2016.2552165
15. Chen, C. and Jensen, M.A., "Secret key establishment using temporally and spatially correlated wireless channel coefficients", ***IEEE Transactions on Mobile Computing***, Vol. 10, No. 2, (2010), 205-215. doi: 10.1109/TMC.2010.114
16. Margelis, G., Fafoutis, X., Oikonomou, G., Piechocki, R., Tryfonas, T. and Thomas, P., "Physical layer secret-key generation with discreet cosine transform for the internet of things", In IEEE International Conference on Communications (ICC), IEEE, (2017), 1-6. doi: 10.1109/ICC.2017.7997419
17. Margelis, G., Fafoutis, X., Oikonomou, G., Piechocki, R., Tryfonas, T. and Thomas, P., "Efficient dct-based secret key generation for the internet of things", ***Ad Hoc Networks***, Vol. 92, (2019), 101744. https://doi.org/10.1016/j.adhoc.2018.08.014
18. Wu, Y., Sun, Y., Zhan, L. and Ji, Y., "Low mismatch key agreement based on wavelet-transform trend and fuzzy vault in body area network", ***International Journal of Distributed Sensor Networks***, Vol. 9, No. 6, (2013), 1-16. https://doi.org/10.1155/2013/912873
19. Zhan, F. and Yao, N., "On the using of discrete wavelet transform for physical layer key generation", ***Ad Hoc Networks***, Vol. 64, (2017), 22-31. https://doi.org/10.1016/j.adhoc.2017.06.003
20. Cheng, L., Li, W., Zhou, L., Zhu, C., Wei, J. and Guo, Y., "Increasing secret key capacity of ofdm systems: A geometric program approach", ***Concurrency and Computation: Practice***

and Experience, Vol. 29, No. 16, (2017), e3966. https://doi.org/10.1002/cpe.3966

21. Poor, H.V. and Schaefer, R.F., “Wireless physical layer security”, ***Proceedings of the National Academy of Sciences***, Vol. 114, No. 1, (2017), 19-26, https://doi.org/10.1073/pnas.1618130114
22. Lai, L., Liang, Y., Poor, H.V. and Du, W., Key generation from wireless channels. In Physical Layer Security in Wireless Communications (pp. 47-92). CRC Press, (2013).
23. Arcidiacono, C. and Simoncini, V., “Approximate nonnegative matrix factorization algorithm for the analysis of angular differential imaging data”, In Adaptive Optics Systems VI, International Society for Optics and Photonics. Vol. 10703, United States, (2018). https://doi.org/10.1117/12.2311681
24. Marques, O. and Vasconcelos, P.B., “Computing the bidiagonal svd through an associated tridiagonal eigenproblem”, In International Conference on Vector and Parallel Processing, Springer, (2016), 64-74. https://doi.org/10.1007/978-3-319-61982-8_8
25. Gillis, N., Mehrmann, V. and Sharma, P., “Computing the nearest stable matrix pairs”, ***Numerical Linear Algebra with Applications***, Vol. 25, No. 5, (2018), 1-19. https://doi.org/10.1002/nla.2153

Persian Abstract

چکیده

در امنیت مبتنی برلایه فیزیکی، استخراج کلید امن مسئله‌ای مهم است که در نسل‌های مخابراتی 5G و نسل‌های مخابراتی آتی، روشی با پیچیدگی کم و در عین حال مقاوم می‌باشد. برخلاف تحقیقات گذشته که زیرکانال‌ها در سیستم‌های OFDM مستقل فرض می‌گردید، در این مقاله به تأثیر همبستگی زیرکانال‌ها پرداخته شده است و یک مدل واقعی برای همبستگی زیرکانال‌ها استفاده شده است. نتایج شبیه‌سازی نشان می‌دهد که که همبستگی زیرکانال‌ها باعث کاهش اطلاعات متقابل تا ۷۲ درصد می‌شود. رویکرد جدیدی برای دست یافتن به اطلاعات متقابل بهینه و استخراج کلید بکار گرفته شده است. بدین منظور از پیش پردازش SVD برای کاهش میزان همبستگی مابین مقادیر اندازه‌گیری شده کانال و همچنین کاهش نویز استفاده گردید. پیچیدگی محاسباتی کم در رویکرد پیشنهادی، روش امیدوار کننده‌ای برای توسعه‌ی شبکه‌های امن و پرسرعت می‌باشد. نتایج شبیه‌سازی نشان می‌دهد که اعمال پیش‌پردازش بر روی مقادیر اندازه‌گیری شده‌ی وابسته، منجر به بهره حداقل ۹ دسی‌بل شده است. همچنین برتری عملکرد SVD نسبت به روش‌های دیگر نظیر PCA, DCT, WT با شکل، نمایش داده شده است.

IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1223-1230

International Journal of Engineering

Journal Homepage: www.ije.ir

AIOSC: Analytical Integer Word-length Optimization Based on System Characteristics for Recursive Fixed-Point Linear Time Invariant Systems

M. Grailoo, B. Alizadeh*

School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

PAPER INFO

Paper history:
Received 13 October 2019
Received in revised form 13 October 2019
Accepted 17 June 2020

Keywords:
Analytical Formulation
Fixed-Point Arithmetic
High-Level Synthesis and Optimization
Linear-Time-Invariant System
Range Analysis
Resource Usage

ABSTRACT

The integer word-length optimization known as range analysis (RA) of the fixed-point designs is a challenging problem in high level synthesis and optimization of linear-time-invariant (LTI) systems. The analysis has significant effects on the resource usage, accuracy and efficiency of the final implementation, as well as the optimization time. Conventional methods in recursive LTI systems suffer from inaccurate range estimations due to dependency to symmetry or non-symmetry of the input range over zero, and involvement with parameter adjustments. The under estimations endanger the range safety, and generate a great error due to overflows. On the other hand, the over estimations increase the hardware costs, as well as weaken the signal, if the over estimated ranges are utilized in down-scaling. Therefore, in this paper, we propose an efficient, safe and more precise RA method to measure the range of both recursive and non-recursive fixed-point LTI systems through analytical formulation. Our main idea is to obtain the input sequences for which variables in the LTI system would be maximum and minimum. By applying these sequences to the system, the upper and lower bounds of the intended variables are obtained as the range. The proposed method enhances the bit-widths accuracy more than 34% in average in comparison with the state-of-the-arts. The results also show about 37% and 6% savings in the area and delay, respectively.

***doi**: 10.5829/ije.2020.33.07a.08*

1. INTRODUCTION

The increasing complexity of modern embedded applications in recent decades has forced design methodologies and tools to move to higher abstraction levels. Raising the abstraction levels, and accelerating automation of the synthesis, optimization and verification processes in addition to reducing the time-to-market, help to reduce the verification time as well as facilitate other flows such as accuracy analysis. In high level optimization, a crucial decision to be made is the datapath word length, including the word length of different registers and functional units. In this work, we concentrate on fixed-point representation due to the preference for fixed-point implementations of digital signal processing (DSP) algorithms over floating-point because of hardware cost reduction. Deciding on factors such as integer width and fractional parts of the circuit has significant effects on the resource consumption, accuracy and efficiency of the final implementation. To have finite-precision fixed-point implementations of such systems, range analysis (RA) is an essential and fundamental design step. The analysis characterizes the integer bit-widths (IB) for all the fixed-point variables such that no overflow and underflow occur [1–10].

In this paper, an analytical integer word-length optimization for recursive LTI systems is proposed. LTI systems are the most important category of DSP applications since they include finite-impulse response (FIR), infinite-impulse response (IIR) digital filters, and

* Corresponding Author Email: *b.alizadeh@ut.ac.ir* (B. Alizadeh)

Please cite this article as: M. Grailoo, B. Alizadeh, AIOSC: Analytical Integer Word-length Optimization Based on System Characteristics for Recursive Fixed-Point Linear Time Invariant Systems, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1223-1230

signal transformations such as Fast Fourier Transform, Discrete Cosine Transform, and Wavelet Transforms [9, 10]. The method not only minimizes the hardware implementation cost, but also reduces the optimization time significantly. In the method, the safe and more precise range is obtained through analytical formulation without any involvement of the parameter adjustments, and without additional iterative operations. The estimations in the method are independent of symmetry or non-symmetry of input range over zero. To do so, the method directly extracts the two input sequences, for which the variables would be maximum and minimum, from the impulse response using the theorem explained in Section 4. The sequences are then applied to the system to obtain the upper and lower bounds of the intended variables. Note that the theorem in this work is also applicable to the feed-forward systems.

The remainder of this paper is organized as follows. Section 2 reviews previous works. Section 3 states our contributions. Section 4 details the proposed range analysis flow through a simple example. Section 5 investigates the experimental results and finally, Section 6 concludes the paper.

2. RELATED WORKS

Several approaches have been introduced to tackle range analysis problems of fixed-point designs which, in general, can be categorized into dynamic and static analyses. Dynamic analysis methods evaluate the system by using input stimulus. This analysis suffers from unsafe, data dependent, and time-consuming estimations, which confine its applicability [8]. Static analysis, however, uses static characteristics of the inputs which are propagated through the system. So, it has recently gained much interest due to safety, no data dependency, and higher efficiency [1–8]. In static analysis, one of the most significant categories is self-validated numerical (SVN) methods. The two most popular SVN methods are interval arithmetic (IA) and affine-arithmetic (AA) [2]. Due to the efficiency of these methods in terms of analysis time, many literatures use them or their extensions to account for RA. The other category of static methods uses more sophisticated approaches such as SMT-based range analysis [7], and hybrid [8] as a combination of IA, AA and AT. These tighter results in the recent addressing methods are obtained at the cost of more analysis time consumption.

Such solutions, however, may not always be adequate, due to being unable to handle recursive circuits, such as IIR filters. Since several fixed-point DSP circuits are based on arithmetic expressions with possible feedbacks, the RA of such circuits, in general, remains still challenging. The main challenge of such systems is to determine final amount of a value when it falls into an infinite loop. In this regard, the methods in [4, 5], utilize L1-norm and L2-norm of impulse responses to compute an inaccurate measurement of the exact range. The L1-norm-based methods in [5] also use the maximum absolute value of the input to obtain the output range. This leads to an over-estimation when the input range is non-symmetric over zero. The over-estimations increase the hardware costs, as well as weaken the signal, if the over-estimated ranges are utilized in down-scaling. The L2-norm-based method in [4] multiplies the maximum absolute value by the L2-norm of the impulse response. The L2-norm-based method under-estimates the ranges when the input is symmetric over zero. The under-estimations endanger the range safety, and generate a great error due to overflows. In order to obtain a tighter range than L1-norm, the method in [3] computes the range by iterative operations of flattening the system, $y[n]$. The analysis will face the problem of adjusting the two parameters to determine the required number of iterations. The parameters are the convergence window size, i.e. w, and the resolution of convergence, i.e. ε. Since the convergence of the algorithm depends on the position of poles and the stability conditions, there is no guarantee to precisely adjust the parameters. So, there is always a probability for an under-estimation in this method which is unacceptable in RA. For comparing our RA method in terms of the precision and hardware cost saving, three methods with the over- and under- estimations are chosen. They include the L1-norm and L2-norm methods due to their prestige and popularity in the scope of analytical range determination of LTI systems. Also, we compare our method with the flattening-based method as an iterative method.

3. OUR CONTRIBUTION

In order to clarify our main contributions, in this section we explain our ideas for efficient RA, obtaining more precise integer bit-widths in a bounded-input, bounded-output (BIBO) stable LTI. Our basic idea in this paper is to analyze the range from the system impulse response without any involvement in any parameter adjustments issues, and iterative operations.

In order to find the output range, we aim to find the input sequences for which the output will be maximum and minimum. To extract the input sequences, we use the impulse response of a system and the input bounds as will be explained in the following. The maximum and minimum input sequences, as well as the input upper and lower bounds are called $InputSeq_{max}[n]$ and $InputSeq_{min}[n]$, as well as x_{max} and x_{min}, respectively. In the following, we only consider $InputSeq_{max}[n]$, and the primary output variable of y which has the impulse response, i.e. $h[n]$, according to Figure 1(a). Similar arguments exist for intermediate variables with different impulse responses.

The output of a system is obtained by convolving an input sequence and its impulse response. In order to obtain the maximum output, we consider a sequence in a state that has the most overlapping with the impulse response as illustrated in Figure 1(b). In the state, the output maximum is obtained when the input would be in the upper bound, where the impulse response is positive, as well as the input would be in the lower bound, where the impulse response is negative as illustrated in Figure 1(b). This input sequence, which we are looking for to maximize the output, i.e. $InputSeq_{max}$, follows the impulse response form such that places in its input upper, i.e. x_{max}, or lower bounds, i.e. x_{min}, where the impulse response is positive or negative, respectively. Since in the other states with less overlapping, the input sequences generate lower output values, they are not investigated. The $InputSeq_{min}$ is also obtained in a similar way in which the input sequence would be x_{min} and x_{max}, when the impulse response is positive or negative, respectively.

The sequences are then getting backward in time and applied to the system, to account for the output upper and lower bounds. These operations will be repeated for each variable. Since there are variables with the same impulse responses, these variables are grouped together in order to reduce the number of repetitive computations. In fact, the variables, with the same impulse response, constitute a group. Hence, our main contribution is a new method for static RA of LTI systems with or without feedbacks, to achieve safe, more efficient, and more accurate range than the state-of-the-art methods.

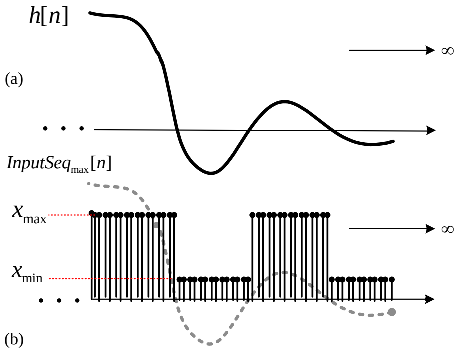


Figure 1. Idea illustration for range analysis of LTI systems: a) impulse response of a system plotted in time domain; b) the input sequence for obtaining upper bound of y when the input is between x_{max} and x_{min}

4. PRPOSED RANGE ANALYSIS

In this section, we propose the RA method, called Analytical Integer Word-length Optimization based on System Characteristics (AIOSC). As mentioned before, RA is crucial for the discrete system design in the implementation of a BIBO stable LTI system. The ranges are used to assign suitable integer bit-widths for all variables such that it is guaranteed that no underflow and overflow happen. Our method finds an input sequence that maximizes the output of a system when it is convolved by the impulse response. The sequence is obtained by following the impulse response form according to Theorem 1. Before introducing the algorithm; we first prove the theorem, which is needed in the rest of this section.

Theorem 1: Two input sequences, in which the BIBO stable LTI system, i.e. $y[n]$, would be maximum and minimum, are $InputSeq_{max}[n]$, and $InputSeq_{min}[n]$, respectively. They are obtained as follows, where $u[n]$ is the unit step function.

$$InputSeq_{max}[n] = x_{max} \times u\big[h[n]\big] + x_{min} \times u\big[-h[n]\big] \quad (1)$$

$$InputSeq_{min}[n] = x_{min} \times u\big[h[n]\big] + x_{max} \times u\big[-h[n]\big] \quad (2)$$

Proof: As discussed in Section 3, the input sequences include only the maximum and minimum of the system input, i.e. x_{max} and x_{min}. Choosing between x_{max} and x_{min} depends on the values of $h[k], k \in \{0,1,\dots,n\}$, as follows:

$$InputSeq_{max}[n] = \begin{cases} x_{max} & if\ h[k] \times x_{max} \geq h[k] \times x_{min} \\ x_{min} & if\ h[k] \times x_{max} < h[k] \times x_{min} \end{cases} \quad (3)$$

$$InputSeq_{min}[n] = \begin{cases} x_{max} & if\ h[k] \times x_{max} \leq h[k] \times x_{min} \\ x_{min} & if\ h[k] \times x_{max} > h[k] \times x_{min} \end{cases} \quad (4)$$

Since $x_{max} \geq x_{min}$, the above relations can be simplified as follows:

$$InputSeq_{max}[n] = \begin{cases} x_{max} & if\ h[k] \geq 0 \\ x_{min} & if\ h[k] < 0 \end{cases} \quad (5)$$

$$InputSeq_{min}[n] = \begin{cases} x_{max} & if\ h[k] \leq 0 \\ x_{min} & if\ h[k] > 0 \end{cases} \quad (6)$$

These relations are equivalent to $InputSeq_{max}[n] = x_{max} \times u\big[h[n]\big] + x_{min} \times u\big[-h[n]\big]$ and $InputSeq_{min}[n] = x_{min} \times u\big[h[n]\big] + x_{max} \times u\big[-h[n]\big]$. The sequences can also be obtained through the Equations of (5) and (6).

4. 1. Range Analysis Flow The proposed flow for RA is shown in Figure 2. It takes the input bounds, i.e. $[x_{min}, x_{max}]$, as inputs, and returns the variable integer bit-widths as outputs. This flow is repeated for each group of the variables. In fact, the variables, with the same impulse response, constitute a group in order to reduce the number of repetitive computations. In Step 1, the impulse response for each group is obtained from its linear constant-coefficient difference equation (LCCDE), if it currently does not exist. In Step 2, the input sequences, i.e. $InputSeq_{max}[n]$ and $InputSeq_{min}[n]$, are found based on Theorem 1. In this step, the function $UnitStep()$ from Mathematica is invoked to apply the unit step function to the impulse response. In Step 3, these sequences are getting backward in time, and applied to the system. This response can be obtained by direct evaluation of the convolution sum of the sequences and the impulse response, as indicated in the figure where "$*$" denotes convolution. However, since the convolution in the time domain corresponds to multiplication in the z-domain, another simple alternative is obtaining the response in the z-domain. So, the z-transform of the sequences, and the impulse response can be created by the function $ZTransform()$ from Mathematica [11]. Then the z-transform of the impulse response is multiplied by the z-transforms of the sequences. Finally, the function $InverseZTransform()$ is invoked to obtain the corresponding results in the time domain, i.e. the $UpperBound\,[n]$ and $LowerBound\,[n]$. In Step 4, the minimum and maximum of the mentioned functions (called a and b) will constitute the final range, i.e. $[a, b]$. To obtain the bit-width including sign bit from the range, the following relation is employed.

$$i = \lceil log_2(max\ (|a|, |b|)) \rceil + \alpha,$$
$$\alpha = \begin{cases} 1 & if\ \ mod(log_2(b)) \neq 0 \\ 2 & if\ \ mod(log_2(b)) = 0 \end{cases} \quad (7)$$

4. 2. Example In order to clarify the flow, let us consider the example of $y[n] = \alpha y[n\ -1] + \beta x[n]$, with $\alpha = 0.8$ and $\beta = 0.5$. The example is a low pass filter, which enjoys wide applications in control systems, Kalman filtering, communication processing to reduce noise, and image averaging. The filter with all the input and intermediate variables, as some vertical rectangles, is shown in Figure 3(a). In this example, the variables x_2 to x_4 offer the same impulse response, which differs from the impulse response of x_1. So the variables are broken down into two groups: x_1 in G_1, and x_2 to x_4 in G_2. For G1, first (according to Step 1) the impulse response $h[n]$ is obtained by using direct and inverse z-transform. In order

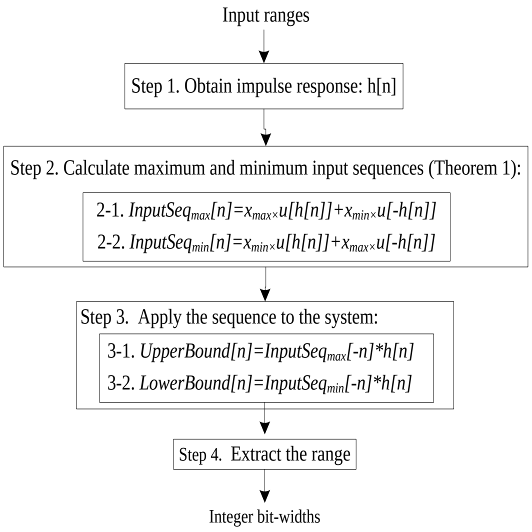

Input ranges
Step 1. Obtain impulse response: h[n]
Step 2. Calculate maximum and minimum input sequences (Theorem 1):
2-1. $InputSeq_{max}[n]=x_{max}\times u[h[n]]+x_{min}\times u[-h[n]]$
2-2. $InputSeq_{min}[n]=x_{min}\times u[h[n]]+x_{max}\times u[-h[n]]$
Step 3. Apply the sequence to the system:
3-1. $UpperBound[n]=InputSeq_{max}[-n]*h[n]$
3-2. $LowerBound[n]=InputSeq_{min}[-n]*h[n]$
Step 4. Extract the range
Integer bit-widths

Figure 2. Proposed range analysis flow

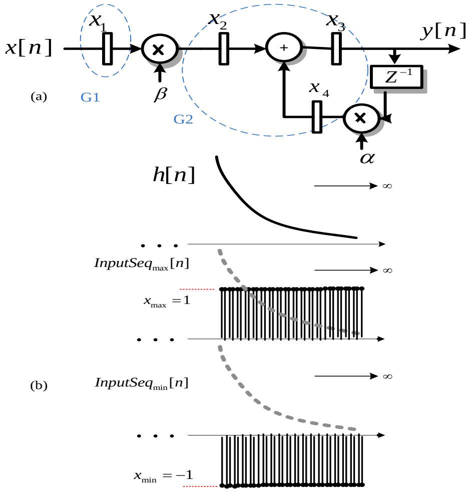


Figure 3. a) one-pole digital filter with intermediate variables; b) the impulse response and input sequences, $InputSeq_{max}$ and $InputSeq_{min}$, of the filter

to obtain $h[n]$, the other way is to solve the difference equation of $y[n]$, when $x[n]$ is replaced by $\delta[n]$, and $y[n]$ is replaced by $h[n]$. The impulse response for the variables in G_1 would be $h[n] = 0.5\ (0.8)^n u[n]$. Second, based on Theorem 1, the input sequences for $x_{max} = 1$ and $x_{min} = -1$ would be $InputSeq_{max}[n] = u[n]$ and $InputSeq_{min}[n] = -u[n]$. The impulse response of $h[n]$ and the input sequences of $InputSeq_{max}[n]$ and $InputSeq_{min}[n]$ are depicted in Figure 3(b). As shown in this figure, $InputSeq_{max}[n]$ gets the maximum input when $h[n]$ is positive, and the minimum input when $h[n]$ is negative. Since in this case, $h[n]$ is always positive, $InputSeq_{max}[n]$ would be $u[n]$, and vice versa for $InputSeq_{min}[n]$. Third, since the backward of the sequences in time domain are also unit step functions, these sequences are applied to the system as follows:

$$UpperBound[n] = \mathcal{Z}^{-1}\left\{\frac{0.5z}{z-0.8} \times \mathcal{Z}\{u[n]\}\right\} = \mathcal{Z}^{-1}\left\{\frac{0.5z}{z-0.8} \times \frac{z}{z-1}\right\} = 2.5 - 2 \times e^{-0.223144n} \tag{8}$$

$$LowerBound[n] = \mathcal{Z}^{-1}\left\{\frac{0.5z}{z-0.8} \times \mathcal{Z}\{-u[n]\}\right\} = \mathcal{Z}^{-1}\left\{\frac{0.5z}{z-0.8} \times \frac{-z}{z-1}\right\} = -2.5 + 2 \times e^{-0.223144n} \tag{9}$$

Finally, when n approaches infinity, the range x_3 is obtained which is $[-2.5, 2.5]$. These values for x_1, x_2, and x_4 are $[-1,1]$, $[-0.5,0.5]$ and $[-2,2]$, respectively. The integer bit-widths for x_1 to x_4 are $IB_{x_1} = 2$, $IB_{x_2} = 1$, $IB_{x_3} = 3$ and $IB_{x_4} = 3$. The obtained output range and integer bit-width by L2-norm for x_3 are $[-0.83, 0.83]$ and $IB_{x_3} = 1$, respectively. It is obvious that these measurements under-estimate the exact ones.

5. EXPERIMENTAL RESULTS

In order to demonstrate the applicability of our proposed method in different types and forms of the recursive LTI systems, as well as the superiority of the method over the state-of-the-arts, we have provided several benchmarks with various forms and types. The forms are direct (DR), parallel (PRL), and cascade (CS), as well as the types are high-pass (HPF), low-pass (LPF), and band-pass (BPF) filters. Bench #3 is a bi-quad eighth-order cascaded structure of four 2nd-order direct-form IIR filters. The last benchmark is also a National Television Systems Committee (NTSC) channel cascaded eighth-order LPF IIR filter with the cutoff frequency of 4.74MHz. The details of the benchmarks such as type, order, numerator, and denominator coefficients are given in Table 1. Our algorithm has been implemented with Mathematica, and run on an Intel 4702MQ core i7 with 8 GBs of main memory, running Linux operating system. For the synthesis process, the tool Xilinx ISE V14.1 on the Virtex-7 FPGAs target has been chosen. The device contains user-programmable elements known as slices, dedicated multiply-and-add units, DSP blocks and embedded RAMs. In order to make fair comparisons, the designs are implemented by using slices and combinatorial elements without any pipelining. The variable indexes in the feedback parts have been numbered in a clockwise direction. In the first experiment, we compare AIOSC with L2-norm-based method (L2-norm) in [4] and flattening-based methods in [3] to show the precision of our method. It is assumed that the primary inputs are symmetric over zero, and lie within the normalized range of $[-1,1]$. The estimated range, bit-widths, and their under-estimation ratio have been reported in Table 2. In the table, the first major column has listed the benchmarks. The second and third major columns include the estimated ranges and bit-widths by the RA methods. Finally, the last column shows the under-estimation ratio of the AIOSC than the state-of-the-art methods. As shown in the table, L2-norm when the

TABLE 1. Range and bit-width evaluation results of AIOSC and L2-norm for the primary output variable

Bench #	Estimated Range			Estimated bit-width			Underestimation Ratio %			
	AIOSC	L2-norm	Flattening-based	AIOSC	L2-norm	Flattening-based	L2-norm/AIOSC		Flattening-based /AIOSC	
							Range	Bit	Range	Bit
1	[-270.89,270.89]	[-103.33,103.33]	[-232.15,232.15]	10	8	9	162	25	17	12
2	[-4.58,4.58]	[-2.18,2.18]	[-4.47,4.47]	4	3	4	109	33	3	0
3 Quad	[-76.23,76.23]	[-25.40,25.40]	[-75.25,75.25]	8	6	8	200	33	2	0
4 NTSC	[-275.12,275.12]	[-73.26,73.26]	[-273.92,273.92]	10	8	10	275	25	5	0
				Average underestimation ratio %			186.5	29	6.75	3

TABLE 2. Benchmark features

Bench#	Type	Form	Order	Numerator Full-precision Coefficient	Denominator Full-precision Coefficient
1	**HPF**	**DR**	2	101.8, -203.4, 101.6	1, -1.967, 0.968
2	**LPF**	**PRL**	3	2,0.1,-0.4	1,0.1,-0.46,0.08
3 Quad	**BPF**	**CS**	8	1, 2, 1 1, -2, 1 1, 2, 1 1, -2, 1	1, a, b 1, -a, b 1, c, d 1, -c, d a=0.47583613785934908, b=0.63399428536347535, c=1.0921588046377746, d=0.87447915380668007
4 NTSC	**LPF**	**CS**	8	1,2,1 1,2,1 1,2,1 1,2,1	1, a, b 1, c, d 1, e, f 1, g, h a=-0.7093449002973562, b=0.19225253081578914, c=-0.22413592126247239, d=0.41113157239125847, e=0.27362911645488941, f=0.66517393946636161, g=0.57030039990570558, h=0.88861236005184185

input is symmetric over zero under-estimates ranges and bit-widths, in all benchmarks. The under-estimations are more in the higher order benchmarks of Quad and NTSC. The ranges and bit-widths under-estimations are about 186% and 29% on average, respectively. Hence the estimations generate a great error due to overflows. Obtaining the exact output range requires the exhaustive simulations by feeding all possible sequences into inputs. The sequences are infinite for recursive filters. So, generating all possible infinite sequences is time consuming and even impossible in high order filters.

In the flattening-based method, the window size and the resolution are considered $(w, \varepsilon) = (10,1)$. As illustrated in the table, the ranges are under-estimated in all benchmarks. The under-estimations in the first benchmark lead to the under-estimated bit-width. In the other benchmarks, if the under-estimated ranges are used in the signal down-scaling, it can cause the overflow in the variables, which encompass the larger numbers. Let us consider the second benchmark. The flattening-based method estimates the maximum absolute range of 4.47 while the output variable can accept the number ± 4.58. In this case, all signals are divided by 4.47 and the output encompasses the number 1.02. The number is more than one which led to the output overflow. So, the flattening-based method under-estimates range and generates a great error.

In the next experiment, we concentrate on the safe methods, i.e. L1-norm-based method (L1-norm) [5], in comparison to AIOSC. In the experiment, it is assumed that the primary inputs are non-symmetric over zero and lie within the range of [9,10]. The bit-widths estimations for primary outputs are depicted in Figure 4. In this figure, the other estimations include the ranges plus the improvements are also shown as some entries of the small tables beside the bit-width bars. As seen in this figure, the L1-norm-based method constitutes over-estimations when the input bound is not symmetric over zero. The range over-estimations in some benchmarks are more than 20 times than the estimated range by AIOSC. If the over-estimated ranges are utilized in down-scaling, the range can strongly weaken the signal. Moreover, the range over-estimations result in an additional integer bit for the all benchmarks. As seen, by increasing the range over-estimations, the excess bits are also growing. The excess bits growing have significant impact on hardware cost. The amount of the impact is investigated in the next experiment. In the experiment, the AIOSC method shows the bit-width improvement of about 34.75% on average.

As mentioned, the effect of inaccuracy in the opposite direction, i.e., over-estimations instead of under-estimations, is on the hardware cost. Whatever the ranges of all intermediate and output variables are more exact, we expect to achieve the smallest bit-widths, leading to a

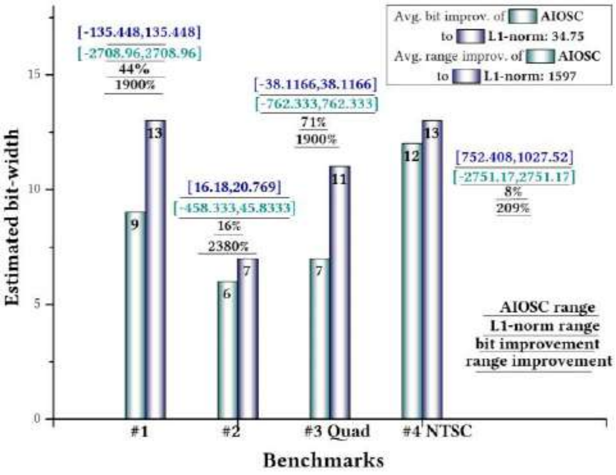


Figure 4. The estimation results of different RA methods

reduction in the circuit area and delay. In order to complete the experiment, the effect of the over-estimations on the area and delay are investigated in Figure 5. This figure indicates the area costs of the benchmarks for the assigned bit-widths obtained by AIOSC and L1-norm when using Xilinx ISE for the synthesis process. In this figure, other results of delay plus area and delay savings are also shown as some entries of the small tables beside the area bars. As illustrated in Figure 5, area and delay almost follow the estimated bit-widths. It means, in the positions that one method has estimated lower bit-widths; the delay and area are pursuing this flow and become less. The area (slice) and delay saving of AIOSC is 37.25% and 5.6% in comparison with L1-norm, respectively.

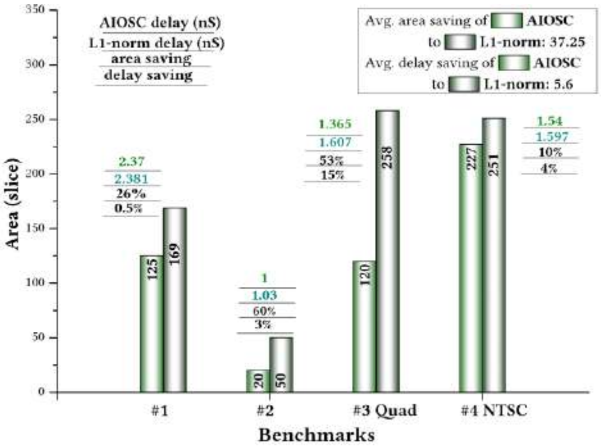


Figure 5. Area and delay comparison of RA methods

6. CONCLUSIONS AND FUTURE WORK

The range analysis plays an important role in high-level synthesis of arithmetic circuits, as it can directly impact the overall design cost and performance. Most of existing analyses on the recursive LTI systems estimate the bounds inaccurately. It leads to produce some great errors or increase the hardware cost. Therefore, in this paper, a new, more accurate and efficient RA method for fixed-point recursive LTI systems was proposed. The method obtained a safe and more precise range and bit-width estimations from the impulse response, without any involvement of the parameter adjustments, and without any additional iterative operations. The proposed method brought advantages of 29% bit-width improvement. It led to 37.25% and 5.6% area and delay saving in comparison with the previous state-of-the-art methods.

As our future work, we are going to extend our method to support the error analysis in LTI systems with feedback for the maximum mismatch (MM), mean square error (MSE) and signal to quantization noise ratio (SQRT) metrics.

7. REFRENCES

1. Grailoo, M., Alizadeh, B. and Forouzandeh, B., “Improved range analysis in fixed-point polynomial data-path”, ***IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems***, Vol. 36, No. 11, (2017), 1925–1929. doi: 10.1109/TCAD.2017.2666607
2. Grailoo, M., Alizadeh, B. and Forouzandeh, B., “UAFEA: Unified analytical framework for IA/AA-based error analysis of fixed-point polynomial specifications”, ***IEEE Transactions on Circuits and Systems II: Express Briefs***, Vol. 63, No. 10, (2016), 994–998. doi: 10.1109/TCSII.2016.2539078
3. Sarbishei, O., Radecka, K., and Zilic, Z., “Analytical optimization of bit-widths in fixed-point LTI systems”, ***IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems***, Vol. 31, No. 3, (2012), 343–355. doi: 10.1109/TCAD.2011.2170988
4. Abbas, M., Gustafsson, O. and Johansson, H., “On the fixed-point implementation of fractional-delay filters based on the Farrow structure”, ***IEEE Transactions on Circuits and Systems I,*** Vol. 60, No. 4, (2013), 926–937. doi: 10.1109/TCSI.2013.2244272
5. Rocher, R., Menard, D., Scalart, P. and Sentieys, O., “Analytical Approach for Numerical Accuracy Estimation of Fixed-Point Systems Based on Smooth Operations”, ***IEEE Transactions on Circuits and Systems I***, Vol. 59, No. 10, (2012), 2326–2339. doi: 10.1109/TCSI.2012.2188938ï
6. Chung, J. and Kim, L. W., “Bit-Width Optimization by Divide-and-Conquer for Fixed-Point Digital Signal Processing Systems”, ***IEEE Transactions on Computers***, Vol. 64, No. 11, (2015), 3091–3101. doi: 10.1109/TC.2015.2394469
7. Kinsman, A. B. and Nicolici, N., “Bit-width allocation for hardware accelerators for scientific computing using SAT-modulo theory”, ***IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems***, Vol. 29, No. 3, (2010), 405–413. doi: 10.1109/TCAD.2010.2041839
8. Pang, Y., Radecka, K., and Zilic, Z., “An efficient hybrid engine to perform range analysis and allocate integer bit-widths for arithmetic circuits”, In Proceedings of the Asia and South Pacific Design Automation Conference, ASP-DAC, (2011), 455–460. doi: 10.1109/ASPDAC.2011.5722233
9. Chernoyarov, O.V., Golpaiegani, L.A., Glushkov, A.N., Lintvinenko, V.P. and Matveev, B.V., “Digital binary phase-shift keyed signal detector”, ***International Journal of Engineering - Transactions A: Basics***, Vol. 32, No. 4, (2019), 510–518. doi: 10.5829/IJE.2019.32.04A.08
10. Kumar, P., and Kumar Chaudhary, S., “Stability and Robust Performance Analysis of Fractional Order Controller over Conventional Controller Design”, ***International Journal of Engineering - Transactions B: Applications*** , Vol. 31, No. 2, (2018), 322–330. doi: 10.5829/ije.2018.31.02b.17
11. Wolframe Mathematica: definitive system for modern technical. [Online]. Available: http://www.wolfram.com/mathematica/

Persian Abstract

چکیده

بهینه‌سازی طول کلمه صحیح یا تحلیل دامنه، یک مساله چالش‌برانگیز در بهینه‌سازی و سنتز سطح بالای سیستم‌های خطی نامتغیر با زمان بازگشتی محسوب می‌گردد. این تحلیل، تاثیر بسزایی بر مصرف منابع، دقت، کارایی و زمان بهینه‌سازی می‌گذارد. روش‌های پیشین، از معایبی چون تخمین نادقیق دامنه شامل تخمین مازاد یا تخمین زیر مقدار واقعی رنج می‌برند. این عدم دقت بدلیل وابستگی روش‌ها به تقارن ورودی نسبت به صفر و همچنین وابستگی به برخی از پارامترها می‌باشد. تخمین‌های زیر مقدار واقعی، سبب ایجاد سرریز و تولید خطاهای بزرگ می‌شود. از طرف دیگر، تخمین مازاد، هزینه سخت افزار را افزایش می‌دهد. همچنین اگر این تخمین مازاد در مقیاس کردن استفاده شود، سبب تضعیف سیگنال می‌گردد. بنابراین در این مقاله یک روش تحلیل دامنه دقیق، کارا و ایمن با روش‌های تحلیلی برای اندازه‌گیری دامنه در سیستم‌های خطی نامتغیر با زمان بازگشتی و غیربازگشتی برای طراحی‌های ممیز ثابت، ارائه می‌شود. ایده اصلی یافتن دنباله ورودی برای هر متغیر است که به ازای آن خروجی سیستم ماکسیسمم و مینیمم گردد. با اعمال این دنباله‌ها به سیستم، محدوده بالا و پایین هر متغیر به عنوان دامنه بدست می‌آید. روش ارائه شده، دقت طول کلمه را بطور متوسط تا بیش از ۳۴٪ در مقایسه با روش‌های قبلی بهبود می‌بخشد. نتایج همچنین ۳۷٪ بهبود در مساحت و ٦٪ بهبود تاخیر را نشان می‌دهد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Channel Estimation and Carrier Frequency Offset Compensation in Orthogonal Frequency Division Multiplexing System Using Adaptive Filters in Wavelet Transform Domain

A. R. Fereydouni, A. Charmin*, H. Vahdati, H. Nasir Aghdam

Department of Electrical Engineering, Islamic Azad University, Ahar Branch, Ahar, Iran

PAPER INFO

Paper history:
Received 10 September 2019
Received in revised form 30 November 2019
Accepted 11 June 2020

Keywords:
Channel Estimation
Per-Tone Equalization Structure
Set-Membership Filtering
Short Cyclic Prefix
Wavelet Packet Transform

ABSTRACT

In this paper, a combination of channel, receiver frequency-dependent IQ imbalance, and carrier frequency offset estimation under short cyclic prefix length is considered in orthogonal frequency division multiplexing system. An adaptive algorithm based on the set-membership filtering algorithm is used to compensate for these impairments. In short CP length, per-tone equalization structure is used to avoid inter-symbol interference. Due to CFO impairment and IQ imbalance in the receiver, we will expand the PTEQ structure to a two-branch structure. This structure has high computational complexity, so using the set-membership filtering idea with variable step size while reducing the average computation of the system can also increase the convergence speed of the estimates. On the other hand, applying wavelet transform on each branch of this structure before applying adaptive filters will increase the estimation speed. The proposed algorithm will be named SMF-WP-NLMS-PTEQ. The results of the simulations show better performance than the usual adaptive algorithms. Besides, estimation and compensation of channel effects, receiver IQ imbalance and carrier frequency offset under short CP can be easily accomplished by this algorithm.

***doi**: 10.5829/ije.2020.33.07a.09*

1. INTRODUCTION

The use of high-performance wireless communication systems is steadily growing [1]. As a multicarrier modulation technique, the orthogonal frequency division multiplexing (OFDM) technique can provide reliable high data rate transmission in different communication scenarios [2, 3].

The tendency to use cutting-edge technology and smaller elements in the same silicon area can also cause problems. For example, replacing the direct conversion architecture (DCA) with a superheterodyne configuration scheme with low cost and lower power utilization in silicon compared to the super heterodyne can lead to an imbalance of Inphase and Quadrature (IQ). IQ imbalance can be described as the difference between the gain, phase, and even frequency response of each orthogonal branch of communication systems [4, 5].

The OFDM direct conversion receiver also suffers from carrier frequency offset and DC offset in addition to IQ imbalance. On the other hand, the OFDM system is very sensitive to the carrier frequency offset (CFO) [6]. In OFDM systems, the presence of some frequency offset will disrupt the orthogonality of the sub-carriers, resulting in inter-carrier interference and a loss of performance [7, 8].

In [9], an empirical mode decomposition (EMD) based adaptive filter for channel estimation in an OFDM system is proposed. In this method, the length of the channel impulse response (CIR) is first approximated. Then, CIR is estimated using an adaptive filter through the received OFDM symbol.

* Corresponding Author Email: *a_charmin@sut.ac.ir* (A. Charmein)

Please cite this article as: A. R. Fereydouni, A. Charmein, H. Vahdati, H. Nasir Aghdam, Channel Estimation and Carrier Frequency Offset Compensation in Orthogonal Frequency Division Multiplexing System Using Adaptive Filters in Wavelet Transform Domain, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1231-1239

In [10], an efficient IQ imbalance compensation scheme is proposed based on pilots; this is achieved through a nonlinear least squares (NLS) analysis of the joint channel and IQ imbalance estimation, and a simple symbol detection procedure. For rigor, the Cramer-Rao lower bounds (CRLBs) for both the IQ imbalance parameters and the channel coefficients are also derived.

In [11], a novel joint channel impulse response estimation and impulsive noise mitigation algorithm based on compressed sensing theory is proposed. In this algorithm, both the channel impulse response and the impulsive noise are treated as a joint sparse vector. Then, the sparse Bayesian learning framework is adopted to jointly estimate the channel impulse response, the impulsive noise, and the data symbols, in which the data symbols are regarded as unknown parameters.

In [12], an adaptive approach to eliminate the receiver frequency IQ imbalance in OFDM systems based on a direct conversion scheme is implemented using decision feedback adaptive filtering.

In [13], the effect of receiver IQ imbalance on the OFDM system is investigated and a design based on training data in both time and frequency domain is presented. In [14], the same method is applied assuming the presence of transmitter IQ. In [15], an integrated structure for estimating transmitter, receiver, and channel IQ under short CP conditions is presented. In this structure, PTEQ structure is used to overcome the inter-symbol-interference (ISI) problem. Short CP increases bandwidth efficiency.

Employing cyclic-prefix (CP) in multi-carrier systems not only protects the signal from inter-symbol-interference but also allows circular interpretations of the channel which simplifies the estimation and equalization techniques. Nevertheless, the CP information is usually discarded at the receiver side [16].

In this paper, short CP is considered for optimal use of the channel. Therefore, PTEQ structure-based algorithms will be used. In addition to effectively estimating the channel, the receiver IQ imbalance and CFO prevent ISI and deterioration of the estimation results. In this paper, to increase the speed of estimation of the adaptive algorithms, in each sub-carrier of the PTEQ output data, a wavelet transform is implemented. Then the NLMS adaptive algorithm is developed in the new output signal. Due to the length of filters required and the number of subcarriers, the implementation of any adaptive algorithm in this structure will require a lot of computation for convergence.

Therefore, computational optimization of the applied algorithms while maintaining the required convergence speed and steady-state error is necessary. In the second step, to reduce the computation of the channel estimation process based on the adaptive algorithm, we need to increase the convergence speed again. For this purpose, we use set-membership filtering in adaptive filters. The use of SMF filtering can improve the estimation speed alone by using variable step-size, optimal, and dependent on noise variance. Also, in some iteration, it prevents partial sub-carriers from being updated and reduces average computation.

Therefore, the implementation of an adaptive algorithm in the PTEQ structure in combination with wavelet transform and set-membership filtering will have several important features. First, it will increase channel utilization efficiency by reducing the length of CP used. Second, the ISI can be eliminated for the short CP utilization and, ultimately, due to SMF and WP transform using, has high speed and accuracy in channel and receiver IQ imbalance estimation. The CFO effect can also be compensated in the proposed structure by existing methods independent of channel effects and IQ imbalances. Therefore, using SMF-WP-NLMS-PTEQ proposed method, good channel equalization can be performed for the data.

This paper first describes the OFDM-based receiver IQ imbalance and channel estimation model. Then the wavelet packet transform is reviewed in Section 2. The proposed adaptive compensation algorithm is presented using SMF-based adaptive filters in the wavelet transform domain and PTEQ structure under short CP in Section 3. Finally, the simulation results are presented in Section 4 and the conclusions in Section 5.

2. CHANNEL ESTIMATION MODEL IN OFDM SYSTEMS

In this section, the compensation of channel distortion under sufficient CP length will be studied and corresponding analytical equations will be introduced to model the channel effect so that based on these equations the proposed method can be presented in the next sections. In sufficient CP length, inter-symbol-interference does not occur, so a first-order filter as a compensator can minimize existing distortions. But in the case of short CP length, a structure called PTEQ is used to compensate and estimate the data.

2. 1. Sufficient CP Length

First, for a sufficient CP, the channel compensation scheme is examined. It is assumed that S represents the OFDM symbol in the frequency domain with$(N \times 1)$ length, where N is the number of subcarriers. Thus the baseband symbol in the time domain can be written as Equation (1).

$$s = P_{CI}F_N^{-1}S \tag{1}$$

where P_{CI} is a cyclic prefix (CP) insertion matrix of length υ to symbol S, and F_N^{-1} represents the inverse matrix of DFT.

When the output symbol in the transmitter passes through the semi-stationary channel with a length of L_{ch}, the received baseband symbol r can be written as Equation (2).

$$r = c \otimes s + n \tag{2}$$

In Equation (2), c represents the baseband channel model. Also, n is Gaussian additive white noise.

The OFDM system is sensitive to the CFO due to changes in the transmitter and receiver local oscillators. Therefore, its effect must be compensated. By applying the CFO effect on the received signal, Equation (3) is obtained.

$$z = r.e^{j2\pi\Delta f.t} \tag{3}$$

when considering the receiver IQ imbalance as the CFO and the channel effect, the received signal can be written as Equation (4).

$$y = y_{1r} \otimes \left(r.e^{j2\pi\Delta f.t}\right) + y_{2r} \otimes \left(r^*.e^{-j2\pi\Delta f.t}\right) = y_{1r} \otimes \left((c \otimes s + n).e^{j2\pi\Delta f.t}\right) + y_{2r} \otimes \left((c \otimes s + n)^*.e^{-j2\pi\Delta f.t}\right) \tag{4}$$

where y_{1r} and y_{2r} are the frequency selective imbalance filter model of the receiver in branches I and Q. In the frequency domain, the CFO will result in a leakage of energy from the desired sub-carrier to all sub-carriers in one OFDM symbol. The energy leakage from sub-carrier$[l]$to other sub-carrier $[l']$ will be in Equation (5) [17].

$$P_\zeta[l'] = e^{j\pi(\zeta-l')\frac{N-1}{N}}\frac{\sin\pi(\zeta-l')}{\sin\frac{\pi}{N}(\zeta-l')} \tag{5}$$

where ζ as normalized CFO, is the actual ratio of CFO namely Δf and the distance between the substrates $\frac{1}{T.N}$ and is obtained as $\zeta = \Delta f.T.N$. and, T is the duration of each bit.

Equation (4) clearly shows that the received signal is first distorted by the channel, then the resulting signal is affected by the CFO, and is eventually re-distorted by the receiver IQ. To compensate for CFO and IQ, the CFO is first estimated. Then the signal itself and its conjugate once are multiplied by $e^{-j2\pi.\Delta\tilde{f}.t}$ and the Equations (6) and (7) are obtained.

$$y_1 = y.e^{-j2\pi\Delta\tilde{f}.t} \tag{6}$$

$$y_2 = (y)^*.e^{-j2\pi\Delta\tilde{f}.t} \tag{7}$$

Straightforwardly, we can derive symbol estimation using a suitable linear combination of two inputs in the frequency domain including $Y_1[l]$ and $Y_2^*[l_m]$ as Equation (8).

$$\tilde{S}[l] = [W_a[l] \quad W_b[l]]\begin{bmatrix} Y_1[l].e^{-j2\pi\Delta\tilde{f}.t} \\ Y_2^*[l_m].e^{-j2\pi\Delta\tilde{f}.t}\end{bmatrix} \tag{8}$$

Equation (8) shows a two-tap FEQ equalizer that with any adaptive algorithm, the coefficients can be learned. The coefficients $W_a[l]$ and $W_b[l]$ can be calculated using the MSE criterion, which is expressed by Equation (9).

$$\min_{W_a[l],W_b[l]} \Xi\left\{\left|\tilde{S}[l] - \begin{bmatrix} W_a[l] \\ W_b[l]\end{bmatrix}^T \begin{bmatrix} Y_1[l].e^{-j2\pi\Delta\tilde{f}.t} \\ Y_2^*[l_m].e^{-j2\pi\Delta\tilde{f}.t}\end{bmatrix}\right|^2\right\} \tag{9}$$

In the above equations, (.)$_m$ represents the mirror operator and is expressed as $Y_m[l] = Y[l_m]$ where:

$$[l_m] = \begin{cases} [N-l+2], for[l] = 2,\ldots,N \\ [l], for[l] = 1 \end{cases} \tag{10}$$

2. 2. Wavelet Packet Transform

Wavelets are transform methods that have received a great deal of attention over the past decades. The wavelet transform is a time-scale representation by which signals are broken down into time and scale functions in terms of basic functions. Wavelet transform has been used extensively for various applications including feature extraction, detection, data compression, signal denoising, etc. [18].

The basis of the wavelet transform is based on two basic functions shown in Equation (11).

$$\begin{aligned} \varphi_{j,k}(t) &= 2^{\frac{j}{2}}\varphi(2^j t - k) \\ \psi_{j,k}(t) &= 2^{\frac{j}{2}}\psi(2^j t - k) \end{aligned} \tag{11}$$

In which they are called the mother wavelet function and the basic scale function, respectively [18].

Using these two basic functions, any arbitrary signal can be written as Equation (12).

$$f(t) = \sum_{-\infty}^{+\infty} c_k\varphi(t-k) + \sum_{k=-\infty}^{+\infty}\sum_{j=0}^{+\infty} d_{j,k}\psi(2^j t - k) \tag{12}$$

Most of the results of wavelet theory are developed using filter banks. The full wavelet packet decomposition in two scales is shown in Figure 1.

3. PROPOSED ALGORITHM

In this section, the proposed algorithm for short CP length will be provided. Reducing CP length improves channel utilization. Short CP lengths in multiple channels will

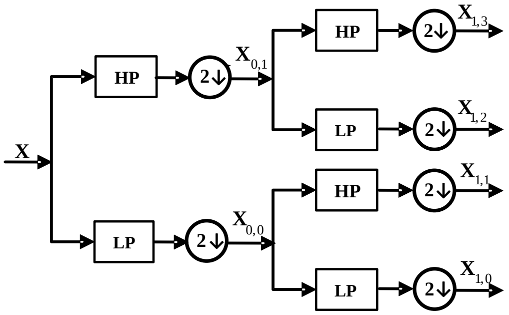


Figure 1. Wavelet packet decomposition in two scales

result in inter-symbol interference. In this paper, to increase the efficiency of channel use, inter-symbol interference is accepted and then using the PTEQ structure in the frequency domain, and simultaneously with an effective estimation of the channel, the inter-symbol interference will be eliminated. The estimation and compensation of the channel effect assuming a short CP length cannot be modeled based on of the equations described in the previous section. The PTEQ compensation structure to overcome the ISI is originally derived from the integration of a time-domain equalizer (TEQ) and a frequency-domain equalizer (FEQ) [17]. The PTEQ structure, which can reduce the effective channel length and compensate for the channel effect in the frequency domain, by making adjustments commensurate with that used in this paper, is illustrated in Figure 2.

A PTEQ is a unified compensation structure, where equalization is performed individually on each subcarrier after taking the DFT of the received signal.

In the PTEQ structure, a multi-tap filter is used for each subcarrier as an equalizer. Using this mechanism, the estimation of data in each subcarrier is based on the optimal design of the coefficients of these filters, showing Equations (13) to (15) of the corresponding equations.

$$\tilde{S}^{(i)}[l] = W^{(i)}[l]\left(F_{ext}^{(i)}[l]z\right) \qquad (13)$$

where the matrix of $F_{ext}[l]$ is defined as Equation (14) and $L' = L^{"} + L_r - 1$. Where, L_r and $L^{"}$ is considered as the length of the receiver imbalance and PTEQ structure branch filters, respectively.

$$F_{ext}[l] = \begin{bmatrix} I_{L^{'}-1} & 0_{L^{'}-1\times N-L^{'}+1} & -I_{L^{'}-1} \\ & 0_{1\times L^{'}-1} & F_N[l] \end{bmatrix} \qquad (14)$$

In this matrix, the first row represents the difference, and the second row the DFT matrix. And again, the MSE criterion according to Equation (15) is used to obtain the PTEQ coefficients.

$$\min_{W_a[l],W_b[l]} \Xi\left\{\left|\tilde{S}[l] - \begin{bmatrix} W_a[l] \\ W_b[l] \end{bmatrix}^T \begin{bmatrix} F_{ext}^{(i)}[l]\,y.\,e^{-j2\pi\Delta\tilde{f}.t} \\ F_{ext}^{(i)}[l]\,y^{*}.\,e^{-j2\pi\Delta\tilde{f}.t} \end{bmatrix}\right|^2\right\} \qquad (15)$$

Equation (15) shows the compensator criteria in the two-branch structure of PTEQ, in which the coefficients can be learned by any adaptive algorithm.

The speed and accuracy of channel estimation are essential. Therefore, before the development of adaptive filters, one of the orthogonal transforms in the proposed structure will be applied. Figure 3 shows the block diagram of the receiver section of the proposed SMF-WP-NLMS-PTEQ algorithm, which is described below.

Due to the benefits of wavelet transform and considering that the convergence speed of the LMS algorithm in this domain is faster than the LMS algorithm applied in the time domain, discrete cosine and Fourier transform domain [19], the wavelet transform in the proposed structure will be used. It has been shown previously in reference [20] and the system identification scenario that the use of adaptive filtering in orthogonal transform domain such as wavelet and DCT can increase convergence speed, and reduce steady-state error. This improvement is due to the decrease in the self-correlation of the input signals. Therefore, to increase the speed of the algorithm in channel estimation, and using the wavelet transform denoising property, in the proposed algorithm, the wavelet transform is first applied to the data behind the branches in each subcarrier. The appropriate adaptive algorithm is then extended to each branch of the PTEQ structure.

After applying the wavelet transform on the PTEQ branches and the data $F_{ext}^{(i)}[l]y$, new data will be shown as $\left(F_{ext}^{(i)}[l]y()_{WP}\right)$.

Figure 3 shows that each subcarrier must be equalized using a multi-tap adaptive filter.

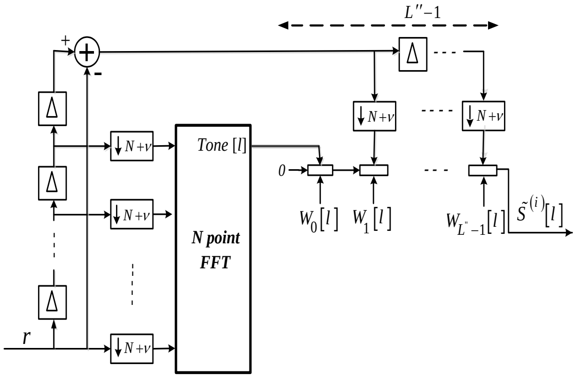


Figure 2. PTEQ-OFDM Receiver Structure

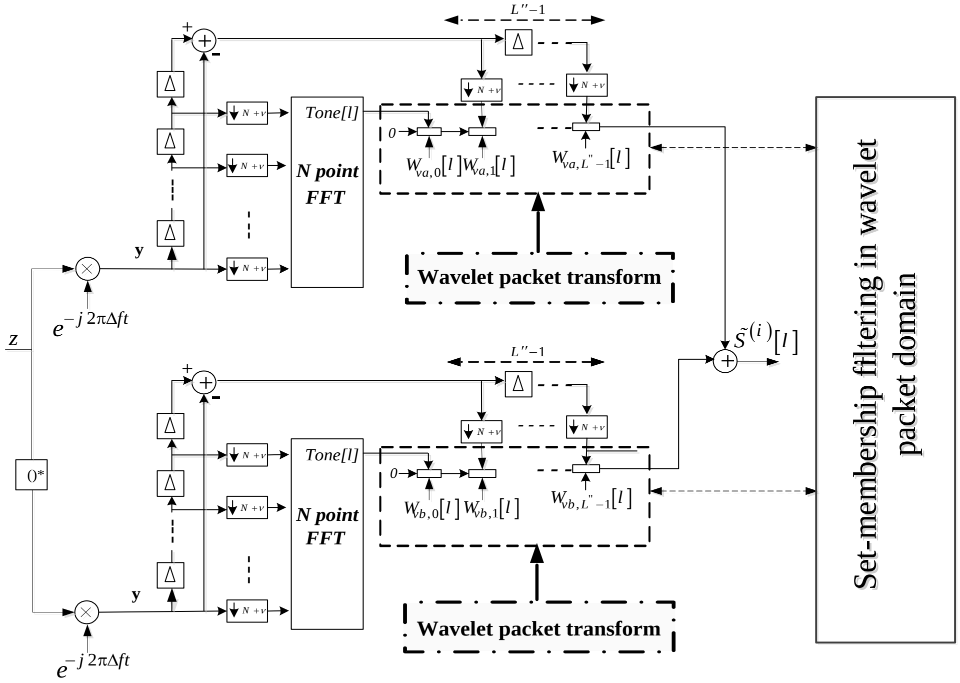


Figure 3. Receiver block diagram of proposed SMF-WP-NLMS-PTEQ algorithm

In this paper, first, the proposed WP-NLMS-PTEQ algorithm is developed for high-speed channel estimation. Then, using set-membership filtering along with wavelet packet analysis is proposed to increase the coefficient training speed and reduce the average computation rate. This algorithm is called SMF-WP-NLMS-PTEQ.

In the SMF method, when the error is less than the predetermined bound value, it is prevented from updating the coefficients and will consequently reduce the amount of computation. On the other hand, the convergence speed is also increased due to the use of the variable and noise dependent step size. The proposed algorithm will consider the estimation problem in the presence of a short CP length.

In the case of short CP, the NLMS-type algorithm is extended to the framework of the PTEQ structure for channel-effect compensation. Here, we obtain the NLMS technique in the PTEQ structure for a given error in the adaptive filter output. In a PTEQ structure, we can derive the output of the adaptive filter as $\tilde{S}^{(i)}[l]$ instant i by using Equation (17), where $W[l] = [W_0[l], W_1[l], \dots, W_{L''-1}[l]]^T$, is defined as an $L'' \times 1$ filter coefficient vector, and $F_{ext}^{(i)}[l]y$, is the $L'' \times 1$ data vector. The NLMS algorithm is computable for each branch in the PTEQ structure and is derived by using the following constrained minimization problem similar to [21]:

$$\min_{W^{(i+1)}[l]} \left\| W_a^{(i+1)}[l] + W_b^{(i+1)}[l] - W_a^{(i)}[l] - W_b^{(i)}[l] \right\|_2^2 \quad (16)$$

The limitation of the above criterion is as follows, in which $S[l]$ is a known pilot data.

$$(F_{ext}^{(i)}[l]\, y. e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_a^{(i+1)}[l] + (F_{ext}^{(i)}[l]\, y^*. e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_b^{(i+1)}[l] = S[l] \quad (17)$$

In Equation (17), $(F_{ext}^{(i)}[l]y)_{WP}$ represents the transform of $(F_{ext}^{(i)}[l]y)$ using the wavelet packet.

The above equations are solved by minimizing the Lagrange coefficients and the cost function of Equation (18):

$$J^{(i)}(l) = \left\| W_a^{(i+1)}[l] + W_b^{(i+1)}[l] - W_a^{(i)}[l] - W_b^{(i)}[l] \right\|_2^2 + \lambda(S^{(i)}[l] - (F_{ext}^{(i)}[l]\, y. e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T\, W_a^{(i+1)}[l] + (F_{ext}^{(i)}[l]\, y^*. e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_b^{(i+1)}[l]) \quad (18)$$

These equations are very complicated. So for simplicity, we assume that in any adaptive filter each branch can estimate the output after the training procedure. Therefore, the above equations can be decomposed into two independent Equations as (19) and (20) to solve the problem in simple form:

$$\min_{W_a^{(i+1)}[l]} \left\| W_a^{(i+1)}[l] - W_a^{(i)}[l] \right\|_2^2 \\ (F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_a^{(i+1)}[l] = S[l] \tag{19}$$

$$\min_{W_b^{(i+1)}[l]} \left\| W_b^{(i+1)}[l] - W_b^{(i)}[l] \right\|_2^2 \\ (F_{ext}^{(i)}[l]\, y^*.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_b^{(i+1)}[l] = S[l] \tag{20}$$

Based on the above decomposed equations and NLMS algorithm, the initial update equations to calculating the filter taps can be written in Equations (21) and (22):

$$W_a^{(i+1)}[l] = W_a^{(i)}[l] + \\ \frac{(F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T (W_a^{(i+1)}[l] - W_a^{(i)}[l])(F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}}{\left\| (F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP} \right\|_2^2} \tag{21}$$

$$W_b^{(i+1)}[l] = W_b^{(i)}[l] + \\ \frac{(F_{ext}^{(i)}[l]\, y^*.e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T (W_{vx}^{(i+1)}[l] - W_{vx}^{(i)}[l])(F_{ext}^{(i)}[l]\, y^*.e^{-j2\pi\Delta\tilde{f}.t})_{WP}}{\left\| (F_{ext}^{(i)}[l]\, y^*.e^{-j2\pi\Delta\tilde{f}.t})_{WP} \right\|_2^2} \tag{22}$$

On the other hand, we have:

$$(F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_a^{(i+1)}[l] - \\ (F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_a^{(i)}[l]) \\ = S[l] - (F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T = e^{(i)}(l) \tag{23}$$

$$(F_{ext}^{(i)}[l]\, y^*.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_{vx}^{(i+1)}[l] - \\ (F_{ext}^{(i)}[l]\, y^*.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T W_{vx}^{(i)}[l] \\ = S[l] - (F_{ext}^{(i)}[l]\, y^*.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP}^T = e^{(i)}(l) \tag{24}$$

Thus, Equations (21) and (22) can be written as (25) and (26). Also, the step size μ is inserted into the equations to control the stability and convergence rate.

$$W_a^{(i+1)}[l] = W_a^{(i)}[l] + \mu \frac{(F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP} e^{(i)}(l)}{\left\| (F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP} \right\|_2^2} \tag{25}$$

$$W_b^{(i+1)}[l] = W_b^{(i)}[l] + \mu \frac{(F_{ext}^{(i)}[l]\, y^*.e^{-j2\pi\Delta\tilde{f}.t})_{WP} e^{(i)}(l)}{\left\| (F_{ext}^{(i)}[l]\, y^*.e^{-j2\pi\Delta\tilde{f}.t})_{WP} \right\|_2^2} \tag{26}$$

In a PTEQ structure, due to the large number of branches and, as a result, the high number of filter coefficients in the branches, the amount of computation needed to train the system is high. Therefore, it is necessary to reduce the computational complexity of this system. Therefore, the SMF-based algorithm for reducing the average computation and increasing the convergence speed is presented in the following section. Also, in the NLMS algorithm, the step size μ is limited to $0 < \mu < 2$.

3. 1. Proposed SMF-WP-NLMS-PTEQ Algorithm

To obtain the SMF-WP-NLMS-PTEQ algorithm in the PTEQ structure, the step size used in Equations (25) and (26) are considered as a variable. The update equations of this algorithm are provided in Equations (27) and (28). Equation (29) also shows the relation of the variable step size.

$$W_a^{(i+1)}[l] = W_a^{(i)}[l] + \alpha^{(i)} \frac{(F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP} e^{(i)}(l)}{\left\| (F_{ext}^{(i)}[l]\, y.\, e^{-j2\pi\Delta\tilde{f}.t})_{WP} \right\|_2^2} \tag{27}$$

$$W_b^{(i+1)}[l] = W_b^{(i)}[l] + \alpha^{(i)} \frac{(F_{ext}^{(i)}[l]\, y^*.e^{-j2\pi\Delta\tilde{f}.t})_{WP} e^{(i)}(l)}{\left\| (F_{ext}^{(i)}[l]\, y^*.e^{-j2\pi\Delta\tilde{f}.t})_{WP} \right\|_2^2} \tag{28}$$

$$\alpha^{(i)} = \begin{cases} 1 - \frac{\gamma}{|e^{(i)}[l]|}, if\gamma > \left| e^{(i)}[l] \right| \\ 0, otherwise \end{cases} , \gamma = \sqrt{5\sigma_n^2} \tag{29}$$

In Equation (29), $\alpha^{(i)}$ is a variable step size andγrepresents the error threshold value which can be calculated based on of noise variance as $\gamma = \sqrt{5\sigma_n^2}$. In the SMF-NLMS algorithm, an upper threshold$\gamma$is assumed to control and constrain the estimation error [22].

In the SMF (Set-membership filtering) method, the coefficients are not updated at times when the error is less than the threshold value, thus reducing the amount of computation, significantly. Also, the algorithm uses a noise-dependent variable step size, which can increase the convergence speed.

4. SIMULATION RESULTS

In this section, the results of several simulations are shown to demonstrate the efficiency of the proposed algorithm. In the simulations, the FFT size is set to 64, the CP length is $\upsilon = 5$, and the data modulation is 64QAM and 16QAM. Multipath channel with $(L_{ch} + 1 = 18)$ taps is used, in which the taps are independently selected and have a complex Gaussian distribution. Also, the upper band of SMF-based algorithms is considered to be $\gamma = \sqrt{5\sigma_n^2}$, where σ_n^2 is the noise variance and it is assumed to be known. In curves that do not use the SMF technique, the adaptive filter step size $\mu = 0.05$ is used. In all simulations, the decomposition level of the wavelet transform packet will be 2. All results are simulated in a Rayleigh multipath channel with 18-tap length.

Figure 4 shows a plot of bit error rate (BER) versus signal-to-noise ratio (SNR) for the proposed SMF and WP-based methods for the NLMS-PTEQ algorithm in the PTEQ structure. This simulation result is derived considering both CFO and receiver frequency selective IQ. The modulation used is 16QAM and the results are compared with LS algorithm with sufficient CP and unequal conditions. The proposed SMF-NLMS-PTEQ, WP-NLMS-PTEQ, and SMF-WP-NLMS-PTEQ algorithms, in addition to the LS algorithm, show good results in terms of BER than the conventional NLMS-PTEQ algorithm. Also in the curves (c), (e), and (f) which used SMF technique, updating was done in 61, 55, and 52% of the iterations, respectively. Therefore, the average computation of the estimator system has also decreased in proportion to its update rate. The best results are obtained when both the SMF and WP techniques are combined. All of these curves are compared to the ideal 16QAM modulation on the AWGN channel. Curve (e) is performed in the absence of the CFO and only in the presence of the receiver and channel IQ and shows better results than the same algorithm (f).

Figure 5 shows a plot of BER versus SNR for the proposed SMF and WP-based methods for the NLMS-PTEQ algorithm in the PTEQ structure. This simulation result is derived considering both CFO and receiver frequency selective IQ. The modulation used is 64QAM and the results are compared with the LS algorithm with sufficient CP and unequal conditions. The proposed SMF-NLMS-PTEQ, WP-NLMS-PTEQ, and SMF-WP-NLMS-PTEQ algorithms, in addition to the LS algorithm, show good results in terms of BER than the conventional NLMS-PTEQ algorithm. Also in the curves (d), (e), (g), and (h) which used SMF technique, updating was done in 67, 63, 58, and 61% of the iterations, respectively. Therefore, the average computation of the estimator system has also decreased in proportion to its update rate. The best results are obtained when both the SMF and WP techniques are combined. All of these curves are compared to the ideal 64QAM modulation on the AWGN channel. The curves (e) and (h) are performed in the absence of the CFO and only in the presence of the receiver IQ and channel effect and show better results than the identical algorithms (d) and (h).

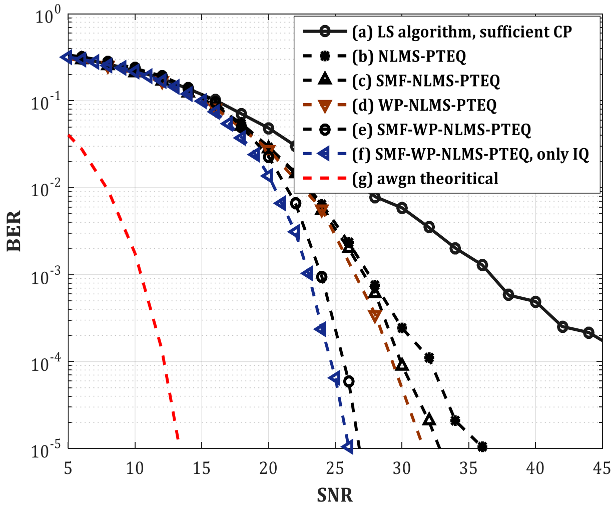


Figure 4. Error performance curve, 16QAM signal, training with NLMS and SMF-WP-NLMS-PTEQ under short CP in presence of CFO and receiver IQ

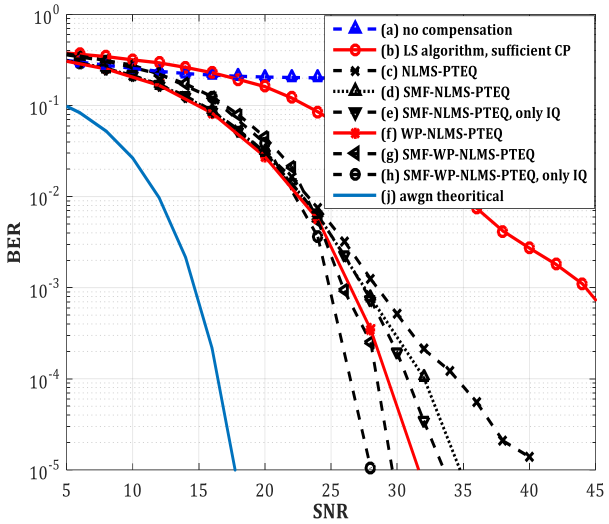


Figure 5. Error performance curve, 64QAM signal, training with NLMS and SMF-WP-NLMS-PTEQ under short CP in presence of CFO and receiver IQ

In Figure 6, the effect of increasing the value of the parameter L'' is investigated. By increasing the value of L'', the performance is improved and ISI can be eliminated. Only 400 symbols are used to train the system in simulated curves. Also in the simulations where the CFO exists, the value of the normalized CFO is taken to be $\zeta = 0.1$. Also, the CFO is first estimated in the presence of IQ imbalances using the NLS algorithm proposed in [23]. The impulse response of the mismatched filters caused by the IQ imbalance in the receiver with 2-tap length $L_r = 2$, are considered $y_{r1} = [0.80.1]$ and $y_{r2} = [0.10.8]$. The branch lengths of the PTEQ structure are also considered to be 20. In the simulated figures, in addition to the normal state of the algorithm, the curves are compared with the LS algorithm used for channel estimation and with sufficient CP length.

As shown in Figure 6, the effect of decreasing L'' is investigated. By decreasing L'', the bit error rate performance will deteriorate and the ISI can not be eliminated in the presence of a short CP. The error performance in the short CP approaches the non-compensation state, and by increasing the branch lengths of the PTEQ structure, the ISI effects of the short CP length are eliminated. In this simulation, the SMF-WP-NLMS-PTEQ algorithm is used for the curves (e) and (f). Also in curves (b), (c), and (d) the SMF-NLMS-PTEQ algorithm is used and the WP technique is not used due to

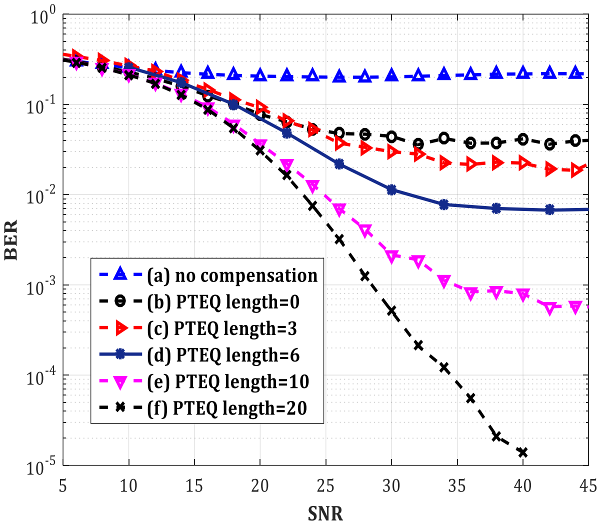


Figure 6. Error performance curve, 64QAM signal, Investigation of PTEQ branch lengths under short CP

the short length of the branches of the structure. However, the curves show that the use of PTEQ structure in short CP conditions is necessary to ISI effect elimination.

5. CONCLUDING REMARKS

In this paper, a new effective adaptive based method has been proposed for the joint estimation of the receiver IQ imbalance, CFO, and channel effects under short CP conditions. In this method, before estimating the channel, the CFO is first estimated and eliminated by existing methods. This algorithm is implemented using the structure of PTEQ. The PTEQ structure is capable of compensating channel and CFO efficiently under short CP length. Two parallel PTEQ structures are used to joint estimation of impairments. The proposed method employs an adaptive algorithm more optimally and its implementation in the wavelet transform domain has increased the convergence speed of the channel estimation. The use of PTEQ and wavelet transform along with the set-membership filtering concept, has made the proposed algorithm well capable of estimating and compensating channel effects in the presence of receiver IQ and CFO in short CP length. The average computational cost of the system has also decreased and the use of the system has become more cost-effective and implementable in real applications. Overall, the simulation results for the proposed 'SMF-WP-NLMS-PTEQ' algorithm under short CP and in the presence of receiver IQ and CFO showed sufficient improvement in BER along with reduced computation, which makes this algorithm nearly ideal in terms of its performance.

6. REFERENCES

1. Murdas, I., “Quadrature amplitude modulation all optical orthogonal frequency division multiplexing-dense wavelength division multiplexing-optical wireless communication system under different weather conditions”, ***International Journal of Engineering - Transactions A: Basics***, Vol. 30, No. 7, (2017), 988–994. doi: 10.5829/ije.2017.30.07a.08
2. Qiao, G., Babar, Z., Ma, L. and Ahmed, N., “Channel Estimation and Equalization of Underwater Acoustic MIMO-OFDM Systems: A Review Estimation du canal et l’égalisation des systèmes MEMS-MROF acoustiques sous-marins: une revue”, ***Canadian Journal of Electrical and Computer Engineering***, Vol. 42, No. 4, (2019), 199–208. doi: 10.1109/CJECE.2019.2897587
3. Xie, H., Wang, Y., Andrieux, G. and Ren, X. “Efficient compressed sensing based non-sample spaced sparse channel estimation in OFDM system”, ***IEEE Access***, Vol. 7, (2019), 133362–133370. doi: 10.1109/ACCESS.2019.2941152
4. Rodriguez-Avila, R., Nunez-Vega, G., Parra-Michel, R. and Mendez-Vazquez, A., “Frequency-selective joint Tx/Rx I/Q imbalance estimation using Golay complementary sequences”, ***IEEE Transactions on Wireless Communications***, Vol. 12, No. 5, (2013), 2171–2179. doi: 10.1109/TWC.2013.040213.120622
5. Manasseh, E., Ohno, S. and Nakamoto, M., “Training symbol design for channel estimation and IQ imbalance compensation in OFDM systems”, ***IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences***, Vol. 95, No. 11, (2012), 1963–1970. doi: 10.1109/VETECS.2012.6239897
6. Miyashita, H., Inamori, M., Sanada, Y. and Ide, T., “IQ imbalance estimation scheme with intercarrier interference self-cancellation pilot symbols in OFDM direct conversion receivers”, In 2012 IEEE 75th Vehicular Technology Conference (VTC Spring), IEEE, (2012), 1–5. doi: 10.1109/VETECS.2012.6239943
7. Nguyen, Q.D., Luu, T.T., Pham, L.H. and Nguyen, T.M., “Inter-Carrier Interference suppression combined with channel estimation for mobile OFDM system”, In 2016 Second Asian Conference on Defence Technology (ACDT), IEEE, (2016), 61–66. doi: 10.1109/ACDT.2016.7437644
8. Simon, E.P., Ros, L., Hijazi, H. and Ghogho, M. Simon, E. P., Ros, L., Hijazi, H. and Ghogho, M., “Joint carrier frequency offset and channel estimation for OFDM systems via the EM algorithm in the presence of very high mobility”, ***IEEE Transactions on Signal Processing***, Vol. 60, No. 2, (2012), 754–765. doi: 10.1109/TSP.2011.2174053
9. Krishna, E. H., Sivani, K., and Reddy, K. A., “Empirical Mode Decomposition based Adaptive Filtering for Orthogonal Frequency Division Multiplexing Channel Estimation”, ***International Journal of Engineering (IJE)***, Vol. 30, No. 10, (2017), 1517–1525. doi: 10.5829/ije.2017.30.10a.13
10. Cheng, H., Xia, Y., Huang, Y., Yang, L. and Mandic, D.P., “Joint channel estimation and Tx/Rx I/Q imbalance compensation for GFDM systems”, ***IEEE Transactions on Wireless Communications***, Vol. 18, No. 2, (2019), 1304–1317.
11. Cheng, H., Xia, Y., Huang, Y., Yang, L. and Mandic, D. P., “Joint Channel Estimation and Impulsive Noise Mitigation Method for OFDM Systems Using Sparse Bayesian Learning”, ***IEEE Access***, Vol. 7, (2019), 74500–74510. doi:

10.1109/ACCESS.2019.2920724

12. McPherson, R.K. and Schroeder, J., “Frequency-selective I/Q imbalance compensation for OFDM receivers using decision-feedback adaptive filtering”, In 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), IEEE, (2012), 188–192. doi: 10.1109/ACSSC.2012.6488986
13. Tandur, D. and Moonen, M., “Decoupled compensation of IQ imbalance in MIMO OFDM systems”, ***IEEE Transactions on Signal Processing***, Vol. 90, No. 5, (2011), 1194–1209. doi: 10.1016/j.sigpro.2010.11.008
14. Tarighat, A., Bagheri, R. and Sayed, A. H., “Compensation schemes and performance analysis of IQ imbalances in OFDM receivers”, ***IEEE Transactions on Signal Processing***, Vol. 53, No. 8, (2005), 3257–3268. doi: 10.1109/TSP.2005.851156
15. Chung, Y. H. and S. M. Phoong, S. M., “OFDM channel estimation in the presence Of transmitter and receiver i/q imbalance”, In 16th European Signal Processing Conference, IEEE., (2008), 830-834. doi: 10.1587/transcom.E95.B.531
16. Ehsanfar, S., Matth´e, M., Chafii, M. and Fettweis, G., “Pilot- and CP-aided Channel Estimation in MIMO Non-Orthogonal Multi-Carriers”, ***IEEE Transactions on Wireless Communications***, Vol. 18, No. 1, (2019), 650-664. doi: 10.1109/TWC.2018.2883940
17. Tandur, D., “Digital compensation of front-end non-idealities in broadband communication systems”, Doctoral dissertation, Ph. D. thesis. (2010)
18. Burrus, C. S., Gopinath, R. A. and Guo, H., Introduction to Wavelets and Wavelet Transforms, a Primer, Prentice Hall New Jersey, (1998).
19. Attalla. S., “The Wavelet Transform-Domain LMS Adaptive Filter With Partial Subband-Coefficient Updating”, ***IEEE Transactions on Circuits and Systems II: Express Briefs***, Vol. 53, No. 1, (2006), 8-12. doi: 10.1109/TCSII.2005.855042
20. “Feng, C., Zhang, L. and Hui, X., “A New Adaptive Filtering Algorithm Based on Discrete Wavelet Transforms”, In 3rd International Congress on Image and Signal Processing, IEEE, (2010), 3284-3286. 10.1109/CISP.2010.5647465
21. Goodwin. G. C., and Sin, K. S., Adaptive Filtering, Prediction, and Control. Englewood Cliffs, NJ: Prentice-Hall, (1984).
22. Cheffi, A., and Djendi. M., “New Two-Channel Set-Membership Partial-Update NLMS Algorithms for Acoustic Noise Reduction”, In International Conference on Signal, Image, Vision, and their Applications, (2018), 1123-1127. doi: 10.1109/SIVA.2018.8661128
23. Xing, G., Shen, M. and Liu, H., “Frequency offset and i/q imbalance compensation for direct-conversion receivers”, ***IEEE Transactions on Wireless Communications***, Vol. 4, No. 2, (2005), 673-680. doi: 10.1109/TWC.2004.842969

Persian Abstract

چکیده

در این مقاله، تخمین توام کانال، عدم توازن IQ گیرنده وابسته به فرکانس و آفست فرکانسی حامل با طول CP کوتاه در سیستم OFDM در نظر گرفته شده است. برای جبران‌سازی این خرابی‌ها، الگوریتم تطبیقی مبتنی بر ترکیب فیلترینگ set-membership مورد استفاده قرار گرفته است. در حالت CP کوتاه برای جلوگیری از تداخل بین سمبلی از ساختار PTEQ استفاده می شود. بخاطر وجود خرابی CFO و عدم توازن IQ در گیرنده، ساختار PTEQ را به حالت دو شاخه ای بسط خواهیم داد. این ساختار پیچیدگی محاسباتی بالایی داشته لذا در چنین شرایطی استفاده از ایده فیلترینگ set-membership با ضریب گام متغیر ضمن کاهش متوسط محاسبات سیستم، می تواند سرعت همگرایی تخمین ها را نیز افزایش دهد. از طرفی بهره‌گیری از تبدیل موجک بر روی هر شاخه این ساختار قبل از بکارگیری فیلترهای تطبیقی، سرعت تخمین را به نوبه ی خود افزایش خواهد داد. الگوریتم پیشنهادی با نام SMF-WP-NLMS-PTEQ ارائه خواهد شد. نتایج شبیه‌سازی‌ها نشان دهنده عملکرد بهتر از الگوریتم‌های تطبیقی معمول بوده است. علاوه بر این تخمین و جبران‌سازی اثرات کانال، عدم توازن IQ گیرنده و آفست فرکانسی تحت CP کوتاه به‌سهولت بوسیله این الگوریتم قابل انجام می‌باشد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Fast Unsupervised Automobile Insurance Fraud Detection Based on Spectral Ranking of Anomalies

Z. Shaeiri[1], S. J. Kazemitabar*[2]

[1] *Son Corporate Group, Tehran, Iran*
[2] *Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran*

PAPER INFO

Paper history:
Received 10 November 2019
Received in revised form 23 December 2019
Accepted 04 January 2020

Keywords:
Spectral Ranking of Anomalies
Auto-insurance Fraud Detection
Random Forest
Deep Learning

ABSTRACT

Collecting insurance fraud samples is costly and if performed manually is very time consuming. This issue suggests the usage of unsupervised models for fraud data collection. One of the accurate methods in this regards is Spectral Ranking of Anomalies (SRA) that is shown to work better than other methods for auto-insurance fraud detection, specifically. However, this approach is not scalable to large samples and is not appropriate for online fraud detection. This is while, real-time fraud management systems are necessary to prevent huge losses. In this study, we propose an implementation methodology which makes it possible to apply the SRA to big data senarios. We exploit the power of spectral ranking of anomalies to create an estimated target variable from the unlabeled dataset. We then use two robust models, namely, random forest and deep neural networks to fit a model based on the estimated labeled training set. Next, the incoming live data are fed to the mentioned trained models for predicting the target variable. Simulation results confirm that the proposed approach has higher speed and acceptable false alarm rate compared to existing related methods.

doi*: 10.5829/ije.2020.33.07a.10*

NOMENCLATURE

$k(s_1, s_2)$	Kernel function	$\delta(s_1, s_2)$	Delta Kronickel function
σ	Kernel width parameter	f	Ranking vector
W	Adjacency matrix	L	Laplacian matrix

1. INTRODUCTION

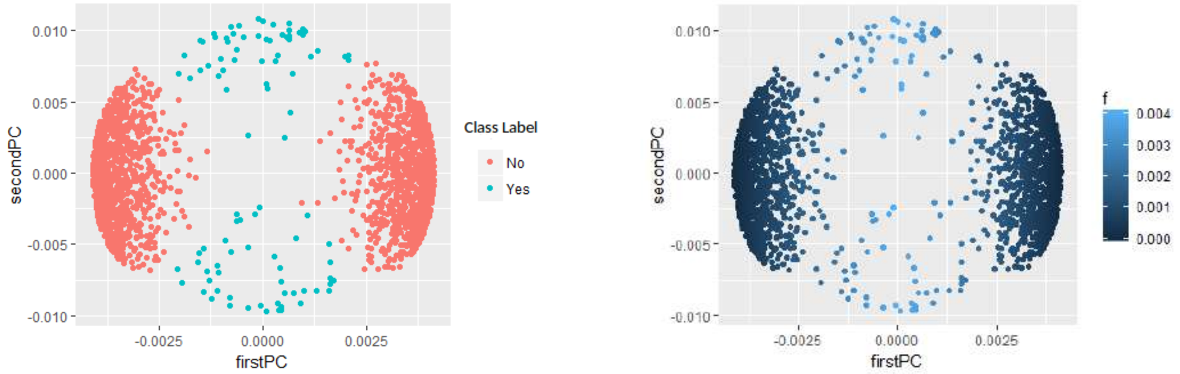
A major concern among insurance companies is the issue of claims fraud. According to ABI (Association of British Insurers) statistics, in 2014, 3.6 million pounds was lost due to claims fraud on a daily basis. According to Tennyson and Salsas-Forn-2002 [1] approximately 21% to 36% of auto-insurance claims are fraudulent but only less than 3% of suspected cases are detected and persecuted. Besides putting the advantages of the insurer at risk, fraud is also harmful for its value chain. It increases the cost of insurance which will be tolerated by the insured parties in the form of increased premium rates. Thus, fraud is detrimental to the very basic dependencies that keep the concept of insurance alive. Fraud detection is essential to prevent huge losses which insurance systems are encountered with. Traditionally, insurance fraud detection was performed by insurance investigators and claim adjusters. Since manually detecting suspicious claims from huge insurance fraud datasets is inefficient, data mining and machine learning methods are used extensively. These methods help the insurers detect fraud prior to reimbursing the customer. This is an essential requirement that machine learning and data mining approaches can handle appropriately. In the literature, auto-insurance fraud detection was most often performed via some supervised models and scarcely with unsupervised methods. Supervised models need to have access to the claims of both classes; fraudulent and legitimate. However, two issues exist that make supervised models difficult to use. First, there are not many labeled data samples available to work with.

*Corresponding Author Institutional Email: *j.kazemitabar@nit.ac.ir* (S. J. Kazemitabar)

Moreover, since the fraud datasets are imbalanced most supervised methods cannot perform very well. More precisely, their classification performance is not equal across both fraud and non-fraud samples.

In [2], the authors show the performance of binary classifiers (logistic) for auto-insurance fraud detection and obtain models for demonstrating misclassification in the target variable which is caused by auditors' mistakes. Performance of various classification techniques for auto-insurance fraud detection is taken into account in [3]. Performances of logistic, k-nearest neighbor, Bayesian learning, multilayer perceptron neural network, support vector machine, naive Bayes, and tree-augmented naive Bayes classification were compared and contrasted in that paper. In [11], the most relevant attributes are chosen from the original dataset by using an evolutionary algorithm based feature selection method. A test set is then extracted from the selected attribute set and the remaining dataset is subjected to the Possibilistic Fuzzy C-Means (PFCM) clustering technique for the under-sampling approach. In [12], suspicious groups have been detected by applying cycle detection algorithms (using both DFS, BFS trees). Afterwards, the probability of being fraudulent for suspicious components was investigated to reveal fraudulent groups with maximum likelihood, and their reviews were prioritized. In [13], a multiple classifier system based on Random Forest, Principle Component Analysis and Potential Nearest Neighbor is proposed. The authors ameliorate the classification accuracy of the ensemble classifier by improving the difference of the base classifiers. The proposed method is then applied to detect automobile insurance fraud and the fraud rules are obtained. Proposing supervised learning models seems inefficient as collecting target variables for auto-insurance fraud datasets -like other insurance fraud datasets- is very costly and time consuming. Moreover, the investigators act significantly different from one another in their fraud diagnosis. In addition, risk assessment is more appealing to insurers than simply accessing binary fraud/non-fraud classification of claims. Considering these facts, unsupervised methods and specifically unsupervised anomaly ranking seems more appropriate for fraud detection problems. In the literature, few unsupervised auto-insurance fraud detection models are proposed. One of the earlier unsupervised approaches in this context is proposed by Brockett et al., in 1998 [4]. In this paper, the authors propose to apply a self-organizing neural network for classification of fraud datasets. In another work, Brockett et al., proposed the PRIDIT methodology (Principal component analysis and RIDIT scoring method) for auto-insurance fraud detection [5]. In these works, the authors have examined their studies over the PIP (Personal Injury Protection) dataset which is provided by AIB (Automobile Insurance Bureau). It is worth noting that in this dataset suspicion of fraud among samples is ranked somehow by the auditors and experts. Thus, it is fair if one considers these works as semi-supervised methods. Fraud datasets often contain categorical and ordinal variables. These variables require pre-processing prior to being used, a process which itself requires expert knowledgedataset. Another point that needs to be emphasized is that since we are attempting to classify or rank the fraud dataset it seems that the dataset should have one major class of normal samples and one small pattern of anomalies.

However, this will not be the case when the dataset contains many categorical variables. Therefore, methods that rely on the implicit assumption that the dataset contains one major pattern may not be successful. In another work [6], the authors proposed an unsupervised method which is based on spectral ranking of anomalies (SRA). Their work is motivated by the observation that there is a connection between unsupervised Support Vector Machine (SVM) optimization formulation and the spectral optimization. They derived a ranking vector



(a) Visualization of data instances (this dataset is a 2-Dimensional synthetic dataset

(b) The second non-principal Eigen-vector of the Laplacian versus the first one

Figure 1. Visualizing the information contained in the synthetic data

which provides the degree of relative abnormality of samples in the dataset. To the best of our knowledge, among different methods in the literature, SRA provides significant results in ranking anomalies for auto-insurance fraud datasets [15-22]. This observation motivates us to focus more on the SRA and to try to implement and apply it on big datasets. As shown by the authors in [6], the performance of this unsupervised method is very close to the supervised techniques such as SVM which serves as an upper bound for unsupervised methods. While it is considerably accurate and effective in auto-insurance fraud detection, this method suffers from high computational complexity and cannot be applied for online fraud detection. In this paper, we propose an online unsupervised methodology for auto-insurance fraud detection. Here, we want to explain the important drawback of the theoretically accurate SRA method, which makes it impossible to apply to big datasets. Implementation of SRA involves large scale matrix multiplication and Eigen decomposition. For real world datasets the similarity/distance matrix is a huge dense square matrix which is always too large to fit in memory. This makes it impossible to implement SRA on big datasets.And even if it were computationally feasible to do, so, its memory requirements are unusually high. To the best of our knowledge, today a handful of technologies provide a solution for applying linear algebra operations on large dense square matrices which also encompass high memory requirements. The main contribution of this paper is thus to facilitate the implementation of SRA for big datasets with low memory resources. First, we transform the unsupervised problem to a supervised one which means to provide labels for a fraction of the data. We apply SRA on this fraction and obtain the anomaly ranking vector. Then, we use this vector as a guide for estimating the target variable for the mentioned small set of samples. The SRA method is very accurate in ranking the data points. We exploit the ranking vector derived from SRA and apply a threshold on it for labeling the points as normal and fraudulent. The concept of SRA is such that it is less affected by imbalanced data. The details of how SRA handles this issue can be found in [6]. In short, it defines a parameter which is the ratio of the two labels and uses that for applying a threshold. Fraud detection is an interactive task which means that expert knowledge is exploited in various steps of the design and implementation. For example, the mentioned threshold value is given by the experts and is fixed.

The organization of the paper is as follows. In Section 2 background of the spectral anomaly ranking method is provided. An overview of the similarity measures used in auto-insurance fraud detection is provided in Section 3. The distance measure and the kernel function that are used in this paper are discussed in subsections 3.1 and 3.2 respectively. In Section 4 the proposed methodology is introduced. Finally, simulation results are provided in Section 5.

2. BACKGROUND OF SPECTRAL RANKING OF ANOMALIES

In [6], an unsupervised method is proposed which uses spectral ranking of anomalies for fraud detection. Motivated by the analogy between unsupervised SVM optimization and the spectral optimization formulation, the authors indicated that spectral optimization can be treated as a relaxation of the unsupervised SVM optimization.They derived a similarity matrix using Hamming distance measure which is appropriate for datasets consisting of categorical and ordinal variables. They demonstrated that the absolute value of the first non-principal Eigen-vector of Laplacian of this similarity matrix provides a measure of anomaly ranking in a bi-class clustering problem. Magnitudes of entries of this non-principal Eigen-vector contain valuable information about the degree by which the corresponding samples (samples in the same positions) are anomalous. An observation is more likely to be an anomaly if the magnitude of its corresponding entry in the non-principal Eigen-vector is larger. It is worth noting that based on the structure of the underlying dataset some possible scenarios may arise. In the SRA a choice of reference is allowed in the anomaly detection process, such that the mass of the minority cluster determines how to generate the ranking. For example, in one of the probable scenarios, the minority cluster does not have a sufficient mass. In this case the anomaly likelihood can be assessed with respect to a single majority class. In another scenario when the minority cluster has sufficient mass, anomaly can be assessed with two main clusters. We refer the readers to [6] for more details about these different scenarios. More details of this method is presented in Figure 1.

2. 1. Overview of Similarity Measures In Auto-insurance Fraud Detection

Traditionally, binary predictor variables were being used in the problem of auto-insurance fraud detection. Examples of such binary variables are coverage (third party liability equals 1 and extended coverage equals 0), deductible (existence of a deductible equals 1, otherwise equals 0), witness (existence of witness equals 1, otherwise equals 0), and so on. However, many of the important categorical predictor variables in the fraud detection problem have more than two categories, e.g., age of the driver. Age can be expressed as a categorical variable with for example 5 number of ordered categories. Some works use natural integer scoring for these variables [5]. In natural integer scoring, one simply assigns for instance the numbers 1, 2, 3, 4, and 5 to the five different categories of a variable.

It is worth noting that this kind of variable transformation can impose unwanted scaling and order, and unintended distribution to the original predictor variable which finally will result in weak and incorrect outcomes. The fact is that many of the predictor variables in a real world dataset are categorical or ordinal. Most of the machine learning approaches are sensitive to the above mentioned pre-processing steps which are applied on the datasets. These pre-processing steps sometimes dramatically change the predictor variables meanings, sometimes cause information loss, and ultimately result in unreliable outcomes. Historically, working on similarity measures dates back to the past century [7]. One of the seemingly appropriate methods in the current problem is similarity measures based on match and mismatch of the nominal values of the variables. This similarity measure is known as nominal value definition derived in [6]. While it is simple, this similarity measure preserves the meaning of the predictor variables. This way, one does not require the pre-processing step for transforming the predictor variables. In this paper, we use this similarity measure for obtaining a dissimilarity matrix.

2. 2. Distance Measure A distance measure is a real-valued function which shows the extent of dissimilarity between two samples in the data space. For each pair of N-dimensional samples $(\boldsymbol{s}_1,\boldsymbol{s}_2)$ Hamming distance is defined as the number of mismatch between the variables divided by total number of variables:

$$d^H(\boldsymbol{s}_1,\boldsymbol{s}_2) = \frac{\sum_{i=1}^{N} \delta(\boldsymbol{s}_1^i,\boldsymbol{s}_2^i)}{\boldsymbol{n}} \tag{1}$$

In this equation $\boldsymbol{s}_1^i$ and $\boldsymbol{s}_2^i$ are the ith element of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ respectively and $\delta(\boldsymbol{s}_1^i,\boldsymbol{s}_2^i)$ is defined as:

$$\delta(\boldsymbol{s}_1^i,\boldsymbol{s}_2^i) = \begin{cases} 1, & \boldsymbol{s}_1^i \neq \boldsymbol{s}_2^i \\ 0, & \boldsymbol{s}_1^i = \boldsymbol{s}_2^i \end{cases} \tag{2}$$

2. 3. Kernel Function To handle complicated relationships among attributes, it is common to transform the data into a usually high dimensional feature space via various kinds of kernel methods [8]. Datasets that are produced based on human activities and behaviors, e.g. insurance claim datasets, naturally contain categorical and ordinal features. As stated above, it is shown that a meaningful treatment for capturing relationships among different features in the datasets containing categorical features is exploiting an appropriate similarity measure. In [9], Couto introduced the Hamming distance kernel for datasets containing categorical features. For each pair of data instances this similarity measure is achieved by considering match and mismatch between the categorical features. In this paper, we use the Gaussian kernel derived from the Hamming distance which is obtained by replacing the Euclidean distance in the standard Gaussian kernel with the Hamming distance. More precisely, the kernel function

$$k(\boldsymbol{s}_1,\boldsymbol{s}_2) = \exp\left(-\frac{d^{\{H\}}(\boldsymbol{s}_1,\boldsymbol{s}_2)}{2\sigma^2}\right) \tag{3}$$

is considered in which $\sigma > 0$ is a constant kernel width parameter.

3. OVERVIEW OF THE PROPOSED METHOD

Spectral clustering is one of the recent clustering techniques that proved its superiority among traditional clustering methods. The spectral ranking of anomalies was applied for fraud detection on insurance claim datasets. It shows considerable improvements in anomaly detection compared with other unsupervised methods such as one-class support vector machine (OC-SVM) and local outlier factor (LOF) when applied on automobile fraud detection dataset as well as various kinds of synthetic datasets [6]. However, SRA is not applicable and appropriate for handling big datasets. The structure of this algorithm makes it difficult to work on big datasets. The SRA requires using all the samples from the beginning. There is no mechanism to exploit it for live data. Like many other precise anomaly detection methods, SRA requires the formation of the similarity/distance matrix. As one knows, this matrix is a dense square matrix of order *M* (the number of records or samples in the dataset). Practically, creating this matrix for real world datasets is resource intensive. When the dense square matrix is big, one cannot possibly afford storing it in a dense way. It probably would not even fit into the memory. On the other hand, implementation of SRA involves large scale matrix multiplication and Eigen decomposition. First, we should find an appropriate way to create the required matrices. It is only at that stage that we can hope to proceed through the rest of the stages. More precisely, just then we may implement the matrix multiplications and the Eigen decomposition steps of the algorithm. We provide a methodology for facilitating SRA for handling big datasets with low memory resources and high performance as well as acceptable false alarm rate. We first use a fraction of the dataset to be processed with the spectral ranking method. Output of the spectral ranking is a ranking vector denoted here as f. Using this ranking vector and a threshold value, the portion of the data with higher risk is labeled. In the SRA, the ratio of the number of fraud cases to the number of normal records is used as the threshold value. Generally, the threshold is chosen experimentally by referring to the domain experts. We have used the same threshold value that is exploited in the SRA method. The result is a two-class labeled dataset in which the samples are tagged as normal and fraudulent. We then use the tagged data and fit a model via supervised learning methods such as random forest or deep learning. This model can then be used to process any incoming data to determine whether it is fraudulent or not.

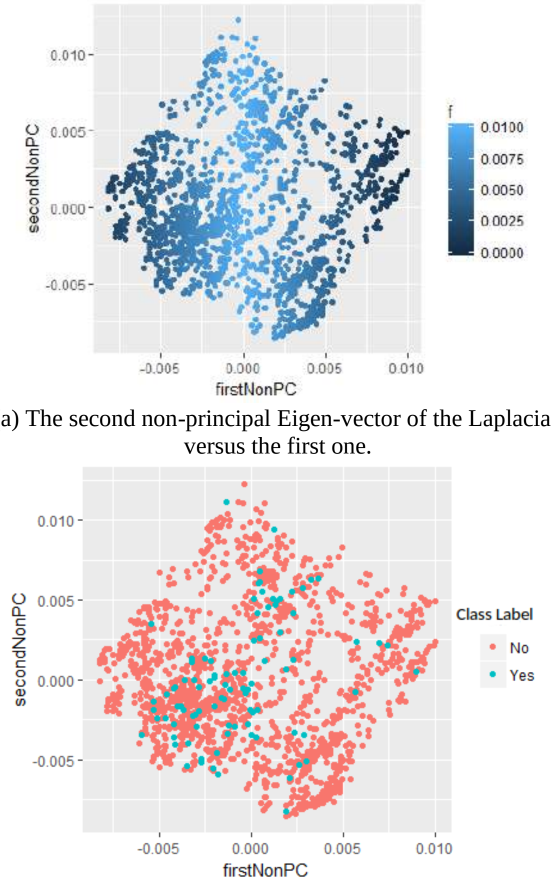
3. 1. Details Let D be the data space consisting of data instances $\boldsymbol{s}_i$ with $i = 1,\ldots,M$ where M is the number of data points in D. The main objective is to cluster the dataset into a number of clusters such that distances are minimized within each cluster. This similarity based clustering can be viewed as a graph cut problem in which each data point is a vertex and each pair of vertices are connected together by an edge with a weight equal to the similarity of the corresponding data points. An adjacency matrix W is derived which summarizes the similarity between each pair of data instances in D. The distance measure which we have used in this paper can handle nominal and ordinal variables as well as numerical variables. The handling of nominal and ordinal variables is achieved by using the general dissimilarity coefficient of Gower [10] in which match and mismatch of the variables entries are considered for deriving the distance measure. The Euclidean distance is used for numeric variables. The numeric variables are rescaled before applying the Euclidean distance. Each rescaled (numeric) variable has range [0,1] exactly. The mentioned distance measure is exerted into the standard Gaussian kernel to obtain the adjacency matrix W. More precisely:

$$W = \exp\left(-\frac{R}{2\sigma^2}\right), \tag{4}$$

where R is the distance matrix with its (i,j)th entry R_{ij} being the distance between data points i and j. Next, we compute the degree matrix D of the vertices which is a diagonal matrix with diagonal entries $d_i = P_j W_{ij}$. From the adjacency matrix W and the degree matrix D the Laplacian matrix L is derived which is a fundamental quantity in the spectral anomaly ranking part of our methodology. We use the following definition of the Laplacian matrix in our analysis:

$$L = I - D^{-0.5} W D^{-0.5}, \tag{5}$$

in which I is the identity matrix. SRA introduces a technique for ordering the data instances based on their anomalous behavior. The main idea of this method is that entries of a non-principal Eigen-vector of the Laplacian matrix provide valuable information for anomaly detection. Let $\lambda_0 < \ldots < \lambda_{M-1}$ be the M Eigen-values of Laplacian matrix L. Associated with each Eigen-value λ_i there is an Eigen-vector v_i . Based on the first non-principal Eigen-vectors of L a ranking vector is derived which provides meaningful distinguishability among normal data instances and abnormal ones. Figure 1 provides an illustrative example of the spectral ranking method in which the distinguishability or clustering strength of the first non-principal Eigen-vector of L is demonstrated for a balanced two-class dataset. The first and the second non-principal Eigen-vectors of the Laplacian, corresponding to the Gaussian kernel are depicted. In Figure 1.a, true output class labels are specified by the color, where red points indicate the normal data instances, while blue points show anomaly cases. Each point in the 2-dimensional Eigen-space corresponds to one point in the original data space. To each point in the Eigen-vectors space a value is assigned, the magnitude of which shows the level of abnormality of the corresponding point. In Figure 1.b, points with larger f (ligther color) are associated with data instances that are more abnormal. Figure 2 shows similar results for the automobile fraud dataset. As can be inferred from this figure, vector f provides a good measure for anomaly since it properly distinguishes the normal data points from the abnormal ones. Despite its precision in ranking the data instances, SRA is a huge resource consumer. This is mainly due to the matrix multiplication operation and the Singular Value Decomposition (SVD) calculation in this method. Thus, SRA is impractical for large datasets and it cannot be applied for online anomaly/fraud detection. In this study, we propose a methodology in which the power of SRA -its accuracy- is exploited in designing an online auto-insurance fraud detection technique. Based on the ranking vector (f) derived from the SRA and by applying a wisely chosen



(a) The second non-principal Eigen-vector of the Laplacian versus the first one.

(b) Visualization of data instances (The auto-insurance fraud dataset)

Figure 2. Visualizing the information contained in the auto-insurance dataset

threshold upon it we estimate the target variables/labels for a small fraction of the dataset. This fraction is selected using random sampling technique. Random sampling is as fair and unbiased as possible since it makes units equally likely to be chosen. It ensures independent selection by gathering as much independent information as possible. Thus, the sample is fair and representative. In the SRA the threshold is chosen approximately based on the prior knowledge, i.e., approximate percentage of normal and anomaly cases. Generally, one can approximately achieve an acceptable threshold by referring to the domain experts. We treat the labeled dataset as a training set. In other words, relying on correctness of the ranking vector f, we transform the unsupervised problem to a supervised one. Data points with f greater than the threshold value are treated as anomaly and the remaining as normal cases. Next, two powerful supervised methods, namely the Random Forrest (RF) and the Deep Learning (DL) classification models are trained using the generated training set and their speeds and accuracies are analyzed. The results show that while the proposed method is simple and straightforward, it is considerably accurate as well as fast.

3. 2. Remarks For computing the similarity/distance matrix, we used the R “daisy” function from “cluster” package which computes all the pairwise dissimilarities between observations in the dataset. This function can handle datasets with mixed type variables (nominal, ordinal and numeric). Table 1 shows the amount of the allocated memory for different sample sizes.

It shows that we face serious memory issues for large sample sizes. The main contribution of this paper is to facilitate implementation of SRA for big datasets with low memory resources. We used random sampling technique for the selection of the small dataset. Results show significant robustness with standard deviation of percentage of accuracy 0.35 for 10 runs. We exchanged the accuracy by speed.

4. SIMULATION RESULTS

Several experiments are conducted on an auto-insurance claim dataset as well as two synthetic datasets. Description of the datasets are given in Table 1. The synthetic datasets contain two normal clusters and anomaly points. Each of the normal clusters consists of 2000 Gaussian distributed data instances. The synthetic dataset 1 contains 200 uniformly distributed point anomalies and synthetic dataset 2 contains two small Gaussian distributed anomaly clusters together with 200 uniformly distributed point anomalies. The auto-insurance claim dataset is collected by Angoss KnowledgeSeeker software from January 1994 to December 1996. This dataset contains 15420 observations where each claim is assigned a label indicating if that claim is a normal case or it is an anomaly. Totally, it contains 14497 normal and 923 fraudulent cases. To the best of our knowledge this dataset is the only labeled auto-insurance fraud dataset available in the academic literature. The predictor variables contained in this dataset are categorical and ordinal including base policy, day of week, number of cars, witness present, and the past number of claims. First, to obtain the target variable, spectral ranking of anomalies is performed by running the SRA on the unlabeled training set. For categorical features, the Hamming distance and for numerical features the Euclidean distance is used as the dissimilarity measure. In the SRA based estimation of the target variable the Gaussian similarity kernel is considered. Using a wisely chosen threshold, we labeled the training dataset by the derived target variable. The threshold is chosen by referring to the prior knowledge and the domain experts. In the training phase, the mentioned labeled dataset which is the output of the previous stage, is used to train the RF and DL models. In the RF model the number of trees are set to 50. The algorithm converges when the 2-tree average is within 0.001 of the prior two 2-tree averages. The DL model is a 5-layer neural network. We used deep learning implementation in R language's H2O package. Hyperbolic tangent is used as the activation function and the number of epochs are set to 1000. First, we will create two independent splits for train (80 percent) and test (20 percent) sets. The models are trained on the train set. The test set is used for ensuring that the model can predict properly on the new datasets. The results are compared with the original SRA model. We have used the ROC curves to compare our results with the SRA. Tracing the ROC curves is a commonly used way for visualizing the performance of binary classifiers. The ROC curves show the trade-off between true-positive and false-positive quantities for different choices of threshold. Therefore, it does not depend on prior knowledge to combine true-positive and false-positive quantities into a unique object. From the ROC plot, one can distinguish the dominant algorithm, i.e. the one that provides a better solution at any false-positive value. We have also calculated the Area Under Curve (AUC) for these methods as another measure of the quality. Simply speaking, AUC summarizes the performance of a binary classifier in a single number. Table 3 contains the

TABLE 1. Memory usage of creating the distance matria

# of records of the sample data	Memory allocated
1542	185 MB
3084	731 MB
4626	1.64 GB
6168	2.9 GB

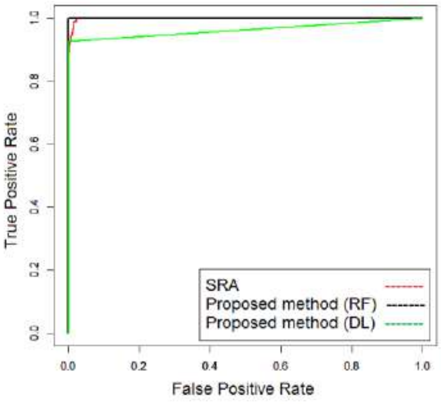
execution time of the methods for the datasets. For the auto-insurance fraud dataset the execution time for the proposed method does not include the time it takes to create the labeled training set. The fact is that by increasing the volume of dataset the gap between the execution time of the proposed method and the SRA will increase dramatically. It can be seen that the execution time of the proposed methodology is significantly less than the execution time of the original SRA method. Figures 3 and 4 show the ROC curves of the supervised SRA and our methodology on the insurance claim dataset and on the two synthetic datasets. From Tables 2 and 3 and Figure 4 it can be inferred that while comparable with the SRA in terms of the accuracy the proposed methodology has higher speed. Table 4 contains comparisons between SRA and two non-spectral ranking methods such as LOF and OC-SVM.

TABLE 2. Description of the datasets

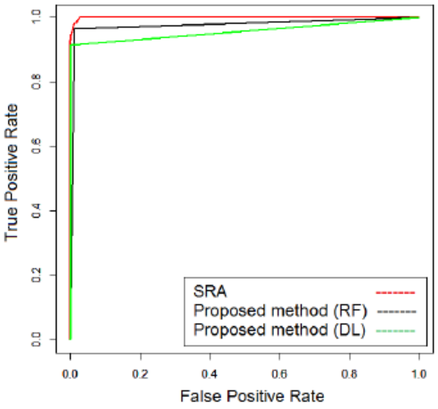
Dataset	# of Normal	# of Anomaly	Description
Synthetic data 1	2000	200	The dataset consists of 2 large normal clusters and 200 point anomalies.
Synthetic data 2	2000	287	The dataset consists of 2 large normal clusters together with 200 point anomalies and 2 small clusters.
Insurance data [1]	14497	923	The dataset is provided by Angoss KnowledgeSeeker software. It consists of 31 categorical features.

TABLE 3. Execution time (s)

Dataset	SRA	Proposed method (RF)	Propose method (DL)
Synthetic data 1	18.43	3.08	2.84
Synthetic data 2	18.86	2.22	3.14
Insurance data [1]	17.31	2.45	3.91



(a) Synthetic dataset 1



(b) *Synthetic dataset 2*

Figure 3. ROC Curves of methods for two synthetic datasets

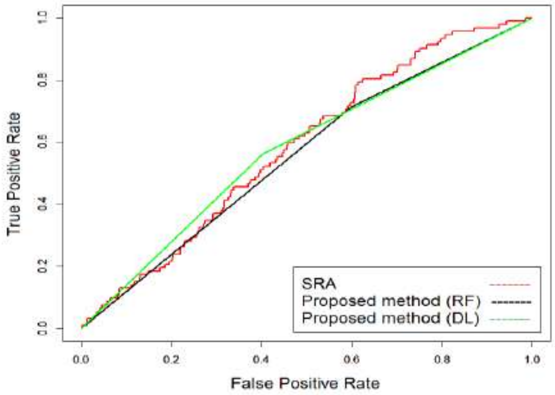


Figure 4. ROC curves of methods for insurance fraud dataset; AUC of SRA, RF, and DL are respectively, 0.6, 0.52, and 0.57

5. CONCLUSION

In this paper, a fast online unsupervised methodology was proposed for the challenging auto-insurance fraud detection problem. The proposition was based on the spectral ranking of anomalies [6]. This method is one of the recently published unsupervised anomaly detection methods which is used for anomaly detection for auto-insurance claim dataset. Despite its high precision among previous unsupervised methods, for very large datasets, a number of large matrix multiplication and Eigen value decomposition stages make this method useless. To tackle this problem, in this study, we proposed to exploit the power of the spectral anomaly ranking approach for generating a training set. The SRA was applied on a small fraction of the unlabeled dataset. The ranking vector generated by the SRA was used for creating the training set. The generated training set was used for training two models, namely, random forest and deep learning models. These trained models were used for estimating the target variable from the unlabeled dataset.

TABLE 1. Summary of the AUC for the auto-insurance fraud detection dataset For entries marked by *, AUC reported is one minus the actual AUC, OS: Overlapping Similarity, AGK: Adaptive Gaussian Kernel, HDK: Hamming Distance Kernel, DISC: DISK similarity

Automobile Fraud Detection Dataset										
Method			OS	AGK				HDK		Disk
				β				λ		
				10	100	1000	3000	0.5	0.8	
LOF	k_{svm}	10	0.53	0.5	0.52	0.58	0.64	0.52	0.53	0.55
		100	0.51	0.51*	0.54	0.58	0.67	0.51	0.52	0.57
		500	0.53	0.52*	0.55	0.59	0.68	0.51	0.51	0.57
		1000	0.53	0.52*	0.53	0.59	0.69	0.5	0.5	0.56
		3000	0.5	0.58*	0.55	0.58	0.69	0.54*	0.55*	0.53
OC-SVM	v_{svm}	0.01	0.51*	0.53*	0.51*	0.54	0.59	0.51*	0.52*	0.53*
		0.05	0.51*	0.53*	0.51*	0.55	0.59	0.52*	0.53*	0.52*
		0.1	0.51*	0.54*	0.51*	0.55	0.59	0.53*	0.54*	0.56*
SRA	mFlag	1	0.73	0.74	0.74	0.66	0.74	0.74	0.74	0.66

The approach is tested on a real auto-insurance claim dataset as well as two synthetic datasets. Results confirm the superiority of the proposed method in terms of accuracy as well as speed and performance.

Modern datasets are rapidly growing in size. Today, a handful of technologies provide solutions for handling large matrix operations. Apache Spark has emerged as a widely used open-source engine which is a fault-tolerant and general-purpose cluster computing framework. It provides APIs in Python, R, Java, and Scala. It also provides an optimized engine that supports general execution graphs. Recently, distributed linear algebra and some new optimization libraries have been developed in Spark. The linalg library consists of fast and scalable implementations of standard matrix computations. Common linear algebra operations such as multiplication, and more advanced operations such as factorizations are implemented in this library [14]. Using these technologies, one can proceed in implementing the SRA for the entire dataset. We suggest using these libraries to implement the SRA algorithm on large datasets.

6. REFERENCES

1. Tennyson. S, and Salsas-Forn. P, “Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives”, ***Journal of Risk and Insurance***, Vol. 69, No. 3, (2002), 289-308, doi: 10.1111/1539-6975.00024.
2. Artis. M, Ayuso. M, and Guillén. M, “Detection of Automobile Insurance Fraud with Discrete Choice Models and Misclassified Claims”, ***Journal of Risk and Insurance***, Vol. 69, No. 3, (2002), 325-340, doi: 10.1111/1539-6975.00022.
3. Viaene. S, Derrig. R. A, Baesens. B, and Dedene. G, “A Comparison of State-Of-The-Art Classification Techniques For Expert Automobile Insurance Claim Fraud Detection”, ***Journal of Risk and Insurance***, Vol. 69, No. 3, (2002), 373-421, doi: 10.1111/1539-6975.00023.
4. Brockett. P. L, Xia. X, and Derrig. R. A, “Using Khonen’s Self Organizing Feature Map to Uncover Automobile Bodily Injury Claim Fraud”, ***Journal of Risk and Insurance***, Vol. 65, No. 2, (1998), 245-274, doi: 10.2307/253535.
5. Brockett . P. L, Derrig. R. A, Golden. L. L, Levine. A, and Alpert. M “Fraud Classification Using Principal Component Analysis of RIDITs”, ***Journal of Risk and Insurance***, Vol. 69, No. 3, (2002) 341-371, doi: 10.1111/1539-6975.00027.
6. Nian. K, Zhang. H, Tayal. A, Coleman. Th, and Li. Y, “Unsupervised Spectral Ranking For Anomaly and Application to Auto Insurance Fraud Detection”, ***Journal of Finance and Data Science***, Vol. 2, No. 1, (2016), 58-75, doi: 10.1016/j.jfds.2016.03.001.
7. Boriah. S, Chandola. V, and Kumar. V, “Similarity measures for categorical data: A comparative evaluation***”, In Proceedings of the eighth SIAM International Conference on Data Mining***, 243-254, (2008).
8. Gartner. T, Le. Q. V, and Smola. A. J, “A short tour of kernel methods for graphs”, ***Technical report, NICTA***, Australia, Canberra, (2006).
9. Couto. J, “Kernel k-means for categorical data”, ***Lecture Notes in Computer Science, Springer***, (2005), 46-56, doi: 10.1007/11552253_5.
10. Gower. J. C, “A General Coefficient of Similarity and Some of Its Properties”, ***Biometrics***, Vol. 27, No. 4, (1971), 857-871, doi: 10.2307/2528823.
11. Subudhi. Sh, and Panigrahi. S, “Detection of Automobile Insurance Fraud Using Feature Selection and Data Mining Techniques”, ***International Journal of Rough Sets and Data Analysis***, Vol. 5, No. 3, (2018), 1-20, doi: 10.4018/IJRSDA.2018070101.
12. Wang. Y, and Xu. W, “Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud”,

Decision Support Systems, Elsevier, Vol. 105, (2018), 87-95, doi: 10.1016/j.dss.2017.11.001.

13. Li. Y, Yan. C, Liu. W, and Li. M, “A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification”, ***Applied Soft Computing, Elsevier***, Vol. 70, (2018), 1000-1009, doi: 10.1016/j.asoc.2017.07.027.
14. Bosagh-Zadeh. R, Meng. X, Ulanov. A, Yavus. B, Pu. L, Venkataraman. Sh, Sparks. E, Staple. A, Zaharia. M, “Matrix computations and optimization in Apache Spark”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 31-38, (2016).
15. Tennyson. Sh, and Salsas-Forn. P, “Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives”, ***The Journal of Risk and Insurance***, Vol. 69, No. 3, (2002), 289-308, doi: 10.1111/1539-6975.00024.
16. Itri. B, Mohamed. Y, Mohamed. Q, and Omar. B, “Performance comparative study of machine learning algorithms for automobile insurance fraud detection”, Third International Conference on Intelligent Computing in Data Sciences (ICDS), (2019).
17. Roy. R, George. K. Th, “Detecting insurance claims fraud using machine learning techniques”, International Conference on Circuit ,Power and Computing Technologies (ICCPCT), (2017).
18. Stephen-Kalwihura. J, and Logesvaran. R, “Auto-Insurance fraud detection: a behavioral feature engineering approach”, ***Journal of Critical Reviews***, Vol. 7, No. 3, (2020), 125-129, doi: 10.31838/jcr.07.03.23.
19. Abdallah. A, Aaizaini-Maarof. M, and Zainal. A, “Fraud detection system: A survey”, ***Journal of Network and Computer Applications***, Vol. 68, (2016), 90-113, doi: 10.1016/j.jnca.2016.04.007.
20. Phua. C, Lee. V, Smith. K, and Gayler. R, “A Comprehensive Survey of Data Mining-based Fraud Detection Research***”, Computers in Human Behavior***, Vol. 28, (2012), 1002–1013, doi: 10.1016/j.chb.2012.01.002.
21. Phua. C, Alahakoon. D, and Lee. V, “Minority report in fraud detection: classification of skewed data”, ***ACM SIGKDD Explorations***, Vol. 6, No. 1, (2004), 50-59, 10.1145/1007730.1007738.
22. H. Hassanpour, and A. Darvishi, “A Geometric View of Similarity Measures in Data Mining”, ***International Journal of Engineering, Transactions C: Aspects*** Vol. 28, No. 12, (2015), 1728-1737. doi: 10.5829/idosi.ije.2015.28.12c.05

Persian Abstract

چکیده

جمع‌آوری کردن نمونه‌های تقلب بیمه هزینه‌بر است و درصورتیکه به طور دستی انجام شود بسیار زمانبر خواهد بود. این امر استفاده از روش‌های یادگیری ماشین بدون ناظر را می‌طلبد. یکی از روش‌های دقیق در حوزه کشف تقلب بیمه خودرو، روش رتبه‌بندی طیفی آنومالی است که دارای دقت بسیار بالاتری نسبت به روش‌های معروف دیگر بوده است. با این حال، این روش در مواجهه با داده های حجیم، مقیاس‌پذیری کافی ندارد و برای کشف برخط آنومالی مناسب نیست. جهت جلوگیری از خسارت‌های هنگفت، سیستم‌های مدیریت تقلب برخط ضروری هستند. در این مطالعه، ما یک متدلوژی پیاده‌سازی را پیشنهاد می‌کنیم که استفاده از الگوریتم رتبه‌بندی طیفی را برای کلان داده ممکن می‌سازد. ما از توانایی روش رتبه بندی طیفی آنومالی، جهت تولید متغیر هدف تخمینی برای داده بدون برچسب استفاده می‌کنیم. سپس، از این داده دارای برچسب تخمینی برای آموزش دو مدل رگرسیون پایدار جنگل تصادفی و شبکه عصبی عمیق استفاده می‌کنیم. در مرحله بعد داده ورودی بدون برچسب، به مدل‌های آموزش‌دیده اعمال می‌شود و برچسب تخمینی به دست می‌آید. نتایج شبیه‌سازی‌ها نشان می‌دهد که روش پیشنهادی دارای سرعت بسیاری زیادی در کنار نرخ مثبت کاذب قابل قبول می‌باشد.

International Journal of Engineering

Journal Homepage: www.ije.ir

A Comparative Analysis of Wavelet-Based FEMG Signal Denoising with Threshold Functions and Facial Expression Classification Using SVM and LSSVM

V. Kehri*, R. N. Awale

Department of Electrical Engineering, VJTI Mumbai, India

PAPER INFO

Paper history:
Received 22 August 2019
Received in revised form 18 March 2020
Accepted 14 April 2020

Keywords:
Facial Electromyogram
Wavelet Transform
Support Vector Machine
Least-square Support Vector Machine

ABSTRACT

This work presents a technique for the analysis of facial electromyogram signal activities to classify five different facial expressions for computer-muscle interfacing applications. Facial electromyogram (FEMG) is a technique for recording the asynchronous activation of neuronal inside the face muscles with non-invasive electrodes. FEMG pattern recognition is a difficult task for the researcher, where classification accuracy is key concerns. Artifacts, such as eyeblink activity and electroencephalogram (EEG) signals interference, can corrupt these FEMG signals and directly affected the classification results. In this work, a robust wavelet-based thresholding technique, which comprised of a wavelet transform (WT) method and the statistical threshold, is proposed to remove the different artifacts from FEMG datasets and improve recognition accuracy rate. A set of five different raw FEMG data set was analyzed. Four wavelet basis functions, namely, haar, coif3, sym3, and bior4.4, were considered. The performance parameters (signal-to-artifact ratio (SAR) and normalized mean square error (NMSE) were utilized to understand the effect of the proposed signal denoising protocol. After denoising, the effectiveness of different statically features has been extracted. Two pattern recognition algorithms support vector machine (SVM) and the least square support vector machine (LSSVM) are implemented to classify extracted features. The performance accuracy of SVM and LSSVM classifier was evaluated and compared to know which classifier is the best for facial expression classification. The results showed that: (i) the proposed technique for denoising, improves the performance parameter results; (ii) The proposed work gives the best 95.24% classification accuracy.

doi*: 10.5829/ije.2020.33.07a.11*

NOMENCLATURE

PT	Proposed threshold	Ci	DWT coefficients
WP,Q (t)	Mother wavelet	K	Length of FEMG signal
βi	Statistical threshold	P	Scaling parameter
βNEW	Proposed threshold	Q	Shifting parameter
SAR	Signal to artifact ratio	std	Standard deviation
NMSE	Normalized mean square	N	Number of wavelet coefficient at each level
WT	Wavelet Transform	Ti	Threshold improvement factor of PT
DWT	Discrete wavelet transform	αi	Universal Threshold
SVM	Support vector machine		
LSSVM	Least square support vector machine		

1. INTRODUCTION

Recently researchers and scientists giving the highest priorities for developing a methodology to interface between electronics-mechanics with biology-medicine and try to improve the lifestyle. This study will help patients who are critically disabled and cannot even move their neck by inventing controlling devices, such as hands-free wheel-chairs. Since for designing such a system, strong human-computer interfaces have been needed [1]. Recognizing the facial expression through

*Corresponding Author Institutional Email: *vakehri@el.vjti.ac.in* (V. Kehri) bioelectrical action and transform into control commands for the system have focused in this study.

FEMG is the accepted standard for measuring facial muscle activity [1-3]. However, FEMG signals are nonlinear and random, generated by the summation of action potentials from thousands of motor units [4]. Surface electrodes acquired the recruitment and firing frequency of action potential.

EMG signals are a noninvasive technique of understanding muscle activity [5]. Due to the nonlinear and random nature of FEMG signals, mathematical tools such as DFT and FFT are failed to provide specific information. For analysis of such a signal, DWT methods were introduced, which provide better time-frequency information [6]. The FEMG signal having amplitude range differs from 0 to 12mv and frequency range differs from 0 to 450Hz respectively.

Table 1 depicts the previous research work which were used FEMG signals to classify either facial expressions or emotions. In all the studies, the number of classes, channels, segmentation with feature extraction method, and classification results were shown. This paper proposed an EMG based technique for recognizing five different facial expressions by proposing the methodology that results in good recognition accuracy. A wavelet-based denoising protocol comprised with a statistical threshold proposed to clean the FEMGs and improve the signal to noise ratio. This study on FEMG signal analysis is carried out in different stages. (1) Proposing a FEMGs denoising protocol, (2) Selecting the informative and discriminative FEMG features (3) Examining SVM and LSVM pattern recognition classifiers and identifying the best one.

2. SUBJECT AND EXPERIMENT SETUP

The ethical committee formed by the electrical department VJTI Institute Mumbai has approved the experimental work for FEMG signal recording.

In this study, the myon made aktos-mini EMG acquisition device depicts in Figure 1 used for FEMG signal recording. The electrodes were cleaned with alcohol, and then a gel is used to increase the conductivity of the electrode. After that, two pairs of non-invasive electrodes are attached to the specified participant’s face in the bipolar configuration as shown in Figure 1(a). The recorded signal sampled at 1KHz sampling frequency. FEMG data were collected from thirty physically fit subjects, including sixteen males and fourteen females in the age group of 18-40. The five facial expressions considered in this work are smiling, closing both eyes, opening the mouth saying ‘a’, raising the eyebrows and keeping the face in a neutral state. Participants were asked to perform each expression for two seconds of time duration. Each expression can be performed twice by each participant with ten seconds rest between them. Hence for each expression, four (2×2) seconds are informative information recorded

TABLE 1. Literature review in the area of FEMG Analysis

Classes	Channels	Segmentation (msec)	Features	Classifier	Accuracy (%)	Ref.
5	3	-	-	Thresholding	-	[7]
5	3	200	MAV	SVM	89.7	[8]
3	4	10	MRMS	LS	100	[9]
6	8	-	AV	Gaussian	92.00	[10]
5	1	400	RMS,FMDZC, MAV	BP, ANN	98.7	[11]
4	-	-	VAR	MLP,KNN	61.0,60.7	[12]
5	1	200	FMD,SC,WL,FMN	SVM	93.75	[13]
4	-	-	Wavelet	SVM	-	[14]
5	2	256	RMS	FCM	90.80	[15]
8	3	200	RMS	FCM, SVM	80.40,91.80	[16]
10	3	256	RMS	FCM	90.4	[17]
5	1	200	8 time domain features	SVM	93.50	[18]
4	-	-	Wavelet coefficients	LSSVM	91.6	[19]
10	3	256	WL, RMS, MAV	FCM	21.5-90.8	[20]
10	3	256	RMS, MAV	LSSVM	19.7-97.1	[21]
2	3	100	MAV	BPANN	80.-90.	[22]

through, data acquisition device. For five different, expressions each of 8000 sample's datasets (2 [no. of channels] ×4 seconds [informative signal] × 1000 [sampling frequency]) are collected from each subject. The band-stop filter with 50 Hz frequency is applied to remove the effect of line frequency noise.

(a) (b)

Figure 1. (a) Setup for FEMG Signal Recording, (b) Wireless Data acquisition System

3. WAVELET TRANSFORM AND WAVELET BASIS FUNCTION

Wavelet Transform (WT) converts the time-domain FEMG signal into its set of basis functions known as wavelets [23, 24]. These wavelet functions are achieved by doing dilation and shifting of the mother wavelet shown in Equation (1) [25].

$$\omega_{P,Q}(t) = w\left(\frac{t-Q}{P}\right) \tag{1}$$

In Equation (1), P indicates a scaling parameter whereas Q shifting parameter [25]. The FEMG datasets, decomposed into multi-level wavelet coefficients in order to get precise information where artifacts are available. Figure 2 shows the DWT decomposition structure.

WT of the FEMG datasets provides the multi-level coefficients which show the correlation between FEMG datasets with wavelet basis functions. Figure 3 depicts some WT basis functions implemented in this study. These wavelet functions resemble the characteristic of eye blinks activity, EEG artifacts, and perform well [26]. The selection of efficient wavelet basis function is considered as a dominant parameter in wavelet denoising for the FEMG signal. For non-stationary signals, biorthogonal is best for decomposing the signal.

In this work, we have implemented and compared symlet3, haar, coif3, and biorthogonal 4.4 wavelets basis function. Figure 3 depicts wavelet basis functions implemented in this work for artifact removal from FEMG data.

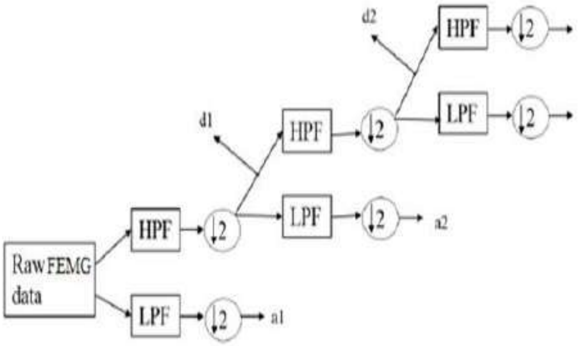


Figure 2. A DWT decomposition structure

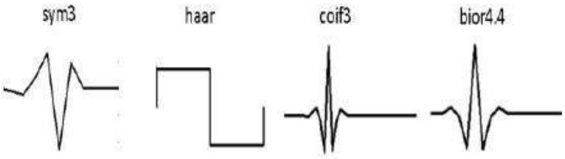


Figure 3. Wavelet basis functions applied for artifact removal from FEMG data

4. PROPOSED METHODOLOGY FOR FEMG SIGNAL DENOISING

The universal threshold (UT) was first suggested by Kumar et al. [26]. Threshold values are determined by the following relation:

$$\beta_i = \alpha_i \sqrt{2 \log K} \tag{2}$$

Here K represents the length of the raw FEMG signal, αi, mean absolute deviation, and βi is the threshold at ith decomposition level.

Statistical threshold (ST) was recommended by Krishnaveni et al. [27], which practically depends on the statistics of the signal. The effective threshold value βi is determined by the following equation:

$$\beta_i = 1.5 * \text{std}\,(C_i) \tag{3}$$

where Ci represents the DWT coefficients at the ith level and factor 1.5 is the approximate value of Gaussian noise.

The proposed threshold (PT) presented in this work depends on the statistics of the FEMG signal characteristic. The PT is adaptive to distinct sub-band by analyzing the wavelet coefficients. Mathematically, PT derived by the superposition of the universal threshold and the statistical threshold. The thresholds values of β_{NEW} determined by the following equation:

$$\beta_{NEW} = T_i * \text{std}\,(C_i) \tag{4}$$

$$T_i = e^{\frac{(\alpha_i - U_i)}{(\alpha_i + U_i)}} \tag{5}$$

Std(Ci) represents the standard deviation of wavelet coefficients at the ith level. Where Ti is the threshold improvement factor, and αi indicates the universal threshold function

$$U_i = \frac{\sum_i |C_i|}{N} \tag{6}$$

where N represents the number of wavelet coefficients at each level. Hard thresholding sets any coefficient greater than the threshold value to zero [28]. In this paper, hard thresholding was implemented, which removes wavelet coefficients (artifacts) if the wavelet coefficient is greater than the PT value. Figure 4 depicts the general steps of the proposed methodology for denoising of FEMG signal.

4. 1. Performance Parameters

The performance of the proposed threshold is based on the two statistical performance parameters. The performance parameters are SAR and NMSE, observed in this work. SAR is a technique to estimate the amount of artifact removal in a FEMG signal after processing with the proposed algorithm [29]. If F is the FEMG signal with artifact and Fˆ is the corrected signal obtained after processing then

$$\mathrm{SAR} = 10\log\left(\frac{\mathrm{std}(\mathrm{F})}{\mathrm{std}(\mathrm{F}-\hat{\mathrm{F}})}\right) \tag{7}$$

NMSE define the difference between F1(j) (signal without artifact) and F2(j) (signal with artifacts) [11]. NMSE is computed in dB using the given equation.

$$NMSE = 20 log E\left\{\frac{\sum[F1\,(j) - F2\,(j)]^2}{\sum[F1\,(j)]^2}\right\} \tag{8}$$

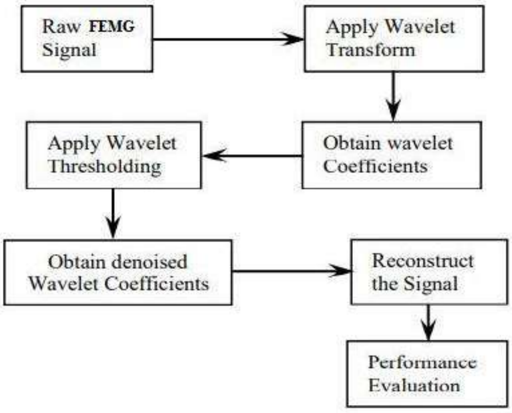

Raw FEMG Signal
Apply Wavelet Transform
Obtain wavelet Coefficients
Apply Wavelet Thresholding
Obtain denoised Wavelet Coefficients
Reconstruct the Signal
Performance Evaluation

Figure 4. Block diagram of the proposed methodology for FEMG signal denoising

5. FEATURE EXTRACTION

Analysis of large numbers of FEMG data is a difficult task for researchers and dimension reduction is a necessary step for further analysis. Feature extraction performs a key role by converting huge datasets into the précised meaning. There are various techniques with numerous complexity in time and frequency domain, which shows different FEMG characteristics [5]. For feature extraction, we implemented a WT method that generates wavelet coefficients [29-36]. An active part of FEMG data containing 8000 samples with a sampling frequency of 1000Hz. The FEMG data then decomposed into 3 levels of decomposition using on wavelet family 'db4'. The frequency bands for decomposed wavelet coefficients are 0 - 120 Hz. Out of every decomposed frequency sub-band, we extracted six different features using wavelet coefficients. According to literature, very rare studies analyzed and compared FEMG frequency domain features [30]. The perfect evaluation and analysis are required to discover the most discriminative FEMGs feature by selecting a range of classifiers. In this paper, the performance accuracy of six widely used frequency domain FEMG features is determined. The wavelet domain extracted features are mean, variance, covariance, standard deviation, energy, and RMS.

6. CLASSIFICATION

For recognition of different facial expressions, extracted features must be classified into accurate classes. The selected classifier must be fast and efficient enough to meet the proper requirement. Here, two pattern recognition algorithms are implemented on extracted features to classify the FEMG datasets. The selection of classifiers will be based on several criteria, such as high-performance accuracy based on literature, processing time, etc. Frequency-domain features can be given to classifiers that classify facial expression. The implemented classifiers were grouped into different kernel machine method. The optimum model of the classifier can be designed by examining a wide range of kernel values in order to find the best performance of the classifier. For this purpose, a 70-30 cross-validation scheme is used to test the parameters and evaluate the classifier performance. In this work, SVM [13, 33] and LSSVM [31] are utilized.

SVM is a nonparametric classifier, and it targets to determine discriminant hyperplanes that differentiate the data into various groups [30]. A multiclass SVM with a polynomial kernel function was implemented. The Polynomial kernel is given good accuracy compared to other nonlinear kernels. The mathematical equation of the polynomial kernel given as follows:

$$f(y_i, y_j) = (y_i * y_j + 1)^c \tag{9}$$

where c represents the degree of the polynomial. The recognition accuracy depends on the degree of the polynomial (c). For FEMG datasets (nonlinear), poly-order should be more than one.

LSSVM is the advanced version/model of SVM classifiers, and it improves the process during the testing and training phase [31]. In this technique, equality constraints were implemented to find the solution to the optimal problem by dealing with a set of equations instead of the quadratic optimization problems [20]. The LSSVM methodology was implemented in this paper formed by the Gaussian kernel function. The Gaussian kernel GK(Ka, Kb) defined as follows:

$$GK\,(k_a, k_b\,) = \exp\left\{-\frac{\|k_a - k_b\|}{2\sigma^2}\right\}; a, b = 1,2,.\,N \tag{10}$$

Here σ shows the width of the Gaussian kernel [37, 38].

7. RESULTS ANALYSIS AND DISCUSSION

Let Y[n] is the recorded FEMG signal (with artifacts) and Ytr[n] is the true FEMG signal (artifact-free). The objective of the wavelet-based proposed thresholding algorithm in this work is to estimate Ytr[n] by efficiently removing artifacts from Y[n]. The proposed algorithm for denoising of FEMG signals as follows:

- Wavelet transforms with different wavelet basis function, decompose the actual FEMG signal Y[n] into wavelet coefficients $\mathbf{W_i^n}$ = [$Wi^{1,}$ Wi^2, ……Wi^n] at each scale i.
- Wavelet coefficients W_i^n at each scale i, was thresholded, by applying appropriate thresholding function shown in Equation 5. The thresholded wavelet coefficients $T_i^n = [T_i^1, T_i^2\ , T_i^3 \ldots\ldots\ldots T_i^n]$ are the estimate of the coefficient values of Y_{tr}[n].
- Denoised (reconstructed) signal was obtained by applying the inverse wavelet transform on the thresholded wavelet coefficients $\mathbf{T_i^n}$.

The performance of the proposed threshold was measured by SAR and NMSE. The methodology that gives the maximum value of SAR and the minimum value of NMSE is more acceptable. Table 2 compares the SAR values on FEMG signals using different threshold and different wavelet basis functions. From Table 2, it is noted that SAR values are maximum when DWT + PT combination is used with all four types of wavelet basis function, indicating DWT + PT effectively reduces the artifacts, while DWT + ST and DWT + UT is conservative. Table 3 compares the NMSE values on FEMG signals using different thresholds and different wavelet basic functions. Based on Table 3, DWT + PT again performs well. The lower value of NMSE shows the best performance. NMSE values are minimum when PT is applied with DWT and a combination of all four basis functions. The proposed algorithm is tested for FEMG signals with artifacts. The sample for an eight-second epoch is shown in Figure 5, along with the reconstructed FEMG signal obtained after PT.

Classification accuracy is the most important parameter to estimate the system performance. The classification accuracy of classifiers is better by applying denoised features, obtain after PT rather than the raw ones. The LSSVM classifier gives the best classification results, which is 95.24%, followed by SVM (91%), respectively. Table 4 depicts the facial expression classification confusion matrix for the LSSVM model. Table 5 depicts the facial expression classification confusion matrix for the SVM Model. The results show that all five facial expressions (five classes) recognized with high accuracy. For comparison of SVM

TABLE 2. SAR on FEMG signals with different threshold and wavelet basis function

Thresh.	sym3	Haar	Coif3	Bior4.4
DWT+UT	1.28±0.63	1.13±0.49	1.31±0.63	1.3 ± 0.62
DWT + ST	1.91±0.85	1.88±0.75	1.9±0.82	1.88±0.73
DWT + PT	**2.53±0.96**	**2.98±0.79**	**2.30±0.85**	**2.28± 0.85**

TABLE 3. NMSE on FEMG signals with different threshold and wavelet basis function

Thresh.	sym3	Haar	Coif3	Bior4.4
DWT+ UT	-5.21±2.7	-4.8±1.90	-5.23±2.7	-5.29±2.35
DWT + ST	-7.97±2.1	--5.9±2.55	-7.5±3.3	-7.32±3.15
DWT+ PT	**-8.82±3.7**	**-8.57±2.6**	**-8.74±3.2**	**-8.87±3.22**

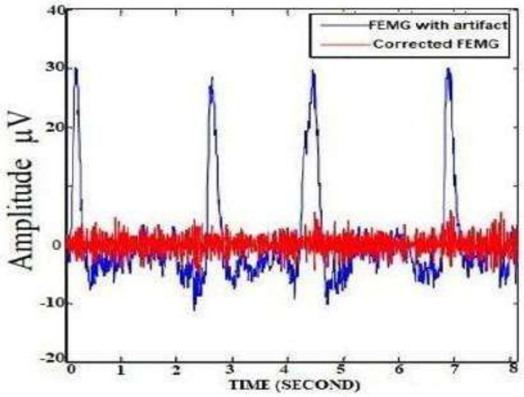


Figure 5. FEMG with an artifact and corrected FEMG signal using PT

TABLE 4. The facial expression classification confusion matrix for the test set using the LSSVM

Facial Expression	**Cass 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Class 1	98.8	1.2	0	0	0
Class 2	4.7	90.5	1.2	1.2	2.4
Class 3	0	2.4	89.3	0	8.3
Class 4	0	0	0	100	0
Class 5	0	0	2.4	0	97.6
Average(%): 95.24					

TABLE 5. The facial expression classification confusion matrix for the test set using the SVM Model

Facial Expression	**Cass 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Class 1	90	0	4	6	0
Class 2	3	95	0	0	2
Class 3	4	0	92	4	0
Class 4	5	0	6	81	8
Class 5	0	3	0	0	97
Average(%): 91.0					

TABLE 6. The classification accuracy obtained from different classifiers

Sr. No.	**Classifier**	**Classification Accuracy (%)**	
		RAW FEMG	Denoised FEMG (After Proposed Threshold)
1	LSSVM	87.3	95.24
2	SVM	86.9	91
3	KNN	85.1	90.1
4	ANN	83.6	89.3
5	LDA	80.5	85.14

with LSSVM, we implemented the same training and test set for the SVM and LSSVM based on recognition model. The results show that the performance of the SVM model was lower than that of the LSSVM model. Table 6 compares the classification accuracy of different classifiers for raw and denoised FEMG signal.

8. CONCLUSION

An effective study based on FEMG signal analysis is presented here in order to provide the best performance to classify five different facial expressions. In this work, FEMG data acquired from myon made wireless data acquisition device has been presented as a representative of the FEMG signal contaminated with artifacts to compare several wavelet-based techniques. The most common artifact in the FEMG signal is due to the effect of EEG interference and eye blink activity. A robust technique is proposed for FEMG signal denoising, including DWT with proposed thresholding (PT), and results show that the proposed method denoise the FEMG signal effectively and enhances the performance of the classifier. Based on the SAR results depicted in Table 2, DWT with PT using haar wavelet is to be more useful than other combinations. Based on analysis and results, DWT with PT using all WT basis functions have also performed satisfactorily for removing different artifacts while preserving original signals.

After denoising of FEMG data, features can be extracted using the WT method with the db8 family. Among all six features, RMS and energy is the most informative feature. Inspection on two classifiers SVM and LSSVM reveals that the LSSVM model has better capability to classify features giving 95.24% classification accuracy. Our study shows the proposed signal denoising protocol can improve the system performance. This presented work also helps to set up a systematic connection between the face muscle and machine. This interface can be applied for designing real real-time processing controlling devices like assistive wheelchairs.

9. REFERENCES

1. Kotzé, P., Eloff, M.M. and Adesina-Ojo, A., "Accessible computer interaction for people with disabilities: The case of quadriplegics, 6th International Conference on Enterprise Information Systems, Porto., (2004). http://hdl.handle.net/10500/465
2. Ferreira, A., Silva, R., Celeste, W., Bastos Filho, T. and Sarcinelli Filho, M., "Human–machine interface based on muscular and brain signals applied to a robotic wheelchair", in Journal of Physics: Conference Series, IOP Publishing. Vol. 90, (2007), 012094. DOI:10.1088/1742-6596/90/1/012094
3. Englehart, K., Hudgin, B. and Parker, P.A., "A wavelet-based continuous classification scheme for multifunction myoelectric control", ***IEEE Transactions on Biomedical Engineering***, Vol. 48, No. 3, (2001), 302-311. DOI: 10.1109/10.914793
4. Englehart, K. and Hudgins, B., "A robust, real-time control scheme for multifunction myoelectric control", ***IEEE Transactions on Biomedical Engineering***, Vol. 50, No. 7, (2003), 848-854. DOI: 10.1109/TBME.2003.813539
5. Kehri, V., Ingle, R., Patil, S. and Awale, R., Analysis of facial emg signal for emotion recognition using wavelet packet transform and svm, in Machine intelligence and signal analysis. 2019, Springer.247-257. DOI: https://doi.org/10.1007/978-981-13-0923-6_21
6. Shenoy, P., Miller, K.J., Crawford, B. and Rao, R.P., "Online electromyographic control of a robotic prosthesis", ***IEEE Transactions on Biomedical Engineering***, Vol. 55, No. 3, (2008), 1128-1135. DOI: 10.1109/TBME.2007.909536

7. Tsui, C.S.L., Jia, P., Gan, J.Q., Hu, H. and Yuan, K., "Emg-based hands-free wheelchair control with eog attention shift detection", in 2007 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE. (2007), 1266-1271. DOI: 10.1109/ROBIO.2007.4522346

8. Firoozabadi, S.M.P., Oskoei, M.A. and Hu, H., "A human-computer interface based on forehead multi-channel bio-signals to control a virtual wheelchair", in Proceedings of the 14th Iranian conference on biomedical engineering (ICBME). (2008), 272-277.

9. Arjunan, S. and Kumar, D.K., "Recognition of facial movements and hand gestures using surface electromyogram (semg) for hci based applications", in 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007), IEEE. (2007), 1-6. DOI: 10.1109/DICTA.2007.4426768

10. Gibert, G., Pruzinec, M., Schultz, T. and Stevens, C., "Enhancement of human computer interaction with facial electromyographic sensors", in Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7. (2009), 421-424. DOI: https://doi.org/10.1145/1738826.1738914

11. Wei, L., Hu, H. and Yuan, K., "Use of forehead bio-signals for controlling an intelligent wheelchair", in 2008 IEEE International Conference on Robotics and Biomimetics, IEEE. (2009), 108-113. DOI: 10.1109/ROBIO.2009.4912988

12. Van den Broek, E., Lisy, V., Janssen, J., Westerink, J., Schut, M. and Tuinenbreijer, K., "Affective man-computer interface: Unveiling human emotions through biosignals", ***Biomedical Engineering Systems and Technologies***, Vol. 52, (2010), 21-47. https://doi.org/10.1007/978-3-642-11721-3_2

13. Wei, L. and Hu, H., "Emg and visual based hmi for hands-free control of an intelligent wheelchair", in 2010 8th World Congress on Intelligent Control and Automation, IEEE. (2010), 1027-1032. DOI: 10.1109/WCICA.2010.5554766

14. Yang, G. and Yang, S., "Emotion recognition of electromyography based on support vector machine", in 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, IEEE. (2010), 298-301. DOI: 10.1109/IITSI.2010.122

15. Hamedi, M., Rezazadeh, I.M. and Firoozabadi, M., "Facial gesture recognition using two-channel bio-sensors configuration and fuzzy classifier: A pilot study", in International Conference on Electrical, Control and Computer Engineering 2011 (InECCE), IEEE. (2011), 338-343. DOI: 10.1109/INECCE.2011.5953903

16. Hamedi, M., Salleh, S.-H. and Swee, T.T., "Surface electromyography-based facial expression recognition in bi-polar configuration", ***Journal of Computer Science***, Vol. 7, No. 9, (2011), 1407-1415. ISSN 1549-3636

17. Hamedi, M., Salleh, S.-H., Tan, T., Ismail, K., Ali, J., Dee-Uam, C., Pavaganun, C. and Yupapin, P., "Human facial neural activities and gesture recognition for machine-interfacing applications", ***International Journal of Nanomedicine***, Vol. 6, (2011), 3461-3472. DOI: 10.2147/IJN.S26619

18. Wei, L., Hu, H. and Zhang, Y., "Fusing emg and visual data for hands-free control of an intelligent wheelchair", ***International Journal of Humanoid Robotics***, Vol. 8, No. 04, (2011), 707-724. DOI: 10.1142/S0219843611002629

19. Yang, Y.G. and Yang, S., "Study of emotion recognition based on surface electromyography and improved least squares support vector machine", ***Journal of Computers***, Vol. 6, No. 8, (2011), 1707-1714. DOI:10.4304/jcp.6.8.1707-1714

20. Hamedi, M., Salleh, S.-H., Noor, A.M., Swee, T.T. and Afizam, I., "Comparison of different time-domain feature extraction methods on facial gestures' emgs", ***Progress In Electromagnetics Research***, Vol. 1897, (2012), 1897-1900. DOI: http://eprints.utm.my/id/eprint/34373/

21. Hamedi, M., Salleh, S.-H., Noor, A., Harris, A.R. and Majid, N., "Multiclass least-square support vector machine for myoelectric-based facial gesture recognition", in The 15th International Conference on Biomedical Engineering, Springer. (2014), 180-183. DOI: 10.1007/978-3-319-02913-9_46

22. Gruebler, A. and Suzuki, K., "Design of a wearable device for reading positive expressions from facial emg signals", ***IEEE Transactions on Affective Computing***, Vol. 5, No. 3, (2014), 227-237. DOI: 10.1109/TAFFC.2014.2313557

23. Chen, Y., Yang, Z. and Wang, J., "Eyebrow emotional expression recognition using surface emg signals", ***Neurocomputing***, Vol. 168, (2015), 871-879. https://doi.org/10.1016/j.neucom.2015.05.037

24. Zikov, T., Bibian, S., Dumont, G.A., Huzmezan, M. and Ries, C., "A wavelet based de-noising technique for ocular artifact correction of the electroencephalogram", in Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology, IEEE. Vol. 1, (2002), 98-105. DOI: 10.1109/IEMBS.2002.1134407

25. M SadeghpourHaji; S. A. Mirbagheri; A. H. Javid; M. Khezri; G. D. Najafpour, "A wavelet support vector machine combination model for daily suspended sediment forecasting", ***International Journal of Engineering C: Aspects***, Vol. 27, No. 6, (2014), 855-864. DOI:10.5829/idosi.ije.2014.27.06c.04

26. Kumar, P.S., Arumuganathan, R., Sivakumar, K. and Vimal, C., "Removal of ocular artifacts in the eeg through wavelet transform without using an eog reference channel", ***International Journal of Open Problems inComputer Science and Mathematics***, Vol. 1, No. 3, (2008), 188-200. https://www.emis.de/journals/IJOPCM/files/IJOPCM(vol.1.3.2.D.8).pdf

27. Krishnaveni, V., Jayaraman, S., Malmurugan, N., Kandasamy, A. and Ramadoss, D., "Non adaptive thresholding methods for correcting ocular artifacts in eeg", ***Academic Open Internet Journal***, Vol. 13, (2004). http://www.acadjournal.com/2004/V13/Part1/p2/

28. Cichocki, A., Shishkin, S.L., Musha, T., Leonowicz, Z., Asada, T. and Kurachi, T., "Eeg filtering based on blind source separation (bss) for early detection of alzheimer's disease", ***Clinical Neurophysiology***, Vol. 116, No. 3, (2005), 729-737. https://doi.org/10.1016/j.clinph.2004.09.017

29. Soomro, M.H., Badruddin, N., Yusoff, M.Z. and Malik, A.S., "A method for automatic removal of eye blink artifacts from eeg based on emd-ica", in 2013 IEEE 9th International Colloquium on Signal Processing and its Applications, IEEE. (2013), 129-134. DOI: 10.1109/CSPA.2013.6530028

30. Kehri, V. and Awale, R., "Emg signal analysis for diagnosis of muscular dystrophy using wavelet transform, svm and ann", ***Biomedical and Pharmacology Journal***, Vol. 11, No. 3, (2018), 1583-1591. DOI: https://dx.doi.org/10.13005/bpj/1525

31. Shen, J. and Zhou, X., "Least squares support vector machine for constitutive modeling of clay", ***International Journal of Engineering***, Vol. 28, No. 11, (2015), 1571-1578. DOI: 10.5829/idosi.ije.2015.28.11b.04

32. Suykens, J., Lukas, L., Van Dooren, P., De Moor, B. and Vandewalle, J., "Least squares support vector machine classifiers: A large scale algorithm", in Proceedings of the European Conference on Circuit Theory and Design, Citeseer. Vol. 10, (1999). https://perso.uclouvain.be/paul.vandooren/publications/SuykensLVDV99.pdf

33. Toledo-Pérez, D.C., Rodríguez-Reséndiz, J., Gómez-Loenzo, R.A. and Jauregui-Correa, J., "Support vector machine-based

emg signal classification techniques: A review", ***Applied Sciences***, Vol. 9, No. 20, (2019), 4402. https://doi.org/10.3390/app9204402

34. Toledo-Pérez, D.C., Martínez-Prado, M.A., Gómez-Loenzo, R.A., Paredes-García, W.J. and Rodríguez-Reséndiz, J., "A study of movement classification of the lower limb based on up to 4-emg channels", ***Electronics***, Vol. 8, No. 3, (2019), 259. https://doi.org/10.3390/electronics8030259

35. Toledo-Pérez, D., Rodríguez-Reséndiz, J. and Gómez-Loenzo, R.A., "A study of computing zero crossing methods and an improved proposal for emg signals", ***IEEE Access***, Vol. 8, (2020), 8783-8790. DOI: 10.1109/ACCESS.2020.2964678

36. Purushothaman, G. and Vikas, R., "Identification of a feature selection based pattern recognition scheme for finger movement recognition from multichannel emg signals", ***Australasian Physical & Engineering Sciences in Medicine***, Vol. 41, No. 2, (2018), 549-559. DOI: 10.1007/s13246-018-0646-7

37. Too, J., Abdullah, A., Saad, N.M., Ali, N. And Zawawi, T.T., "Application of spectrogram and discrete wavelet transform for emg pattern recognition", ***Journal of Theoretical & Applied Information Technology***, Vol. 96, No. 10, (2018), 3036-3047. http://www.jatit.org/volumes/Vol96No10/24Vol96No10.pdf

38. Sui, X., Wan, K. and Zhang, Y., "Pattern recognition of semg based on wavelet packet transform and improved svm", ***Optik***, Vol. 176, (2019), 228-235. https://doi.org/10.1016/j.ijleo.2018.09.040

Persian Abstract

چکیده

این کار یک تکنیک برای تجزیه و تحلیل فعالیت های سیگنال الکترومیوگرام صورت برای طبقه بندی پنج حالت مختلف صورت برای برنامه های کاربردی واسط رایانه-عضله ارائه می دهد. الکترومیوگرام صورت (FEMG) روشی برای ضبط فعال سازی ناهمزمان نورون در عضلات صورت با الکترودهای غیر تهاجمی است. شناخت الگوی FEMG برای محقق کاری دشوار است ، که در آن دقت طبقه بندی نگرانی های کلیدی است. مصنوعات ، از قبیل فعالیت چشم بند و تداخل سیگنال های الکتروانسفالوگرام (EEG) ، می توانند این سیگنال های FEMG را خراب کرده و به طور مستقیم بر نتایج طبقه بندی تأثیر بگذارند. در این کار ، یک تکنیک آستانه محور مبتنی بر موجک ، که از یک روش تبدیل موجک (WT) و آستانه آماری تشکیل شده است ، برای حذف آثار باستانی مختلف از مجموعه داده های FEMG و بهبود میزان دقت تشخیص استفاده شده است. مجموعه ای از پنج مجموعه داده مختلف FEMG خام مورد تجزیه و تحلیل قرار گرفت. چهار تابع پایه موجک ، یعنی haar، coif3، Sym3 و bior4.4 در نظر گرفته شد. از پارامترهای عملکرد (نسبت سیگنال به مصنوع (SAR) و میانگین خطای مربع عادی (NMSE) برای درک تأثیر پروتکل دیوایزینگ سیگنال پیشنهادی استفاده شده است.پس از denoising ، اثربخشی ویژگی های مختلف استاتیک استخراج شده است. دستگاه بردار پشتیبانی الگوریتم ها (SVM) و ماشین بردار کمترین مربع پشتیبانی (LSSVM) برای طبقه بندی ویژگی های استخراج شده اجرا می شوند.علت عملکرد SVM و طبقه بندی کننده LSSVM مورد بررسی قرار گرفت و مقایسه شد تا بدانیم کدام طبقه بندی بهترین برای طبقه بندی بیان صورت است. نتایج نشان داد که: (۱) روش ارائه شده برای نانوایی کردن ، پارامتر عملکرد را بهبود می بخشد ؛ (ب) کار ارائه شده بهترین دقت طبقه بندی ۹۵.۲۴٪ را نشان می دهد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Efficient Parallelization of a Genetic Algorithm Solution on the Traveling Salesman Problem with Multi-core and Many-core Systems

M. Abbasi*, M. Rafiee

Department of Computer Engineering, Engineering Faculty, Bu-Ali Sina University, Hamedan, Iran

PAPER INFO

Paper history:
Received 10 November 2019
Received in revised form 23 December 2019
Accepted 04 January 2020

Keywords:
Genetic Algorithm
Parallel
Multi-core
Many-core
Traveling Salesman Problem

ABSTRACT

Efficient parallelization of genetic algorithms (GAs) on state-of-the-art multi-threading or many-threading platforms is a challenge due to the difficulty of scheduling hardware resources regarding the concurrency of threads. In this paper, for resolving the problem, a novel method is proposed, which parallelizes the GA by designing three concurrent kernels, each of which are running some dependent effective operators of GA. The proposed method can be straightforwardly adapted to run on many-core and multi-core processors by using Compute Unified Device Architecture (CUDA) and Threading Building Blocks (TBB) platforms. To efficiently use the valuable resources of such computing cores in concurrent execution of the GA, threads that run any of the triple kernels are synchronized by a considerably fast switching technique. The offered method was used for parallelizing a GA-based solution of Traveling Salesman Problem (TSP) over CUDA and TBB platforms with identical settings. The results confirm the superiority of the proposed method to state-of-the-art methods in effective parallelization of GAs on Graphics Processing Units (GPUs) as well as on multi-core Central Processing Units (CPUs). Also, for GA problems with a modest initial population, though the switching time among GPU kernels is negligible, the TBB-based parallel GA exploits the resources more efficiently.

***doi**: 10.5829/ije.2020.33.07a.12*

1. INTRODUCTION

Meta-heuristic optimization algorithms [1-3] like Genetic Algorithms (GAs) have been widely used in science and engineering problems [3-5]. GA is considered as a class of evolutionary algorithms that are used for finding approximate solutions in search [6], optimization problems [7, 8], image processing [9], optimizing artificial neural networks [10], scheduling [11, 12], and rule-based systems [13].

The main difficulty with using GA is the considerable iterations of the genetic algorithm [14]. In such cases, increasing the number of generations would raise the rate of crossovers and mutations. These, in turn, expressively increase the time complexity of the organized algorithm. Therefore, many researchers try to examine parallelization methods for GA on multi-core systems as well as many-core systems [15]. A common approach is to migrate the computation of fitness, mutation, crossover, and selection functions to parallel machines [16]. An interesting point is the efficiency of the deployed parallel kernels in using the computational resources of systems for accelerating GA. Although different approaches with different complexities have been presented for parallel programming on multi-core and many-core systems, a few of them have been carried out to quantitatively assess and compare the efficiency of parallel kernels on multi-core machines with that of many-core systems. Also, due to the intricate design of parallel kernels, none of them have efficiently utilized the computational resources of the parallel systems to accelerate GA.

A pilot study in the domain of efficient parallelization of GA is the research of Zhu et al. [17], which combines the mechanism of Threading Building Blocks (TBB) and Message Passing Interface (MPI) platforms to parallelize a GA-based solution of the Traveling Salesman Problem (TSP). Consistent with the results obtained from

*Corresponding Author Institutional Email: *abbasi@basu.ac.ir* (M. Abbasi) implementation on different datasets in one hundred generations, the achieved acceleration rate on four processing cores in 144 and 1889 cities is 2.1 and 2.55, respectively. Studies directed by Fujimoto et al. [18] and Chen et al. [19] can be considered as preliminary studies, which present two different methods for parallelizing GA computations on the Compute Unified Device Architecture (CUDA) platform. None of those studies can fully exploit the computational resources of the Graphics Processing Units (GPUs). Other studies, like [20, 21] and [22-25], have incoherently executed parallel kernels for GA-based solutions of TSP on GPUs.

The most recently conducted study in this trend is the study conducted by Saxena et al. [26], which compares the efficiency of Open Multi-Processing (OpenMP) and CUDA through running parallel GA-based optimization kernels on multi-core Central Processing Units (CPUs) and GPUs. Unfortunately, this study does not offer a typical experimental setting for all parallel kernels. As a consequence, their shallow results cannot be used for any decision and comparison as to the efficiency of those parallelization platforms.

Our investigation shows that all of the researches in the field of GA parallelization have focused on the unclear design of parallel algorithms. Also, none of the recent studies have inspected the approaches for efficient parallelization of GA operators on multi-core platforms like TBB and CUDA. The different models of parallel processing of the threads and diverse approaches of synchronization of threads in different blocks of GPU have not been exactly studied in any of them.

Motivated by the above mentioned problem, this paper proposes an efficient method for parallelizing the key operators of the genetic algorithm. The proposed parallelization method is based on the structure of multi-core CPUs and many-core GPUs. It shows that by concise use of the parallel resources of a multi-core CPU, the efficiency of multi-core parallelization can be higher than that of a many-core GPU. Also, the presented study inspects the effect of different parameters of GA-based solutions of TSP on the performance of the parallel kernel on both multi-core systems as well as many-core systems.

TSP is an NP-complete problem. The main idea of TSP is to find the shortest path among a set of cities, provided that each city is visited only once, and the source city should be revisited at the end.

The rest of the paper is structured as follows. Section 2 reviews the structure of CUDA and TBB platforms. In addition, the GA solution of the TSP is described in the third section. Then, the offered approach for parallel implementation of a GA is discussed. Experiments and their corresponding analyses are explored in the following section. Section 5 concludes the paper and proposes some future directions.

2. BACKGROUND

This section gives a brief overview of the related concepts, including the architecture of GPU in the CUDA platform and the main ingredients of TBB architecture.

2. 1. CUDA The graphics processing unit is a tool dedicated to display graphic images at workstations, game consoles, or personal computers [27, 28]. CUDA provides features for developers to use the hardware capabilities of Nvidia graphics cards in non-graphical programs and speed up the execution speed of complex algorithms using GPU capabilities. CUDA supports the main factors involved in computing from two different points of view: host and device. The host performs the main program while the machine aids in processing. A typical scenario is that the CPU is considered as the host and the GPU is considered as a help to the processor.

Any program written in CUDA can consist of several kernels. Each kernel is implemented by a grid of several blocks. Each block is made of several threads. These threads are responsible for implementing the program [29].

2. 2. TBB Intel Threading Building Blocks (Intel® TBB) is a common C++ library for writing parallel shared-memory programs. Using this library provides benefits including synchronized containers, scalable memory allocator, work-stealing task scheduler, low-level synchronization primitives. This library is considered as the best tool for task-based parallelism. The details of scheduling by this efficient library can be found in many resources, such as [30, 31].

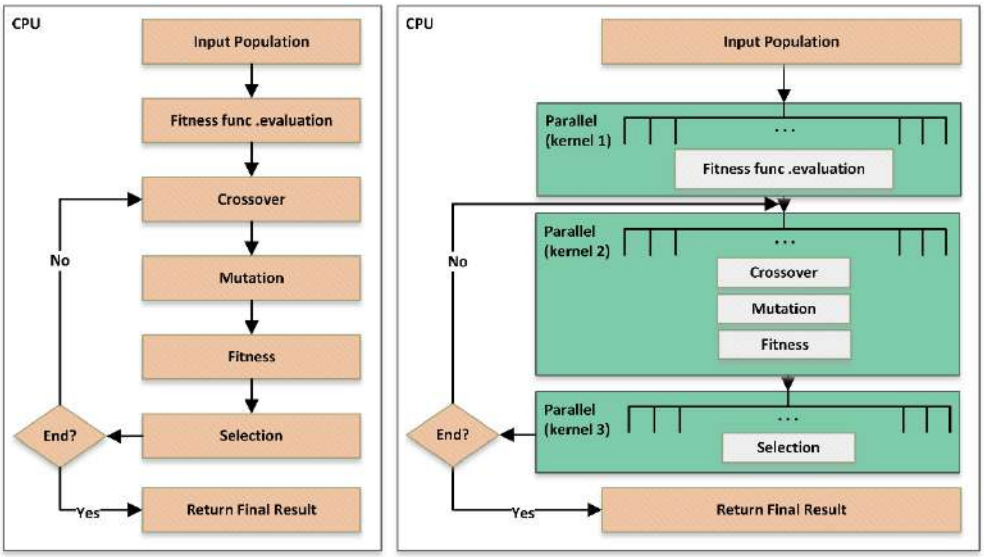
3. THE PROPOSED APPROACH

The common method for solving the TSP with a genetic algorithm is shown by a flowchart in Figure 1(a). This algorithm is used in a lot of applications in different areas of science and engineering [32].

Population: each chromosome contains a fixed number of genes. In this case, each city is represented by a gene, and each chromosome is a permutation of cities.

The fitness function of the initial population: for each chromosome, the function produces a non-negative integer value, which indicates fitness and individual aptitude of each chromosome. In the calculating the fitness of each chromosome in TSP, a matrix containing the coordinates of cities is used. The distance between every pair of cities in a chromosome is computed according to the following equation [33]:

$$f(x) = \left(\sum_{i=2}^{n} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}\right) + \sqrt{(x_1 - x_n)^2 + (y_1 - y_n)^2} \tag{1}$$



(a) Sequential Genetic Algorithm, | (b) Parallel Genetic Algorithm

Figure 1. Sequential and parallel approaches for Genetic Algorithm

In the above equation, x_i and y_i denote the coordinate of the i-th city of a chromosome.

Crossover: this operator exchanges the information between the paired chromosomes and also controls the convergence speed of the genetic algorithm with a probability. This probability value P_c is called the crossover rate. For doing a crossover, a parent and a random position between the parent's genes are considered. Then, all the genes at both sides of the parent chromosome with respect to the specified position are moved to form a new chromosome.

Mutation: This operator produces the new chromosome by randomly changing one of the genes with a low probability. The overall probability of mutation on a chromosome is called mutation rate, which is denoted by P_m.

Calculating the fitness of a generation: the fitness function of the population is calculated using crossover and mutation operators.

Selection: there are diverse methods for selecting the best chromosome and transferring it to the next generation. Commonly, the tournament method is used for selection. In this method, two chromosomes are randomly selected from the population. Then, a random number between zero and one is selected as r. Next, two fitted chromosomes or the ones which are less adapted are selected as the parents. These two chromosomes are then returned to the initial population and again participated in the selection process. Finally, the selected chromosomes are recognized as the next generation and are sent to the next round of algorithm implementation [16]. The statistical analysis of the operators of genetic algorithms was thoroughly investigated in many research studies, including [34-36]. This issue has a substantial effect on designing an efficient GA [37].

The general assembly of the sequential and parallel genetic algorithm is demonstrated in Figure 1(b). In all of the offered parallel kernels, the original population is identical. The aim of the proposed kernels for parallel GA is to assess and compare the efficiency of parallelized codes using CUDA and TBB. In the sequential method, all the operations of the genetic algorithm are performed consecutively. But, in the parallel method, the important operations of the GA are executed in parallel, which decreases the runtime of the GA. These operators are fitness, crossover, mutation, and selection. Figure 1(b) demonstrates the flowchart of the proposed method for parallelizing GA. Regarding the structure of GPUs, only the threads in one CUDA block can be synchronized. The synchronization is a crucial mechanism in the genetic algorithm, where some operators need to be executed in sequence. The most important benefit of our method over recent methods like [19] is offering a technique for solving the problem of synchronizing threads which can synchronize the threads of more than one block. For this purpose, we form three different kernels, each of which corresponds to a different function of GA. These kernels would be duplicated on different CUDA blocks at the same time. By switching between kernels, all threads coincide. The chief challenge in switching among kernels is to minimize the associated cost. Given that objective, the result of the calculations of each kernel is stored in the global memory of the GPU, and no data would be exchanged at each switching step between the host and the device. Consequently, the time required for switching among kernels is negligible. The process of each kernel is described below.

In the proposed method, first, the fitness of the chromosomes in the initial population is calculated concurrently by the first parallel kernel. In this kernel, each thread calculates the fitness of a chromosome. Next, to produce a new generation, the operators of crossover, mutation, and fitness functions are applied concurrently by the second kernel. In this kernel, each thread should form a new child chromosome using dissimilar operators of the GA. The third parallel kernel is responsible for the selection operator. This operator selects the finest chromosomes of the current generation to be passed to the next generation. In this kernel, each selects a chromosome. This cycle is repeated, and at the end of each cycle, the condition of the termination of the generation of the new population is checked. If the condition of the termination of rounds is met, this cycle stops, and the best answer is returned from the populations as a result. Otherwise, the second and third kernels that perform the creation of a new generation, as well as the selections, will continue running to the end of the pre-specified number of iterations. In the next section, we inspect the performance of the proposed method on TSP.

4. IMPLEMENTATION AND PERFORMANCE EVALUATION

In this section, first, the hardware characteristics of the computer system and the values of the diverse parameters of the GA, are described. Then, the performance of the proposed kernels is examined based on several metrics.

4. 1. Conditions and Environments of Implementation The experiments were executed on an Intel (R) Core i5-7600 3.50GHz computer with 8 GB of random-access memory that was equipped with an NVIDIA GeForce GTX 960 graphics card. This GPU has 1024 cores, and its base and boost clocks are 1,127MHz, and 1,178MHz, respectively. The frameworks used in this implementation were created by C++ CUDA 8.0 (V8.0.61) for many-core GPU and TBB version 2018 for multi-core CPU. The number of threads in the CUDA platform and multi-core CPUs was set equal to the number of chromosomes and the number of cores, respectively. Hence, the number of cores in the TBB-based parallel kernel is remarkable. To investigate this effect, we executed the TBB-based parallel kernels on dual-core and quad-core CPUs.

Crossover and mutation probability were set to 0.8 and 0.02, respectively. The probability of selection operator, providing that the operator may select a sample with less fitness, is 0.8. Note that the crossover operator is a single-parent and single-point one, and the mutation operator is of the movement type.

The standard datasets of PKA379, rbx711, and xit1083 from VLSI data were used in implementing the TSP. The datasets consisted of 379, 711, and 1083 geographical regions of cities [38]. TSP was solved with 1024 to 16384 populations by using 100, 200 to 3000 rounds of the GA. The number of offsprings produced in the crossover was between 10%-50% of the size of the preliminary population. In the next section, we analyze the results which were acquired from ten times repeatition of the experiments.

4. 2. Performance Evaluation To assess the proposed methods and compare their performances, we analyze the influence of parameters like population size, number of generations, number of crossover-mutation, and chromosomes size on the performance of each method.

Evaluation metrics are the running time of the GA in the proposed methods and the speedup of the parallel versions of the serial method. First, we examine the cost of switching between the proposed kernels over the CUDA platform in diverse scenarios. Table 1 shows the time required for solving TSP with 370 cities using the proposed set of tertiary kernels with different sizes of population and different numbers of generations. For each case, the time required for switching among kernels, as well as their execution time, is reported. Table 2 shows the switching and execution time of the kernels in 100 generations, with dissimilar numbers of cities and different sizes of the population. In these two tables, the ratio of offsprings is 50% of the population. As explained in Section 3, since there is no data transmission between kernels and generations (from GPU to CPU and vice versa), growing the number of generations has not any effect on the switching time. But by increasing the number of cities and the size of the population, switching time rises due to the transfer cost of the preliminary population from CPU to GPU. However, the switching time is small compared with the kernel execution time. The sum of the switching time and the time of kernel computations represent the total computation time for a parallel kernel in the CUDA platform.

The plots in Figure 2 demonstrate the effect of increasing the size of the initial population and the number of generations on the running time of the serial method, the parallel method on the CUDA platform, and the parallel methods formed by exploiting TBB on dual-core/quad-core CPUs. In this experiment, the size of the new population generated from the crossover operation is 50 % of the initial population. By growing the number of generations, the running time of TSP is grown in all methods.

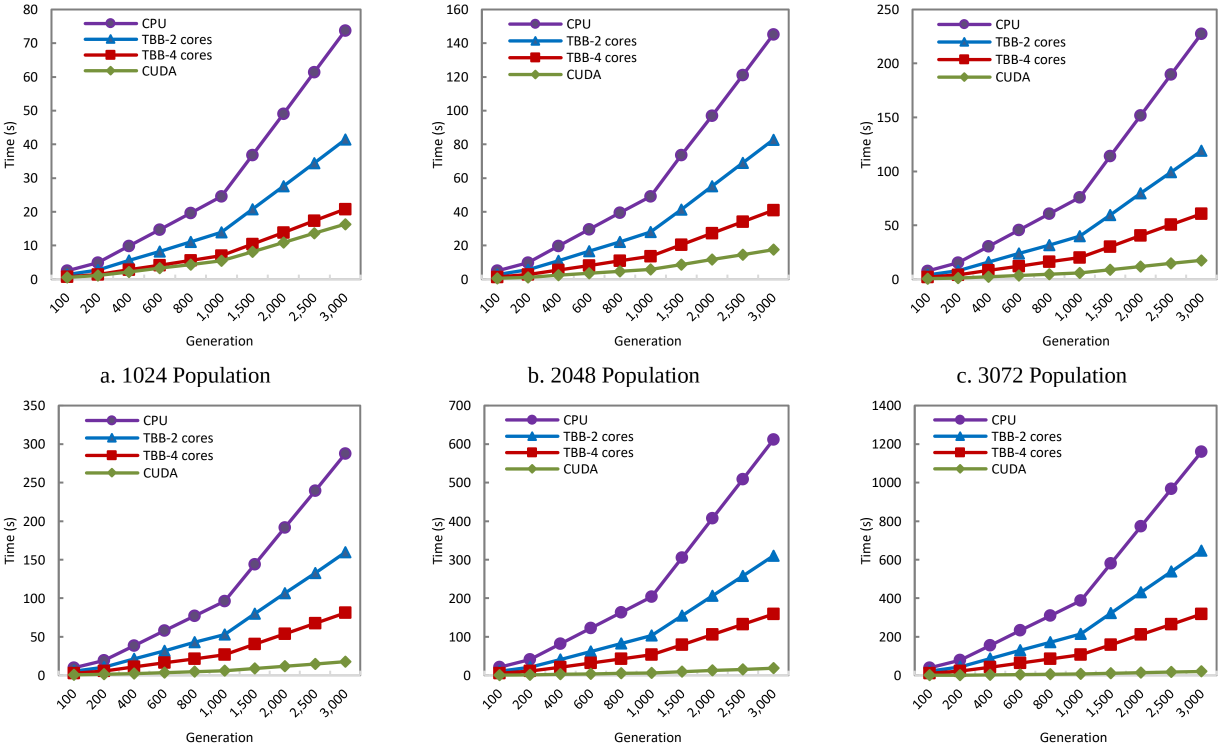
A notable point in this experiment is the effect of the size of the population on the execution time of the CUDA-based parallel code as compared to the TBB-

TABLE 1. Switching and kernel computation time (ms) in CUDA- 379 cities (pka379)

Generation	Population							
	1024		5120		10240		16384	
	Switch	Kernel	Switch	Kernel	Switch	Kernel	Switch	Kernel
100	0.195128	548.545	0.78577	600.4875	1.50617	573.9525	2.3287	672.3125
400	0.18966	2181.9425	0.79366	2388.99	1.52645	2261.02	2.37945	2655
1000	0.196124	5450.35	0.78552	5960.375	1.45712	5638.075	2.29706	6624.425
3000	0.194644	16340.6	0.78178	17871.925	1.48930	16972.8	2.32974	19876.325

TABLE 2. Switching and kernel computation time (ms) in CUDA- 100 generations

Cities	Population							
	2048		9216		32768		131072	
	Switch	Kernel	Switch	Kernel	Switch	Kernel	Switch	Kernel
379 (pka379)	0.341525	588.4825	1.3636	567.645	5.520225	1756.985	19.6385	6283.225
711 (rbx711)	0.586467	1089.205	2.5615	1164.0825	8.821775	3291.925	34.5882	11756.775
1083(xit1083)	0.866655	1677.325	3.8843	1787.5275	12.68052	5065.65	52.2602	18439.525



a. 1024 Population

b. 2048 Population

c. 3072 Population

d. 4096 Population

e. 8192 Population

f. 16384 Population

Figure 2. The running time of the algorithm over different numbers of generations and different sizes of population

based parallel codes running on dual-core and quad-core machines. As shown in Figure 2(a), the running time of CUDA-based kernel is worse than TBB-based kernel on dual-core and quad-core CPUs. In this experiment, all resources of CPUs have been completely exploited while in the GPU, only up to 1024 threads have been used.

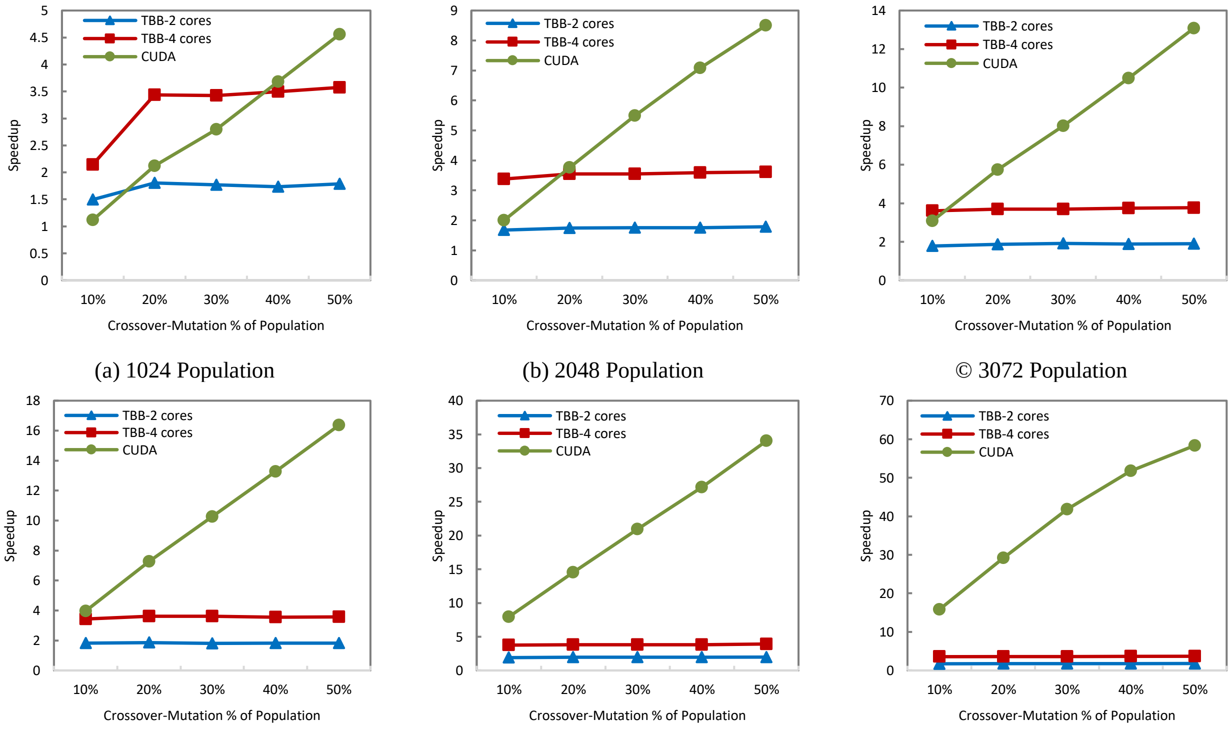
Growing the size of populations increases the concurrency, and improves the performance of CUDA kernels. For example, in the case of having a population with the size of 4096, the CUDA-based parallel TSP code has been better than TBB-based TSP. The degree of this dominance reaches a maximum when the size of the population hits 16384. In this state, the maximum number of GPU resources, or equivalently the maximum number of threads, have been used synchronously.

Figure 3 demonstrates the plots of speedup of three parallelization methods, namely TBB-based TSP kernel on dual-core and Quad-core CPUs and CUDA-based GPU kernel concerning the sequential code in different sizes of population and offsprings created by crossover. As a common setting in this experiment, the upper bound of the number of generations of the GA is set to 100. In all cases, the speedup of TBB on four cores is more than the speedup of TBB on two cores. In addition, the overall rate of the speedup has not been changed with the size of the population and the number of offsprings. This is while in the CUDA-based TSP kernel, the speedup has changed with both of these parameters. The reason is that both parameters affect the utilization of the resources of GPU. The maximum speedup values obtained in the execution of TBB-based TSP on dual-core and quad-core CPUs are 1.98 and 3.92, respectively, while the maximum speedup of CUDA-based TSP on 1024 cores is 58.35.

The effect of changing the number of chromosomes on the execution time of the parallel TSP codes is illustrated in Figure 4. In the corresponding experimentation, the number of generations and the number of offsprings resulting from each crossover process is a constant value while the size of the population is variable. In this experiment, three different datasets with 379, 711, and 1083 cities have been used. The size of offsprings is set to be 50% of the primary population for 100 generations.

As shown in Figure 4, in all cases, the CUDA-based parallel code is the fastest as compared to the other methods. This state hits when CUDA computes the GA solution of TSP with 16,384 population; in this case, each thread computes a chromosome. As the population grows, each thread should compute more than one chromosome, which increases the running time. This is shown in the speedup plots of Figure 4. Another notable point in this experiment is the impotence of changes in the number of genes on a chromosome (number of cities) in the overall execution time of the three parallel kernels of the GA.

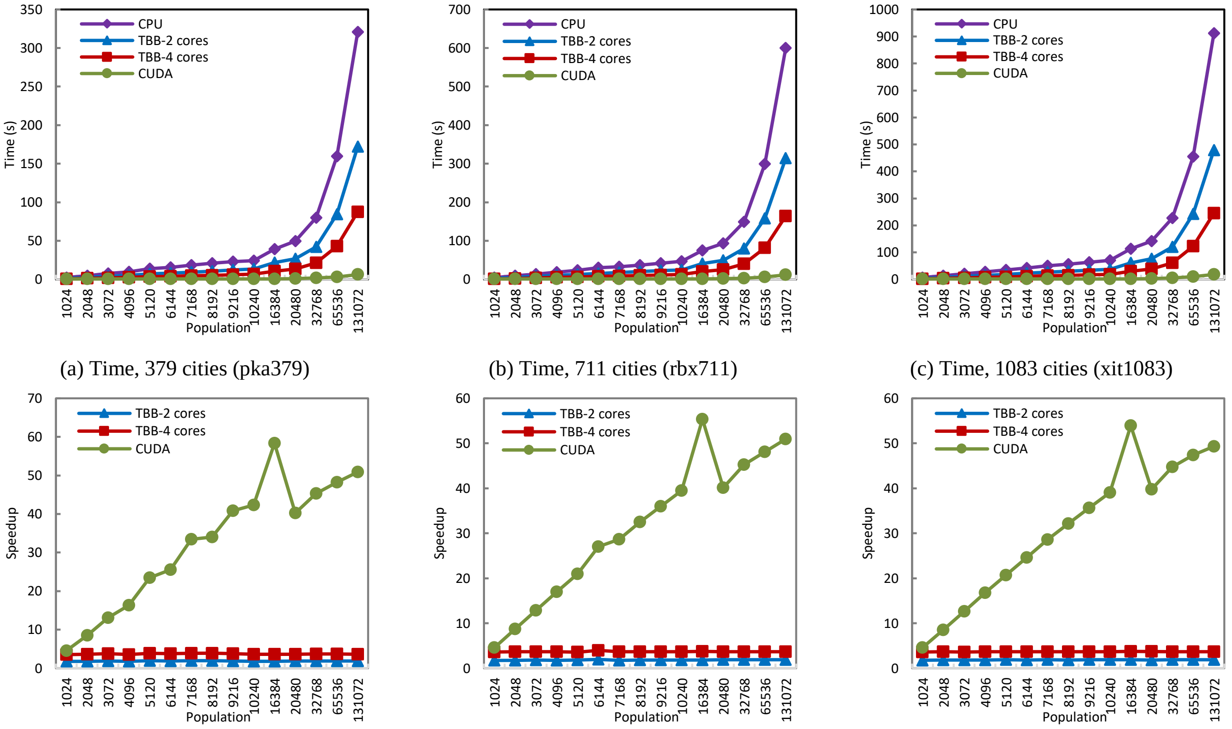
Regarding the ability of CUDA to associate the maximum number of threads for computations of a kernel, the use of CUDA seems to produce higher performance; but, in the following, the results of our experiments show that the TBB-based implementation exploits the computational resources of the system more efficiently.



(a) 1024 Population (b) 2048 Population © 3072 Population

(d) 4096 Population (e) 8192 Population (f) 16384 Population

Figure 3. The speedup of the parallel methods for different ratios of offsprings on diverse populations in a TSP with 379 cities



(a) Time, 379 cities (pka379)

(b) Time, 711 cities (rbx711)

(c) Time, 1083 cities (xit1083)

(d) Speedup, 379 cities (pka379)

(e) Speedup, 711 cities (rbx711)

(f) Speedup, 1083 cities (xit1083)

Figure 4. The impact of population size on the speedup of parallel methods

4. 3. Efficiency Evaluation and Comparison

To intuitively show the effectiveness of the proposed parallelization method, the efficiency of the proposed kernels is computed and compared in Table 3 with the recent advanced methods. As shown in Table 3, the efficiency of the proposed parallel GA on a GPU with 1024 cores is 0.057, which is higher than all opponents. Also, the efficiency of the proposed parallel GA code on a quad-core CPU using the TBB platform is the highest as compared to the other parallel GA solutions of TSP on multi-core CPUs. Also, the efficiency of multi-core parallelization of the GA solution of TSP, using the TBB platform is 0.9975, which is significantly more than that of the CUDA-based parallelization. This result shows that the proposed parallelization method not only outstands any present many-core parallelization of GA solution of TSP, but also confirms that the proposed method could more effectively use the computational resources of the multi-core CPUs and consequently reach higher efficiency.

TABLE 3. Comparing the efficiency of state-of-art parallelizations of GA solution of TSP

Method	Year	Reference	# of Cores	Speed up	Efficiency
Multi-core	2013	Zhu [17]	4	2.55	0.6375
	2019	Saxena [26]	4	2	0.5
	2020	**Proposed Method**	**4**	**3.99**	**0.9975**
Many-core	2011	Fujimoto [18]	240	13.3	0.055
	2011	Chen [19]	448	1.44	0.003
	2016	Kang [21]	2048	6.014	0.003
	2017	Moumen [22]	1280	25.07	0.019
	2019	Saxena [26]	128	5.66	0.044
	2020	**Proposed Method**	**1024**	**58.35**	**0.057**

5. CONCLUSION

In this paper, a novel method for efficient parallelization of genetic algorithms on multi-core and many-core systems was presented. The proposed method efficiently executes parallel kernels on the multi-core and many-core systems, each corresponding to a different operator of GA, by using TBB and CUDA platforms, respectively. The proposed method was tested for parallelizing a GA-based solution of the TSP on CUDA and TBB platforms with the same settings, including the same number of primary population and generations as well as the same ratio of population created by crossover and mutation operators on the same data set. The performance of these two platforms was assessed based on different metrics including the running time and speedup of the parallel GA over each of them.

From the results, we have drawn the following conclusions that crucially represent the real and novel contribution of our work. First, the highest speedup of the parallel algorithm on the GPU, the quad-core, and the dual-core processor are 58.35, 3.99, and 1.99, respectively. Second, the performance of a parallel GA on a GPU-like many-core processor is much higher than that of a multi-core processor, but in a low initial population, parallelization resources in multi-core processors are more efficiently utilized than in the GPU-like many-core systems. Third, the efficiency of the proposed parallelization of the GA solution of TSP on a CUDA-based many-core platform is the highest as compared to state-of-art parallel solutions. Fourth and the most important finding is that the proposed TBB-based parallelization of the GA solution of TSP achieves the highest level of efficiency in exploiting the computational resources of the system for parallel execution of GA.

The CPU/GPU clusters have recently been considered as high-performance accelerators for computation-intensive programs. Therefore, future studies would study how to adapt the proposed parallelization method of GA to best use the resources of the GPU cluster computations.

6. REFERENCES

1. Fathollahi-Fard, A.M., Hajiaghaei-Keshteli, M., and Tavakkoli-Moghaddam, R., 'The Social Engineering Optimizer (Seo)', ***Engineering Applications of Artificial Intelligence***, (2018), Vol. 72, 267-293. https://doi.org/10.1016/j.engappai.2018.04.009
2. Fard, A.F. and Hajiaghaei-Keshteli, M., 'Red Deer Algorithm (Rda); a New Optimization Algorithm Inspired by Red Deers' Mating', in, International Conference on Industrial Engineering, IEEE., (2016). https://doi.org/10.1007/s00500-020-04812-z
3. Mohammadzadeh, H., Sahebjamnia, N., Fathollahi-Fard, A., and Hahiaghaei-Keshteli, M., 'New Approaches in Metaheuristics to Solve the Truck Scheduling Problem in a Cross-Docking Center', ***International Journal of Engineering-Transactions B: Applications***, 2018, Vol. 31, No. 8, 1258-1266. https://doi.org/10.5829/ije.2018.31.08b.14
4. Fathollahi-Fard, A., Hajiaghaei-Keshteli, M., and Tavakkoli-Moghaddam, R., 'A Lagrangian Relaxation-Based Algorithm to Solve a Home Health Care Routing Problem', ***International Journal of Engineering Transactions A: Basics,*** Vol. 31, No. 10, (2018), 1734-1740. https://doi.org/10.5829/ije.2018.31.10a.16
5. Hajiaghaei-Keshteli, M., Abdallah, K., and Fathollahi-Fard, A., 'A Collaborative Stochastic Closed-Loop Supply Chain Network Design for Tire Industry', ***International Journal of Engineering Transactions A: Basics,*** Vol. 31, No. 10, (2018), 1715-1722. https://doi.org/10.5829/ije.2018.31.10a.14
6. López-González, A., Campaña, J.M., Martínez, E.H., and Contro, P.P., 'Multi Robot Distance Based Formation Using Parallel Genetic Algorithm', ***Applied Soft Computing***, (2020), Vol. 86, p. 105929. https://doi.org/10.1016/j.asoc.2019.105929
7. Munroe, S., Sandoval, K., Martens, D.E., Sipkema, D., and Pomponi, S.A., 'Genetic Algorithm as an Optimization Tool for the Development of Sponge Cell Culture Media', ***In Vitro Cellular & Developmental Biology-Animal***, Vol. 55, No. 3, (2019), 149-158. https://doi.org/DOI: 10.1007/s11626-018-00317-0
8. Sin, I.H. and Do Chung, B., 'Bi-Objective Optimization Approach for Energy Aware Scheduling Considering Electricity Cost and Preventive Maintenance Using Genetic Algorithm', ***Journal of Cleaner Production***, Vol. 244, (2020), 118869. https://doi.org/10.1016/j.jclepro.2019.118869
9. Yasmin, S.: 'Linear Colour Image Processing in Hypercomplex Algebra Guided by Genetic Algorithms', University of Essex, 2019
10. Lima, A.A., de Barros, F.K., Yoshizumi, V.H., Spatti, D.H., and Dajer, M.E., 'Optimized Artificial Neural Network for Biosignals Classification Using Genetic Algorithm', J***ournal of Control, Automation and Electrical Systems***, Vol. 30, No. 3, (2019), 371-379. https://doi.org/10.1007/s40313-019-00454-1
11. Rajagopalan, A., Modale, D.R., and Senthilkumar, R., Optimal Scheduling of Tasks in Cloud Computing Using Hybrid Firefly-Genetic Algorithm', Advances in Decision Sciences, Image Processing, Security and Computer Vision, (Springer, 2020) ISBN: 978-3-030-24318-0. https://doi.org/10.1007/978-3-030-24318-0_77
12. Rajesh, K., Visali, N., and Sreenivasulu, N., Optimal Load Scheduling of Thermal Power Plants by Genetic Algorithm', *Emerging Trends in Electrical, Communications, and Information Technologies*, (Springer, 2020) ISBN: 978-981-13-8942-9. https://doi.org/10.1007/978-981-13-8942-9_33
13. Arif, M.H., Li, J., Iqbal, M., and Liu, K., 'Sentiment Analysis and Spam Detection in Short Informal Text Using Learning Classifier Systems', ***Soft Computing***, Vol. 22, No. 21, (2018), 7281-7291. https://doi.org/10.1007/s00500-017-2729-x
14. Talbi, E.-G., 'A Unified View of Parallel Multi-Objective Evolutionary Algorithms', ***Journal of Parallel and Distributed Computing***, Vol. 133, (2019), 349-358. https://doi.org/10.1016/j.jpdc.2018.04.012
15. Nayak, S. and Panda, M., Hardware Partitioning Using Parallel Genetic Algorithm to Improve the Performance of Multi-Core Cpu', ***Advances in Intelligent Computing and Communication***, (Springer, 2020) ISBN: 978-981-15-2774-6
16. Giap, C.N. and Ha, D.T., 'Parallel Genetic Algorithm for Minimum Dominating Set Problem', in, Computing, Management and Telecommunications (ComManTel), 2014 International Conference on, (IEEE, 2014). https//doi.org/10.1109/ComManTel.2014.6825598
17. Zhu, J. and Li, Q., 'Application of Hybrid Mpi+ Tbb Parallel Programming Model for Traveling Salesman Problem', in, Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing, (IEEE, 2013). https://doi.org/10.1109/GreenCom-iThings-CPSCom.2013.408
18. Fujimoto, N. and Tsutsui, S., 'A Highly-Parallel Tsp Solver for a Gpu Computing Platform', in, International Conference on Numerical Methods and Applications, (Springer, 2010). https://doi.org/10.1007/978-3-642-18466-6_3
19. Chen, S., Davis, S., Jiang, H., and Novobilski, A., Cuda-Based Genetic Algorithm on Traveling Salesman Problem', Computer and Information Science 2011, (Springer, 2011). https://doi.org/10.1007/978-3-642-21378-6_19
20. Sánchez, L.N.G., Armenta, J.J.T., and Ramírez, V.H.D., 'Parallel Genetic Algorithms on a Gpu to Solve the Travelling Salesman Problem', Difu100ci@ Revista en Ingeniería y Tecnología, UAZ, 2015, 8, (2). https://doi.org/10.1007/978-3-662-45049-9_96

21. Kang, S., Kim, S.-S., Won, J., and Kang, Y.-M., 'Gpu-Based Parallel Genetic Approach to Large-Scale Travelling Salesman Problem', ***The Journal of Supercomputing***, 2016, Vol. 72, No. 11, 4399-4414. https://doi.org/10.1007/s11227-016-1748-1

22. Moumen, Y., Abdoun, O., and Daanoun, A., 'Parallel Approach for Genetic Algorithm to Solve the Asymmetric Traveling Salesman Problems', in, Proceedings of the 2nd International Conference on Computing and Wireless Communication Systems, (ACM, 2017) https://doi.org/10.1145/3167486.3167510

23. Cekmez, U., Ozsiginan, M., and Sahingoz, O.K., 'Adapting the Ga Approach to Solve Traveling Salesman Problems on Cuda Architecture', in, Computational Intelligence and Informatics (CINTI), 2013 IEEE 14th International Symposium on, (IEEE, 2013) https://doi.org/10.1109/CINTI.2013.6705234

24. Radford, D. and Calvert, D., 'A Comparative Analysis of the Performance of Scalable Parallel Patterns Applied to Genetic Algorithms and Configured for Nvidia Gpus', ***Procedia Computer Science***, Vol. 114, (2017), 65-72. https://doi.org/10.1016/j.procs.2017.09.009

25. Li, C.-C., Lin, C.-H., and Liu, J.-C., 'Parallel Genetic Algorithms on the Graphics Processing Units Using Island Model and Simulated Annealing', ***Advances in Mechanical Engineering***, Vol. 9, No. 7, (2017), 12-25. https://doi.org/10.1177%2F1687814017707413

26. Saxena, R., Jain, M., Sharma, D., and Jaidka, S., 'A Review on Vanet Routing Protocols and Proposing a Parallelized Genetic Algorithm Based Heuristic Modification to Mobicast Routing for Real Time Message Passing', ***Journal of Intelligent & Fuzzy Systems***, Vol. 36, No. 3, (2019), 2387-2398. https://doi.org/10.3233/JIFS-169950

27. NVIDIA. NVIDIA CUDA (Compute Unified Device Architecture) Programming Guide, Available: http://docs.nvidia.com/cuda/pdf/CUDA_C_Programming_Guide.pdf, (Accessed July 2020).

28. Jam, S., Shahbahrami, A., and Ziyabari, S., 'Parallel Implementation of Particle Swarm Optimization Variants Using Graphics Processing Unit Platform', ***International Journal of Engineering Transactions A: Basics***, Vol 30, No. 1, (2017), 48-56. https://doi.ir/10.5829/idosi.ije.2017.30.01a.07

29. Yip, C.M. and Asaduzzaman, A., 'A Promising Cuda-Accelerated Vehicular Area Network Simulator Using Ns-3', in, *P*erformance Computing and Communications Conference (IPCCC), 2014 IEEE International, (IEEE, 2014) https://doi.org/10.1109/PCCC.2014.7017048

30. Kim, C.G., Kim, J.G., and Lee, D.H., 'Optimizing Image Processing on Multi-Core Cpus with Intel Parallel Programming Technologies', ***Multimedia Tools and Applications***, Vol. 68, No. 2, (2014), 237-251. https://doi.org/10.1007/s11042-011-0906-y

31. Reinders, J. 'Intel threading building blocks: outfitting C++ for multi-core processor parallelism', O'Reilly Media. Inc, 2007.

32. Hougardy, S. and Wilde, M., 'On the Nearest Neighbor Rule for the Metric Traveling Salesman Problem', ***Discrete Applied Mathematics***, (2014). https://doi.org/10.1016/j.dam.2014.03.012

33. Groba, C., Sartal, A., and Vázquez, X.H., 'Solving the Dynamic Traveling Salesman Problem Using a Genetic Algorithm with Trajectory Prediction: An Application to Fish Aggregating Devices', ***Computers & Operations Research***, Vol. 56, (2015), 22-32. https://doi.org/10.1016/j.cor.2014.10.012

34. Hussain, A. and Muhammad, Y.S., 'Trade-Off between Exploration and Exploitation with Genetic Algorithm Using a Novel Selection Operator', ***Complex & Intelligent Systems***, (2019), 1-14. https://doi.org/0.1007/s40747-019-0102-7

35. Hassanat, A., Prasath, V., Abbadi, M., Abu-Qdari, S., and Faris, H., 'An Improved Genetic Algorithm with a New Initialization Mechanism Based on Regression Techniques', ***Information***, Vol. 9, No. 7, (2018), 167. https://doi.org/10.3390/info9070167

36. Doughabadi, M.H., Bahrami, H., and Kolahan, F., 'Evaluating the Effects of Parameters Setting on the Performance of Genetic Algorithm Using Regression Modeling and Statistical Analysis', ***Journal of Industrial Engineering***, *University of Tehran*, (2011), 61-68.

37. Contreras-Bolton, C. and Parada, V., 'Automatic Combination of Operators in a Genetic Algorithm to Solve the Traveling Salesman Problem', ***PloS One***, Vol. 10, No. 9, (2015), e0137724-e0137724. https://doi.org/10.1371/journal.pone.0137724

38. VLSI-TSP-Collection, Available: http://www.math.uwaterloo.ca/tsp/vlsi/index.html, (Accessed July 2020).

Persian Abstract

چکیده

موازی‌سازی کارآمد الگوریتم‌های ژنتیک روی بسترهای multi-threading یا many-threading موجود به دلیل دشواری زمانبندی منابع سخت‌افزاری با توجه به همزمانی نخ‌ها، موضوعی چالش برانگیز است. در این مقاله، برای حل این مشکل یک روش جدید ارائه شده است که الگوریتم ژنتیک را با طراحی سه کرنل همزمان که هرکدام تعدادی از عملگرهای موثر وابسته به یکدیگر از الگوریتم ژنتیک را اجرا می‌کنند موازی می‌سازد. روش پیشنهادی را می‌توان به راحتی با استفاده از بسترهای Compute Unified Device Architecture (CUDA) و Threading Building Blocks (TBB) برای اجرا روی پردازنده‌های many-core و multi-core تطبیق داد. برای استفاده بهینه از منابع ارزشمند این نوع پردازنده‌ها در اجرای موازی الگوریتم ژنتیک، پردازش‌های نخی که یکی از کرنل‌های سه‌گانه را اجرا می‌کنند، توسط مکانیسم جابجایی پرسرعتی هماهنگ می‌شوند. روش پیشنهادی، برای موازی‌سازی مسئله فروشنده دوره‌گرد مبتنی بر الگوریتم ژنتیک با استفاده از پلتفرم‌های CUDA و TBB با تنظیمات یکسان آزمایش شده است. نتایج، کارایی روش پیشنهادی در موازی‌سازی الگوریتم ژنتیک را روی واحد پردازش گرافیکی و همچنین روی واحد پردازش مرکزی تایید می‌کنند. علاوه بر این، در مسئله‌های ژنتیک با جمعیت اولیه متوسط، با اینکه زمان جابجایی بین کرنل‌های واحد پردازش گرافیکی بسیار ناچیز است اما، الگوریتم ژنتیک موازی مبتنی بر TBB از منابع بطور موثرتری استفاده می‌کند.

International Journal of Engineering

Journal Homepage: www.ije.ir

Repeated Record Ordering for Constrained Size Clustering

R. Mortazavi*

School of Engineering, Damghan University, Damghan, Iran

PAPER INFO

Paper history:
Received 15 March 2020
Received in revised form 09 April 2020
Accepted 17 June 2020

Keywords:
Constrained Clustering
Microaggregation
Data Privacy

ABSTRACT

One of the main techniques used in data mining is data clustering, which has many applications in computer science, biology and social sciences. Constrained clustering is a type of clustering in which side information provided by the user is incorporated into current clustering algorithms. One of the well researched constrained clustering algorithms is called microaggregation. In a microaggregation technique, the algorithm divides the dataset into groups containing at least k members, where k is a user-defined parameter. The main application of microaggregation is in Statistical Disclosure Control (SDC) for privacy preserving data publishing. A microaggregation algorithm is qualified based on the sum of within-group squared error, SSE. Unfortunately, it has been proven that the optimal microaggregation problem is NP-Hard in general, but the special univariate case can be solved optimally in polynomial time. Many heuristics exist for the general case of the problem that are founded on the univariate case. These techniques order multivariate records in a sequence. This paper proposes a novel method for record ordering. Starting from a conventional clustering algorithm, the proposed method repeatedly puts multivariate records into a sequence and then clusters them again. The process is repeated until no improvement is achieved. Extensive experiments have been conducted in this research to confirm the effectiveness of the proposed method for different parameters and datasets.

doi*: 10.5829/ije.2020.33.07a.13*

1. INTRODUCTION

Nowadays, there is a considerable demand for real-world datasets in various data mining tasks. However, the privacy of involved entities usually agitates data owners about the usage of such information [1, 2]. Privacy preserving data publishing is the task that addresses the problem. The problem is also investigated in research communities of the Internet of Things (IoT) [3, 4] and Statistical Disclosure Control (SDC). Usually, the privacy requirement in terms of Disclosure Risk (DR) is formalized using a computational privacy model, which can then be realized by an implementation method. The main idea of different solutions is based on changing the original data records to preserve the privacy of involved entities. Such changes decrease the utility of published data which is stated by Information Loss (IL). It is desired to minimize both the competing indices of DR and IL, which is a challenging multi-objective optimization task [5].

One of the most famous computational privacy models is called k-anonymity [6]. In a k-anonymous dataset, for each set of identifying attributes, there exist at least k records. Therefore, an intruder who knows some attributes of an entity cannot limit its record data to a small group, i.e., a group with less than k members. Microaggregation is a perturbative approach to realize k-anonymity. It was initially developed for numerical data volumes, while it can also be used for other types of datasets [7]. A microaggregation technique tries to cluster the dataset records into groups with at least k members and then aggregates them into their centroids. The centroids are then substituted for the original records and published for public usage. In other words, the original entries are masked using their associated centroids. The replacement decreases the details of the published values, which results in IL. For microaggregation algorithms, IL is usually quantified in terms of the sum of within-group

*Corresponding Author Institutional Email: *r_mortazavi@du.ac.ir* (R. Mortazavi)

Please cite the article as: R. Mortazavi, Repeated Record Ordering for Constrained Size Clustering, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1266-1273

squared error (SSE).

Unfortunately, it has been proven that given the privacy parameter k, the optimal microaggregation problem is NP-hard in general [8]. Still, the univariate instance can be optimally solved in polynomial time using the Mukherjee and Hansen Microaggregation (MHM) algorithm [9]. Some heuristic approaches try to map the general multivariate microaggregation problem to the univariate case [10, 11]. For example, the NPN-MHM algorithm [10] traverses all records in a Nearest Point Next fashion starting from the farthest record from the dataset centroid to put them in a sequence and then applies MHM on the output ordering. Similarly, MDAV-MHM [10] clusters the dataset using a traditional microaggregation algorithm, Maximum Distance to Average Vector (MDAV) [7] and then visits all records, group by group. Mortazavi et al. proposed Improved MHM (IMHM[‡]) [11], which accelerates MHM and uses it in multivariate microaggregation. However, existing techniques are not general and usually suffer from increased IL when the dataset has an internal structure and is naturally clustered. For instance, NPN-MHM is more useful in anonymizing datasets with very separated clusters, but for clustered data with moderate gaps or skewed data, the CBFS–MHM and MDAV–MHM produce the best results [10]. Similarly, IMHM [11] is more successful when k is small and the dataset is clustered, but for homogeneous datasets, it produces more useful anonymized versions when k is large. Additionally, comparing the results of some recent heuristics with proved lower bounds of the problem [12] shows large gaps in some cases.

The primary contribution of this paper is to propose an innovative ordering technique in which multivariate data records are ordered in a sequence while considering the internal structure of the dataset using a conventional clustering method. Additionally, it is shown that the process of converting the output of a clustering algorithm to a sequence can be repeated that in turn results in considerable improved IL. Extensive experiments in this research show the advantage of the proposed method in terms of data utility in comparison with similar previous techniques.

The remainder of the paper is structured as follows. Section 2 formalizes the microaggregation problem. Section 3 reviews some related microaggregation algorithms. Section 4 describes the proposed method. Experimental results are reported in Section 5. Finally, Section 6 concludes the paper.

2. MICROAGGREGATION PROBLEM

In this section, the problem of microaggregation is formalized. Assume a dataset T of n numerical records in a d-dimensional space, i.e., $T = \{x_1, x_2, \ldots, x_n\}$ where $x_i \in \mathbb{R}^d$. Given an input value k as the privacy parameter, the microaggregation algorithm aims to partition the whole dataset T into c non-overlapping groups $G_1, \ldots, G_c$ each with at least k members. The objective of microaggregation techniques as an optimization problem is to minimize the SSE, which aims to obtain clusters of similar records. This measure is shown in Equation (1).

$$SSE = \sum_{p=1}^{c} \sum_{j=1}^{|G_p|} \left(x_{pj} - \overline{x_p}\right)^T \left(x_{pj} - \overline{x_p}\right) \quad (1)$$

In Equation (1), x_{pj} is record j of group G_p, and $\overline{x}_p$ denotes the centroid of G_p, i.e., $\overline{x}_p = \sum_{j=1}^{|G_p|} x_{pj} / |G_p|$. The value is usually divided by the Sum of Squares Total (SST) to normalize IL. SST is related to the dataset itself and is invariant to the microaggregation algorithm or the privacy model parameters. It is formulated in Equation (2).

$$SST = \sum_{i=1}^{n} (x_i - \bar{x})^T (x_i - \bar{x}) \quad (2)$$

In Equation (2), $\bar{x}$ is the centroid of the whole dataset, i.e., $\bar{x} = \sum_{i=1}^{n} x_i / n$. The normalized measure $IL = SSE/SST * 100\%$ is always between 0 and 100%, where lower values of IL indicate less utility degradation due to microaggregation.

3. RELATED WORKS

It was shown by Domingo-Ferrer and Mateo-Sanz that in an optimal constrained size clustering, each group contains at most $2k - 1$ records [13]. A polynomial-time technique was developed by Hansen and Mukherjee for univariate microaggregation that is called MHM [9]. The MHM first sorts univariate records and then creates a directed acyclic graph in which each arc in the graph matches a valid group that may be a cluster in the optimal solution. The authors showed that the optimal univariate microaggregation problem is reduced to computing the shortest path in the graph. A cluster exists in the optimal partition if its equivalent arc is in the computed shortest path. The complexity of the technique is $O(\max(n \log n, k^2 n))$. Mortazavi *et al.* introduced an improved implementation of the MHM called IMHM [11] that makes use of incremental weight computation of graph arcs to improve the complexity of graph construction to $O(kn)$ operations. The authors generalized the application of IMHM for multivariate datasets in an iterative optimization process, but the experiments show that the user has to carry out different experiments with multiple parameters, which is a time-consuming task.

[‡] The pseudo-code of the IMHM is described briefly in the Appendix.

The optimal property of MHM provides a hopeful tactic to solve the challenging problem of the multivariate microaggregation. However, sorting multivariate records for optimal microaggregation is not well-defined. Therefore, different heuristics are devised in literature to sequence multivariate records. Domingo-Ferrer *et al.* [10] proposed some heuristics, such as the Nearest Point Next MHM (NPN-MHM), MDAV-MHM, and Centroid-Based Fixed-Size MHM (CBFS-MHM) to order records and form a sequence of them. Then, MHM is applied to records on the path. However, their reports show that their approach is usually far from optimal, especially for clustered datasets. Monedero *et al.* used two projection methods, i.e., Principal Component Analysis (PCA) and Z-score, to reduce the dimension of the underlying dataset to one [14]. In the PCA technique, the first principal component of the dataset is utilized to sort data records. The Z-score algorithm orders multivariate records based on the sum of their Z-scores. Again, there is a significant distance to optimal solutions in both methods. Soria-Comas and Domingo-Ferrer presented a method to satisfy the differential privacy requirement [15] through univariate microaggregation [9]. Additionally, Mortazavi and Jalili introduced the Fast Data-oriented Microaggregation algorithm (FDM) [16] that produces an optimal assignment of records with respect to their Travelling Salesman Problem (TSP[§]) tour for a continuous range of the privacy parameter k. However, the running time to compute the TSP tour of multivariate records is considerable. More recently, Khomnotai *et al.* devised the Iterative Group Decomposition (IGD) technique [17] to refine the solution of a microaggregation algorithm by either shrinking or decomposing its clusters. Unfortunately, none of the mentioned methods can achieve near-optimal solutions. They are usually useful for particular datasets with pre-specified data distribution or very limited ranges of k. Moreover, the methods in the literature are somehow hard-coded with complex parameters that limit their flexibility in practice. It is therefore desired to devise a general method that can produce more useful anonymized datasets, which is addressed in the next section.

4. PROPOSED MICROAGGREGATION ALGORITHM

In this section, the Repeated record Ordering heuristic for multivariate Microaggregation, RepOrdMic is detailed. Briefly, the algorithm accepts an initial clustering of records and traverses all records group-by-group to complete a sequence (ordering) of all records. In each group, all records are visited using a TSP heuristic, and then the nearest unexplored group is processed. After all records were added to the sequence, the IMHM is utilized to produce a (constrained) clustering. The process is repeated until no significant improvement is achieved. Algorithm 1 shows the pseudo-code of the proposed method[**]. The algorithm accepts the normalized dataset T, the privacy parameter k, and an initial clustering label lbl_{in} as inputs, and produces the perturbation error SSE and labels of assigned records to constrained size groups lbl_{out} as outputs. The function initially creates an empty sequence Seq to store the total ordering of multivariate records in Step 1. Step 2 finds the farthest record x_f from the whole dataset centroid and then stores it in the current record x_c in Step 3. In Steps 4 to 9, all records in T are visited group-by-group and the order of visiting them is saved in Seq. In Step 4, the group label of x_c is considered as the current group, G_c. If the current group has only one member, the algorithm continues to process other groups (Step 6-1). Otherwise, the algorithm looks for the most distant point from the current record x_c among current group members and adds it to the end of Seq. Other records in the current group will be inserted between these two group members. The process utilizes an idea inspired by the nearest insertion heuristic to solve TSP for entering all records in the current group to the Seq. Steps 7 to 9 repeatedly choose an unseen record with minimum distance to its nearest neighbor among the current group members in Seq. Then, they insert it between the two consecutive records (Figure 1) for which such an insertion causes the minimum increase in the total sequence length of Seq. In other words, the insertion has to minimize $\text{len}(Seq) = \sum_{i=1}^{|Seq|-1} D(Seq[i+1],$ $Seq[i])$ where $D(.)$ denotes the Euclidean distance operator. For example, by inserting x_t between two records x_i and x_{i+1}, $D(x_i, x_t)$ and $D(x_t, x_{i+1})$ are added to the total length of Seq, but the distance between x_i and x_{i+1}, i.e., $D(x_i, x_{i+1})$ is subtracted from the total length. These cost values are computed in Step 9, and the minimum one is added to Seq in Step 10. The process continues until all records in the current group are added to Seq. In Step 12, if no unseen record remains, the algorithm goes to Step 14, otherwise the nearest unseen record in the whole dataset to the last record in Seq is chosen as the current record in Step 12-2, and the process of record ordering continues from Step 3. After ordering all records, the algorithm applies the optimal univariate microaggregation algorithm IMHM on it to produce a clustering that satisfies the size constraint and computes its SSE in Step 13. This clustering can be fed again to the algorithm to reorder all records and improve SSE (Step 15). If SSE is not decreased significantly, the function terminates, and the last computed SSE and the final

[§] Please recall that given a list of points, the TSP is to find the shortest possible route that visits each point and returns to the origin [16].

[**] An illustrative example of the algorithm execution on a small dataset is provided in the supplementary document of the paper.

clustering labels are returned as the algorithm outputs in Step 16.

4. 1. Analysis of the Algorithm The idea of using an initial clustering makes it possible to capture the inherent structure of the underlying dataset. Moreover, processing all records in a group-by-group fashion enables the process to focus on records of the current group rather than the whole dataset that makes the algorithm more efficient. The proposed method consists of repeated steps of record ordering in a sequence and applying IMHM. The SSE of each iteration of the loop is not worse than its value in the previous iteration since the ordering procedure meets records in a group-by-group manner and IMHM does not change the order of records in each group. Therefore, the SSE of each iteration improves gradually, or no significant decrement occurs at the last step, and the algorithm stops. It is also notable that the algorithm stops necessarily since the lower-bounded SSE cannot improve infinitely[††].

Algorithm 1. The pseudo-code of the RepOrdMic

Input: $T = \{x_1, x_2, \dots, x_n\}$: original dataset, k: the privacy parameter, lbl_{in}: initial clustering labels

Output: SSE: microaggregation error, lbl_{out}: the label of records to groups with at least k members

```
1       Initialize the sequence $Seq$ to NULL
2       Find the farthest record $x_f$ from the dataset centroid
3       $x_c \leftarrow x_f$
4       Set the group label of $x_c$ as the current group $G_c$, i.e., $G_c \leftarrow lbl_{in}[x_c]$
5       Add $x_c$ to the end of $Seq$
6       If the current group size is less than 2
6-1         Goto Step 12.
        Else
6-2         Find the farthest record from $x_c$ in the current group and add it to the end of $Seq$.
7       Foreach consecutive records of $G_c$ in $Seq$, $x_i$, and $x_{i+1}$
8           Foreach record $x_t$ in the $G_c$ that is not in $Seq$
9               Compute the cost of $x_t$ addition to $Seq$ at position $i$, i.e., $cost_{t,i} \leftarrow D(x_i, x_t) + D(x_t, x_{i+1}) - D(x_i, x_{i+1})$.
10      Find the minimum cost, say $cost_{t^*,i^*}$ and insert $x_{t^*}$ between $x_{i^*}$ and $x_{i^*+1}$.
11      If there exists any record of the current group that is not in $Seq$, Goto Step 7.
12      If there is not any unseen group
12-1        Goto Step 14
        Else
12-2        From any unseen groups, find the nearest record to the last record in $Seq$, set it as the current record $x_c$, and Goto Step 3.
13      Apply IMHM to $Seq$, to compute information loss and new labels, and save them in $SSE$ and $lbl_{out}$, respectively.
14      If $SSE$ is improved significantly
15          $lbl_{in} \leftarrow lbl_{out}$
            Goto Step 2
        Else
16          Return the last $SSE$ and $lbl_{out}$
```

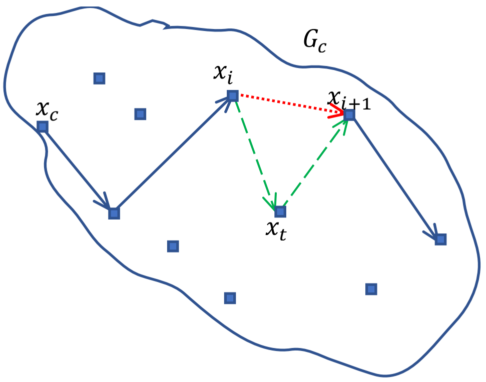


Figure 1. Adding $\boldsymbol{x_t}$ between $\boldsymbol{x_i}$ and $\boldsymbol{x_{i+1}}$ in the current group $\boldsymbol{G_c}$. Dashed green and dotted red lines indicate inclusion and removal, respectively

The space complexity of the proposed method for $|T| = n$ is $O(n)$ that is used for storing the ordering of records in Seq and computing the optimal clustering in IMHM. However, the runtime complexity is considerable. It is assumed that an initial clustering is provided before the first iteration. This clustering can be used for multiple values of the privacy parameter k and has to be computed once so that its execution time can be safely discarded. Finding the farthest record from the dataset centroid in Step 2 requires $O(n)$ operations. In the following steps of the algorithm, each iteration orders records of each group in a sequence. Except for the first step that processes the initial clustering labels, the following clustering results have $O(k)$ records in each group since they are the output of IMHM. Therefore, for each of $O(k)$ records in a clustering with $O(n/k)$ groups, the cost values can be computed in $O(k)$ operations. The process has to be repeated $O(k)$ times in each group to cover all records in the group, thus $O(k^3)$ operations are

[††] Note that the number of different orderings is limited and the algorithm visits each ordering at most once, so it has to stop. In practice, the iterations are broken when no significant improvement in SSE is achieved.

needed for each group. Finding the nearest unseen record of other groups is accomplished in $O(n)$ and is repeated $O(n/k)$ times. Hence, the ordering is completed in $O\left(\frac{n}{k}.k^3+n.\frac{n}{k}\right)=O\left(n\,.k^2+\frac{n^2}{k}\right)$ computations. The univariate microaggregation algorithm can be implemented efficiently in $O(nk)$ operations. As a result, for l iterations, the algorithm complexity is $O\big(l.n.(k^2+k+n/k)\big)$. However, it is notable that the whole process is an offline task, and the runtime is usually not a bottleneck, but the quality of the produced clustering in terms of SSE is more important.

As a side note, for an efficient implementation of the ordering process, it can be seen that after inserting a record to sequence Seq, most of the previously computed cost values remain the same and can be reused, but a small number of them has to be computed or updated.

5. EXPERIMENTAL RESULTS

A prototype of the proposed method was implemented in Microsoft Visual C++ 2019 in release mode. All evaluations are conducted within Windows 10 operating system on a regular laptop with Intel Core i5-8265U 1.60 GHz CPU and 8 GB of main memory. For initial clustering, different outputs of the *k*-means clustering algorithm are used for a number of clusters between 1 annd 200. Additionally, the iterations are broken when $\Delta SSE < 1e-7$.

Experiments were performed on three real-world benchmark datasets that are usually used for the evaluation of microaggregation algorithms. Benchmark datasets that are used in related previous studies [16, 17] are described in Table 1. All datasets contain numeric attributes without any missing values.

Table 2 shows information loss for various values of k. The proposed algorithm, namely RepOrdMic, is compared with MDAV [11], MDAV-MHM [10], IMHM [11], and IGD [17] methods. The results show that IL increases when k becomes greater for all microaggregation algorithms. For instance, $IL_{MDAV-MHM}$ for Census and $k = 3$ is 5.65, which is increased to 14.22 for $k = 10$. Additionally, IL of Tarragona dataset is generally higher than the other two datasets since it is known as a sparse dataset, which increases the cost of the anonymization process. The classic methods MDAV [11] and MDAV-MHM [10] are reported as reference techniques but do not produce any competing results in total. Similarly, the results of IGD are always worse than the winner methods. The IMHM [11] is more successful in the anonymization of Tarragona and Census for $\boldsymbol{k} = \mathbf{10}$, while the difference between IMHM and RepOrdMic is negligible in these cases. RepOrdMic, the proposed method, achieves the best results in other cases. For example, for EIA and k=10, RepOrdMic has improved the outputs of MDAV, MDAV-MHM, IMHM, and IGD by 48.18%, 47.21%, 8.29%, and 1.48%, respectively. In brief, the results indicate that RepOrdMic is successful in producing more useful datasets in 7 out of 9 experiment sets.

TABLE 1. Standard benchmark datasets for microaggregation comparison [16]

Dataset name	Number of data records (n)	Number of numeric attributes (d)
Tarragona	834	13
Census	1080	13
EIA	4092	11

TABLE 2. Information loss comparison for various standard datasets. Best $\boldsymbol{IL}$ values are bolded

Dataset	$\boldsymbol{k}$	MDAV	MDAV -MHM	IMHM	IGD	RepOrdMic
Tarragona	3	16.93	16.93	16.93	15.60	14.80
	5	22.46	22.46	22.18	21.31	21.13
	10	33.19	33.19	30.78	32.87	31.13
Census	3	5.69	5.65	5.37	5.33	5.01
	5	9.09	9.09	8.42	8.37	7.94
	10	14.16	14.22	12.23	12.65	12.74
EIA	3	0.48	0.41	0.374	0.39	0.369
	5	1.67	1.26	0.76	0.76	0.75
	10	3.84	3.77	2.17	2.02	1.99

All elapsed times of the proposed method, excluding initial clustering, read and write disk operations, and (de)normalization are shown in Table 3. These values are reported for 200 different number of initial clusters from 1 to 200 that are produced by *k*-means seeded with 0. The runtime of the algorithm for EIA is much larger than the other two cases since EIA is a large clustered numeric volume that makes it difficult and time-consuming for the algorithm to satisfy the privacy requirement. The experiments also show a decreasing runtime trend when $\boldsymbol{k}$ increases since for small values of $\boldsymbol{k}$, the runtime of Step 12-2 in Algorithm 1 dominates the runtime of other parts, but the behavior changes when $\boldsymbol{k}$ becomes larger. In brief, given an initial clustering, the algorithm is efficient and general; it usually terminates in a reasonable time regardless of the privacy parameter and underlying distribution or structure of the dataset.

Notably, the experiments can be extended to evaluate other important issues about the proposed method such as its effect on the proximity-based attack [18], its application for anonymization of complex structures such as graphs [19], and the impact of using other initial clustering techniques such as x-means [20] or consensus clustering [21].

TABLE 3. The runtime of RepOrdMic for 200 initial clustering labels.

Dataset	*k*	Total time (sec)	Total Iterations	Avg Iterations per clustering	Time per iteration (msec)
Tarragona	3	5.4	2525	12.63	2.14
	5	4.6	2755	13.78	1.67
	10	5	2988	14.94	1.67
Census	3	7.3	2368	11.84	3.08
	5	6.6	2504	12.52	2.64
	10	5.7	2527	12.64	2.26
EIA	3	51.3	1643	8.22	31.22
	5	36.8	1811	9.06	20.32
	10	25.9	1922	9.61	13.48

6. CONCLUSIONS

This paper presents a novel microaggregation algorithm based on the repeated ordering of multivariate records and mapping them to optimal univariate microaggregation algorithm IMHM. The process of ordering and applying IMHM is repeated until no significant improvement is achieved in terms of SSE. The output quality of the proposed method is usually better than similar methods in terms of IL. Extensive experiments on real-world datasets for different values of the privacy parameter k confirm that the algorithm is an efficient and general approach for practical usages. A promising extension of the proposed technique for future study is the way the records are ordered in a sequence.

7. REFERENCES

1. Erfani, S.H. and Mortazavi, R., “A Novel Graph-modification Technique for User Privacy-preserving on Social Networks”, ***Journal of Telecommunications and Information Technology***, Vol. 3, (2019), 27-38. DOI: 10.26636/jtit.2019.134319
2. Mortazavi, R. and Erfani, S.H., “An effective method for utility preserving social network graph anonymization based on mathematical modeling”, ***International Journal of Engineering, Transactions A: Basics*** Vol. 31, No. 10, (2018), 1624-1632. DOI: 10.5829/ije.2018.31.10a.03
3. Bypour, H., Farhadi, M. and Mortazavi R., “An Efficient Secret Sharing-based Storage System for Cloud-based Internet of Things”, ***International Journal of Engineering***, ***Transactions B: Applications***, Vol. 32, No. 8, (2019), 1117-1125. DOI: 10.5829/ije.2019.32.08b.07
4. Gheisari, M., Wang, G., Bhuiyan, M.Z.A. and Zhang W., “Mapp: A modular arithmetic algorithm for privacy preserving in IoT”, In IEEE International Symposium on Parallel and Distributed Processing with Applications and IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), (2017), 897-903. DOI: 10.1109/ISPA/IUCC.2017.00137
5. Mortazavi, R. and Jalili, S., “Preference-based anonymization of numerical datasets by multi-objective microaggregation”, ***Information Fusion***, Vol. 25, (2015), 85-104. https://doi.org/10.1016/j.inffus.2014.10.003
6. Sweeney, L., “k-anonymity: A model for protecting privacy”, ***International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems***, Vol. 10, No. 5, (2002), 557-570. https://doi.org/10.1142/S0218488502001648
7. Domingo-Ferrer, J. and Torra, V., “Ordinal, continuous and heterogeneous k-anonymity through microaggregation”, ***Data Mining and Knowledge Discovery***, Vol. 11, No. 2, (2005), 195-212. https://doi.org/10.1007/s10618-005-0007-5
8. Oganian, A. and Domingo-Ferrer, J., “On the complexity of optimal microaggregation for statistical disclosure control”, Statistical Journal of the United Nations Economic Commission for Europe, Vol. 18, No. 4, (2001), 345-353.
9. Hansen, S.L. and Mukherjee, S., “A polynomial algorithm for optimal univariate microaggregation”, ***IEEE Transactions on Knowledge and Data Engineering***, Vol. 15, No. 4, (2003), 1043-1044. DOI: 10.1109/TKDE.2003.1209020
10. Domingo-Ferrer, J., Martínez-Ballesté, A., Mateo-Sanz, J.M. and Sebé, F., “Efficient multivariate data-oriented microaggregation”, ***The VLDB Journal***, Vol. 15, No. 4, (2006), 355-69. https://doi.org/10.1007/s00778-006-0007-0
11. Mortazavi, R., Jalili, S. and Gohargazi, H., “Multivariate microaggregation by iterative optimization”, ***Applied Intelligence***, Vol. 39, No. 3, (2013), 529-544. https://doi.org/10.1007/s10489-013-0431-y
12. Aloise, D., Hansen, P., Rocha, C. and Santi, É., “Column generation bounds for numerical microaggregation”, ***Journal of Global Optimization***, Vol. 60, No. 2, (2014), 165-182. https://doi.org/10.1007/s10898-014-0149-3
13. Domingo-Ferrer, J. and Mateo-Sanz, J.M., “Practical data-oriented microaggregation for statistical disclosure control”, ***IEEE Transactions on Knowledge and Data Engineering***, Vol. 14, No. 1, (2002), 189-201.
14. Monedero, D.R., Mezher, A.M., Colomé, X.C., Forné, J. and Soriano, M., “Efficient k-anonymous microaggregation of multivariate numerical data via principal component analysis”, ***Information Sciences***, Vol. 503, (2019), 417-443. https://doi.org/10.1016/j.ins.2019.07.042
15. Soria-Comas, J. and Domingo-Ferrer, J., “Differentially private data publishing via optimal univariate microaggregation and record perturbation”, ***Knowledge-Based Systems***, Vol. 153, (2018), 78-90. https://doi.org/10.1016/j.knosys.2018.04.027
16. Mortazavi, R. and Jalili, S., “Fast data-oriented microaggregation algorithm for large numerical datasets”, ***Knowledge-Based Systems***, Vol. 67, (2014), 195-205. https://doi.org/10.1016/j.knosys.2014.05.011
17. Khomnotai, L., Lin, J.L., Peng, Z.Q. and Santra, A.S., “Iterative Group Decomposition for Refining Microaggregation Solutions”, ***Symmetry***, Vol. 10, No. 7, (2018), 262-274. https://doi.org/10.3390/sym10070262
18. Mortazavi, R. and Jalili, S., “Fine granular proximity breach prevention during numerical data anonymization”, ***Transactions on Data Privacy***, Vol. 10, No. 2, (2017), 117-144.
19. Mortazavi, R. and Erfani, S.H., “GRAM: An efficient (k, l) graph anonymization method”, ***Expert Systems with Applications***, Vol. 153, (2020), In press. https://doi.org/10.1016/j.eswa.2020.113454
20. Pelleg, D. and Moore, A.W., “X-means: Extending k-means with efficient estimation of the number of clusters”, In Proceedings of the Seventeenth International Conference on Machine Learning (ICML), (2000), 727-734.
21. Ünlü, R. and Xanthopoulos, P., “Estimating the number of clusters in a dataset via consensus clustering”, ***Expert Systems with Applications***, Vol. 125, (2019), 33-39. https://doi.org/10.1016/j.eswa.2019.01.074

8. APPENDIX

THE IMHM ALGORITHM

The appendix presents the pseudo-code of IMHM in brief. More details about the algorithm can be found in [9,11]. The main idea of IMHM is to calculate grouping errors incrementally to improve the time complexity of MHM [9]. The pseudo-code of the IMHM is provided in Algorithm 2. The algorithm accepts original records (as an ordered set) T, the privacy parameter k, and outputs SSE and record labels. The trivial case of $n < 2k$ is handled in Step 1 in which all records are assigned to one group. In Step 2, a directed acyclic graph $M(V, E)$ with $V = \{v_0, v_1, \dots, v_n\}$ is initialized (v_0 is a dummy node and v_i represents X_i in T for $0 < i \leq n$). In the following steps, some directed arcs are added to M. The weight of each arc equals to the SSE of grouping records between the start and end nodes of the arc. In Steps 6-7, the $Centroid$ and SSE of the first group are calculated. The results are stored in Steps 8-9 for later usage. Then, other records are added to the cluster until the size of the group reaches the limit of $2k - 1$ or no more record remains for addition. The weight of each arc is computed progressively in Steps 11-21. In Step 22, the shortest path from v_0 to v_n is saved in SP. The microaggregation error SSE and assignment A are computed in Steps 23 and 24-28, respectively. Finally, the values are returned in Step 29.

Algorithm 2. The pseudo-code of the Improved MHM (IMHM) [11]

Input: $T = (X_1, X_2, \dots, X_n)$: dataset of original records in order, k: the privacy parameter

Output: SSE: microaggregation error, A: the assignment of records to clusters with at least k members

```
1   If $n < 2k$, assign all records to the same cluster, calculate its $SSE$ and Return $SSE$ and the assignment
2   Initialize the directed acyclic graph $M(V, E), |V| = n + 1$
3   For $i \leftarrow 0$ To $n - k$
4       For $j \leftarrow i + k$ To $min\ (n,\ i + 2k - 1)$
5           If $i = 0$ And $j = i + k$
6               $CurrentCentroid \leftarrow$ MEAN ($X_1$ to $X_k$)
7               $CurrentSSE \leftarrow$ calculate $SSE$
8               $BaseCentroid \leftarrow CurrentCentroid$
9               $BaseSSE \leftarrow CurrentSSE$
10          Elseif $j = i + k$
11              $Delta \leftarrow X_{i+k} - X_i$
12              $CurrentCentroid \leftarrow BaseCentroid + Delta/k$
13              $CurrentSSE \leftarrow BaseSSE + \frac{\sum_{l=1}^{d} Delta[l].\left((1-k)Delta[l] + 2k(X_{i+k}[l] - CurrentCentroid[l])\right)}{k}$
14              $BaseCentroid \leftarrow CurrentCentroid$
15              $BaseSSE \leftarrow CurrentSSE$
16          Else
17              $s \leftarrow j - i$        // the group size
18              $OldCentroid \leftarrow CurrentCentroid$
19              $CurrentCentroid \leftarrow OldCentroid + (X_j - OldCentroid)/s$
20              $CurrentSSE \leftarrow CurrentSSE + \sum_{l=1}^{d}(X_j[l] - CurrentCentroid[l])(X_j[l] - OldCentroid[l])$
21          Draw a directed edge $e = (v_i, v_j)$ and set the weight $w(v_i, v_j) \leftarrow CurrentSSE$.
22  Compute $SP$ as the shortest path from $v_0$ to $v_n$ in $M(V, E)$
23  $SSE \leftarrow$ The length of $SP$
24  $ClusterCounter \leftarrow 1$
25  Foreach edge $e = (v_i, v_j) \in SP$
26      Foreach $v_m,\ i < m \leq j$
27          Assign $X_m$ to $G_{ClusterCounter}$       // $A[m] \leftarrow ClusterCounter$
28      $ClusterCounter \leftarrow ClusterCounter + 1$
29  Return $SSE$ and $A$
```

Persian Abstract

چکیده

خوشه‌بندی یکی از روش‌های اصلی در داده کاوی است که کاربردهای فراوانی در علوم کامپیوتری، زیست شناسی و علوم اجتماعی دارد. خوشه‌بندی مقید نوعی خوشه‌بندی است که در آن اطلاعات اضافی ارائه شده توسط کاربر در طی خوشه‌بندی دخالت داده می‌شود. یکی از انواع الگوریتم‌های مورد پژوهش در زمینه خوشه‌بندی مقید، الگوریتم زیرتجمیع است. در ریزتجمیع، الگوریتم خوشه بندی باید مجموعه داده را به گروه هایی با حداقل $\boldsymbol{k}$ عضوتقسیم کند که $\boldsymbol{k}$ یک پارامتر تعریف شده توسط کاربر است. کاربرد اصلی ریزتجمیع در کنترل افشای آماری است که برای انتشار داده‌ها با حفظ حریم خصوصی کاربرد دارد. کیفیت الگوریتم ریزتجمیع بر اساس مجموع مربع خطاهای داخل گروه اندازه گیری می‌شود. متاسفانه ثابت شده که ریزتجمیع بهینه در حالت کلی یک مسئله بهینه سازی $\boldsymbol{NP}$-سخت است، اما نسخه تک بعدی آن به صورت بهینه و در زمان چند جمله‌ای قابل حل است. روش های ابتکاری زیادی براساس تبدیل به نسخه تک متغیره پیشنهاد شده است. این روشها باید داده‌ها چند بعدی را در یک دنباله مرتب کنند. این مقاله روش جدیدی برای مرتب سازی داده ها پیشنهاد می‌کند. با شروع از یک خوشه بندی اولیه، روش پیشنهادی به صورت تکراری رکوردهای چند بعدی را در یک دنباله مرتب کرده و آنها را خوشه‌بندی می‌کند. این فرایند تا زمانی که بهبودی حاصل نشود ادامه می‌یابد. مجموعه وسیعی از آزمایش‌های تجربی انجام شده در این تحقیق نشان از برتری روش پیشنهادی برای پارامترها و مجموعه داده‌های مختلف دارد.

International Journal of Engineering

Journal Homepage: www.ije.ir

A Location-Routing Model for Assessment of the Injured People and Relief Distribution under Uncertainty

H. Beiki[a], S. M. Seyedhosseini*[b], V. R. Ghezavati[a], S. M. Seyedaliakbar[a]

[a] *School of Industrial Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran*
[b] *Department of Industrial Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran*

PAPER INFO

Paper history:
Received 13 April 2020
Received in revised form 20 May 2020
Accepted 01 June 2020

Keywords:
Assessment of Injured People
Disaster Relief
Location-Routing Problem
Response Phase
Uncertainty

ABSTRACT

Throughout history, nature has exposed humans to destructive phenomena such as earthquakes, floods, droughts, tornadoes, volcanic eruptions, and tropical and marine storms. The large scale of damages and casualties caused by natural disasters around the world has led to extensive applied research in the field of preparation and development of a comprehensive system for disaster management to minimize the resulting casualties and financial damages. Based on this motivation and challenges to the field, this research designs an integrated relief chain to optimize simultaneously the preparedness and response phases of disaster management. Decisions to improve the supply chain include locating distribution centers of relief supplies; the amount of inventory stored in facilities in pre-disaster phase, locating temporary care centers and transportation points of the injured, how to allocate relief services to the affected areas, and routing of the vehicles used to distribute relief supplies and evacuate the injured. The results show that decreasing the capacity of distribution centers increases the amount of shortage of supplies and increasing the capacity of these centers reduces the amount of shortage of supplies.

***doi**: 10.5829/ije.2020.33.07a.14*

1. INTRODUCTION

In today's world, because natural disasters such as earthquakes, floods, droughts, storms, volcanic eruptions and so on are sweeping the globe, the importance of disaster management and strategies to accelerate supply and response to demand created at the time of disaster is felt more than before [1]. In fact, the unpredictable nature and devastating effects of such events force communities to foresee and develop appropriate plans to minimize disaster damage and casualties. Disaster management is a continuous process involving operations to prepare for a disaster, respond to it as soon as it occurs, and support and rebuild damaged infrastructure after a disaster [2]. Disaster management activities are usually categorized into four phases of preparedness, response, recovery, mitigation, and proper planning for each phase can lead to better preparation, less vulnerability or future disaster prevention. After the occurrence of a disaster, we face major problems, such as a shortage of inventory to supply and transport many of the critical supplies including food, clothing, medicine, equipment, and personnel to the affected areas [3]. Emergency relief efforts should also focus on finding and rescuing survivors. Logistics and rescue personnel must therefore be transported quickly and efficiently to maximize the rescue rate of affected people and minimize the cost of operations [4]. However, as the transportation of logistics to the affected areas is carried out under uncertainty and as the relevant logistics information changes during disaster response, planning for the disaster will face significant complexities [5]. Therefore, the rules and strategies of relief organizations should be dynamic and flexible due to the sudden and unexpected changes and an information support system can bring this dynamism and adapt plans to the new information obtained [6, 7]. In relief logistics problems, finding the location of critical supplies before the occurrence of a disaster is one of the most important logistics strategies to reduce delivery times and operating costs, prior to the accident [8–10]. Pre-location of the

*Corresponding Author Email: *seyedhosseini@iust.ac.ir* (S. M. Seyedhosseini)

Please cite this article as: H. Beiki, S. M. Seyedhosseini, V. R. Ghezavati, S. M. Seyedaliakbar, A Location-Routing Model for Assessment of the Injured People and Relief Distribution under Uncertainty, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1274-1284

facilities not only enables faster response but also creates a better preparation plan and improves distribution costs. Also, a good logistics support operation requires tactical level decisions to transport logistics to the affected areas and the affected people to hospitals and care centers [10–15]. Therefore, efficient transportation systems along the relief supply chain are of the utmost importance in order to respond appropriately to the critical conditions occurred.

Based on recent advances in this research area, the location-routing problem has been one of the most important location problems from the viewpoint of integrated logistics systems analysts [15–19]. This problem gives analysts a stronger perspective on the mutual relationships between facility locations and vehicle routes; allows for a centralized operational plan where delays are eliminated and limited resources are allocated as best as possible. In fact, the fundamental difference between this comprehensive and integrated approach and classic location problems is that in the location-routing method, after determining the location of the facilities, the routes between the facilities and the customers are examined as a tour, while in the traditional method; it is assumed that there are direct routes between the customer and the facilities. Having a conclusion about aforementioned contributions, the main novelty of this paper is to develop a new location-routing model for the assessment of injured people and relief distribution under uncertainty.

The rest of this research will be as follows: In section 2, we will review the literature on this subject. In section 3, the problem statement will be mentioned and in section 4, we will introduce the parameters of the problem and the mathematical model. Section 5 will outline the solution approach. Section 6 will describe the case study and present the computational results. Finally, section 7 will present conclusion, managerial insights and future research proposals.

2. LITERATURE REVIEW

The literature is rich in using the application location-routing models for different supply chain concepts [16–22]. To close to nature of real-world application, uncertainty modeling is an active keyword for logistics network from recently-published papers [23–32]. Here, we review some recent and most relevant works related to our scope of research. For example, in 2019, Paul and Zhang [1] presented a multi-objective hybrid optimization model for the routing-location problem of mobile units of medical services. They presented three heuristic and metaheuristic algorithms to solve the model and applied real data to validate these algorithms. Paul and Wang [2] formulated the location and distribution of relief supplies during the occurrence of a flood under different probabilistic scenarios using two-stage stochastic programming. The model assumed that there are several types of relief centers, such as: regional relief centers, local relief centers, etc. that support each other and demand points if necessary. The objective was to minimize costs and a heuristic method was used to solve it.

In another recent paper, Nagurney et al., [12] proposed a two-stage multi-criteria uncertain programming model for locating pre-disaster emergency response and distribution centers for efficient emergency logistics in times of a disaster. They also presented a goal programming that in the first stage, determined the location, capacity of the facilities and the quantity of supplies stored in each facility; and in the second stage, a transportation problem was solved with two main assumptions: the capacity of the routes was infinite, but it was possible that in a scenario this route would not be possible and nodes were as storages for supplies. Fathollahi-Fard et al., [18] developed a two-stage stochastic programming model to develop a closed-loop logistics network design for the application of water distribution network. They developed an adaptive Lagrangian relaxation to solve a case study in the west Azerbaijan province of Iran. Another closed-loop logistics under uncertainty was developed by Abdi et al., [16]. They considered the demand, returned productions and prices as the uncertain parameters and applied a financial risk model to evaluate the application fruit industry in Iran. Another contribution was a comparison with whale optimization algorithm as a recent meta-heuristic with genetic algorithm, simulated annealing and particle swarm optimization based on the assessment metrics of Pareto fronts.

In another different research, Davoodi et al. [6] developed a deterministic and static location-routing approach for deploying pre-disaster establishment of suppliers in such a way as to maximize the probability of supplying the demand of affected points through supplier facilitation considering the transportation network failures. With regards to order allocation and the selection of suppliers, Safaeian et al., [21] developed a Zemin’s fuzzy model to consider the uncertain parameters of this problem. Their significant contribution was to apply a non-dominated sorting genetic algorithm to find an interaction between the total cost, quality of products, prices and satisfactions of customers as the objective functions. As another supplier selection assessment, Feng et al., [28] developed a new hybrid fuzzy grey TOPSIS method to provide a comprehensive analysis for a case study in China.

Fathollahi-Fard et al., [24] developed a bi-objective logistics network for the application home healthcare organizations. Their model as a variant of location-routing model optimizes the total cost and the environmental pollution simultaneously. They also provided a comparison with new and state of the art modified simulated annealing algorithms. In another multi-objective model, Torabi et al. [10] presented a multi-objective approach for the location of emergency shelters as well as determination of evacuation routes during the preparedness and response phases. Since a route or a shelter may be unusable due to a fire, the backup route or shelter is provided for each building.

As another logistics network under uncertainty, Noham and Tzur [7] integrated the problem of pre-disaster facility location, inventory, and routing by presenting a two-stage probabilistic planning model. The objective was to determine the location and number of local distribution centers and their inventory levels to ensure rapid and efficient response in times of a disaster. In the first stage, the design variables were determined based on the available information and in the second stage, these variables were estimated with the objective of optimizing the total demand met and the total transportation costs based on the existing information. In addition, Loree and Aros-Vera [3] presented an integrated supply chain logistics model for controlling the flow of multiple relief supplies in the response network. The model considered the optimal locations for several layers of temporary facilities, as well as the optimal routes for delivering and loading relief supplies. Based on a collaborative closed-loop logistics for water supply chain, Torabi et al. [33] proposed a stochastic programming and applied Lagrangian relaxation to address it.

In 2019, Liu et al. [13] proposed a multiple optimization algorithm for the capacitated location-routing problem. In this study, the capacitated location-routing problem was divided into two facility location problem and the vehicle routing problem with multiple warehouses, so that the second problem was a sub-problem of the first problem. Mehranfar et al., [34] proposed a production-distribution logistics network considering carbon tax under uncertainty. A novel hybrid whale optimization algorithm was developed to address their problem.

As the last example of multi-objective optimization in this literature, Li et al. [25] presented a three-objective transportation location model for the disaster response phase. The objectives of this model included reducing the transportation time of relief supplies, reducing the number of rescuers needed to open and operate the established distribution centers, and reducing the number of unmet demand. Finally, the epsilon constraint exact approach was proposed to solve the model. At last but not least, Haghi et al., [8] proposed a three-level stochastic programming model for the disaster response phase. This model aimed to maximize effectiveness and fairness in relief distribution by locating facilities, allocating resources, and last-mile distribution of relief supplies. In this problem, demand, the number of transportation vehicles and accessibility of uncertain communication infrastructure were considered.

Based on the aforementioned works and to keep this research area active, the following contributions can fill the research gaps in this research area:

- Designing an integrated four-level relief chains including suppliers, distributors, affected areas and a variety of care centers with the aim of minimizing unmet demand and uncared people.
- Simultaneous consideration of strategic and operational decisions related to disaster preparedness and response phases.
- Simultaneous optimization of facility location, resource allocation, relief distribution and evacuation of the injured problems assuming demand uncertainty and facility availability and so on.
- Using the data and results of earthquake damage estimation in district one of Tehran city to validate the model under real conditions.

3. PROBLEM STATEMENT

As noted before, as human health suffers most damages in disaster situations, medical service planning and management in emergency situations are of utmost importance. This is especially important in the early hours after the occurrence of a disaster because the efficient planning and management of medical and pharmaceutical supplies can save the injured.

Further, any kind of planning in disaster situations without considering the inseparable features of these situations, is not efficient. These features include issues such as uncertainty. In the event of a disaster, potential damage, the location of temporary centers and consequently the amount of demand in the affected areas is highly uncertain. Therefore, the location should be done such that it can effectively cover the demand points. In the real world, we often face uncertain supply, demand and costs during disaster response. Considering the uncertainties arising from disaster situations in the design and analysis of the model presents many challenges.

Our objective in this study is to present a mathematical model for integrating location and routing decisions in uncertain and changeable situations arising from the occurrence of a disaster. In fact, we formulate the disaster relief logistics location-routing problem as a linear integer scenario-based multi-objective model. In this study, a multi-objective model is presented to coordinate the distribution of emergency medical supplies and emergency evacuation activities of the injured. In order to have an effective relief distribution, the proposed model also seeks to locate a number of temporary relief centers near the affected areas. In this model, the major source of uncertainty, which is an inherent characteristic of emergency situations, is considered. This source of uncertainty includes the unpredictability of the time and location of the disaster, the amount of demand and potential damage to the infrastructure which is discussed in this paper through scenario-based planning considering different scenarios for the disaster. Based on the explanations given, this paper attempts to optimize the necessary strategic decisions in disaster situations by presenting a multi-period multi-objective mathematical model.

4. MATHEMATICLA FORMULATION

This section provides the assumptions, notations and mathematical model.

4. 1. Assumptions

- The number and location of suppliers, affected areas and existing care centers are fixed and determined.
- Potential locations for the establishment of relief distribution centers and the injured transportation points are identified.
- The capacity of relief centers is constrained and varies based on their size (small, medium, large).
- Each distribution center is only able to serve the area in which it is located.
- The capacity of vehicles to carry different kinds of supplies and the injured is constrained and determined.
- Each vehicle can start and end its route from different locations.

4. 2. Notations

- **Sets**

N: Set of network nodes (o,p$\in N$) (I$\cup J \cup K \cup L \subset N$)
I: Set of suppliers
J: Set of candidate distribution points
K: Set of affected areas
M: Set of temporary care centers
H: Set of hospitals
C: Set of supplies
W: Set of the injured
S: Set of possible scenarios
L: Set of the size of distribution centers
V: Set of vehicles

- **Parameters:**

p_s: Probability of occurrence of scenario s
ω_c: Priority of meeting the demand for supply c
ω'_w: Priority of serving the injured person w
P: The number of temporary care centers to be established
v_c: Volume of each unit of supply c
w''_c: Weight of each unit of supply c
cap_l: Capacity of distribution center size l
$capm_w$: Capacity of a temporary care center for the injured type w
$capv_v$: Volume capacity of vehicle v for transporting supplies (in cubic meters)
$capw_v$: Weight capacity of vehicle v for transporting supplies (in kilograms)
$Capl_v$: Capacity of vehicle v to carry the injured
$Caps_{oc}$: The amount of supply c that can be supplied from the supplier node o∈I
av_{vs}: Number of vehicles type v available in scenario s
dc_{ocs}: Demand for supply c in affected node o∈kin scenario s
dw_{ows}: Number of affected people type w in affected node o∈kin scenario s
ρ_{os}: Percentage of accessibility of facilities including suppliers, distributors and care centers located in node o in scenario s
sb_{wo}: Number of beds available at node o∈H for the injured type w
δ_{opvs}: 1, If the vehicle v can travel the axis of o to pin scenario s; otherwise it is zero.
ac_{cv}: 1. If vehicle v is capable of carrying supply c; otherwise it is zero.
aw_{wv}: 1, If vehicle v is capable of carrying the injured type w; otherwise it is zero.
ab_{ops}: 1, If the facility at the location o∈Jis able to serve the facility located in p∈kin scenario s; otherwise it is zero.

- **Decision variables:**

Q_{opc}: The quantity of supply c supplied by the supply node o and stored in the node p
x^v_{opcs}: The quantity of supply c transported from supply node o to p by vehicle v in scenario s
y^v_{opws}: Number of the injured type w transported from node o∈k to node p∈H by vehicle v in scenario s
y'^s_{opw}: Number of the injured type w transported from node o∈k to node p∈Min scenario s
N_{opvs}: The number of vehicles type v passing the route (o,p)in scenario s
Ux_{ocs}: The amount of shortage for supply c in affected node $o \in k$in scenario s
Uy_{ows}: Number of the injured type w that are not served yet in affected node $o \in k$in scenario s.
z_{ol}: 1, If the distribution center size l is established in the node o∈J; otherwise it is zero.
z'_{os}: 1, If a temporary care center is established in scenario s in location o∈M; otherwise it is zero.

4. 3. Mathematical Model

$$MinZ_1 = \sum_w \sum_{o\in k} \sum_s \omega'_w . Uy_{ows} \qquad (1)$$

$$MinZ_2 = \sum_c \sum_{o\in k} \sum_s \omega_C . Ux_{ocs} \qquad (2)$$

$$\sum_{o\in I} \rho_{os} . Q_{opc} + \sum_{o\in I} \sum_v x^v_{opcs} - \sum_{o\in K} ab_{pos} . \sum_v x^v_{pocs} \geq 0 \qquad \forall p \in j,c,s \qquad (3)$$

$$\sum_{p\in j}\sum_{v} ab_{pos}\,.\,x_{pocs}^{v} + Ux_{ocs} \geq dc_{ocs} \quad \forall o \in k, c, s \qquad (4)$$

$$\sum_{p\in H}\sum_{v} y_{opws}^{v} + \sum_{p\in M} y'^{s}_{opw} + Uy_{ows} = dw_{ows} \quad \forall o \in k, w, s \qquad (5)$$

$$\sum_{o\in N} x_{opcs}^{v} \leq M.\sum_{l} Z_{pl} \quad \forall p \in j, c, s \qquad (6)$$

$$\sum_{p\in N} x_{opcs}^{v} \leq M.\sum_{l} Z_{ol} \quad \forall o \in j, c, s \qquad (7)$$

$$\sum_{w}\sum_{o\in K} y'^{s}_{opw} \leq M.\sum_{o\in K} z'_{os} \quad \forall p \in M, s \qquad (8)$$

$$\sum_{o\in N}\sum_{p\in N} x_{opcs}^{v} \leq M.\,ac_{cv} \quad \forall v, c, s \qquad (9)$$

$$\sum_{o\in N}\sum_{p\in N} y_{opws}^{v} \leq M.\,aw_{wv} \quad \forall v, w, s \qquad (10)$$

$$\sum_{o,c} \upsilon_{c}.\,Q_{opc} \leq \sum_{l} Cap_{l}.\,Z_{pl} \quad \forall p \in j \qquad (11)$$

$$\sum_{p\in j} Q_{opc} \leq caps_{oc} \quad \forall o \in I, c \qquad (12)$$

$$\sum_{V}\sum_{p\in j} x_{opcs}^{v} \leq \rho_{os}.\,caps_{oc} \quad \forall o \in I, c, s \qquad (13)$$

$$\sum_{V}\sum_{o\in k} y_{opws}^{v} \leq \sum_{o\in k} \rho_{os}.\,sb_{wp} \quad \forall p \in H, w, s \qquad (14)$$

$$\sum_{V}\sum_{o\in k} y'^{s}_{opw} \leq \sum_{o\in k} capm_{w}.\,z'_{os} \quad \forall p \in M, w, s \qquad (15)$$

$$\sum_{v}\sum_{c} \upsilon_{c}.\,x_{opcs}^{v} \leq \sum_{v} Capv_{v}.\,N_{opvs} \quad \forall o \in N, p \in N, s \qquad (16)$$

$$\sum_{v}\sum_{c} w''_{c}.\,x_{opcs}^{v} \leq \sum_{v} capw_{v}.\,N_{opvs} \quad \forall o \in N, p \in N, s \qquad (17)$$

$$\sum_{v}\sum_{w} y_{opws}^{v} \leq \sum_{v} Capl_{v}.\,N_{opvs} \quad \forall o \in N, p \in N, s \qquad (18)$$

$$\sum_{(o,p)\in N} N_{opvs} \leq av_{vs} \quad \forall v, s \qquad (19)$$

$$N_{opvs} \leq M.\,\delta_{opvs} \quad \forall o \in N, p \in N, s, v \qquad (20)$$

$$\sum_{l} Z_{ol} \leq 1 \quad \forall o \in J \qquad (21)$$

$$\sum_{o\in M} z'_{os} = p \quad \forall s \qquad (22)$$

$$z'_{os}, z_{ol} \in \{0,1\}, Ux_{ocs}, Uy_{ows}, N_{opvs}, y'^{s}_{opw}, x_{opcs}^{v}, y_{opws}^{v}, Qopc \geq 0 \qquad (23)$$

Contrary to previous works, the first objective function does not seek to minimize the number of the injured served; in fact it seeks to increase the level of service to the injured. The second objective function also seeks to minimize the shortage of relief supplies in the affected areas. It should be noted that these objective functions do not have the same priority and are optimized hierarchically.

Constraint (3) relates to the flow of relief supplies in distribution centers and ensures that the quantity of supplies delivered by each distribution center to the affected areas should be less than the inventory available at those centers. Constraint (4) shows the flow of relief supplies in the affected areas and indicates the amount of shortage in each affected area. Constraint (5) relates to the flow of the injured in the affected areas and also determines the number of the injured waiting in each affected area to be served. Constraints (6) and (7) ensure that the inflow and outflow of supplies in distribution centers is only possible if these centers are established. Constraint (8) also guarantees that the transportation of the injured to temporary care centers is only possible if these centers are established.

Constraints (9) and (10), respectively, indicate the ability of vehicles to carry different kinds of relief supplies and to carry different kinds of the injured. Constraint (11) guarantees that it is possible to store all kinds of relief supplies in distribution centers only if these centers are established and that the amount of these supplies is less than the capacity of the distribution centers. Constraint (12) indicates the capacity of suppliers to deliver relief supplies to distribution centers before the occurrence of a disaster. Constraint (13) guarantees that suppliers are able to deliver relief supplies after the occurrence of a disaster. Constraints (14) and (15) also ensure the consideration of the capacity of existing care centers and temporary care centers after the occurrence of a disaster. Constraints (16) and (17) guarantee the consideration of the volume and weight capacity of vehicles to carry different kinds of relief supplies. Constraint (18) ensures the consideration of the capacity of vehicles to carry different kinds of the injured. Constraint 19 indicates that the number of vehicles available in each scenario is constrained. Constraint 20 also guarantees that vehicles can only move on the network arcs if they are available after the occurrence of a disaster. Constraint 21 indicates that only one of the sizes of distribution centers can be established at each point. Constraint 22 guarantees that a determined number of temporary care centers will be established. Constraint 23 specifies the type of decision variables.

5. SOLUTION APPROACH

Given that the proposed model is presented in the humanitarian space, their objective functions cannot be compared with each other in terms of priority. For example, the objective functions related to evacuation and rescue of the injured have a higher priority than the distribution of relief supplies. Also, time-related objective functions are more important than cost functions. Therefore, since the objective functions of the proposed model are hierarchically prioritized, one of the multi-objective optimization methods called lexicographic approach has been used to solve the problem. Based on this approach, the first objective function is assumed without considering the other objective functions. Then, this objective function is optimized based on the optimal value obtained from solving the model, the constantf_1. Therefore, the initial objective function is added to the model as an additional constraint $\sum_w \sum_{o\in k} \sum_s \omega_w . Uy_{ows} \leq f_1$. Then, the initial model is solved by assuming that a constraint is added to the problem in order to minimize the second objective function.

6. CASE STUDY AND RESULTS

Tehran is one of Asia's most densely populated and earthquake-prone cities. Evidence shows that severe earthquakes could result severe damages if they happen in this city. District one of Tehran city is one of the busiest and most sensitive parts of Tehran, surrounded by two Mosha and Ray faults. According to statistical data, this area is 200 square kilometers with a population of 620000 people. In this section, the performance of the proposed model in this area is investigated. Figure 1 shows Tehran’s earthquake-prone zones along with the map of the case study.

Here, we solve this case study in Tehran. In this regard, Table 1 shows the probability of occurrence of each scenario in the case study.

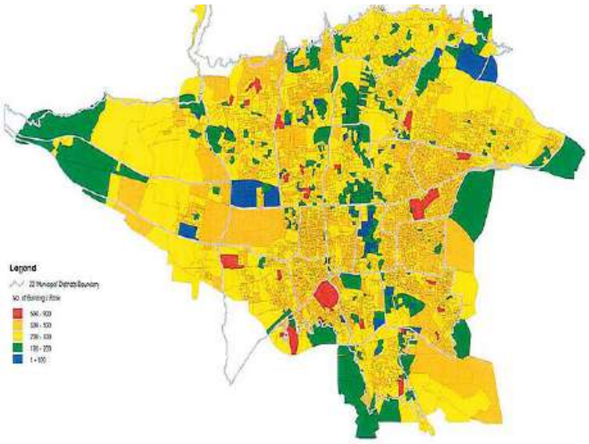


Figure 1. The map of the case study

TABLE 1. The probability of occurrence of each scenario

Scenario	Mosha fault		Ray fault	
Time of occurrence	night	day	night	day
Probability of occurrence	0.0614	0.2036	0.0305	0.0465
Severity of occurrence	6.8		6.2	

Table 2 shows the set of candidate points for the establishment of distribution centers. These centers can be established in three sizes: small, medium and large, each with different establishment costs.

Table 3 shows the bases of suppliers in District 1 of Tehran city. In this study, we have assumed that the affected areas include damaged and old areas of the city. These areas are listed in Table 4. The set of available care centers in the district is shown in Table 5.

TABLE 2. Candidate points for the establishment of distribution centers

No.	Distribution center
1	Niavaran base
2	Jamaran Base
3	Dezashib base
4	Tajrish base
5	Elahieh base
6	Chizar base
7	Velenjak base
8	Aqdasieh base

TABLE 3. Supplier bases

No.	Supplier base
1	Hekmat base
2	Farmanieh base
3	Evin base
4	Zafaranieh base

TABLE 4. Affected areas

No.	Affected area	No	Affected area
1	Kamranieh	6	Pasdaran
2	Sa’dabad	7	Aqdasieh
3	Darakeh	8	Dezashib
4	Jamshidieh	9	Andarzgu
5	Darband	10	Kashanak

TABLE 5. Care centers

No.	Hospital	Reception capacity	No	Hospital	Reception capacity
1	Sasan	1500	6	Mahak	5000
2	Chamran	5000	7	Jamaran	10000
3	Nikan	8000	8	505 Artesh	12000
4	ShohadaTajrish	4800	9	Nurafshar	10000
5	Farhangian	6000	10	Ramtin	4000

Three types of relief supplies including tents, water and food are considered in this study. To calculate the demand for these supplies during the occurrence of an earthquake, it is assumed that during the first 100 hours (golden time) of the disaster response, one tent will be delivered to each affected family and two quotas of water and food will be delivered to each person per day. The percentages of availability of facilities in each of these areas were calculated using the percentage of destruction of buildings and are presented in Table 6.

Table 7 shows the parameters needed for relief supplies including their weight, volume and cost. Table 7 shows information on priority of meeting demand, supply costs, etc. for each type of relief supplies. It is also assumed that the cost of maintaining inventory and the cost of providing each supply for the post-disaster phase is similar to the pre- disaster phase.

Table 8 shows the information and parameters related to the injured. Table 9 shows the capacity of suppliers for relief supplies.

TABLE 6. Percentage of facility availability

Affected area	Percentage of facility availability	Affected area	Percentage of facility availability
1	25	6	75
2	40	7	19
3	58	8	60
4	25	9	65
5	48	10	80

TABLE 7. Parameters required for relief supplies

Commodities	Volume (m^3)	Weight (kg)	Supply cost (10^3\$)
Water	0.0038	1.5	0.003
Tent	0.18	3	0.05

TABLE 8. Parameters needed to serve the injured

Injured type	Priority
Mild	0.15
medium	0.45
Dire	0.65

TABLE 9. Capacity of suppliers

No.	Supplier base	Water	Food	Tent
1	Hekmat base	600000	500000	40000
2	Farmanieh base	550000	450000	40000
3	Evin base	750000	800000	40000
4	Zafaranieh base	850000	70000	50000

Table 10 shows the capacity of distribution centers. The results of this exact solution are as follows: 36 people are not served and there are 420 units of supply shortage in various affected areas. Table 11 shows the distribution centers established at each of the potential locations with their optimal capacity.

Table 12 shows the quantities of supplies transported from suppliers to distribution centers after the occurrence of a disaster in a defined scenario. Table 13 shows the number of the injured transported from affected areas to existing hospitals in the district.

After solving our model, we examine the model's sensitivity to the parameters of facility capacity, the number of established temporary care centers, and the number of established distribution centers. Decreasing the capacity of distribution centers increases the amount of shortage of supplies and increasing the capacity of these centers reduces the amount of shortage of supplies (Figure 2). As can be seen, due to the 30 percent increase, the objective function has decreased to 980, and by 35 percent decrease, the objective function increases to 3500.

Figure 3 shows the sensitivity analysis of both objective functions relative to the changes in the capacity of vehicles to carry different kinds of relief supplies and the injured. This figure shows that changes in the number and capacity of vehicles can increase or decrease the number of unserved injured in the network. Also, a part of the shortage of relief supplies is related to the weight and volume capacity of vehicles, so that as the capacity of vehicles increases, the amount of shortage of supplies will reach zero.

TABLE 10. Capacity of relief distribution centers

Scale	Capacity
Small	4000
Medium	6000
Large	8000

TABLE 11. Distribution centers established and the amount of their inventory

No.	Scale of distribution center	Amount of inventory in distribution centers		
		Tent	Water	Food
1	small	40771	131065	131065
2	medium	55972	386512	402568
3	large	46962	303987	820040
4	large	49856	294605	294625
5	medium	55802	401612	401612
6	medium	48887	280923	280923
7	large	70858	856006	70926
8	small	40460	130460	129565

TABLE 12. Quantity of relief supplies provided by suppliers in the post-disaster phase

Supplier	Commodities	Distribution center							
		1	2	3	4	5	6	7	8
1	Tent	6200	1200	650	0	450	230	0	0
	Water	0	0	1200	720	0	0	1500	1000
	Food	14200	0	0	1500	2000	100	0	1000
2	Tent	0	1000	0	550	1000	0	0	100
	Water	13500	600	0	780	4000	0	0	5000
	Food	1100	0	20000	120	1500	15000	0	8000
3	Tent	0	0	0	110	100	0	300	700
	Water	0	0	0	250	500	0	1500	230
	Food	0	2500	0	410	0	0	8000	1500

TABLE 13. Number of the injured sent to existing care centers

Affected area	Hospitals									
	1	2	3	4	5	6	7	8	9	10
1	-	350	-	-	230		600	75		-
2	-	-	760	-	-	400	-	-	100	460
3	210	-	-	-	-	-	130	35	-	-
4	-	-	-	800	-	150		-	-	-
5	-	810	-	-	-	-	120	-	200	-
6	-	-	-	560	530	-	-	-	-	460
7	-	360	-	-	-	80	-	100	-	-
8	-	-	-	650	-	-	-	-	300	410
9	400	-	-	-	-	630	200	-	-	-
10	-	-	160	-	100	-		100	-	150

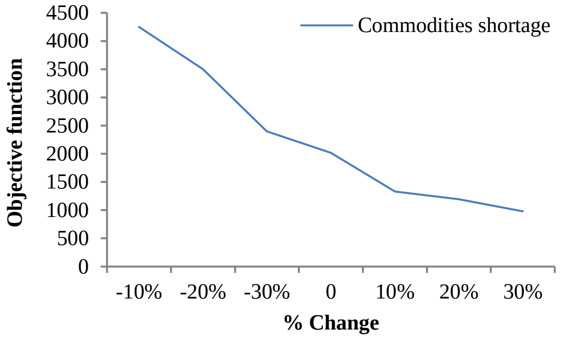


Figure 2. Sensitivity analysis of the objective function of the shortage of supplies relative to the capacity of the relief distribution centers

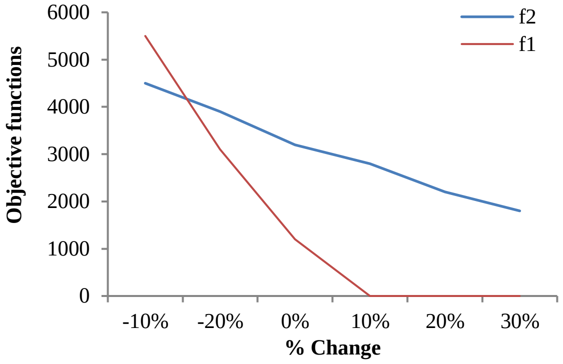


Figure 3. Changes of both objective functions relative to vehicle capacity changes

7. CONCLUSION, MANAGERIAL INSIGHTS AND FUTURE DIRECTIONS

In this study, it was attempted to design a comprehensive and integrated disaster relief model, so as to create a suitable model for disaster management programs. Also, the proposed relief chain structure for one of Tehran's districts was studied to evaluate its effectiveness in complex and flexible disaster situations, especially in an earthquake-prone metropolis such as Tehran with high population, low capacity of passages, unreliable construction and lack of relief facilities.

Finally, the results show that decreasing the capacity of distribution centers increases the amount of shortage of supplies and increasing the capacity of these centers reduces the amount of shortage of supplies. Objective functions also indicate that 36 people have been left unserved and that there are 420 units of shortage of supplies in different affected areas.

This research provides some practical implications from the model. First of all, the objective functions are not affected by each other and each of them individually improves the flow of the injured and relief supplies in the relief network. By optimizing the allocation and routing of different vehicles in the network to distribute relief supplies and evacuate the injured, this model reduces the amount of shortage of supplies and minimizes the number of the injured waiting to be served. Also, an increase in the number and capacity of vehicles can also have a significant impact on the values of the objective functions, but this managerial decision must be made considering system constraints.

The rest of managerial insights can be referred into the road restoration and relief distribution which has not been adequately addressed in past literature. The proposed relief model can provide reliable and fast communications to improve disaster rescue efficiency and road repair, as well as satisfy the victims' psychological needs. The integrated model provides an effective response decision with less total relief time, and higher rescue efficiency especially for large-scale disasters.

At last but not least, this study opens several new directions into this research area. Following suggestions are highly recommended for future studies.

- Choosing the most appropriate vehicle routing policy (such as vehicle routing with time windows), is one of potential directions of this paper.
- Adding more sustainability [35] and or resiliency [36] dimensions to the proposed model opens several new avenues for future works.
- Using heuristic and meta-heuristic solution methods to optimize the problem in a very large scale is highly recommended. We can especially suggest red deer algorithm [37] or social engineering optimizer [38] as two well-known and recent meta-heuristics [39].
- Considering the complete time of the disaster and dividing it into different periods to consider the dynamism of the disaster situation, is another good continuation of this work.

8. REFERENCES

1. Paul, J. A. and Zhang, M., “Supply location and transportation planning for hurricanes: A two-stage stochastic programming framework”, ***European Journal of Operational Research***, Vol. 274, No. 1, (2019), 108–125. doi:10.1016/j.ejor.2018.09.042
2. Paul, J. A. and Wang, X. (Jocelyn), “Robust location-allocation network design for earthquake preparedness”, ***Transportation Research Part B: Methodological***, Vol. 119, (2019), 139–155. doi:10.1016/j.trb.2018.11.009
3. Loree, N. and Aros-Vera, F., “Points of distribution location and inventory management model for Post-Disaster Humanitarian Logistics”, ***Transportation Research Part E: Logistics and Transportation Review***, Vol. 116, (2018), 1–24. doi:10.1016/j.tre.2018.05.003
4. Fathalikhani, S., Hafezalkotob, A., and Soltani, R., “Government intervention on cooperation, competition, and coopetition of humanitarian supply chains”, ***Socio-Economic Planning Sciences***, Vol. 69, (2020). doi:10.1016/j.seps.2019.05.006
5. Cao, C., Li, C., Yang, Q., Liu, Y., and Qu, T., “A novel multi-objective programming model of relief distribution for sustainable disaster supply chain in large-scale natural disasters”, ***Journal of Cleaner Production***, Vol. 174, (2018), 1422–1435. doi:10.1016/j.jclepro.2017.11.037
6. Davoodi, S. M. R. and Goli, A., “An integrated disaster relief model based on covering tour using hybrid Benders decomposition and variable neighborhood search: Application in the Iranian context”, ***Computers and Industrial Engineering***, Vol. 130, (2019), 370–380. doi:10.1016/j.cie.2019.02.040
7. Noham, R. and Tzur, M., “Designing humanitarian supply chains by incorporating actual post-disaster decisions”, ***European Journal of Operational Research***, Vol. 265, No. 3, (2018), 1064–1077. doi:10.1016/j.ejor.2017.08.042
8. Haghi, M., Fatemi Ghomi, S. M. T., and Jolai, F., “Developing a robust multi-objective model for pre/post disaster times under uncertainty in demand and resource”, ***Journal of Cleaner Production***, Vol. 154, (2017), 188–202. doi:10.1016/j.jclepro.2017.03.102
9. Gu, J., Zhou, Y., Das, A., Moon, I.M., and Lee, G., “Medical relief shelter location problem with patient severity under a limited relief budget”, ***Computers and Industrial Engineering***, Vol. 125, (2018), 720–728. doi:10.1016/j.cie.2018.03.027
10. Torabi, S. A., Shokr, I., Tofighi, S., and Heydari, J., “Integrated relief pre-positioning and procurement planning in humanitarian supply chains”, ***Transportation Research Part E: Logistics and Transportation Review***, Vol. 113, (2018), 123–146. doi:10.1016/j.tre.2018.03.012
11. Hajiaghaei-Keshteli, M., Mohammadzadeha, H., and Fathollahi Fard, A. M., “New Approaches in Metaheuristics to Solve the Truck Scheduling Problem in a Cross-docking Center”, ***International Journal of Engineering, Transactions B: Applications***, Vol. 31, No. 8, (2018), 1258–1266. doi:10.5829/ije.2018.31.08b.14
12. Nagurney, A., Salarpour, M., and Daniele, P., “An integrated financial and logistical game theory model for humanitarian organizations with purchasing costs, multiple freight service providers, and budget, capacity, and demand constraints”, ***International Journal of Production Economics***, Vol. 212, (2019), 212–226. doi:10.1016/j.ijpe.2019.02.006
13. Liu, Y., Lei, H., Wu, Z., and Zhang, D., “A robust model predictive control approach for post-disaster relief distribution”, ***Computers and Industrial Engineering***, Vol. 135, (2019), 1253–1270. doi:10.1016/j.cie.2018.09.005
14. Fathollahi-Fard, A. M., Hajiaghaei-Keshteli, M., and Tavakkoli-Moghaddam, R., “A Lagrangian Relaxation-based Algorithm to Solve a Home Health Care Routing Problem”, ***International Journal of Engineering, Transactions, A: Basics***, Vol. 31, No. 10, (2018), 1734–1740. doi:10.5829/ije.2018.31.10a.16
15. Abdalzaher, M. S. and Elsayed, H. A., “Employing data communication networks for managing safer evacuation during earthquake disaster”, ***Simulation Modelling Practice and Theory***, Vol. 94, (2019), 379–394.

doi:10.1016/j.simpat.2019.03.010

16. Abdi, A., Abdi, A., Fathollahi-Fard, A.M., and Hajiaghaei-Keshteli, M., "A set of calibrated metaheuristics to address a closed-loop supply chain network design problem under uncertainty", ***International Journal of Systems Science: Operations and Logistics***, (2019), 1–18. doi:10.1080/23302674.2019.1610197

17. Fathollahi-Fard, A.M., Hajiaghaei-Keshteli, M., Tian, G. and Li, Z., "An adaptive Lagrangian relaxation-based algorithm for a coordinated water supply and wastewater collection network design problem", ***Information Sciences***, Vol. 512, (2020), 1335–1359. doi:10.1016/j.ins.2019.10.062

18. Fathalikhani, S., Hafezalkotob, A., and Soltani, R., "Cooperation and coopetition among humanitarian organizations: A game theory approach", ***Kybernetes***, Vol. 47, No. 8, (2018), 1642–1663. doi:10.1108/K-10-2017-0369

19. Fu, Y., Tian, G., Fathollahi-Fard, A.M., Ahmadi, A. and Zhang, C., "Stochastic multi-objective modelling and optimization of an energy-conscious distributed permutation flow shop scheduling problem with the total tardiness constraint", ***Journal of Cleaner Production***, Vol. 226, (2019), 515–525. doi:10.1016/j.jclepro.2019.04.046

20. Tavana, M., Abtahi, A.R., Di Caprio, D., Hashemi, R. and Yousefi-Zenouz, R., "An integrated location-inventory-routing humanitarian supply chain network with pre- and post-disaster management considerations", ***Socio-Economic Planning Sciences***, Vol. 64, (2018), 21–37. doi:10.1016/j.seps.2017.12.004

21. Safaeian, M., Fathollahi-Fard, A.M., Tian, G., Li, Z. and Ke, H., "A multi-objective supplier selection and order allocation through incremental discount in a fuzzy environment", ***Journal of Intelligent and Fuzzy Systems***, Vol. 37, No. 1, (2019), 1435–1455. doi:10.3233/JIFS-182843

22. Fathollahi-Fard, A. M., Hajiaghaei-Keshteli, M., and Mirjalili, S., "A set of efficient heuristics for a home healthcare problem", ***Neural Computing and Applications***, Vol. 32, No. 10, (2020), 6185–6205. doi:10.1007/s00521-019-04126-8

23. Noyan, N. and Kahvecioğlu, G., "Stochastic last mile relief network design with resource reallocation", ***OR Spectrum***, Vol. 40, No. 1, (2018), 187–231. doi:10.1007/s00291-017-0498-7

24. Fathollahi-Fard, A.M., Govindan, K., Hajiaghaei-Keshteli, M. and Ahmadi, A., "A green home health care supply chain: New modified simulated annealing algorithms", ***Journal of Cleaner Production***, Vol. 240, (2019), 118200. doi:10.1016/j.jclepro.2019.118200

25. Li, H., Zhao, L., Huang, R., and Hu, Q., "Hierarchical earthquake shelter planning in urban areas: A case for Shanghai in China", ***International Journal of Disaster Risk Reduction***, Vol. 22, (2017), 431–446. doi:10.1016/j.ijdrr.2017.01.007

26. Bahadori-Chinibelagh, S., Fathollahi-Fard, A. M., and Hajiaghaei-Keshteli, M., "Two Constructive Algorithms to Address a Multi-Depot Home Healthcare Routing Problem", ***IETE Journal of Research***, (2019), 1–7. doi:10.1080/03772063.2019.1642802

27. Khojasteh, S. B. and Macit, I., "A Stochastic Programming Model for Decision-Making Concerning Medical Supply Location and Allocation in Disaster Management", ***Disaster Medicine and Public Health Preparedness***, Vol. 11, No. 6, (2017), 747–755. doi:10.1017/dmp.2017.9

28. Feng, Y., Zhang, Z., Tian, G., Fathollahi-Fard, A.M., Hao, N., Li, Z., Wang, W. and Tan, J., "A Novel Hybrid Fuzzy Grey TOPSIS Method: Supplier Evaluation of a Collaborative Manufacturing Enterprise", ***Applied Sciences***, Vol. 9, No. 18, (2019), 3770. doi:10.3390/app9183770

29. Fathollahi-Fard, A. M., Niaz Azari, M., and Hajiaghaei-Keshteli, M., "An Improved Red Deer Algorithm to Address a Direct Current Brushless Motor Design Problem", ***Scientia Iranica***, (2019) doi:10.24200/sci.2019.51909.2419

30. Ghasemi, P., Khalili-Damghani, K., Hafezalkotob, A. and Raissi, S., "Stochastic optimization model for distribution and evacuation planning (A case study of Tehran earthquake)", ***Socio-Economic Planning Sciences***, Vol. 71, (2019) doi:10.1016/j.seps.2019.100745

31. Fathollahi-Fard, A.M., Ranjbar-Bourani, M., Cheikhrouhou, N. and Hajiaghaei-Keshteli, M., "Novel modifications of social engineering optimizer to solve a truck scheduling problem in a cross-docking system", ***Computers and Industrial Engineering***, Vol. 137, (2019). doi:10.1016/j.cie.2019.106103

32. Ghasemi, P., Khalili-Damghani, K., Hafezalkotob, A. and Raissi, S., "Uncertain multi-objective multi-commodity multi-period multi-vehicle location-allocation model for earthquake evacuation planning", ***Applied Mathematics and Computation***, Vol. 350, (2019), 105–132. doi:10.1016/j.amc.2018.12.061

33. Torabi, N., Tavakkoli-Moghaddam, R. and Najafi, E., "A Two-Stage Green Supply Chain Network with a Carbon Emission Price by a Multi-objective Interior Search Algorithm", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 32, No. 6, (2019), 828–834. doi:10.5829/ije.2019.32.06c.05

34. Mehranfar, N., Hajiaghaei-Keshteli, M., and Fathollahi-Fard, A. M., "A Novel Hybrid Whale Optimization Algorithm to Solve a Production-Distribution Network Problem Considering Carbon Emissions", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 32, No. 12, (2019), 1781–1789. doi:10.5829/ije.2019.32.12c.11

35. Liu, X., Tian, G., Fathollahi-Fard, A.M. and Mojtahedi, M., "Evaluation of ship's green degree using a novel hybrid approach combining group fuzzy entropy and cloud technique for the order of preference by similarity to the ideal solution theory", ***Clean Technologies and Environmental Policy***, Vol. 22, No. 2, (2020), 493–512. doi:10.1007/s10098-019-01798-7

36. Safaei, A. S., Farsad, S., and Paydar, M. M., "Robust bi-level optimization of relief logistics operations", ***Applied Mathematical Modelling***, Vol. 56, (2018), 359–380. doi:10.1016/j.apm.2017.12.003

37. Fathollahi-Fard, A. M., Hajiaghaei-Keshteli, M., and Tavakkoli-Moghaddam, R., "The Social Engineering Optimizer (SEO)", ***Engineering Applications of Artificial Intelligence***, Vol. 72, (2018), 267–293. doi:10.1016/j.engappai.2018.04.009

38. Fathollahi-Fard, A. M., Hajiaghaei-Keshteli, M., and Tavakkoli-Moghaddam, R., "Red deer algorithm (RDA): a new nature-inspired meta-heuristic", ***Soft Computing***, (2020), 1–29. doi:10.1007/s00500-020-04812-z

39. Fathollahi-Fard, A.M., Ahmadi, A., Goodarzian, F. and Cheikhrouhou, N.,, "A bi-objective home healthcare routing and scheduling problem considering patients' satisfaction in a fuzzy environment", ***Applied Soft Computing Journal***, Vol. 93, (2020). doi:10.1016/j.asoc.2020.106385

Persian Abstract

چکیده

در طول تاریخ، طبیعت بسیاری از بلایای طبیعی نظیر زلزله، سیل، خشکسالی، گردباد، طوفان‌های دریایی و سونامی را به بشر تحمیل کرده است. مقیاس گسترده خسارات و تلافات ناشی از بلایای طبیعی در سراسر جهان باعث شده است که تحقیقات گسترده‌ایی در زمینه تهیه‌ی یک سیستم جامع برای مدیریت بلایا انجام شود تا تلافات و خسارات مالی به حداقل برسد. بر اساس این انگیزه و چالش‌های موجود در این زمینه، این پژوهش یک زنجیره امداد یکپارچه را طراحی کرده است تا همزمان آمادگی و پاسخگویی برای مدیریت بحران بهینه کند. تصمیمات مربوط به بهینه‌سازی زنجیره تامین شامل استقرار مراکز توزیع منابع امدادی، میزان موجودی در تسهیلات در مرحله قبل از فاجعه، مکانیابی مراکز مراقبت موقت و نقاط حمل و نقل مصدومان، نحوه تخصیص خدمات امدادی، حمل و نقل مصدومان، مسیریابی وسائل نقلیه مورد استفاده برای توزیع منابع امدادی و تخلیه مصدومان هستند. نتایج نشان می‌دهد که کاهش ظرفیت مراکز توزیع باعث افزایش کمبود منابع می‌شود. و افزایش ظرفیت این مراکز می‌تواند راه‌حل مناسبی باشد.

IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1285-1292

International Journal of Engineering

Journal Homepage: www.ije.ir

A Statistical Method for Sequential Images–based Process Monitoring

M. A. Fattahzadeh, A. Saghaei*

Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

PAPER INFO

Paper history:
Received 09 December 2019
Received in revised form 24 April 2020
Accepted 11 June 2020

Keywords:
Control Chart
Image Processing
Real-Time Monitoring
Ridge Regression
Statistical Process Control

ABSTRACT

Today, with the growth of technology, monitoring processes by the use of video and satellite sensors have been more expanded, due to their rich and valuable information. Recently, some researchers have used sequential images for defect detection because a single image is not sufficient for process monitoring. In this paper, by adding the time dimension to the image-based process monitoring problem, we detect process changes (such as the changes in the size, location, speed, color, etc.). The temporal correlation between the images and the high dimensionality of the data make this a complex problem. To address this, using the sequential images, a statistical approach with RIDGE regression and a Q control chart is proposed to monitor the process. This method can be applied to color and gray images. To validate the proposed method, it was applied to a real case study and was compared to the best methods in literature. The obtained results showed that it was more effective in finding the changes.

doi*: 10.5829/ije.2020.33.07a.15*

NOMENCLATURE

T	Time	$\|\cdot\|_2$	L_2 norm
N	Number of the in-control samples	r_t	Residual matrix at time t
B_t	Foreground image at time t (object image)	R_{new}	Vectorized form of r_t for new sample
$\hat{B}_t$	Predicted foreground image at time t	Q_{new}	Q-chart statistic
F	Transfer matrix	q	Degrees of freedom in χ^2 distribution
γ	Tuning parameter	p	The coefficient given to the Q-chart statistic

1. INTRODUCTION

Control chart, as a statistical tool, is widely used in monitoring quality characteristic(s) of a process or product. It was first introduced by Walter A. Shewhart in the 1920s. Shewhart used only one quality characteristic, such as length or weight, to monitor a process. Later, researchers developed multivariate models based on multiple characteristics. Woodall et al. [1] presented profile monitoring, which is used in many practical situations. Concurrently with the quality control methods, sensor technology, including image sensors, was developed. Process monitoring by these sensors has various applications in manufacturing processes, natural phenomena, medical decision-making, and sports activities. Many image-based methods have been developed for defect and fault detection [2–6]. However, some processes cannot be monitored by image-based methods, and we need to use sequential images. Of course, some researchers used sequential images but their purpose was only to detect faults in one image [7].

In this paper, we address sequential images–based process monitoring problem for processes that need more than one image. In this problem, the objective is to detect process changes that occur, for example, in the position, speed, shape, and color by using at least two images. Sequential images can be used in many contexts, e.g., in Welding (Figure 1a), Fabric texture (Figure 1b), Eddy phenomenon (Figure 1c), and Solar flare (Figure 1d) [7–9].

This problem has complex characteristics, including 1) high dimensionality, where some sequential images

*Corresponding Author Email: *a.saghaei@srbiau.ac.ir* (A. Saghaei)

Please cite this article as: M. A. Fattahzadeh, A. Saghaei, A Statistical Method for Sequential Images–based Process Monitoring, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1285-1292

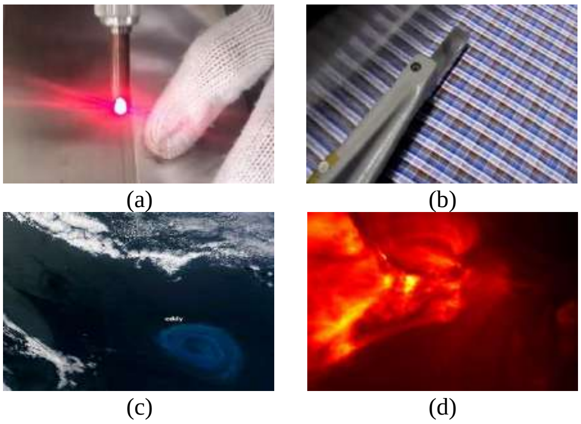


Figure 1. Applications of the sequential images-based processes; a) Welding and laser welding, b) Fabric texture, c) Eddy phenomenon (Image courtesy of the NASA Earth Observatory), d) Solar flare

are at least 0.5M pixels; 2) the correlation between the images; 3) spatial and temporal structure: pixels are spatially correlated within an image and, most of the time, are temporally correlated across the sequential images [7].

Our proposed methodology, inspired by data stream monitoring [7], can handle both grayscale and color images. We present a new method that predicts the behavior of objects in the next image and uses the prediction error to monitor the shape, color, and speed changes of objects.

The remainder of this paper is organized as follows. Section 2 provides a review of the relevant literature. Section 3 introduces the proposed method. In Section 4, we illustrate and evaluate how our proposed method can find changes in a real case. In Section 5, the paper is concluded and some future research areas are provided.

2. LITERATURE REVIEW

In this paper, a common procedure in machine-vision systems [10] is used to monitor the changes in the sequential images. In this procedure, first, image data are collected by the corresponding sensor. Then, the data are preprocessed, by background removal, compressing, denoising, etc.. After that, a set of monitoring features are extracted from each image. Finally, the extracted features are monitored by statistical or engineering methods.

Zou et al. [11] developed a powerful method for monitoring independent data streams. They assumed that the data streams were independent, and therefore, ignored their spatial and temporal structures. The monitoring of data streams with a temporal trend was addressed by Xiang et al. [12], and Qiu and Xiang [13]. They used nonparametric regression with longitudinal techniques. However, they did not address the spatial structure. Yan et al. [7] proposed a novel methodology based on spatio-temporal decomposition for data streams monitoring. Using the lasso method, the image and profile were decomposed to the functional mean, sparse anomalies, and random noises. For the validation of this method, solar activity and steel rolling process were used. They detected the defects in the process, seams in the process of steel rolling, and flares in solar activities. Bračun and Sluga [8] used a stereo monitor for the welding track sensing system. They measured the position of the arc in the 3D space and in the time sequence. They used two high-speed cameras, calculated the center of the arc, and finally saved the direction of motion. Similar to Yan et al. [7], they addressed defect anomalies. This method was designed only for the welding path and is inefficient for other processes. Faghmous et al. [9] provided a parameter-free spatio-temporal model for the detection and trace of an eddy in an ocean. This method was based on finding extreme points in the neighborhood.

For monitoring simple multivariate problems, multiple methods such as T^2, Q-chart, MCUSUM, EWMA, etc. [14–17] can be used. But according to Megahed et al. [18], for monitoring matrix or tensor data like image and satellite data, the monitoring methods are divided into several groups including profile-based and multivariate techniques, multivariate image analysis (MIA) control charts, and spatial control charts.

In the first group, multivariate control charts are used with a feature extraction method. For example, Wang and Tsung [19] modeled the relationship between a baseline and a sampled image using Q-Q plots and used profile-monitoring for changes detection. However, in their method, the information about pixel locations was ignored. In MIA, the features of each color channel were extracted by using partial least squares regression or principal component analysis (PCA). Yan et al. [3] used Low-Rank Tensor Decomposition (LRTD) for monitoring an image-based process. They used multi-linear PCA (MPCA) to extract features and proposed a combined control chart, based on T^2 and Q charts.

The spatial control charts use non-overlapping windows. These windows move across an image to obtain spatial information [20]. Jiang et al. [21] used ANOVA technique for spatial monitoring of the LCD panels and exponentially weighted moving average (EWMA) control chart for detecting the defects in grayscale images.

Some processes are not static and have specific movements. For example, in the fabric weaving process, if the production speed changes or even if the machine is stopped, it can affect the quality of the produced fabric. Image-based methods cannot detect these changes because they monitor the process using only one image. They cannot detect factors such as changes in speed or pattern because they do not pay attention to the time dimension.

The main contribution of this paper is the adding of the time dimension to image-based process monitoring and developing sequential images-based methods for detecting process changes. This can, in addition to the changes in shape and color, detect the changes in the movement pattern such its speed, acceleration, and direction. So, we propose a general statistical method with ridge regression applying previous images in monitoring. A real case study is proposed to evaluate the performance of the proposed method. Also, we compared the proposed method with some image-based methods in literature. Besides the mentioned applications in Figure 1, the sequential images-based process monitoring has various applications, including manufacturing (such as glass forming, steel rolling), traffic control (such as identifying high-risk behaviors), food industries (such as bread baking), and etc.

3. PROPOSED METHOD

An overview of the proposed method is shown in Figure 2. It consists of four steps. The first step is image acquisition, in which the input data as the sequential images are required to be divided into separate images. The next step is the data preprocessing, where first the background is removed and then if the image is noisy, the noise is removed. Step three is feature extraction, in which we estimate a transfer matrix that can predict the next image. The differences between the predicted image and the actual image constitute the residuals matrix. In the final step, these residuals are monitored by the multivariate control charts.

The proposed method has several variables and parameters, shown in Nomenclature.

3. 1. Image Acquisition In this problem, the input data is in the form of sequential images. Therefore, they must be separated without changing the sequence of the images.

3. 2. Preprocessing The images obtained from the previous step consists of two parts: the foreground (or object) and the background. The background should be removed and only the foreground be remain because the aim is the monitoring of the object changes. The selection of the algorithm we should use for the background removal depends on many factors, such as the data type, whether the background is dynamic or static, whether it is smooth or non-smooth, whether the camera is fixed or not, etc. For example, for removing a static background captured with a fixed camera, background subtraction is a suitable method.

Sometimes the images after the background removal are noisy. This noise should be eliminated. Several methods can be used for this, such as median filtering and Gaussian smoothing. See [22] for more details about the noise and noise removal. The obtained foreground image after denoising at time *t* is denoted by B_t.

3. 3. Feature Extraction In this step, it is assumed that *N* in-control samples are available and each foreground image at time *t* + 1 (B_{t+1}) can be predicted by the previous foreground image (B_t). Although more previous images can be used for prediction, but for simplicity, only the latest one is used. The predicted foreground image at time *t* + 1 is denoted by $\hat{B}_{t+1}$, which can be used by Equation (1).

$$\hat{B}_{t+1} = B_t \times F \tag{1}$$

To calculate $\hat{B}_{t+1}$ using $\mathrm{B_t}$, we need a transfer matrix, illustrated by F. In Equation (2), γ and $\|.\|_2$ denotes respectively the tuning parameter and L_2 norm. This transfer matrix is estimated by N sequential in-control samples.

$$argmin_{F_t} \sum_{t=1}^{N-1} \left\|\hat{B}_{t+1} - B_{t+1}\right\|_2^2 + \gamma\|F\|_2 \tag{2}$$

This equation is a ridge formulation. Since both parts of Equation (2) are differentiable, we use differentiation to optimize this equation. Thus, F could be optimized by Equation (3). In Equation (3), *I* is defined as the identity matrix.

$$F = \left(\sum_{z=1}^{N-1} B_{t,z} * {B_{t,z}}^T + \gamma I\right)^{-1} \sum_{z=1}^{N-1} B_{t,z} * B_{t+1,z} \tag{3}$$

The residuals as the difference between the actual and predicted images are obtained by Equation (4) for all images. Let's define r_t as a residual matrix.

$$r_t = B_t - \hat{B}_t \qquad \forall\, t = 2, 3, \dots, N \tag{4}$$

3. 4. Monitoring The in-control samples are used to calculate the transfer matrix and control limits. Here, we use the Q-chart for monitoring the residuals. The Q-chart statistic is based on the residual matrix obtained from Equation (4).

For each new sample, the estimated transfer matrix is used for the calculation of the residuals and plotting the monitoring statistic on the designed Q-chart.

The residual vector of the new sample is represented by $R_{new} = vec(r_t)$. The Q-chart statistic (Q_{new}) is obtained using Equation (5).

$$Q_{new} = \|R_{new}\|_2^2 \tag{5}$$

The residuals are assumed to follow a multivariate normal distribution, and therefore Q_{new}/p follows a χ_q^2 distribution, where *q* denotes the degrees of freedom and p is the coefficient given to the statistic to follow the known distribution function χ_q^2. The parameters p and q can be estimated by the moments method [23]. They can be obtained by solving Equations (6) and (7).

$$E(Q_{new}) = pq \tag{6}$$

$$Var(Q_{new}) = 2qp^2 \tag{7}$$

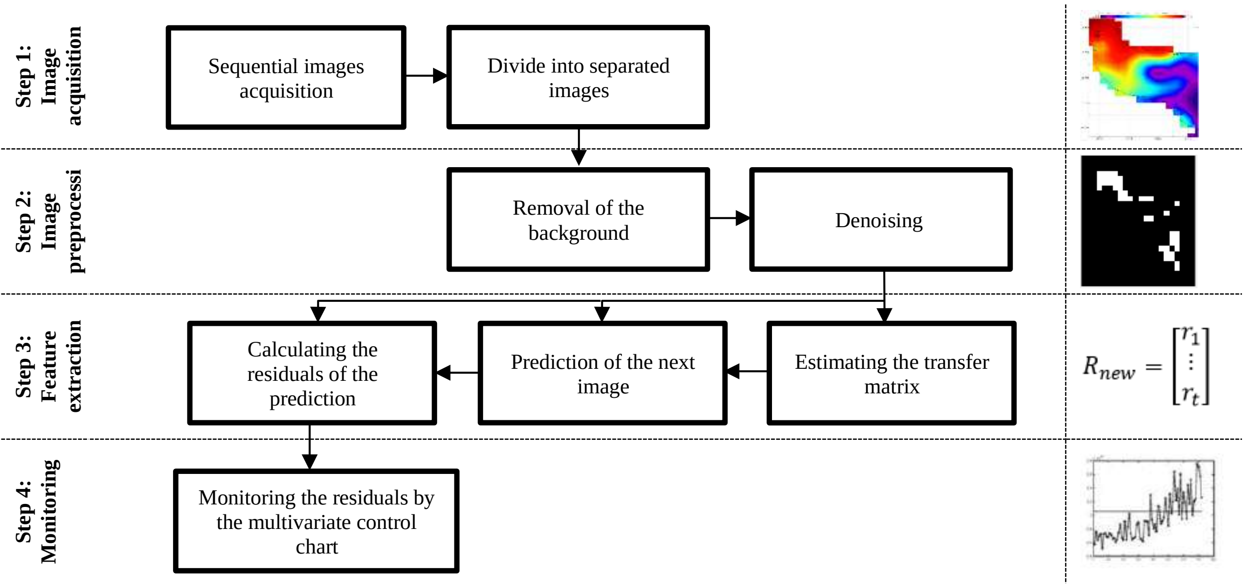

Step 1: Image acquisition
Sequential images acquisition
Divide into separated images
Step 2: Image preprocessi
Removal of the background
Denoising
Step 3: Feature extraction
Calculating the residuals of the prediction
Prediction of the next image
Estimating the transfer matrix
$R_{new} = \begin{bmatrix} r_1 \\ \vdots \\ r_t \end{bmatrix}$
Step 4: Monitoring
Monitoring the residuals by the multivariate control chart

Figure 2. The overview of proposed method

The Q-chart control limit can be calculated by $(1-\alpha)\%$ of the χ^2 distribution with q degrees of freedom.

4. CASE STUDY

4. 1. Case Description

Mesoscale eddies are coherent rotating vortices of water with a span of 25–250 kilometers (Figure 3) lasting 10 to 100 days [9]. Eddies are critical phenomena with an important role in dominating the ocean's kinetic energy. They are responsible for the transport and mixing of heat, salt, nutrients, and energy across an ocean or sea [24]. Moreover, they have a significant impact on terrestrial and marine ecosystems [25]. The creation or growth of eddy provides a large amount of food for phytoplankton and provides them with growth opportunity. This can cause serious damage to the region's ecosystem, such as massive aquatic mortality. Moreover, it can stop tourism activities in the region. In this case, the growth of phytoplankton must be counteracted. For example, coral reefs were destroyed over seven thousand years due to the high growth of phytoplankton, enhanced by an eddy [26]. Eddy changes are urgently needed to be discovered because phytoplankton populations impose damage to the ecosystem and make changes in water flows, which transmit pollution and sea anomalies damaging the maritime tourism industry.

Thus, in this paper, to validate the proposed method, the eddies in the Oman Sea are monitored and unusual behaviors of eddy properties are detected. Position, velocity, size, and height are defined as eddy properties [27].

4. 2. Implementing the Proposed Method

4. 2. 1. Image Acquisition

In step (1), the input data are converted to separate sequential images. We use satellite data from the AVISO dataset. This dataset is publicly available online at https://las.aviso.altimetry.fr/las/UI.vm. In this dataset, Latitude and Longitude of the Oman Sea specified by 22° to 27° and 56° to 60° respectively, on a 0.25° (~28km) grid. These data are weekly sequential matrices with a size of 20 × 16 pixels collected from 1993/01/01 to 2018/12/31. Thus, we need to separate them without disrupting their order.

The value of each pixel represents the sea surface height (SSH) in meter. A sample of 25 years, consisting of 1357 weeks, is used and the first 3 years are considered as the in-control sample.

Most of the time, SSH data are reported as a vector, so we need to reshape them into a rectangle.

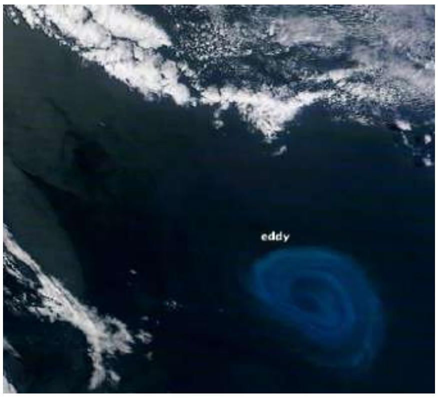


Figure 3. A mesoscale eddy

4. 2. 2. Preprocessing In step (2), the data is preprocessed for further analysis. The output of the previous step is illustrated in Figure 4.

Here, we need to separate the foreground from the background, where eddies are our foreground, required to be monitored. One of the best methods for separating eddies from the background is the method introduced by Faghmous et al. [9]. Figure 5 illustrates the eddies extracted from the image shown in Figure 4. The outputs of this algorithm are not noisy, so we do not need any denoising.

4. 2. 3. Feature Extraction The main objective in step 3 is the estimation of the transfer matrices. It is assumed here that eddies behavior is independent in each month. Therefore, one transfer matrix should be calculated for each month (i.e., a transfer matrix is used to predict the images of each month) and one transfer matrix should be calculated for when the month changes, to predict the first week of the new month (this is due to seasonal changes). Therefore, we need to find 24 transfer matrices, which are of two types; Type I: transfer matrices for each month of the year (totally, 12 transfer matrices), and Type II: transfer matrices for when the month changes (totally, 12 transfer matrices). To estimate these transfer matrices, Equation (2) is made by $\gamma = 0.01$ and 3 sets of $N = 4$ in-control samples for Type I (for example, for estimating transfer matrix of April, we use the weekly data of the Aprils in the three in-control years), and 3 sets of $N = 2$ in-control samples for Type II.

When the prediction of the eddy matrix is obtained, the difference between the actual images and the predictions constitutes the residual matrix.

4. 2. 4. Monitoring In step 4, we encounter two different sample sizes, each of which requires two control charts. The in-control samples are used to calculate the control limits of the Q-charts. Then, for each new sample, first, the transfer matrix is selected, then the residuals are calculated, and finally, the statistic is calculated and plotted on the Q-chart.

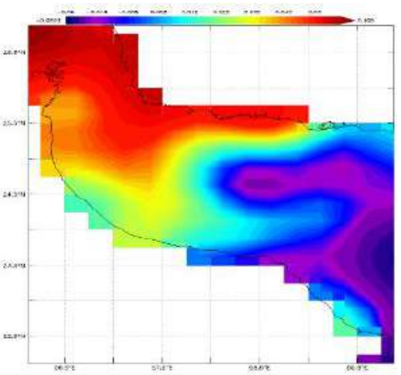


Figure 4. The original image 2008/09/01

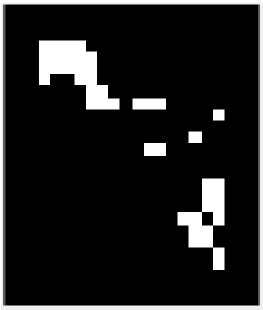


Figure 5. Extracted eddies from the original image 2008/09/01

The control charts for both types are shown in Figure 6. In these two control charts, only the data for the in-control years and two new years (2007 and 2014) are shown. The complete control charts (from 1993 to 2018) are available at bit.ly/2pqZlhi.

4. 3. The Proposed Method Results To evaluate the performance of the proposed method, we examine some of the storms that occurred at that time. One of the out-of-control samples is 2007/06/04. This out-of-control state was caused by the Gonu storm. This storm entered Iran at 2007/06/04 and the control chart detected an out-of-control sample at exactly the same date. The large quantities in the future weeks were due to the storm.

Out-of-control samples were observed in 2014/06/16 and 2014/06/23, which were occurred due to the Nanauk storm. It was started at 2014/06/10, one day after the last day the satellite recorded its information. Regarding the weekly structure of the data, the proposed method was able to correctly identify the out-of-controlled state.

The Nilofar Storm, which began operating in the Arabian Sea at 2014/10/25 towards the Oman Sea, was active until the 2014/10/31. The control charts discovered the first out-of-control sample on 2014/11/03 and the Nilofar Storm may be the reason for other out-of-control states in the coming weeks.

In Table 1, all storms of the Oman Sea and important storms in the west of the Arabian Sea with speed more than 100 km/h in peak time are indexed because it is assumed that these storms can affect the Oman Sea eddies. The proposed method discovered all storms in the Oman Sea and in the north and middle of the Arabian Sea. Also, it discovered most of the storms in the south of the Arabian Sea.

4. 4. Comparison Study To demonstrate the performance of the proposed method, some statistical feature extraction methods, kernel-PCA (KPCA) and PCA, are applied. To monitor the extracted features of PCA and KPCA methods, the T^2 control chart is used. Also, based on our best knowledge, the LRTD method [18] is the most powerful method that can accurately detect the smallest changes in the images, considering the

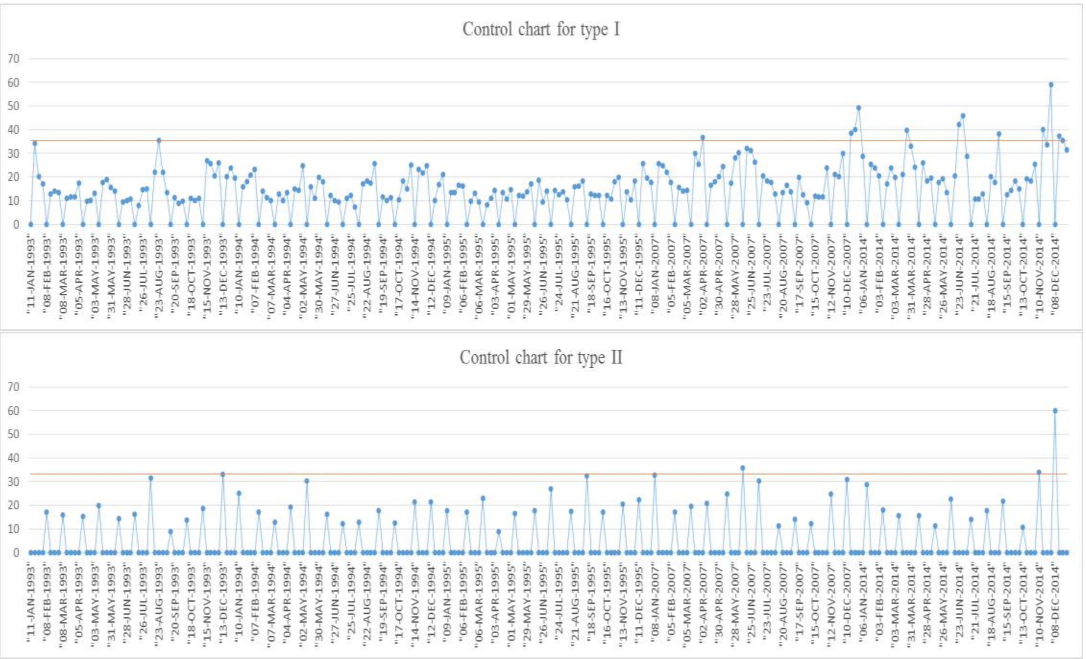


Figure 6. Q control charts for eddy monitoring (Type I and Type II)

seasonal effects. The reason why we use these image-based methods in this comparison study is that the existing sequential image-based methods have been designed either for a specific process or for fault detection in images. Therefore, we cannot use these methods.

As shown in Table 1, twelve major storms occurred in the specified time and place. Among the methods used for comparison, the results of PCA and KPCA methods are very weak and identified less than 6 cases. The LRTD method performed relatively well and identified 8 cases, while the proposed method was able to identify 10 out of the twelve cases, representing its better performance.

The reason why these methods, and even the LRTD method which is a powerful method, performed worse than the proposed method, is that these methods have not considered the time dimension. Ignoring the time dimension in the monitoring of eddies leads to the loss of eddies motion information. Eddies can move, rotate, be resized, and have a variable height over time, and neglecting this information causes that time-dependent changes do not be identified correctly.

TABLE 1. Storms in the Oman Sea and important storms near the Oman Sea

Year	Date	Speed in Peak (km/h)	Place	Proposed method	LRTD	KPCA-T^2	PCA-T^2
1998	11 - 17 DEC	100	Middle of the Arabian Sea	□	□	□	×
2007	1 – 7 JUN	235	The Oman Sea	□	□	□	×
2010	30 MAY – 7 JUN	155	The Oman Sea	□	□	□	□
2014	10 – 14 JUN	85	North of the Arabian Sea	□	□	×	×
2014	25 – 31 OCT	205	North of the Arabian Sea	□	×	×	□
2015	7 – 12 JUN	85	North of the Arabian Sea	□	□	×	×
2015	28 OCT – 4 NOV	215	South of the Arabian Sea	□	□	□	□
2015	5 – 10 NOV	175	South of the Arabian Sea	□	□	□	×
2016	6 - 18 DEC	130	South of the Arabian Sea	□	×	×	×
2018	21 – 27 MAY	175	South of the Arabian Sea	×	×	×	×
2018	6 – 15 OCT	140	South of the Arabian Sea	×	×	×	□
2018	10 – 20 NOV	110	South of the Arabian Sea	□	□	×	×

5. CONCLUSION

Image data are being increasingly used for monitoring different processes, such as manufacturing ones. Some processes are not static and have motion patterns. So, they cannot be monitored with a single image. For the image analysis of non-static processes and detecting changes such as speed, acceleration, and direction, adding time dimension and considering sequential images is essential and improves the results. On the other hand, the temporal correlation between the images, made by adding the time dimension, requires new analytical methods that can handle this correlation and provide better results than the image-based methods.

In this paper, we proposed a novel method combining RIDGE regression and multivariate control chart for sequential image–based process monitoring. In the proposed method, the features are extracted by a statistical method using RIDGE modeling and then the Q control chart is utilized for the monitoring of the residuals. We applied the proposed method to a case study, where the changes in the Oman Sea eddies made by storms were monitored and out-of-control samples were discussed. The proposed method was compared with the LRTD, KPCA, and PCA methods and the results showed that the proposed method, because of considering the time dimension and temporal effect between the images, performed better than the image-based methods and was able to detect more variations.

One important and challenging research topic that needs further study is finding the root cause of out-of-control samples and identifying their underlying factors, including the shape, number, and position of objects. Also, for simplicity, we considered only one previous image for prediction, which can be extended by considering more previous images.

6. REFERENCES

1. Woodall, W. H. Woodall, W.H., and Spitzner, D.J., Montgomery, D.C. and Gupta, S., "Using Control Charts to Monitor Process and Product Quality Profiles", ***Journal of Quality Technology***, Vol. 36, No. 3, (2004), 309-320. doi: 10.1080/00224065.2004.11980276
2. Bui, A.T., and Apley, D.W., "A monitoring and diagnostic approach for stochastic textured surfaces", ***Technometrics***, Vol. 60, No. 1, (2018), 1-13. doi: 10.1080/00401706.2017.1302362
3. Yan, H., Paynabar, K., and Shi, J., "Image-based process monitoring using low-rank tensor decomposition", ***IEEE Transactions on Automation Science and Engineering***, Vol, 12, No. 1, (2015), 216-227. doi: 10.1109/TASE.2014.2327029
4. Prats-Montalbán, J.M. and Ferrer, A., "Statistical process control based on Multivariate Image Analysis: A new proposal for monitoring and defect detection", ***Computers & Chemical Engineering***, Vol. 71, (2014), 501-511. doi: 10.1016/j.compchemeng.2014.09.014
5. Yu, H., MacGregor, J.F., Haarsma, G. and Bourg, W., "Digital imaging for online monitoring and control of industrial snack food processes", ***Industrial & Engineering Chemistry Research***, Vol.42, No. 13, (2003), 3036-3044. doi: 10.1021/ie020941f
6. Pereira, A. C., Reis, M. S., and Saraiva, P. M., "Quality control of food products using image analysis and multivariate statistical tools", ***Industrial & Engineering Chemistry Research***, Vol.48, No. 2, (2009), 988-998. doi:10.1021/ie071610b
7. Yan, H., Paynabar, K., and Shi, J., "Real-Time Monitoring of High-Dimensional Functional Data Streams via Spatio-Temporal Smooth Sparse Decomposition", ***Technometrics***, Vol. 60, No. 2, (2018), 181-197. doi:10.1080/00401706.2017.1346522
8. Bračun, D., and Sluga, A., "Stereo vision based measuring system for online welding path inspection", ***Journal of Materials Processing Technology***, Vol. 223, (2015), 328-336., doi:10.1016/j.jmatprotec.2015.04.023
9. Faghmous, J. H., Frenger, I., Yao, Y., Warmka, R., Lindell, A., and Kumar, V., "A daily global mesoscale ocean eddy dataset from satellite altimetry", ***Scientific Data***, Vol. 2, (2015), 150028. doi: 10.1038/sdata.2015.28
10. Duchesne, C., Liu, J.J., and MacGregor, J.F., "Multivariate image analysis in the process industries: A review", ***Chemometrics and Intelligent Laboratory Systems***, Vol. 117, (2012), 116–128. doi: 10.1016/j.chemolab.2012.04.003
11. Zou, C. Wang, Z., Zi, X., and Jiang, W.,"An efficient online monitoring method for high-dimensional data streams", ***Technometrics***, Vol. 57, No. 3, (2015), 374-387. doi:10.1080/00401706.2014.940089
12. Xiang, D., Qiu, P., and Pu, X., "Nonparametric regression analysis of multivariate longitudinal data", ***Statistica Sinica***, Vol. 23, No.2, (2013), 769-789. doi:10.5705/ss.2011.317
13. Qiu, P., and Xiang, D., "Univariate dynamic screening system: An approach for identifying individuals with irregular longitudinal behavior", ***Technometrics***, Vol. 56, No. 2, (2014), 248-260. doi:10.1080/00401706.2013.822423
14. Rasay, H., Fallahzaded, M.S., and Zaremehrjerdi, Y., "Application of multivariate control charts for condition based maintenance", ***International Journal of Engineering, Transactions A: Basics***, Vol. 31, No. 4, (2018), 597-604. doi:10.5829/ije.2018.31.04a.11
15. Akhavan Niaki, S.T., and Moeinzadeh, B.,"A multivariate quality control procedure in multistage production systems", ***International Journal of Engineering***, Vol. 10, No. 4, (1997), 191-208. http://www.ije.ir/article_71187.html
16. Akhavan Niaki, S.T., Houshmand, A.A., and Moeinzadeh, B., "On the performance of a multivariate control chart in multistage environment", ***International Journal of Engineering***, Vol. 14, No. 1, (2001), 49-64. http://www.ije.ir/article_71286.html
17. Abdella, G., Yang, K., and Alaeddini, A.,"Effect of location of explanatory variable on monitoring polynomial quality profiles", ***International Journal of Engineering-Transactions A: Basics***, Vol. 25, No. 2, (2012), 131-140. doi: 10.5829/idosi.ije.2012.25.02a.03
18. Megahed, F.M., Woodall, W.H., and Camelio, J.A., "A review and perspective on control charting with image data", ***Journal of Quality Technology***, Vol. 43, No. 2, (2011), 83–98. doi: 10.1080/00224065.2011.11917848
19. Wang, K., and Tsung, F., "Using profile monitoring techniques for a data-rich environment with huge sample size", ***Quality and Reliability Engineering International***, Vol. 21 No. 7, (2005), 677–688. doi:10.1002/qre.711
20. 10.Megahed, F.M., Wells, L.J., Camelio, J.A., and Woodall, W.H., "A spatiotemporal method for the monitoring of image data", ***Quality and Reliability Engineering International***, Vol. 28, No. 8, (2012), 967–980. doi:10.1002/qre.1287
21. Jiang, B. C., Wang, C. C., and Liu, H. C., "Liquid crystal display surface uniformity defect inspection using analysis of variance

and exponentially weighted moving average techniques," ***International Journal of Production Research***, Vol. 43, No. 1, (2005), 67–80. doi:10.1080/00207540412331285832

22. Hambal, A.M., Pei, Z. and Ishabailu, F.L., "Image noise reduction and filtering techniques", ***International Journal of Science and Research***, Vol. 3, (2017), 2033-2038. doi: 10.21275/25031706
23. Nomikos, P., and MacGregor, J. F., "Multivariate SPC charts for monitoring batch processes", ***Technometrics***, Vol. 37, No. 1 (1995), 41–59. doi:10.1080/00401706.1995.10485888
24. Fu, L.L., Chelton, D.B., Le Traon, P.Y., and Morrow, R., "Eddy dynamics from satellite altimetry", ***Oceanography***, Vol. 23, No. 4, (2010), 14-25. doi: 10.5670/oceanog.2010.02
25. Faghmous, J.H., Le, M., Uluyol, M., Kumar, V., and Chatterjee, S.,"A parameter-free spatio-temporal pattern mining model to catalog global ocean dynamics", In IEEE 13th International Conference on Data Mining, IEEE, (2013), 151-160. doi:10.1109/ICDM.2013.162
26. Rahul, P.R.C., Salvekar, P.S., Sahu, B.K., Nayak, S., and Kumar, T.S., "Role of a cyclonic eddy in the 7000-year-old mentawai coral reef death during the 1997 indian ocean dipole event", ***IEEE Geoscience and Remote Sensing Letters***, Vol. 7, No. 2, (2009), 296-300. doi: 10.1109/LGRS.2009.2033950
27. Vic, C., Roullet, G., Carton, X., and Capet, X., "Mesoscale dynamics in the Arabian Sea and a focus on the Great Whirl life cycle: A numerical investigation using ROMS", ***Journal of Geophysical Research: Oceans***, Vol. 119, No. 9, (2014), 6422-6443. doi: 10.1002/2014JC009857

Persian Abstract

چکیده

امروزه با رشد تکنولوژی، پایش فرآیندها با بکارگیری سنسورهای تصویری و ماهواره‌ای به دلیل اطلاعات غنی و ارزشمند، گسترش پیدا کرده است. به تازگی برخی محققان برای کشف عیوب در تصویر، از تصاویر متوالی استفاده کرده‌اند، چرا که تحلیل فرآیند به صورت مستقل و با استفاده از یک تصویر امکان‌پذیر نیست. این مقاله با افزودن بعد زمان به مسئله پایش فرآیند به کشف تغییرات حالت فرآیند (مانند اندازه، محل، سرعت، رنگ و ...) می‌پردازد. همبستگی زمانی بین تصویرها و ابعاد بالای داده‌ها باعث شده که این مسئله یک مسئله پیچیده محسوب شود. در این مقاله یک روش آماری با رویکرد مدل‌سازی رگرسیون تیغه‌ای و نمودار کنترل Q برای پایش فرآیند بر پایه تصاویر متوالی پیشنهاد شده است. این روش می‌تواند در تصاویر رنگی و خاکستری بکار برده شود. برای اعتبارسنجی مدل یک مثال واقعی استفاده شده است و عملکرد آن با بهترین روش‌های موجود در ادبیات موضوع مقایسه شده است. نتایج حاصله نشان داد که روش پیشنهادی از اثربخشی بیشتری برخوردار بوده و توانسته تغییرات بیشتری را شناسایی نماید.

International Journal of Engineering

Journal Homepage: www.ije.ir

Time Series Forecasting of Bitcoin Price Based on Autoregressive Integrated Moving Average and Machine Learning Approaches

M. Khedmati*[a], F. Seifi[a], M. J. Azizi[b]

[a] *Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran*
[b] *Daniel J. Epstein department of industrial and systems engineering, University of Southern California, Los Angeles, United States*

PAPER INFO

Paper history:
Received 19 January 2020
Received in revised form 13 March 2020
Accepted 11 June 2020

Keywords:
Time Series Forecasting
Machine Learning
Bitcoin
Multivariate Models

ABSTRACT

Bitcoin as the current leader in cryptocurrencies is a new asset class receiving significant attention in the financial and investment community and presents an interesting time series prediction problem. In this paper, some forecasting models based on classical like ARIMA and machine learning approaches including Kriging, Artificial Neural Network (ANN), Bayesian method, Support Vector Machine (SVM) and Random Forest (RF) are proposed and analyzed for modelling and forecasting the Bitcoin price. While some of the proposed models are univariate, the other models are multivariate and as a result, the maximum, minimum and the opening daily price of Bitcoin are also used in these models. The proposed models are applied on the Bitcoin price from December 18, 2019 to March 1, 2020 and their performances are compared in terms of the performance measures of RMSE and MAPE by Diebold-Mariano statistical test. Based on RMSE and MAPE measures, the results show that SVM provides the best performance among all the models. In addition, ARIMA and Bayesian approaches outperform other univariate models where they provide smaller values for RMSE and MAPE.

***doi**: 10.5829/ije.2020.33.07a.16*

1. INTRODUCTION

Time series forecast plays an important role in many fields such as economics, finance, business intelligence, meteorology, and telecommunication [1]. As such, time series forecasting has been an active area of research since 1950s and many empirical and theoretical studies are conducted [2-4]. As an early attempt, researchers tried to use linear combination of historical data and hence, most of the traditional statistical models including moving average, exponential smoothing, and autoregressive integrated moving average (ARIMA) have linear structure [5]. However, in the late 1970s, it became increasingly clear that linear models, per se, are not adapted for many applications like stochastic series [1]. Therefore, nonlinear models like autoregressive conditional heteroscedastic (ARCH) and general autoregressive conditional heteroscedastic (GARCH) were introduced. In last two decades, Machine Learning (ML) models have established themselves as serious rivals of classical models in forecasting literature [6-9]. ML models are examples of potentially nonparametric and nonlinear models which use only historical data to learn the stochastic dependency between the historical date and future [1].

There is a growing interest on financial time series forecasting in recent years because it plays a significant role in investment decisions. Generally, financial time series have noise characteristic due to the unavailability of complete information while their non-stationary characteristic originates in the distributional changes over time. In other words, financial time series forecasting is a relatively challenging task [10].

Recently, Cryptocurrencies (i.e. digital monetary systems stored in an encrypted block-chain) have received significant attention in the financial community [11]. The current supposed leader of Cryptocurrencies, Bitcoin, presents an interesting time series rising in a market that is in its transient stage [12]. In this paper, we propose some forecasting models based on ARIMA and ML methods to forecast the price of Bitcoin. The proposed ML approaches include Kriging, Artificial

Corresponding Author Email: khedmati@sharif.edu* (M. Khedmati)

Please cite this article as: M. Khedmati, F. Seifi, M.J. Azizi, Time Series Forecasting of Bitcoin Price Based on Autoregressive Integrated Moving Average and Machine Learning Approaches, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1293-1303

Neural Networks (ANNs), Bayesian model, Support Vector Machines (SVMs), and Random Forest (RF). Moreover, we use the opening, maximum, and minimum daily price in addition to the closing price to improve the prediction.

The rest of the paper is organized as follows. Section 2 provides a literature review on Bitcoin price forecasting studies. The ARIMA approach is explored in detail in Section 3. In Section 4, the proposed ML models are studied, and their performance are compared in Section 5. Finally, the concluding remarks constitute Section 6.

2. LITERATURE REVIEW

The studies in this literature are divided into two categories; one uses the Bitcoin features to predict its price while the other one has an economic point-of-view. This paper focuses on the latter. This category consists of two subcategories including the *classical methods* of transformations and the *Machine Learning* models on data like opening, maximum, minimum and gold prices.

2. 1. Classical Approach In this category, Chu et al. [13] applied statistical analysis on the exchange rate log-returns of Bitcoin versus the US Dollar. This paper compared 15 popular financial parametric distributions on the log returns and concluded that the generalized hyperbolic distribution provides the best results. The financial capabilities of Bitcoin are studied in some papers including Dyhrberg [14]. They showed several similarities of Bitcoin to gold and dollar, indicating hedging capabilities and advantages of Bitcoin as a medium of exchange. Autoregressive approaches are also studied in which for example, Hencic and Gouriéroux [15] used the mixed causal-noncausal autoregressive process with Cauchy errors to predict the Bitcoin price. In addition, Ho et al. [16] compared ARIMA, recurrent and multilayer feed-forward networks showing that the first two outperform the last model. Some of the studies use transformations where, for example, Delfin-Vidal and Romero-Melendez [17] used a continuous wavelet transform analysis on the price volatility across different time and investment horizons.

In another spectrum of studies, the authors used Bitcoin attributes and economical tools. For example, Kristoufek [18] addressed the price changes focusing on possible sources of the change, ranging from fundamentals to speculative or technical sources. This work examined how interconnections behave in time with different scales (frequencies). In another study, Kristoufek [19] used a similar approach and studied the relationship between digital currencies, such as Bitcoin, and Google Trends or Wikipedia search queries. This study showed not only that they are connected, but also there exists a pronounced asymmetry between the effect of an increased interest in the currency when it is above or below its trending value. Another work that looked into these types of interconnections is Garcia et al. [20] that used the data from social media and search engines. They studied the links between social signals and Bitcoin prices through a social feedback cycle and found two main positive feedback loops indicating a strong connection.

2. 2. Machine Learning Approaches The Bayesian and linear regression variants have been used extensively to predict the Bitcoin price. For instance, Shah and Zhang [21] employed Bayesian regression as the latent source model and devised a simple strategy for trading Bitcoin. Another example is Greaves and Au [22] which applied linear regression, logistic regression, SVM, and ANN on the block-chain network-based features of the price. In a slightly altered approach, Madan et al. [23] proposed two phases; first, they used over 25 characteristics of the price and payment network over 5 years to predict the sign of future changes using Binomial General Linear Model (GLM), SVM, and RF models. Afterward, for the second phase, they merely focused on the Bitcoin price data, alone.

Deep learning approaches have also been employed for Bitcoin price predictions. Almeida et al. [24] focused on the prediction of the price trend for the next day based on the previous days' price and volume using an ANN model. McNally et al. [12] showed that nonlinear deep learning models including Bayesian optimized Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network can outperform ARIMA. A number of research efforts including Sin and Wang [25], Radityo et al. [26], Indera et al. [11], Jang and Lee [27] and Rahimi and Khashei [28] focused on applying specialized deep learning models. Reinforcement learning models also can be helpful due to the feedback interacting behavior of the price forecast problem. For instance, Lee et al. [29] proposed to predict the price movements using Inverse Reinforcement Learning (IRL) and Agent-Based Modelling (ABM). Their model reproduces synthetic yet realistic rational agents in a simulated market.

As we mentioned in some instances, utilizing a social network or search engine data is effective for modelling and forecasting where, this is shown in Matta et al. [30]. They investigated the relation between the spread of the Bitcoin price and volumes of tweets or Web Search media results, particularly those with a positive sentiment. They explored significant cross-correlations, especially on the Google Trends data.

Sentimental analysis on Twitter feeds can reveal fundamental economic variables and technological factors. Georgoula et al. [31] used this fact to study the relationship between Bitcoin prices and the information derived from the tweets.

3. DATA DESCRIPTION

In this paper, we use the daily Bitcoin exchange rate data (the closed price of Bitcoin) from December 18, 2019, to March 1, 2020, from [32]. The closing prices are the common target of prediction in the literature. Also, note that since the data of this period has a one-time growth, based on the suggestion of economists, we do not use that data. A statistical summary of these data is presented in Table 1. In this table, Mean, SD, Min, and Max represent the mean, standard deviation, minimum, and maximum. In addition, the data is plotted in Figure 1, which obviously suggests non-stationarity of the process.

In the multivariate models, in addition to the Bitcoin price, the maximum, minimum and opening daily prices are also used. Here, we use 75 observations (days) for training the models and the last 44 records as test dataset for one-step-ahead prediction.

4. METHODOLOGY

In this paper, the ARIMA model as the classic method and ANN, SVM, RF, Bayesian method, and Kriging as the machine learning models are chosen for forecasting of Bitcoin price. Then, we analyze the models, compare the results of them and select the best model based on the performance measures.

TABLE 1. Statistical summary of the daily Bitcoin exchange rate data, N = 75

Variable	Mean	SD	Min	Max
Close price	8660.36	1022.66	6967.00	10333.00
Open price	8634.43	1049.73	6613.50	10336.00
High price	8807.38	1038.49	7197.60	10482.60
Low price	8478.33	1007.36	6462.20	10229.30

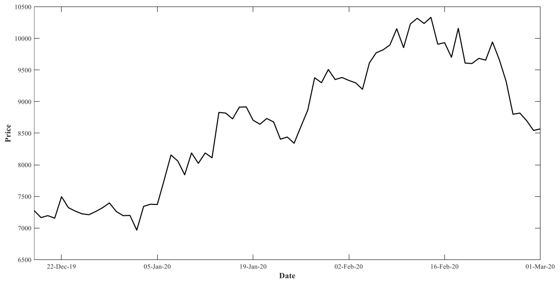


Figure 1. Bitcoin price from December 18, 2019 to March 1, 2020

4. 1. Classic Method

ARIMA models developed by Box and Jenkins [33] have been widely used for time series forecasting. An ARIMA model is usually linear, combined by several previous observations and random errors, and the prediction model is created as a function of the historical data and the errors [34]. The conventional ARIMA (p, d, q) formulation is described as:

$$\Phi(B)(1-B)^d y_t = \delta + \Theta(B)\varepsilon_t \qquad (1)$$

in which δ is a constant term, $\Phi(B)$ is the autoregressive coefficient function, $\Theta(B)$ is the moving average coefficient function, ε_t is the error term at time *t* and *d* is the order of integration terms. If the time series is stationary, then *d* is zero and the model simplifies to ARMA (Autoregressive Moving Average). We first examine the stationarity of the data using Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [35]. In this test, the null and the alternative hypothesis are as below:

Hypothesis 0. y_t is a unit root process.

Hypothesis 1. y_t is trend (or level) stationary or $\sigma_u^2 = 0$

where,

- $y_t = \beta_t + (r_t + \kappa) + e_t$ and y_t are observations during the time
- $r_t = r_{t-1} + u_t$ is a random walk process with $r_0 = \kappa$
- *t* represents time
- $u_t \sim NID(0, \sigma_u^2)$

The value of the KPSS test statistic for Bitcoin data (resp. P-value) is 1.7376 (resp. 0.01), and therefore the null hypothesis cannot be rejected with a 95% confidence level. Now, the sources of non-stationarity should be identified. This can be due to the existence of a trend in the data or changes in their variance. To test the former hypothesis (trend) on the data, we use the Mann-Kendall test [36-38] because of its relevance to the matter. In this test, the null and the alternative hypotheses are as below:

Hypothesis 0. No monotonic trend.

Hypothesis 1. The monotonic trend is present.

The P-value of this test for Bitcoin data is 1.697e-9. Hence, we reject the null hypothesis and use differencing on the data to remove the trend. After differencing, the Mann-Kendall test is used again for the integrated data where, the related P-value is changed to 0.4553. This P-value represents the lack of trend in the transformed data. Figure 2 shows the time series data after differencing.

Now, we must check the variance stability for transformed data. To do this, the Breusch-Pagan [39] test is used based on the following hypotheses.

Hypothesis 0. Variance is homoscedastic.

Hypothesis 1. Variance is not homoscedastic.

The P-value of this test for Bitcoin data is 2.2e-16, which does not lead to the rejection of the null hypothesis and hence, we have a stable variance. In addition, the P-value of KPSS test for the transformed data is 0.1. Therefore, we cannot reject the null hypothesis of KPSS test and accordingly, the transformed data is stationary.

Once the data is stationary, it is time to select the model in which the Extended Sample Auto Correlation

Function (ESACF) is used for this purpose. Based on ESACF matrix, an ARMA (1,1) is appropriate for the transformed data. Note that since differencing is used to convert the data into a stationary time series, the resulting model is ARIMA (1,1,1) and its parameters for forecasting the price on March 1, 2020, are:

$$(1-0.6034B)(1-B)y_t=-31.27+(1-0.4086B)\varepsilon_t \qquad (2)$$

Figure 3 shows the residual plots of this model and Figure 4 shows the ACF (Autocorrelation Function) while PACF (Partial Autocorrelation Function) plots. Since there is no specific trend in these diagrams, it can be assumed that the model is properly selected and can be used for prediction.

There are standard measures for evaluating the performance of forecasting models where, Woschnagg and Cipan [40] address some of these methods. In this paper, we use RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) to evaluate the accuracy of the proposed models. This is because RMSE is a beneficial measure for comparing the accuracy of the models and MAPE is relatively easy to interpret. Specifically, the formulas are:

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n}\left(\hat{y}_t-y_t\right)^2} \\ MAPE &= \frac{1}{n}\sum\nolimits_{t=1}^{n}\left|\left(\hat{y}_t-y_t\right)\right| \end{aligned} \qquad (3)$$

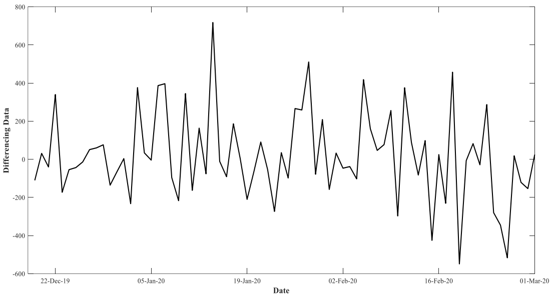


Figure 2. Bitcoin price with first order differencing

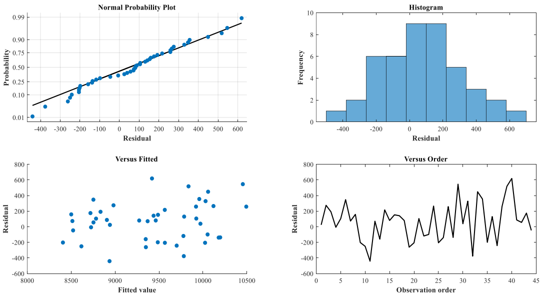


Figure 3. Residual plots

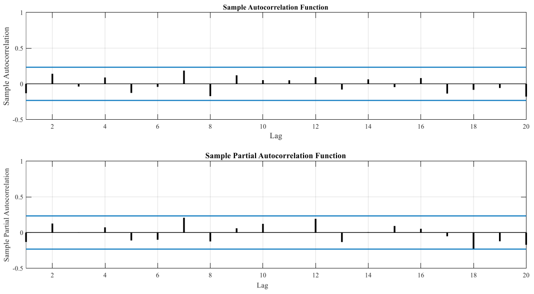


Figure 4. The ACF (top) and PACF (down) plots of residuals

Following an online schema, we update our model parameters after removing the oldest point and adding the new one. The RMSE and MAPE for the proposed ARIMA model are 253.4513 and 2.2286%, respectively.

4. 2. Machine Learning Methods

For the proposed multivariate ML models, the variables of opening, maximum, and minimum daily prices are used in addition to the closing prices, while the univariate model just uses the closing prices. The univariate models include Kriging, ANN, and Bayesian methods and multivariate models include ANN, SVM, RF, and Bayesian methods. This section ends with the comparison of the models with each other.

4. 2. 1. Kriging

Kriging is the interpolation of unknown values in a stochastic function with a linear weighted set of observed values [41]. Krige, an African mining engineer, invented this method to define the exact location of mining rocks in the 1950s [42]. The main idea of this meta-model is to use a weighted mean of outputs in such a way that the weights depend on the interspace between forecasting point and observed points. The optimal weights give minimum prediction error variance and the predictions are the Best Likelihood Unbiased Estimators (BLUE). Due to these properties, Kriging is an optimal interpolator [43], i.e. Kriging meta-models traverse through all the members of the experimental environment. This model is mainly used for prediction purposes in addition to sensitivity analysis and robust optimization. Generally, Kriging is classified into six categories, namely Simple, Ordinary, Co-Kriging, Universal, Blind and Stochastic [43]. To the best of our knowledge, there are just a few studies that use Kriging for time series forecasting. For example, Cellura et al. [44] applied a neural Kriging method to the spatial estimation of wind speed for energy planning in Sicily and, Liu et al. [45] used Kriging for prediction of wind speed.

In this paper, we use a univariate ordinary Kriging (OK) model with a simple parameter tuning, which indicated that using the last 5 days' data provides the best performance. Ordinary Kriging interpolates the one-step-ahead-forecast (y_n+1) using a set of n existing records (y_i; i =1, …, n). We suppose that the mean output is the unknown variable and the prediction is expressed as follows:

$$y_{n+1} = \lambda + \gamma \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{y}^T - \mathbf{1}\lambda\right) \tag{4}$$

where, $\lambda = \left(\mathbf{1}^T \boldsymbol{\Gamma}^{-1}\mathbf{1}\right)^{-1} \mathbf{1}^T \boldsymbol{\Gamma}^{-1}\boldsymbol{y}^T$, **1** is a $n \times 1$ vector of ones, $\boldsymbol{\Gamma} = cov\left(y_i, y_{i'}\right); i = 1,\ldots,n,\ i' = 1,\ldots,n$ a $n \times n$ matrix of covariance between the data points, $\boldsymbol{\gamma} = cov\left(y_i, y_{n+1}\right)$ a $n \times 1$ vector including covariance between the data y_i and the prediction y_{n+1}, and finally, ***y***, a $n \times 1$ vector of the Bitcoin prices. The prediction variance is calculated as:

$$S^2_{y_{n+1}} = \sigma^2\left(1 - \boldsymbol{\gamma}\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}^T + \frac{1 - \mathbf{1}^T \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}^T}{\mathbf{1}^T \boldsymbol{\Gamma}^{-1}\mathbf{1}}\right) \tag{5}$$

where, $\sigma^2 = \frac{1}{n}\left(\boldsymbol{y}^T - \mathbf{1}\lambda\right)^T \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{y}^T - \mathbf{1}\lambda\right)$ [46]. We use ooDACE toolbox of MATLAB as in [47]. The proposed model cares less about the past data compared to the new data, and only the data from the past 5 days is used to fit the model. In this regard, the Kriging model is very similar to a moving average model but the covariance matrix helps to obtain the parameters. The RMSE and MAPE of this model are 298.1863 and 2.6768%, respectively.

4. 2. 2. Bayesian Method The basic theory of prediction using Gaussian processes goes back to Wiener [48] and Kolmogorov [49] in the 1940s. Indeed, Lauritzen [50] discusses the relevant work by Danish astronomer T. N. Thiele from 1880 [51]. The Bayes rule is one of those simple but profound ideas that underlie statistical thinking [52]. However, finding an appropriate prior distribution for the data is a difficult task, where it makes the Bayesian analysis more complicated than other models to utilize.

As a result of the central limit theorem, the Gaussian processes are flexible enough to have a good performance as a prior distribution on many data sets with a large number of data [53]. The goal of the Bayesian forecasting is to compute the distribution P(y_{n+1}|*D*, n+1) of output y_{n+1} given a test input *n+1* and a set of *n* training records D ={(*i*, y_i)| *i* =1, …, n}. We use the Bayes rule to obtain the posterior distribution for the (*n+1*)th Gaussian process outputs. The prediction conditioned on the observed outputs has Gaussian distribution [54]; that is:

$$P(y_{n+1} \mid D, n+1) \sim N\left(\mu_{y_{n+1}}, \sigma^2_{y_{n+1}}\right) \tag{6}$$

where, the mean and variance are given by $\mu_{y_{n+1}} = \boldsymbol{\gamma}^T \boldsymbol{\Gamma} \boldsymbol{y}^T$ and $\sigma^2_{y_{n+1}} = cov\left(y_{n+1}, y_{n+1}\right) - \boldsymbol{\gamma}^T \boldsymbol{\Gamma} \boldsymbol{\gamma}$, respectively. Since the variance is predicted for the (*n+1*)th Gaussian process output, the confidence interval of the prediction, in addition to the point prediction, can be obtained [53]. We use both univariate and multivariate Bayes models and the results in Table 2 show that the multivariate model has better accuracy than the univariate one.

4. 2. 3. Artificial Neural Network Artificial neural networks (ANN) can recognize future patterns of the time series. ANNs are universal and very flexible function approximators, first used in the fields of cognitive science and engineering. One of the similar works on ANN in the financial studies is Kaastra and Boyd [55], which provides an eight-step procedure to design an ANN forecasting model. They also discuss the trade-offs in parameter selection, some common pitfalls, and points of disagreement among practitioners. In addition, Azoff [56] takes the reader of their book beyond the 'black-box' approach to neural networks and provides the knowledge that is required for their proper design and use in financial markets forecasting with an emphasis on futures trading. Zhang [57] proposed a hybrid methodology that combines both ARIMA and ANN models to benefit the strength of ARIMA and ANN models in the linear and nonlinear modeling.

The feed-forward neural networks are used widely in the literature [58-59], where, we propose a feed-forward neural network with a single hidden layer and lagged inputs to forecast univariate time series. The historical data is the input of the neural network model, while the output is the forecast value. The hidden layer stores an appropriate transfer function which is used for processing the data from the input nodes. The model is expressed as:

$$y_{n+1} = w_0 + \sum\nolimits_{j=1}^{Q} w_j g(w_{0j} + \sum\nolimits_{i=0}^{P} w_{i,j} y_{n-i}) \tag{7}$$

where, *P* is the number of input nodes, *Q* the number of hidden nodes, *g* an activation function, {w_j, *j* = 0, 1, …, *Q*} a vector of weights from the hidden layer to output nodes, {$w_{i,j}$, *i* =1, 2, …, *P*, *j* = 0, 1, …, *Q*} are the weights between the input to hidden nodes and $w_{0,j}$ are the weights for each output between input and hidden layer [34]. To select the number of previous observations, features that should be in the model (feature selection) and the number of the nodes in the hidden layer, we use parameter tuning

TABLE 2. RMSE and MAPE for Bayesian model

Model	RMSE	MAPE (%)
Univariate	245.9793	2.0551
Multivariate	192.8742	1.5251

TABLE 3. RMSE and MAPE for ANN model

Model	RMSE	MAPE (%)
Univariate	321.2871	2.8218
Multivariate	252.0465	2.0865

on the training data set with R language. The results of univariate and multivariate ANN models are represented in Table 3.

4. 2. 4. Support Vector Machine The motivation for using the support vector machines (SVMs) in time series forecasting is the ability of this methodology to accurately forecast time series data when the underlying processes are typically nonlinear, non-stationary and not defined a-priori. This model is also shown to outperform other non-linear techniques including neural-network-based non-linear prediction techniques such as multi-layer perceptrons [60].

The general idea of SVM for regression (or SVR) is to generate the regression function by applying a set of high dimensional linear functions. Then, it uses a minimization on the upper bound of the generalization error [34]. The inputs are mapped into a high dimensional nonlinearly feature space (F), wherein the features are correlated linearly with the outputs. The SVR formulation considers the following linear estimation function [10]:

$$f(\boldsymbol{t}) = \boldsymbol{w}^T \phi(\boldsymbol{t}) + \boldsymbol{b} \tag{8}$$

where $\boldsymbol{w}$ is the weight vector, $\boldsymbol{b}$ the bias term vector, $\phi(\boldsymbol{t})$ denotes a mapping function in the feature space and $\boldsymbol{w}^T\phi(\boldsymbol{t})$ the dot production in the feature space F. Various cost functions such as the Laplacian, Huber's Gaussian and Vapnik's linear ε-Insensitivity can be used in the SVR formulation. Among these, the Vapnik's linear ε-Insensitivity loss function is the most commonly adopted [38], which is given in Equation (9).

$$\left|y_i - f(t_i)\right|_\varepsilon = \begin{cases} 0 & if \left|y_i - f(t_i)\right| \le \varepsilon \\ \left|y_i - f(t_i)\right| - \varepsilon & otherwise \end{cases} \tag{9}$$

where, ε is a precision parameter representing the radius of the tube located around the regression function *f*(*t*). Accordingly, linear regression *f*(*t*) is estimated by simultaneously minimizing $||\boldsymbol{w}||^2$ and the sum of the linear ε-Insensitivity losses as bellow:

$$R = \frac{1}{2}\left\|\boldsymbol{w}\right\|^2 + c\left(\sum_{i=1}^{n}\left|y_i - f(t_i)\right|_\varepsilon\right) \tag{10}$$

in which the constant *c* controls the weight of approximation error and size of weights vector, $||\boldsymbol{w}||$. Increasing c, potentially decreases the approximation error with a trade-off that controls the overfitting. Minimizing the risk R is equivalent to the model given in Equations (11a-11d) [5].

$$\text{minimize } R = \frac{1}{2}\left\|\boldsymbol{w}\right\|^2 + c\left(\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right)\right) \tag{11-a}$$

subject to

$$\left(\boldsymbol{w}^T t_i + b\right) - y_i \le \varepsilon + \xi_i \quad \forall i = 1,\ldots,n \tag{11-b}$$

$$y_i - \left(\boldsymbol{w}^T t_i + b\right) \le \varepsilon + \xi_i^* \quad \forall i = 1,\ldots,n \tag{11-c}$$

$$\xi_i, \xi_i^* \ge 0 \qquad \forall i = 1,\ldots,n \tag{11-d}$$

By using the Lagrangian multipliers and Karush-Kuhn-Tucker conditions, the following dual Lagrangian model is obtained as [61]:

$$\text{maximize } L\left(\alpha, \alpha^*\right) \tag{12-a}$$

$$\text{subject to } \sum_{i=1}^{n}\left(\alpha_i^* - \alpha_i\right) = 0 \tag{12-b}$$

$$0 \le \alpha_i \le C \quad \forall i = 1,\ldots,n \tag{12-c}$$

$$0 \le \alpha_i^* \le C \quad \forall i = 1,\ldots,n \tag{12-d}$$

with the following definitions:

$$\begin{aligned} L\left(\alpha, \alpha^*\right) &= -\varepsilon\sum_{i=1}^{n}\left(\alpha_i^* + \alpha_i\right) + \sum_{i=1}^{n}\left(\alpha_i^* - \alpha_i\right)y_i \\ &- \frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i^* - \alpha_i\right)\left(\alpha_j^* - \alpha_j\right)K\left(t_i, t_j\right) \end{aligned} \tag{13}$$

Note that the Lagrangian multipliers in Equation (13) satisfy the equality $\alpha_i\alpha_i^* = 0$ and the optimal weights for the regression is $\left(\boldsymbol{w}^T\right)^* = \sum_{i=1}^{n}\left(\alpha_i^* - \alpha_i\right)K\left(\boldsymbol{t}, t_i\right)$. Hence, the general form of the SVR-based regression function can be written, according to Vapnik [61] as follows:

$$f(\boldsymbol{t}) = \sum_{i=1}^{n}\left(\alpha_i^* - \alpha_i\right)K\left(\boldsymbol{t}, t_i\right) + \boldsymbol{b} \tag{14}$$

where, $K\left(\boldsymbol{t}, t_i\right)$ is the kernel function which is proportional to the inner product of two points, t_i and t_j, in the feature space $\phi(t_i)$ and $\phi(t_j)$; that is, $K\left(t_i, t_j\right) = \phi(t_i)\phi(t_j)$. Although several choices for the kernel function are available, the most widely used is the Radial Basis Function (RBF) $K(t_i, t_j) = \exp(-||t_i - t_j||^2/2v^2)$, where, *v* denotes the width of the RBF [10]. We use the "SVM" package and "tune" function of R for parameter tuning. The results of univariate and multivariate SVM models are represented in Table 4.

4. 2. 5. Random Forest Random Forest (RF) is an ensemble learning method for classification and regression that proceeds by constructing an aggregation of decision trees [62]. Random decision forests adjust for the decision trees' habit of overfitting to their training set. In the context of time series, the changes in the future data are dynamic and a regression might not be an excellent choice however, we can tune it at the current time of a sliding window. An example of a random forest application in time series is the work in Kane et al. [54]. They showed that using random forest enhances the predictive ability over existing time series models for the prediction of infectious disease outbreaks in bird populations. Using RF for one-step-ahead time series forecasting is straightforward and similar to the application of RF in the regression models. Let *f* be the model function which will be used for y_{n+1}, given y_1, …, y_n. If we use *k* lagged variables, the predicted y_{n+1} is obtained based on the following equation for $t = n+1$ [63]:

$$y_t = f\left(y_{t-1},\ldots,y_{t-k}\right), t = k+1,\ldots,n+1 \quad (15)$$

We use a training set of size *n*-*k*. In each training sample, the dependent variable is y_t, for $t = k+1, \ldots, n+1$, while the predictor variables are $y_{t-1}, \ldots, y_{t-k}$. When *k* increases, the size of the training set *n*-*k* decreases. The training set, which includes *n*-*k* samples, is created using the "CasesSeries" function of the "rminer" R package [64-65]. The RMSE and MAPE of this method are obtained as 237.0867 and 2.0787%, respectively.

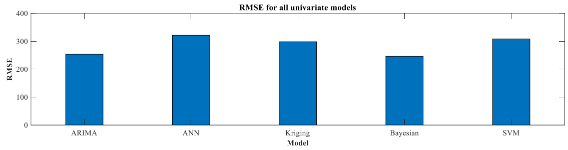


Figure 5. RMSE for all univariate models

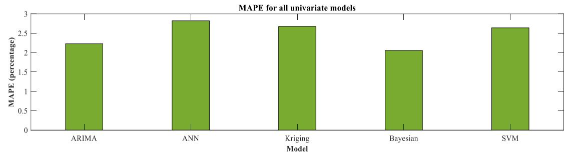


Figure 6. MAPE for all univariate models

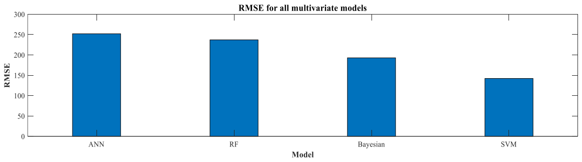


Figure 7. RMSE for all multivariate models

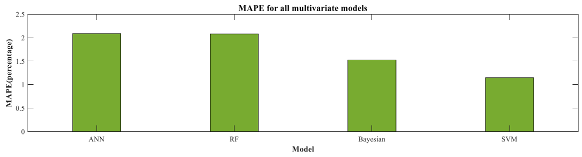


Figure 8. MAPE for all multivariate models

TABLE 4. RMSE and MAPE for SVM model

Model	RMSE	MAPE (%)
Univariate	308.2759	2.6397
Multivariate	142.1580	1.1469

5. PERFORMANCE COMPARISON

In this section, the models are compared based on the results of the previous section. The results of univariate models are shown in Figures 5 and 6. In this regard, considering the univariate models, Bayesian has shown the best performance with the smallest values of RMSE and MAPE among other models. In addition, the results of the multivariate models are shown in Figures 7 and 8. The results show that among the multivariate models, SVM has fantastic performance and its accuracy is significantly better than the other multivariate models. Furthermore, we plot the cumulative RMSE for all the univariate and multivariate models in Figures 9 and 10, respectively. Based on the results, in the univariate case, the Bayesian and ARIMA models and in the multivariate case, the SVM model outperform other models consistently through time.

A more systematic way to compare two time series models is to look at their errors in pairs where the Diebold-Mariano test is used for this purpose. Let y_t as the actual value for time point *t* and $\hat{y}_{it}$ be the prediction of the model *i* for time *t*. Then, the error for model *i* in time *t* is defined as $e_{it} = \hat{y}_{it} - y_t$. Now, the loss differential between the forecast of two models is defined as $d_t = g(e_{1t}) - g(e_{2t})$ where, $g(e_{it})$ shows the loss function. In this regard, we have [66]:

$$\sqrt{T}(\bar{d} - \mu) \rightarrow N(0, 2\pi f_d(0)) \quad (16)$$

where, $\bar{d} = \sum_{t=1}^{T} d_t \big/ T$ and $f_d(0) = \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau) \big/ 2\pi$ is the spectral density of the loss differential at frequency zero. In addition,

$$\gamma_d(\tau) = E[(d_t - \mu)(d_{t-\tau} - \mu)] \quad (17)$$

is the autocovariance of the loss differential at lag τ. Under the null hypothesis that the two forecasts have the same accuracy, the Diebold-Mariano test statistic is defined as:

$$DM = \bar{d} \Big/ \sqrt{\left(2\pi \hat{f}_d(0) \big/ T\right)} \quad (18)$$

in which, $\hat{f}_d(0) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{(T-1)} I(\frac{k}{h-1}) \hat{\gamma}_d(k)$ is the estimate of $f_d(0)$ with:

$$\hat{\gamma}_d(k) = \frac{1}{T}\sum\nolimits_{t=|k|+1}^{T}(d_t - \bar{d})(d_{t-|k|} - \bar{d}) \tag{19}$$

and $I(x) = 1$ if $|x| \leq 1$ and 0, otherwise. In practice, if we set $M = T^{1/3}$, an appropriate estimate of $2\pi f_d(0)$ is obtained by $\sum\nolimits_{k=-M}^{M}\hat{\gamma}_d(k)$, and hence we have:

$$DM = \bar{d}\Big/\sqrt{\left(\sum\nolimits_{k=-M}^{M}\hat{\gamma}_d(k)\Big/T\right)} \tag{20}$$

Under the null hypothesis, the test statistics DM is asymptotically $N(0,1)$ distributed and the null hypothesis is rejected if $|DM| > Z_{\alpha/2}$.

We use this test for the univariate and multivariate models in which the pairwise P-value of Diebold-Mariano statistic for univariate and multivariate time series are represented in Tables 5 and 6, respectively. In this paper, we applied one sided null hypothesis; that is, the value at row *i* and column *j* is the P-value testing if model *i* is better than model *j*. For example, in Table 5, 0.0283 is the P-value of the hypothesis which indicates that the performance of the ARIMA model is better than the Kriging model.

We use a 95% confidence level and follow a round robin policy. The results show that the ARIMA and Bayesian models have the same accuracy and outperform other univariate models. It should be noted that despite the better performance of the ARIMA model, it is harder to fit because of its complex data pre-processing including the differencing and various data transformations. However, machine learning models are faster than the ARIMA model with relatively high accuracy. For example, the accuracy of the Bayesian model is the same as the accuracy of ARIMA, while it provides the results much faster than the latter. In other words, although machine learning models need hyper-parameter tuning, this is easier than the pre-processing step of the ARIMA model. In addition, the effect of hyper-parameter tuning on the performance of ML models is less than the effect of pre-processing on the performance of the ARIMA model.

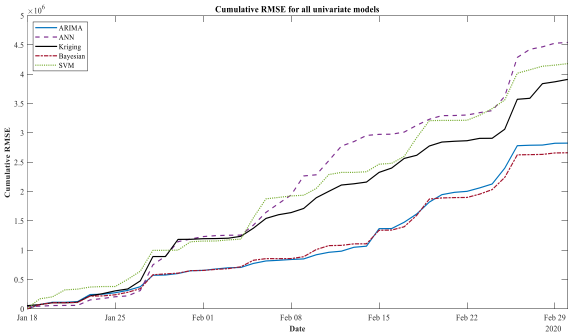


Figure 9. Cumulative RMSE for univariate models

Based on the results in Table 6, the SVM model has the best accuracy between multivariate models which confirms the comparison based on RMSE and MAPE criteria. Hence, we identify the SVM model as the best multivariate model. The outputs of the best univariate and multivariate models are shown in Figures 11 and 12.

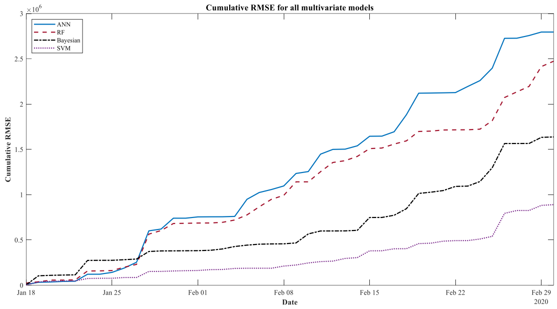


Figure 10. Cumulative RMSE for multivariate models

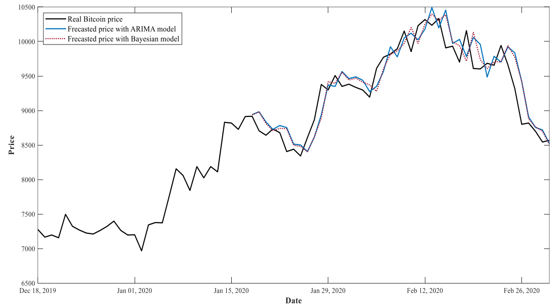


Figure 11. Real Bitcoin price with forecasted them form best univariate models

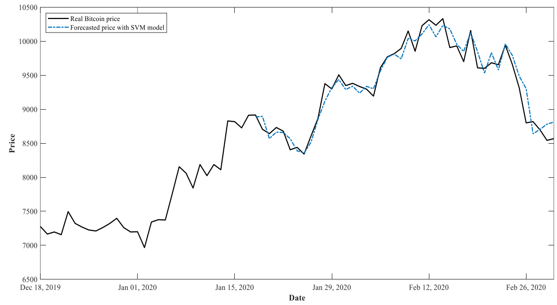


Figure 12. Real Bitcoin price with forecasted them form SVM model

TABLE 5. P-value of DM test for all univariate models

	ARIMA	Kriging	Bayesian	SVM	ANN
ARIMA	-	0.0283	0.7793	0.0158	0.0137
Kriging	0.9716	-	0.9881	0.3418	0.1403
Bayesian	0.2207	0.0119	-	0.0020	0.0064
SVM	0.9841	0.6582	0.9980	-	0.3195
ANN	0.9862	0.8597	0.9935	0.6805	-

TABLE 6. P-value of DM test for all multivariate models

	Bayesian	RF	SVM	ANN
Bayesian	-	0.0294	0.998	0.0062
RF	0.9706	-	1	0.1921
SVM	0.0020	4e-05	-	8e-05
ANN	0.9841	0.6582	0.998	-

6. CONCLUDING REMARKS

In this paper, the time series forecasting approaches of ARIMA as a classical model and five machine learning models including Kriging, artificial neural network (ANN), Bayesian method, support vector machine (SVM) and random forest (RF) are proposed for modeling and forecasting of Bitcoin price. The proposed models included the univariate and multivariate models in which the ARIMA, ANN, Kriging and Bayesian models are used as univariate while ANN, SVM, RF, and Bayesian models are proposed for multivariate case. Then, these models are applied on the Bitcoin price from December 18, 2019, to March 1, 2020, where, in addition to the Bitcoin price, the maximum, minimum and opening price of Bitcoin for the same period is also used in the multivariate models. Comparing the performance of the proposed models in terms of the RMSE and MAPE measurements, it is concluded that ARIMA and Bayesian provide better results compared to other univariate models since they have smaller RMSE and MAPE values compared to other models. However, the SVM outperforms all the univariate and multivariate models and is selected as the best model where its performance measures of RMSE and MAPE are much smaller than the values of all other models.

As a recommendation for future research, one can consider the effect of variables other than the ones related to the Bitcoin price in the multivariate models for possible improvement in the forecasting. Another avenue for future research can be designing a way for hyper-parameter optimization of the investigated models.

7. REFERENCES

1. Bontempi, G., Ben Taieb, S., and Le Borgne, Y. A., Machine Learning Strategies for Time Series Forecasting. ***European Business Intelligence Summer School***, Vol. 138, (2012), 62-77. doi: 10.1007/978-3-642-36318-4_3
2. Stock, J. H. and Watson, M. W., A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. ***National Bureau of Economic Research***, Vol. 14, No. 1, (1996), 11-30. doi: 10.3386/w6607
3. Lemke, C. and Gabrys, B., Meta-learning for time series forecasting and forecast combination. ***Neurocomputing***, *73*(10-12), (2010), 2006-2016. doi: 10.1016/j.neucom.2009.09.020
4. Ahmadia, E., Abooiea, M. H., Jasemib, M., and Zare Mehrjardi, Y., A Nonlinear Autoregressive Model with Exogenous Variables Neural Network for Stock Market Timing: The Candlestick Technical Analysis. ***International Journal of Engineering. Transactions C: Aspects***, Vol. 29, No. 12, (2016), 1717-1725. doi: 10.5829/idosi.ije.2016.29.12c.10
5. Patel, J., Shah, S., Thakkar, P., and Kotecha, K., Predicting stock market index using fusion of machine learning techniques. ***Expert Systems with Applications***, Vol. 42, No. 4, (2015), 2162-2172. doi: 10.1016/j.eswa.2014.10.031
6. Ahmed, N. K., Atiya, A. F., Gayar, N. E., and El-Shishiny, H., An Empirical Comparison of Machine Learning Models for Time Series Forecasting. ***Econometric Reviews***, Vol. 29, No. 5-6, (2010), 594-621. doi: 10.1080/07474938.2010.481556
7. Palit, A. K. and Popovic, D., *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications.* Springer, (2005).
8. Neshat, N., An Approach of Artificial Neural Networks Modeling Based on Fuzzy Regression for Forecasting Purposes. ***International Journal of Engineering. Transactions B: Applications***, Vol. 28, No. 11, (2015), 1651-1655. doi: 10.5829/idosi.ije.2015.28.11b.13
9. Werbos, P. J., Generalization of backpropagation with application to a recurrent gas market model. ***Neural Networks***, Vol. 1, No. 4, (1988), 339-356. doi: 10.1016/0893-6080(88)90007-X
10. Lu, C., Lee, T., and Chiu, C., Financial time series forecasting using independent component analysis and support vector regression. ***Decision Support Systems***, Vol. 47, No. 2, (2009), 115-125. doi: 10.1016/j.dss.2009.02.001
11. Indera, N. I., Yassin, I. M., Zabidi, A., and Rizman, Z. I., Non-linear autoregressive with exogeneous input (narx) Bitcoin price prediction model using pso-optimized parameters and moving average technical indicators. ***Journal of Fundamental and Applied Sciences***, Vol. 9, No. 3, (2017), 791-808. doi: 10.4314/jfas.v9i3s.61
12. McNally, S., Roche, J., and Caton, S., Predicting the Price of Bitcoin Using Machine Learning. 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), (2018). doi: 10.1109/PDP2018.2018.00060
13. Chu, J., Nadarajah, S., and Chan, S., Statistical Analysis of the Exchange Rate of Bitcoin. ***PLoS ONE***, Vol. 10, No. 7, (2015), doi: 10.3386/t0090
14. Dyhrberg, A. H., Bitcoin, gold and the dollar – A GARCH volatility analysis. ***Finance Research Letters***, Vol. 16, (2016), 85-92. doi: 10.1016/j.frl.2015.10.008
15. Hencic, A. and Gouriéroux, C. .Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rates. ***Econometrics of Risk***, Vol. 583, (2014), 17-40. doi: 10.1007/978-3-319-13449-9_2
16. Ho, S. L., Xie, M., Goh, T. N., A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. ***Computers & Industrial Engineering***, Vol. 42, (2002), 371-375. doi: 10.1016/S0360-8352(02)00036-0
17. Delfin-Vidal, R. and Romero-Meléndez, G., The Fractal Nature of Bitcoin: Evidence from Wavelet Power Spectra. ***Trends in Mathematical Economics***, (2016), 73-98. doi: 10.1007/978-3-319-32543-9_5
18. Kristoufek, L., What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. ***PLoS ONE***, Vol. 4, No. 10, (2015). doi: 10.1371/journal.pone.0123923
19. Kristoufek, L., Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. ***Scientific Reports***, Vol. *3*, (2013), 1-7, doi: 10.1038/srep03415.

20. Garcia, D., Tessone, C. J., Mavrodiev, P., and Perony, N., The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. ***Journal of the Royal Society Interface***, Vol. 11, No. 99, (2014), 1-8. doi: 10.1098/rsif.2014.0623

21. Shah, D. and Zhang, K., Bayesian regression and Bitcoin. 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), (2014), 409-414. doi: 10.1109/ALLERTON.2014.7028484

22. Greaves, A. and Au, B., Using the Bitcoin transaction graph to predict the price of Bitcoin, (2015). http://snap.stanford.edu/class/cs224w-2015/projects_2015/.

23. Madan, I., Saluja, S., and Zhao, A., Automated Bitcoin trading via machine learning algorithms, (2015). http://cs229.stanford.edu/proj2014/.

24. Almeida, J., Tata, S., Moser, A., and Smit, V. Bitcoin prediciton using ANN. ***Neural Networks***, Vol. 7, (2015), 1-12.

25. Sin, E. and Wang, L., Bitcoin price prediction using ensembles of neural networks. 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), (2017), 666-671. doi: 10.1109/FSKD.2017.8393351

26. Radityo, A., Munajat, Q., and Budi, I., Prediction of Bitcoin exchange rate to American dollar using artificial neural network methods. International Conference on Advanced Computer Science and Information Systems (ICACSIS), (2017), 433-438. doi: 10.1109/ICACSIS.2017.8355070

27. Jang, H. and Lee, J., An Empirical study on modeling and prediction of Bitcoin prices with Bayesian neural networks based on blockchain information. ***IEEE Access***, Vol. 6, (2017), 5427-5437. doi: 10.1109/ACCESS.2017.2779181

28. Rahimi, Z. H., Khashei, M., A least squares-based parallel hybridization of statistical and intelligent models for time series forecasting. ***Computers & Industrial Engineering***, (2018), doi: https://doi.org/10.1016/j.cie.2018.02.023.

29. Lee, K., Ulkuatam, S., Beling, P., and Scherer, W. . Generating synthetic Bitcoin transactions and predicting market price movement via inverse reinforcement learning and agent-based modeling. ***Journal of Artificial Societies and Social Simulation***, Vol. 21, No. 3, (2018), 1-5. doi: 10.18564/jasss.3733

30. Matta, M., Lunesu, I., and Marchesi, M., Bitcoin Spread Prediction Using Social and Web Search Media. UMAP Workshops, (2015).

31. Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D., and Giaglis, G. M., Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices. ***MCIS 2015 Proceedings***, Vol. 20, (2015). doi: 10.2139/ssrn.2607167

32. www.investing.com/crypto/bitcoin/historical-data. (n.d.).

33. Box, P. and Jenkins, G., Time series analysis: forecasting and control. Wiley, (1976).

34. Shi, J., Guo, J., and Zheng, S. . Evaluation of hybrid forecasting approaches for wind speed and power generation time series. ***Renewable and Sustainable Energy Reviews***, Vol. 16, No. 5, (2012), 3471-3480. doi: 10.1016/j.rser.2012.02.044

35. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y., Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? ***Journal of Econometrics***, Vol. 54, No. 1-3, (1992), 159-178.

36. Mann, H. B. , Non-parametric tests against trend. ***Econometrics***, Vol. 13, (1945), 163-171. doi: 10.2307/1907187

37. Gilbert, R. O., Statistical methods for environmental pollution monitoring. John Wiley & Sons, (1987).

38. Kendall, M. G., *Rank correlation methods*, 4th Edition, Charles Griffin, London, (1975).

39. Breusch, T. S. and Pagan, A. R., A simple test for heteroskedasticity and random coefficient variation. ***Econometrica***, Vol. 47, No. 5, (1979), 1287-1294. doi: 10.2307/1911963

40. Woschnagg, E. and Cipan, J., Evaluating forecast accuracy. UK Ökonometrische prognose, department of economics, university of Vienna, (2004). http://homepage.univie.ac.at/ robert.kunst/procip.pdf.

41. Ankenman, B., Nelson, B. L., and Staum, J., Stochastic Kriging for simulation metamodeling. ***Operations research***, Vol. 58, No. 2, (2010), 371-382. doi: 10.1109/WSC.2008.4736089

42. Kleijnen, J. P., Kriging metamodeling in simulation: A review. ***European Journal of Operational Research***, Vol. 192, No. 3, (2009), 707-716. doi: 10.1016/j.ejor.2007.10.013

43. Journel, A. G., Fundamentals of geostatistics in five lessons. *American Geophysical Union, Washington, D.C.* , (1989).

44. Cellura, M., Cirrincione, G., Marvuglia, A., and Miraoui, A., Wind speed spatial estimation for energy planning in Sicily: A neural Kriging application. ***Renewable Energy***, Vol. 33, No. 6, (2008), 1251-1266. doi: 10.1016/j.renene.2007.08.012

45. Liu, H., Shi, J., and Erdem, E., Prediction of wind speed time series using modified Taylor Kriging method. ***Energy***, Vol. 35, No. 12, (2010), 4870-4879. doi: 10.1016/j.energy.2010.09.001

46. Azizi, M. J., Seifi, F., Moghadam, S., A robust simulation optimization algorithm using Kriging and particle swarm optimization: Application to surgery room optimization. ***Communications in Statistics-Simulation and Computation***, (2019), doi: 10.1080/03610918.2019.1593452

47. Couckuyt, I. Dhaene, T. Demeester, P., ooDACE Toolbox: A Flexible Object-Oriented Kriging Implementation. ***Journal of Machine Learning Research***, Vol. 15, (2014), 3183-3186.

48. Kolmogorov, A., Interpolation und Extrapolation von stationären zufälligen Folgen. ***Izv. Akad. Nauk SSSR Ser. Mat.***, Vol. 5, No. 1, (1941), 3-14.

49. Lauritzen, S. L., Time Series Analysis in 1880: A Discussion of Contributions Made by T.N. Thiele. ***International Statistical Institute (ISI)***, Vol. 49, No. 3, (1981), 319-331. doi: 10.2307/1402616

50. Rasmussen, C. E. and Williams, C. K. I., Gaussian Processes *in Machine Learning.* MIT press, (2006).

51. Efron, B., *Large-Scale Inference Empirical Bayes Methods for Estimation, Testing, and Prediction.* Vol. 1, Cambridge University Press, (2012).

52. Brahim-Belhouari, S., and Bermak, A., Gaussian process for nonstationary time series prediction. ***Computational Statistics & Data Analysis***, Vol. 47, No. 4, (2004), 705-712. doi: 10.1016/j.csda.2004.02.006

53. Kane, M. J., Price, N., Scotch, M., and Rabinowitz, P. . Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics, (2014), Vol. 15:276, doi: 10.1186/1471-2105-15-276

54. Kaastra, I. and Boyd, M., Designing a neural network for forecasting financial and economic time series. ***Neurocomputing***, Vol. 10, No. 3, (1996), 215-236.

55. Azoff, E. M., *Neural Network Time Series Forecasting of Financial Markets.* New York: John Wiley & Sons, (1994).

56. Zhang, G. P., Time series forecasting using a hybrid ARIMA and neural network model. ***Neurocomputing***, Vol. 50, (2003), 159-175. doi: 10.1016/S0925-2312(01)00702-0

57. Khashei, M. and Bijari, M., An artificial neural network (p, d, q) model for time series forecasting. ***Expert Systems with Applications***, Vol. 37, No. 1, (2010), 479-489. doi: 10.1016/j.eswa.2009.05.044

58. Zhang, G. P. and Qi, M., Neural network forecasting for seasonal and trend time series. ***European Journal of Operational Research***, Vol. 160, No. 2, (2005), 501-514. doi: 10.1016/j.ejor.2003.08.037

59. Sapankevych, N. I., Sankar, R., Time Series Prediction Using Support Vector Machines: A Survey. ***IEEE Computational Intelligence Magazine***, Vol. 4, No. 2, (2009), 24-38. doi: 10.1109/MCI.2009.932254

60. Vapnik, V. N., The Nature of Statistical Learning Theory. Springer, (2000).

61. Ho, T. K., Random decision forests. *In Proceedings of 3rd International Conference on Document Analysis and Recognition*, Vol. 1, (1995), 278-282. doi: 10.1109/ICDAR.1995.598994

62. Tyralis, H. and Papacharalampous, G. . Variable Selection in Time Series Forecasting Using Random Forests. ***Algorithms***, Vol. 10, No. 4, (2017), 114. doi: 10.3390/a10040114

63. Liaw, A. and Wiener, M., Classification and regression by random forest. ***R News***, Vol. 2, No. 3, (2002), 18-22.

64. Cortez, P., Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool. ***Advances in Data Mining. Applications and Theoretical Aspects***, (2010), 572-583. doi: 10.1007/978-3-642-14400-4_44

65. Diebold, F. X., Comparing Predictive Accuracy. ***Journal of Business & Economic Statistics***, Vol. 13, No. 3, (1995), 253-263.

Persian Abstract

چکیده

بیت‌کوین به عنوان رمزارز پیشرو، دسته‌ی جدیدی از دارائی‌هاست که در فضای مالی و سرمایه‌گذاری توجه زیادی را به خود جلب کرده است و در نتیجه یک مسئله جذاب پیش‌بینی سری زمانی را ارائه می‌کند. در این مقاله، برخی از روش‌های پیش‌بینی کلاسیک نظیر ARIMA در کنار روش‌های یادگیری ماشین شامل کریگینگ، شبکه‌ی عصبی مصنوعی، روش بیزین، ماشین بردار پشتیبان و جنگل تصادفی جهت مدل‌سازی و پیش‌بینی ارزش بیت‌کوین به کار برده شده‌اند. در این مقاله، برخی از مدل‌ها به صورت تک متغیره و برخی دیگر به صورت چندمتغیره با بهره‌گیری از بیشترین و کمترین ارزش روز و قیمت بازگشایی بازار بیت‌کوین برازش شده‌اند. مدل‌های ارائه شده، بر روی ارزش بیت‌کوین در بازه‌ی زمانی ۱۸ دسامبر ۲۰۱۹ تا ۱ مارس ۲۰۲۰ اعمال شده و عمل‌کرد آنها بر اساس معیارهای ریشه‌ی میانگین مربعات خطا و میانگین قدرمطلق درصد خطا در کنار آزمون آماری Diebold-Mariano با یک‌دیگر مقایسه شده است. برمبنای معیارهای ریشه‌ی میانگین مربعات خطا و میانگین قدرمطلق درصد خطا، نتایج نشان می‌دهد که ماشین بردار پشتیبان دارای بهترین عملکرد در بین تمامی مدل‌هاست. همچنین، مدل‌های ARIMA و بیزین با در اختیار داشتن مقادیر پایین ریشه‌ی میانگین مربعات خطا و میانگین قدرمطلق درصد خطا، بهترین عمل‌کرد را در بین مدل‌های تک‌متغیره دارا هستند.

International Journal of Engineering

Journal Homepage: www.ije.ir

A Bi-level Meta-heuristic Approach for a Hazardous Waste Management Problem

Z. Saeidi-Mobarakeh[a], R. Tavakkoli-Moghaddam*[b], M. Navabakhsh[a], H. Amoozad-Khalili[c]

[a] *School of Industrial Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran*
[b] *School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran*
[c] *Department of Industrial Engineering, Nowshahr Branch, Islamic Azad University, Nowshahr, Iran*

PAPER INFO

Paper history:
Received 02 January 2020
Received in revised form 07 February 2020
Accepted 06 March 2020

Keywords:
Hazardous Waste Management
Location-routing Problem
Bi-level Programming
Meta-heuristic Algorithm

ABSTRACT

This study concentrates on designing a medical waste management system with a hierarchical structure, including a local government and a waste management planner. The upper-level seeks to design and control the waste management facilities by minimizing the environmental risks related to the disposal of medical waste. While, the lower-level model is to determine the waste collection plans by only minimizing its total operational costs. Therefore, this study develops a bi-level mathematical model, in which the benefits of the both stakeholders are taken into account. As this problem poses difficulty in searching for the optimal solution, a bi-level meta-heuristic approach based on the Genetic Algorithm (GA) is employed for solving the problem. Finally, a case study is conducted to show that the proposed model and solution approach are practical and efficient.

***doi**: 10.5829/ije.2020.33.07a.17*

NOMENCLATURE

Index Set:			
G	Waste generation nodes	F	Locations of available facilities
F'	Candidate locations for establishing new facilities	O	Vehicle depots
W	Waste types	K	Vehicles for waste collection
Parameters			
$fc_{f'}$	Cost of establishing a new facility in location $f' \in F'$ (\$)	d_{ij}	Element $i-j$ of distance matrix $(i, j \in (O \cup G \cup F \cup F'))$
tc	The cost of transferring waste (\$ per km per kg)	$vc_{wf'}$	Per unit cost of implementing capacity for processing waste type *w* in new established facility $f' \in F'$ (\$ per kg)
n_i	Number of people living around the location $i \in (F \cup F')$	n'_{ij}	Number of people living along with transportation link $i-j$ $(i, j \in (O \cup G \cup F \cup F'))$
r_w	Comparative risk of waste type *w* before treating	r'_w	Comparative risk of waste type *w* after treating ($r'_w \ll r_w$)
q_{wg}	Amount of waste type *w* generated in generation node *g* (kg)	c_w	Vehicle capacity compatible with waste type *w* (kg)
$e_{wf'}$	1 denotes the compatibility of waste type *w* with a facility at location $f' \in F'$; 0 denotes their incompatibility	e'_{wk}	1 denotes the compatibility of waste type *w* with vehicle *k*; 0 denotes their incompatibility
τ	Toll coefficient of hazardous waste transportation (\$)	Γ	Available budget (\$)
M	A sufficiently large number		
Decision Variables			
$X_{f'}$	1 denotes the establishment of a facility at location $f' \in F'$; 0, otherwise	Y_{ijk}	1 denotes the order of visiting node *j* by vehicle *k* just after node *i*; 0, otherwise $(i, j \in (O \cup G \cup F \cup F'))$
$U_{wf'}$	Capacity of facility $f' \in F'$ for processing waste type *w* (kg)	V_{wf}	Capacity of available facility $f \in F$ reserved for disposing of waste type *w* (kg)
L_{ik}	Load of vehicle *k* after visiting node $i \in (O \cup G)$		

*Corresponding Author Email: *tavakoli@ut.ac.ir* (R. Tavakkoli-Moghaddam)

Please cite this article as: Z. Saeidi-Mobarakeh, R. Tavakkoli-Moghaddam, M. Navabakhsh, H. Amoozad-Khalili, A Bi-Level Meta-heuristic Approach for a Hazardous Waste Management Problem, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1304-1310

1. INTRODUCTION

Finding a suitable framework for a hazardous waste management system is among the most important urban management decisions. Both the level of risk faced by the populations and the cost of the waste collection operations are severely affected by the system configuration. The majority of the formulations found in the literature contain variables for locating facilities and routing waste collection trucks in a single level, in which they are needed to be determined by the government authorities at the same time. But in practice, these decentralized decisions are taken by two different Decision-Makers (DMs) with conflicting interests and different amounts of power. Keeping in mind these differences, the government policymakers must be aware of the behavior of the users of the waste management network, the carriers, and consider their behaviors in strategic and tactical decision-making. This leads to bi-level programming, where each DM independently maximizes its interest, while is affected by the actions from the other DM under a hierarchy [1].

This paper formulates a bi-level programming problem for deciding about the hazardous waste management system. The upper level of the model reflects the network design problem faced by the government authorities for minimizing the total risk. The lower-level problem seeks to optimize the waste collection activities in terms of the total costs. To tackle this problem, a non-linear mathematical formulation is developed and then owing to its complexity, a bi-level meta-heuristic algorithm is employed.

This paper is structured as follows. A brief literature review is presented in Section 2. The details of our problem and its formulation are given in Section 3. A detailed description of the bi-level meta-heuristic algorithm is provided in Section 4. Next, we describe a real-word example of our problem in Section 5. Section 6 gives the computational results. Finally, Section 7 presents concluding remarks and future research directions.

2. LITERATURE REVIEW

Shih and Lin [2] formulated the problem of planning and scheduling the medical waste collection from multiple hospitals. They proposed a two-phased approach for solving the problem in which the first phase partitions a set of hospitals into waste collection vehicles and the second phase determines the visiting period of each vehicle. Shih and Chang [3] extended the previous work by developing a computer program for the dynamic programming of a large-scale waste collection problem. Shih and Lin [4] proposed a multi-objective optimization problem, in which, in addition to minimization of the cost, two objectives on minimization of transportation risk and balancing of workload for collection system workers were considered. They applied a compromise programming approach for the integration of these objectives. Chaerul et al. [5] employed a goal programming approach for solving a planning model in healthcare waste management with multiple objectives under different relative importance. Alagöz and Kocasoy [6] used a commercial vehicle routing package to identify the most feasible routes in terms of efficiency and economy for the collection of healthcare wastes from the temporary storage rooms and transporting them to the final disposal areas. Shi et al. [7] developed an optimization model with the goal of cost minimization for a medical waste reverse logistics network and used a genetic algorithm to solve this problem. Nolz et al. [8] formulated a collector-managed inventory-routing problem for developing a sustainable logistics system to organize the collection of medical waste. Social objectives in terms of minimizing the public health risks and satisfying pharmacists and local authorities were included in the formulation. Hachicha et al. [9] examined the off-site treatment problem of infectious waste from several medical facilities in a planned stream sterilization disposal center. Alshraideh and Qdais [10] considered a route scheduling model with time windows and stochastic demands in a case study of medical waste collection. The authors considered a chance constraint regarding a pre-defined service level and applied a genetic algorithm to solve the problem. The proposed model by Mantzaras and Voudrias [11] aimed to calculate the optimal location and size of the treatment facilities and transfer stations, the number and capacity of all waste collection and transfer vehicles and their optimal path at minimum cost. Wichapa and Khokhajaikiat [12] proposed a two-stage location-routing problem for infectious medical waste including, (1) a multi-objective facility location with the aim of cost minimization and global priority weights maximization and (2) a vehicle routing problem to minimize transportation costs that was solved using a hybrid genetic algorithm.

As can be understood from the literature review, this is the first study that formulated medical waste management system using a bi-level mathematical formulation. The capabilities of bi-level programming can significantly help to formulate the conflicts between decision-makers in this field.

3. MODEL DEVELOPMENT

3. 1. Problem Definition Our study seeks to design a medical waste management system based on the interests of two groups of non-cooperative DMs. The upper-level DM is to establish new facilities to minimize the environmental side effects of unprocessed medical waste on the people living near treatment facilities and separating the collection of medical waste from general waste and meeting demands of the system for treatment processes. The Environmental Organization with environmental concerns is the upper-level DM and its measures must be funded by the Health and Medical Education Minister as responsible for the production of these wastes. The lower-level problem represents the routing decisions optimization in the designed medical waste management system along with optimization of decisions regarding the amount of waste that must be treated and processed in facilities. Municipality authority as a contractor with the ministry performs the decision making in the stated level for waste collection and aims at the cost-efficiency. To illustrate the details of medical waste management processes, the following explanations are presented:

- Collection: It is important to note that various types of infectious medical wastes are produced by several waste generation nodes in different regions of the considered area. To collect these wastes, a set of vehicles are available in a central depot and are planned in several routes of the waste generation units. It is necessary to note that the vehicle fleet is heterogeneous and includes vehicles with differences in their compatibility with diverse characteristics of hazardous wastes. Furthermore, the partial collection is not allowed by the vehicles or in better words, for collecting each type of waste, the generation nodes must be visited only once. During the planned routes, if the capacity of facilities run out or the available capacity becomes lower than the remaining amount in other generation nodes, the vehicles decide to dump their loads at appropriate facilities. After unloading the collected wastes, the vehicles must move back to the depot.
- Treatment: It is assumed that the current structure of the waste management network does not involve any treatment center. Accordingly, a considerable part of generated wastes is dumped while no treatment processes are exerted, and the requirements of waste treatment are not satisfied by the present system. It is expected that installing a new integrated facility, which comprises both modern treatment and disposal technologies at an available place results in a more centralized waste management processes with lower consequences.

3. 2. Model Formulation

This study presents a bi-level mathematical model that can consider the conflict of stakeholders as shown below:

$$\begin{aligned}\text{Min } f_1 = & \sum_{w\in W}\sum_{f\in F} r_w \; n_f \; V_{wf} \\ & + \sum_{w\in W}\sum_{f'\in F'} r'_w \; n_{f'} \; U_{wf'} \\ & + \sum_{w\in W}\sum_{k\in K}\sum_{i\in G}\sum_{j\in(G\cup F\cup F')} r_w \; e'_{wk} \; L_{ik} \; n'_{ij} \; Y_{ijk}\end{aligned} \tag{1}$$

$$\sum_{f'\in F'}\left(fc_{f'} \; X_{f'} + \sum_{w\in W} vc_{wf'} \; U_{wf'}\right) \le \Gamma \tag{2}$$

$$U_{wf'} \le M \; X_{f'} \quad ;\forall w \in W, f' \in F' \tag{3}$$

$$X_{f'} \in \{0,1\} \text{ and } U_{wf'}, V_{wf} \ge 0 \tag{4}$$

where for given $\{U_{wf'}, V_{wf}, X_{f'}\}$, $\{Y_{ijk}\}$ solves

$$\begin{aligned}\text{Min } f_2 = & \, tc \sum_{k\in K}\sum_{i\in(O\cup G\cup F\cup F')}\sum_{j\in(O\cup G\cup F\cup F')} d_{ij} \; Y_{ijk} \\ & + \tau \sum_{k\in K}\sum_{i\in(O\cup G\cup F\cup F')}\sum_{j\in(O\cup G\cup F\cup F')} n'_{ij} \; Y_{ijk}\end{aligned} \tag{5}$$

$$\sum_{i\in O}\sum_{j\in G} Y_{ijk} = 1 \quad ;\forall k \in K \tag{6}$$

$$\sum_{i\in(O\cup G)} Y_{ijk} - \sum_{i'\in(G\cup F\cup F')} Y_{ji'k} = 0 \quad ;\forall j \in G, k \in K \tag{7}$$

$$\sum_{i\in G} Y_{ijk} - \sum_{i'\in O} Y_{ji'k} = 0 \quad ;\forall j \in (F \cup F'), k \in K \tag{8}$$

$$\sum_{k\in K}\sum_{i\in(O\cup G)} e'_{wk} \; Y_{ijk} = 1 \quad ;\forall j \in G, w \in W \tag{9}$$

$$\begin{aligned}L_{ik} - L_{jk} + \sum_{w\in W} e'_{wk} \, c_w \, Y_{ijk} \le \sum_{w\in W} e'_{wk} \left(c_w - q_{jw}\right) \\ ;\forall i, j \in G, k \in K\end{aligned} \tag{10}$$

$$\sum_{w\in W} q_{iw} \; e'_{wk} \le L_{ik} \le \sum_{w\in W} e'_{wk} \, c_w \quad ;\forall i \in G, k \in K \tag{11}$$

$$Y_{ijk}\left(\sum_{w\in W} q_{jw} \; e'_{wk}\right) \le L_{ik} \quad ;\forall i \in O, j \in G, k \in K \tag{12}$$

$$\begin{aligned}L_{jk} \le \sum_{w\in W} e'_{wk}\left(c_w + \sum_{i\in D}\left(q_{jw} - c_w\right) Y_{ijk}\right) \\ ;\forall j \in G, k \in K\end{aligned} \tag{13}$$

$$\sum_{k\in K}\sum_{i\in G} Y_{ijk} \; e'_{wk} \; L_{ik} \le e_{wj} \, U_{wj} \quad ;\forall j \in F', w \in W \tag{14}$$

$$\sum_{k\in K}\sum_{i\in G} Y_{ijk}\ e'_{wk}\ L_{ik}\ \le V_{wj} \qquad ;\forall\, j\in F, w\in W \tag{15}$$

$$Y_{ijk}\ge 0 \qquad ;\forall i, j\in\left(O\cup G\cup F\cup F'\right), k\in K \tag{16}$$

The objective function (1) in the upper-level problem aims at minimizing the risk imposed on the population and includes three terms. The undesirability of the facilities (newly established and available) is focused on the first two terms, while the last term minimizes the risk associated with unprocessed medical wastes. The limitation of the budget allocated to establishing new facilities is controlled in Equation (2). According to Equation (3), waste processing in a newly established facility is subjected to its establishment. The type of upper-level decision variables in terms of non-negativity and integrality is determined by Equation (4). The objective function (5) in the lower-level problem targets the minimization of the cost of waste management activities. The lower-level objective function comprises of two terms defined on the waste collection cost and the transportation toll charges for traffic routes. Equation (6) emphasizes on this assumption that the waste management network includes only a central depot where all collection vehicles depart from there. The continuity of the routes traveled by the waste collection vehicles is ensured by formulating Equations (7) and (8). In this study, split pickups are not permitted and Equation (9) imposes this constraint into our formulation. Equations (10)-(13) incorporate two significant features into our formulation: sub-tour elimination and capacity of vehicles. The input flow to the facilities (available and newly established) must not exceed the capacity of the facilities; Equations (14) and (15) guarantee this issue. Analogous to Equation (4), Equation (16) limits the domain of the lower-level decision variable.

4. SOLUTION METHODOLOGY

This section describes a bi-level meta-heuristic approach, which uses the Genetic Algorithm (GA) as the main and subsidiary algorithm. At the first level of this algorithm, the GA generates solutions with information about the new established facilities and their capacities and then these solutions are moved into the second level to construct the collection routes. Since the fitness value of a solution in the upper-level depends on the decisions at the lower-level, we first describe the lower-level GA structure and then turn into the GA at the upper-level.

4. 1. Second Level GA

The chromosome representation is an important part of implementing the GA because it enables the algorithm to cover all aspects of a problem. Concerning the special form of the constraints, this paper applies the permutation representation for encoding the solutions into the chromosome. As can be understood from Figure 1, each chromosome is illustrated with a matrix with a dimension of $W\times G$ genes, in which W and G are the number of waste types and waste generation nodes, respectively. Evaluation of the fitness value for each chromosome is done based on the decoding procedure. To construct the collection routes associated with each waste type, the genes of the associated row are added to the routes respectively, until the vehicle's capacity will be exceeded. This procedure is continued to cover all waste generation nodes.

In the GA, the evolution process of a population over successive generations is done through genetic operators, including crossover and mutation. In one hand, the crossover operator blends genetic information between parents and makes children that inherit their parents' promising characteristics. The one-point crossover is employed in the second level of the proposed algorithm in determining the collection routes. One the other hand, the mutation operator tries to maintain an adequate diversity of the population and prevents falling into a local optimum trap. A simple mutation operator is used in this study that its basis is a random selection of a pair of genes and replacement of them.

4. 2. GA in the First Level

In the second level, we have a matrix with a dimension of $W\times\left(F'+F\right)$ genes in which W is the number of waste types and F' and F are the potential and available facilities, respectively (see Figure 2 with only potential facilities). Each gene gets a random number between [0, 1]. In establishing new facility, the priority is with the facility that the corresponding column has a greater value. Concerning the budget constraint, a new facility with a certain capacity is established and excess waste is disposed of in the available facilities. In other words, the collection routes of the second level are terminated at a treatment facility with sufficient capacity for unloading the collected waste. Note that the priority of allocating the waste types to the new established facility is determined based on the value of their respective rows. The first level GA applies the same crossover and mutation operators to create an offspring population.

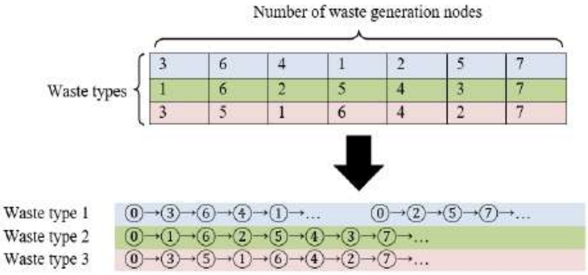

Number of waste generation nodes
Waste types
3 6 4 1 2 5 7
1 6 2 5 4 3 7
3 5 1 6 4 2 7
Waste type 1 0→3→6→4→1→… 0→2→5→7→…
Waste type 2 0→1→6→2→5→4→3→7→…
Waste type 3 0→3→5→1→6→4→2→7→…

Figure 1. Solution representation in the second level GA

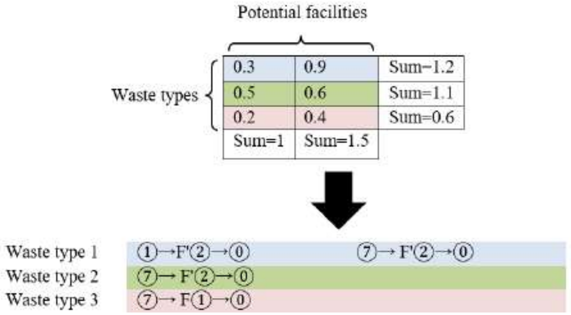


Figure 2. Solution representation in the first level GA

5. CASE STUDY

To demonstrate the application of the proposed model, one of the major cities in Iran, which is dealing with waste management issues in the past decade, is considered as the real-life case study. Isfahan is located in the center of Iran and is one of the highly populated and industrialized cities of the country. Besides the inefficient waste management system, Isfahan is suffering from lack of a proper medical waste management system. According to official guidelines and policies, medical centers should treat their wastes and turn them into general waste before dumping them in landfill sites. This has led to a conflict between the municipality of the city, and the Department of Environment in which both believe that medical centers should be responsible for the waste treatment processes. Therefore, the ministry has allocated a budget for the infrastructure needed to revise the current waste management system. The city requires an exclusive waste management system to cover medical waste all over the city. In the other words, medical waste even after treatment should not be considered as general waste. In this regard, we have included all of the medical centers and hospitals all over the city of Isfahan in the study. The average waste generation rates for medical centers and hospitals are also extracted from the literature and acknowledged using interviews with Department of Environment experts [13].

To design a medical waste management system, all candidate locations for facilities all over the city are determined, and also critical areas for constructions are considered. It is noteworthy that other aspects such as geomorphological criteria are taken into consideration in determining the candidate locations. In Figure 3, the identified locations are shown.

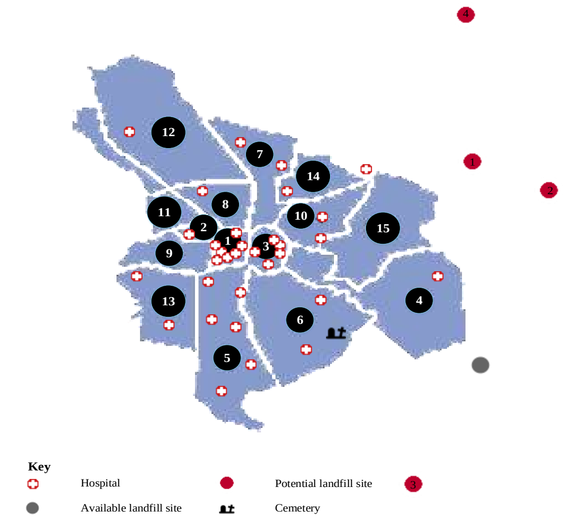


Figure 3. Geographical distribution of hospitals and facilities in Isfahan

6. COMPUTATIONAL RESULTS

6. 1. Parameter Tuning

Parameter tuning is one of the important steps in implementing meta-heuristic algorithms since the parameters of the algorithm play a significant role in the quality of the obtained results and the running time of the algorithm as well. There are various methods for parameter tuning which are mainly based on the design of experiments. Design of experiments helps tune the algorithm parameters using a limited number of experiments which offers good accuracy in a limited time. This study uses the Taguchi method for parameter tuning. The influencing parameters of GA are namely, population size (N_p), the maximum number of generations ($Max\_iteration$), crossover rate (P_c) as well as mutation rate (P_m). The determined experiments using the Taguchi method based on the defined levels of each parameter are summarized in Table 1. It is noteworthy that the Minitab statistical package is used for implementing the Taguchi method in this study.

6. 2. Performance Evaluation of the Bi-level Meta-heuristic Algorithm

This section is dedicated to the obtained results from the proposed mathematical model. The obtained results indicate the optimal location of facilities along with other decision variables. Figure 4 illustrates how the bi-level algorithm converges to the optimal solution. According to this figure, the designed system's effects on the people living near the waste management facilities as the upper-level objective is equal to 611,120 person × ton. According to the findings,

TABLE 1. Parameters and their optimal values

Algorithm	Parameter			
	N_p	$Max\_iteration$	P_c	P_m
First level GA	100	125	0.8	0.1
Second level GA	75	50		

also the total amount of the objective for the lower-level design is equal to $ 38.526. Furthermore, Figure 4 compares the performance of the GA against the well-known Particle Swarm Optimization (PSO). As can be observed, the GA ahieves a better solution by maintaing fine balances between the crossover and mutation operators in searchig the solution space. The main reason for the superiority of the GA to PSO in this study is that the updating process during the iterations of PSO is continuous and requires solution representation with a continuous structure. Notably, the bi-level meta-heuristic algorithms are implemented using MATLAB software on a Core i7 computer with 8 GB of RAM and 2.1 GHz CPU. Table 2 shows the obtained optimal results from the proposed mathematical formulation which was solved using the tuned GA algorithm.

Following are the main findings of this research:

- The obtained results indicate that the designed medical management system for the city of Isfahan requires two trucks for the collection of infectious waste. The model has balanced the utilization of these truck to obtain the most efficient workload.
- Considering the conflict between objectives of the model which is the optimization of both environmental and economic aspects, the model has decided to impose more costs to minimize the environmental issues. That's why the optimum routing decisions for the designed medical waste management system in some districts are longer than what is expected.

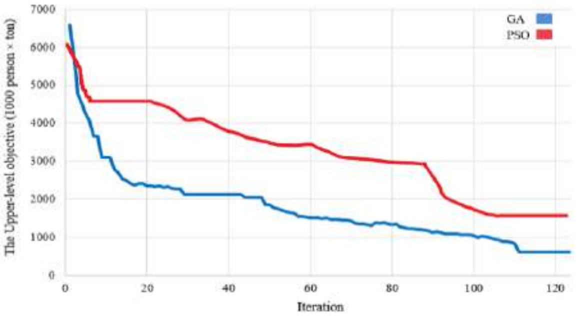


Figure 4. Convergence to the optimal solution

TABLE 2. Optimal waste collection routes

	Waste type			
	Infectious		**Sharp**	**Pathological**
Optimal routes	D14	D5	D14	D14
	↓	↓	↓	↓
	D7	D3	D7	D7
	↓	↓	↓	↓
	D12	D6	D12	D8
	↓	↓	↓	↓
	D2	D4	D8	D12
	↓	↓	↓	↓
	D8	P2	D2	D2
	↓		↓	↓
	D13		D13	D13
	↓		↓	↓
	D1		D6	D5
	↓		↓	↓
	D5		D5	D1
	↓		↓	↓
	D3		D1	D3
	↓		↓	↓
	D10		D3	D10
	↓		↓	↓
	D14		D10	D14
	↓		↓	↓
	F'2		D14	D4
			↓	↓
			D4	D6
			↓	↓
			F'2	C

7. CONCLUSIONS

Considering the benefits of different stakeholders, this paper proposed a mathemathical formulation framework for optimization of the medical waste management system. The upper-level model seeks to design and control the facilities in the medical waste management system by minimizing the environmental risks, while the lower-level model is to determine the waste collection plans by minimizing the total operational costs. As the developed problem is proven to be NP-hard, bi-level meta-heuristic algorithms based on the GA and PSO were employed for solving the problem. Finally, the performance of the developed mathematical formulation and solution approaches were tested in a real medical waste management system. Although the GA offers a high-quality solution, the exact solution methodologies can be approached in the future study.

8. REFERENCES

1. Chang, J.S. and Mackett, R.L., "A bi-level model of the relationship between transport and residential location",

Transportation Research Part B: Methodological, Vol. 40, No. 2, (2006), 123-146. https://doi.org/10.1016/j.trb.2005.02.002

2. Shih, L.-H. and Lin, Y.-T., "Optimal routing for infectious waste collection", ***Journal of Environmental Engineering***, Vol. 125, No. 5, (1999), 479-484. https://doi.org/10.1061/(ASCE)0733-9372(1999)125:5(479)
3. Shih, L.-H. and Chang, H.-C., "A routing and scheduling system for infectious waste collection", ***Environmental Modeling & Assessment***, Vol. 6, No. 4, (2001), 261-269. https://link.springer.com/article/10.1023/A:1013342102025
4. Shih, L.-H. and Lin, Y.-T., "Multicriteria optimization for infectious medical waste collection system planning", ***Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management***, Vol. 7, No. 2, (2003), 78-85. https://doi.org/10.1061/(ASCE)1090-025X(2003)7:2(78)
5. Chaerul, M., Tanaka, M. and Shekdar, A.V., "Resolving complexities in healthcare waste management: A goal programming approach", ***Waste Management & Research***, Vol. 26, No. 3, (2008), 217-232. https://doi.org/10.1177%2F0734242X07076939
6. Alagöz, A.Z. and Kocasoy, G., "Improvement and modification of the routing system for the health-care waste collection and transportation in istanbul", ***Waste Management***, Vol. 28, No. 8, (2008), 1461-1471. https://doi.org/10.1016/j.wasman.2007.08.024
7. Shi, L., Fan, H., Gao, P. and Zhang, H., "Network model and optimization of medical waste reverse logistics by improved genetic algorithm", in International Symposium on Intelligence Computation and Applications, Springer., (2009), 40-52. https://link.springer.com/chapter/10.1007/978-3-642-04843-2_6
8. Nolz, P.C., Absi, N. and Feillet, D., "A stochastic inventory routing problem for infectious medical waste collection ," ***Networks***, Vol. 63, No. 1, (2014), 82-95. https://doi.org/10.1002/net.21523
9. Hachicha, W., Mellouli, M., Khemakhem, M. and Chabchoub, H., "Routing system for infectious healthcare-waste transportation in tunisia: A case study", ***Environmental Engineering and Management Journal***, Vol. 13, No. 1, (2014), 21-29.
10. Alshraideh, H. and Qdais, H.A., "Stochastic modeling and optimization of medical waste collection in northern jordan", ***Journal of Material Cycles and Waste Management***, Vol. 19, No. 2, (2017), 743-753. https://link.springer.com/article/10.1007%2Fs10163-016-0474-3
11. Mantzaras, G. and Voudrias, E.A., "An optimization model for collection, haul, transfer, treatment and disposal of infectious medical waste: Application to a greek region", ***Waste Management***, Vol. 69, (2017), 518-534. https://doi.org/10.1016/j.wasman.2017.08.037
12. Wichapa, N. and Khokhajaikiat, P., "Solving a multi-objective location routing problem for infectious waste disposal using hybrid goal programming and hybrid genetic algorithm", ***International Journal of Industrial Engineering Computations***, Vol. 9, No. 1, (2018), 75-98. DOI: 10.5267/j.ijiec.2017.4.003
13. Davoodi, R., Eslami Hasan Abadi, S., Sabouri, G., Salehi, M., Ghooshkhanei, H., Rahmani, S., Soltanifar, A., Zare Hoseini, M., Asadi, M. and Gharaeian Morshed, M., "Medical waste management in the second largest city of iran (mashhad) with three-million inhabitants", ***Journal of Patient Safety & Quality Improvement***, Vol. 2, No. 4, (2014), 160-164. http://psj.mums.ac.ir/article_3401.html

Persian Abstract

چکیده

این تحقیق روی یک مسئله‌ی مدیریت پسماند خطرناک با ساختار سلسه مراتبی، شامل دولت محلی و برنامه‌ریز مدیریت پسماند تمرکز دارد. سطح بالای این فرآیند سلسه مراتبی در تلاش برای طراحی و کنترل زیرساخت‌های مدیریت پسماند با هدف کمینه‌سازی ریسک‌های زیست‌محیطی مرتبط با دفع پسماند است. از سوی دیگر، مدل ارائه شده برای سطح پایین، به دنبال تعیین طرح‌های جمع‌آوری پسماند صرفا با هدف کمینه‌سازی هزینه‌های عملیاتی است. بنابراین این تحقیق، یک فرمول ریاضی دو سطحی برای توصیف مسئله پیشنهاد می‌کند که منافع هر دو ذینفع در نظر گرفته شود. به دلیل پیچیدگی ساختار مسئله در دستیابی به جواب بهینه، یک الگوریتم فراابتکاری دو سطحی بر پایه‌ی الگوریتم ژنتیک برای حل مسئله مورد استفاده قرار گرفت. در پایان، یک مطالعه‌ی موردی به منظور نمایش کاربردی و کارا بودن مدل و رویکرد حل پیشنهادی صورت پذیرفت.

International Journal of Engineering

Journal Homepage: www.ije.ir

Identifying Tools and Methods for Risk Identification and Assessment in Construction Supply Chain

H. Hernadewita, B. I. Saleh*

Industrial Engineering, Mercu Buana University, Jakarta, Indonesia

PAPER INFO

Paper history:
Received 30 January 2020
Received in revised form 26 March 2020
Accepted 11 June 2020

Keywords:
Construction Supply Chain
Risk identification and assessment method
Analytical Hierical Process
Hazard and Operational
Failure Mode Effect Analysis
Supply Chain Operation Reference

ABSTRACT

The construction project is a business full of risk in every process due to its complexity, changes, and involvement from various stakeholders. One of the critical risks in the construction project is in the supply chain. Identifying and assessing the risk with the right tools and methods in that area will inevitably affect the success of the project. Unfortunately, the research for the tools and methods in a construction supply chain is still limited and scattered. This research objective is to analyze the gap between literature and to create improvement in tools and methods for risk identification and assessment in the construction supply chain. This research will use the systematic literature review method in finding and investigating the tools and methods. The four methods that were found are: Analytical Hierical Process (AHP), Failure Mode Effect Analysis (FMEA), Supply Chain Operation Reference (SCOR), and Hazard and Operational (HAZOP). Strength and weakness with their potential use as tools and methods for identifying and assessing the construction supply chain risk then summarized. The use of SCOR combined with FMEA methods has shown to be practical tools and methods for identifying and assessing the construction supply chain risk.

***doi**: 10.5829/ije.2020.33.07a.18*

1. INTRODUCTION

Construction is a business that consists of risk in every process and exposing to more risks due to their complexity, changes, and various involvement from the stakeholders. Construction is also a project-based business that is temporary, schedule-based, and resource constraint, and failure to create proper risk management will affect the business tremendously. One of the risks that have the most effects is the risk that is associated with supply chain activity, therefore mitigating the risk for the supply chain is the most critical factor to achieve project success [1].

The supply chain is a flow of information, cost, and material that produce value in the form of products or service and delivered to "customers." A construction supply chain is formed by much more complex information, products, and cost that is delivered to the customer as a final product or semi-products. The process of the supply chain is task-based that can acts series or parallel depending on the activity that is affected [3]. Vrijhoef and Koskela [2] have characterized the construction supply chain by the following elements: a) materials for construction works were delivered to a construction site and build inside what called "construction factory," b) the construction supply chain is typified with instability and separation from the design and build, c) a project will produce a new product, little similarities among the products, however, the process could be the same. Gosling and Naim [4,48,49] have constructed and structured the supply chain families based on engineering to order, buy to order, make to order, make to stock, assemble to order, make to stock, and ship to stock structures. The construction supply chain was the most complex system because it involves lots of decision-makers and stakeholders. Uncertainty in the networks has increased within the chains, and more complex the networks are, the more uncertainty and risk will be. A general issue that usually happens in a construction supply chain is

*Corresponding Author Institutional Email: *Irawan_saleh@yahoo.co.id* (B.I. Saleh)

the flow of material, communication in the internal company, project communication, and complexity [5].

Supply Chain Management (SCM) is integration in the business process and improvement of the value within the chain. SCM is aiming to improve productivity and competitiveness, value-added, and profitability for the company and also to the whole supply chain networks, including the end-user. Supply Chain Risk Management (SCRM) by simple definition is a methodology to separate, identify, and mitigate the risk, and ensuring the continuity of the process to achieve profitability[6]. However, the definition of SCRM is still debatable among the researcher. Jüttner, Peck, and Martin [7], for example, defined SCRM as the identification and management of supply chain risk through a coordinated approach among the member of the supply chain to reduce the whole vulnerability. Till now, there still no final consensus on the definition of Supply Chain Risk Management.

Risk identification and assessment are part of the risk management body of knowledge supported by ISO 31000:2009. There is a sixth step standard process in managing the risk, which is the identification, assessment, management, controlling, and communication (Figure 1). Risk identification and assessment is a vital part of the process, where they act as the frontier and responsible for the next phase. Risk identification and assessment are also used in identifying risk types and factors [8].

Research for SCRM has constantly increased. Unfortunately, the majority of industries that have been studied were based on manufacturing, and only a few research has touched the construction. However, now, the construction supply chain has become an exciting topic to discuss, especially in risk management. Furthermore, much research, both in qualitative and quantitative ways, have developed. However, the papers are still scattered and required more effort in modifying the tools and methods to use in practice.

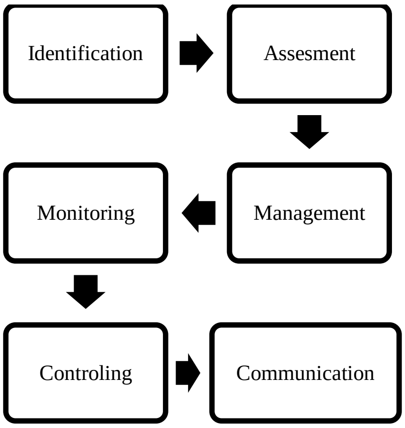


Figure 1. Process of Risk Management

The objective of this paper is to analyze and bridge the gap between literature and to create improvement in tools and methods for risk identification and assessment in the construction supply chain. This paper is organized and divided by sections; Section 2 shows the literature review methodology, describing how to select the literature. Section 3 is analyzing the tools and methods (include the strength and weaknesses) for risk identification and assessment in Supply Chain Risk in the Construction. Section 4 will discuss Proposed Tools and Methods for Risk Identification and Assessment in Construction Supply Chain, and we will conclude this paper in Section 5.

2. LITERATURE REVIEW METHODOLOGY

This paper will follow the methodology from the Systematic Literature Review (SLR)–which is the standard method for investigating a specific subject, which consists of four steps, as seen in Figure 2 below:

In the first stage of this study, papers were selected from the peer-reviewed journal with trustful databases, such as Elsevier, Springer-Link, Francis & Taylor, Inderscience, Emerald, International Journal of Engineering (IJE), Journal of Industrial Engineering and Management (JIEM), International Journal of Industrial Engineering: Theory, Applications, and Practice (IJIETAP), Project Management Journal, Journal of Modern Project Management. Google Scholar also included with careful selection of the journal based on their SCOPUS index. With years of publication range from 2000 -2019.

To achieve the objective of this paper, we will use the keywords “Construction Supply Chain Risk Management," “Construction Supply Chain Risk Identification,” and “Construction Supply Chain Risk Assessment.” These keywords are put in the advanced search where it does not just search in the title but also will search in contents, abstracts, and keywords. The

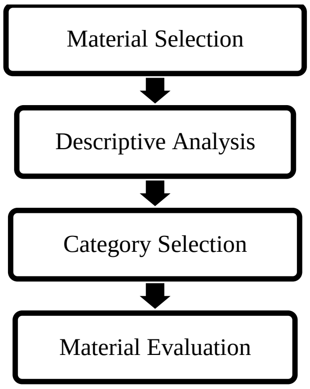


Figure 2. Four steps of Systematic Literature Review

keywords “Construction Supply Chain” combined with the function “And” with “Risk Management” to performs a search for “Construction Supply Chain Risk Management," “Construction Supply Chain” combined with the function “And” with “Risk Identification” to conducts a search for “Construction Supply Chain Risk Identification,” also “Construction Supply Chain” combined with the function “And” with “Risk Assessment” to performs a search for “Construction Supply Chain Risk Assessment” The years then input in advance filter menu for 2000 -2019. The summary of keywords and search location, as seen in Table 1 below:

The category of the papers will be selected based on the criteria : (1) tools and methods (2) Research Type (Case Study, Literature Study, and Survey) (3) Industries (Manufacturing and Construction). The final phase is to analyze the weakness and advantages of the methods, and from the analysis, new or improved methods will be proposed.

3. CATEGORY FOR TOOLS AND METHOD SELECTION REVIEW

Thirty-five journals selected and then categorized based on the tools and methods, research type, and industries.

Figures 3, 4 and 5 summarizes numbers of tools and methods being used, numbers of industries and numbers of research type based on the findings from the selected journal.

Based on Figure 3, there are four methods identified: Analytical Hierarchy Process (AHP), Failure Mode and Effect Analysis (FMEA), Supply Chain Operational Reference (SCOR) model, and Hazard and Operability (HAZOP) analysis. The tables below summarize the tools and methods with their reference.

3. 1. Review of AHP for Risk Identification and Assessments

Gaudenzi and Borghesi [11] have introduced the use of the AHP in their paper to identify and assess risk in the supply chain. They have successfully created a model that can identify a panel of risk indicators that were applied in various levels of the

TABLE 1. List of Keywords and Search Location

Keywords	Search Location
Construction Supply Chain AND Risk Management Construction Supply Chain AND Risk Identification Construction Supply Chain AND Risk Assessment	Elsevier, Springer-Link, Francis & Taylor, Inderscience, Emerald, International Journal of Engineering (IJE), Journal of Industrial Engineering and Management (JIEM), International Journal of Industrial Engineering: Theory, Applications, and Practice (IJIETAP), Project Management Journal, Journal of Modern Project Management. Google Scholar with a selective journal based on the SCOPUS index.

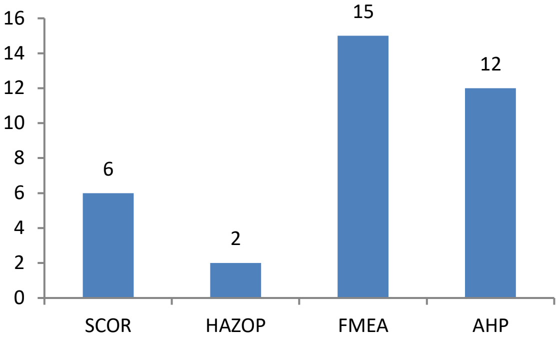


Figure 3. Tools and Methods with total published journal

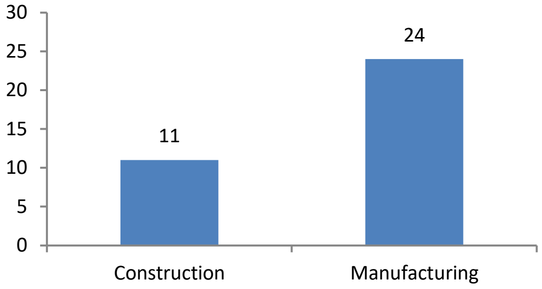


Figure 4. Industries researched with total published journal

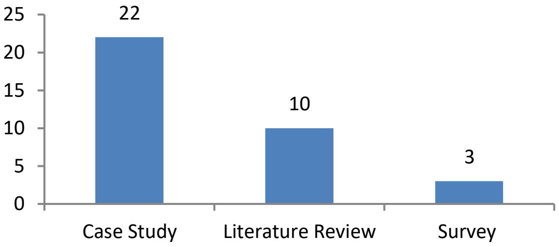


Figure 5. Research type with total published journal

TABLE 2. Summary of tools and methods and references article

Tools and Methods	Industries	References
Analytical Hierarchy Process (AHP)	Manufacturing	[9 -14]
	Construction	[15-20]
Failure Mode and Effect Analysis (FMEA)	Manufacturing	[21-32]
	Construction	[33-35]
Supply Chain and Operational Reference (SCOR)	Manufacturing	[36-39]
	Construction	[40-41]
Hazard and Operational (HAZOP)	Manufacturing	[42-43]

chain. Sharma and Bhat [14] have used AHP by classification of the risk factor in the hierarchy and rated all risks in the pairwise comparison matrix. The papers have successfully shown how to calculate the matrices of AHP and rank the risk prioritization. There are a number of research that successfully combined the AHP models by other methods to identify and assess the risk. Li et al. [12] have used the AHP-fuzzy comprehensive evaluation model, which based on the combination of AHP and fuzzy mathematical theory, to assess the risks in the supply chain. Dong and Cooper [10] have also developed the orders-of-magnitude AHP (OM-AHP) that was enabled to compare tangible and intangible elements that influence supply chain risks, and also succeed in creating a risk assessment based on their probability and consequence severity. AHP is a fascinating method to discuss and to apply in risk identification and assessment for the supply chain, but the research is mostly used in manufacturing industries.

AHP in the construction supply chain was mainly used to assessing supplier or material selection [36, 20, 18]. There is no significant research that AHP was used in risk identification and assessment in the construction supply chain.

In general, the phases of AHP are: 1) defining the objective(s) and preferable solution(s), and 2) creating a hierarchical structure based on the main objective (Figure 6).

Create a pairwise comparison matrix that describes relative contributions or influences of each element to goals or criteria on the same level. To get higher accuracy, it requires a full decomposition until it reaches the end. Some levels are developed from goal, decompose to criterion, and alternatives. The second phase is to set up priority or judgment. Prioritization is being done at every level of the hierarchy. A pairwise judgment matrix constructs by element and element and also compared to their next level using the nine-point rating that has been developed by Saaty [9].

The relative weights were then calculated by the right eigenvector (*w*) corresponding to the largest eigenvalue (λmax), as shown in Equation (1):

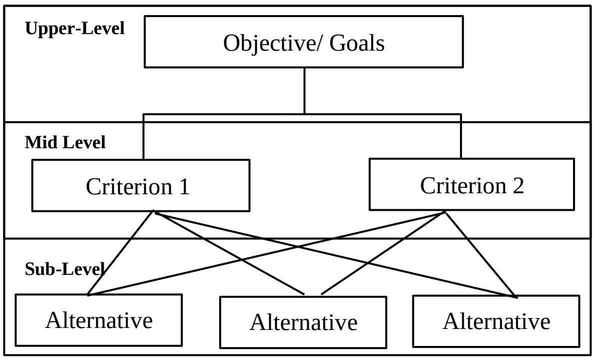


Figure 6. A standard hierarchical structure sample for AHP

$$A_w = \lambda_{max} w \quad (1)$$

The matrix was said consistently if matrix A has a rank of one and $\lambda_{max} = n$, and weights can be obtained by normalizing rows or columns in A. Then, the measure of consistency, called Consistency Index(*CI*), as deviation or degree of consistency is calculated using the following Equation (2):

$$CI = (\lambda_{max} - n)/\ (n-1) \quad (2)$$

The final Consistency Ratio (CR) is then calculated to see whether the evaluation is sufficiently consistent; the calculation is based on Equation (3):

$$CR == \frac{\mathrm{CI}}{RI} \quad (3)$$

where, RI is Random Index, if CR ≤ 10%, then inconsistency is acceptable, however, if the CR is ≥ 10%, the procedure then to be repeated to improve the CR [11-14].

The advantages in using AHP methods in risk identification and assessment in the construction supply chain are: (1) It is a flexible and straightforward model; (2) The evaluation model will be based on the expert judgment from a variety of discipline; (3) Details of risk can be presented detail in level; (4) It can measure the consistency of judgments/decision

However, in the construction supply chain, the AHP methods have some weaknesses, including: (1) Construction supply chain is the most sophisticated model of the supply chain, and the complexity will make AHP become unrealistic methods to run with; (2) Subjective matters on the expert judgments will be the constraint of AHP, wherein construction will require efforts from all project member to get consensus; (3) It will require help from the computational assistance to speed up the process; (4) There is no certainty based on the statistics, where AHP is only a mathematical model.

3. 2. Review of Scor for Risk Identification and Assessments

Only a few studies have been identified for SCOR methods in the identification and assessment of supply chain risk. Faisal, Banwet, and Shankar [37] have shown to mitigate the risk in the supply chain using the SCOR model and analytical network process. [38] has briefly described the use of the SCOR model for evaluating the risks and combined with AHP. Lemghari, Okar, and Sarsri [39] identified the limitation and benefit of the SCOR model in automotive industries. Cheng et al.[40] have comprehensively discussed the use of the SCOR model on the construction supply chain and successfully modeled the construction supply chain based on the SCOR and evaluating the processes performance. Pan, Lee, and Chen [41] have also used the SCOR model to improve the supply chain system in construction. All of the research has not used the newest version of the

SCOR, as seen in Figure 6.

Building SCOR methods for the construction supply chain is as follows: (1) The material needs to be identified as engineering to order, buy to order, make to order, make to stock, assemble to order, make to stock, or ship to stock; (2) SCOR level 1 (Figure 6) and level 2 models are created based on the material; (3) Create a Level 3 SCOR Business Model; (4) The last one is to identify the risk in the processes. The general model of level in the SCOR can be seen in Figure 7.

The advantages of using SCOR for risk identification and assessment in Construction Supply Chain are: (1) The business process is identified based on the organization of the material; (2) It is a standardized method in modeling the supply chain based on the business process; (3) Risk can be identified within the process in the supply chain. Although the SCOR model seems to be a powerful method, it also has weaknesses and limitations, including: (1) The models in Level 1-3 in SCOR model are based on the knowledge of the process; (2) Creating the process required involvement with the experienced and skilled team that knows Construction Supply Chain; (3) It requires training and experience in developing the SCOR model.

3. 3. Review of Fmea for Risk Identification and Assessments

The Failure Mode Effect Analysis (FMEA) has gainedpopularity in the risk management tools and reached the supply chain risk management for years. FMEA is a hybrid tool derived from the Fault Tree Analysis (FTA) methods. It was one of the popular methods and has been used by professionals and researchers in risk identification and assessment. Curkovic, Scannell, and Wagner [24] have made a study of how FMEA is used in managing risk in the company and how familiar the stakeholder is in the methodology. Shinha, Whitman, and Malzahn [30] have used FMEA for risk assessment in supply chain risk management. FMEA has also been widely used in assessing the performance of the supplier, logistics, and material in the supply chain. In the construction projects, FMEA was mostly used in assessing the project risk. Rohmah et al. [28] have used in their paper fuzzy FMEA methods to assess the risks.

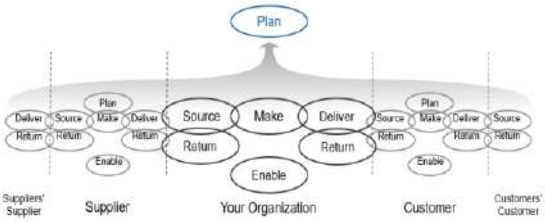
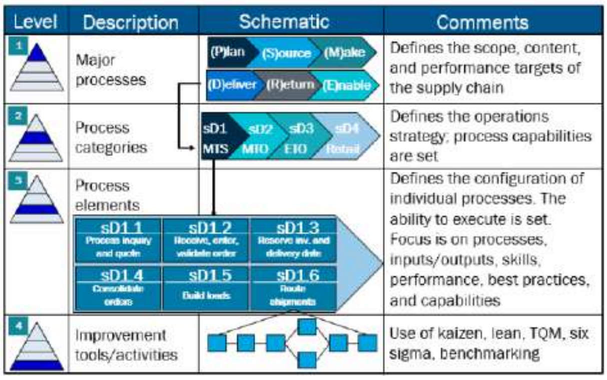


Figure 6. SCOR version 12.0 Level 1 – APICS



Level	Description	Schematic	Comments
1	Major processes	(P)lan (S)ource (M)ake (D)eliver (R)eturn (E)nable	Defines the scope, content, and performance targets of the supply chain
2	Process categories	sD1 MTS sD2 MTO sD3 ETO sD4 Retail	Defines the operations strategy; process capabilities are set
3	Process elements	sD1.1 Process inquiry and quote sD1.2 Receive, enter, validate order sD1.3 Reserve inv. and delivery date sD1.4 Consolidate orders sD1.5 Build loads sD1.6 Route shipments	Defines the configuration of individual processes. The ability to execute is set. Focus is on processes, inputs/outputs, skills, performance, best practices, and capabilities
4	Improvement tools/activities		Use of kaizen, lean, TQM, six sigma, benchmarking

Figure 7. Four levels of SCOR version 12.0 business processes-APICS

Unfortunately, significant research focusing on using FMEA for risk identification and assessment in the construction supply chain has not yet been identified. However, FMEA was used to identifying and assessing the risk in the whole construction project processes and also mostly used to identify the risk from the supplier, logistics, product, and material in the construction supply chain [33 -35].

The steps in FMEA are as follows: A table is generated as standard FMEA table (Table 3), FMEA team that consists of experts in the area need to be assembled for justification and judgment, the process in the supply chain then be listed in the table comprehensively. Failure mode(s) are listed in the table for every process steps, that are: (1) Listing the effects of failure; (2) Inputing the severity rating based on the agreed scale, with scale 1-10 (from low to high severity); (3) Identification of the potential cause of failure, input the occurrence factor from the potential cause of failure with scale 1- 10 (from low to a high probability); (4) Identification of the control to detect the risks, input the rating for detection, usually with scale 1-10, (5) Calculating Risk Priority Number (RPN) based on Equation (4), and (6) input the recommended actions for mitigation [27].

$$RPN = Severity\ factor \times Occurence\ or\ Probability\ factor \times Detection\ factor \tag{4}$$

FMEA has offered several advantages for identification and assessment in construction supply chain risk, which are: (1) It is a simple method and commonly is used practically in assessing the supply chain and project risks; (2) It is an early identification to identify and mitigate the risks; (3) Risk prioritization can be identified; (4) Create a sense of belonging in each of department for the risks; Can capture most of the risks.

Some of the weaknesses of using FMEA for risk identification and assessment in the construction supply chain are: (1) Factors in severity, occurrence/probability, and detection were based on the agreement. Therefore, the numbers are not statistically correct and somehow potentially bias; (2) The identification of risk will be based on the knowledge of the experts, which will limit the risk and potentially losing some of the critical risks.

3. 4. Review of Hazop for Risk Identification and Assessments

The Hazard and Operability (HAZOP) was initially developed in the chemical process industry and has now been widely used to assess the risk associated with health, safety, and the environment in the process and manufacturing industries. Trough the years, the researcher has widely spread the use of HAZOP into several risk management process and have touched the SCRM. Adithya, Srinivasan, Karimi [42] have used the Hazard and Operability (HAZOP) method to identify the risk involved in the supply chain, by following the general rule in HAZOP methods where risks are drawn using a diagram. The diagram itself is following the process flow diagram. Mitkowski and Zenka-Podlaszewka [43] have successfully transferred the HAZOP method from the process to supply chain management and identified the risk in the supply chain. The search for HAZOP as a method in Supply Chain Risk Management in Construction has come to a disappointment. Most of the literature showing the use of HAZOP for assessing the design and processing of the construction.

The first step in using HAZOP is creating what was called Work Flow Diagram (WFD) and Supply Chain Flow Diagram (SCFD), forms of the diagram similar to Process Flow Diagram (PFD). The WFD describes the sequence of works of one or more activities. Examples of WFD is shown in Figure 8.

Supply Chain Flow Diagram (SCFD) is showing the connections between the chain. It contains the flow of material and information along the chain. After the Work Flow Diagram and Supply Chain Diagram are created, a risk analysis process is conducted by a specialized team.

The advantages using HAZOP for risk identification and assessment in the construction supply chain are: (1) Flow processes are described comprehensively; (2) The risk in every chain is identified; (3) It is a systematic model to identifying the risks. However, HAZOP has several weaknesses, especially in the construction supply chain, that are: (1) The flow of material, information, and cost in the supply chain are complicated, where every chain can intervene one another, and that can cause more issue in the HAZOP model; (2) The diagram still does not have a common standard, there will be variety in creating the diagram; (3) It will take time and effort in describing one process to another.

TABLE 3. A standard FMEA Table

Process Input	Failure Mode/Risk	Effect(s) of Failure	Severity (1-10)	Potential Cause(s)/ Mechanism(s) of Failure	Occurance /Probability (1-10)	Current Process Controls	Detection (1-10)	RPN (Risk Piority Number)	Recommended Action(s)
What is the process step or feature under investigation?	In what ways could the step or feature go wrong?	What is the impact on the customer if this failure is not prevented or corrected?	Scale 1-10 based on the severity	What causes the step or feature to go wrong? (how could it occur?)	Scale 1-10 based on the occurrence/ probability	What controls exist that either prevent or detect the failure?	Scale 1-10 based on the detection	RPN = Sev x Occ x Detc	What are the recommended actions for reducing the occurrence of the cause or improving detection?

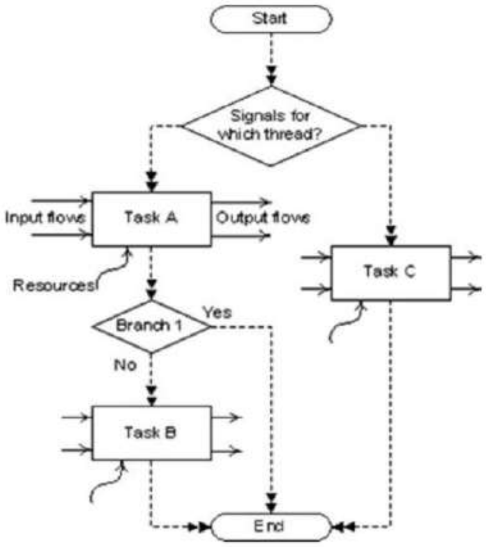


Figure 8. Structure of Work Flow Diagram (WFD) [42]

4. PROPOSED TOOLS AND METHODS FOR RISK IDENTIFICATION AND ASSESSMENT IN CONSTRUCTION SUPPLY CHAIN

In general, the four methods that have been described in this paper are not directly implied and gained their popularity for identifying and assessing the risk in the construction supply chain. However, to assess the applicability of those methods, we will look back on their advantages and weaknesses compared with the nature of the construction supply chain, driven by the complexity of the processes, the structure of the materials, and considerable stakeholder involvement. AHP and HAZOP models have their weakness in their flexibility to withstand the complexity of the supply chain process. Moreover, it will be unrealistic and impractical to use those methods in significant and complex construction projects. It will consume time and effort, and sometimes losing the significant risk that needs to be recorded.

On the other hand, the SCOR method has several advantages that are: the models can describe the supply chain behavior in every step of the processes; showing the risk in each level in the business process; it has become a universal standard tool in describing supply chain process. FMEA also has its advantages in its simplicity, and it can capture most of the risk, and model the risk prioritization in simple mathematical methods. FMEA is a popular model in supply chain risk management and has gain familiarity in construction projects. One of the weaknesses of using FMEA is that it requires a correct input of the process so that the risk can be identified and assessed correctly.

Based on these reviews, we proposed a method for identifying and assessing the risk in the construction supply chain by using SCOR and FMEA methods. The SCOR methods will be used as the first phase. This consists of three steps: (1) Identifying the material based on engineering to order, buy to order, make to order, make to stock, assemble to order, make to stock, and ship to stock; (2) Identifying the level 1 and 2 models in SCOR based on the category of material; (3) Creating a level 3 business process in the SCOR model.

Then, next steps are using FMEA model to identify and assess the risk, with following steps: (1) Processing from level 3, then, including the process column in FMEA table; (2) continue to follow the steps of identifying the risk/failure, effects of the risk, severity factor, potential cause, probability/occurrence factor, process control, detection factor, calculating the risk priority number, and the risk mitigation; (3) prioritizing Risk and selected by their criticality using other assistance tools (i.e., Pareto chart). The frameworks can be seen in Figure 9:

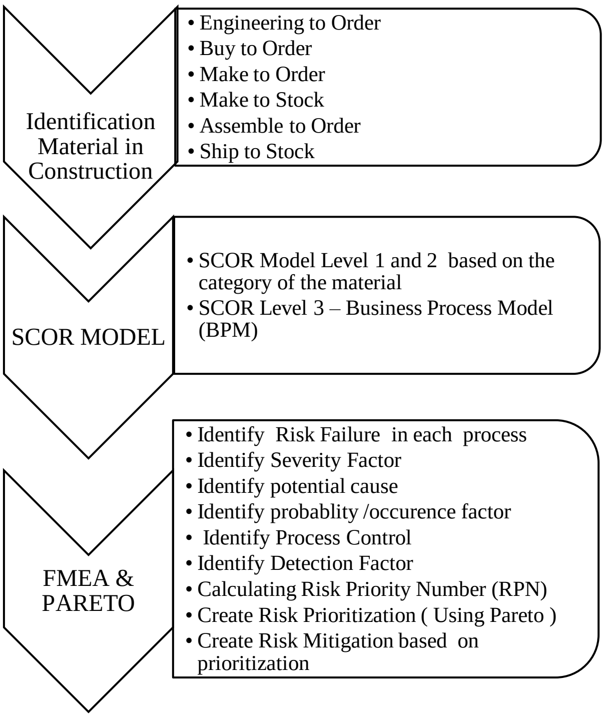

Identification Material in Construction
• Engineering to Order
• Buy to Order
• Make to Order
• Make to Stock
• Assemble to Order
• Ship to Stock
SCOR MODEL
• SCOR Model Level 1 and 2 based on the category of the material
• SCOR Level 3 – Business Process Model (BPM)
FMEA & PARETO
• Identify Risk Failure in each process
• Identify Severity Factor
• Identify potential cause
• Identify probablity /occurence factor
• Identify Process Control
• Identify Detection Factor
• Calculating Risk Priority Number (RPN)
• Create Risk Prioritization (Using Pareto)
• Create Risk Mitigation based on prioritization

Figure 9. Proposed frameworks for risk identification and assessment in the Construction Supply Chain

5. CONCLUSION

This paper has provided a systematic literature review for risk identification and assessment frameworks in the construction supply chain. Articles published in 2000-2019 are collected. 35 articles were selected and reviewed thoroughly. We have summarized the four methods in risk identification and assessment in the supply chain, which are: Analytical Hierarchy Process (AHP), Supply Chain Operating Reference (SCOR), Failure Mode and Effect Analysis (FMEA) and Hazard and Operational (HAZOP). However, the four methods have not yet being used directly for identifying and assessing risk in the construction supply chain. To select the best models, we explore each of the methods to fit in the construction project, and we have found that combining the SCOR model and FMEA will be the best and efficient methods of risk identification and assessment. Further research opportunities for applicating this method is still open, where case studies based on these methods are required.

6. ACKNOWLEDGMENTS

The Authors would like to thanks and show their gratitude to Mercu Buana University for having given the opportunity to do the research and to other lecturers for valuable inputs and comments during this research.

7. REFERENCES

1. Aloini, D., Dulmin, R., Mininno, V. and Ponticelli, S., “Supply chain management: A review of implementation risks in the construction industry”. ***Business Process Management Journal***, Vol. 18, No. 5, (2012), 735-761, DOI: https://doi.org/10.1108/14637151211270135.

2. Vrijhoef, R. and Koskela, L., "The four roles of supply chain management in construction". ***European Journal of Purchasing and Supply Management***, Vol. 6, No. 3-4, (2000), 169-178, DOI: https://doi.org/10.1016/S0969-7012(00)00013-7

3. Pujawan, I. N. and Geraldin, L. H., "House of risk: A model for proactive supply chain risk management". ***Business Process Management Journal***, Vol. 15, No. 6, (2009), 953-967, DOI: https://doi.org/10.1108/14637150911003801.

4. Gosling, J. and Naim, M. M., "Engineer-to-order supply chain management: A literature review and research agenda". ***International Journal of Production Economics***, Vol. 122, No. 2, (2009), 741-754, DOI: https://doi.org/10.1016/j.ijpe.2009.07.002.

5. Thunberg, M., and Fredriksson, A., "Bringing planning back into the picture–How can supply chain planning aid in dealing with supply chain-related problems in construction?" ***Construction Management and Economics***, Vol. 36, No. 8, (2018), 425-442, DOI: https://doi.org/10.1080/01446193.2017.1394579

6. Faizal, K., and Palaniappan, P. K. "Risk Assessment and Management in Supply Chain". ***Global Journal Of Research In Engineering.*** (2014), 14, 19-30. https://engineeringresearch.org/index.php/GJRE/article/view/1130

7. Jüttner, U., Peck, H., and Martin, C.,"Supply Chain Risk Management: Outlining an Agenda for Future Research". ***International Journal of Logistics : Research & Applications,*** 6. (2003), 197-210, DOI: https://doi.org/10.1080/13675560310001627016

8. Ho, W., Zheng, T., Yildiz, H., and Talluri, S. "Supply chain risk management: A literature review". ***International Journal of Production Research***, Vol. 53, No. 16, (2015), 5031-5069, DOI: https://doi.org/10.1080/00207543.2015.1030467

9. Badea, A., Prostean, G., Goncalves, G. and Allaoui, H., "Assessing Risk Factors in Collaborative Supply Chain with the Analytic Hierarchy Process (AHP)". ***Procedia-Social and Behavioral Sciences***, Vol. 124, (2014), 114-123, DOI: https://doi.org/10.1016/j.sbspro.2014.02.467.

10. Dong, Q. and Cooper, O., "An orders-of-magnitude AHP supply chain risk assessment framework". ***International Journal of Production Economics***, Vol. 182, (2016), 144-156, DOI: https://doi.org/10.1016/j.ijpe.2016.08.021.

11. Gaudenzi, B., and Borghesi, A., "Managing Risks in the Supply Chain Using the AHP Method", ***The International Journal of Logistics Management***, Vol. 17, No. 1, (2006), 114-136, DOI: https://doi.org/10.1108/09574090610663464.

12. Li, M., Du, Y. Wang, Q., Sun, C., Ling, X., Yu, B., and Xiang, Y., "Risk assessment of supply chain for pharmaceutical excipients with AHP-fuzzy comprehensive evaluation". ***Drug Development and Industrial Pharmacy***, Vol. 42, No. 4, (2016), 676-684, DOI: https://doi.org/10.3109/03639045.2015.1075027.

13. Samvedi, A., Jain, V, and Chan , F.T.S., "Quantifying Risks in a Supply Chain through Integration of Fuzzy AHP and Fuzzy TOPSIS", ***International Journal of Production Research***, Vol. 51, No. 8, (2013), 2433-2442, DOI: https://doi.org/10.1080/00207543.2012.741330.

14. Sharma, S. K. and Bhat, A., "Identification and assessment of supply chain risk: Development of AHP model for supply chain risk prioritization". ***International Journal of Agile Systems and Management***, Vol. 5, No. 4, (2012), 350-369, DOI: https://doi.org/10.1504/IJASM.2012.050155.

15. Amade, B., Akpan, E. O. P., Ukwuoma, Ononuju, C. N. and Okore, O. L., " A Supply Chain Management (SCM) Framework for Construction Project Delivery in Nigeria: An Analytical Hierarchy Process (AHP) Approach". ***PM World Journal***, Vol. 7, No. 4, (2018), 1-26. Retrieved from: https://pmworldlibrary.net/article/a-supply-chain-management-scm-framework-for-construction-project-delivery-in-nigeria-an-analytical-hierarchy-process-ahp-approach/

16. Soo Yong, K.,and Nguyen, V.T,"An AHP Framework for Evaluating Construction Supply Chain Relationships", ***KSCE Journal of Civil Engineering,*** 22.5 (2018), 1544-56, DOI: https://doi.org/10.1007/s12205-017-1546-1

17. Rahimi, Y., Tavakkoli-Moghaddam, R., Shojaie, S. and Cheraghi, I., "Design of an innovative construction model for supply chain management by measuring agility and cost of quality: An empirical study". ***Scientia Iranica***, Vol. 24, No. 5, (2017), 2515-2526, DOI: https://doi.org/10.24200/sci.2017.4388.

18. Shirude, A., Shelake, K., Mahavarkar, P. and Bokil, S., "Smart Decision Making Technique in Construction Supply Chain Management for Infrastructure Engineering Projects". 4th National Level Construction Techies Conference Advances in Infrastructure Development and Transportation Systems in Developing India, (2018), 44-47. http://ijermce.com/specissue/may/8.pdf

19. Wang, T. K., Zhang, Q., Chong, H. Y. and Wang, X., "Integrated supplier selection framework in a resilient construction supply chain: An approach via analytic hierarchy process (AHP) and grey relational analysis (GRA)". Sustainability (Switzerland), Vol. 9, No. 2, (2017), DOI: https://doi.org/10.3390/su9020289.

20. Waris, M., Shrikant, P., Mengal, A., Soomro, M.I., Mirjat, N.H., Ulah, M., Azlan, Z. S., and Khan, A., "An application of analytic hierarchy process (AHP) for sustainable procurement of construction equipment: Multicriteria-based decision framework for Malaysia". ***Mathematical Problems in Engineering***, (2019), DOI : https://doi.org/10.1155/2019/6391431

21. Angara, R. A., "Implementation of Risk Management Framework in Supply Chain: A Tale from a Biofuel Company in Indonesia". ***SSRN Electronic Journal***, (2012), DOI: https://doi.org/10.2139/ssrn.1763154.

22. Ariyanti, F. D. and Andika, "A Supply chain risk management in the indonesian flavor industry: Case study from a multinational flavor company in Indonesia". ***Proceedings of the International Conference on Industrial Engineering and Operations Management***, Vol. 8-10 March, (2016), 1448–1455. Retrieved from: http://ieomsociety.org/ieom_2016/pdfs/401.pdf

23. Chen, P. S. and Wu, M. T., "A modified failure mode and effects analysis method for supplier selection problems in the supply chain risk environment: A case study". ***Computers and Industrial Engineering***, Vol. 66, No. 4, (2013), 634-642, DOI: https://doi.org/10.1016/j.cie.2013.09.018.

24. Curkovic, S., Scannell, T. and Wagner, B., "Using FMEA for Supply Chain Risk Management. ***Managing Supply Chain Risk***, No. 2, (2015), 25-42, DOI: https://doi.org/10.1201/b18610-3.

25. Giannakis, M. and Papadopoulos, T., "Supply chain sustainability: A risk management approach". ***International Journal of Production Economics***, Vol. 171, (2016), 455-470, DOI: https://doi.org/10.1016/j.ijpe.2015.06.032.

26. Li, S. and Zeng, W., "Risk analysis for the supplier selection problem using failure modes and effects analysis (FMEA)". ***Journal of Intelligent Manufacturing***, Vol. 27, No. 6, (2016), 1309-1321, DOI: https://doi.org/10.1007/s10845-014-0953-0.

27. Neghab, A. P., Siadat, A., Tavakkoli-Moghaddam, R. and Jolai, F., "An integrated approach for risk-assessment analysis in a manufacturing process using FMEA" 2011 IEEE International Conference on Quality and Reliability, ICQR 2011, (2011), 366–370, DOI : https://doi.org/10.1109/ICQR.2011.6031743.

28. Rohmah, D., Urianty, M. Dania, W.A.P, and Dewi, I.A, "Risk Measurement of Supply Chain Organic Rice Product Using Fuzzy Failure Mode Effect Analysis in MUTOS Seloliman

Trawas Mojokerto", ***Agriculture and Agricultural Science Procedia***, Vol. 3, (2015), 108-113, DOI: https://doi.org/10.1016/j.aaspro.2015.01.022.

29. Simba, S., Niemann, W., Kotzé, T. and Agigi, A., "Supply chain risk management processes for resilience: A study of South African grocery manufacturers". ***Journal of Transport and Supply Chain Management***, Vol. 11, No. 0, (2017), 1-13, DOI: https://doi.org/10.4102/jtscm.v11i0.325.

30. Sinha, P. R., Whitman, L. E. and Malzahn, D., "Methodology to mitigate supplier risk in an aerospace supply chain". ***Supply Chain Management***, Vol. 9, No. 2, (2004), 154-168, DOI: https://doi.org/10.1108/13598540410527051.

31. Sutrisno, A., Kwon, H. M., Lee, T.-R. Jiun-S. and Ae, J. H." Improvement Strategy Selection in FMEA – Classification, Review and New Opportunity Roadmaps". ***Operations and Supply Chain Management: An International Journal***, Vol. 6, No. 2, (2014), 54, DOI: https://doi.org/10.31387/oscm0140088.

32. Teng, S. G., Ho, S. M., Shumar, D. and Liu, P. C., "Implementing FMEA in a collaborative supply chain environment". ***International Journal of Quality and Reliability Management***, Vol. 23, No. 2, (2006), 179-196, DOI: https://doi.org/10.1108/02656710610640943.

33. Azambuja, M. and Chen, X., "Risk assessment of a ready-mix concrete supply chain". Construction Research Congress 2014: Construction in a Global Network - Proceedings of the 2014 Construction Research Congress, No. 2003, (2014), 1695-1703, DOI : https://doi.org/10.1061/9780784413517.0173.

34. Bajpai, P., Kalra, M. and Sunyna B, "A. Supply Chain Management of Road Projects in India using FMEA and ISM Technique". ***Indian Journal of Science and Technology***, Vol. 9, No. S1, (2016), 1-5, DOI: https://doi.org/10.17485/ijst/2016/v9is1/105699.

35. Mecca, S and Masera, M, "Technical risk analysis in construction by means of FMEA methodology", In: 15th Annual ARCOM Conference**,** Vol. 2, (1999), 425-34. https://pdfs.semanticscholar.org/5542/a0f5eae1b684b17481d5805b739e25ffa518.pdf

36. Abolghasemi, M., Khodakarami, V. and Tehranifard, H., "A new approach for supply chain risk management: Mapping SCOR into Bayesian network". ***Journal of Industrial Engineering and Management***, Vol. 8, No. 1, (2015), 280-302, DOI : https://doi.org/10.3926/jiem.1281.

37. Faisal, M. N., Banwet, D. K. and Shankar, R., "Management of Risk in Supply Chains: SCOR Approach and Analytic Network Process". ***Supply Chain Forum: An International Journal***, Vol. 8, No. 2, (2007), 66-79, DOI: https://doi.org/10.1080/16258312.2007.11517183.

38. Jiang, B., Li, J. and Shen, S., "Supply Chain Risk Assessment and Control of Port Enterprises: Qingdao port as case study". ***Asian Journal of Shipping and Logistics***, Vol. 34, No. 3, (2018), 198-208, DOI: https://doi.org/10.1016/j.ajsl.2018.09.003.

39. Lemghari, R., Okar, C. and Sarsri, D., "Benefits and limitations of the scor® model in automotive industries". MATEC Web of Conferences, Vol. 200, (2018), DOI: https://doi.org/10.1051/matecconf/201820000019.

40. Cheng, J. C. P., Law, K. H., Bjornsson, H., Jones, A. and Sriram, R. D." Modeling and monitoring of construction supply chains". ***Advanced Engineering Informatics***, Vol. 24, No. 4, (2010), 435-455, DOI: https://doi.org/10.1016/j.aei.2010.06.009.

41. Pan, N. H., Lee, M. L. and Chen, S. Q., "Construction material supply chain process analysis and optimization". ***Journal of Civil Engineering and Management***, Vol. 17, No. 3, (2011), 357-370, DOI : https://doi.org/10.3846/13923730.2011.594221.

42. Adhitya, A., Srinivasan, R. and Karimi, I. A., "Supply chain risk management through HAZOP and dynamic simulation". ***Computer Aided Chemical Engineering***, Vol. 25, (2008), 37-42, DOI: https://doi.org/10.1016/S1570-7946(08)80011-9.

43. Mitkowski, P. T. and Zenka-Podlaszewska, D., "HAZOP method in identification of risks in a CPFR supply chain". ***hemical Engineering Transactions***, Vol. 39, No. Special Issue, (2014), 445-450, DOI: https://doi.org/10.3303/CET1439075.

44. Bennett, J. and Ormerod, R. N., "Simulation applied to construction projects", ***Construction Management and Economics***, Vol. 2, No. 3, (1984), 225-263, DOI: https://doi.org/10.1080/01446198400000021

45. Dallasega, P., Rojas, R. A., Bruno, G. and Rauch, E., "An agile scheduling and control approach in ETO construction supply chains". ***Computers in Industry***, Vol. 112, (2019), 103-122, DOI: https://doi.org/10.1016/j.compind.2019.08.003.

46. Ennouri, W., "Risk Management Applying Fmea-Steg Case Study". ***Polish Journal of Management Studies***, Vol. 11, No. 1, (2015), 56-67. https://pjms.zim.pcz.pl/resources/html/article/details?id=167187

47. Gao, Q., Guo, S., Liu, X., Manogaran, G., Chilamkurti, N., and Kadry, S., "Simulation analysis of supply chain risk management system based on IoT information platform". (2019), 1-25, DOI: https://doi.org/10.1080/17517575.2019.1644671.

48. Gosling, J., Naim, M. and Towill, D., "Identifying and categorizing the sources of uncertainty in construction supply chains". ***Journal of Construction Engineering and Management***, Vol. 139, No. 1, (2013), 102-110, DOI: https://doi.org/10.1061/(ASCE)CO.1943-7862.0000574.

49. Gosling, J., Towill, D. and Naim, M., "Learning how to eat an elephant: Implementing supply chain management principles". Association of Researchers in Construction Management, ARCOM 2012 - Proceedings of the 28th Annual Conference, Vol. 1, No. September, (2012), 633–643. Retrieved from: https://pdfs.semanticscholar.org/0df9/433baad6c1c8ceb6f4f06b7a93f02eab8ac4.pdf

50. Green, S. D., Fernie, S. and Weller, S., "Making sense of supply chain management: A comparative study of aerospace and construction". ***Construction Management and Economics***, Vol. 23, No. 6, (2005), 579-593, DOI: https://doi.org/10.1080/01446190500126882.

51. Jones, M." Supply chain management in construction". ***Construction Project Management: An Integrated Approach***, Vol. 4, No. 3, (2005), 308-339, DOI: https://doi.org/10.4324/9780203006986

52. Kenley, C. R., "Requirements Risk Assessment - Integrating QFD and Risk Assessment". ***INCOSE International Symposium***, Vol. 14, No. 1, (2004), 1615-1623, DOI: https://doi.org/10.1002/j.2334-5837.2004.tb00599.x.

53. London, K. and Singh, V., "Integrated construction supply chain design and delivery solutions". ***Architectural Engineering and Design Management***, Vol. 9, No. 3, (2013), 135-157, DOI: https://doi.org/10.1080/17452007.2012.684451.

54. Loosemore, M., Alkilani, S. and Mathenge, R., "The risks of and barriers to social procurement in construction: a supply chain perspective". ***Construction Management and Economics***, (2019), 1-18, DOI: https://doi.org/10.1080/01446193.2019.1687923.

55. Ouabouch, L. and Paché, G., "Risk management in the supply chain: Characterization and empirical analysis". ***Journal of Applied Business Research***, Vol. 30, No. 2, (2014), 329-340, DOI : https://doi.org/10.19030/jabr.v30i2.8401.

56. Ritchie, B. and Brindley, C., "An emergent framework for supply chain risk management and performance measurement". ***Journal of the Operational Research Society***, Vol. 58, No. 11, (2007), 1398-1411, DOI: https://doi.org/10.1057/palgrave.jors.2602412.

57. Shojaei, P. and Haeri, S. A. S., "Development of supply chain risk management approaches for construction projects: A grounded theory approach". ***Computers and Industrial Engineering***, Vol. 128, (2019), 837-850, DOI: https://doi.org/10.1016/j.cie.2018.11.045.

58. Shu, T., Chen, S., Wang, S. and Lai, K. K., "GBOM-oriented management of production disruption risk and optimization of supply chain construction". ***Expert Systems with Applications***, Vol. 41, No. 1, (2014), 59-68, DOI: https://doi.org/10.1016/j.eswa.2013.07.011.

59. Tah, J. H. M., and Carr, V., "Towards a framework for project risk knowledge management in the construction supply chain". ***Advances in Engineering Software***, Vol. 32, No. 10-11, (2001), 835–846, DOI: https://doi.org/10.1016/S0965-9978(01)00035-7.

60. Tazehzaded, M., Rezaei, A., and Kamali, S., "Supply Chain Risk Management in Canadian Construcion Industry". 11th International Congress on Civil Engineering 8-10 May 2018, University of Tehran, (2018), 1-7. Retrieved from: https://www.researchgate.net/publication/327645157_Supply_Chain_Risk_Management_in_Canadian_Construction_Industry

Persian Abstract

چکیده

پروژه‌ی ساخت‌وساز به دلیل پیچیدگی، تغییرات، و درگیری ذی‌نفعان مختلف، مشاغل پُرخطر در هر فرایند است. یکی از خطرات مهم در پروژه‌ی ساخت‌وساز در زنجیره‌ی تأمین است. شناسایی و ارزیابی ریسک با ابزارها و روش‌های مناسب در آن منطقه ناگزیر بر موفقیت پروژه خواهد بود. متأسفانه تحقیقات در مورد ابزارها و روش‌های موجود در زنجیره‌ی تأمین ساخت‌وساز هنوز محدود و پراکنده است. هدف این تحقیق، تحلیل شکاف بین ادبیات و ایجاد پیشرفت در ابزارها و روش‌های شناسایی ریسک و ارزیابی در زنجیره‌ی تأمین ساخت‌وساز است. در این تحقیق از روش مرور ادبیات منظم در یافتن و بررسی ابزارها و روش‌ها استفاده خواهد شد. چهار روشی که برای این تحلیل پیدا شد عبارتند از: فرایند تحلیلی سلسله‌مراتبی (AHP)، تحلیل امکان بروز خطا و اثرات آن (FMEA)، مرجع عملکرد زنجیره‌ی تامین (SCOR) و خطر و عملیاتی (HAZOP). سپس، نقاط قوت و ضعف با استفاده‌ی بالقوه از آنها به عنوان ابزار و روش های شناسایی و ارزیابی ریسک زنجیره‌ی تأمین ساخت‌وساز خلاصه می‌شوند. استفاده از SCOR همراه با روش‌های FMEA نشان داده است که ابزار و روش‌های عملی برای شناسایی و ارزیابی ریسک زنجیره‌ی تأمین ساخت‌وساز است.

International Journal of Engineering

Journal Homepage: www.ije.ir

A New Model of Equivalent Modulus Derived from Repeated Load CBR Test

A. Salmi*[a], L. Bousshine[a]. K. Lahlou[b]

[a] *Structural and Material Mechanics team, National High School of Electricity and Mechanics, Hassan II University of Casablanca, Casablanca, Morocco*
[b] *Research Team in Construction Engineering, LaGCHEC Laboratory, Hassania School of Public Works, Casablanca, Morocco*

PAPER INFO

Paper history:
Received 23 February 2020
Received in revised form 10 Mars 2020
Accepted 12 June 2020

Keywords:
Modulus
CBR
Granular Material
Least-squares Method
Stiffness
Strain Gauge
Resilient Behavior

ABSTRACT

This paper presents a new model of equivalent modulus derived from the Repeated Load CBR (RL-CBR) test without strain gauge. This model is an updated version of Araya et al. model (2011), the update consists of using the vertical strain as weighting factor instead of vertical displacement in the mean vertical and horizontal stresses calculation. The accuracy of equivalent modulus was improved by decreasing the relative error from 25% to 3%. The extra-large mold adopted by Araya et al. is used with a thickness of 8 mm instead of 14.5 mm. In experimental investigations, equivalent modulus may be calculated from experimental data and model parameters estimated by finite element (EF) simulation. There are five model parameters when the RL-CBR test is used, and three parameters when the strain gauge is not used. Model parameters are determined in two steps. First, the FE simulation of the RL-CBR test is conducted using various loading conditions (i.e., plunger penetration) and various quality ranges of unbound granular materials (UGM). In the second step, the non-linear multidimensional regression is accomplished to fit the equivalent modulus to Young's modulus. The influence of FE analysis inputs is investigated to find the optimal inputs set that make the best compromise between the model accuracy and the calculation time consumption. The calculation of model parameters is carried out based on the optimal set data. Results from the new model and those from Araya et al. model are compared and have shown the improved accuracy of the developed model.

doi*: 10.5829/ije.2020.33.07a.19*

NOMENCLATURE		$abs(.)$	Absolute value function
CBR	Californian Bearing Ratio	SDS	Standard deviation sum
RL-CBR	Repeated load CBR	RD	Relative deviation (10^{-3})
RLT	Repeated Load Triaxial	RE	Relative error (10^{-2})
LSM	Least-Squares Method	n	Number of data line in a set of analyses
UGM	Unbound Granular Materials	E_i	Young's (exact) modulus for the ith dat line (MPa)
M_r	Resilient modulus (MPa)	E_{eqi}	Equivalent modulus for the the ith dat line (MPa)
q	Deviator stress (kPa)	**Greek Symbols**	
u	Resilient plunger penetration (mm)	σ_{r1}, σ_{r3}	Resilient axial and confining pressure (kPa)
f	Constant factor (= 2 or π/2)	$\epsilon_{r1}, \epsilon_{r3}$	Recoverable axial and radial micro-strains
E	Young's modulus (MPa)	σ_p	Mean stress under plunger or plate (MPa)
E_{eq}	Equivalent modulus (MPa)	ν	Poisson's ratio (-)
d	Plunger or plate diameter	σ_{vi} and σ_v	Nodal and mean vertical stress (kPa)
k_1 to k_3	Model parameters	σ_{hi} and σ_h	Nodal and mean horizontal stress (kPa)
u_{vi}	Nodal vertical displacement (mm)	ϵ_{vi}	Nodal vertical strain (-)
k_{g1} to k_{g5}	Model parameters	ε_{hm}	Lateral micro-strain of mold at mid-height (-)

1. INTRODUCTION

Stiffness modulus of soils and Unbound Granular Materials (UGM) are main input data in the Mechanistic-Empirical (M-E) design process of flexible pavements adopted in recent decades by many countries. However, its evaluation is a challenge in road engineering. In pavement engineering, many correlations allow for the estimation of granular materials modulus based on the

*Corresponding Author: *abdelaziz.salmi@ensem.ac.ma* (A. Salmi)

Please cite this article as: A. Salmi, L. Bousshine. K. Lahlou, A New Model of Equivalent Modulus Derived from Repeated Load CBR Test, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1321-1330

CBR index employed worldwide. Nevertheless, using elastoplastic simulation, a recent study [1] shows that the CBR index depends on other parameters, such as yield stresses in compression and compressibility index. At times, this modulus independent of Young's modulus. The Repeated Load Triaxial (RLT) test is the most accepted and widely used test in research laboratories to study the resilient and permanent behavior of these materials [2]. However, the configuration and equipment required for this test are technically complex and very expensive. Therefore, it is not part of every laboratory's facilities, especially, those in developing countries [3–5]. To overcome this challenge, the French standard of M-E road pavement design [6] adopts a modulus based on empirical classification of UGM. In the RLT test, the specimen is loaded by a confining pressure, σ_3, and an axial deviator stress, q, defined by Equation (1):

$$q = \sigma_{r1} - \sigma_{r3} \tag{1}$$

To simulate traffic load repetition, q is usually a periodic stress, but σ_3 may be periodic for Variable Confining Pressure test variant (VCP) or not periodic for Constant Confining Pressure test variant (CCP). For the RLT test, the specimen has a diameter of 160 mm or 300 mm and a height of 320 mm or 600 mm, respectively. The RL-CBR test was validated based on resilient modulus derived from a large CCP triaxial test and equivalent modulus derived from the RL-CBR test with strain gauge [2, 7]. A steel mold with a 250 mm internal diameter, 200 mm height, and wall thickness of 14.5 mm was used in the experimental validation program. In this study, we consider the same mold, except that the wall thickness is taken equal to 8 mm. This choice was made in order to reduce the mold mass for practical use in experimental tests.

Experimental characterization of granular materials is a large research field; many empirical and theoretical models were proposed to describe their resilient behavior, as reported in existing reviews of findings in the field [8-10]. Their mechanical behavior depends on multiple parameters [11, 12]. Many tests can be employed to characterize the resilient behavior of unbound granular materials [13, 14]. The use of the uniaxial compressive test is also possible in the case of cohesive or bound granular materials [15,16].When the RLT test is used, the resilient modulus is evaluated by Equations (2) and (3) for VCP variant and CCP variant, respectively, according to the European standards [17].

$$M_r = \frac{\sigma_{r1}^2+\sigma_{r1}\sigma_{r3}-2\sigma_{r3}^2}{\sigma_{r1}\epsilon_{r1}+\sigma_{r1}\epsilon_{r3}-2\sigma_{r3}\epsilon_{r3}} \tag{2}$$

$$M_r = \frac{\sigma_{r1}}{\epsilon_{r1}} \tag{3}$$

In the framework of the RL-CBR test, the stiffness of UGM is evaluated by a resilient modulus designed by Araya [2] as an "equivalent modulus", while Molenaar [4] calls it "effective modulus". Both nomenclatures were justified by the fact that the stress state throughout the specimen is not uniform. Therefore, the resilient modulus may vary throughout it due to stiffness-stress dependency for soils and granular materials. The equivalent or effective modulus is just a bulk measurement of the specimen stiffness, rather than an intrinsic material's characteristic. This approach is also adopted by Albayati et al. [18] to calculate equivalent modulus of asphalte concrete layers. The expression used is inspired by Boussinesq's Equation (4), valid in the case of the elastic isotropic semi-infinite solid loaded by a circular plunger. A detailed review on this equation is given by Timoshenko and Goodier [19]. For the RL-CBR test, a mold with finite dimensions is employed. Araya [2] suggested modifying the Equation (4) into Equation (5). Three model parameters are introduced. They are determined by the LSM (cf. 2. 2) applied on the RL-CBR test numerical data analysis. The LSM is a statistical method widely used to find the best fit for a set of inputs and outputs data points. Recently it's served for Shen and Zhou [20] improved the constitutive modelling of clay in drained and undrained conditions. When the equivalent modulus is calculated by Equation (5), a value of Poisson's ratio, ν, should be specified. It is generally taken equal to 0.35 for soils and UGM in pavement design [6]. This test variant has demonstrated its ability to study the effect of moisture content, dry density and stress level in the experimental investigations of Haghighi et al. [21] using the staged RL-CBR test.

$$E = \frac{f(1-\nu^2)\sigma_p\frac{d}{2}}{u} \tag{4}$$

$$E_{eq} = \frac{k_1(1-\nu^{k_2})\sigma_p\frac{d}{2}}{u^{k_3}} \tag{5}$$

In case of the RL-CBR test with strain gauge, Araya et al. [7] used the nodal vertical displacements as weighting factor to estimate the mean vertical and horizontal stresses using Equations (6) and (7). Four transfer functions, Equations (8), (9), (10) and (11) are used to establish a regression fit between the mean vertical and horizontal stresses, with Poisson's ratio and equivalent modulus on one side and the RL-CBR test outputs on the other side. The use of this approach makes possible the comparison between resilient and equivalent moduli derived from the RLT and the RL-CBR tests, respectively [2,7]. In this model, the average weighted vertical and horizontal stresses are derived from nodal vertical and horizontal stresses using Equations (6) and (7). The nodal displacements through symmetry axis are used as a weighting factor. In this paper, the nodal strains are used as a weighting factor as in Equations (12) and (13) and were shown to be the most appropriate for accurate estimation of equivalent modulus. This change improves the accuracy of the initial Araya et al. model (cf. 3. 2).

$$\sigma_v = \frac{\sum \sigma_{vi} u_{vi}}{\sum u_{vi}}, \quad \sigma_h = \frac{\sum \sigma_{hi} u_{vi}}{\sum u_{vi}} \qquad (1)\ \&\ (2)$$

$$\sigma_v = k_{g1}\sigma_p \qquad (3)$$

$$\nu = k_{g2}(\frac{\varepsilon_{hm}}{\sigma_p}) \qquad (4)$$

$$\sigma_h = k_{g3}\varepsilon_{hm}\exp\,(\mathrm{k_{g4}}/\nu) \qquad (5)$$

$$E_{eq} = \frac{k_{g5}(\sigma_v - 2\nu\sigma_h)}{u} \qquad (6)$$

$$\sigma_v = \frac{\sum \sigma_{vi} \epsilon_{vi}}{\sum \epsilon_{vi}}, \quad \sigma_h = \frac{\sum \sigma_{hi} \epsilon_{vi}}{\sum \epsilon_{vi}} \qquad (7)\ \&\ (8)$$

Characteristics of used materials and research methodology are presented in section 2. After that, the optimal set of parameters is determined based on model estimation accuracy. This set will be used to validate the modified Araya et al. model at the end of this paper.

2. MATERIALS AND METHODS

2. 1 Materials In the present study, a linear elastic behavior is considered for the granular materials. Large quality ranges of UGM were studied by varying the Young's modulus value from 25 to 1000 MPa and Poisson's ratio from 0.15 to 0.45. The plunger penetration, used in the present study, varies from 0.1 mm to 3 mm. It should be noted that this penetration is smaller than 1 mm in previous experimental investigations [2,22].

In the first set, we consider Young's modulus from 25 to 1000 MPa (40 values), Poisson's ratio from 0.15 to 0.45 (4 values), and the plunger penetrations from 0.1 mm to 3 mm (30 values). In total, there were 4,800 simulations of the RL-CBR test, which would take a long time to calculate. As a consequence, we had to reduce the number of simulations by choosing the ones that offer an optimal and accurate estimation of the model parameters. Optimized lengths of Young's moduli and the plunger penetration lists were determined by reducing both lists' lengths. Thus, the model's accuracy is kept at an acceptable level. Tables 1 and 2 summarize parameters sets employed at this stage of the study. The flowchart research methodology is summarized in Figure 1.

2. 2 Methods

2. 2. 1. Finite Element Model of the RL-CBR Test Finite element simulation of the RL-CBR test is performed with CAD software. As geometry, loading, and boundary conditions are in an axisymmetric disposition, a plane axisymmetric approach is used in the modelling process of the RL-CBR set-up. A linear elastic material behavior is assumed for the steel mold with an elastic modulus of 210 GPa and a Poisson's ratio of 0.3 (rather than the 0.2 used by Araya et al. [7]), in addition to the granular material using the various elastic characteristics presented in Tables 1 and 2. As for the standard CBR test, the RL-CBR is a strain-controlled test with a uniform downward displacement of the plunger through material specimen [1,7,23]. The contact between the plunger and the specimen is assumed to be frictionless. The hard pressure-overclosure is adopted for the normal contact property between the mold and the specimen. For the tangential interaction, frictionless contact was chosen. These considerations mean that neither penetration nor friction between both parts will take place when local contact between them is established. The use of frictionless contact assumes that the internal mold surface is very smooth; a demolding oil is used in test preparation to replicate the smoothness. The same normal and tangential interaction models are considered for the plunger-specimen contact. Figure 2 Shows the axisymmetric model used in finite element analysis of the RL-CBR test. CAX8R, 8-node biquadratic axisymmetric quadrilateral reduced integration elements type is used for specimen mesh, because it offers an accurate FE analysis of a 3D problem using plane axisymmetric model [2].

TABLE 1. Sets of parameters tested to reduce the Young's modulus list length

Set	*E* step (MPa)	*ν* step (-)	*u* step (mm)	Number of simulations
Set 1	25	0.1	0.1	4800
Set 2	50	0.1	0.1	2400
Set 3	100	0.1	0.1	1200
Set 4	125	0.1	0.1	960
Set 5	200	0.1	0.1	600
Set 6	250	0.1	0.1	480
Set 7	500	0.1	0.1	240

TABLE 2. Sets of parameters tested to reduce the list of plunger penetrations u

Set	*E* step (MPa)	ν step (-)	*u* step (mm)	Number of simulations
Set 5	200	0.1	0.1	600
Set 8	200	0.1	0.2	300
Set 9	200	0.1	0.3	200
Set 10	200	0.1	0.5	120
Set 11	200	0.1	0.6	100
Set 12	200	0.1	1	60

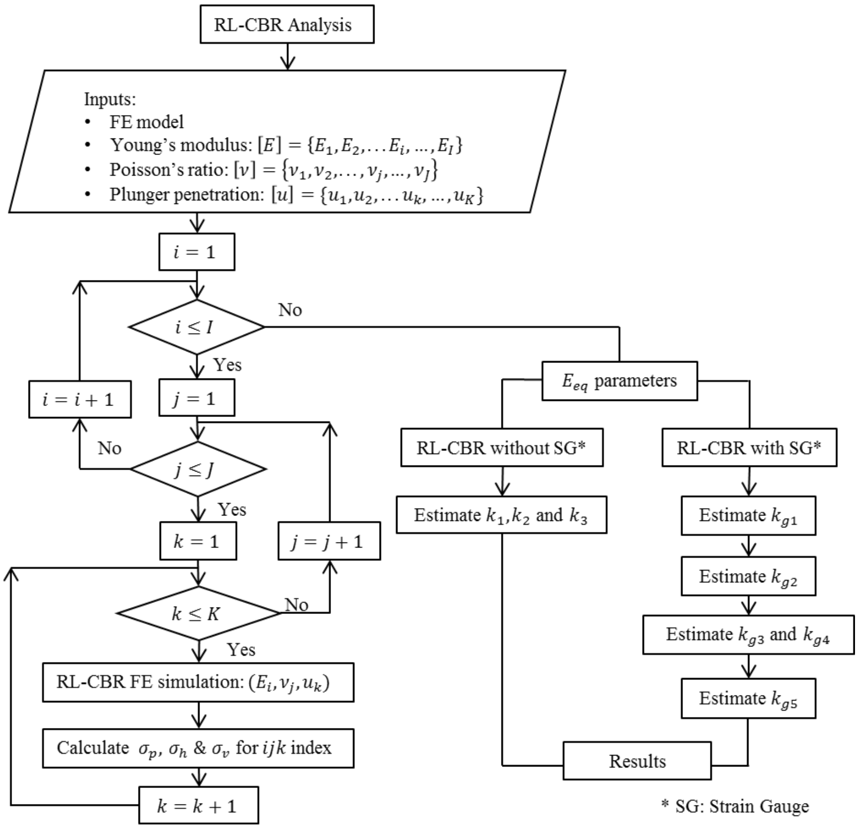

RL-CBR Analysis
Inputs:
• FE model
• Young's modulus: $[E] = \{E_1, E_2, \ldots E_i, \ldots, E_I\}$
• Poisson's ratio: $[\nu] = \{\nu_1, \nu_2, \ldots, \nu_j, \ldots, \nu_J\}$
• Plunger penetration: $[u] = \{u_1, u_2, \ldots u_k, \ldots, u_K\}$
$i = 1$
$i \le I$
No
Yes
$i = i + 1$
$j = 1$
No
$j \le J$
Yes
$k = 1$
$j = j + 1$
$k \le K$
No
Yes
RL-CBR FE simulation: (E_i, ν_j, u_k)
Calculate σ_p, σ_h & σ_v for ijk index
$k = k + 1$
E_{eq} parameters
RL-CBR without SG*
RL-CBR with SG*
Estimate k_1,k_2 and k_3
Estimate k_{g1}
Estimate k_{g2}
Estimate k_{g3} and k_{g4}
Estimate k_{g5}
Results
* SG: Strain Gauge

Figure 1. Flowchart of research methodology

A sensitive study of mesh size is undertaken to validate the model. A uniform mesh with a size from 0.25 mm to 8 mm for 8-node biquadratic axisymmetric quadrilateral reduced integration elements type is adopted. The granular material has a Young's modulus of 1000 MPa and a Poisson's ratio of 0.25. The plunger penetration was chosen equal to u=0.2 mm. Figure 3 shows the variation of the obtained mean stress under the plunger vs. elements number in the model (in logarithmic scale). By reducing the mesh size from 8 mm to 0.25 mm, the mean stress was decreased from 3.65 MPa to 3.59 MPa with excessive start-slope that decreases to be very soft for fine mesh. The adopted mesh for this study illustrated in Figure 2 gives a mean stress under the plunger of 3.66 MPa. This value is very close to the fine mesh value, 3.95 MPa. The mesh is refined under and near the plunger where there is the stress concentration phenomenon and is made increasingly coarse away from this area. The model counts 780 elements instead of 442000 elements for the fine mesh model.

2. 2. 1. Least-Squares Method After conducting a set of RL-CBR numerical analyses (cf.2. 1), the data was organized as illustrated in Table 3. In the FE analysis of the RL-CBR test, Young's modulus E, Poisson's ratio ν

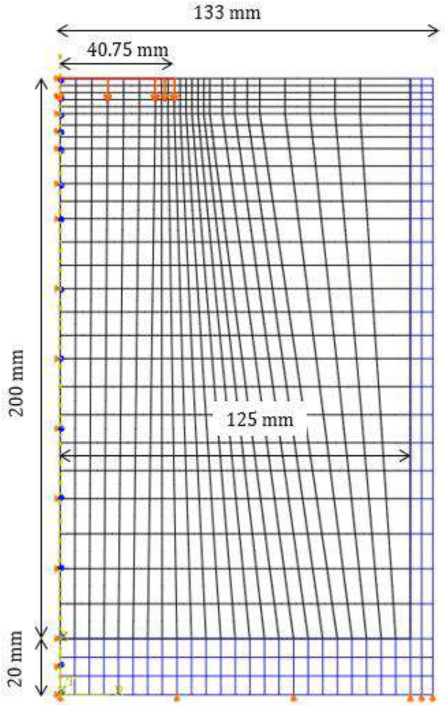


Figure 2. Axisymmetric finite element model of RL-CBR test with 8 mm thick mold

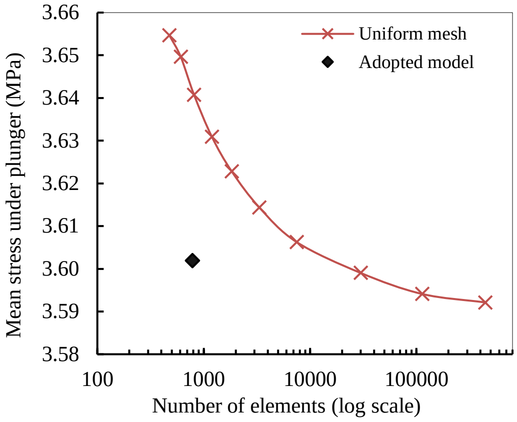


Figure 3. Mean stress under the plunger vs. number of elements (E = 1000 MPa, ν = 0.25, u =0.2 mm)

TABLE 3. Table of variables for a set of analysis

Data line	Response E (MPa)	Explanatory variables		
		ν(-)	u (mm)	σ_p (MPa)
1	E_1	ν_1	u_1	σ_{p1}
2	E_2	ν_2	u_2	σ_{p2}
…	…	…	…	…
i	E_i	ν_i	u_i	σ_{pi}
…	…	…	…	…
n	E_n	ν_n	u_n	σ_{pn}

and plunger penetration u are the input parameters, and the mean stress under the plunger σ_p is the output parameter. In the regression analysis, with respect to Equation (5), elastic modulus is considered as the response variable and other parameters (ν, u, σ_p) as explanatory variables. However, the purpose is to derive an equivalent modulus expression to be used in the experimental characterization of granular materials usingthe RL-CBR test where the stiffness is researched. Resilient plunger penetration and mean resilient stress are measured during the test.

For a set of analyses, the main goal is to find the three model's parameters: k_1, k_2, and k_3 that allow the estimation of the response variable given explanatory variables. These parameters should give an equivalent modulus as close as possible to the initial elastic modulus for each variable's combination. The non-linear multivariate regression problem is solved by the LSM. The LSM consists of researching parameters that minimizes the Squared Deviations Sum (SDS) defined by Equation (14). The General Reduced Gradient (GRG) algorithm is applied to solve this non-linear optimization problem, which is summarized in Equation (15).

$$SDS(k_1;k_2;k_3) = \sum_{i=1}^{i=n}(E_i - E_{eqi})^2 \tag{14}$$

$$\begin{array}{l} \text{Minimize} \quad SDS(k_1;\ k_2;\ k_3) \\ \text{Without technological constraints} \end{array} \tag{15}$$

3. RESULTS AND DISCUSSION

3. 1. Optimal Parameters Set

Parameters of Equation (5) were determined for each data set presented in Tables 1 and 2. The values of these parameters resulting from simulations used to reduce the length of the Young's modulus list are summarized in Table 4 with the determination coefficient R^2 for each set. Table 4 shows that the first parameter was decreased by increasing the step between consecutive values of Young's modulus from 1.653 for 25 MPa step set to 1.641 for 500 MPa step set. The second parameter was increased from 0.978 to 0.998, while the third parameter remained invariant for all seven sets. With respect to the determination coefficient, all correlations seem to be good except that of the 7th set, where the determination coefficient decreased to 0.990. However, R2 values presented in Table 4 cannot be used to compare the accuracy of the derived solutions because of the significant differences between the sample size of each set. To do this comparison, SDS's values are calculated for each solution with respect to the finite elements simulations' results of the largest sample (i.e., set 1). Figure 4 presents the variation of the SDS's RD when parameters derived from ith set are used (SDS_i) with respect to the SDS obtained when parameters derived from the 1st set are used (SDS_1). RD for ith set parameters is defined by Equation (16). Figure 4 shows that RD increases by increasing the Young's modulus step. For parameters obtained from set 7, RD is about 13 ‰, which is four times the RD obtained when parameters derived from set 6 are used and seven times the RD resulting from the use of set 5 parameters. By comparing the relative deviations related to the use of each set with respect to set 1, results show that set 5 offers an accurate estimation of the model parameters. This means that the length of Young's modulus list is divided by 8, keeping the accuracy of the estimation at the same level. This reduction may optimize the analysis time and facilitates the estimation of model parameters for other test configurations.

$$RD(set_i; set_1) = 1000\frac{(SDS_i - SDS_1)}{SDS_1} \tag{16}$$

The same approach is used to reduce the length of plunger penetrations list. The starting set is set 5 (chosen above). For subsequent sets this penetration step is increased from 0.1 mm to 1 mm. Table 2 presents adopted plunger penetration steps for each set. Table 5 summarizes obtained parameters in this part of the study. It is noted that the values of the parameters do not change much, and for the last four sets, they do not change at all.

These results indicate that reducing the length of the plunger penetration list does not significantly influence the values of the parameters of the model. This can be explained by the contact conditions between mold and specimen that make the problem linear. Then, the ratio of mean stress σ_{pi} by plunger penetration u_i was seen to be constant for a given specimen for each Poisson's ratio. Figure 5 shows the constancy of RD evaluated when obtained parameters from sets 5 and 8 to 12 are used. Only the parameters resulting from set 8 gave less accurate estimations. For the other cases, the same level of accuracy is maintained. To choose the appropriate set that makes the best compromise between the model accuracy and the calculation time, the non-linear character of the optimization problem had to be taken into account. Accordingly, the lengths of lists for all parameters must be at least 3, which is the case for set 12. To gain accuracy for other cases of simulation considering other analysis conditions, set 11 is chosen. In this set, plunger penetration takes 5 values, from 0.6 mm to 3 mm.

TABLE 4. Model parameters for sets used to reduce the list length of Young's modulus

Set	$k_1(-)$	$k_2(-)$	$k_3(-)$	R^2
Set 1	1.653	0.978	1.001	0.996
Set 2	1.652	0.979	1.001	0.996
Set 3	1.651	0.982	1.001	0.995
Set 4	1.650	0.983	1.001	0.995
Set 5	1.648	0.986	1.001	0.995
Set 6	1.647	0.988	1.001	0.995
Set 7	1.641	0.998	1.001	0.990

TABLE 5. Model parameters for sets used to reduce the list length of plunger penetrations

Set	$k_1(-)$	$k_2(-)$	$k_3(-)$	R^2
Set 5	1.648	0.986	1.001	0.995
Set 8	1.648	0.987	1.001	0.995
Set 9	1.647	0.987	1.000	0.995
Set 10	1.647	0.987	1.000	0.995
Set 11	1.647	0.987	1.000	0.995
Set 12	1.647	0.987	1.000	0.995

Moreover, Poisson's ratio takes four values from 0.15 to 0.45 in 0.1 steps. To reduce the number of values considered here, we removed the first value, and we saw the accuracy of the solution of the problem (15). The obtained solution is: $k_1 = 2.062$, $k_2 = 0.698$ and $k_3 = 1.000$ with $R^2 = 0.998$. When these values are used and compared to set 1 data, the SDS is 3 times higher than the minimal SDS of set 1. After removing the second value (0.25), the SDS is 27 times the minimal SDS. These tests show that the reduction of Poisson's ratio list length affects the accuracy of the model, so the initial Poisson's ratio list is maintained as in the set 11. In practical use of Equation (5), the accuracy of the model must be considered in estimating equivalent modulus using Relative Error (RE) defined by Equation (17) and given in Table 6 for various combinations of Poisson's ratio and equivalent modulus.

$$RE\ (E; E_{eq}) = \frac{abs(E-E_{eq})}{E} * 100 \qquad (17)$$

To compare the accuracy of this solution with previous studies, Table 7 summarizes the values of parameters of present and previous studies and the ratio of SDS per the reference SDS obtained for set 1 data when parameters for the set 11 are used. After this comparison, the use of Araya's solution [2] induced a model half as accurate as the solution of the current study. It is noted that for this study, a particular model for the specimen-mold contact

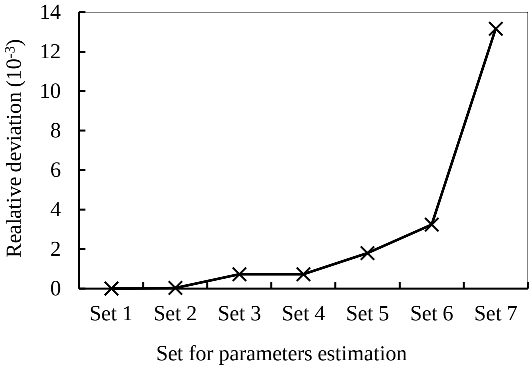


Figure 4. Comparison of obtained model parameters' accuracy to reduce Young's modulus list length

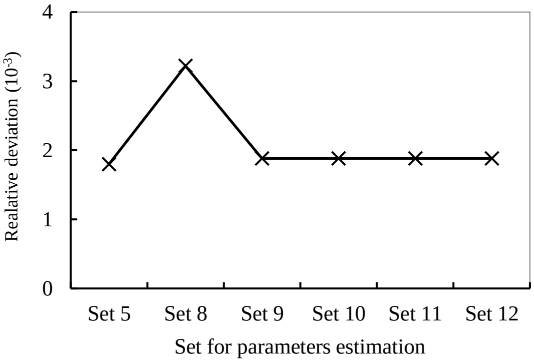


Figure 5. Comparison of accuracy of model parameters used to reduce the list length of plunger penetrations

TABLE 6. Relative error of equivalent modulus derived from RL-CBR test without strain gauge

ν(-)E_{eq} (MPa)	0.15	0.25	0.35	0.45
200	2.1	3.3	5.2	0.8
400	3.1	4.5	1.0	1.6
600	3.8	3.0	1.8	4.0
800	5.3	1.9	3.8	6,4
1000	2.0	3.6	5.8	2.3

TABLE 7. Comparison with previous studies for the RL-CB without strain gauge case

Study	Tangential contact	Normal contact	k_1	k_2	k_3	$^{SDS}/_{SDS11}$
Present study	Frictionless	hard	1.647	0.987	1.000	1
Salmi et al. [24]	Frictionless	exponential	1.377	1.552	1.056	6
Araya et al. [2]	intermediate friction	exponential	1.513	1.104	1.012	2

TABLE 8. Estimated parameters for present and previous studies

Parameter	Present study	Araya et al. [7]
k_{g1} (-)	0.478	0.368
k_{g2} (kPa^{-1})	62.027	120.927
k_{g3} (kPa)	31.552	43.898
k_{g4} (-)	0.095	0.072
k_{g5} (mm)	0.139	0.144

is considered: hard for normal contact and frictionless for the tangential one.

For equivalent modulus derived from the RL-CBR with strain gauge using Equations (8) to (11), model parameters are estimated based on the set 11 data (results shown in Table 8). For comparison purposes, the values of the parameters found by Araya et al. [7] are also summarized in the same table. Regression fit of the four transfer functions to the simulation results shows a good correlation with R^2 = 0.999 for average vertical and horizontal stress, 0.975 for Poisson's ratio and 0.952 for the equivalent modulus. The difference between the values of the parameters in this study and that of Araya et al. [7] can be attributed to the mold thickness (8 mm instead of 14.5 mm), tangential-normal contact model (i.e. frictionless-hard instead of intermediate friction-exponential), and the accuracy of the finite element model. Particularly, the thickness effect is notable for the estimation of the Poisson's ratio: the ratio of parameters of the model for Araya et al. [7] $k_{g2}(\text{Ara})$ per that of the present study $k_{g2}(\text{ps})$ is approximately equal to the ratio of thicknesses in both studies: $t(\text{Ara})$ and $t(\text{ps})$, respectively, as expressed by Equations (18) and (19).

$$\frac{k_{g2}\,(\text{Ara})}{k_{g2}(\text{ps})} = 1.942, \frac{t\,(\text{Ara})}{t\,(\text{ps})} = 1.813 \qquad (18)\&(19)$$

3. 2 A New Model for Equivalent Modulus

Even if the parameters of Araya et al. model adopted for the RL-CBR test with strain gauge are derived with good determination coefficients (R^2), it is found that this model sometimes makes inaccurate estimations of equivalent modulus, especially for $\nu = 0.45$, where the relative error was between 20% and 25% for many simulations.

The suggested model consists of keeping the same transfer functions, Equations (8)-(11), but nodal displacement is replaced by nodal strain in the calculation of vertical and horizontal weighted average stresses. Equations (6) and (7) are replaced by Equations (12) and (13) in the new model, named Modified Araya et al. model for the RL-CBR test with strain gauge. Table 9 summarizes estimated parameters for the new model with determination coefficient R^2 for each parameter. It's noted that the second parameter, $k_{g2,}$ is the same for both models, since the estimation of Poisson's ratio does not depend on the mean vertical and horizontal stresses expressions as in Equation (9).

The plot of equivalent moduli estimated by Araya et al. and modified Araya et al. models versus exact elastic modus used in finite element analyses of the RL-CBR test is shown in Figure 6. For all analyses, the predictions of the developed model are more accurate than those estimated using Araya et al. model. Maximum relative error of estimated equivalent modulus for various

TABLE 9. Modified Araya et al. model parameters

Parameter	Value	R^2
k_{g1} (-)	0.431	1.000
k_{g2} (kPa-1)	62.027	0.975
k_{g3} (kPa)	18.848	0.996
k_{g4} (-)	0.140	
k_{g5} (mm)	0.140	0.999

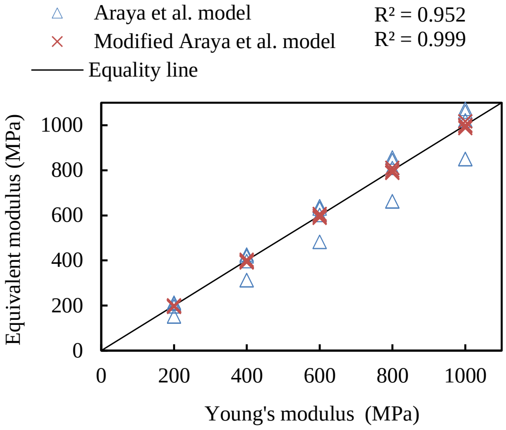


Figure 6. Predicted equivalent modulus vs. Young's modulus used in FE analyses

Poisson's ratios are shown in Figure 7 for both models. The obtained relative error for Araya et al. model is usually higher than 5%, except for $\nu = 0.35$, where it is lower than 2%. Meanwhile, it does not reach 3% for modified Araya et al.'s, even for $\nu = 0.45$, where the mean relative error for Araya et al. model was equal to 20% and the maximum one was equal to 25%. It's noted that for $\nu = 0.35$, both models' estimations are in the same accuracy level with a RE less than 2%. Meanwhile, Poisson's ratio is estimated using the same expression (4) for both models. The plot of mean estimated Poisson's ratio vs. real Poisson's ratio for various Young's moduli is shown in Figure 8.

The Poisson's ratio is overestimated for high and low values (i.e., 0.15 and 0.45), while it is underestimated for intermediate values (i.e., 0.25 and 0.35) as shown in Figure 8. Additionally, the relative error of Poisson's ratio higher than 0.25 is below 10%. Accordingly, the predicted Poisson's ratio for soils and unbound granular materials around $\nu = 0.35$ can be used in road pavement and geotechnical engineering. The modified model has demonstrated a good accuracy compared to its initial version in the equivalent modulus estimation for all Poisson's ratio. For experimental investigations, the intrinsic relative error of the model should be added to that of the equipment used. Then, the final results can be analyzed carefully. Table 10 shows relative error, for various combinations of Poisson's ratio and equivalent modulus to be considered in experimental investigations. For cases with values different than the ones given, linear interpolation can be utilized to estimate the corresponding RE.

The Poisson's ratio is overestimated for high and low values (i.e., 0.15 and 0.45), while it is underestimated for intermediate values (i.e., 0.25 and 0.35) as shown in Figure 8. Additionally, the relative error of Poisson's ratio higher than 0.25 is lower than 10%. Accordingly, the predicted Poisson's ratio for soils and unbound granular materials around $\nu = 0.35$ can be used in road pavement and geotechnical engineering. The modified model has demonstrated a good accuracy compared to its initial version in the equivalent modulus estimation for all Poisson's ratio. For experimental investigations, the model's intrinsic relative error should be added to that of the equipment used. Then, the final results can be analyzed carefully. Table 10 shows relative error, for various combinations of Poisson's ratio and equivalent modulus to be considered in experimental investigations. For cases with values different than the ones given, linear interpolation can be utilized to estimate the corresponding RE.

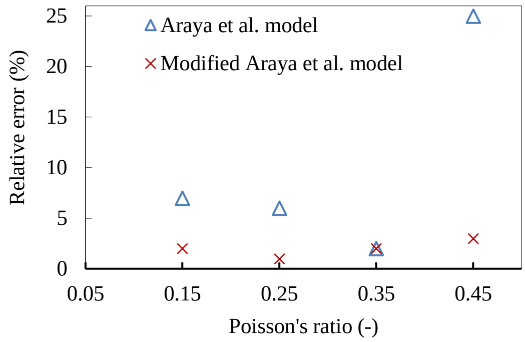


Figure 7. Equivalent modulus maximum relative error for Araya et al. and modified Araya et al. models

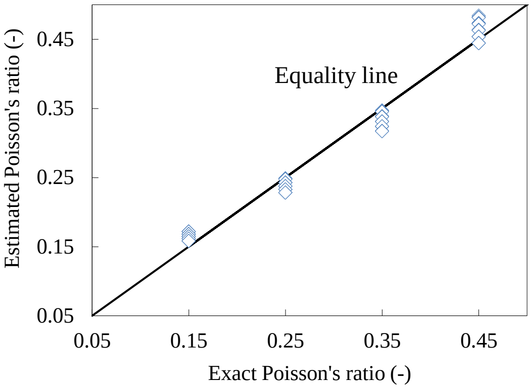


Figure 8. Estimated Poisson's ratio vs. exact Poisson's ratio

TABLE 10. Relative error (%) of Modified Araya et al. model

ν (-) E_{eq} (MPa)	0.15	0.25	0.35	0.45
200	2.1	0.7	1.1	1.2
400	1.1	0.8	1.9	1.3
600	0.4	2.5	1.4	0.1
800	3.2	1.6	0.2	1.7
1000	1.8	0.4	1.4	0.5

4. CONCLUSIONS AND OUTLOOKS

The paper presented a finite element simulation of the RL-CBR test with an 8 mm thick extra-large mold. The contact between the granular specimen and steel mold was assumed to be frictionless. In case of the RL-CBR test without strain gauge, the model parameters used to calculate equivalent modulus are estimated. For the repeated load CBR test with strain gauge parameters of Araya et al. and modified Araya et al. model are estimated. The comparison between the estimations of both models showed that the accuracy was remarkably increased through this modification, especially for materials with high and low Poisson's ratio. For material with Poisson's ratio of 0.35, the accuracy of estimation is kept in the same level.

This research will continue, in one side, by investigating the effect of the contact type between the specimen and mold on parameters of the model and, in another side, by an experimental validation of the derived equivalent modulus. This validation will be based on the resilient modulus derived from the RLT laboratory test and reaction modulus derived from plate and Westergaard in-situ tests.

Equivalent modulus may be used as a comparative tool for UGM quality ranging. But its use in the M-E design of pavements requires other laboratory and full-scale investigations.

5. ACKNOWLEDGMENT

- Abdelaziz Salmi is grateful to Hassan II academy of sciences and technology for its financial support of this research in the framework of the scholarships program for winners of CGST competition.
- The authors thank Y. Waddad for his support on the use of the CAD software and M. Farah, M. Esskely, H. El Fatihi, R. Ibourk El Idrissi, E. El Bahjaoui and F. Kassou for the review of the manuscript.

6. REFERENCES

1. Mendoza C., Caicedo B. "Elastoplastic framework of relationships between CBR and Young's modulus for granular material". ***Road Materials and Pavement Design***. Vol. 19, No. 8, (2017). DOI:10.1080/14680629.2017.13475 17.

2. Araya A.A. "Characterization of unbound granular materials for pavements", PhD thesis, Delft University of Technology, (2011).

3. Araya A.A., Huurman M., Molenaar A.A.A. "Integrating traditional characterization techniques in mechanistic pavement design approaches". In 1st Congress of Transportation and Development Institute, Chicago, Illinois,United States, March 13-16 2011, American Society of Civil Engineers, (2011), 596-606, DOI: 10.1061/41167 (398)57

4. Molenaar A.A.A. "Repeated load cbr testing, a simple but effective tool for the characterization of fine soils and unbound materials". In Transportation Research Board 87th Annual Meeting, January 13-17 2008, Washington, United States, (2008).

5. Molenaar A.A.A. "Characterization of Some Tropical Soils for Road Pavements". ***Transportation Research Record: Journal of the Transportation Research Board***, 2007, Vol. 1989-2, No. 1, 186-193. DOI: 10.3141/1989-63

6. AFNOR. NF P 98-086- "Dimensionnement structurel des chaussées routières- application aux chaussées neuves". 2011.

7. Araya A.A, Huurman M, Molenaar A.A.A, Houben LJM. "Investigation of the resilient behavior of granular base materials with simple test apparatus", ***Materials and Structures***, (2012), Vol. 45, No. 5, 695–705. DOI:10.1617/ s11527-011-9790-1

8. Han Z., Vanapalli S.K. "State of the art: prediction of resilient modulus of unsaturated subgrade soils". ***International Journal of Geomechanics***, (2016), Vol. 16, No.4, 1–15. DOI:10.1061/(ASCE)GM.1943-5622.0000631.

9. Fredrick L., Ulf I., Andrew D. "State of the art. I: resilient response of unbound aggregates", ***Journal of Transportation Engineering***, (2000), Vol. 126, No. 1, 66–75. DOI: 10.1061/(ASCE)0733-947X(2000)126:1(66)

10. Saada A., Townsend F., "State of the art: laboratory strength testing of soils", in Laboratory Shear Strength of Soil, ed. R. Yong and F. Townsend (West Conshohocken, PA: ASTM International), (1981), 7-77. DOI:10.1520/STP28744S

11. Minassian G.H. "Behavior of granular materials under cyclic and repeated loading", Ph.D thesis, University of Alaska Fairbanks, United States, (2003).

12. Gidel G. "Comportement et valorisation des graves non traitées calcaires utilisées pour les assises des chaussées souples", Thèse de doctorat, Université Bordeaux I, France, (2001).

13. Semmelink C.J. "The use of DRTT K-mould to determine the elastic and shear properties of pavement materials", Report No. 89/149, Department of Transport, South Africa, (1991).

14. Edwards P., Thom N., Fleming P.R., Williams J. "Testing of unbound materials in the nottingham asphalt tester springbox", ***Transportation Research Record: Journal of the Transportation Research Board***, (2005), Vol. 1913, No. 1, 32-40. DOI: 10.1177/0361198105191300104

15. Jalili M, Ghasemi M.R., Pifloush A.R., "Stiffness and strength of granular soils improved by biological treatment bacteria microbial cements". ***Emerging Science Journal***, (2018), Vol. 2, No. 4, 219-227, DOI: 10.28991/esj-2018-01146

16. Ogundipe O.M., Adekanmi J.S., Akinkurolere O.O., Ale P.O., "Effect of compactive efforts on strength of laterites stabilized with sawdust ash". ***Civil Engineering Journal***, (2019), Vol. 5, No. 11, 2502–25014, DOI: 10.28991/cej-2019-03091428

17. CEN. "EN 13286-7- Unbound and hydraulically bound mixtures - Part 7 : Cyclic load triaxial test for unbound mixtures", European committee for standardization, (2004).

18. Albayati A.H, Al-Mosawe H, Fadhil A.T., Allawi A.A. "Equivalent Modulus of Asphalt Concrete Layers", Vol. 4, No. 10, ***Civil Engineering Journal***, (2018), 2264-2274. DOI:10.28991/cej-03091156

19. Timoshenko S, Goodier J.N. "Theory of elasticity", ed. McGraw-Hill, (1951), 532.

20 J. Shen et X. Zhou, "Least Squares Support Vector Machine for Constitutive Modeling of Clay", ***International Journal of Engineering Transactions B: Applications***, Vol. 28, No. 11, (2015), 1571-1578.

21. Haghighi H., Arulrajah A., Mohammadinia A., Horpibulsuk S. "A new approach for determining resilient moduli of marginal pavement base materials using the staged repeated load CBR test

method". ***Road Materials and Pavement Design***, Vol. 19, No. 8 (2017), 1848-1867. DOI: 10.1080/ 14680629.2017.1352532

22. Abid A.N., Salih A.O., Nawaf E.A. "The influence of fines content on the mechanical properties of aggregate subbase course material for highway construction using repeated load CBR test". ***Al-Nahrain Journal for Engineering Sciences***, Vol. 20, No. 3, (2017), 615-24. URL: nahje.com/index.php/main/article/view/252.

23. Gu J. "Computational modeling of geogrid reinforced soil foundation and geogrid reinforced base in flexible pavement", PhD thesis, Louisiana State University, (2011).

24 Salmi A., Bousshine L., Lahlou K., "Two unbound granular materials stiffness analysis with staged repeated load CBR test," in 14th Congress of Mechanics, Rabat, Morocco, 16-19 Apr. 2019, DOI: 10.1051/matecconf/201928606003.

Persian Abstract

چکیده

این پژوهش یک مدل جدید از مدول معادل حاصل از آزمایش بار تکراری CBR (RL-CBR) بدون استفاده از کرنش‌سنج را ارائه می‌دهد. این مدل یک نسخه‌ی به‌روز شده از مدل آریا و همکاران (۲۰۱۱) است. مدل، به‌روزرسانی شده شامل استفاده از کرنش عمودی به عنوان ضریب وزنه به جای جابه‌جایی عمودی در محاسبه‌ی میانگین تنش‌های عمودی و افقی است. دقت مدول معادل با کاهش خطای نسبی از ۲۵٪ به ۳٪ بهبود یافت. قالب فوق‌العاده بزرگی که توسط آریا و همکاران پذیرفته شده است، به جای ضخامت ١٤.٥ میلی‌متر از ضخامت ۸ میلی متر استفاده می‌شود. در تحقیقات تجربی، مدول معادل ممکن است از داده‌های تجربی و پارامترهای مدل تخمین زده شده توسط شبیه‌سازی المان محدود (EF) محاسبه شود. هنگام استفاده از آزمون RL-CBR پنج، و در صورت عدم استفاده از کرنش‌سنج، سه پارامتر مدل وجود دارد. پارامترهای مدل در دو مرحله مشخص می‌شوند. ابتدا، شبیه‌سازی FE از آزمایش RL-CBR با استفاده از شرایط بارگذاری مختلف (به عنوان مثال، نفوذ سمبه) و دامنه‌های مختلف کیفیت مواد گرانول نچسبیده (UGM) انجام می‌شود. در مرحله‌ی دوم، رگرسیون چندبعدی غیرخطی انجام می‌شود که برابر با مدول یانگ می‌باشد. تأثیر ورودی‌های تحلیل FE برای یافتن مجموعه‌های ورودی بهینه که بهترین سازش بین دقت مدل و صرف زمان محاسبه را دارند، بررسی شده است. محاسبه‌ی پارامترهای مدل بر اساس داده‌های مجموعه بهینه انجام می‌شود. نتایج مدل جدید و نتایج مدل آریا و همکاران با هم مقایسه شده و دقت بهبودیافته‌ی مدل توسعه‌یافته را نشان داده است.

International Journal of Engineering

Journal Homepage: www.ije.ir

Study on Iraqi Bauxite Ceramic Reinforced Aluminum Metal Matrix Composite Synthesized by Stir Casting

M. A. Aswad*, S. H. Awad, A. H. Kaayem

Department of Ceramics Engineering and Building Materials, Faculty of Materials Engineering, University of Babylon, Babylon, Iraq

PAPER INFO

Paper history:
Received 14 March 2020
Received in revised form 30 March 2020
Accepted 20 May 2020

Keywords:
Alumium-based Composite Material
Iraqi Bauxite
Stir Casting
Crack Opening Displacement
Brinell Hardness
Fracture Toughness
Mechanical Properties

ABSTRACT

For the past decades researchers are showing immense interest to investigate the natural advantage of preparation of composites from minerals such as bauxite particles, and proved their effectiveness as cost effective reinforcing agents in fabrication of high performance composites. This study, is a new attempt in using the Iraqi natural bauxite powder with different proportions (2, 4 and 6 wt%) in preparation of aluminum metal matrix composites (AMMCs) using stir casting and Mg additives. In experimental work, the bauxite stones were crashed and milled, then the powder was fired at 1400 °C. The powders were characterized using particle size, XRD and XRF analysis. The AMMCs casts were machined, polished, preheated, and their properties were characterized using hardness measurements, microstructural observations, and calculation of their Young's modulus, Poisson's ratio and fracture toughness. Also, their fracture toughness were evaluated by means of crack mouth opening displacement (CMOD) measurements from extensometer recordings. The results proved the successful production of AMMCs with improved fracture toughness, hardness and elastic modulus properties using Mg and Iraqi fired bauxite added at 2 and 4 wt% by stir casting. Moreover, results from CMOD measurements showed the effect of addition bauxite particles at 2 and 4 wt% in increasing "maximum load at failure" and "critical CMOD at critical load" of the matrix materials to about " 25 and 44%" , and " 32 and 47%", respectively. Also, at these ratios, the calculated fracture toughness of the matrix materials by means of K_{IC}, and young modulus showed improvement at about "22 and 69%", and "8 and 12%", respectively. Addition of bauxite at 6% could not give the required improvement in the fracture toughness despite its effects in recording the highest improvements in hardness (57%) and elastic modulus (22%) due to the brittle behavior of AMMCs at this ratio.

***doi**: 10.5829/ije.2020.33.07a.20*

NOMENCLATURE

COD	Crack Opening Displacement (mm)	ν	Poisson's Ratio
AMMCs	Aluminum Metal Matrix Composites	*E*	Young' Modulus (GPa)
XRD	X-ray Diffraction	PSA	Particle Size Analysis
XRF	X-ray Fluorescence	Al	Aluminum
Mg	Magnesium	CT	Compact Tension
HB	Brinell Hardness (MPa)	SEM	Scanning Electron Microscope
KIC	Fracture Toughness (GPa.m$^{1/2}$)	CMOD	Crack Mouth Opening Displacement (mm)

1. INTRODUCTION

For the past decades ceramic-reinforced aluminum metal matrix composites AMMCs materials have proved their dominance in many applications such as field of automobile and marine due to their superior properties such as high strength-to-weight ratio, corrosion resistance and superior tribological properties [1-3].

Also, particles reinforced aluminum metal matrix composites (AMMCs) have received increasingly importance in advanced applications in comparison to fiber reinforced composites due to their superior

*Corresponding Author Email: *mohsin.aswad@gmail.com* (M. A. Aswad)

Please cite this article as: M. A. Aswad, S. H. Awad, A. H. Kaayem, Study on Iraqi Bauxite Ceramic Reinforced Aluminum Metal Matrix Composite Synthesized by Stir Casting, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1331-1339

properties such as low cost and relatively isotropic properties [4].

The selection of suitable matrix and reinforcement materials are being the main challenge in the fabrication of AMMCs [1]. Currently, researchers are showing immense interest to strengthening of AMMCs with ceramic materials such as silicon carbide, boron carbide and zirconium oxide [3, 5]. Many studies have investigated the natural advantage of preparation of composites from minerals such as bauxite particles, and proved their effectiveness as cost effective reinforcing agents in fabrication of high performance composites [6, 7].

Iraqi bauxite, is a heterogeneous material found naturally in the form of large blocks in the western desert and in Al-Anbar in Iraq as the main sources for alumina and aluminum production. It consists of aluminum hydroxide minerals, and mixtures of aluminosilicate, iron oxide, silica, titania, and other impurities [8-11].

In view of aforementioned issues, the present work is mainly concentrated on the development of the new trends in use of Iraqi natural bauxite particles in preparation of AMMCs by stir casting process, and study the effect of these particles on the mechanical properties (fracture toughness by means of crack mouth opening displacement, CMOD measurements from extensometer recordings, hardness and Young's modulus).

Stir casting has proved its superiority by means of technical and economical consideration in manufacturing of AMMCs due to its advantages of flexibility, and applicability to large quantity production, and large size components [2].

2. EXPERIMENTAL PROCEDURES

2. 1. Preparation and Charaterization of Starting Materials

2. 1. 1. Bauxite Particles Iraqi bauxite rocks were manually kibbled using mortar to get the quasi finished powder. Then the powder was washed, dried and milled for 8 hours by using ball mill at speed 350 rpm. Then, the powder was fired at 1400 °C to avoid moisture and some impurities. XRD was used to characterize the Iraqi bauxite ceramic powder. The particle size distribution of bauxite powder was determined using particle size analysis (PSA) process. Chemical Composition of the oxides contents and other components in fired and unfired bauxite powder were determined using X-ray Fluorescence, XRF.

2. 1. 2. Matrix Material Al wires (99.7% purity) were used in this study in casting of matrix material. The Al wires were cut into Ø 3mm and (10-15) mm pieces, cleaned and washed, then dried in order to be ready for preparation of AMMCs samples by stir casting. Table 1 shows the results of X-ray for the chemical composition of the Al wire used as matrix material. Mg powder was used as additives in AMMCs.

2. 2. Preparation of AMMCs by Stir Casting

2. 2. 1. Stir Casting Process Table 2 shows the ratios of matrix, reinforcement, and Mg additives used in preparation of AMMCs specimens in the present study.

For fabrication of AMMC composites through stir casting route**,** Al wires were accommodate into the graphite crucible in an electrical furnace type (20122 MILANO). The temperature of furnace was raised slowly above liquidus temperature to melt the Al wires, and it then was maintained at 700°C. Mg additives (Ø 56.05) covered with foil were added to the melt to improve the wettability between matrix and reinforcement; then, the slag was removed. The weighted reinforcements of bauxite were covered with aluminum foil and pressed carefully in order to realize particles from air, then preheated at 300°C for 15 minutes to avoid any moisture contains. The temperature was slowly reduced to below the liquidus temperature of the matrix material. The semisolid molten was mixed for 7 minutes with stirrer blade rotated at a speed of 870 rpm, in order to obtain uniform ceramic particles distribution; thereby, improve wetting and permeability of the reinforcements in the liquid matrix. After that, the temperature was raised slowly again above liquidus temperature (850°C) to increase fluidity of molten metal. Finally, the melt was poured into a preheated casting mould. After casting, the casts were processed to prepare the specimens for tests. Figure 1 shows the flowchart of the experimental approach.

2. 2. 2. Casts Machining and Treating In this process, samples for the compact tension test were prepared according to the American Standard E399 [12], the casts were machined to the final dimensions of

TABLE 1. Chemical composition of the Al wires

%Al	%Si	%Fe	%Cu	%Mn	%Mg
0.70	0.06	0.12	0.01	0.01	0.02
%Zn	%Ti	%B	%V	%Cr	%Others
0.03	0.01	0.005	0.01	0.01	0.015

TABLE 2. Chemical Composition of AMMCs Specimens

Sample code	(Wt%) Al	(wt%) Mg	(wt%) Bauxite
S0	98	2	----
S1	96	2	2
S2	94	2	4
S3	92	2	6

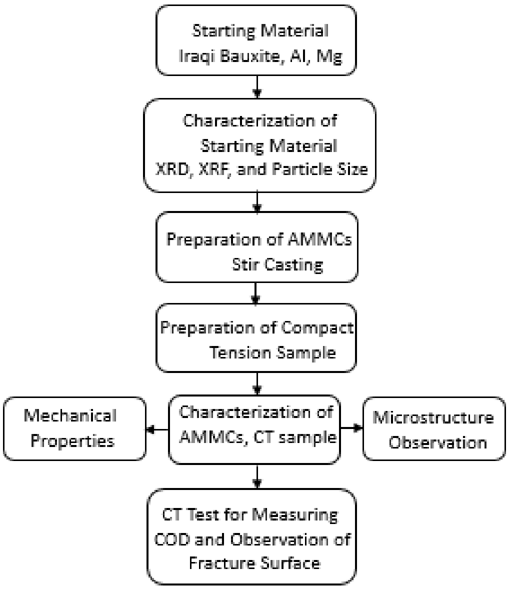

Starting Material
Iraqi Bauxite, Al, Mg
Characterization of Starting Material XRD, XRF, and Particle Size
Preparation of AMMCs Stir Casting
Preparation of Compact Tension Sample
Mechanical Properties
Characterization of AMMCs, CT sample
Microstructure Observation
CT Test for Measuring COD and Observation of Fracture Surface

Figure 1. Flowchart of experimental procedures

(36.3×37.8×15.3 mm^3). The casts were first pre-machined using facing milling machine to get the final dimensions. After milling process, the samples were ground using (YMP-2Machine) with (180-2000 grit) SiC grinding papers, and polished. Then, the notches (10mm length) and pre-crack (2mm length) in the samples were made using Wire Electrode Discharge Machining (WEDM) CNC machine. The Ø=7.65 mm two holes were made using drilling machine type (WDM Z5050) as shown in Figure 2. After that, the casts were subjected to heat treatment at 350 °C for 2 hours for stress reliving.

2. 2. 3. Compact Tension Test The compact tension test was achieved on a universal testing machine type (WAW-200). Figure 3 shows the setup loading parts and the arrangement of the samples, extensometer, and other components of the system in the compact tension test. The samples were fixed between the lower and upper jaws of the machine with the aid of clamping tool. The electronic extensometer(YYU-10/50-111276 Japan) was fixed on the samples for recording its extension during the test, it was fixed to be in front to the crack mouth. The results obtained from the system after loading with range (0.1KN) were (load, time, and crack mouth opening displacement (CMOD)). CMOD is a term for measurement of the crack opening displacement (COD) near the crack mouth. COD concept is widely used in fracture toughness evaluation because of it is superficially simple, plausible, and readily visualized in its ideal aspect.

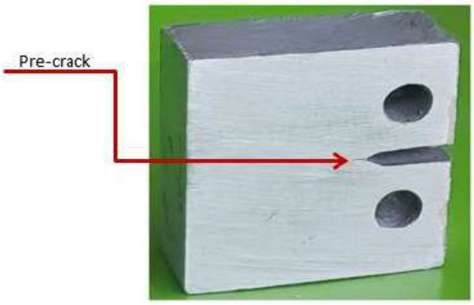


Figure 2. Compact tension of AMMCs Sample

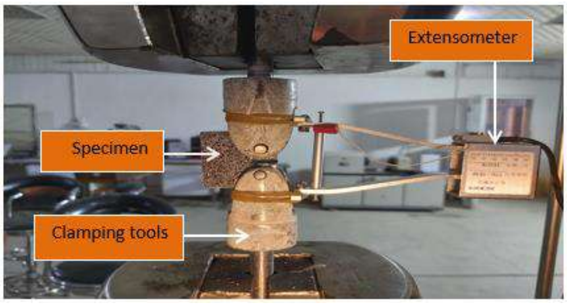


Figure 3. Compact tension sample arrangement during uniaxial tensile test

2. 3. Characterization of AMMCs

2. 3. 1. Calculation of Mechanical Properties

The ultrasonic device (CSI type CCT-4) was used for calculation of Young's modulus and Poisson's ratio of AMMCs samples by using equations 1 and 2 [12].

$$\nu = \frac{1-2\left[\frac{C_s}{C_l}\right]^2}{2-2\left[\frac{C_s}{C_l}\right]^2} \qquad (1)$$

$$E = 2\,\rho C_S^2 (1+\nu) \qquad (2)$$

where:

C_L is speed of sound of longitudinal, C_s is speed of sound of shear, ρ is density of the specimen, and E is Young's modulus and ν is Poisson's ratio.

The Brinell hardness of the AMMCs specimens was achieved using the hardness tester type (WILSON HARDNESS UH-250). An average of three measurements was adopted in recording of hardness results.

The fracture toughness of compact tension specimen was measured according to ASTM E 399 using Equation (3).

$$K_{IC} = \left(\frac{P}{BW}\right)^{1/2} \times f\left(\frac{a}{w}\right) \qquad (3)$$

where:

$$f\left(\frac{a}{w}\right) = \frac{[(2+a/w)(0.886+4.64(a/w)-13.32(a/w)2+14.72(a/w)3-5.6(a/w)4)]}{(1-a/w)^{3/2}}$$

where:

K_{IC}= Fracture Toughness (MPa.m$^{1/2}$), P= Maximum force (kN), B= width of the specimen (m), W= height of specimen (m), a= crack length (m).

2. 3. 2. Microstructural Characterization The SEM images were used to observe the microstructure of the AMMCs specimens and the crack region using Scanning Electron Microscope model (FIE, TESCAN MIRA3). The SEM was used to observe the distribution of bauxite within the microstructure of the AMMCs specimens.

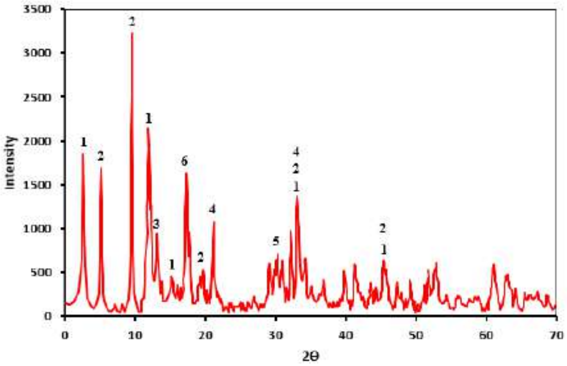
3. RESULTS AND DISCUSSION

3. 1. XRF Table 3 shows the results of XRF analysis of unfired bauxite powder and fired bauxite powder heated at (1400 °C) with soaking time of (3h). The result showed that the percentages of alumina increased after firing bauxite which used as reinforcement phase with decreasing of most impurities. The major oxides found in Iraqi bauxite were alumina, silica, titania, iron oxide which confirmed using XRF. Iraqi bauxite was used to strengthen the AMMCs due to presence these oxides which have good properties for example stiffness and hardness.

3. 2. XRD Bauxite powders were examined before and after the firing. Figure 4 and Table 3 show that the bauxite rocks consist mainly of Kaolinite, Quartz, Anatase, Boehmite, Calcite, and Gibbsite. Figure 5 shows the XRD patterns of bauxite after firing and the major peak at 25.75° is alumina phase due to the bauxite has the highest percentage of alumina and silica. While changed as the proportion of alumina increased which represented by the highest peaks. Also, peaks of some components reduced or almost disappeared after firing, and through the combustion process, a higher percentage of alumina was obtained, which could be further utilized for the reinforcement of the composite material. Furthermore, the high mullite phase amount observed in fired bauxite can make it more suitable for the fabrication of composite materials with improved properties and uniform distribution of reinforcements [13].

TABLE 3. XRF results of unfired and fired bauxite

Bauxite Compound	Unfired Bauxite	Fired Bauxite
%Al_2O_3	54.70	68.39
%SiO_2	21.31	25.23
%Fe_2O_3	1.48	1.94
%CaO	0.05	0.07
%MgO	< 0.02	< 0.02
%TiO_2	2.59	3.36
%SO_3	0.04	0.02
%Cl	0.02	< 0.02
%P_2O_5	0.02	0.02
%K_2O	< 0.02	< 0.02
%Na_2O	< 0.02	< 0.02
%L.O.I	0.78	0.93



1. Kaolinite 2. Boehmite 3. Gibbsite 4. Anatase 5. Quartz 6. Calcite

Figure 4. XRD patterns of unfired bauxite

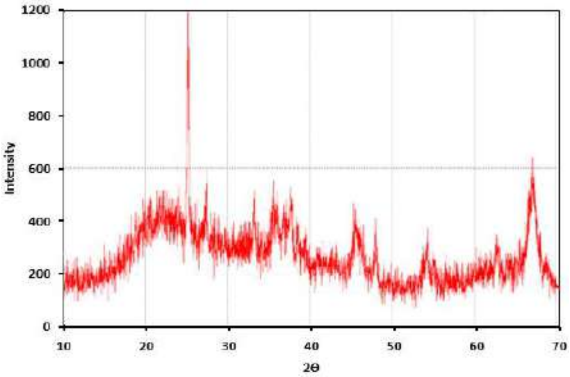


Figure 5. XRD patterns of fired bauxite

3. 3. Particle Size Analysis The average particle size of fired bauxite powders distribution was (0.979µm) as shown in Figures 6. The small particular size provides the particles of distribution more uniformity during the casting process and hence good results for the properties of composite material due to increase in the surface area of fired bauxite powders. At any rate, the distribution of the particle size is effected by milling time and the presence of the agglomeration [14]. Also, a small particle

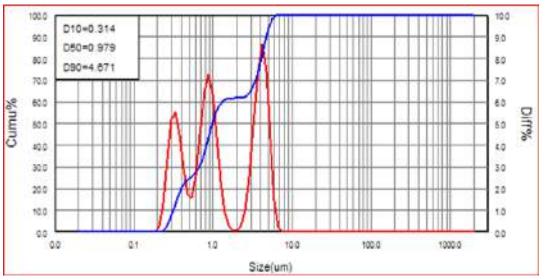


Figure 6. Particle size distribution of fired bauxite powder

size of the reinforcements will provide them a fair homogenous distribution in the matrix which can lead to microstructural advantage [1]. Anyhow, compared to nano-sized particles, micro particles can have a more positive influence on the relative density of AMMCs and their properties due to higher wettability and clustering problems of nano-sized particles [15].

3. 4. Microstructural Observation

The microstructure of the different specimens are shown in Figure 7 using SEM**.** A fair homogenous distribution of bauxite particles was observed in the matrix alloy which leads to more microstructural advantage. The SEM images proved that continuous stirring is essential in the fabrication of bauxite reinforced AMMCs composites due to the tendency of particles to form agglomeration or clusters sites at higher reinforcement content [1, 16]. It can also observed that the composites are free from any type of defect in casting.

It is clearly shown that the use of stir during casting of composite induced an acceptable distribution of the reinforcing particles which may improve its mechanical properties. Anyhow, the precise presence of every elements of Al matrix and reinforcement would be confirmed from the EDX analysis [1]. In spite of an acceptable distribution of the particles, there were little agglomerations of the particles in the matrices. This may be reflected on the properties of the specimen’s .The proper choice of the parameters for stir casting (stirring speed, time, blade's design, stirring temperatures) played a vital role in the uniform distribution of the bauxite particles in Al matrix. The interface characteristics between matrix metals and reinforcements and microstructure are strongly affected the properties of composites. The microstructural observations shown in Figure 7 were clearly characterized by relatively non-homogeneous distribution of bauxite particles in Al matrix and clustering. Probably, during composites processing the variation of contact time between the molten Al and reinforcement, resulted in poor wetting behavior of particles in the molten and high surface tension, thereby could lead to such distribution. The high bauxite addition in samples S4, resulted in microstructure characterized by more non- homogenization of reinforcement particles in comparison to other AMMCs samples. Generally, the microstructure of all samples were characterized by porosities, because that the bauxite particles introduced air in the melt entrapped between the particles when they added in the melt. Therefore higher porosity can attributed to increase in bauxite particles addition.

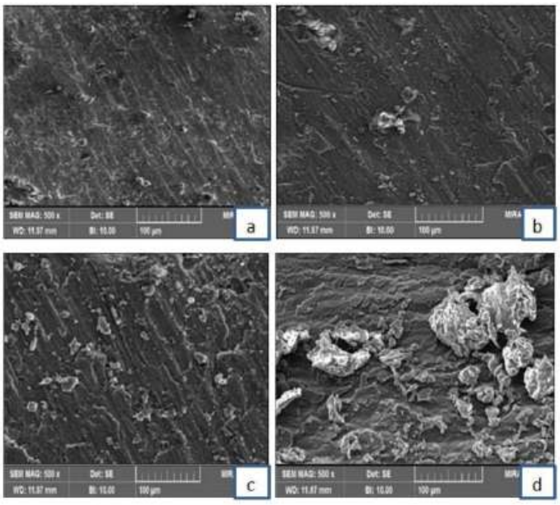


Figure 7. Microstructure observation of AMMACs specimens: (a) S0, (b) S1, (c) S2, and (d) S3 using SEM

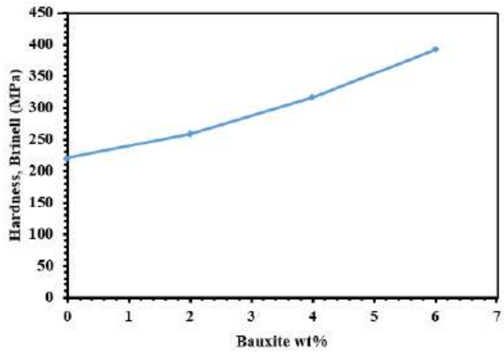


Figure 8. Effect of the bauxite on the hardness of AMMCs Specimens

3. 5. Results of Mechanical Properties

3. 5. 1. Hardness Results

The results from Brinell hardness test are presented in the Figure 8. Each value was considered as an average of three reading. The micro-hardness of AMMCs increased with an increasing of reinforcement. Addition of bauxite particles in to Al melt provided supplementary substrate for the solidification to activate there by decreasing the grain size and increasing the nucleation rate. High hardness of bauxite particles could act as barriers to dislocation motion or the matrix motion. Also, the hardness of the AMMC composites enhanced with an increase in weight percentage of bauxite reinforcements due to the resistance provided to the indentation by hard bauxite particles. The increment percentage of Brinell hardness were 17.6, 35.3 and 56.8% for the S1, S2, and S3, respectively comparing with unreinforced specimens (S0). The results are in agreement with other works [17]. Furthermore, the microstructure of composite is strongly affected by stirring time and stirring speed which can cause the change in structure and hardness results of fabricated AMMCs. Also, the particle agglomerate is more when operated with less stirring time and at low stirring speed. However, increase in speed and stirring time can give better particles distribution [16].

3. 5. 2. Results of Fracture Toughness, Young's Modulus and Poisson's Ratio (K_{IC}, E, and ν) The results of fracture toughness, ultrasonic method for determination of (E,ν) are shown in Table 4. The clear increasing of fracture toughness, Young's modules and Poisson's ratio can be related to effect of sinterability of MgO which hinder the grain growth and enhanced the densification of the sample while at 6% the fracture toughness was decreased due to the brittle behavior of AMMCs at this percent [18-20].

3. 6. Time and Load Relationship From the results recorded using the extensometer device, the values of the load projected with time and the deformations of the material along the fracture area were obtained. The relationship between time and load is shown in Figure 9. It showed load increasing with time and reached a critical value as the maximum value of the projected load. After this value the load decreased, thereby indicating the occurrence of the crack and failure of the material. In general, this maximum load increased from 4.74 kN for sample S0 to 5.9 kN and 6.8 kN after increasing bauxite particles to 2and 4 wt%, respectively. Again, the reinforcement using bauxite particles proved its effects in improvement of fracture toughness of the composite materials. Anyhow, sample S3 expressed brittle behavior, which resulted in decreasing of maximum load to 5 kN.

TABLE 4. Results of fracture toughness, Young's modulus, and Poisson's ratio

Sample	KIC (MPa.m$^{1/2}$)	E (GPa)	N
S0	19	69	0.33
S1	23	74	0.33
S2	32	77	0.22
S3	16	84	0.31

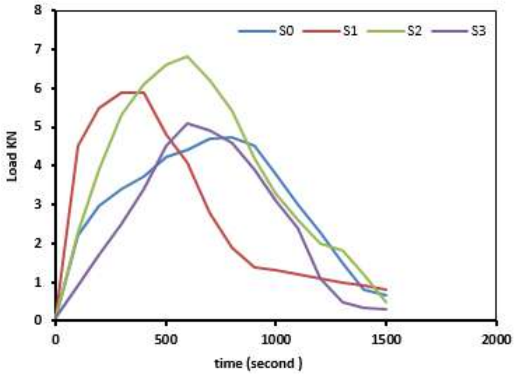


Figure 9. Load versus time of AMMCs specimens

3. 7. COD and Time Relationship CMOD results at mouth notch were recorded using the clip gauge that fixed on the mouth of notch region, as described in Figure 3. Figure 10 shows the results of time versus CMOD from extensometer recordings. It can be observed, that the area of CMOD decreased with increasing the bauxite particles from 2 to 4 wt%, denoting the strengthening influences in the AMMCs samples. Anyhow, S3 samples (6%wt) could not give the required results because of their behavior similar to those of brittle material.

3. 8. COD and Load Relationship The relationship between CMOD with applied load are shown in Figure 11. Generally, the curves can be divided into three stages. The first stage showed a sudden increase or a linear increase in the load to a certain limit with CMOD, denoting the materials resistance to the applied load. The second stage represented the critical stage in the behavior, where there was a continuous increasing in the load to critical point of maximum load and critical CMOD. Furthermore, the reinforcement with bauxite particles proved its effects in improving the fracture toughness properties by means of small critical CMOD at high critical load. Also, the critical CMOD values were 5.4, 3.7 and 2.9 mm for sample S0, S1 and

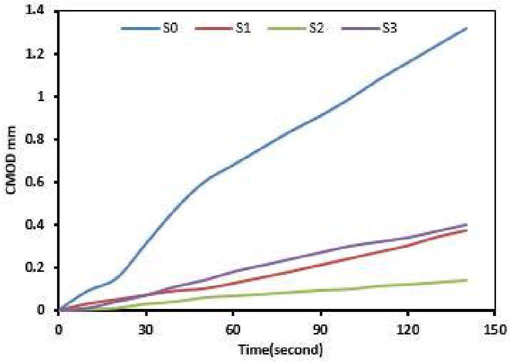


Figure 10. Results of time versus CMOD from extensometer recordings

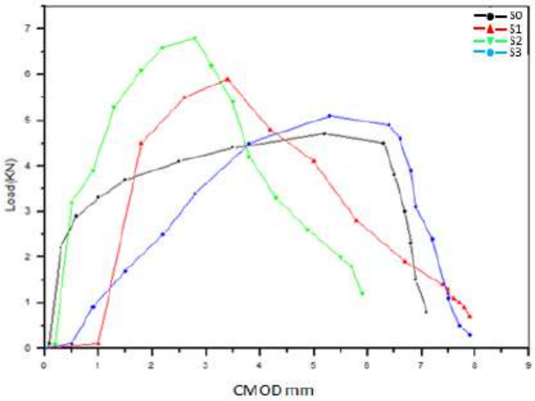


Figure 11. Results of Load versus CMOD

S2, respectively. Again, the brittle behavior of S3 sample recorded values of CMOD and load better than those of samples S0, but not better than those of samples S1 and S2. The three stage represented the failure stage, where the load decreased with increasing of CMOD to the failure point. It can be also divided to three parts. The first part, the relationship is linear until reach the peak load (maximum load). The second part start when move away from peak load where the crack opening displacement (CMOD) increase with decreasing the load. The final part represent the tail of curve shows the crack opening displacement (CMOD) increase while the load gradually decreased. The critical (CMOD) are 5.4, 3.7 and 2.9 mm of S0, S1 and S2, respectively. The critical (CMOD) of S1 and S2 is less than of S0 that result from addition of bauxite that enhancing the mechanical properties of sample. While the critical CMOD of S3 increased due to the brittle behavior produced from addition (6wt %) of bauxite.

3. 9. Fracture Results

Figures 12 to 15 show the SEM images for the crack regions for different samples. In general, the images could prove the role of reinforcement with bauxite particles in improving the fracture toughness of AMMC$_S$ by means of cracks propagation and their behavior in the composite materials.

Figure 12 shows the fracture nature of a metal-based alloy. It can be clearly observed that the magnesium granules located around the fracture area, and due to the high ductility of aluminum, high deformation can be observed in the fracture region.

For samples S1, Figure 13 shows less fracture progresses with less deformations than those of samples S0. Also, the bauxite particles presented in the way of fracture progress, could cause obstruction of the fracture progress, and thereby, increasing the fracture toughness**.**

For samples S2, Figure 14 shows a clear ablution in the bauxite granules that are more presented in the fracture area and thus hinder the progress of the fracture. This increases in the time period of failure resulting from changing in the fracture trajectory due to the accumulation of the bauxite particles. The bauxite percentages of 2 and 4% in both samples S1 and S2, respectively leads to enhance the fracture toughness of AMMCs due to stain localization ahead of crack tip (i.e. large plastic zone ahead of fractured particle).

Increasing the bauxite content to 6% in the samples S3 causes an increase in the brittleness of the composite material and hence rapid progression of fracture as shown in Figure 15. The high percentage of bauxite caused to increase the high density of dislocation of the aluminum metal matrix composite which leads to increase the brittle phase at this percentage.

Generally, the results of mechanical properties obtained in the present work can be comparable to those of other works in the literature although they recorded hardness results higher than those of the present study. Furthermore, the study presented in this paper focused on

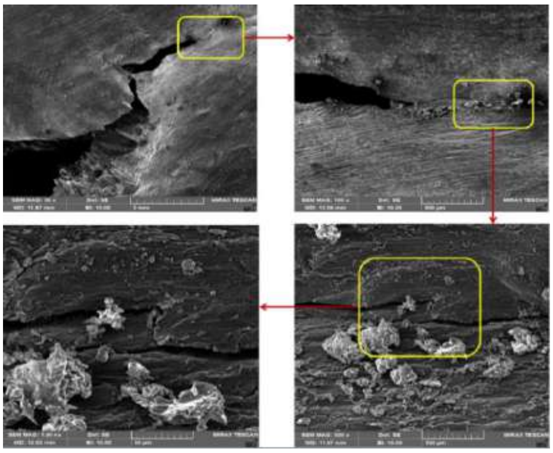


Figure 12. SEM images of samples S0 at different magnifications

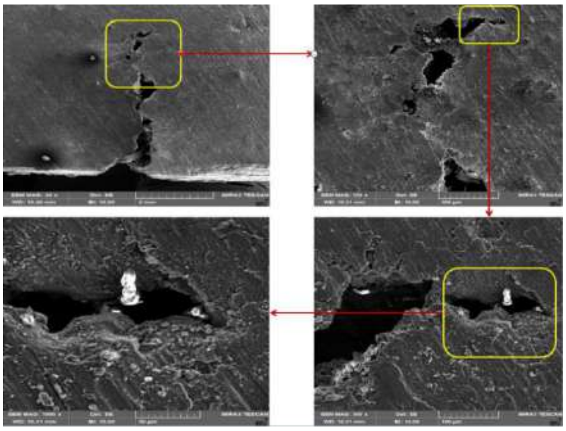


Figure 13. SEM images of samples S1 at different magnifications

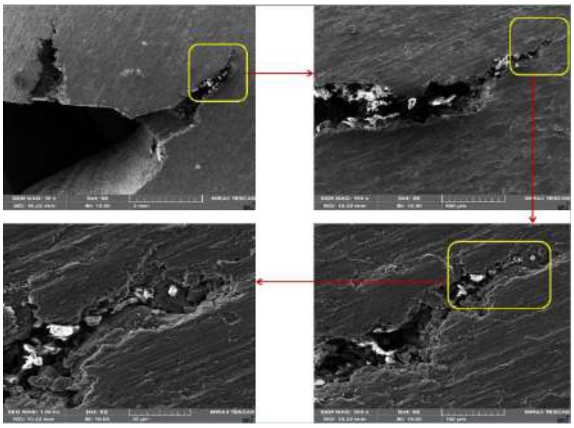


Figure 14. SEM images of samples S2 at different magnifications

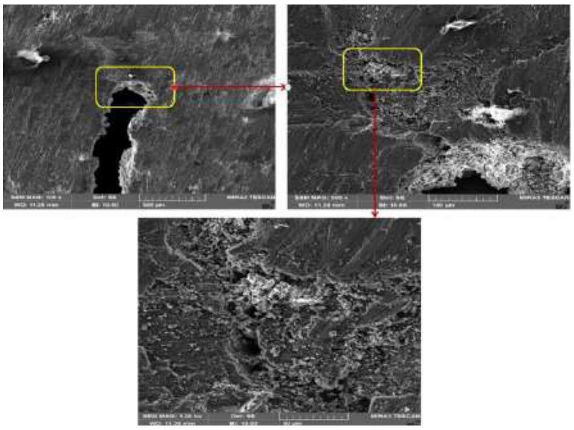


Figure 15. SEM images of samples S3 at different magnifications

development new process in fabrication of AMMCs through relatively green and natural resources of reinforcements, which eliminates using of conventional or industrial reinforcements, hence concerning for the costs saving and human safety. The other works [2, 3, 5, 16] used many types of reinforcement like, Al_2O_3, SiC, TiC, Gr, TiO_2, and B_4C, where these reinforcements have more advantages over its matrix material and still having the cost barrier. Also, other studies in the literature used industrial waste martials like red mud [1,16] , and industrial and agro waste like fly ash, groundnut shell, rice husk ash and bagasse ash [16, 21] in reinforcement of AMMC composites. Anyhow, compared to the present study ,most of works aforementioned above focused on the investigation of mechanical properties (hardness and ultimate tensile strength), tribological properties, and thermal expansion, and they did not investigate the fracture toughness using the methods which adopted in the present study (calculated and measured by means of crack mouth opening displacement CMOD measurements from extensometer recordings) and Young's modulus .The results of present study might promote the future works in reinforcement of metal matrix composites using natural bauxite and study of their accurate fracture mechanics using CMOD measurements and accurate observation of crack initiation, propagation and failure.

4. CONCLUSIONS

The outstanding results concluded from the study are list as follows:

1. Iraqi natural bauxite reinforced AMMCs with improved mechanical properties and fracture toughness can be prepared using stir casting route.
2. The claimed improvements can be attributed to uniform distribution of bauxite particles in the matrix, their ceramic properties, and their obstructing to the crack initiation.
3. Bauxite addition at 2, and 4 wt% could give improvements in AMMCs hardness at about 17 and 43%, respectively.
4. CMOD measurements showed the effect of addition bauxite particles at 2 and 4% in increasing "maximum load at failure" and "critical CMOD at critical load" of the matrix materials to about " 25 and 44%" , and " 32 and 47%", respectively.
5. The calculated fracture toughness by means of K_{IC}, and young modulus at additions of 2 and 4 wt% showed improvement at about " 22 and 69%" ,and " 8 and 12%", respectively.
6. Addition of bauxite at 6 wt% could not give the required improvement in the fracture toughness despite its effects in recording the highest improvements in hardness (57%) and elastic modulus (22%) due to the brittle behavior of AMMCs at this ratio.

5. REFERENCES

1. Samal, P., Raj, R., Mandava, R.K. and Vundavilli, P.R., Effect of red mud on mechanical and microstructural characteristics of aluminum matrix composites, in Advances in materials and manufacturing engineering. 2020, Springer.75-82. DOI: 10.1007/978-981-15-1307-7_8
2. Imran, M. and Khan, A.A., "Characterization of al-7075 metal matrix composites: A review", ***Journal of Materials Research and Technology***, Vol. 8, No. 3, (2019), 3347-3356. DOI: 10.1016/j.jmrt.2017.10.012
3. Singh, J., Jawalkar, C. and Belokar, R., "Analysis of mechanical properties of amc fabricated by vacuum stir casting process", ***Silicon***, (2019), 1-11. DOI: 10.1007/s12633-019-00338-8
4. Sahu, P.S. and Banchhor, R., "Effect of different reinforcement on mechanical properties of aluminium metal matrix composites", ***Research Journal of Engineering***, Vol. 6, No., (2017), 39-45. http://www.isca.in/IJES/Archive/v6/i6/7.ISCA-RJEngS-2017-142.php
5. Soltani, S., Khosroshahi, R.A., Mousavian, R.T., Jiang, Z.-Y., Boostani, A.F. and Brabazon, D., "Stir casting process for manufacture of al–sic composites", ***Rare Metals***, Vol. 36, No. 7, (2017), 581-590. https://link.springer.com/article/10.1007/s12598-015-0565-7
6. Chen, Y., Li, B., Shi, Y. and Ouyang, S., "Natural advantages of preparation of composites from minerals: Effect of bauxite addition on the microstructures and properties of fe-al2o3 based composites", ***Materials***, Vol. 12, No. 9, (2019), 1456. DOI: 10.3390/ma12091456
7. Zabihi, O., Ahmadi, M., Ferdowsi, M.R.G., Li, Q., Fakhrhoseini, S.M., Mahmoodi, R., Ellis, A.V. and Naebe, M., "Natural bauxite nanosheets: A multifunctional and sustainable 2d nano-reinforcement for high performance polymer nanocomposites", ***Composites Science and Technology***, Vol. 184, No., (2019), 107868. DOI: 10.1016/j.compscitech.2019.107868
8. Al-Amer, E.M.H. and Al-Kadhemy, M.F.H., "Improving the physical properties of iraqi bauxite refractory bricks", ***Al-Nahrain Journal of Science***, Vol. 18, No. 3, (2015), 67-73. https://www.iasj.net/iasj?func=article&aId=140601

9. Njoya, D., Elimbi, A., Fouejio, D. and Hajjaji, M., "Effects of two mixtures of kaolin-talc-bauxite and firing temperatures on the characteristics of cordierite-based ceramics", ***Journal of Building Engineering***, Vol. 8, No., (2016), 99-106. DOI: 10.1016/j.jobe.2016.10.004

10. Imad, S., Ali, N.M. and Abood, T.W., "Improving the physical and mechanical properties of fireclay refractory bricks by added bauxite", ***Journal of Engineering***, Vol. 25, No. 4, (2019), 18-28. DOI: 10.31026/j.eng.2019.04.02

11. Mustafa, M.M., Al-Bassam, K.S. and Al-Ani, T.M., "Karst bauxite deposits of north hussainiyat area, western desert, iraq: An overview", ***Iraqi Bulletin of Geology and Mining***, Vol., No. 8, (2019), 125-146. https://www.iasj.net/iasj?func=article&aId=167115

12. Astm, E., "399-90:" Standard test method for plane-strain fracture toughness of metallic materials", ***Annual book of ASTM standards***, Vol. 3, No. 01, (1997), 506-536. https://www.astm.org/DATABASE.CART/HISTORICAL/E1820-01.htm.

13. Pasha, M.B. and Kaleemulla, M., "Processing and characterization of aluminum metal matrix composites: An overview", ***Reviews on Advanced Materials Science***, Vol. 56, No. 1, (2018), 79-90.

14. Shotorbani, A.R. and Saghai, H.R., "Modeling and implementing nonlinear equations in solid-state lasers for studying their performance", ***Emerging Science Journal***, Vol. 2, No. 2, (2018), 78-84. DOI: 10.28991/esj-2018-01130

15. Ünal, T.G. and Diler, E.A., "Properties of alsi9cu3 metal matrix micro and nano composites produced via stir casting", ***Open Chemistry***, Vol. 16, No. 1, (2018), 726-731.

16. Kant, S. and Verma, A.S., "Stir casting process in particulate aluminium metal matrix composite: A review", ***International Journal of Mechanics and Solids***, Vol. 9, No. 1, (2017), 61-69.

17. Daoud, A., Abou El-Khair, M. and Abdel-Azim, A., "Effect of al 2 o 3 particles on the microstructure and sliding wear of 7075 al alloy manufactured by squeeze casting method", ***Journal of Materials Engineering and Performance***, Vol. 13, No. 2, (2004), 135-143. DOI: 10.1361/10599490418325

18. Aswad, M.A., Awad, S.H. and Kaayem, A.H., "Study the effect of bauxite ceramic powder on the fracture mechanics of aluminum alloy using digital image correlation", Vol., No.

19. El Amri, A., Haddou, M.E.Y. and Khamlichi, A., "Thermal-mechanical coupled manufacturing simulation in heterogeneous materials", ***Civil Engineering Journal***, Vol. 2, No. 11, (2016), 600-606. DOI: 10.28991/cej-2016-00000062

20. Arasaretnam, T. and Kirudchayini, A., "Studies on synthesis, characterization of modified phenol formaldehyde resin and metal adsorption of modified resin derived from lignin biomass", ***Emerging Science Journal***, Vol. 3, No. 2, (2019). DOI: 10.28991/esj-2019-01173

21. Samal, P., Mandava, R.K. and Vundavilli, P.R., "Dry sliding wear behavior of al 6082 metal matrix composites reinforced with red mud particles", ***SN Applied Sciences***, Vol. 2, No. 2, (2020), 313.

Persian Abstract

چکیده

طی دهه های گذشته محققان علاقه زیادی به تحقیق در مورد مزیت طبیعی تهیه کامپوزیت ها از مواد معدنی مانند ذرات بوکسیت نشان داده اند و اثربخشی آنها را به عنوان تقویت کننده های مقرون به صرفه در ساخت کامپوزیت های با کارایی بالا ثابت کرده اند. این مطالعه ، تلاش جدیدی در استفاده از پودر بوکسیت طبیعی عراق با نسبت های مختلف (۲ ، ٤ و ٦ درصد وزنی) در تهیه کامپوزیت های ماتریس فلزی آلومینیوم (AMMC) با استفاده از ریخته گری و مواد افزودنی منیزیم است. در کار آزمایشی ، سنگهای بوکسیت خرد شده و آسیاب شدند ، سپس این پودر با دمای ١٤٠٠ درجه سانتیگراد حرارت دید. پودرها با استفاده از اندازه ذرات ، تجزیه و تحلیل XRD و XRF مشخص شدند. خصوصیات ریخته گری AMMC به ماشینکاری ، جلا ، پیش گرم شده و با استفاده از اندازه گیری های سختی ، مشاهدات ریزساختاری و محاسبه مدول Young ، نسبت پواسون و مقاومت به شکست آنها مشخص شد. همچنین ، ضخامت شکستگی آنها با استفاده از اندازه گیری جابجایی بازشدن دهان (CMOD) از ضبط های extensometer مورد ارزیابی قرار گرفت. نتایج نشان داد که تولید موفقیت آمیز AMMC با مقاومت به سختی شکستگی ، سختی و خاصیت مدول الاستیک با استفاده از Mg و بوکسیت حریق عراقی با ۲ و ٤ درصد وزنی با ریخته گری اضافه شده است. علاوه بر این ، نتایج حاصل از اندازه گیری CMOD ، اثر ذرات بوکسیت اضافی را در ۲ و ٤ درصد وزنی در افزایش "حداکثر بار در شکست" و "CMOD بحرانی در بار بحرانی" مواد ماتریس به حدود "۲۵ و ٤٤٪" ، و "نشان داد. به ترتیب ۳۲ و ٤٧٪ ". همچنین ، در این نسبت ها ، محکم بودن شکستگی مواد ماتریس با استفاده از KIC ، و مدول یانگ به ترتیب در حدود "۲۲ و ٦٩٪" و "۸ و ۱۲٪" بهبود نشان داده است. افزودن بوکسیت با ٦٪ نمی تواند با وجود تأثیرات آن در ثبت بالاترین پیشرفت در سختی (۵۷٪) و مدول الاستیک (۲۲٪) به دلیل رفتار شکننده AMMC ها در این نسبت ، بهبود مورد نیاز را ایجاد نکند.

International Journal of Engineering

Journal Homepage: www.ije.ir

Life Prediction of Carbon Fiber Reinforced Polymers using Time Temperature Shift Factor

T. A. Hafiz*[a,b]

[a] Department of Mechanical Engineering, College of Engineering, Taif University, Al-Hawiah, Kingdom of Saudi Arabia
[b] Bristol Composites Institute (ACCIS), Department of Aerospace Engineering, Queen's Building, University of Bristol, Bristol, United Kingdom

PAPER INFO

Paper history:
Received 07 April 2020
Received in revised form 01 May 2020
Accepted 23 May 2020

Keywords:
Carbon Fiber Reinforced Polymers
Dry and Wet Coupons
High Temperature
Moisture
Storage Modulus
Time Temperature Shift Factor

ABSTRACT

The properties of Carbon Fiber–Reinforced Polymers (CFRP) are greatly affected under extreme environmental conditions. This paper reports an experimental study to determine the response of IM7-carbon/977-2 cycom epoxy laminates under different humidity and temprature conditions. Short-term 3-point bending creep tests using Dynamic Mechanical Analysis (DMA) were used to test the dry and saturated samples at various temperature levels. The dry coupons were tested at the room temperature (RT) and at 60-120 °C with 20 °C increment and then at 130 °C, 150-180 °C with 10 °C increment for each next test. The saturated (wet) coupons were tested at RT, 40 - 120 °C with 10 °C increment in temperature for each next test and at 145 °C, 150 °C, and 160 °C. The time-temperature shift factor (TTSF) was applied and it is shown that the viscoelastic behavior of the invetigated IM7-carbon/977-2 epoxy laminates, is accurately predicted through the use of TTSF. It has also been shown that determining the viscoelastic behavior at elevated temperatures helps to predict temperature below the glass transition temperature using TTSF. The long-term life of the material is relatively easily predicted using TTSF by conducting traditional short-term laboratory tests.

***doi**: 10.5829/ije.2020.33.07a.21*

1. INTRODUCTION

Fiber Reinforced Polymers (FRPs) are light weight and possess high specific strength and stiffness. Therefore, their usage is continuously increasing in many primary and secondary load carrying structural applications such as aerospace, missiles, aircrafts, automobiles, etc. However, the properties of these materials are greatly affected when exposed to high temperature and moisture. Therefore, it is crucial to understand the behavior of these materials for their widespread usage. Hence, there is a need to develop such testing procedures and protocols to be used to evaluate life-times properties of materials in extreme service conditions. Therefore, many researchers have attempted to characterize FRPs under different testing conditions [1]. FRPs have long expected life in many applications. Therefore, it can be expected that the properties of these material in service may change over the long period of time. For example, it is expected that FRPs used in a highspeed commercial aircraft may last for 20 years with 2500 days (6.85 years) of the time at elevated temperature [2]. With services, lifetime measured for such a long time, it is not possible to conduct real time experimental tests under different testing conditions. Even if it becomes possible to do testing for this long period, considering advances in technology, this will be very impractical as the material under consideration for testing for couple of decades might have been totally changed that one might plan to initiate testing today. Due to these reasons, accelerated testing methodologies are getting more and more attention [2].

Stress-Number of cycle to failure (S-N) curve is very common among the researchers community for fatigue life prediction. In simple words, the curve assumes that fatigue life of the material is dependent on number of cycles. Simply increasing the loading frequency in the fatigue testing, the testing can be accelerated. A

*Corresponding Author Institutional Email: *Htali@tu.edu.sa* (H. T. Ali) reduction in testing time by a factor of 120 through this has been reported in the literature [2]. However, there is limitation in this methodology that generally it does not consider the viscoelasticity or oxidation phenomenon in the fatigue response. Unlike metals, the polymers have strong viscoelastic behavior. Therefore S-N curve is not very helpful in accurately predicting the fatigue life of FRPs. As there is great dependency of viscoelastic materials on time, hence changing other parameters will not help either. As the accelerated testing of viscoelastic deformation is well known, therefore Time-Temperature Shift Factor (TTSF) is used for such materials [3]. The procedure of this principle is very simple. Time is effectively accelerated by elevating the temperature. In this way, this principle overcomes the difficulty faced by many accelerating methods to find the link between the accelerated test with the real-life operation.

The working function of TTSF is that measurements of material compliance are taken, at several different temperatures for a small time period. These values are then plotted against the log of time. TTSF is utilized to shift the values for compliance measurements on the log time scale. Shifting the values on the log time scale generates a curve which is named as master curve. The compliance of the material for any time range can then be predicted using the master curve. The acceleration in the testing due to elevated temperatures is established by compliance values overlapping curves at different temperatures [2]. To draw the master curve using TTSF, one curve is taken as reference and hence does not move along the log time scale. Whereas rests of the curves obtained during specific testing regime are shifted to match with the fixed curve. Therefore, it should be kept in mind that master curve predicts the time-temperature deformation for the temperature associated with the first curve. By noting the magnitude of the shifts required to form the master curve, a relation between the horizontal/vertical shifts and temperature/modulus is developed. Therefore, the entire master curve is moved to other temperatures and moduli to predict the life of the viscoelastic materials. This curve is also useful to predict creep compliance by shifting it to appropriate temperatures. Although the application of TTSF was initially for non-destructive characterization of material properties over the time, it has now been extended to characterize the polymeric matrix composites deformation properties [4-8]. Models have been developed to predict the shift factor with change in temperature.

Miyano et al. [9] measured the effect of different constant strain-rates and temperatures on the compressive strength of resin impregnated carbon fibre strands. Using the procedure for the tensile strength reported in literature [10], they obtained master curve for compressive strength. They predicted the tensile and compressive strengths of unidirectional Carbon Fiber–Reinforced Polymers (CFRP) materials along the longitudinal direction under constant strain-stress loadings by making use of Rosen’s strength formula [11-12] for elastic matrix composites.

Despite the many attempts, the durability of materials is generally not understood comprehensively under the exact nature of environmental attacks such as moisture and higher temperature. It is well understood now that higher temperature not only causes degradation but also changes significantly the performance and failure mechanisms of CFRP [13-15].

The aim of this study is therefore to conduct a comprehensive experimental study to consider the effect of different significant parameters such as moisture uptake and high temperature to predict the life performance of aerospace graded IM7-carbon/977-2 epoxy laminates using Time Temperature Shift Factor (TTSF).

Following the introduction section, the structure of the manuscript is designed as follows: In section 2, the theoretical background of TTSF is detailed along with supporting mathematical expressions. In section 3, sample preparation and procedure are detailed. Results and discussions are presented in section 4 while in section 5 concluding remarks are highlighted, followed by references.

2. THEORATICAL BACKGROUND

Viscoelastic materials deform slowly when exposed to an external force but return to their original configuration when the external force is removed. Wiechert model is very helpful to understand the overall response of such materials with viscoelastic properties. The model helps to grasp the phenomenon of distribution of relaxation time. The model is used to calculate the relaxation modulus as given below:

$$C(\mathrm{t}) = E_e + \sum_i E_i \, exp\left(\frac{-t}{\tau_i}\right)$$

where the notations used have their usual meanings. i.e C(t), Ee, Ei and τi are the viscoelastic relaxation modulus, the equilibrium modulus, elastic modulus and relaxation time, respectively.

The expression can then be used to derive expression for E' and E''.

$$E' = E_e + \sum_i E_i \frac{\omega^2 \tau_i^2}{1+\omega^2 \tau_i^2}$$

$$E'' = \sum_i E_i \frac{\omega \tau_i}{1+\omega^2 \tau_i^2}$$

where E', E'' and ω are the storage, loss modulus and applied circular frequency, respectively. Moduli are derived in terms of the relaxation distribution.

It is worth nothing from expression for E'' that when the applied circular frequency is infinitely small then in turn E''will be very small. Hence the loss modulus is neglected mostly. Therefore, in this case, resulting stress and strain have no lag phase. If the case remains valid under negligible or very small frequency, then it can be assumed that E', $\approx$ C (t). If the expression remains valid then Miyano et al. [13] verified that the relaxation modulus can be used to obtain creep compliance S(t). It must however be noted that the creep response takes longer to get to its equilibrium than relaxation response. It is worth noting that increasing the temperature reduces the viscoelastic relaxation time for viscoelastic materials. Therefore, increase in temperature accelerates the process effectively for such materials. For thermorheologically simple materials, the relaxation modes of such materials can be calculated in terms of temperature by defining time-temperature stated below:

$$\tau_i(T) = a_T\tau_i(T_0)$$

where T_0 and a_T are reference temperature and temperature-dependent horizontal shift factor, respectively. Therefore, the viscoelastic behavior of the material at T_0 is helpful to calculate the storage modulus at temperature T as:

$$E'(\omega,T) = E_e(T_0) + \sum_i E_i(T_0)\frac{(a_T\omega)^2\tau_i^2(T_0)}{1+(a_T\omega)^2\tau_i^2(T_0)}$$

In the storage modulus equation, frequency ω is replaced with time as ω=1/t. Therefore, time-temperature superimposed master curve is obtained by shifting the response curves at different temperatures along the logarithmic time axis as shown in Figure 1 [15].

Scaling factors are temperature dependent; in logarithmic axis, time scaling represents horizontal shift, a_{T0} while modulus scaling represents a vertical shift, b_{T0}.

$$\log a_{T0}\ (T) = \log t - \log t'$$

$$\log b_{T0}\ (T) = \log D_c\ (t,T) - \log D_c(t', T_0)$$

where Dc is the creep compliance measured from the deflection of the specimen's center.

Hence, the life of composite materials is relatively easily estimated through the application of TTSF by conducting traditional short-term laboratory tests.

3. EXPERIMENTAL METHODLOGY

3. 1. Material and Sample Preparation

IM7-carbon/977-2 epoxy Composite system was selected for the current research work. 977-2 is an epoxy blended with a thermoplastic polymer for toughening purposes. This material is extensively employed in primary aeronautical structures, such as fuselages and wings.

Unidirectional IM7-carbon/977-2 epoxy prepreg of 450 mm length and 200 mm width were cut and then five layers of the prepreg were laid down manually. To avoid the air entrap between the plies, vacuum was applied after every layer. After laying down the plies, the laminate was then vacuum bagged. The manufacturer recommended curing data cycle was strictly followed for curing the material in autoclave. After autoclave curing, laminates were cut in desired rectangular strips of 50 x 15 x 1 as shown in Figure 2, using diamond saw cutter. The samples were divided into two (minimum 40 test samples in each half). Half of the coupons were stored in desiccator until just before the time of testing. The remaining half of the coupons were used for moisture uptake. Water bath was used for this purpose. The water bath temperature was kept at 25 ˚C. As it was noted that moisture uptake rate was very slow hence the temperature was increased to 80 ˚C. This accelerated the moisture uptake as expected. The moisture uptake in samples in the water bath were monitored regularly and were kept in the bath until all the samples had moisture content of about 1%.

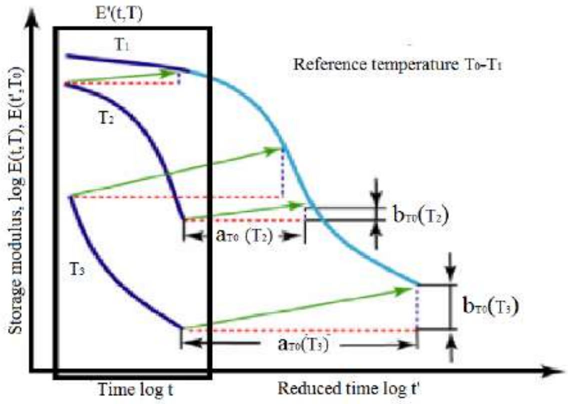


Figure 1. Shifting of storage modulus using TTSF

3. 2. Dynamic Mechanical Analysis (DMA)

For all the tests reported in this manuscript, Dynamic Mechanical Analysis (DMA) was applied using 3-point bending to investigate the viscoelastic effect on CFRPs samples. The testing reporting in this manuscript was performed according to ASTM standard. DMA is used to deform the coupon mechanically and then the sample response is measured. DMA facilitates to monitor the deformation response of the sample using time or temperature as a function.

Dimensions of the dry and saturated coupons were kept same. The length, width and thickness for the tested samples were kept as 50, 15 and 1 mm respectively. The schematic of the test samples and the test setup, showing

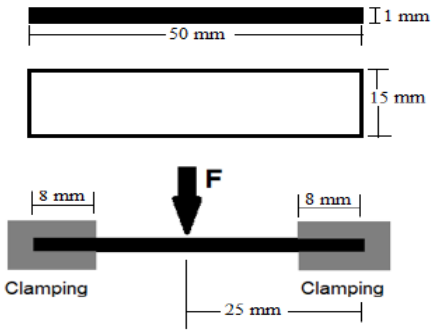


Figure 2. Schematic of the testing and set-up

the width, length and thickness dimensions is shown in Figure 2. The testing time for dry as well as saturated coupons were chosen to be 2000 seconds (33.33 minutes). However, to estimate the full trend of the samples under creep, it was decided to run the dry sample test at 130 °C for 20000 seconds (5.5 hours). It was noted that for the sample, load started to drop below zero after just running for 3836 seconds (63.9 minutes) as is obvious from Figure 3.

Dry coupons were tested at RT, and at 60 - 120 °C with 20 °C increment in temperature and then at 130°C, 150°C - 180 °C with 10 °C increment in temperature as plotted in Figure 3. Saturated (wet) coupons were tested at RTRT, 40 - 120 °C with 10 °C increment in temperature for each next test and at 145 °C, 150 °C, and 160 °C as plotted in Figure 4. It is clear that the wet/saturated samples have shown lower storage modulus compared to storage modulus for the dry coupons for almost all of the temperatures tested except for 60 °C but even in this particular test there is a slight difference between the storage moduli of the coupons.

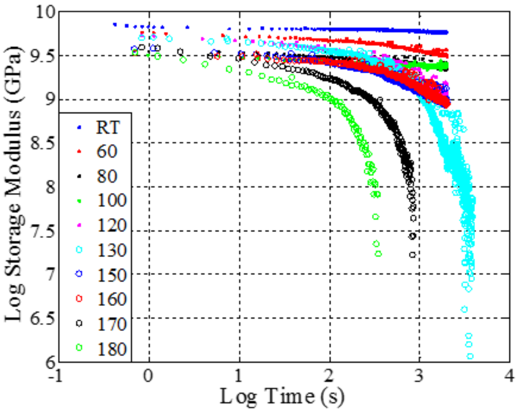


Figure 3. Storage modulus from relaxation tests for different temperature ranges for the dry CFRP coupons

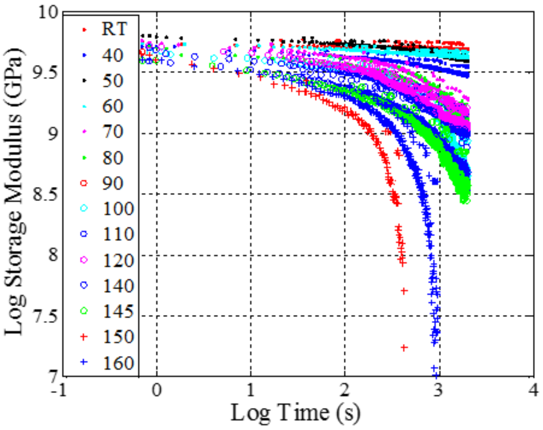


Figure 4. Storage modulus from relaxation tests for different temperature ranges for the wet CFRP coupons

4. TEST RESULTS AND DISCUSSIONS

Following major steps were applied to obtain the master curve for the dry and wet coupon tests performed under 3-point bending test using DMA.

1. Perform creep relaxation tests in a prescribed temperature range

2. Fix a reference temperature and obtain the constants for the reference temperature test(s) via a nonlinear regression on

$$Log[E_s(t,T_{RT})] = \beta_1^{RT}\left[\frac{1}{2}-\frac{1}{\pi}\tan^{-1}[\beta_2^{RT}(Logt-\beta_3^{RT})]\right]$$

where Es (t,T_{RT}) is the room temperature (RT) storage modulus after time t and β_i ^{RT} are coefficients that are estimated from a least square regression performed on the experimental data; T_{RT} is the reference temperature. The resulting fit for dry and saturated coupons is also shown in Figures 5 and 6.

3. Estimate the horizontal and vertical shift factors at the various temperatures via minimising the quadratic error w.r.t. to the "shifted" equation

$$Log[E_s(t,T)] = Log[b_{RT}(T,H)] + \beta_1^{RT}\left[\frac{1}{2}-\frac{1}{\pi}\tan^{-1}[\beta_2^{RT}(Logt-Log[a_{RT}(T,H)]-\beta_3^{RT})]\right]$$

The time-temperature shift factor implies that there exists a rescaled time t', also known as "reduced" time, given by

$$t' = \frac{t}{a_{RT}(T)}$$

where a_{RT}(T) is the "horizontal" shift corresponding to the temperature T. The storage modulus at time t' and reference temperature T_{RT} is related as in the equation

$$E_s(t,T) = b_{RT}(T)E_s(t',T_{RT})$$

where b_{RT}(T) is the "vertical" shift factor. Hence, the storage modulus is easily found at any time and temperature from a reference curve, if the corresponding shift factors are known. Horizontal and verticle shift factors are determined via the regression. The verticle shift value was close to unity. Therefore, the application of TTSF to Cycom 977-2 does not require the introduction of a vertical shift, since the regression yields b_{RT}(T) values very close to unity for the whole temperature range considered. However, this was not the case with horizontal shift factor, a_{RT}(T) as the values for a_{RT}(T) varied significantly. Therefore, Arrhenius relation was applied with two different activation energy levels.

4. Verify that the "shifted" data collapse on the master curve

The "wet" coupons considered in this work had moisture content of about 1%. Using the data plotted in Figure 4 for wet coupons, the regression yields again a unit value for the vertical shift factor, b_{RT}(T). However, the horizontal shift factor, $a_{RT}(T)$ is highly affected by the moisture content, as shown in Figure 7. On average, the horizontal shift factor is halved by moisture content of 1% for temperatures exceeding 60°C. Below this temperature, the difference becomes even larger, approaching one order of magnitude. Hence, humidity accelerates the material creep, as expected. In Figure 7, it is also worth observing that the glass transition temperature, corresponding to a sudden drop of the horizontal shift factor, decreases with the moisture content. Therefore, humidity plays an important role in the material creep accelaration.

In Figures 5 and 6, Matlab routine was programmed for plotting the master curve making use of "nlinfit" command which estimates the coefficients of a nonlinear regression function, using least squares estimation. The master curve for dry test case was used as "reference" and the values for $ß_1^{RT}$, $ß_2^{RT}$ and $ß_3^{RT}$ as used in the expression were found to be 10.072, 3.51 and 5.991, respectively. This is very clear from the Figures 5 and 6 that the data fit very well to the master curve for the dry CFRP samples.

It is worth noting that the effect of humidity is accounted for in the activation energies and characteristic temperatures in the expression of the horizontal shift factor. It is clear that the storage modulus of coupons which were kept in water and absorbed water (i.e. wet samples) decreased at higher rates than those samples which were not kept in water (i.e. dry samples) as expected, with temperature approaching glass transition region. The temperature at which 30–50 carbon chains start to move. At the glass transition temperature, the amorphous regions experience transition from rigid state to more flexible state making the temperature at the border of the solid state to rubbery state. This can be concluded that the plasiticization effect of the absorbed water has led to a higher rate decrease in storage modulus for the wet samples.

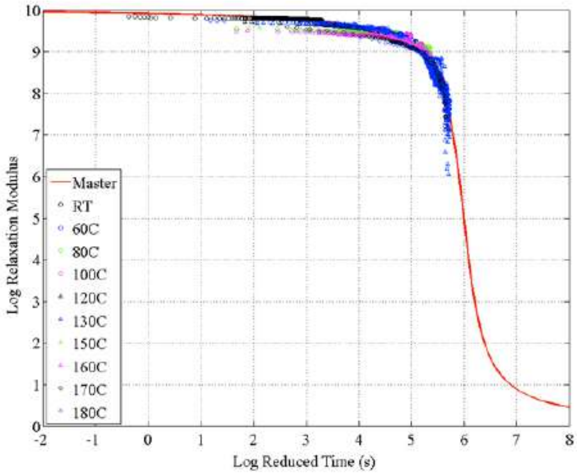


Figure 5. Master curve of creep relaxation modulus (dry samples)

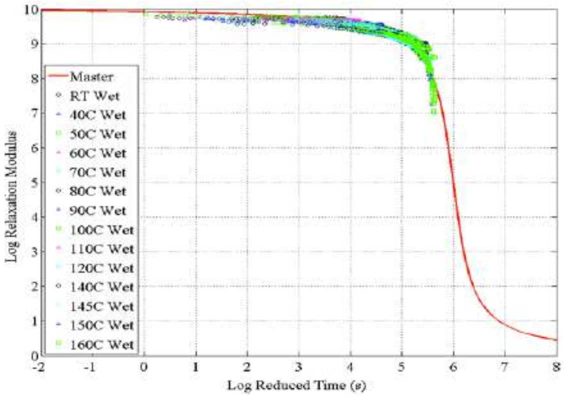


Figure 6. Master curve of storage modulus (wet samples)

When the experimental data for 1% H_2O were plotted; Arrehnius type regression with two different slopes was obtained as given in the expressions below and shown in Figure 7.

$$a_{RT}(T,H) = \frac{\Delta H_1}{R}\left[\frac{1}{T} - \frac{1}{T_0}\right] \qquad T \leq T_1$$

$$a_{RT}(T,H) = \frac{\Delta H_2}{R}\left[\frac{1}{T} - \frac{1}{T_2}\right] \qquad T \geq T_1$$

A sharp drop of the horizontal temperature shift is noticed close to the glass transition temperature. This has to be attributed to the fact that the material comes into a "rubbery" like state due to this temperature transition and because of the materials being in rubbery state it will further enhance creep effect creep. It is also worth observing that the glass transition temperature, corresponding to a sudden drop of the horizontal shift factor, decreases more with the moisture content.

It is not clear why a "kink" like appearance has observed in the horizontal shift factor for T =T_1 . This might be attributed to the fact that the transition temperature T_1 corresponds to a phase transition of the blend as 977-2 is an epoxy blended with a thermoplastic polymer for toughening purposes.

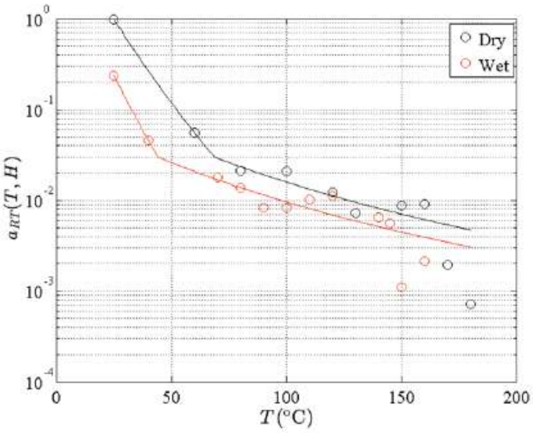


Figure 7 Horizontal time-temperature shift factor for experimental data of the dry and wet IM7/977-2

The master curves obtained for the storage modulus allow predicting the strength of the material in the full range of environmental conditions considered. The key outcome of this research is that the visco-elastic response of the material has a massive impact on the strength and that environmental effects strongly influence the visco-elastic response. It has also been observed that the shift factor associated with temperature and humidity can be suitably described by characteristic activation energies, which are inherent material properties.The residual strength properties of the fully saturated material have been found to be extremely low, particularly in terms of transverse tension. This poses significant challenges for the design of composite structures and highlights the importance of predicting the actual moisture content in service, as well as inhibiting the moisture ingress in composite materials via suitable surface protections.

5. CONCLUISON

In this study, time-temperature shift factor (TTSF) has been applied to study of the temprature and water absorption effect on the long-term viscoelastic response of IM7-carbon/977-2 epoxy composite materials.

The 3-point bending tests on dry as well as wet coupons using DMA show that the water uptake by the samples have affected the storage modulus, *E*s, of the IM7-carbon/977-2 epoxy composite at temperatures below Tg. It has been shown that the storage modulus of the wet samples were affected more severely and decreased at higher rates than the dry samples, with temperature approaching the glass transition region. This higher rate decrease in storage modulus for the wet coupons can be associated with plasiticization effect. However, no significant vertical shift factor is obtained from the investigated samples analysis. When the data were plotted for horizontal shift factor then Arrehnius type regression with two different slopes was obtained. All the experimental data for the dry as well as wet coupons fit very well by using the principle of TTSF and it is concluded that the properties of these composites can be relatively easily predicted through the application of TTSF by conducting traditional short-term laboratory tests. Future work for assessing the real hazard posed by environmental factors on the durability of composites in service should be focussed on the characterisation/prediction of the effects of cyclic temperature and moisture. This study has proven the viability of the timetemperature-humidity shift principles for steady environmental conditions.

6. REFERENCES

1. Hafiz, T. A., “Influence of Temperature and Moisture on the Compressive Strength of Carbon Fiber Reinforced Polymers”, ***International Journal of Engineering, Transactions B: Applications,*** Vol. 33, No. 5, (2020), 916-922. DOI: 10.5829/IJE.2020.33.05B.24
2. Reeder, J.R. “Prediction of Long-Term Strength of Thermoplastic Composites Using Time- Temperature Superposition”, (2002), NASA/TM-2002-211781
3. Yeow, Y. T., D. H. Morris and H. F. Brinson. “Time-Temperature Behavior of a Unidirectional Graphite/Epoxy Composite,” in Composite Materials: Testing Design (fifth Conference), ASTM STP 674, S. W. Tsai, Ed. Philadelphia: American Society for Testing and Materials (ASTM), (1979), 263-281. https://doi.org/10.1520/STP36913S
4. Rodriquez, P. I. (1993). “On the Design of Structural Components Using Materials with Time-Dependent Properties,” NASA TM108428.
5. Miyano, Y., Kanemitsu, M., Kunio, T. and Kuhn, H. (1986). “Role of Matrix Resin on Fracture Strengths of Unidirectional CFRP”, ***Journal of Composite Materials***, Vol. 20, 520-538. https://doi.org/10.1177%2F002199838602000602
6. Miyano, Y., McMurray, M. K., Kitade, N., Nakada, M. and Mohri, M. (1994). “Role of Matrix Resin on the Flexural Static Behavior of Unidirectional Pitch-based Carbon Fiber Laminates”, ***Advanced Composite Materials,*** Vol. 4, 87-99. https://doi.org/10.1163/156855194X00222
7. Miyano, Y., McMurray, M. K., Enyama, J. and Nakada, M. (1994). “Loading Rate and Temperature Dependence on Flexural Fatigue Behavior of a Satin Woven CFRP Laminate”, ***Journal of Composite Materials***, Vol. 28, (1994), 1250-1260. https://doi.org/10.1177%2F002199839402801305
8. Miyano, Y., Nakada, M., Kudoh H. and Muki, R. (1999). “Prediction of tensile fatigue life under temperature environment for unidirectional CFRP”, ***Advanced Composite Materials***, Vol. 4, (1999), 235-246. https://doi.org/10.1163/156855199X00236
9. Miyano, Y., Nakada, M., Watanabe, N., Murase, T. and Muki, R. “Time–temperature superposition principle for tensile and compressive strengths of unidirectional CFRP”, Proceedings of 2003 SEM Annual Conference & Exposition on Experimental and Applied Mechanics (SEM 2003), Charlotte: 147. (2003).
10. Rosen, B.W. (1964). “Tensile failure of fibrous composites”, The American Institute of Aeronautics and Astronautics (AIAA), Vol. 2, 1985-1991. https://doi.org/10.2514/3.2699
11. Rosen, B.W., Dow, N.F. and Hashin, Z. (1964). “Mechanical properties of fibrous composites”, NASA contract report, 31-35.
12. Zhao, Y., Wang, Z., Seah, L. Keey., Chai, G.B., “Long-Term Viscoelastic Response of E-glass/Bismaleimide Composite in Seawater Environment”, ***Applied Composite Materials***, Vol. 22, (2015), 693-709. https://doi.org/10.1007/s10443-014-9431-2
13. Miyano, Y., Nakada, M., Cai, H., “Formulation of long-term creep and fatigue strengths of polymer composites based on accelerated testing methodology”, ***Journal of Composites Materials***, Vol. 42, (2008), 1897-1920. https://doi.org/10.1177%2F0021998308093913
14. Cysne Barbosa, A.P., Ana, P.P.F., Guerra E.S.S., Arakaki K.F., Tosatto, M., Costa, M.C.B. and Melo, J.D.D., “Accelerated aging effects on carbon fiber/epoxy composites, ***Composites Part: B Engineering***, Vol. 110, (2017), 298-306. https://doi.org/10.1016/j.compositesb.2016.11.004
15. Korkees, F., Arnold, C. and Alston, S., “Water absorption and low-energy impact and their role in the failure of ±45° carbon fibre composites”, ***Polymer Composites***, Vol. 39, No. 8, (2018), 1-12. https://doi.org/10.1002/pc.24269

Persian Abstract

چکیده

خواص پلیمرهای تقویت شده با فیبر کربن (CFRP) تا حد زیادی در شرایط شدید محیطی تحت تأثیر قرار می گیرند. در این مقاله مطالعه تجربی به منظور تعیین پاسخ لمینتهای اپوکسی اکسید کربن IM7-karbon / 977-2 در شرایط مرطوب و شرایط مختلف ارائه شده است. از آزمون خمش ۳ نقطه ای کوتاه مدت با استفاده از آزمون خزش دینامیکی مکانیکی (DMA) برای آزمایش نمونه های خشک و اشباع شده در سطوح مختلف دما استفاده گردید. کوپنهای خشک در دمای اتاق RT و در دمای ٦۰- ۱۲۰ درجه سانتیگراد با افزایش ۲۰ درجه سانتیگراد و سپس در دمای ۱۳۰ درجه سانتیگراد ، ۱۵۰ تا ۱۸۰ درجه سانتیگراد با افزایش ۱۰ درجه سانتیگراد برای آزمایش بعدی مورد استفاده قرار گرفت. کوپنهای اشباع شده (مرطوب) در RT، ٤۰ - 120 ° C با ۱۰ درجه سانتیگراد افزایش دما برای هر آزمایش بعدی و در ١٤٥ درجه سانتیگراد ، ۱۵۰ درجه سانتیگراد و ١٦۰ درجه سانتیگراد مورد آزمایش قرار گرفتند. ضریب تغییر زمان دما (TTSF) استفاده شد و نشان داده شده است که رفتار ویسکوالاستیک لمینت های اپوکسی داخلی IM7-کربن / ۹۷۷-۲ درونشویه ، با استفاده از TTSF با دقت پیش بینی می شود. همچنین نشان داده شده است که تعیین رفتار ویسکوالاستیک در دماهای بالا به پیش بینی دما در زیر دمای انتقال شیشه با استفاده از TTSF کمک می کند. عمر طولانی مدت مواد با استفاده از TTSF با انجام آزمایشات سنتی آزمایشگاهی کوتاه مدت پیش بینی می شود.

International Journal of Engineering

Journal Homepage: www.ije.ir

Experimental Study of Management Systems for Emission Reduction in Ignition Engines

A. Fekari[a], N. Javani*[b], S. Jafarmadar[c]

[a] *Department of Mechanical Engineering, Bonab Branch, Islamic Azad University, Bonab, Iran*
[b] *Faculty of Mechanical Engineering, Yildiz Technical University, Besiktas, Istanbul, Turkey*
[c] *Department of Mechanical Engineering, Urmia University, Urmia, Iran*

PAPER INFO

Paper history:
Received 04 January 2020
Received in revised form 20 January 2020
Accepted 05 March 2020

Keywords:
Engine Management System
Exhaust Gases
Exhaust Gases Emission
Spark Advance

ABSTRACT

The present study aims to investigate the amount of exhaust gases emissions of a 4-cylinder gasoline-ignition engine. An experimental study of an ignition engine management system has been conducted for emissions optimization, using Winols specialized software. In order to achieve a steady state conditions in the experiments, the temperature of the water and engine oil before each test reached the engine's working temperature (90°C) to allow various parts of the engine to remain stable and the tests are performed in in-line engine operation. Two sets of tests with idle (850-900 rpm) and mid-range (2500 rpm) are considered. Experiments were performed for three identical engines with different mileages and obtained results were discussed. According to the obtained results, after applying changes to the engine management system, a 22% reduction in the unburned hydrocarbon emission in both cases was obtained. Furthermore, it is found that 31 and 5% reduction in carbon monoxide emissions in the idle and mid-range were obtained, respectively. As a result of applying these changes, there was a reduction of 1.4% in NOx emission in the idle case and a decrease in about 19% at 2500 rpm.

***doi**: 10.5829/ije.2020.33.07a.22*

NOMENCLATURE

AFR	Air-fuel ratio	P_{air}	Inlet manifold air pressure
AVL	Anstalt fur Verbrennungskraft machinen list	T_{air}	Intake manifold air temperature
TNM	Tarahan Novin Madar programmer	P_{ex}	Exhaust manifold gas pressure
PM	Particulate material	T_{ex}	Exhaust manifold gas temperature
RPM	Revolution per minute	MAP	Manifold air pressure

1. INTRODUCTION

Environmental protection is one of the most important global concerns and international organizations are trying to tackle the unfavorable side effects. In this regard, reducing air pollution has become a global concern, due to increased fuel consumption, especially in transportation section. Diesel engines and spark ignition for various vehicles are one of the main sources of air pollution. The exhaust gases of spark ignition engines include nitrogen oxide (NO) and NO_2 known as NOx, carbon monoxide (CO) and organic components, which include unburned or incomplete combusted hydrocarbons (HC). The amount of these gases depend on engine design and its operating conditions. The spark ignition engine is typically close to the stoichiometric state, or inclined to a rich fuel, to ensure smooth and reliable operation of the engine [1, 2].

Many researchers have conducted studies about the effect of gasoline fuel in internal combustion engines [3]. Jafarmadar et al. [4] stated that NOx emissions and soot formation depend intensely on engine temperature and equivalence ratio. There are specific controlling methods to reduce the emissions and to achieve acceptable average values [1]. Hsiao-Chi Chuang et al. [5] have studied the effects of vehicle exhaust particles

*Corresponding Author's Institutional Email: *njavani@yildiz.edu.tr* (N. Javani)

on cardiovascular health. Mohebbi et al. [6] have studied external exhaust gas recirculation (EGR) method and deduced that it will decrease combustion temperature and also decrease in NOx emissions. Kousoulidou et al. [7] used a Portable Emission Measurement System (PEMS) to develop and validate automobile emission factors. Six different vehicles were studied; under different driving standards and two different driving paths in the region of Lombardy, Italy were tested.

Twigg [8] has investigated the role of catalysts in controlling the exhaust emissions and development of the catalyst in this field. Lozhkin [9] conducted experiments on 13 gasoline-powered and 2 diesel-powered vehicles, ranging from Euro 0 to Euro 5 using gas analyzers to study the percentage of increased exhaust emissions in an environment with poor air-conditioning conditions. Vehicle emissions reduction using alternative fuels have been investigated for the aim of pollutant emission regulations [10]. Another effort to reduce emissions in diesel engines is the employment of light fuels (such as alcohol and diesel) as a diesel fuel supplement, which simultaneously reduces nitrogen oxide (NOx) and particulate material (PM) [11, 12].

There are a few experimental studies in the literature about the affecting parameters in the emissions of engine exhaust gases in the engine management systems. There have also been efforts in various researches to reduce one specific emission by using various additives in the fuel or using various equipment. The purpose of the present study is to reduce all engine exhaust emissions by considering engine performance at minimal cost and without adding extra equipment. Achieving this goal through software optimization of the engine management system makes this investigation different from the other studies in the literature.

2. SYSTEM DESCRIPTION

In the current study, by using an AVL gas analyser, a 4-cylinder ignition engine (EF7- without catalytic converter) exhaust emissions is evaluated. The EF7 series are designed jointly by Iran Khodro powertrain company (IPCO) and F.E.V Gmbh of Germany which was produced in 2008. Tested engine in the current study and the site weather properties are summarized in Table 1.

The main reason for selecting EF7 engine in the current study is its popularity and mass production in the regional market. Domestic automakers have problems to supply catalytic converter due its high costs. On the other hand, this engine is working on the majority of passenger cars without catalytic converter in Iran. Anstalt fur Verbrennungskraft machinen list (AVL) gas analysis device, engine testing device and Tarahan Novin Madar programmer (TNM) were employed for the related measurements. Furthermore, the effect of air-fuel ratio, spark advance, fuel injection time were investigated. In order to obtain their work maps from the engine control unit, the TNM programmer is used for analyzing the program code.

TABLE 1. Test engine properties and site weather properties

Parameter	Value
Model– Type	EF7 – Spark ignited
Each cylinder volume	412.433 cc
Cylinder number	4
Valve number of each cylinder	4
Max power	84 kW @ 6000 rpm
Max torque	155 N.m @ 3500-4500 rpm
Stroke	85 mm
Bore	78.6 mm
Fuel	Gasoline
Charging method	Natural aspirated (sequential)
Combustion ratio	11:1
Test site air pressure	870 mbar
Test site air temperature	300 K

3. RESEARCH METHODOLOGY

The TNM programmer device is used to read the program code of the engine management system by OBD connector. The extracted file format of TNM programmer device is binary that is capable of importing into Winols software. Winols software is also employed for modifications. Then, the air-fuel ratio, spark advance and reduction in starting temperature of the engine cooling fan are changed step by step and their effects are studied in detail to get optimized mode of reduced emissions by measuring the engine exhaust emission. The actual air-fuel ratio (AFR_{actual}) and ideal (stoichiometric) air-fuel ratio (AFR_{ideal}) is called equivalence air-fuel ratio or lambda (λ) which are defines as $AFR = \frac{m_{air}}{m_{fuel}}$ and $\lambda = \frac{AFR_{actual}}{AFR_{ideal}}$.

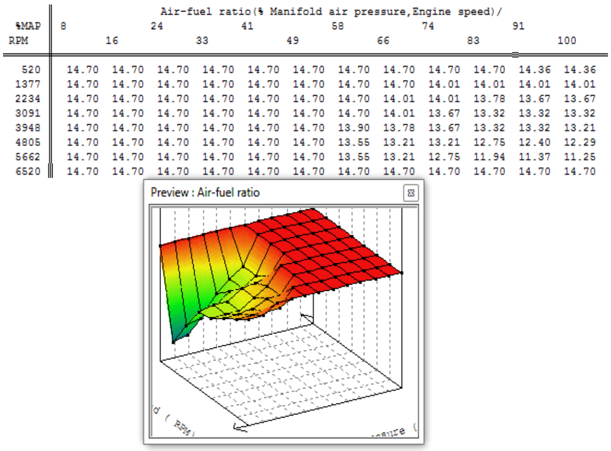
The spark ignition engine typically operates close to the stoichiometric ratio, or somehow close to a rich fuel to ensure the smooth and reliable operation of the engine [1]. The chemical reaction for the combustion of octane-air mixture is as follows [13]:

$$C_8H_{18}+20(O_2+3.76\ N2) = 8\ CO_2+9\ H_2O + 7.5\ O_2+ 75.2\ N_2 \quad (1)$$

Rich ratios produce less efficiency but the output power is higher power and the burning temperature is usually a lower temperature. The ratios above the stoichiometry are called lean, where the efficiency is higher, but more nitrogen oxide is produced [2]. The exact setting of the ignition timing has a significant effect on the engine performance and resulted emissions [14]. Due to the fact that the fuel burns at a limited speed and this combustion is also dependents on the environment, spark timing needs to be changed in the engine operating conditions, which also indicates the spark ignition values in the engine control unit. Also, changing spark timing will affect the engine emissions [15]. The maximum cylinder pressure and temperature and engine exhaust emissions are affected by ignition [16]. In the conducted tests, authors have concluded that by increasing the engine temperature, the brake specific fuel consumption will decrease. In addition, nitrogen oxide production with respect to the engine temperatures has been reported. It has also been pointed out that at high temperature of the engine results in increased friction losses and emission [17, 18].

4. RESULTS AND DISCUSSIONS

In the present study, an exhaust gas analyzer is employed to measure the exhaust emissions from an exhaust system of a 4-cylinder ignition engine of a gasoline EF7 model. Furthermore, the effect of air-fuel ratio, spark advance and fuel injection time are considered as the affecting parameters. To obtain their operation values based on the engine control unit, the TNM programmer is used to read the programming codes. Winols software is also used to change the values and thereafter, the effects of parameters variation are studied in detail.



```
                  Air-fuel ratio(% Manifold air pressure,Engine speed)/
 %MAP  ║ 8          24          41          58          74          91
RPM    ║      16          33          49          66          83          100
═══════╬═══════════════════════════════════════════════════════════════════════
   520 ║ 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.36 14.36
  1377 ║ 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.01 14.01 14.01 14.01
  2234 ║ 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.01 14.01 13.78 13.67 13.67
  3091 ║ 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.01 13.67 13.32 13.32 13.32
  3948 ║ 14.70 14.70 14.70 14.70 14.70 14.70 13.90 13.78 13.67 13.32 13.32 13.21
  4805 ║ 14.70 14.70 14.70 14.70 14.70 14.70 13.55 13.21 13.21 12.75 12.40 12.29
  5662 ║ 14.70 14.70 14.70 14.70 14.70 14.70 13.55 13.21 12.75 11.94 11.37 11.25
  6520 ║ 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70 14.70
```

Figure 1. Air-fuel ratio sample map of the studied engine management system in winols software

First, the required parameters in the original mode are obtained. Figure 1 shows the sample map of the air-fuel ratio of the considered engine for its management system study. This is for comparison and optimization purposes using the gas analyzer and the diagnostic tool of the engine and sensors installed in both the idle and the mid-range operating conditions (Table 2).

By increasing and decreasing the air-fuel ratio from 2.5 to 25 percent compared to the original state, the obtained values were compared with the original mode. According to Figure 2, it can be seen that by increasing the amount of air-fuel ratio from 2.5 to 12.5 percent, there is a noticeable decrease in the emissions of hydrocarbons, which can be due to the leaning of the air and fuel mixture. However, the problem here is the increase of NOx gas, which is due to an increase in combustion temperature.

By shifting to the values beyond 12.5%, there will be an increase in hydrocarbon gases. That is why values over the 12.5% range are ignored due to the engine performance (combustion instability starts which is measured with COVimep). Regarding the NOx variations, as it can be seen from Figure 2, it can be found that by increasing the amount of air-fuel ratio a significant increase in the emission of NOx; which is due to the leaning of the air and fuel mixture, causing high combustion temperature.

According to the results illustrated in Figure 3, it can be seen that by increasing the amount of air-fuel from 2.5 to 12.5 percent, the carbon monoxide will reduce. This is mainly because of the increasing oxygen and leaning of the mixture of air and fuel. When the air-fuel ratio exceeds 12.5 percent, the amount of carbon monoxide produced will continue to decrease, which is due to a disruption in the proper functioning of the engine. The carbon dioxide variations depend on

TABLE 2. Parameters in original mode for the considered engine

Parameter	Idle speed	Mid-range
AFR	14.21	13.77
Lambda	0.97	0.94
HC (ppm)	79	47
CO (%)	2.34	1.29
CO_2 (%)	14.24	15.62
O_2 (%)	0.71	0.68
NOx (ppm)	33	75
P_{air} (mbar)	400	280
T_{air} (K)	328.15	328.15
P_{ex} (mbar)	880	925
T_{air} (K)	553.15	553.15

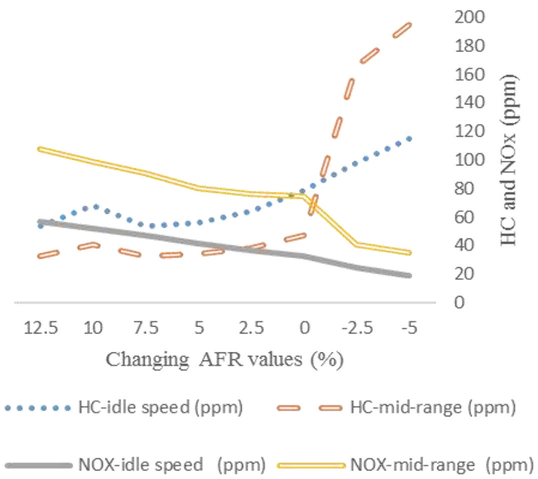


Figure 2. HC and NOx variations compared with changing the AFR values in engine idle speed and mid-range

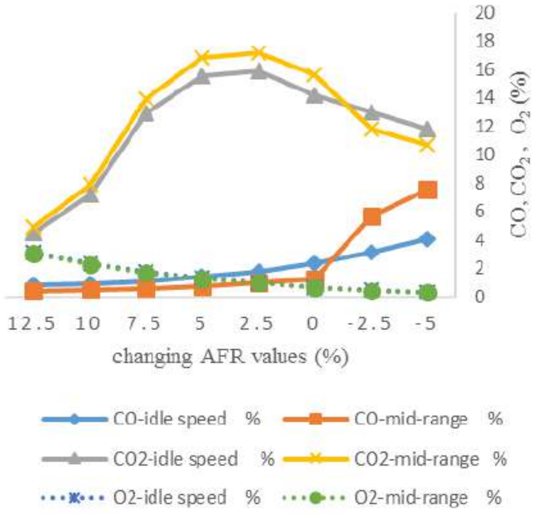
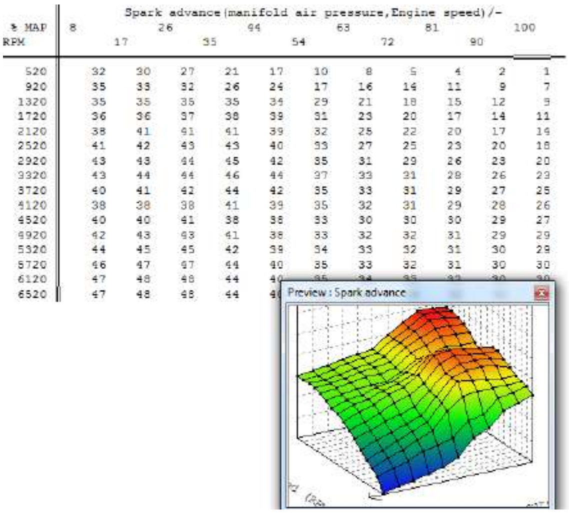


Figure 3. CO, CO_2 and O_2 variations comparing with changing AFR values in engine idle speed and mid-range

AFR and after getting a maximum value, decreases as it can be seen from Figure 3. With an increase in the air-fuel ratio, the amount of CO_2 initially increases due to an increase in the oxygen, but subsequently the carbon dioxide level decreases due to an increase in engine temperature and leaner mixture of air and fuel. Figure 4 shows the sample map of the spark advance of the considered engine for its management system study. From the testing modes, we selected modes 2.5 to 12.5 percent increase in the air-fuel ratio and a decrease of 2.5 percent, based on their reasonable emissions and engine performance. Also increasing and decreasing of the spark timing by 3, 6, and 9 percents, lead to 0.75, 1.5 and 3 degrees of spark timing advance, respectively.

Later on, the values obtained with the original case in terms of reducing the amount of emissions, especially unburned hydrocarbons, carbon monoxide and NOx were compared. It should be noted that by making changes beyond mentioned range in the engine, an impairment in the proper performance of the engine, including the creation of a knock in the engine are observed. Due to the resulted big dataset in different modes, here only number of conducted tests are explained.

Figure 5 reveals that with increasing spark advance, in cases where the air-fuel ratio increases, a reduction for emissions can be deduced. This is due to the enhancement in the time of the flame, resulting in a complete combustion.


Spark advance(manifold air pressure,Engine speed)/-

% MAP / RPM	8	17	26	35	44	54	63	72	81	90	100
520	32	30	27	21	17	10	8	5	4	2	1
920	35	33	32	26	24	17	16	14	11	9	7
1320	35	35	35	35	34	29	21	18	15	12	9
1720	36	36	37	38	39	31	23	20	17	14	11
2120	38	41	41	41	39	32	25	22	20	17	14
2520	41	42	43	43	40	33	27	25	23	20	18
2920	43	43	44	45	42	35	31	29	26	23	20
3320	43	44	44	46	44	37	33	31	28	26	23
3720	40	41	42	44	42	35	33	31	29	27	25
4120	38	38	38	41	39	35	32	31	29	28	26
4520	40	40	41	38	38	33	30	30	30	29	27
4920	42	43	43	41	38	33	32	32	31	29	29
5320	44	45	45	42	39	34	33	32	31	30	29
5720	46	47	47	44	40	35	33	32	31	30	30
6120	47	48	48	44	40	35	34	35	32	30	30
6520	47	48	48	44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Figure 4. Spark advance sample map of the studied engine management system in winols software

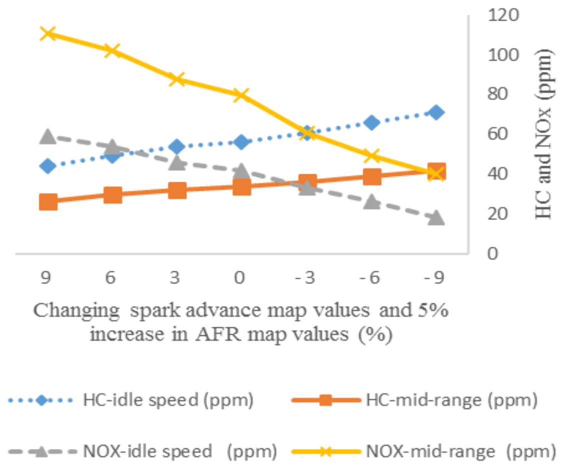


Figure 5. HC and NOx variations comparing with changing spark advance values and 5% increase in AFR values in engine idle speed and mid-range

However, the NOx production is increasing due to the increased combustion temperature. With reducing spark advance values, the amount of emissions of unburned hydrocarbons has begun to increase, which is due to the ignition delay, resulting in an incomplete combustion, but the production of NOx is decreasing because of reducing combustion temperature.

According to Figure 6, it is observed that with increasing spark advance values; there will be a reduction in the carbon monoxide emissions for the increased air-fuel ratios. This is mainly due to complete combustion and the reaction of the existing oxygen with carbon monoxide. Accordingly, with reducing spark advance values, the carbon monoxide emissions increase since the combustion is incomplete.

It can be inferred from Figure 7 that, in case where the air-fuel ratio is reduced, increasing spark advance will result in unburned hydrocarbons emission with decreased NOx generation. At the same time, due to the richness of fuel and air mixture, emission is more than the initial value. With reducing spark advance values, the rate of unburned hydrocarbons and nitrogen oxides production will increase which can be explained as the effect of combustion delay, resulting in incomplete ignition and increased combustion temperature. By reducing spark advance values, more increase in carbon monoxide emissions can be observed which is due to a delay in combustion and because of incomplete combustion. For the test conditions, the starting temperature of the engine cooling fan was reduced by 5 degrees. Figure 8 show the sample NOx variations comparing with changing AFR values in different test engines. By reducing the cooling fan starting temperature, a slight increase in the production of unburned hydrocarbons is observed. Also, there is a reduction in NOx production due to the engine temperature decrease. NOx generation is decreased for AFR ratios enhancement. The spark advance values changing and the richness or leaning of the fuel play a significant role in this regard.

No particular changes were observed in the production of carbon monoxide by reducing the starting temperature of the cooling fan. It should be noted that the difference observed in the idle speed and mid-range is due to the change in air-fuel ratio, the throttle position openness, air manifold pressure and spark advance. According to the results, in order to reduce all exhaust emissions, it is observed that in conditions of 5% increase in the air-fuel ratio values, a 3% reduction in the amount of spark advance values and 5°C reduction in the starting temperature of the engine cooling fan, the amount of unburned hydrocarbons, carbon monoxide and NOx will decrease 22.4, 31 and 1.4%, respectively, in the idle range and a decrease of 23.3, 5.3 and 18.8%,

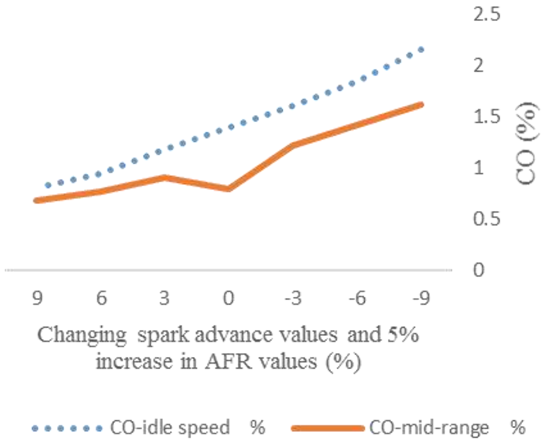


Figure 6. CO variations comparing with changing spark advance values and 5% increase in AFR values in engine idle speed and mid-range

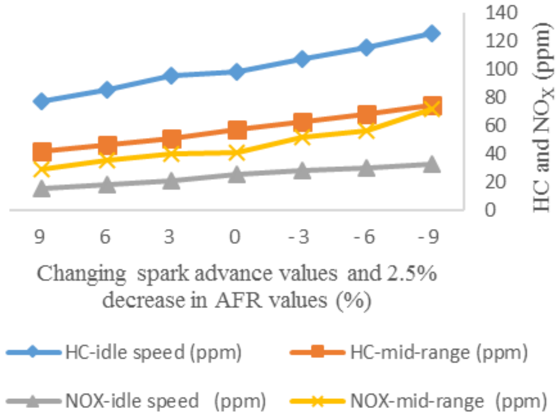


Figure 7. HC and NOx changing comparing with changing spark advance values and 2.5% decrease in AFR values in engine idle speed and mid-range

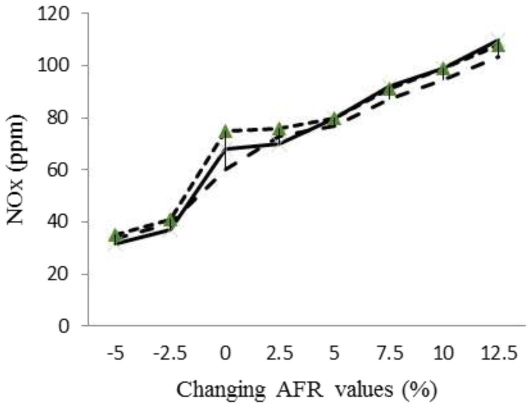


Figure 8. Sample NOx variations comparing with changing AFR values in different test engines mid-range

respectively at 2500 rpm. Also, in the conditions of 12.5% increase in air-fuel ratio values, 9% reduction in spark advance and 5°C reduction in the operating temperature of the engine cooling fan, then, the unburned hydrocarbons, carbon monoxides and NOx will decrease 13, 46.5 and 44.9%, in idle range and 14.2, 27.31 and 46.5 at 2500 rpm, respectively. Brake-specific fuel consumption (BSFC) reduces more than 1.4% in idle speed and 1.2% in medium speed in the experiments at the optimum mode.

5. CONCLUSION

An experimental study of the EF7 engine management system has been conducted for emissions optimization, using Winols specialized software, TNM programmer device, AVL gas analyzer and etc. Two cases of idle (850-900 rpm) and mid-range (2500 rpm) are considered in the tests and air-fuel ratio, spark advance and starting temperature of engine cooling fan are considered as effective parameters. The key findings of the investigation are summarized as follows:

- Increasing the amount of air-fuel ratio will decrease the emissions of hydrocarbons and carbon monoxide.
- Increasing spark advance leads to a reduction in unburned hydrocarbons and carbon monoxide while there the NOx emission will increase.
- Due to reducing the cooling fan starting temperature, a minor increase in the production of unburned hydrocarbons is observed. Also, there is a reduction in NOx production due to the engine temperature decrease.
- Optimized emission conditions can be achieved by applying software changes in engine management system at a minimal cost. Improving the equipment will obviously have favorable outcomes, but the results show that when there are no changes in the system, modifying and applying control unit's program with TNM or similar programmer devices will lead in decreased emissions of the engine.

6. REFERENCES

1. Fundamentals, I.C.E. and Heywood, J.B., "Mcgraw-hill series in mechanical engineering". 1988, Singapore. https://doi.org/10.1016/B978-0-12-809943-8.00002-9
2. Huang, D.-Y., Jang, J.-H., Lin, P.-H. and Chen, B.-H., "Effect of ignition timing on the emission of internal combustion engine with syngas containing hydrogen using a spark plug reformer system", ***Energy Procedia***, Vol. 105, (2017), 1570-1575. https://doi.org/10.1016/j.egypro.2017.03.499
3. Ebrahimi, R. and Mercier, M., "Experimental study of performance of spark ignition engine with gasoline and natural gas", ***International Journal of Engineering, Transactions B: Applications*** Vol. 24, No. 1, (2011), 65-74. http://www.ije.ir/article_71888.html
4. Jafarmadar, S. and Zehni, A., "Multi-dimensional modeling of the effects of spilt injection scheme on performance and emissions of idi diesel engines", ***International Journal of Engineering-Transactions C: Aspects***, Vol. 25, No. 2, (2012), 135. doi:10.5829/idosi.ije.2012.25.02c.07
5. Chuang, H.-C ,.Lin, L.-Y., Hsu, Y.-W., Ma, C.-M. and Chuang, K.-J., "In-car particles and cardiovascular health: An air conditioning-based intervention study", ***Science of the Total Environment***, Vol. 452, (2013), 309-313. https://doi.org/10.1016/j.scitotenv.2013.02.097
6. Mohebbi, A., Jafarmadar, S., Pashae ,J. and Shirnezhad, M., "Experimental studying of the effect of egr distribution on the combustion, emissions and performance in a turbocharged di diesel engine", ***International Journal of Engineering***, ***Transactions A: Basics***, Vol. 26, No. 1, (2013), 73-82. doi:10.5829/idosi.ije.2013.26.01a.10.
7. Kousoulidou, M., Fontaras, G., Ntziachristos, L., Bonnel, P., Samaras, Z. and Dilara, P., "Use of portable emissions measurement system (PEMS) for the development and validation of passenger car emission factors", ***Atmospheric Environment***, Vol. 64, (2013), 329-338. DOI: 10.1016/j.atmosenv.2012.09.062
8. Twigg, M.V., "Progress and future challenges in controlling automotive exhaust gas emissions", ***Applied Catalysis B: Environmental***, Vol. 70, No. 1-4, (2007), 2-15. https://doi.org/10.1016/j.apcatb.2006.02.029
9. Lozhkin, V., Lozhkina, O. and Dobromirov, V., "A study of air pollution by exhaust gases from cars in well courtyards of saint petersburg", ***Transportation Research Procedia***, Vol. 36, No., (2018), 453-458. https://doi.org/10.1016/j.trpro.2018.12.124
10. Datta, A. and Mandal, B.K., "A comprehensive review of biodiesel as an alternative fuel for compression ignition engine", ***Renewable and Sustainable Energy Reviews***, Vol. 57, (2016), 799-821. https://doi.org/10.1016/j.rser.2015.12.170
11. Ghadikolaei, M.A., "Effect of alcohol blend and fumigation on regulated and unregulated emissions of ic engines—a review", ***Renewable and Sustainable Energy Reviews***, Vol. 57, (2016), 1440-1495. doi: 10.1016/j.rser.2015.12.128
12. Imran, A., Varman, M., Masjuki, H.H. and Kalam, M.A., "Review on alcohol fumigation on diesel engine: A viable alternative dual fuel technology for satisfactory engine performance and reduction of environment concerning emission", ***Renewable and Sustainable Energy Reviews***, Vol. 26, (2013), 739-751. https://doi.org/10.1016/j.rser.2013.05.070
13. Madu, K., "Thermodynamic properties and heat generation potential of octane fuel, under stoichiometric condition", ***Equatorial Journal of Chemical Sciences***, Vol. 2, No. 2, (2018), 14-19. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3197885
14. Meshram, W.H., Chincholkar, S.P., Somankar, V.I. and Suryawanshi, J.G., "Ignition timing investigation on the performance and emissions of spark ignition engine fuelled with gasoline. International Conference On Emanations in Modern Technology and Engineering (ICEMTE-2017), ISSN: 2321-8169.
15. Tunka, L. and Polcar, A., "Effect of various ignition timings on combustion process and performance of gasoline engine", ***Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis***, Vol. 65, No. 2, (2017), 545-554. https://doi.org/10.11118/actaun201765020545
16. Gailis, M. and Pirs, V., "Effect of ignition timing on emissions of spark ignition engine using e85 fuel", ***Research for Rural Development***, Vol. 1, (2014), 212-218. https://www.researchgate.net/publication/287272988_Effect_of_

ignition_timing_on_emissions_of_spark_ignition_engine_using_E85_fuel

17. Hossain, A.K., Smith, D.I. and Davies, P.A., "Effects of engine cooling water temperature on performance and emission characteristics of a compression ignition engine operated with biofuel blend", ***Journal of Sustainable Development of Energy, Water and Environment Systems***, Vol. 5, No. 1, (2017), 46-57. DOI: 10.13044/j.sdewes.d5.0132

18. Virale, A.G. and Nitnaware, P.T., "Experimental analysis of jacket cooling of si engine and study of operating parameters and emissions", ***International Journal of Advances in Engineering & Technology***, Vol. 10, No. 1, (2017), 113. ISSN: 22311963

Persian Abstract

چکیده

مطالعه حاضر با هدف بررسی میزان آلایندگی گازهای خروجی یک موتور احتراقی ٤ سیلندر بنزینی انجام شده است. یک مطالعه تجربی از یک سیستم مدیریت موتور احتراقی برای بهینه سازی آلایندگی ها، با استفاده از نرم افزار تخصصی Winols انجام شده است. به منظور دستیابی به شرایط پایدار در تست ها، دمای آب و روغن موتور قبل از هر آزمایش به دمای کاری موتور (۹۰ درجه سانتیگراد) می رسد تا قطعات مختلف موتور به حالت پایا برسند.. دو حالت دور آیدل (۸۵۰-۹۰۰ دور در دقیقه) و متوسط (۲۵۰۰ دور در دقیقه) در نظر گرفته شده است. آزمایشات برای سه موتور یکسان با کارکردهای مختلف انجام می شود و نتایج به دست آمده مورد بحث قرار می گیرد. با توجه به نتایج به دست آمده، پس از اعمال تغییرات در سیستم مدیریت موتور، در هر دو دور مذکور، شاهد کاهش ۲۲ درصدی آلاینده های هیدروکربن هستیم. علاوه بر این، میزان آلاینده ی مونوکسید کربن ۳۱٪ و ۵٪ در دور آیدل و متوسط کاهش یافت. در نتیجه اعمال این تغییرات، کاهش میزان آلاینده ی NOx در دور آیدل ۱.٤٪ و در دور ۲۵۰۰ حدود ۱۹٪ کاهش می یابد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Effect of Non-uniform Magnetic Field on Non-newtonian Fluid Separation in a Diffuser

S. M. Moghimi, M. Abbasi*, M. Khaki Jamei, D. D. Ganji

Department of Mechanical Engineering, Sari Branch, Islamic Azad University, Sari, Iran

PAPER INFO

Paper history:
Received 11 February 2020
Received in revised form 18 April 2020
Accepted 11 June 2020

Keywords:
Diffuser
Non-uniform Magnetic Field
Separation Point
Second Grade

ABSTRACT

The purpose of the present study is to investigate the boundary layer separation point in a magnetohydrodynamics diffuser. As an innovation, the Re value on the separation point is determined for the non-Newtonian fluid flow under the influence of the non-uniform magnetic field due to an electrical solenoid, in an empirical case. The governing equations including continuity and momentum are solved by applying the semi-analytical collocation method (C.M.). The analysis revealed that for specific values of De from 0.4 to 1.6, α from 20° to 2.5° and Ha from zero to 8, the Re value on the separation point is increased from 52.94 to 1862.78; thus, the boundary layer separation postponed. Furthermore, the impact of the magnetic field intensity on the separation point is analyzed from the physical point of view. It is observed the wall shear stress increases by increasing magnetic field intensity that leads to delaying the boundary layer separation.

***doi**: 10.5829/ije.2020.33.07a.23*

NOMENCLATURE

u_{max}	Inlet velocity (m/s)	Ha	Hartmann number
u_r	r-component of the velocity (m/s)	**Greek Symbols**	
r	Coordinate in the direction of flow (m)	σ	Electrical conductivity $(\Omega.m)^{-1}$
P	Pressure (Pa)	α	Diffuser half angle (radian)
Re	Reynolds number	θ	Angle between center line and wall (radian)
t	Time (s)	ρ	Density (kg/m^3)
T	Rivlin Ericksin tensors	η	Arbitrary variable
I_o	Identity tensor	$F(\eta)$	Dimensionless velocity
R	Residual	α_1	First material constants
De	Deborah number	α_2	Second material constants
J	Density of electric current (A/m^2)	τ	Shear stress (pa)
I	Electric current (A)	μ_o	Permeability of free space (Tm/A)
B	Total magnetic field (T)	μ_r	Relative permeability
B_0	External magnetic field (T)	μ	Permeability of a specific medium (Tm/A)
B	Induced magnetic field (T)	μ_f	Dynamical viscosity (pa.s)
E	Electric field (V/m)		

1. INTRODUCTION

The interaction of the magnetic field on the elec-trically conducting fluid flow is called magnetohy-drodynamics (MHD). Investigation of the way the magnetic field affects the boundary layer is an interesting subject for researchers. The study of non-Newtonian fluid flow has been appealing many interests, due to its numerous industrial and engineering applications including polymer technology, condensed matter physics, astrophysics, geophysics, environmental, biophysics and molten plastics, etc. Furthermore, according to

*Corresponding Author Email: *mmortezaabbasi@gmail.com* (M. Abbasi)

Please cite this article as: S. M. Moghimi, M. Abbasi, M. Khaki Jamei, D. D. Ganji, Effect of Non-uniform Magnetic Field on Non-newtonian Fluid Separation in a Diffuser, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1354-1363

literature, the subject of the MHD fluid flows have been considered in many research works using different methods [1-5], and fluid flow in the nozzle has been investigated with different methods as well [6, 7]. Makinde [8] investigated the effect of arbitrary magnetic Reynolds number on a steady flow of an incompressible conducting viscous liquid in divergent channels under the influence of an externally applied homogeneous magnetic field. Nourazar et al. [9] studied the MHD flow of non-Newtonian Casson fluid in a stretching/shrinking divergent channel. Their results showed that by increasing the stretching parameter or Reynolds number, due to additional drag force acting on the plate at large values of stretching parameter, the velocity increased. Ara et al. [10] investigated the non-Newtonian fluid flow between the non-parallel walls. It was found that although the angular velocity of micro constituents in the flow increases, the fluid velocity decreases, as the vortex viscosity parameter values augmentation is associated with divergent channel. In addition, the spin gradient viscosity and micro-inertia density both enhance the micro rotation profiles. Umar Khan et al. [11] revealed that Dufour and Soret influence on the second-grade flow in diverging channels with stretchable walls. The results showed that the maximum velocity is obtained in the middle section of the divergent channel. Hayat et al. [12] studied the effects of a second-grade fluid flow in a divergent channel, and showed that the dimensionless velocity is a function of Reynolds number and divergent angle which reduces the velocity. However, when Deborah number increases, the velocity is also increased. The effect of non-uniform magnetic intensity on separation in a diffuser with Newtonian fluid flow is provided by Moghimi et al. [13]. The results indicated that by increasing the magnetic field intensity, the Lorentz force increased; then, the separation point was delayed.

According to the above discussion, although the numerous researches have focused on the MHD and non-Newtonian flows, but the investigation on the flow separation and determination of the separation point for the MHD non-Newtonian flows in a diffuser imposed with the non-uniform magnetic field intensity has not been studied yet. Hence, the aim and novelty of the present study is to compute the separation point for the MHD flows of a second-grade non-Newtonian fluids in divergent channels in the presence of the non-uniform magnetic field for different values of Re. The semi-analytical collocation method (C.M.) is applied for solving the problem. In addition, the effect of problem parameters on the separation point is presented graphically. Hereafter, the problem will be introduced and governing equations presented. Then, the solution and collocation method will be explained. Thereafter, the results will discussed, and finally, conclusions will be provided.

2 .PROBLEM STATEMENT

A diffuser with half angle α subjected to a non-uniform external magnetic field is shown in Figure 1. A laminar, incompressible, steady and second-grade non-Newtonian fluid flows through the diffuser.

The r and theta (θ) axes are defined as usual polar coordinates. The r-axis corresponds to the direction of the flow and lies on centerline. The flow velocity distribution is considered. The general governing equations for the defined problem are introduced according to literature [11, 14, 15]:

$$\vec{\nabla}.\vec{V}=0 \tag{1}$$

$$\rho\frac{D\vec{V}}{Dt}=-\vec{\nabla}S+\vec{J}\times B \tag{2}$$

$$\vec{J}=\sigma(\vec{E}+\vec{V}\times\vec{B}) \tag{3}$$

$$S=-PI_o+\tau \tag{4}$$

$$\tau=\mu_f\,T_1+\alpha_1\,T_2+\alpha_2\,T_1^2 \tag{5}$$

$$T_1=\nabla V+\nabla V^T \tag{6a}$$

$$T_2=\frac{D\vec{V}}{Dt}+T_1\,\nabla V+\nabla V^T\,T_1 \tag{6b}$$

Equations (1) to (3) are continuity, momentum and Ohm's law, respectively. $\vec{V}$, $\vec{J}$, E, σ, $I_\circ$, μ_f, τ, T_1, T_2, ∇, α_1 and α_2 are the velocity vector, the electric current density, the electric field intensity, electrical conductivity, the pressure, identity tensor, the dynamic viscosity, the shear stress, first and second Rivlin Ericksin tensors, Laplacian operator and material constants, respectively [11, 16]. The total magnetic field is B=B$_o$+b, where B$_o$ and b are the external and induced

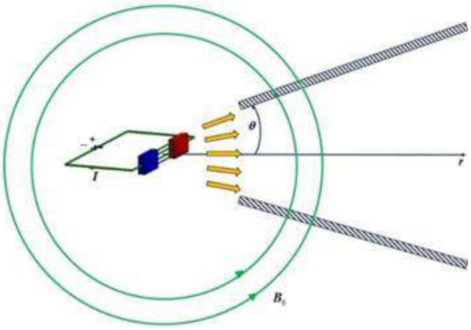


Figure 1. Schematic of a diffuser under the influence of a non-uniform magnetic field

magnetic fields, respectively. In the problem at hand, the magnetic Reynolds number is assumed small enough, so that the induced magnetic field can be neglected [15].

The cell electric current enters a solenoid, including parallel wires, by a distributer. The wires current is collected by a collector and returned back to the cell. Due to the electrical current in solenoid wires, a magnetic field is generated around the solenoid with intensity proportional to the distance from solenoid center, as shown in Figure 1. The applied magnetic field (B_0) can be obtained as:

$$B_{\circ} = \frac{\mu I}{2\pi r}\ , \quad \mu = \mu_{\circ} . \mu_r \tag{7}$$

where μ, μ_r and $\mu_{\circ}$ are the specific medium permeability, the relative permeability, and the permeability of the free space, respectively. $\mu_{\circ}$ is constant and equal to $4\pi\times10^{7}$ (Tm/A), r the distance from the center of solenoid and *I* the electric current. According to Equation (7), the magnetic field intensity is a function of the radius and current, thus, it can be regulated by changing the electrical current magnitude.

From Equations (1) and (2), the continuity and momentum equations in r and θ direction are given as [11, 16]:

$$\frac{1}{r}\frac{\partial r u_r}{\partial r} = 0 \tag{8a}$$

$$\begin{aligned} & u_r \frac{\partial u_r}{\partial r} = -\frac{\partial P}{\rho \partial r} + \frac{\mu_f}{\rho}(2\frac{\partial^2 u_r}{\partial r^2} + \frac{2}{r}\frac{\partial u_r}{\partial r} \\ & + \frac{1}{r^2}\frac{\partial^2 u_r}{\partial \theta^2} - \frac{2u_r}{r^2}) - \frac{\alpha_1}{\rho}(\frac{2u_r}{r}\frac{\partial^2 u_r}{\partial r^2} + \\ & 2\frac{\partial u_r}{\partial r}\frac{\partial^2 u_r}{\partial r^2} - \frac{1}{r^3}(\frac{\partial u_r}{\partial \theta})^2 + 2u\frac{\partial^3 u_r}{\partial r^3} - \\ & \frac{1}{r^2}\frac{\partial u_r}{\partial r}\frac{\partial^2 u_r}{\partial r \partial \theta}) - \frac{\alpha_1}{\rho}(\frac{1}{r^2}\frac{\partial^2 u_r}{\partial r \partial \theta} + \\ & \frac{\partial u_r}{\partial r}\frac{\partial^2 u_r}{\partial \theta^2} - \frac{2u_r}{r^3}\frac{\partial^2 u_r}{\partial \theta^2} - \frac{u_r}{r^2}\frac{\partial^3 u_r}{\partial r^2 \partial \theta} + \\ & 2\frac{{u_r}^3}{r^3} - \frac{2u_r}{r^2}\frac{\partial u_r}{\partial r}) - \frac{\sigma}{\rho}u_r {B_{\circ}}^2 \end{aligned} \tag{8b}$$

$$\begin{aligned} & u_r \frac{\partial u_r}{\partial r} = -\frac{\partial P}{\rho \partial r} + \frac{\mu_f}{\rho}(2\frac{\partial^2 u_r}{\partial r^2} + \frac{2}{r}\frac{\partial u_r}{\partial r} \\ & + \frac{1}{r^2}\frac{\partial^2 u_r}{\partial \theta^2}\frac{2u_r}{r^2}) - \frac{\alpha_1}{\rho}(\frac{2u_r}{r}\frac{\partial^2 u_r}{\partial r^2} + \\ & 2\frac{\partial u_r}{\partial r}\frac{\partial^2 u_r}{\partial r^2} - \frac{1}{r^3}(\frac{\partial u_r}{\partial \theta})^2 + 2u\frac{\partial^3 u_r}{\partial r^3} - \\ & \frac{1}{r^2}\frac{\partial u_r}{\partial r}\frac{\partial^2 u_r}{\partial r \partial \theta}) - \frac{\alpha_1}{\rho}(\frac{1}{r^2}\frac{\partial^2 u_r}{\partial r \partial \theta} + \frac{\partial u_r}{\partial r}\frac{\partial^2 u_r}{\partial \theta^2} - \\ & \frac{2u_r}{r^3}\frac{\partial^2 u_r}{\partial \theta^2} - \frac{u_r}{r^2}\frac{\partial^3 u_r}{\partial r^2 \partial \theta} + 2\frac{{u_r}^3}{r^3} - \frac{2u_r}{r^2}\frac{\partial u_r}{\partial r}) \\ & -\frac{\sigma}{\rho}u_r {B_{\circ}}^2 \end{aligned} \tag{8c}$$

where u_r is a radial velocity in direction of flow through diffuser, and μ_f is the dynamic viscosity. The boundary conditions are defined as follows:

$$\theta = 0: \qquad u_r = u_{max} \tag{9a}$$

$$\theta = 0: \qquad \frac{\partial u_r}{\partial \theta} = 0 \tag{9b}$$

$$\theta = \alpha, -\alpha: \qquad u_r = 0 \tag{9c}$$

Integrating Equation (8a) and using Equation (9b) gives:

$$u_r = \frac{1}{r} f(\theta) \tag{10}$$

To eliminate pressure gradient, Equations (8b) and (8c) are differentiated with respect to θ and r, respectively, and then subtracting from each other. Afterwards, by employing the similarity dimensionless variables as [17]:

$$\eta = \frac{\theta}{\alpha} \tag{11a}$$

$$F(\eta) = \frac{u_r}{u_{max}} = \frac{1}{r}\frac{f(\theta)}{u_{max}} \tag{11b}$$

In addition, after some simplifications, the governing equations reduces to:

$$\begin{aligned} & F''' + 2\alpha \operatorname{Re} FF' + (4\alpha^2 - Ha^2)F' + \\ & 4De(FF''' + 4\alpha^2 FF') = 0 \end{aligned} \tag{12a}$$

$$\operatorname{Re} = \frac{u_{max} r\alpha}{\nu} \tag{12b}$$

$$Ha = B_{\circ} r\alpha\sqrt{\frac{\sigma}{\mu_f}} \tag{12c}$$

$$De = -\frac{\alpha_1 u_{max}}{r\mu_f} \tag{12d}$$

Here, Re denotes the Reynolds number, Ha is the Hartmann number and De represents Deborah number which are dimensionless physical quantities. In the present investigation, the skin friction coefficient is the intended physical quantity which is calculated as [12]:

$$C_f = \frac{1}{\operatorname{Re}} F'(1) \tag{13}$$

The Re value pertinent to separation is shown with Re_{Sep}. It happens when C_f becomes zero. On the other hand, according to Equation (13), when $F'(1) = 0$ the separation occurs.

Furthermore, the boundary conditions are transformed into the dimensionless form as follows:

$$\eta = 0: \qquad F(0) = 1 \tag{14a}$$

$$\eta = 1: \qquad F(1) = 0 \tag{14b}$$

$$\eta = 0: \qquad F^{'}(0) = 0 \tag{14c}$$

2. 1. The Problem Solution Two different methods are conducted to solve for the current problem. In the first method, the numerical method is employed to achieve dimensionless velocity distribution data. By using trial and error, the Re value related to $F'(1) = 0$ for different values of Ha, De and α is obtained and called $\mathrm{Re}_{\mathrm{Sep}}$.

In the other method, the Collocation Method (C.M.) [17-19], described in the following section, is used to obtain the dimensionless velocity distribution relation. The velocity distribution is plotted using the obtain relation. It is expected to see a reverse flow to occur for some Re values greater than $\mathrm{Re}_{\mathrm{Sep}}$.

2. 1. 1. Collocation Method The semi-analytical collocation method is used for solving the governing equation. In this method, a trial function is approximated which satisfies the boundary conditions. By inserting the approximated trial function into the governing equations, it is expected that the residuals to be zero. However, due to the approximate solution, the residual does not exactly equal to zero, but it approaches zero. Obviously, as much as the residual value becomes closer to zero, the approximated function has more accuracy.

In this investigation, a polynomial trial function is applied to approximate the solution as follows:

$$\begin{aligned} F(\eta) = {} & c_{18}\eta^{18} + c_{17}\eta^{17} + c_{16}\eta^{16} + c_{15}\eta^{15} + \\ & c_{14}\eta^{14} + c_{13}\eta^{13} + c_{12}\eta^{12} + c_{11}\eta^{11} + c_{10}\eta^{10} + \\ & c_{9}\eta^{9} + c_{8}\eta^{8} + c_{7}\eta^{7} + c_{6}\eta^{6} + c_{5}\eta^{5} + c_{4}\eta^{4} \\ & + c_{3}\eta^{3} + c_{2}\eta^{2} + c_{1}\eta^{1} + c_{0} \end{aligned} \tag{15}$$

Equation (15) has 19 unknown coefficients which have to be determined. First, in order that Equation (15) satisfied the boundary conditions described in Equation (14), the boundary conditions are inserted in Equation (15) which yields:

$$c_0 = 1 \tag{16a}$$

$$c_1 = 0 \tag{16b}$$

$$1 + c_2 + c_3 + c_4 + \ldots = 0 \tag{16c}$$

Thus, Equation (15) can be rewritten as:

$$\begin{aligned} F(\eta) = {} & c_{18}\eta^{18} + c_{17}\eta^{17} + c_{16}\eta^{16} + c_{15}\eta^{15} \\ & + c_{14}\eta^{14} + c_{13}\eta^{13} + c_{12}\eta^{12} + c_{11}\eta^{11} + \\ & c_{10}\eta^{10} + c_{9}\eta^{9} + c_{8}\eta^{8} + c_{7}\eta^{7} + c_{6}\eta^{6} + \\ & c_{5}\eta^{5} + c_{4}\eta^{4} + c_{3}\eta^{3} + c_{2}\eta^{2} + 1 \end{aligned} \tag{17}$$

Therefore, Equation (17) is the approximated trial function which satisfied the problem boundary conditions. As can be seen in Equation (17), there is 16 unknown coefficients. Hence, 16 equations are required for determining the coefficient. For determining the unknown coefficient, the following procedure is conducted using the collocation method. First, Equation (17) is replaced in Equation (12a) for the case of Re=350, De=0.8, Ha=4 and $\alpha = 5^{\circ}$, which gives:

$$\begin{aligned} & R(c_2, c_3, \ldots, c_{18}, \eta) = 6c_3 + 24c_4\eta + \\ & 60c_5\eta^2 + 120c_6\eta^3 + 210c_7\eta^4 + \\ & 336c_8\eta^5 + \ldots + 4896c_{18}\eta^{15} + 2\alpha\,\mathrm{Re} \\ & (2c_2\eta^1 + 3c_3\eta^2 + \ldots + 18c_{18}\eta^{17})(1 + \\ & c_2\eta^2 + c_3\eta^3 + \ldots + c_{18}\eta^{18}) + (4\alpha^2 - \\ & \mathrm{Ha}^2)(2c_2\eta^1 + 3c_3\eta^2 + \ldots + 17c_{17}\eta^{16} + \\ & 18c_{18}\eta^{17}) + 4\mathrm{De}[(1 + c_2\eta^2 + c_3\eta^3 + \ldots + \\ & c_{18}\eta^{18})(6c_3 + 24c_4\eta + \ldots + 4896c_{18}\eta^{15}) \\ & + 4\alpha^2(2c_2\eta^1 + 3c_3\eta^2 + \ldots + 18c_{18}\eta^{17})(1 + \\ & c_2\eta^2 + c_3\eta^3 + \ldots + c_{18}\eta^{18})] \end{aligned} \tag{18}$$

It is known that η domain is between zero and one $(0 < \eta < 1)$. In the collocation method, the domain is divided into finite intervals the values of which depends on the number of the trial function for unknown coefficients. Therefore, the domain is divided into 16 intervals as the residual become zero:

$$R(\frac{1}{17}) = 0,\ R(\frac{2}{17}) = 0,\ \ldots,\ R(\frac{16}{17}) = 0 \tag{19}$$

By inserting each value of the interval in Equation (18), 16 equations are obtained. By solving the set of equations, the unknown coefficients are computed and the approximated trial function is calculated as:

$$\begin{aligned} & F(\eta) = \\ & -2.567900139\eta^{18} + 16.03918792\eta^{17} \\ & -41.62834552\eta^{16} + 52.06904837\eta^{15} \\ & -11.88853492\eta^{14} \text{-} 68.20484287\eta^{13} \\ & +130.2747688\eta^{12} - 133.8834906\eta^{11} \\ & +93.23789869\eta^{10} - 47.17891722\eta^{9} \\ & +18.12684323\eta^{8} - 4.992575696\eta^{7} \\ & -0.04929084927\eta^{6} - 0.16058893\eta^{5} \\ & +1.844560443\eta^{4} - 0.0012939369\eta^{3} \\ & -2.036526779\eta^{2} + 1 \end{aligned} \tag{20}$$

Here, $F(\eta)$ is the dimensionless velocity distribution, according to Equation (10b).

2. 1. 1. Numerical Method In order to validate the results of the collocation method, the numerical 4th order Runge-Kutta method [20] is used to solve Equation (12a).Thus, Equation (12a) is rewritten as:

$$F^{'''}(\eta) = -\frac{A_1 F(\eta)F^{'}(\eta) + A_2\, F^{'}(\eta)}{(1 + A_3\, F(\eta))} \tag{21}$$

where A1, A2 and A3 are constant values equal to ($2\alpha Re+16\alpha^2 De$), ($4\alpha^2$-Ha^2) and (4De), respectively. Now, to solve Equation (12a) by numerical method, it can be rewritten as a set of equations:

$$F'(\eta) = G(\eta), F(0) = 1 \tag{22a}$$

$$G'(\eta) = H(\eta), G(0) = 0 \tag{22b}$$

$$H'(\eta) = -\frac{A_1 F(\eta)\, G(\eta) + A_2\, G(\eta)}{(1 + A_3\, F(\eta))}, H(0) = ? \tag{22c}$$

To solve the above set of equations, three initial values of F(0), G(0) and H(0) are required which H(0) is not determined. Thus, we assume a value for H(0) and then solve using the shooting method as far as the boundary condition F(1)=0 is satisfied. Now, the set of equations (22) is solved using the 4th order Runge-Kutta method.

No reference in literature exactly coincides with the defined problem (especially the non-uniform magnetic intensity); hence, for validation, the results from two methods will be compared. Besides, in Figure 2, for verification, the results for a special case (Ha=0, De=0.8, Re=40, and α=5°) is plotted and compared with the that of Ref. [12]. It can be seen that there is good agreement between th two. The results for the dimensionless velocity obtained from the two different solutions, i.e. the numerical method and C.M., are presented in Table 1. The comparison of the results shows the reasonable agreement between them.

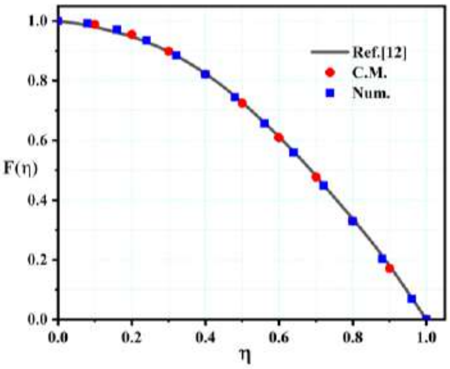


Figure 2. Comparison of the present study result with Ref. [12] result

TABLE 1. The dimensionless velocity distribution F(η) comparison for C.M. and numerical method at Ha=4, De=0.8, Re_{Sep}=568.31 and α=5°

η	Re=350		
	C.M.	**Numerical**	**Diff.**
0.0	1.0000	1.0000	0.0000
0.1	0.9798	0.9798	0.0000
0.2	0.9213	0.9214	0.0001
0.3	0.8307	0.8307	0.0000
0.4	0.7166	0.7167	0.0001
0.5	0.5893	0.5894	0.0001
0.6	0.4585	0.4585	0.0000
0.7	0.3318	0.3318	0.0000
0.8	0.2139	0.2138	0.0001
0.9	0.1048	0.1048	0.0000
1.0	0.0000	0.0000	0.0000
	Re=450		
0.0	1.0000	1.0000	0.0000
0.1	0.9731	0.9731	0.0000
0.2	0.8963	0.8963	0.0000
0.3	0.7805	0.7806	0.0001
0.4	0.6411	0.6410	0.0001
0.5	0.4948	0.4949	0.0001
0.6	0.3563	0.3563	0.0000
0.7	0.2355	0.2355	0.0000
0.8	0.1371	0.1371	0.0000
0.9	0.0601	0.0601	0.0000
1.0	0.0000	0.0000	0.0000
	Re=568.31		
0.0	1.0000	1.0000	0.0000
0.1	0.9635	0.9635	0.0000
0.2	0.8611	0.8612	0.0001
0.3	0.7118	0.7119	0.0001
0.4	0.5414	0.5414	0.0000
0.5	0.3754	0.3755	0.0001
0.6	0.2334	0.2335	0.0001
0.7	0.1255	0.1256	0.0001
0.8	0.0532	0.0532	0.0000
0.9	0.0126	0.0128	0.0002
1.0	0.0000	0.0000	0.0000
	Re=650		
0.0	1.0000	1.0000	0.0000
0.1	0.9562	0.9562	0.0000

0.2	0.8348	0.8348	0.0000
0.3	0.6618	0.6618	0.0000
0.4	0.4715	0.4715	0.0000
0.5	0.2958	0.2957	0.0001
0.6	0.1557	0.1557	0.0000
0.7	0.0595	0.0594	0.0001
0.8	0.0047	0.0047	0.0000
0.9	-0.0139	-0.0140	0.0001
1.0	0.0000	0.0000	0.0000

2. 2. Results and Discussion

In Tables 2a-2d, the results from C.M. are presented. It can be seen that as Ha increases while De and α decrease, the value of Re, in which the separation occurs, increases. For example, in Table 2b, when the Ha increases from zero to 4 (at De=0.8 and α=5$^{o)}$, the value of Re_{Sep} is increased about 77.2%, and the De increases from 0.4 to 1.6 (at Ha=4 and α=5$^{o)}$, the value of Re_{Sep} is increased about 62.4%. In Table 2a-2d α decreases from 20^o to 2.5^o at Ha=4 and De=0.8, the value of Re_{Sep} is increased about 72.4%. It should be noted that the Re values smaller than Re_{Sep} (for the specified problem), the separation does not occur.

TABLE 2. The $Re_{Sep.}$ values for some Ha, and De at the separation point ($F'(1)=0$) for different angle of diffuser

Ha \ De	0.4	0.8	1.2	1.6
α=2.5°				
0	445.65	642.65	834.79	1024.27
1	473.56	672.11	865.32	1055.62
2	558.55	761.61	957.96	1150.67
3	704.58	914.78	1116.06	1312.52
4	918.74	1138.2	1345.74	1546.88
5	1211.27	1442.0	1656.59	1862.78
6	1593.83	1839.4	-	-
α=5°				
0	222.30	320.57	416.41	510.94
1	236.25	335.29	431.68	526.61
2	278.73	380.04	477.99	574.13
3	351.74	456.61	557.03	655.04
4	458.80	568.31	671.85	772.20
5	605.04	720.19	827.25	930.14
6	796.29	918.89	1029.97	1135.39
7	1036.9	1171.17	1287.93	1396.34
8	1327.9	1480.98	1607.72	1721.39
α=10°				
0	110.09	158.77	206.25	253.08
1	117.06	166.13	213.88	260.91
2	138.29	188.49	237.03	284.66
3	174.77	226.76	276.52	325.09
4	228.26	282.57	333.90	383.65
5	301.33	358.46	411.55	462.57
6	396.90	457.75	512.86	565.14
7	517.19	583.83	641.77	695.54
8	662.65	738.69	801.60	857.99
α=20°				
0	52.94	76.37	99.22	121.76
1	56.41	80.04	103.03	125.67
2	67.00	91.19	114.58	137.52
3	85.19	110.28	134.28	157.70
4	111.86	138.12	162.91	186.91
5	148.29	175.97	201.64	226.29
6	195.97	225.49	252.17	277.46
7	256.02	288.41	316.49	342.52
8	328.68	365.73	396.28	423.59

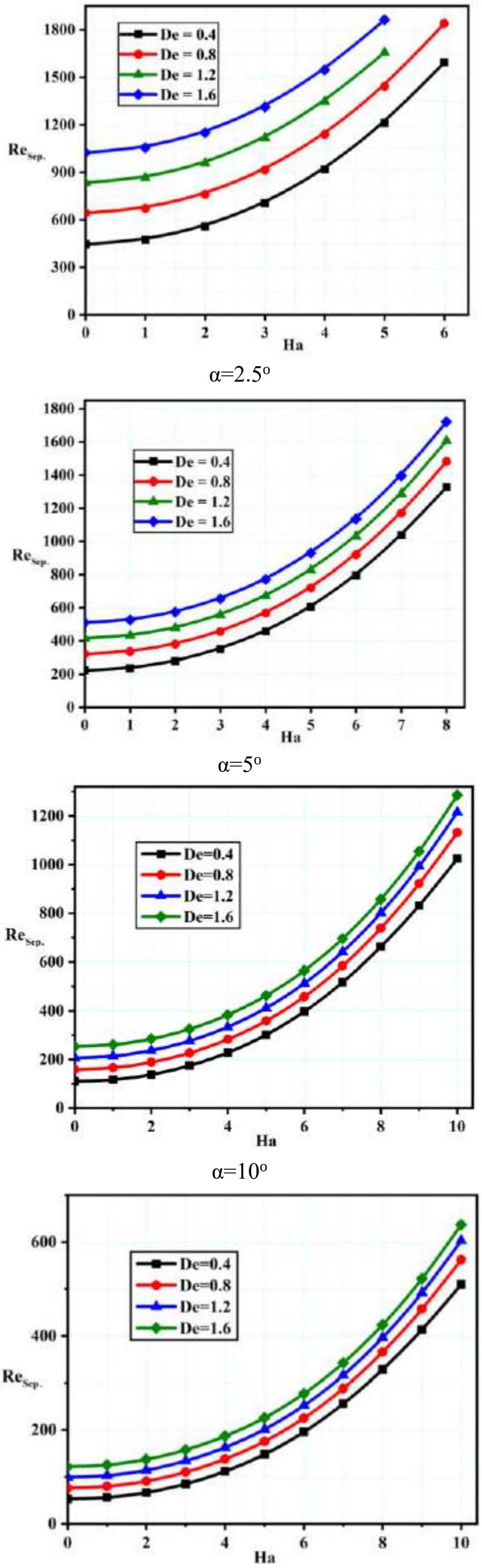
Figure 3 demonstrates the variation of Re_{Sep} with respect to Ha for different De values at four channel divergent angles. The second-grade fluid definition, μ_f, which is the first viscosity coefficient (the dynamic viscosity), has higher weight than the second viscosity coefficient α_1. With increasing μ_f while De decreases, the momentum near the wall decreases more rapidly, leads to the earlier separation.

The influence of different dimensionless quantities, the Re, De, Ha, and α are shown in Figures 4 and 5.

As seen in Equation (12), the skin friction coefficient (the wall shear stress) is proportional to $F'(\eta)$.The effect of Re on $F(\eta)$ and $F'(\eta)$ are shown in Figures 4a and 5a, respectively. As shown in Table 2b, the separation Reynolds number value is 568.31. Therefore, as the Reynolds number. reaches this value, $F(\eta)$ is getting tangent to the wall,while on the contrary, $F'(\eta)$ value becomes zero on the wall. When the Reynolds number continues to increase, the reverse flow develops.

Figures 4b and 5b show variations of $F(\eta)$ and $F'(\eta)$ versus De. These figures are for Re=568.31, Ha=4 and $\alpha=5^\circ$.

According to Table 2b, the separation occurs at De=0.8. With increasing De, the shear stress at the wall increases and the momentum along the flow decreases



α=2.5°

α=5°

α=10°

α=20°

Figure 3. The effect of Ha and De on separation point for different angle of diffuser

and approaches zero. As shown in Figure 4b, the flow over the wall gets tangent and according to Figure 5b, becomes zero. Proceeding along the wall, the reverse flow develops.

The effect of Hartman number on $F(\eta)$ and $F'(\eta)$ are shown in Figures 4c and 5c, for De=0.8, Re=568.31 and $\alpha = 5^{\circ}$. According to Table 2b, the separation occurs at Ha=4 and the reverse flow is observed for Ha less than 4. As shown in Figure 4c, for Ha=4 the flow velocity is tangent to the wall, and according to Figure 5c, the value of $F'(1)$ is zero. With increasing Ha, i.e. magnetic field intensity augmentation, the momentum near the wall increases such that the separation will not occur. As can be seen, the reverse flow will not happen. In physical aspect, the magnetic Lorentz force is proportional to the velocity and its value in the centerline region is more than in the wall region. Thus, for a special mass flow value, when the magnetic field intensity increases, the momentum near the wall increases, which leads to the separation.

It is expected that with increasing the channel divergent angle, the separation is expedited. This phenomenon can be observed in Figures 4d and 5d. According to Table 2c, for Ha=4, Re=282.57 and De=0.8, the separation happens at $\alpha = 10^{\circ}$. With increasing the channel divergent angel, the reverse flow occurs.

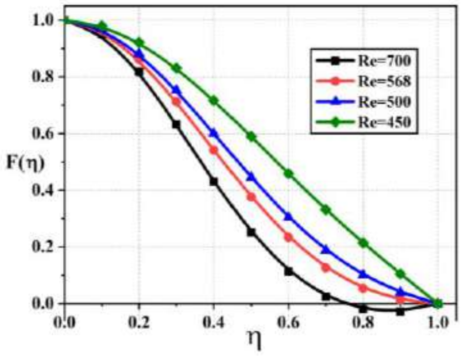


Figure 4a. The Re effect on F(η) for De=0.8, Ha=4 and α=5°

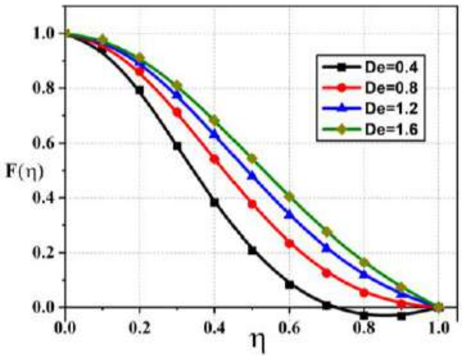


Figure 4b. Debora number effect on F(η) for Re=568.31, Ha=4 and α=5°

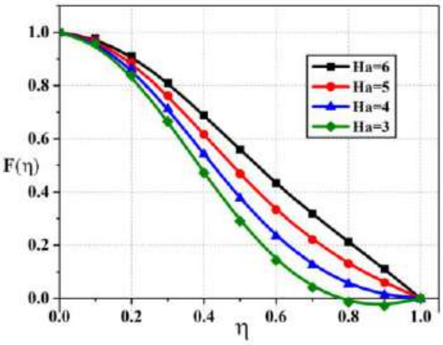


Figure 4c. Hartmann number effect F(η) for Re=568.31, De=0.8 and α=5°

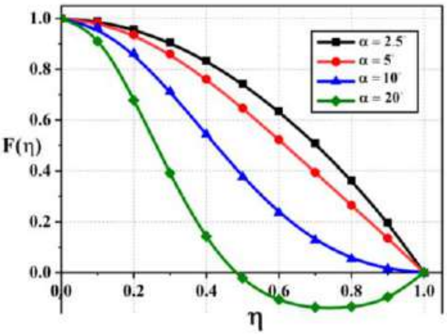


Figure 4d. Half divergent angle effect on F(η) for Re=282.57, De=0.8 and Ha=4

Figure 6 shows the Cf values versus Re for De=0.8 at different values of Ha . Where the Cf value becomes zero, the Re value corresponds to the one for separation. The effect of Ha augmentation in this case is similar for Figures 4c and 5c, as discussed earlier.

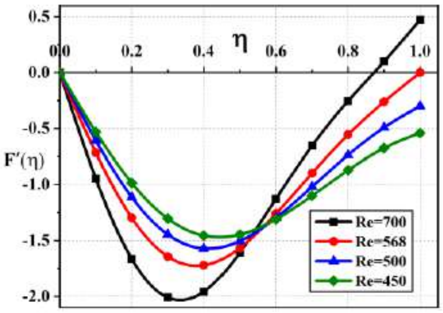


Figure 5a. The Re effect on F'(η) for De=0.8, Ha=4 and α=5°

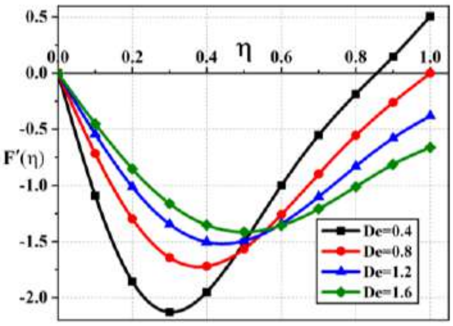


Figure 5b. Debora number effect on F'(η) for Re=568.31, Ha=4 and α=5°

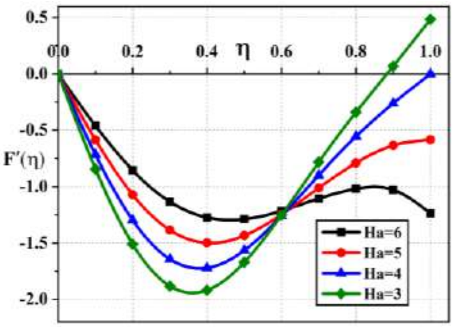


Figure 5c. Hartmann number effect F'(η) for Re=568.31, De=0.8 and α=5°

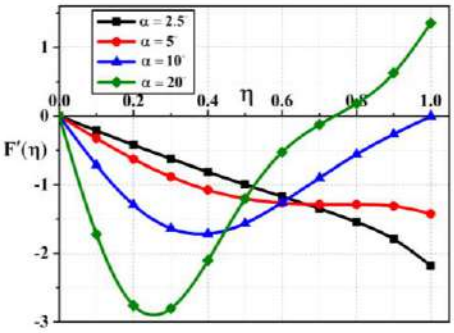


Figure 5d. Half divergent angle effect on F'(η) for Re=282.57, De=0.8 and Ha=4

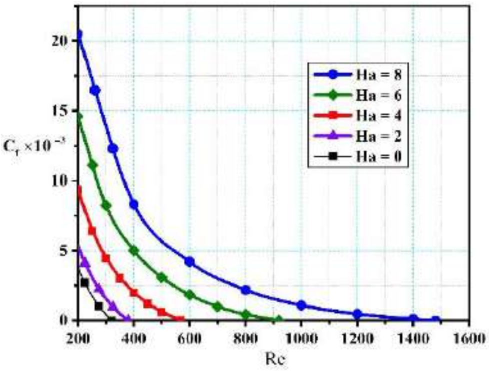


Figure 6. The Ha effect on Friction Factor C_f for De=0.8 and α=5°

3. CONCLUSIONS

In this article, the separation point of a non-Newtonian second-grade fluid flow through a diffuser was investigated. The imposed magnetic field intensity and viscous fluid degree effects on the flow were studied. As an innovation, for non-Newtonian fluid flow, the Re value on the separation point is determined under the influence of the non-uniform magnetic field due to an electrical solenoid, at an empirical case. The effect of material parameter of second-grade fluid and magnetic field intensity on the skin friction coefficient is opposite and by increasing of the magnetic field intensity (Ha), the Lorentz force increases which leads to delay in the separation. It was observed when $\alpha=5^{\circ}$ and Ha changes from 0 to 8, for De=0.4, Re_{sep} increases 498%, for De=0.8, Re_{sep} increases 362%, for De=1.2, Re_{sep} increases 286% and for De=1.6, Re_{sep} increases 231%. This augmentation was also observed for other values of α. Moreover, in a diffuser with certain conditions, by adjustable magnetic field intensity, one can increase the magnitude of the Lorentz force up to a level that the separation does not occur on the walls. Furthermore, the analysis indicated that the dimensionless velocity distribution becomes flatter when Ha, De and α increases. Besides the effects of Re, De and Ha on the skin friction coefficient C_f was presented.

4. REFERENCES

1. Cruze, D., Hemalatha, G., Jebadurai, S.V.S., Sarala, L., Tensing, D., and Christy, S.J.E., "A review on the magnetorheological fluid, damper and its applications for seismic mitigation", ***Civil Engineering Journal***, Vol.4, No. 12, (2018), 3058-3074. DOI: 10.28991/cej-03091220

2. Dirbude, S.B., and Maurya, V.K., "Effect of Uniform Magnetic Field on Melting at Various Rayleigh Numbers", ***Emerging Science Journal***, Vol. 3, No. 4, (2019), 263-273. DOI: 10.28991/esj-2019-01189

3. Ghadikolaei, S., Hosseinzadeh, K., and Ganji, D., "Analysis of unsteady MHD Eyring-Powell squeezing flow in stretching channel with considering thermal radiation and Joule heating effect using AGM", ***Case studies in thermal engineering***, Vol. 10, (2017), 579-594. DOI: 10.1016/j.csite.2017.11.004

4. Sears, F.W., Zemansky, M.W., Young, H.D., and Freedman, R.A., "Sears and Zemansky's University Physics: With Modern Physics", Addison-Wesley, New York, (2012). ISBN: 0321696867, 9780321696861

5. Taheri, M.H., Abbasi, M., and Jamei, M.K., "An integral method for the boundary layer of MHD non-Newtonian power-law fluid in the entrance region of channels", ***Journal of the Brazilian Society of Mechanical Sciences and Engineering***, Vol. 39, No. 10, (2017), 4177-4189. DOI: 10.1007/s40430-017-0887-5

6. Sengupta, A.R., Gupta, R., and Biswas, A., "Computational Fluid Dynamics Analysis of Stove Systems for Cooking and Drying of Muga Silk", ***Emerging Science Journal***, Vol. 3, No. 5, (2019), 285-292. DOI: 10.28991/esj-2019-01191

7. Su, C., and Cheng, Y.-h., "Numerical and Experimental Research on Convergence Angle of Wet Sprayer Nozzle", ***Civil Engineering Journal***, Vol. 4, No. 9, (2018), 1985-1995. DOI: 10.28991/cej-03091132

8. Makinde, O., "Effect of arbitrary magnetic Reynolds number on MHD flows in convergent-divergent channels", ***International Journal of Numerical Methods for Heat & Fluid Flow***, Vol. 18, No. 6, (2008), 697-707. DOI:10.1108/09615530810885524

9. Nourazar, S., Nazari-Golshan, A., and Soleymanpour, F., "On the expedient solution of the magneto-hydrodynamic Jeffery-Hamel flow of Casson fluid", ***Scientific reports***, Vol. 8, No. 1, (2018), 1-16. DOI: 10.1038/s41598-018-34778-w

10. Ara, A., Khan, N.A., Naz, F., Raja, M.A.Z., and Rubbab, Q., "Numerical simulation for Jeffery-Hamel flow and heat transfer of micropolar fluid based on differential evolution algorithm", ***AIP Advances***, Vol. 8, No. 1, (2018), 1-17. DOI: 10.1063/1.5011727

11. Khan, U., Ahmed, N., and Mohyud-Din, S.T., "Soret and Dufour effects on Jeffery-Hamel flow of second-grade fluid between convergent/divergent channel with stretchable walls", ***Results in physics***, Vol. 7, (2017), 361-372. DOI: 10.1016/j.rinp.2016.12.020

12. Hayat, T., Nawaz, M., and Sajid, M., "Effect of heat transfer on the flow of a second-grade fluid in divergent/convergent channel", ***International Journal for Numerical Methods in Fluids***, Vol. 64, No. 7, (2010), 761-776. DOI: 10.1002/fld.2170

13. Moghimi, S.M., Abbasi, M., Jamei, M.K., and Ganji, D.D., "Boundary-layer separation in circular diffuser flows in the presence of an external non-uniform magnetic field", ***Mechanical Sciences***, Vol. 11, No. 1, (2020), 39-48. DOI: 10.5194/ms-11-39-2020

14. Christov, I.C., "Stokes' first problem for some non-Newtonian fluids: Results and mistakes", ***Mechanics Research Communications***, Vol. 37, No. 8, (2010), 717-723. DOI:10.1016/j.mechrescom.2010.09.006

15. Roidt, M., and Cess, R., "An approximate analysis of laminar magnetohydrodynamic flow in the entrance region of a flat duct", ***Journal of Applied Mechanics***, Vol. 29, No. 1, (1962), 171-176. DOI: 10.1115/1.3636451

16. Khan, U., Ahmed, N., and Mohyud-Din, S.T., "Thermo-diffusion and diffusion-thermo effects on flow of second grade fluid between two inclined plane walls", ***Journal of Molecular Liquids***, Vol. 224, (2016), 1074-1082. DOI: 10.1016/j.molliq.2016.10.068

17. Kazemi, K.B., "Solving differential equations with least square and collocation methods", Master thesis, George Washington University, USA, (2004).

18. Ebrahimi, S., Abbasi, M., and Khaki, M., "Fully developed flow of third-grade fluid in the plane duct with convection on the walls", ***Iranian Journal of Science and Technology, Transactions of Mechanical Engineering***, Vol. 40, No. 4, (2016), 315-324. DOI: 10.1007/s40997-016-0031-7

19. Gavabari, R.H., Abbasi, M., Ganji, D., Rahimipetroudi, I., and Bozorgi, A., "Application of Galerkin and Collocation method to the electrohydrodynamic flow analysis in a circular cylindrical conduit", ***Journal of the Brazilian Society of Mechanical Sciences and Engineering***, Vol. 38, No. 8, (2016), 2327-2332. DOI: 10.1007/s40430-014-0283-3

20. Hussain, K., Ismail, F. and Senu, N., "Runge-Kutta Type Methods for Directly Solving Special Fourth-Order Ordinary Differential Equations", ***Mathematical Problem Engineering***, Vol. 2015, (2015), 1-11. DOI: 10.1155/2015/893763

Persian Abstract

چکیده

هدف از این مطالعه، بررسی نقطه‌ی جدایش لایه‌ی مرزی برای جریان هیدرودینامیک مغناطیسی در یک دیفیوزر است. به عنوان نوآوری، برای جریان سیال غیرنیوتونی تحت تأثیر میدان مغناطیسی نایک‌نواخت که توسط یک سلونوئید الکتریکی ایجاد شده، مقدار عدد رینولدز در نقطه‌ی جدایش تعیین می‌شود. معادلات حاکم شامل معادلات پیوستگی و ممنتوم با استفاده از روش نیمه‌تحلیلی تلفیقی حل می‌شوند. تحلیل نتایج نشان می‌دهد که برای مقادیر ویژه هنگامی که عدد دبورا از ۰/٤ تا ١/٦، نصف زاویه‌ی واگرایی از ۲۰ تا ۲/۵ درجه و عدد هارتمن از صفر تا ۸ تغییر می‌کند، مقدار عدد رینولدز نقطه‌ی جدایش از ۵۲/۹٤ به ۱۸٦۲/۷۸ افزایش یافته که سبب تاخیر در جدایش لایه‌ی مرزی می‌شود. علاوه بر این، از دیدگاه فیزیکی نیز تأثیر شدت میدان مغناطیسی بر جدایش لایه‌ی مرزی بررسی شده است. همچنین، مشاهده می‌شود که با افزایش شدت میدان مغناطیسی، تنش برشی روی دیوار افزایش می‌یابد که منجر به تاخیر در جدایش لایه‌ی مرزی می‌شود.

International Journal of Engineering

Journal Homepage: www.ije.ir

Meshing Error of Elliptic Cylinder Gear Based on Tooth Contact Analysis

D. Changbin*, L. Yongping, W. Yongqiao

School of Mechanical and Electronical Engineering, Lanzhou University of Technology, Lanzhou, China

PAPER INFO

Paper history:
Received 21 February 2020
Received in revised form 09 March 2020
Accepted 12 July 2020

Keywords:
Elliptical Cylinder Gear
Static Contact
Finite Element Analysis
Dynamic Meshing Characteristics
Meshing Error

ABSTRACT

In order to study the dynamic meshing characteristics of the elliptic cylinder gear and obtain the meshing error of the gear transmission system, the two-dimensional static contact analysis of the gear tooth surface is carried out using ANSYS software, and the key parts of the contact area of the tooth surface are determined. Then, the dynamic meshing model of the elliptic cylinder gear is established and the dynamic contact process under load is simulated by ANSYS LS-DYNA software. The distribution law of effective plastic strain, effective stress and pressure of the driving and driven wheels are obtained. On this basis, the distribution law of meshing error is obtained by calculation. The results show that the distribution of stress, strain and tooth surface pressure during tooth meshing is related to the position of the tooth on the elliptical pitch curve. The position of the tooth on the pitch curve and the load it bears has a certain influence on the meshing error. The results of this research can provide some guidance for subsequent study of transmission error of non-circular gears, gears modification and engineering applications.

doi: *10.5829/ije.2020.33.07a.24*

1. INTRODUCTION

Non-circular cylinder gears are distinguished from cylinder gears by their non-circular pitch curves. As one of the simplest non-circular cylinder gears, the elliptic cylinder gears are compact and widely used in robots, printers, fans and hydraulic motors [1]. In view of its wide range of applications, it is of practical value to study the meshing characteristics and meshing errors of elliptic cylinder gears. At present, static contact or incomplete gear model are mostly used to analyze the meshing characteristics of gears, while the tooth profiles of elliptic cylinder gears are different. In view of the above situation, this paper analyzes and establishes the dynamic meshing model of the elliptic cylinder gear to simulate the actual working condition of the gear under the load. Compared with static contact analysis, dynamic contact analysis is more accurate.

The small tooth profile error, manufacturing error, installation error and center-to-center error will cause the involute gear to generate a lot of vibration and noise during the meshing process [2], and those errors are transmitted and accumulated by teeth meshing. Therefore, it is necessary to study the meshing errors existing in the gear meshing process. As a kind of gear transmission error, the meshing error has received extensive attention in recent years, and a lot of research results have been accumulated for the gear transmission error. Among them, Sainte et al. [3], aimed at spur gear and helical gear, studied the relationship between the dynamic transmission error and the load it bears. Sungho et al.[4] obtained the dynamic transmission error of the gear by simulating the loaded tooth contact, and proposed the dynamic transmission error of the gear as a basis for fault diagnosis. Hotait et al. [5] studied the relationship between dynamic transmission error and dynamic stress coefficient in spur gear pairs. Wang et al. [6] studied the dynamic transmission error of eccentric cylinder gears under consideration of time-varying backlash caused by gear eccentricity and load variation. Yu et al. [7] proposed a double eccentric gear model for calculating the transmission error caused by the gear eccentricity error and the transmission error caused by

*Corresponding Author's Email: *lutdcb@126.com* (D. Changbin) the eccentricity error can be predicted by obtaining the transmission error of the designed gap compensation device. Deng et al.[8] proposed a model for synchronously measuring the gear transmission curve in the time-scale domain, and solved the jump error of the transmission error curve in the time domain and applied it to the measurement of the hypoid gear transmission error. Zou et al. [9] established an analytical calculation model for transmission error of the involute gear pairs with double eccentricity error, and studied the dynamic transmission error of the eccentric gear. Wu et al. [10] studied the effects of installation error, eccentricity error and meshing error on the uniform load characteristics of the system by establishing a nonlinear dynamic model of the planetary gear train. Chang et al. [11] studied the gear meshing error and pointed out that the deformation during the gear meshing can be divided into two parts: linear macroscopic deformation and nonlinear local deformation, and studied the influence of different error types and distribution forms on system vibration.

Most practical problems are difficult to get accurate solutions for. While the finite element method not only has high calculation accuracy, but also can adapt to a variety of complex shapes.Therefore, it has become an effective engineering analysis method. As far as the current dynamic analysis methods are concerned, most of them adopt the finite element simulation method [12-14]. The most widely used analysis software in the gear meshing analysis process is LS-DYNA, which is probably the most famous universal explicit dynamic analysis program in the world. It can simulate various complex problems in the real world, and is especially suitable for solving nonlinear dynamic shock problems such as high-speed collisions, explosions, and metal forming of various 2D and 3D nonlinear structures. At the same time, it can solve heat transfer, fluid, and fluid-solid coupling problems [15-17].

The above literatures are all concerned with the transmission error of gears, and there are few reports on the study of meshing errors between teeth. Generally, the gear transmission error is analyzed for the displacement or rotation angle between the input and output shafts, and the deformation between the shaft and the gear is not considered. Considering the meshing error of the gear pair, the problem of load deformation and meshing of the gear under the meshing error state is analyzed, which makes the gear meshing more in line with the actual situation. Therefore, the paper proposes that the difference of strain, stress and pressure at the same meshing unit between the driving and driven wheels is used as the meshing error. Taking a pair of elliptic cylinder gear as the research object, the meshing error of the elliptic cylinder gear is studied by simulating the actual meshing process of the gear.

This work is organized as follows. Section 2 introduces the mathematical model and finite element analysis theory of elliptic cylinder gear tooth surface. Section 3 introduces the static contact analysis of tooth surface. Section 4 proposes the dynamic meshing model of elliptic cylinder gear and introduces the dynamic contact analysis of gears and the distribution law of meshing error. Finally, Section 5 provides the conclusions. Figure 1 shows the flow chart of the research process.

2 ELLIPTIC CYLINDER GEAR TOOTH SURFACE MODEL AND FINITE ELEMENT ANALYSIS THEORY

2. 1. Elliptic Cylinder Gear Tooth Surface Model

The tooth surface of the elliptic cylinder gear is an involute surface, the end face of the tooth is an involute, and the tooth profile of the elliptic cylinder gear can be determined by the involute of the tooth profile. The vector equation of the tooth profile is [1]:

$$\boldsymbol{r}_f = \boldsymbol{r}_g + \boldsymbol{an} \quad (1)$$

where $\boldsymbol{r}_f$ is the radial diameter of any point n on the tooth profile. $\boldsymbol{r}_g$ is the radial diameter of the pitch curve at the intersection of the normal and pitch curve on point n of the tooth profile. $\boldsymbol{an}$ indicates that the direction is consistent with the normal direction of the tooth profile, and the length is equal to the section curve and the tooth profile vector of distance. The tooth profile of the elliptic cylinder gear can be divided into two points, which are higher than the pitch curve and

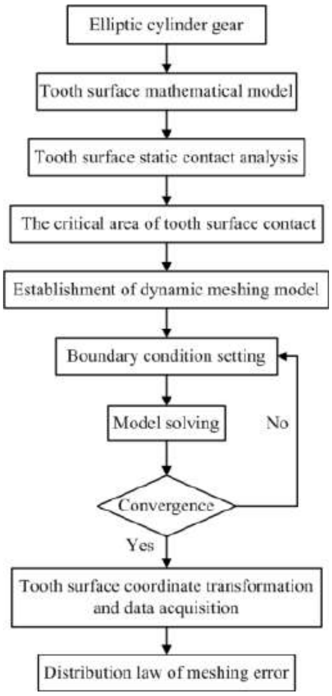

Elliptic cylinder gear
Tooth surface mathematical model
Tooth surface static contact analysis
The critical area of tooth surface contact
Establishment of dynamic meshing model
Boundary condition setting
Model solving
No
Convergence
Yes
Tooth surface coordinate transformation and data acquisition
Distribution law of meshing error

Figure 1. The flow chart of the research process

lower than the pitch curve. For the two-part tooth profile equation, there are different methods [1]. Figures 2 and 3 are the equations of the tooth profile above the pitch curve and below the pitch curve respectively, where the right tooth profile angle is $\theta - \mu \pm \alpha_u$, and the left tooth profile angle is $\mu - \theta \pm \alpha_u$ (+,- representing above the pitch curve point and below the pitch curve point, respectively).

The equation of the left and right tooth profiles is extended to the three-dimensional space to obtain the tooth surface equation of the elliptic cylinder gear, wherein the right tooth surface equation is:

$$\begin{cases} x_R = r_g \cos\theta \pm an\cos(\theta - \mu + \alpha_\mu) \\ y_R = r_g \sin\theta \pm an\sin(\theta - \mu + \alpha_\mu) \\ z_R = z_i \end{cases} \tag{2}$$

And the left tooth surface equation of the elliptic cylinder gear is:

$$\begin{cases} x_L = r_g \cos\theta \mp a'n'\cos(\mu - \theta - \alpha_u) \\ y_L = r_g \sin\theta \mp a'n'\sin(\mu - \theta - \alpha_u) \\ z_L = z_i \end{cases} \tag{3}$$

where, (x_R, y_R, z_R) represents the right gear surface coordinate, (x_L, y_L, z_L) represents the left gear surface coordinate, z_i refers to the direction of the tooth line and is equal to the width of the tooth, α_μ the profile pressure angle, μ the angle between the pitch curve diameter and the pitch curve tangent of point *a* [18].

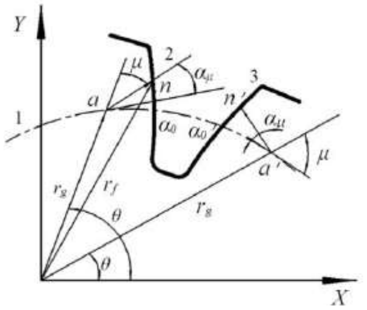


Figure 2. Tooth profile above the pitch curve

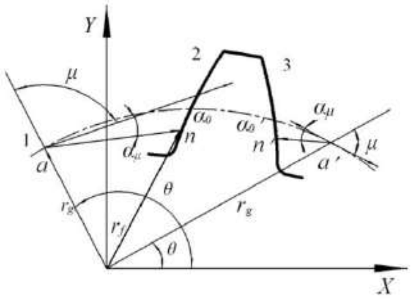


Figure 3. Tooth profile below the pitch curve

2. 2. Finite Element Analysis Theory

The basic motion equation of the gear system dynamics analysis is：

$$M\ddot{U} + C\dot{U} + KU = F \tag{4}$$

where M, C, K and F are the mass matrix, damping matrix, stiffness matrix. U, $\dot{U}$ and $\ddot{U}$ are the displacement vector, velocity, and acceleration of the nodes, respectively.

As one of the post-processing softwares of ANSYS LS-DYNA, LS-PREPOST uses the central difference method to solve the motion differential equation of dynamic problems. The essence of the central difference method is to replace the differential with the difference, namely:

$$\begin{cases} \ddot{U}_t = \dfrac{1}{\Delta t^2}(U_{t-\Delta t} - 2U_t + U_{t+\Delta t}) \\ \dot{U}_t = \dfrac{1}{2\Delta T}(-U_{t-\Delta t} + U_{t+\Delta t}) \end{cases} \tag{5}$$

Taking Equation (5) into dynamic differential Equation (4), the system of linear equations can be obtained

$$\bar{M}U_{t+\Delta t} = \bar{R}_t \tag{6}$$

$$\bar{M} = \frac{1}{\Delta t_2}M + \frac{1}{2\Delta t}C \tag{7}$$

$$\bar{R}_t = F_t - (K - \frac{2}{\Delta t^2}M)U_t - (\frac{1}{\Delta t^2}M - \frac{1}{2\Delta t}C)U_{t-\Delta t} \tag{8}$$

where $\bar{M}$ is the effective mass matrix and $\bar{R}_t$ is the effective load vector.

According to Equation (6), the state parameters of $t + \Delta t$ time can be calculated from the state quantities of $t - \Delta t$ and t time, which is the characteristic of display algorithm.

The time step of the display calculation is limited by the element size, and a reasonable and accurate finite element model needs to be established to improve the calculation accuracy [18]. In contrast, LS-PREPOST software is used to simulate the dynamic meshing process of the gear by establishing a non-linear contact unit in the meshing area. On this basis, gear deformation, transmission error, and gear stress can be obtained at the same time, which avoided the complicated writing and calculation of finite element programs, and can be applied here for research on the meshing error of elliptic cylindrical gear. Figure 4 shows the finite element model of elliptic cylinder gear meshing and the parameters of the elliptic cylinder gear are shown in Table 1.

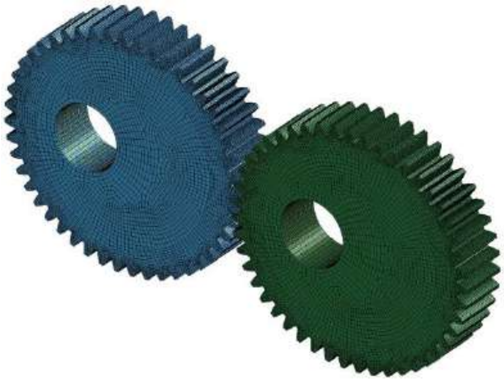


Figure 4. The meshing finite element model of elliptic cylinder gear

TABLE 1. Elliptic cylinder gear design parameters

Parameter	Value
Module m(mm)	3
Number of teeth z	47
Center distance a(mm)	145
Addendum coefficient *ha**	1
Top clearance coefficient *C**	0.25
Tooth width B(mm)	30
Eccentricity *e*	0.3287
Pressure angle(°)	20
Equation of pitch curves	$r = \frac{64.667}{1 \pm 0.3287 \cos\varphi}$

3. ANALYSIS OF STATIC CONTACT MESHING CHARACTERISTICS OF ELLIPTIC CYLINDER GEARS

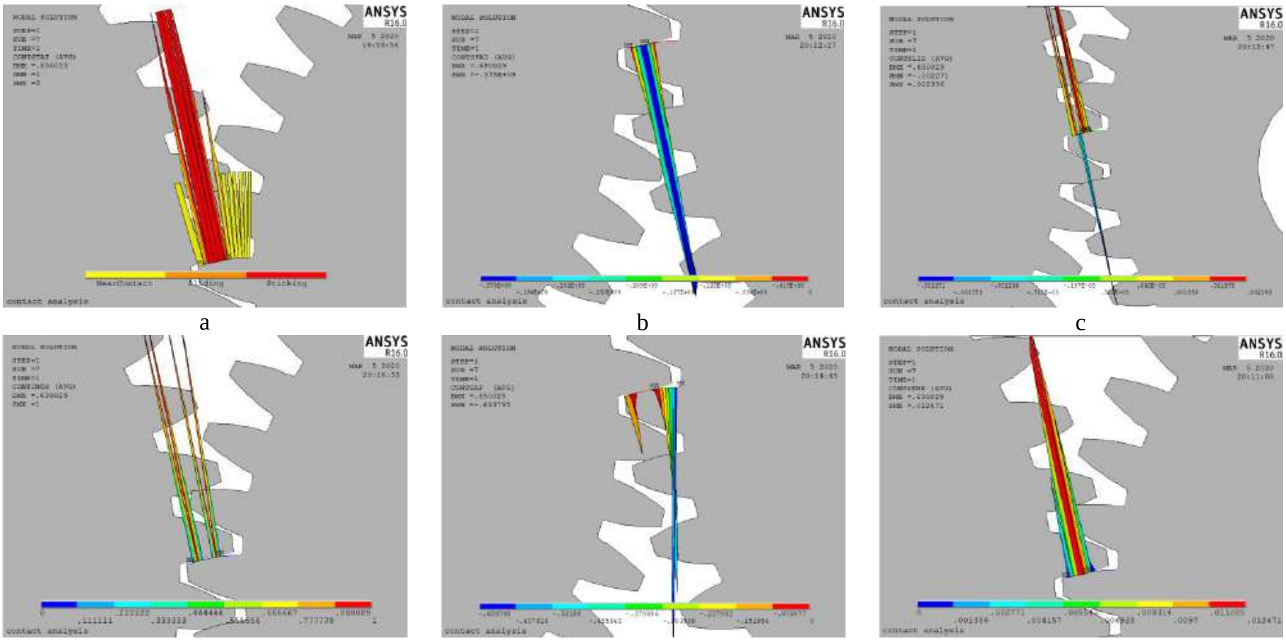
For the elliptic cylinder gears having different tooth profiles, it is difficult to analyze the distribution of stress, strain and pressure during the tooth meshing process. Static analysis should be carried out first to find the key parts of the stress, strain and pressure distribution during the meshing process. The distribution of the contact state, tooth surface friction, slip distance, contact vibration, tooth meshing contact gap and tooth contact penetration of a single tooth under the condition of 2D contact is obtained by ANSYS software, as shown in Figure 5.

In Figure 5a, the pitch curve and the vicinity of the root are in the sticking state of driving wheel (left gear), which means that the amount of wear here is the largest during the tooth engagement. The tooth surface friction distribution state in Figure 5b shows the largest near the pitch curve, and the transition from the pitch curve to the root and the tip of the tooth is reduced. Figure 5c shows the slip distance during the tooth meshing, and the slip between the driving and driven wheels is also the largest near the pitch curve. Due to certain vibration generated during the tooth meshing process, the friction between the two tooth surfaces during the meshing process will transfer the vibration to the non-contact area, the other part except the pitch curve, as shown in Figure 5d. The vibration amplitude of the top and root regions is larger, and the vibration amplitude near the pitch curve is smaller. Figure 5e shows the distribution of the contact gap during the tooth meshing process. In the vicinity of the pitch curve, the contact gap is almost 0, indicating the teeth are in meshing state. During the meshing process, the frictional force in the vicinity of the pitch curve are larger than that in the top and root of the tooth. That is to say, the amount of wear in the vicinity of the pitch curve is also the largest. In order to prove this, the tooth meshing contact penetration as shown in Figure 5f is obtained. Here the meshing penetration in the vicinity of the pitch curve is larger than the top and the root portion of tooth. Therefore, in the process of subsequent analysis of the tooth meshing error, the area near the pitch curve should be mainly analyzed.

The relationship between cumulative iteration number and absolute convergence norm can be used to show the convergence under certain convergence criteria. The convergence criterion of the finite element simulation used in this analysis is 2-norm, as show in Figure 6, the purple line represents the residual, and the blue line represents the convergence criterion. When the residual is dipped below the convergence criterion, it represents convergence.

4 DYNAMIC MESHING CHARACTERISTICS ANALYSIS OF ELLIPTIC CYLINDER GEARS

When simulating the meshing process, the following boundary conditions should be set: the inner ring of the rigid body shaft hole drives the gear body to rotate, the gear material is Solid-164 flexible body, the inner hole of the shaft is Shell-163 rigid body, density is $7.8\times10^3 \text{kg}/m^3$, the modulus of elasticity is 201Gpa and Poisson's ratio is 0.3. The driving and driven wheels are limited to X, Y, Z three-direction moving degree of freedom and X, Y rotation degrees of freedom, and the driving speed of the driving wheel is 600r/min. In the process of solving the tooth meshing model, the time step and the scale factor of the calculation time step are too large to interrupt the simulation, while the generation of negative volume is mostly caused by grid distortion, which is related to mesh quality and material and load conditions. Therefore, the appropriate time



a b c

d e f

Figure 5. Static analysis of elliptic cylinder gear meshing a tooth contact state, b tooth surface friction, c meshing slip distance, d tooth contact vibration, e tooth contact gap, f tooth contact penetration

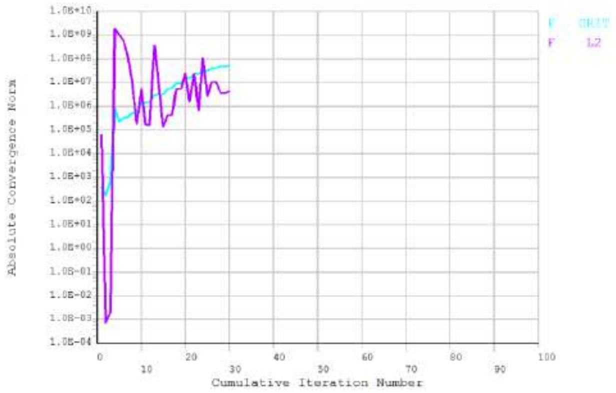


Figure 6. Convergence of finite element analysis

step should be taken to avoid the negative volume. The debug time step scale factor TSSFAC is taken the value of 0.5, the time step DT2MS values -2×10^{-7} can complete the analog tooth engagement. After meshing, the number of driving wheel nodes is 205,716, the number of units is 210,211, the number of driven wheel nodes is 181,740, and the number of units is 186,180. [18]

The tooth surface contact area during the meshing of the elliptic cylinder gears exhibits an elliptical shape, which is similar to the tooth surface contact area of the spur gear, indicating that the stress and strain of the this region is greatest during the tooth meshing process. Therefore, in the subsequent analysis of the meshing error of the elliptic cylinder gear, the meshing elements of the driving and driven wheel pitch curves and the middle section are selected as the research objects. Since the tooth profiles of the elliptic cylinders are different, it is difficult to analyze the data of all the teeth. For this reason, six teeth are selected as the research object in the analysis process, selecting No.1 tooth near the shaft hole and No.5, No.10, No.15, No.20 and No.24 teeth in the clockwise direction, as shown in Figure 7.

4. 1. Comparative Analysis of Effective Plastic Strain of Teeth

Gear wear is very complicated, and it is related to various conditions such as working conditions, lubrication conditions, tooth surface roughness, etc. In the process of simulating, the degree of wear of the gear can be reflected by amount of the effective plastic strain of the tooth surface. By collecting

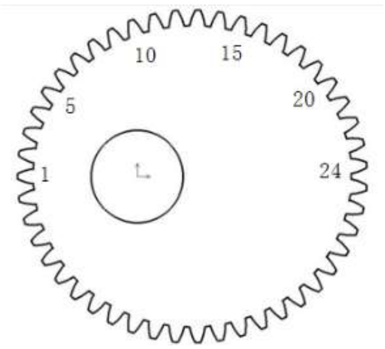


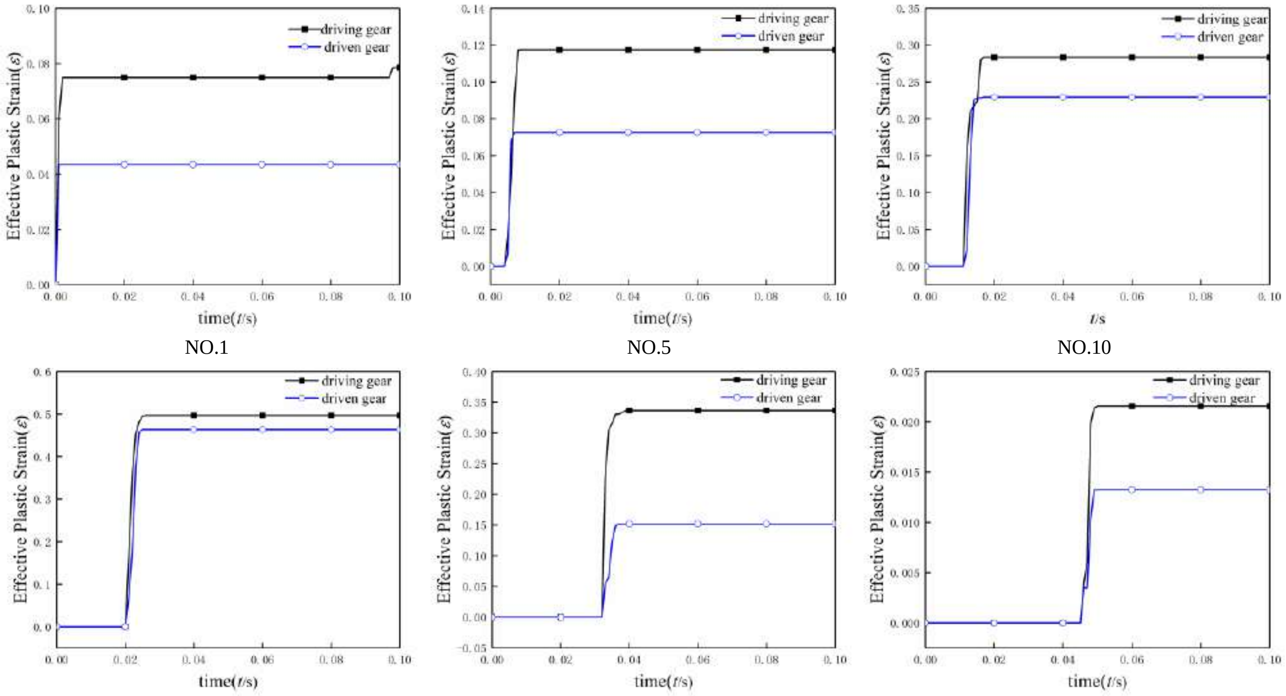
Figure 7. Tooth Number

the meshing data of the same meshing unit of driving and driven gear, the variation law of effective plastic strain during the tooth meshing process can be obtained, as shown in Figure 8. The effective plastic strain of the driving wheel is larger than that of the driven wheel, and will change due to the position of the teeth on the pitch curve. The effective plastic strain of the teeth at the ends of the short axis and its vicinity is larger than that at both ends of the long axis and its vicinity, which indicates that the amount of wear of the driving wheel is greater than that of the driven wheel during the tooth meshing process.

The presence of lubricant between the teeth during meshing reduces the wear between the teeth and accelerates the transmission of the heat generated by the meshing between the teeth, but does not eliminate the wear between the primary and driven wheels. The presence of the impact causes the driving wheel to wear more than the driven wheel. Therefore, it is conceivable to distribute the gear with higher manufacturing precision at the driving wheel during the mounting process.

4. 2. Comparative Analysis of Tooth Effective Stress

Through the solution of the finite element model of the elliptic cylinder gear, the distribution area of the effective stress of the tooth surface is obtained, as shown in Figure 9, which shows that the contact area of the tooth surface in the process of gear engagement is also elliptical and consistent with the analysis in the literature [18-22], which further verifies the accuracy of the analysis method proposed in article.



NO.1 NO.5 NO.10

NO.15 NO.20 NO.24

Figure 8. Effective plastic strain of the tooth data collection point

Figure 10 shows the distribution of the effective stress of the six teeth. The effective stress curve of the driving wheel is smoother, and there is a certain impact in that of driven wheel. The occurrence of the impact is synchronized with the driving wheel, but the amplitude is larger than that of the driving wheel, which indicates that there is a certain meshing impact in the tooth meshing process. In the initial meshing phase, the effective stress curves of the driving and driven wheels have a large impact (biting impact), and the meshing

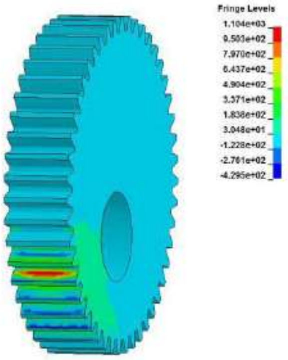
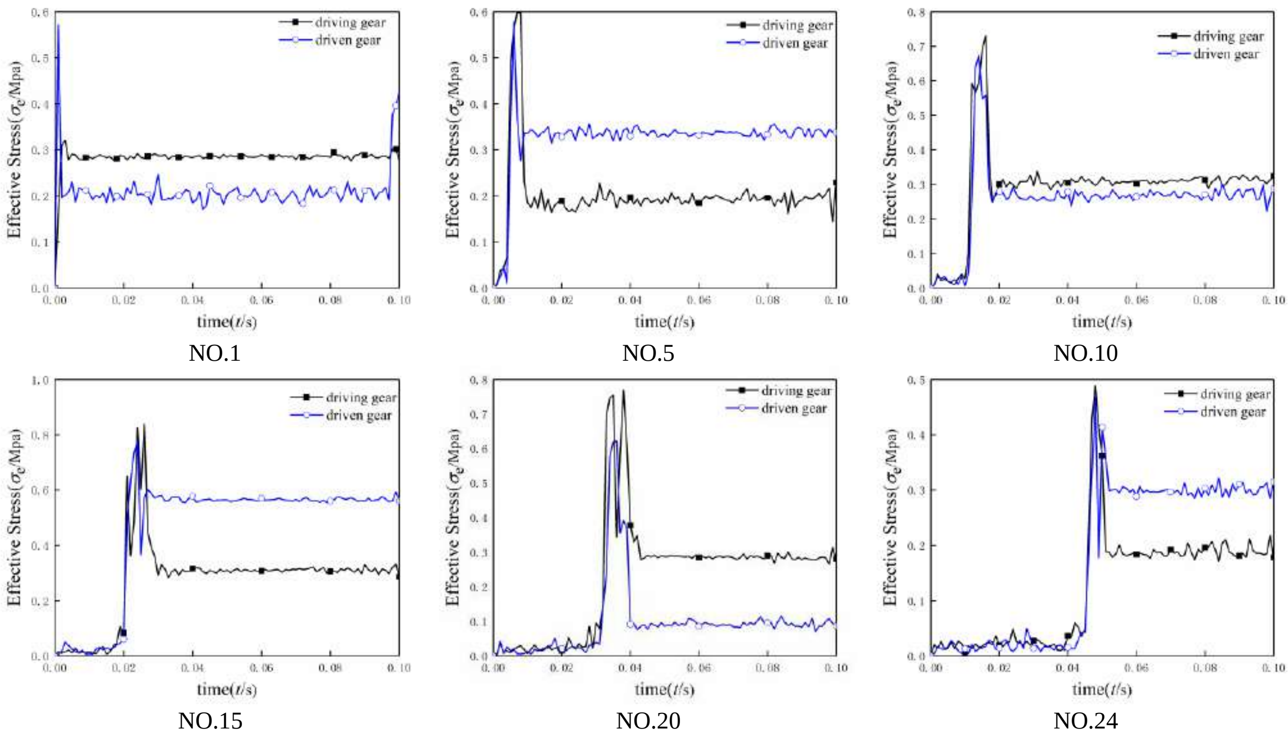


Figure 9. Effective stress distribution of tooth surface

time also changes with the position of the teeth on the pitch curve. The impact will have a certain backward movement with the position of the teeth. After the withdrawal of the mesh, the stress on the teeth does not decrease to 0 due to the residual stress, but fluctuates within a certain range of values. The residual stress experienced by the driving and driven wheels will vary with the position of the teeth on the pitch curve. In the theoretical state, the interaction of forces and the interaction of deformation indicate that the effective stress between the teeth of the same pair should be the same. However, the presence of the flank clearance and power loss will cause the stresses and strains of the driving and driven teeth to be different during the meshing process, and the difference can be approximated to the calculated value of the meshing error of the teeth [11]. The residual stresses, produced in the contact region, is a key step in determination of the continuing integrity of engineering and structural components, and the literature [23-24] have been described in detail, and will not be described here.

Table 2 shows the fluctuation values of tooth effective stress. The fluctuation values of No. 1 tooth and No. 5 tooth are several times larger than that of the other teeth. The reason is that the center of the gear rotation is located at the left focus of the elliptical curve, and the driving gear mesh with the teeth near the focus are located near the right focus of the elliptical curve. The tooth at the position and the left focus has the smallest force arm and the largest meshing force, which causes the material deformation to increase and the effective stress to increase. The increase in the amount of tooth wear during the meshing process causes the meshing error to become large.



NO.1

NO.5

NO.10

NO.15

NO.20

NO.24

Figure 10. Effective stress of tooth data collection point

TABLE 2. Fluctuation value of tooth equivalent stress

Number	Δ Driving	Δ Driven	Fluctuation value
1	19.174	74.185	55.011
5	66.621	40.712	25.909
10	53.787	63.102	9.315
15	45.988	37.352	80636
20	36.704	44.126	7.422
24	45.84	36.684	9.156

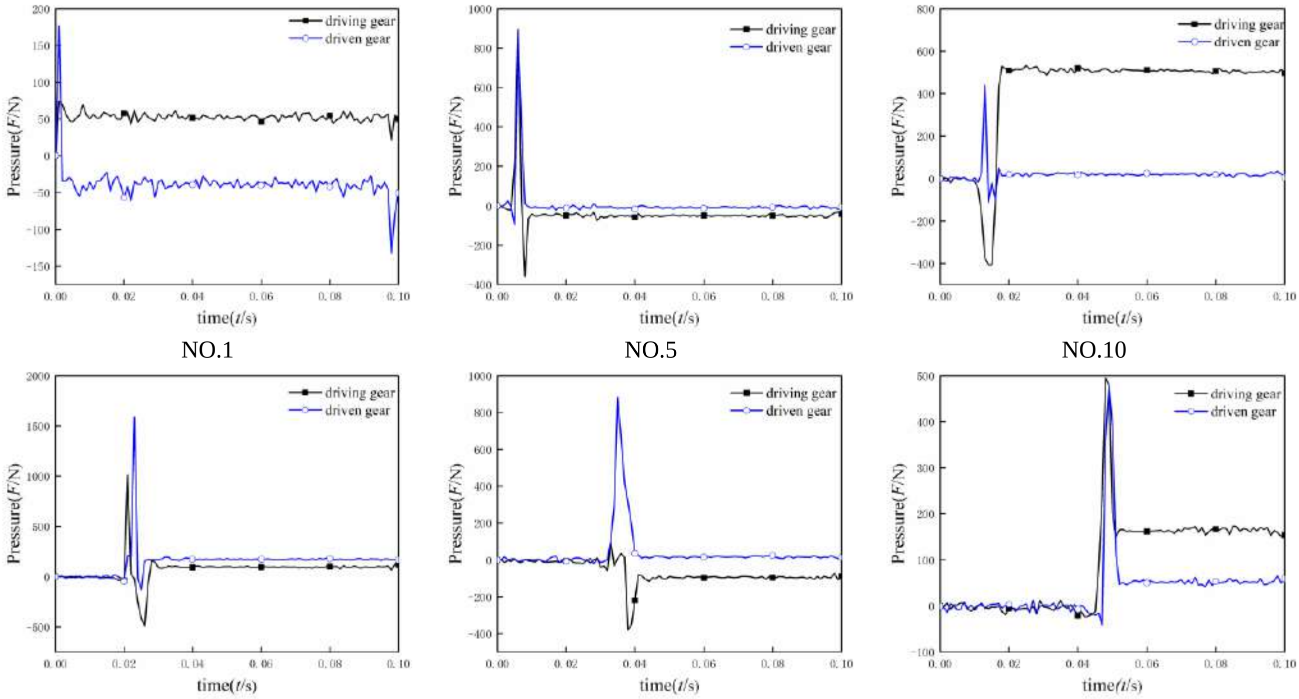
4. 3. Comparative Analysis of Tooth Surface Contact Pressure

Figure 11 shows the trend of the pressure on the same engagement unit of the driving and driven wheels. The pressure curve is always continuous, which indicates that the teeth have good contact characteristics and no tooth separation occurs. In the initial meshing stage, due to the existence of the meshing impact, there will be a certain collision between the teeth, which is specifically shown as the impact phenomenon on the left side of the pressure curve. Since the teeth are different in position on the elliptical pitch curve, the meshing time will also be slightly different. Therefore, the phenomenon of impact on the left side of Figure 11 will have a certain right shift.

In Figure 11, only No.1 and No.24 teeth have a slight impact on the pressure curve, and the other tooth pressure curves have only a small range of fluctuations. The reason for this phenomenon is that No. 1 tooth and No. 24 tooth are respectively located at both ends of the long axis in the elliptical pitch curve. Due to the change of the radius of the pitch curve during the tooth meshing process, the teeth will generate a certain meshing vibration and impact, and the pressure curve will be reflected in the form of small fluctuations. The fluctuation of the pressure between the driving and driven wheels as shown in Table 3, was obtained by collecting the gear meshing data. In Table 3, except for No. 1 tooth and No. 20 tooth near the end points of the long axis, the fluctuation values of the other four teeth are relatively stable. In contrast, the fluctuation value of the driving wheel is larger. The fluctuation of the surface pressure is reflected in the form of vibration and impact in the specific working conditions, and the inconsistency of the pressure fluctuation between the driving and driven wheels indicates that the meshing error between the teeth also changes at the moment momentarily, and is decreased with the transition of the long axis of the elliptic pitch curves to the short axis.

4. 4. The Influence of Load on Tooth Meshing Error

In order to study the influence of the load on the tooth meshing error, the change in effective plastic strain, effective stress and surface pressure of No. 1 tooth under normal load and alternating load were obtained respectively, as shown in Figure 12. Under the condition of alternating load, the effective plastic strain of the driving and driven wheels is larger than that of the normal load, which indicates that the meshing conditions of the teeth are more complicated under the condition of alternating load. In terms of effective stress, the alternating load, due to its time-varying, causes the material strains to cancel each other during the tooth meshing process, which makes the effective stress under the alternating load condition lower than the normal load condition. During the tooth meshing process, the effective stress and surface pressure change under



NO.1 NO.5 NO.10

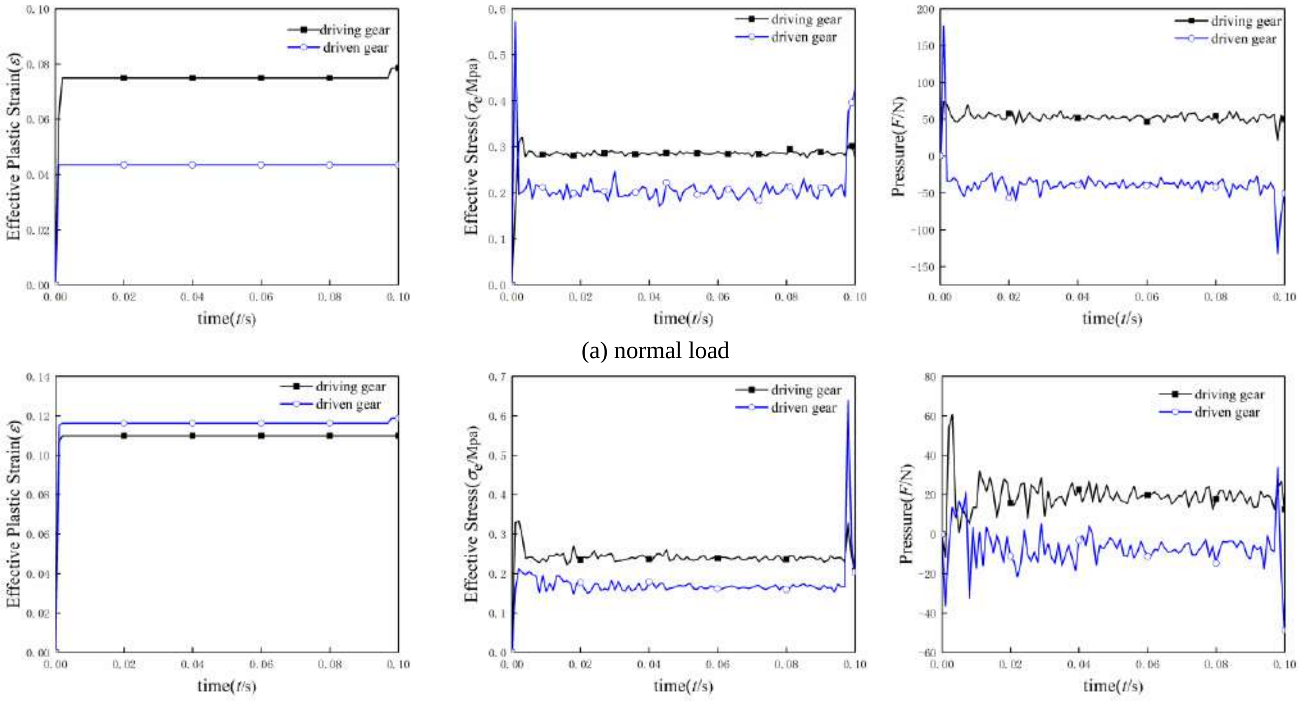
NO.15 NO.20 NO.24

Figure 11. Tooth surface pressure of the data collection point

TABLE 3. Fluctuation value of tooth surface pressure

Number	Δ Driving	Δ Driven	Fluctuation value
1	26.187	38.774	12.557
5	39.472	31.26	8.212
10	45.767	42.882	2.885
15	43.858	37.72	6.138
20	42.365	16.422	25.943
24	19.345	24.768	5.428

normal load conditions are relatively stable. In contrast, under the alternating load condition, there is a slight impact on the effective stress curve and the pressure impact on the tooth surface increases, which indicates that the meshing error between the teeth under the alternating load condition will increase. To verify the above conclusion, the tooth surface pressure fluctuation values under the two load conditions were obtained, as shown in Table 4. Through the comparative analysis of the value of pressure fluctuation of the tooth surface, it can be known that the meshing error of the teeth is increased due to the time variation of the load under the alternating load condition.



(a) normal load

(b) alternating load

Figure 12. Distribution of strain, stress and pressure of teeth under two load conditions

TABLE 4. Fluctuation values of tooth surface pressure under different load conditions

	Common load	Alternating load
Δ Driving	23.576	26.187
Δ Driven	37.647	38.744
Fluctuation values	14.071	12.557

5 CONCLUSIONS

For the elliptic cylindrical gear pair, the static analysis of the tooth surfaces that mesh with each other was first performed to determine the key parts of the gear meshing process. Based on this, the difference between the stress, strain, and pressure of the tooth surfaces of the driving and driven gears during the gear meshing process is proposed as a reference for describing the gear meshing error, and a dynamic meshing model of the elliptic cylinder gear is established. The distribution laws of effective plastic strain, effective stress, surface pressure and meshing error during gear meshing are obtained. The conclusions are as follows:

a) The static contact analysis of the elliptic cylinder gear pair shows that the region near the pitch curve is the key part in the process of gear meshing.

b) The effective plastic strain, effective stress, pressure on the tooth surface and the meshing error will vary with the position of the tooth on the pitch curve. Therefore, in the process of gear installation, the tooth modification can be considered or the gear with higher manufacturing precision can be distributed at the driving wheel to avoid this phenomenon.

c) Load has a certain influence on meshing error. In contrast, the alternating load causes meshing condition to be more complicated due to its time-varying characteristics, and will increase the meshing error of the gear.

d) The dynamic meshing model and meshing error analysis method proposed in this paper can be applied to other types of noncircular gears

6. ACKNOWLEDGEMENTS

This research received specific grant from the National Natural Science Foundation of China (No. 51765032)

7. REFERENCES

1. Litvin, F. L., "Noncircular gears design and generation", ***Cambridge University Press***, 2009.
2. Li R. F., Wang J. J., "Dynamics of gear system", ***Beijing: Science Press***, 1997.
3. Sainte M. N., Velex P., Roulois G., Caillet J., "A study on the correlation between dynamic transmission error and dynamic tooth loads in spur and helical gears", ***Journal of Vibration and Acoustic***, Vol. 139, (2017) ,1-10. http://dx.doi.org/10.1115/1.4034631
4. Sungho P., Seokgoo K., Joo-Ho C., "Gear fault diagnosis using transmission error and ensemble empirical mode decomposition", ***Mechanical Systems and Signal Processing***, No. 108, (2018), 262–275. http://dx.doi.org/10.1016/j.ymssp.2018.02.028
5. Hotait M. A., Kahraman A., "Experiments on the relationship between the dynamic transmission error and the dynamic stress factor of spur gear pairs", ***Mechanism and Machine Theory***, Vol. 70, No. 6 (2013), 116-128. http://dx.doi.org/10.1016/j.mechmachtheory.2013.07.006
6. Wang G. J., Chen L., Li Y., Zou S. D., "Research on the dynamic transmission error of a spur gear pair with eccentricities by finite element method", ***Mechanism and Machine Theory***, Vol. 109, (2017), 1-13. http://dx.doi.org/10.1016/j.mechmachtheory.2016.11.006
7. Yu L., Wang G. J., Zou. S. D., "The experimental research on gear eccentricity error of backlash-compensation gear device based on transmission error", ***International Journal of Precision Engineering and Manufacturing***, Vol. 19, No. 1 (2018), 5-12. http://dx.doi.org/10.1007/s12541-018-0001-7
8. Deng X. Z., Xu A. J., Zhang J., Li H. B., Xu K,. "Analysis and experimental research on the gear transmission error based on pulse time spectrum", ***Chinese Journal of Mechanical Engineering***, Vol. 50, No. 1 (2014), 85-90. http://dx.doi.org/10.3901/JME.2014.01.085
9. Zou S. D., Wang G. J., "Research on transmission error of dual-eccentric gears", ***Journal of University of Electronic Science and Technology of China***, No. 6, (2017), 157-162. http://dx.doi.org/10.3969/j.issn.1001-0548. 2017.06.027
10. Wu S. J., Peng Z. M., Wang X. S.,Zhu W. L., Li H. W., Qian B., Cheng Y., "Impact of mesh errors on dynamic load sharing characteristics of compound planetary gear sets," ***Chinese Journal of Mechanical Engineering***, Vol. 3, No. 40, (2016), 1-6. http://dx.doi.org/10.3901/JME.2015.03.029
11. Chang L. H., Liu G., Wu L. Y., "Determination of composite meshing errors and its influence on the vibration of gear system", ***Chinese Journal of Mechanical Engineering***, Vol.51, No.1 (2015), 123-130. http://dx.doi.org/10.3901/JME.2015.01.123
12. Sumar H. S., Athanasius P. B, Jamari., Jamari G. R., "Simulation of excavator bucket pressuring through finite element method", ***Civil Engineering Journal***, Vol. 4, No. 3, (2018), 478-487. http://dx.doi.org/10.28991/cej-0309189
13. Massoud H. E., Mahyar N., Farzad T., "Position control of a flexible joint via explicit model predictive control: an experimental implementation", ***Emerging Science Journal***, Vol. 3, No.3 (2019) 146-156. http://dx.doi.org/10.28991/esj-2019-01177
14. Abbas A. L., Mohammed A. H., Abdul-Razzaq K. S., "Finite Element Analysis and Optimization of Steel Girders with External Prestressing", ***Civil Engineering Journal***, Vol. 4, No. 3, (2018) 1490-1500. http://dx.doi.org/10.28991/cej-0309189
15. Li Q., Wu S., Q., Yan J. B, "Dynamic contact emulate analysis of logarithmic spiral bevel gear with ANSYS/LS-DYNA", ***Applied Mechanics and Materials***, Vol. 86, (2011), 531-534. http://dx.doi.org/10.4028/www.scientific.net/AMM.86.531
16. Tong H., Liu Z. H., Yin L., Jin Q., "The dynamic finite element analysis of shearer's running gear based on LS-DYNA", ***Advanced Materials Research***, Vol.402 (2012), 753-757. http://dx.doi.org/10.4028/www.scientific.net/AMR.402.753
17. Ma Y., Dong X. H., Huang Q., "Tooth root stress analysis of gear processed by full radius hob based on ANSYS/LSDYNA", ***Advanced Materials Research***, Vol.479-491 (2012), 1409-1413. http://dx.doi.org/10.4028/www.scientific.net/AMR.479-481.1409
18. Dong C. B., Liu Y. P., Wei Y. Q., "Dynamic Meshing Characteristics of Elliptic Cylinder Gear Based on Tooth Contact Analysis", International Journal of Engineering, Transactions A: Basics, Vol. 33, No. 4 (2020) 676-685. http://dx.doi.org/10.5829/IJE.2020.33.04A.19

Reze Kashyzadeh K., Farriahi G. K., Shariyat M., Ahmadian M. T., "Experimental and Finite Element Studies on Free Vibration of Automotive Steering Knuckle". International Journal of Engineering, Transactions B: Applications, Vol. 30, No. 11 (2017) 1776-1783. http://dx.doi.org/10.5829/ije.2017.30.11b.20

Francisco S. M., Jose L. I., Victor R. C., "Numerical tooth contact analysis of gear transmissions through the discretization and adaptive refinement of the contact surfaces", ***Mechanism and Machine Theory***, Vol. 101 (2016), 75-94. http://dx.doi.org/10.1016/j.mechmachtheory.2016.03.009

19. Argyris J., Fuentes A., Litvin F. L., "Computerized integrated approach for design and stress analysis of spiral bevel gears", ***Computer Methods in Applied Mechanics and Engineering***, Vol. 191, No. 11-12 (2002), 1057-1095. http://dx.doi.org/10.1016/S0045-7825(01)00316-4
20. Wang Y. M., Shao J. P., Wang X. G., Zhao X. Z., "Thermomechanical coupled contact analysis of alternating meshing gear teeth surfaces for marine power rear transmission system considering thermal expansion deformation", ***Advances in Mechanical Engineering***, Vol. 10, No. 1 (2018), 1-12. http://dx.doi.org/10.1177/1687814017753910
21. Fawwahi G. H., Faghidian S. A., Smith D. J., "An inverse approach to determination of residual stresses induced by shot peening in round bars", ***International Journal of Mechanical Sciences***, Vol. 51, No. 9-10 (2009), 726-731. http://dx.doi.org/10.1016/j.ijmecsci.2009.08.004
22. Faghidian S. A., Goudar D., Farrahi G. H., David J. S., "Measurement, analysis and reconstruction of residual stresses", ***Journal of Strain Analysis for Engineering Design***, Vol. 47, No. 4 (2012), 254-264. http://dx.doi.org/10.1177/0309324712441146

Persian Abstract

چکیده

به منظور بررسی ویژگی‌های جفت‌شوندگی پویای چرخ‌دنده‌های استوانه‌ای با مقطع بیضوی و به دست آوردن خطای انطباق در سیستم انتقال دنده، تحلیل تماس ایستای دو بعدی سطح دندانه‌ی دنده و از نرم افزار ANSYS استفاده شد و قسمت‌های کلیدی سطح تماس دندده‌ها تعیین شدند. سپس، مدل جفت‌شوندگی پویای چرخ دنده‌های استوانه‌ای بیضوی تولید شده و با استفاده از نرم افزار ANSYS LS-DYNA، فرآیند تماس پویا زیر بار شبیه‌سازی شد. قانون توزیع فشار موم‌سان مؤثر، تنش و فشار مؤثر چرخ‌های راننده و محرک به دست آمد. بر این اساس، قانون توزیع خطای جفت‌شوندگی از طریق محاسبه به دست می‌آید. نتایج نشان می‌دهد که توزیع تنش، کرنش و فشار سطح دندانه در حین درگیری دندانه با موقعیت آن در منحنی بیضوی گام مرتبط است. موقعیت دندانه روی منحنی گام و باری که تحمل می‌کند، تأثیر خاصی بر خطای جفت‌شوندگی دارد. نتایج این تحقیق می‌تواند راهنمایی‌هایی برای مطالعه‌ی بعدی در مورد خطای انتقال دنده‌های غیردایره‌ای، اصلاح دنده‌ها و کاربردهای مهندسی آنها ارائه دهد.

International Journal of Engineering

Journal Homepage: www.ije.ir

Single-vehicle Run-off-road Crash Prediction Model Associated with Pavement Characteristics

M. Akbari[a], G. Shafabakhsh*[a], M. R. Ahadi[b]

[a] *Faculty of Civil Engineering, Semnan University, Semnan, Iran*
[b] *Transportation Research Institute, Ministry of Roads and Urban Development, Tehran, Iran*

PAPER INFO

Paper history:
Received 12 March 2020
Received in revised form 30 April 2020
Accepted 12 June 2020

Keywords:
Highway Segmentation
Pavement Physical Characteristics
Crash Prediction Models
Run-off-road Accidents
Functional Forms
Nonlinear Negative Binomial Regression

ABSTRACT

This study aims to evaluate the impact of pavement physical characteristics on the frequency of single-vehicle run-off-road (ROR) crashes in two-lane separated rural highways. In order to achieve this goal and to introduce the most accurate crash prediction model (CPM), authors have tried to develop generalized linear models, including the Poisson regression (PR), negative binomial regression (NBR)), and non-linear negative binomial regression models. Besides exposure parameters, the examined pavement physical characteristics explanatory variables contain pavement condition index (PCI), international roughness index (IRI) and ride number (RN). The forward procedure was conducted by which the variables were added to the core model one by one. In the non-linear procedure and at each step, 39 functional forms were checked to see whether the new model gives better fitness than the core/previous model. Several measurements were taken to assess the fitness of the model. In addition, other measurements were employed to estimate an external model validation and an error structure. Results showed that in PR and NBR models, variables coefficients were not significant. Findings of the suggested nonlinear model confirmed that PCI, as an objective variable, follows the experts anticipation (i.e., better pavement manner associates with less ROR crashes). Finally, it should be noted that the roughness variable was insignificant at the assumed significance level, so it had no contribution to ROR crashes. The results imply that improving the pavement condition leads to a more probable decrease in the ROR crashes frequency.

doi*: 10.5829/ije.2020.33.07a.25*

1. INTRODUCTION

Crash prediction models (CPMs) describe a mathematical relation between accident frequencies and defined explanatory variables. These models have been used to give an estimate of the expected crash frequency based on the characteristics of accident factors.

Among all types of accidents, the run-off-road (ROR) crash is one of the most severe crash occurrences. Many types of research have been done to increase the effectiveness and reliability of the existing safety plans related to ROR crashes. Some of which are related to roadside features [1-3].

However, there are several studies that include road infrastructure and geometry features [4-6].

Among all these studies, the influence of pavement condition and the riding quality aspects related to pavement physical characteristics have been neglected.

The main target of the recent study is to fill these research gaps by developing a CPM that incorporates the pavement condition and related riding quality measures to the frequency of ROR crash occurrence in two-lane separated rural highways. We will further aim to develop a way to formulate the relationship between ROR crash frequency and the following explanatory variables:

- Exposure variables (annual average daily traffic (AADT) and segment length),
- Pavement condition variable (pavement condition index (PCI)),
- Riding quality variables (international roughness index (IRI) and ride number (RN)).

*Corresponding Author Institutional Email: *ghshafabakhsh@semnan.ac.ir* (G. Shafabakhsh)

Please cite this article as: M. Akbari, G. Shafabakhsh, M. R. Ahadi, Single-vehicle Run-off-road Crash Prediction Model Associated with Pavement Characteristics, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1375-1386

Two categories of models were used to achieve this goal, the generalized linear models (GLMs) and the nonlinear stochastic models. So, it will be recognized if the nonlinear model can give better results or if the GLM ones are significant.

1. 1. Literature Review

The Primitive researches about pavement influences on accident rate go back to the 1970s. The first Pavement-Accident models were developed based on standard multiple linear-regression equations.

Later, Chan et al. [7] studied the effects of asphalt pavement conditions on traffic accidents. They developed twenty-one negative binomial (NB) models for different crash types. Those models examined the impact of the pavement characteristic variables, including rut depth (RD), international roughness index (IRI), and present serviceability index (PSI) on crash frequency rates. The primary outcome is that the RD models are insignificant for all types of accidents while the IRI and PSI parameters give a suitable fitness and execute well on the total crash data. The deficiency of their study might be related to conjunctions between single and multi-vehicle crash types in modeling.

Jiang et al. [8] correlated pavement management and traffic parameters to accident frequency occurrence. Results approved the significant impact of PSI and PDI on accident frequency. The frequency models have a direct relationship with the road roughness, i.e., less pavement roughness can be associated with less accident frequency.

Akbari et al. [6] studied the impact of pavement condition index (PCI) variable on the ROR crashes. They concluded that a unit increase in PCI variable reduces about 1.93% in frequency of ROR crashes.

It should be noted that the study on the impact of pavement characteristics on accident frequency (specially ROR crashes) are rare and this study carried out to fulfill this research gap.

2. DATA DESCRIPTION

The Semnan province has a critical situation not only in Iran (because of locating along with the internal East-West corridor), but also in Asian highways networks (as located in Asian Highway Route No. 1 (AH1) also known as the Silk Road).

The Semnan to Tehran link had some segments with the critical deteriorated pavement. High travel demand caused by heavy trucks are the reasons for such conditions that affected the pavement quality.

This Province does not have a good position in the ROR crash occurrence ranking. The presence of severe pavement distresses, and high ROR crashes are the main reasons for selecting this roadway. An executive report revealed that 42 percent of fatal crashes due to ROR crashes are related to Isfahan, Fars, Semnan and Kerman provinces and another document highlighted that 64.5 percent of all accidents which occurred in Semnan province were related to ROR crashes.

There are three types of data that are generally available for empirical analysis: time series, cross-section and pooled. Cross-sectional data, in statistics is a type of data collected by observing many subjects (such as road segments) at one point or period of time. The analysis might also have no regard to differences in time. Analysis of cross-sectional data usually consists of comparing the differences between selected subjects.

The cross-sectional sample provides a snapshot of those observed subjects, at that one point in time. Note that no one knows based on one cross-sectional sample if the dependent variable is increasing or decreasing; the model can only describe the current situation. Cross-section differs from time series, in which the same small-scale or aggregate entity is observed at various points in time. Another type of data, pooled data, combines both cross-sectional and time series data ideas and looks at how the subjects change over a time series.

Some variables like geometric design and roadside features usually keep their characteristics for years, but some variables such as pavement characteristics, gradually deteriorate every year. The main restriction of data gathering for this study is to survey pavement condition characteristics, especially for a continuous bases in several years. Analyzing the pavement condition index (PCI) for a certain road segments during a continuous years is not only too expensive but also time consuming. So in this study, the cross-sectional data of one calendar year were considered to describe the current situation.

Although a high correlation between some of the variables exists, none of them were eliminated. Significance and goodness-of-fit measurements were used to consider the model stability instead of considering their correlations [9].

In multiple regression analysis, the nature and significance of the relations between the explanatory and independent variables are often of particular interest. Multicollinearity is a common problem when estimating linear or generalized linear models, including logistic regression and Cox regression. When the predictor variables are correlated among themselves, multicollinearity among them is said to exist which leading to unreliable and unstable estimates of regression coefficients. Most data analysts know that multicollinearity is not a good thing. But many do not realize that there are several situations in which multicollinearity can be safely ignored. In this situation, the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data [10]. Multicollinearity does not reduce the predictive power or reliability of the model as a whole, at least within the sample data set; it only affects calculations regarding individual predictors.

That is, a multivariate regression model with collinear predictors can indicate how well the entire bundle of predictors predicts the outcome variable, but it may not give valid results about any individual predictor, or about which predictors are redundant with respect to others.

As it discussed in previous paragraphs, this issue appears in linear or generalized linear models. These types of models are a part of our study but the main part of our modelling is refer to nonlinear negative binomial regression. So this issue is not a matter to our models comparison study.

Table 1 shows the statistical details of contributed variables. As discussed previously, the cross-sectional data of one Iranian calendar year were considered to describe the current situation of examined roadway.

2. 1. Accident Data

The single-vehicle ROR crash is applied to those events in which an errant vehicle is encroaching the roadway. This encroachment may lead to three types of outcomes: non-strike incidence, crash with non-fixed objects at roadside, and collisions involving roadside features and fixed-objects. All of these accidents are considered in this study.

The accident data were collected from FAVA department of Rahvar Police Office. During the study period (i.e., 20 March 2011 to 20 March 2012) among all 373 single- and multi-vehicle crashes, there were 166 single-vehicle ROR crash occurrences in the Semnan to Tehran roadway.

Shafabakhsh et al. [11] studied the spatial analysis of frequency of rural accidents and concluded that the accidents are considerable within 30 km from Garmsar.

2. 2. Exposure

The lack of exposure parameters leads definitely to uncertainty in any CPM outputs. Researchers have emphasized that AADT is one of the most critical exposure variables for any crash prediction model one of the many exposure units is a million vehicle-kilometer of travel per year. The exposure function for the i^{th} segment is given by Eq. (1). The scale of traffic data variable can be estimated as AADT/1000 [12, 13].

$$Exposure_i = \frac{AADT_i \times 365 \times Length_i}{10^6} \tag{1}$$

Homogenous segments give better results and have more significant variables. Furthermore, short road sections have undesirable impacts on the linear regression models outputs. Therefore, long segments are more suitable and recommendable.

The distance between Semnan and Tehran is 150 kilometer, which is a two-lane separated rural highway. So, the link is separated into 55 homogeneous segments based on its pavement condition.

2. 3. Pavement Surface Characteristics Data

2. 3. 1. Pavement Condition Index (PCI)

The PCI is a numerical indicator that exemplifies the present condition of a pavement by considering the observed distresses. The PCI for the roadway is estimated from the collected visual survey distress data and is based on the standard practice of roads PCI surveys, ASTM D6433-11 [14].

In this study, the pavement data were collected by a car equiped to a spcial instrument that continuously takes photographs while passing the lanes. Captured photos cover a 2*3.8 square meters area with a 0.5-meter longitudinal coverage to fit consecutive margins. The 100-meter-longitudinal merged photos are processed using Road Mapper® Software, and then the procedure of PCI calculating will be performed (see Figure 1). After assessing the PCI of each 100-meter units, the PCI of defined homogenous segments is calculated by taking the average of PCIs from involved 100-meter units. The summary of defined segments PCI is given in Table 1.

2. 3. 2. International Roughness Index (IRI)

The international roughness index (IRI) is a mathematical transformation of a longitudinal profile that represents pavement roughness results in vehicle vibration. This model relates the movements of a simulated quarter-car to a single longitudinal wheel path profile at a constant speed of 80 km/h. This method is certificated in ASTM E1926-08 [15]. The IRI value is zero for newly constructed roadway, and it grows up to 12 for deteriorated ones.

TABLE 1. Details of selected variables for each homogenous segment in the study

Variables and Abbreviations		Units	Min.	Max.	Mean	Std. Dev.
Exposure	Annual Average Daily Traffic (AADT)	veh/day	9570	10380	9830	380
	Segment Length	km	0.7	6.7	2.74	1.53
Pavement Condition Index (PCI)		%	12.5	99.38	69.35	30.4
International Roughness Index (IRI)		m/km	1.39	9.47	3.67	2.29
Riding Number (RN)		1 to 5	0.72	3.79	2.56	1.05
Run-Off-Road (ROR) Crashes Frequency		No.	0.0	6.0	1.16	1.26

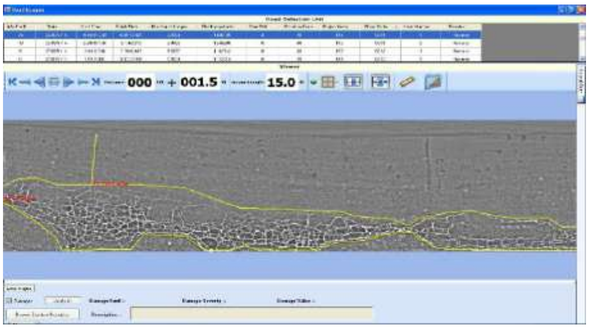


Figure 1. A view of Road Mapper Software interface and typical measurement of pavement distresses

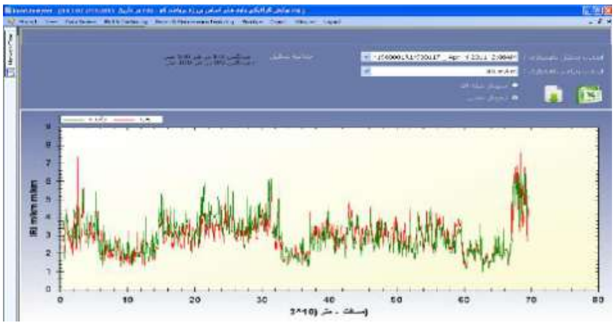


Figure 2. A The view of Road Analyzer Software interface and typical illustration of IRI oscillation

2. 3. 3. Ride Number (RN) The ride number (RN) is defined as the ride-ability index of a pavement that ranges from 0 to 5 for roadways indicator of impassable to perfect conditions, respectively. The RN depends on the physical properties of specific kinds of measurement tools and is based on the nature of the longitudinal profile. Therefore, the RN is a time-independent parameter.

In this study, the longitudinal profile of the roadway is picked up by an instrumented vehicle. After that, the Road Analyzer® software is used to analyze the collected raw data. The variables of IRI and RN will then be calculated based on the ASTM E1926-08 (see Figure 2) [15], together with ASTM E1489-08 [16]. The summary of calculated IRI and RN values are given in Table 1.

3. METHODOLOGY AND MODELS FRAMEWORK

The methodological issues and processes of crash prediction models (CPMs) should be discussed as the following procedures:

- Collecting of explanatory variables' data,
- Offering appropriate functional/model forms,
- Running suggested models,
- Evaluating Goodness-Of-Fit,
- Estimating models' error.

3. 1. Generalized Linear Regression Models Framework (GLMs) GLMs supply a class of fixed-effect regression models for dependent variables such as crash counts [17].

3. 1. 1. Poisson Regression Model Form This model develops a model of accident frequency variable Y, which follows a Poisson distribution with a mean value λ. Equation (2) describes the probability function of such models.

$$P(Y_i = y_i) = f_{Y_i}(y_i;\lambda_i) = \frac{e^{-\lambda_i}.\lambda_i^{y_i}}{y_i!} \quad (2)$$

where y_i is the number of ROR crashes on the i^{th} road segment, $P(y_i)$ is the probability of y_i count of ROR occurrences on the i^{th} road segment and λ_i is the expected value of y_i.

The model is developed through a linear predictor η_i to convert the nonlinear function $E(y_i)$ into a generalized linear one that is given in Equation (3).

$$\eta_i = g(\lambda_i) = \log(\lambda_i) = \beta_0 + \sum_{i=1}^{k} \beta_i.x_i \quad (3)$$

3. 1. 2. Negative Binomial Model Form Since accident data usually have different variances and mean values, the negative binomial (NB) regression models are recommended. Although the forms of generalized linear predictor and logarithm link function for NB and Poisson regression models are similar, they have differences shown below [18]:

- The expected value (y_i) conforms to the NB distribution.
- The error term is added to the NB model.

The expected value ($E(y_i)$) for the NB regression model, is given in Equation (4). The variance function of the NB model reassigns as Equation (5) [19].

$$E(y_i) = \lambda_i = \exp(\beta_i.X_i + \varepsilon_i) \quad (4)$$

$$Var(y_i) = \lambda_i + \lambda_i^2/\phi \quad (5)$$

The variables coefficients, β_i, and dispersion parameters, κ, are obtained by using the maximum likelihood estimation (MLE) method. The GENMOD procedure in SAS represents one of the suitable tools for this purpose. The λ_i function for both the Poisson and NB regression models are redefined by adding the exposure parameter, and therefore, the η_i function would be rewritten as Equation (6).

$$\eta_i = \log(Exposure) + \beta_0 + \sum_{i=1}^{k} \beta_i.x_i \quad (6)$$

3. 1. 3. Decision Criteria to Select GLM Type

The dispersion parameter, σ_d, is estimated based on the Poisson error structure (Equation (7)). The DOF is described as the subtraction of the number of observations from that of the model variables. The Pearson χ^2 is defined in Equation (8).

$$\sigma_d = \frac{Pearson\chi^2}{\text{degree of freedom (DOF)}} \tag{7}$$

$$Pearson\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{E}(Y_i))^2}{Var(Y_i)} \tag{8}$$

where y_i is the number of occurred accidents at each segment, $\hat{E}(y_i)$ is the number of expected accidents at the segment *i*, and $Var(Y_i)$ is the variance of the dependent variable.

3. 2. Nonlinear Negative Binomial Regression Model Form

The employed forward procedure to develop nonlinear CPM is shown in Figure 3. The structure of nonlinear CPM was given in Equation (9).

$$E(y) = f(Exposure) \times g(PCI) \times h(IRI) \times l(RN) \tag{9}$$

where $E(y)$ is the expected ROR crash frequency, and *f*, *g*, *h*, and *I* are the proposed functional forms for each independent variable.

The NLMIXED procedure in SAS maximizes an approximation of the log-likelihood integrated over the random effects. The log-likelihood functions for the negative binomial distribution, which must be maximized, are defined in Equations (10) and (11) [9]:

$$L(y,\mu,\kappa) = \sum_i l_i \tag{10}$$

$$\begin{aligned} l_i &= y_i.\log(\kappa.\mu_i) - (y_i + 1/\kappa).\log(1+\kappa.\mu_i) \\ &+ \log\left(\frac{\Gamma(y_i + 1/\kappa)}{\Gamma(y_i+1).\Gamma(1/\kappa)}\right) \end{aligned} \tag{11}$$

where L is the log-likelihood function, l_i is an individual contribution to the log-likelihood, y_i is the response, μ_i is an estimated mean, and κ is the dispersion parameter.

3. 3. Functional Form Selection and Goodness-of-Fit Statistics

Two groups of functional forms should be investigated to determine the goodness-of-fit (GOF) for the developed nonlinear CPM. The first one is the core functional form, and the second group includes the expanded functional forms. The recommended criteria to estimate the models' goodness-of-fit by NLMIXED procedure in SAS are $-2LL$, AIC, AICC, and BIC. All of which benefit the same rule; smaller values indicate better data fitness [20].

By estimating the likelihood ratio test (LRT) (Equation (12)) and the scaled deviance values, deciding on whether the core model is improved by adding new independent variables will depend on how close these values are to $\chi^2_{0.1,(df_2-df_1)}$ at the 90% confidence level. The LRT and scaled deviance measures are used to test the null hypothesis which remarks all coefficients of the added variables are zero (i.e. $H_0 : \beta_{p+1} = 0$) [4, 13].

$$LRT = -2\log\left(\frac{\text{Likelihood of core model}}{\text{Likelihood of alternative model}}\right) \tag{12}$$

3. 4. Model Error Estimates

The error statistics are practically beneficial to investigate how well the model fits the data. The Mean Absolute Error (MAE) criterion provides a measure of the average misprediction of the model.

In Equation (13), a value close to zero implies that the CPM predicts well the observed data [21-24].

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |O_i - E_i| \tag{13}$$

where n is the number of observations, O_i, is the measure of performance observed from the field data (i.e., the observed ROR crashes) and E_i is the measure of performance, which was estimated by the CPMs (i.e., estimated ROR crash by proposed CPMs).

Mainly, Root Mean Squared Error (RMSE) is the root of Mean Squared Prediction Error (MSPE), which is a standard indicator of the models error. The equation of RMSE is given by Equation (14). Results that are closer to zero represent better data fitness [25-27].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (O_i - E_i)^2} \tag{14}$$

4. MODELING RESULTS AND DISCUSSION

The procedure of developing models is a well-defined forward procedure. Then by considering a proper GOF measure for each functional form, the statistical significance of a new model could be examined. There are three steps for estimating the variables coefficients and their dispersion values. At each step, these two essential notes must be considered:

- The new model contains exposure variables and an added variable that has the best GOF.
- In nonlinear CPMs, that is necessary to control the suggested 39 functional forms for all the new variables.

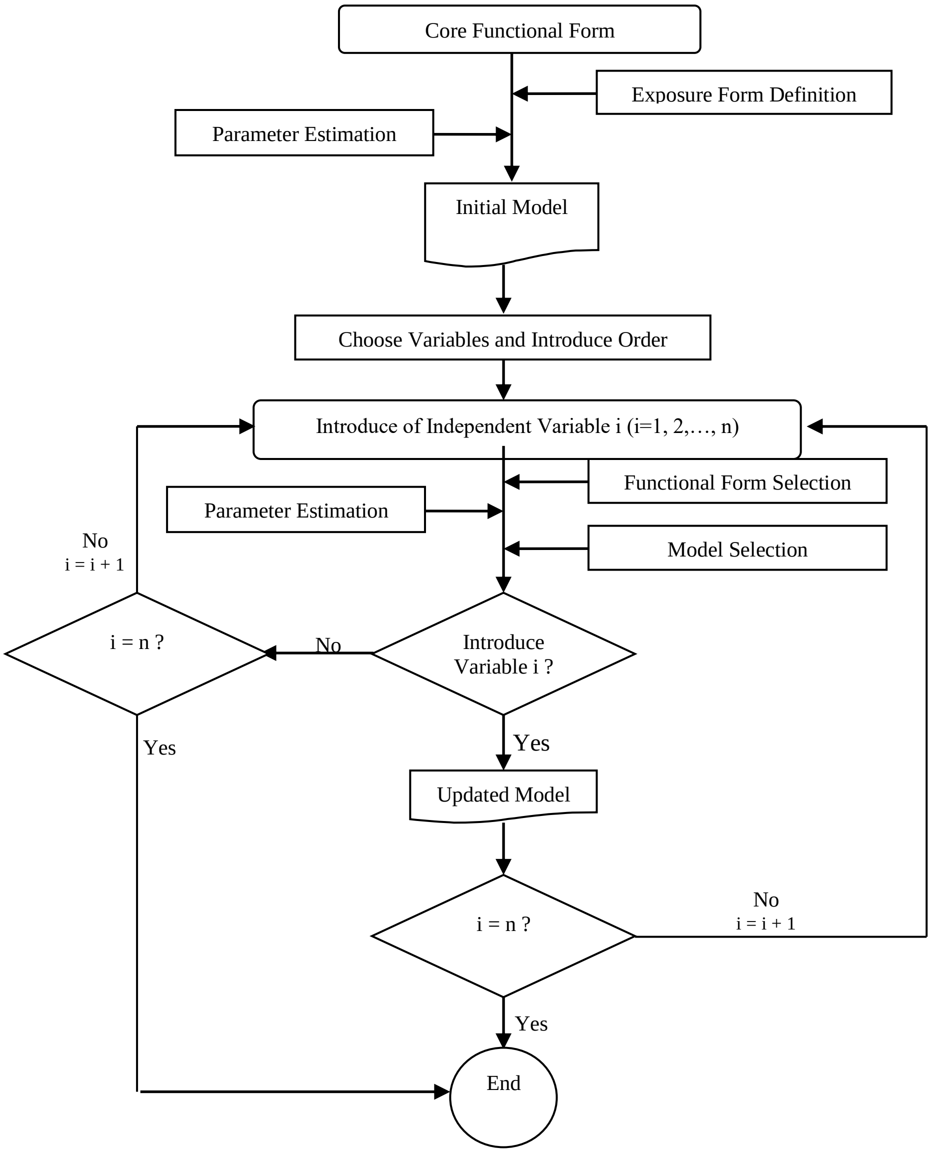


Figure 3. The nonlinear modelling procedure overview

4. 1. Parameter Estimates

4. 1. 1. Generalized Linear CPMs

Based on Equation (6), and concerning both gathered ROR crashes and independent variables data, the Poisson regression (PR) and NB regression (NBR) models are developed (see Table 2).

The variables coefficient indicates the effect of a variable change on the accident frequency value. A change in a given variable by an amount of one percent corresponds to a change in accident frequency by the value of $(\exp(\beta_i)-1)\times 100\%$ changes. The p-values column in Table 2 show that all variables are not statistically significant at the level of 90%. In this regard, the necessity of using nonlinear forms of regression models can be revealed.

The criteria for assessing GLMs' GOF are the scaled deviance ratio and the σ_d value. The obtained DOF is 51. In Table 3, the scaled deviance ratio for PR and NBR are estimated to be equal to 1.6575 and 1.0327, respectively. Further, the σ_d values for PR and NBR are estimated to be equal to 2.7346 and 1.6287, respectively. Comparing these values confirms that the NB regression model is well-fitted to the data much more than the Poisson model. This result reflects the effect of dispersion parameter consideration in the NBR model, which is estimated to be equal to 0.5114.

4. 1. 2. Nonlinear Negative Binomial CPM

First step: The first step in modeling the procedure is the formation of the core model which only contains the exposure variables. The core model performs as a reference model and provides the base structure for introducing an alternative model which includes a newly added variable. The remained explanatory variables are added continuously to update the core model. By using the NLMIXED procedure in SAS, the coefficient of exposure variable converges to the value of 135.39 after 15 iterations.

The calculated $-2LL$ is equal to 172.5 and the BIC criterion is 180.5; these values will be the references to decide how the newly introduced independent variables can improve the CPM. The dispersion parameter value of the model is 0.5579.

Second step: PCI, IRI, and RN are separately added on to the core model, and 39 candidate functions are chosen as link equations between ROR crashes and independent variables. Rather than intrinsic functions, their coefficients, GOF criteria and dispersion values are embedded in Table 4 which examined the PCI alternative link equations.

As shown in Table 4, the 13th functional form has the lowest GOF values among the other functions which means that it has the best fitness for the study data. The β_{1-1} value is the coefficient of exposure variable which should be reassigned for each function.

These functional forms and mentioned procedures are implemented for IRI and RN variables. The results of the second step are listed in Table 5. In this table, the model name FF_{2-1} is the abbreviation of the suggested functional form for the first variable's model (i.e., PCI's model) in the second step.

The p-values of coefficients emphasize that the PCI and RN parameters are significant at the assumed level of 90% but the IRI variable does not satisfy this criterion. Although the $g(PCI)$ and $g(RN)$ functions improved the core model, the PCI and RN variables will remain for the next step. The comparison of GOF criteria shows that the LRT values for all models are equal to 0.5,

TABLE 2. Analysis of parameter estimates about GLMs criteria for assessing GLMs' GOF

	Estimate		Standard Error		Wald 95% Confidence Limits				Chi-Square		Pr > ChiSq	
	PR	NBR	PR	NBR	PR		NBR		PR	NBR	PR	NBR
Intercept	3.4120	3.5032	2.0822	2.6779	-0.669	7.4931	-1.7455	8.7518	2.69	1.71	0.1013	0.1908
PCI	-0.0196	-0.019	0.0116	0.0137	-0.042	0.0030	-0.046	0.0078	2.88	1.95	0.0897	0.1630
IRI	0.145	0.1389	0.233	0.2986	-0.311	0.6016	-0.446	0.7242	0.39	0.22	0.5336	0.6418
RN	0.8469	0.8572	0.6978	0.8748	-0.520	2.2145	-0.8573	2.5717	1.47	0.96	0.2249	0.3271
Dispersion	-	0.5114	-	0.2492	-	-	0.0231	0.9998	-	-	-	-

TABLE 3. Parameters for assessing GLMs' GOF

	Value		Value/DF	
	PR	NBR	PR	NBR
Deviance	84.5309	52.6702	1.6575	1.0327
Pearson Chi-Square	139.4632	83.0612	2.7346	1.6287
Log Likelihood	-64.6943	-60.5226	-	-

TABLE 4. Examined functional forms for PCI independent variable and estimated GOF criteria of the second step

No.	Functional Forms for h(PCI)	Variables' coefficients				Goodness-of-fit criteria				κ Value
		β_{1-1}	β_{2-1}	β_{2-2}	β_{2-3}	-2LL	AIC	AICC	BIC	
1	$1+\beta_{2-1}.\log(x)$	156.66	-0.0331			172.5	178.5	179.0	184.5	0.5552
2	$1+\beta_{2-1}.x$	148.51	-0.0013			172.4	178.4	178.9	184.5	0.5519
3	$\beta_{2-1}+\beta_{2-2}.x$	6.0946	24.368	-0.0313		172.4	180.4	181.2	188.5	0.5519
4	$x^{\beta_{2-1}}$	156.10	-0.0348			172.5	178.5	179.0	184.5	0.5554
5	$\beta_{2-1}.x^{\beta_{2-2}}$	12.4941	12.494	-0.0348		172.5	180.5	181.3	188.5	0.5554
6	$x+\beta_{2-1}.x^2$	5.8231	-0.008			172.2	178.2	178.7	184.2	0.5453

7	$\beta_{2-1}.x+\beta_{2-2}.x^2$	1.6451	3.5398	-0.0285		172.2	180.2	181.0	188.2	0.5453
8	$1+\beta_{2-1}.x+\beta_{2-2}.x^2$	1.7490	3.2892	-0.0265		172.2	180.2	181.0	188.2	0.5450
9	$\beta_{2-1}+\beta_{2-2}.x+\beta_{2-3}.x^2$	2.9694	18.9458	1.2102	-0.01	172.0	182.0	183.3	192.1	0.5412
10	$x^2+\beta_{2-1}.x^3$	5.84E-7	-1862.7			228.4	234.4	234.9	240.5	2.4266
11	$\beta_{2-1}.x^2+\beta_{2-2}.x^3$	-0.1718	-0.9051	0.0085		176.3	184.3	185.1	192.4	0.6445
12	$1+\beta_{2-1}.x^2+\beta_{2-2}.x^3$	-2.3E-8	-3.3E6	22577		188.0	196.0	196.8	204.0	0.9728
13	$x.e^{\beta_{2-1}.x}$	8.4118	-0.0193			172.0	178.0	178.5	184.0	0.5428
14	$\beta_{2-1}.x.e^{\beta_{2-2}.x}$	2.9003	2.9003	-0.0193		172.0	180.0	180.8	188.0	0.5428
15	$1+\beta_{2-1}.x.e^{\beta_{2-2}.x}$	135.39	0.8451	-9.7402		172.5	180.5	181.3	188.5	0.5579
16	$\beta_{2-1}.x+e^{\beta_{2-2}.x}$	1.5034	1.5319	-5.5338		182.5	190.5	191.3	198.5	0.8479
17	$x^{\beta_{2-1}}.e^{\beta_{2-2}.x}$	4.4560	1.2209	-0.0232		172.0	180.0	180.8	188.0	0.5422
18	$e^{\beta_{2-1}.x}$	148.13	-0.0013			172.4	178.4	178.9	184.5	0.5522
19	$\beta_{2-1}.e^{\beta_{2-2}.x}$	12.1707	12.1707	-0.0013		172.4	180.4	181.2	188.5	0.5522
20	$1/(1+\beta_{2-1}.x^9)$	Err	Err	-	-	Err	Err	Err	Err	Err
21	$1/(1+e^{-(\beta_{2-1}+\beta_{2-2}.x)})$	135.39	1.0000	1.001		172.5	180.5	181.3	188.5	0.5579
22	$1/(1+e^{-\beta_{2-1}.x})$	135.39	1.0007			172.5	178.5	179.0	184.5	0.5579
23	$\beta_{2-1}.x/\sqrt{1+x^2}$	11.6373	11.6373			172.5	178.5	179.0	184.5	0.5579
24	$\beta_{2-1}.x/\sqrt{1+\beta_{2-2}.x^2}$	4.0430	4.0430	0.0141		172.5	180.5	181.3	188.5	0.5602
25	$\beta_{2-1}.e^{-\beta_{2-2}.e^{-\beta_{2-3}.x}}$	11.6357	11.6357	1.0000	1.000	172.5	182.5	183.7	192.5	0.5579
26	$(1-\beta_{2-1}.e^{-2.x})/(1+\beta_{2-1}.e^{-2.x})$	135.39	1.0000	1.0000		172.5	180.5	181.3	188.5	0.5579
27	$2.\beta_{2-1}.e^{-x}/(1+\beta_{2-1}.e^{-2.x})$	1.22E7	1.22E7	1.0000		5114	5123	5123	5131	1.1E7
28	$\beta_{2-1}.x/(1+\beta_{2-2}.x)$	16.0773	16.0773	1.8891		172.5	180.5	181.3	188.5	0.5586
29	$1/(1+\beta_{2-1}.x)$	147.74	0.00135			172.4	178.4	178.9	184.5	0.5525
30	$1/(1+\beta_{2-1}.x^2)$	146.07	0.00001			172.4	178.4	178.9	184.4	0.5499
31	$\beta_{2-1}/(1+\beta_{2-2}.x^2)$	12.0861	12.0861	0.00001		172.4	180.4	181.2	188.4	0.5500
32	$1/(1+\beta_{2-1}.x^3)$	145.90	1.62E-7			172.4	178.4	178.8	184.4	0.5480
33	$1/(1+\beta_{2-1}.x^4)$	145.78	1.75E-9			172.4	178.4	178.8	184.4	0.5468
34	$1/(1+\beta_{2-1}.x^5)$	145.48	1.8E-11			172.3	178.3	178.8	184.4	0.5463
35	$1/(1+\beta_{2-1}.x^{\beta_{2-2}})$	135.39	-0.9308	-6.6334		172.5	180.5	181.3	188.5	0.5579
36	$1/(1+x^{\beta_{2-1}})$	135.39	-8.4585			172.5	178.5	179.0	184.5	0.5579
37	$1/(1+\beta_{2-1}.x^6)$	144.97	1.8E-13			172.4	178.4	178.8	184.4	0.5464
38	$1/(1+\beta_{2-1}.x^7)$	144.31	1.7E-15			172.4	178.4	178.8	184.4	0.5469
39	$1/(1+\beta_{2-1}.x^8)$	140.66	1.1E-17			172.4	178.4	178.9	184.4	0.5514

and they have the same priority to be selected for the next step. By considering the values $\chi^2_{0.1,1}$, these new models do not satisfy the significance requirement. It is because the LRT values are less than 2.71. Nevertheless, this small improvement rejects the null hypothesis and in the next step by adding other remained variables, CPMs will fit quite well to the data. It is essential to mention that the user must carefully interpret LRTs when the values fall under the null hypothesis boundary.

Third step: At this stage, the second variable is added to previous models. It should be noted that the adding order of these variables changes the results because of the effect of functional form selection at each stage. By considering the GOF criteria of the given models in Table 6, it appears that the FF_{3-1} model has the best fit among the others.

The models name FF_{3-1} is the abbreviation of the suggested functional form for the new variable (i.e., RN) at the third step. The coefficients' p-value emphasizes that the coefficients of the FF_{3-1} model are significant at the level of around 90%, but the coefficients of the FF_{3-2} model do not satisfy this criterion. The LRT value for the FF_{3-1} model is 2.1 (concerning the FF_{2-1} model) and compared to $\chi^2_{0.1,1}$ distribution with one DOF (i.e., the value of 2.71), the new model significantly improves the prior model. The final form of nonlinear CPM is defined as Equation (15).

$$\begin{aligned} &ROR - Accident_{nonlinear\ CPM} \\ &= 6.4521 \times Exposure \times (PCI \times e^{-0.03331.PCI}) \times (e^{0.4735.RN}) \end{aligned} \tag{15}$$

4. 2. Safety Effects of Independent Variables

All GOF criteria and p-values of the suggested non-linear model (i.e., Model FF_{3-1} which is represented in Equation (15) reflect that all variables have a statistically significant effect on the ROR accident occurrence. The independent variables coefficients of the model clarify how these explanatory parameters are associated with the total frequency of single-vehicle ROR crashes. The positive coefficient reflects the direct relationship between the accident frequency and the independent variable. On the contrary, it has a reverse effect if the coefficient is negative. Based on this point of view, the CPM's variables will be discussed below.

4. 2. 1. PCI Variable The negative sign for PCI's coefficient predicts that segments with higher PCI values will probably have lower ROR crash rates. The exponential coefficient for PCI is equal to 0.9672 indicates that the ROR crash rate is reduced by a factor of 0.967. The percentage change in accident rate associated with 1% increase in the PCI is -3.31%. The 1.01 multiplier refers to the form of Model FF_{3-1}; in other words, it shows a 1% increase in PCI in addition to its exponential coefficient increment.

TABLE 5. Estimated coefficients and GOF criteria for independent variables of the second step

Model name	Variable	Functional Forms for g(.)	Variables' coefficients		Goodness-of-fit criteria				κ (p-value)
			β_{1-1} (p-value)	β_{2-1} (p-value)	-2LL	AIC	AICC	BIC	
FF_{2-1}	g(PCI)	$PCI.e^{\beta_{2-1}.PCI}$	8.4118 (0.0195)	-0.0193 (0.001)	172.0	178.0	178.5	184.0	0.5428 (0.0397)
FF_{2-2}	g(IRI)	$1/(1+\beta_{2-1}.IRI^4)$	143.06 (<.0001)	0.000082 (0.5877)	172.0	178.0	178.5	184.1	0.5573 (0.0353)
FF_{2-3}	g(RN)	$RN.e^{\beta_{2-1}.RN}$	161.99 (0.0330)	-0.3956 (0.0197)	172.0	178.0	178.4	184.0	0.5546 (0.0363)

TABLE 6. Estimated coefficients and GOF criteria for independent variables of the third step (p-values are given in parenthesis)

Model name	Variables' order	Functional Forms for h(.)	Variables' coefficients			Goodness-of-fit criteria				κ (p-value)
			β_{1-1} (p-value)	β_{2-1} (p-value)	β_{3-1} (p-value)	-2LL	AIC	AICC	BIC	
FF_{3-1}	g(PCI)×h(RN)	$e^{\beta_{3-1}.RN}$	6.4521 (0.032)	-0.033 (0.005)	0.4735 (0.110)	169.9	177.9	178.7	185.9	0.5115 (0.04)
FF_{3-2}	g(RN)×h(PCI)	$1+\beta_{3-1}.PCI$	140.51 (0.053)	-0.063 (0.839)	-0.0069 (0.011)	170.4	178.4	179.2	186.4	0.5154 (0.04)

This outcome is logical and compatible with experts expectations and drivers experiments. Distresses force driver to change his direction rapidly, so this fast reaction leads to careless steering most of the time and a ROR crash is consequently expectable.

4. 2. 2. RN Variable The functional form of Model FF3-1 clarifies that as the riding quality grows up, the ROR crashes increases. Referring to Equation (15), the RN's exponential coefficient is equal to 1.6056 and an increase in the RN value by a unit will result in a 60.56% increase in the ROR accident rate. The RN's coefficient sign is positive; that means if the RN value increases, higher accident rates will be expected. As mentioned in section 2.3.3., RN is a subjective riding quality criterion. This definition confirms that the riders anticipations and reactions have an important role in the ROR accident occurrence. The higher RN values allow the rider to drive monotonously and inattentively; by considering such behaviors, the results of the model will be acceptable and consistent with what has been expected.

4. 3. Models' Error Structure As mentioned before, to justify the external validation and the overall accuracy of CPM, It is necessary to investigate the error statistics. The MAE value equals to 1.09 and confirms the acceptability of model fitness on the data. Furthermore, the RMSE value is equal to 1.5 which approves the accuracy of suggested nonlinear negative binomial CPM.

5. CONCLUSION

This study aimed to investigate the extent of the pavement physical characteristics and riding quality effects on the occurrence of single-vehicle ROR crashes on two-lane separated rural roadways. For this purpose, three types of databases were obtained, including AADT, physical pavement features (including length, PCI, IRI and RN of segments) and ROR crash data. By obtaining the raw data and processing them, 55 homogenous segments with a total length of 150.5 km were categorized to develop reliable CPM. The examined CPMs were classified into two different models; the generalized linear regression models (GLMs) and the nonlinear multivariate negative binomial regression model. These regression models were developed to estimate the variables coefficients and other GOF criteria.

Literature reviews showed that researchers agree with developing CPMs which fits well on the data with small counts such as ROR crashes. The results of developed GLMs imply that the models variables are not significant at the assumed level. In order to achieve this goal non-linear forms of negative binomial multivariate regression models were investigated. The proposed model indicated that the explanatory variables are complicated enough to use a non-linear model for rural roads accidents.

Based on GOF criteria, the proposed non-linear model (i.e., Equation (15)) statistically provides significant improvements in the model fit. The p-values of variables coefficients show that the PCI and RN variables are significant but IRI could not satisfy this criterion in modeling. Also, the given values of error estimate measures imply that Model FF_{3-1} provides quite reliable predictions of ROR crash occurrence related to pavement physical characteristics.

The sign and coefficient of variables reasonably follow the expected outcomes. The proposed model shows that an increment in the PCI (which is an objective riding quality criterion) will cause a drop in the ROR crash frequency. Based on the suggested functional form of Eq. (15), a unit improvement in PCI (as a pavement manner criterion) corresponds to a 3.31 percent reduction in single-vehicle ROR crashes. This effectiveness emphasizes the importance of pavement management and maintenance programs. Therefore, the road safety authorities are responsible for periodically controlling the pavement condition and following the scheduled maintenance and repairing programs.

The RN's coefficient represents a different attitude; however, such interpretations become acceptable if we consider the concept of subjective riding quality that refers to the drivers anticipation. The variables coefficient shows that a unit increase in RN value associates with a 60.56 percent increase in the ROR crash rate. The RN is calculated from longitudinal profile measurements and is used to estimate subjective ride quality. Hence, that is not used to measure any pavement roughness or distress. This variable defines the comfort level of riding and relates to the nature of the longitudinal roadway profile. Highway engineers usually prefer to construct infrastructures that have high grades of RN. However, the results of this study reveal that these high grades most likely associate with higher single-vehicle ROR crashes. That is since the drivers seem to not expect a dangerous situation and they usually follow their steadily riding, which might result in driving weakness and/or drowsiness.

6. REFERENCES

1. Fitzpatrick, C.D., Harrington, C.P., Knodler Jr. M.A., Romoser, M.R.E., "The influence of clear zone size and roadside vegetation on driver behavior", ***Journal of Safety Research,*** Vol. 49, (2014), 97-104. doi: 10.1016/j.jsr.2014.03.006

2. Carrigan, C.E., Ray, M.H., "A new approach to run-off-road crash prediction", Proceeding of the 96th Annual Meeting Compendium of TRB, Washington D.C., United States, (2017). https://www.roadsafellc.com/linked/carrigan17h.pdf

3. Torre, F.L., Tanzi, N., Yannis, G., Dragomanovits, A., Richter, T., Ruhl, S., Karathodorou, N., Graham, D., "Accident prediction in

European countries: Development of a practical evaluation tool", Proceeding of the 7th Transport Research Arena (TRA), Vienna, Austria, (2018).

4. Theofilatos, A., Yannis, G., "A review of the effect of traffic and weather characteristics on road safety", ***Accident Analysis and Prevention,*** Vol. 72, (2014), 244-256. doi: 10.1016/j.aap.2014.06.017

5. Ambros, J., Valentová, V., Sedoník, J., "Developing updatable crash prediction model for network screening: Case study of Czech two-lane rural road segments", ***Journal of the Transportation Research Board,*** Vol. 2583, (2016), 1-7. doi: 10.3141/2583-01

6. Akbari, M., Shafabakhsh, Gh., Ahadi, M.R., "Evaluating the safety effects of pavement condition index (PCI) on frequency of run-off-road accidents", ***Journal of Transportation Infrastructure Engineering,*** Vol. 1, No. 3, (2015), 47-61. doi: 10.22075/jtie.2015.316

7. Chan, C.Y., Huang, B., Yan, X., Richards, S., "Investigating effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS)", ***Journal of Advanced Transportation,*** Vol. 44, No. 3, (2010), 150-161. doi: 10.1002/atr.129

8. Jiang, X., Huang, B., Zaretzki, R.L., Richards, S., Yan, X., "Estimating safety effects of pavement management factors utilizing Bayesian random effect models", ***Traffic Injury Prevention,*** Vol. 14, No. 7, (2013), 766-775. doi: 10.1080/15389588.2012.756582

9. Jafari, R., Hummer, J.E., "Safety effects of access points near signalized intersections", Proceeding of the 92nd Annual Meeting Compendium of TRB, Washington D.C., United States, (2013).

10. Ashuri, A., Amiri, A., "Drift change point estimation in the rate and dependence parameters of autocorrelated poisson count processes using MLE approach: An application to IP counts data", ***International Journal of Engineering, Transactions A: Basics,*** Vol. 28, No. 7, (2015), 1021-1030. doi: 10.5829/idosi.ije.2015.28.07a.08

11. Shafabakhsh, Gh., Famili, A., Akbari, M., "Spatial analysis of data frequency and severity of rural accidents", ***Transportation Letters,*** Published Online: 08 Mar 2016, (2016). doi: 10.1080/19427867.2016.1138605

12. Roque, C., Cardoso, J.L., "Investigating the relationship between run-off-the-road crash frequency and traffic flow through different functional forms", ***Accident Analysis and Prevention,*** Vol. 63, (2014), 121-132. doi: 10.1016/j.aap.2013.10.034

13. van Petegema, J.W.H.(J.H.), Wegman, F., "Analyzing road design risk factors for run-off-road crashes in the Netherlands with crash prediction models", ***Journal of Safety Research,*** Vol. 49, (2014), 121-127. doi: 10.1016/j.jsr.2014.03.003

14. ASTM D6433-11, "Standard practice for roads and parking lots pavement condition index (PCI) surveys", American Standard, (2011). doi: 10.1520/D6433

15. ASTM E1926-08, "Standard practice for computing international roughness index (IRI) of roads from longitudinal profile measurements", American Standard, (2015). doi: 10.1520/E1926

16. ASTM E1489-08, "Standard practice for computing ride number (RN) of roads from longitudinal profile measurements made by an inertial profile measuring device", American Standard, (2013). doi. 10.1520/E1489

17. Basu, S., Saha, P., "Regression models of highway traffic crashes: A review of recent research and future research needs", ***Procedia Engineering,*** Vol. 187, (2017), 59-66. doi: 10.1016/j.proeng.2017.04.350

18. Wood, A.G., Mountain, L.J., Connors, R.D., Maher, M.J., Ropkins, K., "Updating outdated predictive accident models", ***Accident Analysis and Prevention,*** Vol. 55, (2013), 54-66. doi: 10.1016/j.aap.2013.02.028

19. Ye, Z., Zhang, Y., Lord, D., "Goodness-of-fit testing for accident models with low means", ***Accident Analysis and Prevention,*** Vol. 61, (2013), 78-86. doi: 10.1016/j.aap.2012.11.007

20. Hosseinpour, M., Yahaya, A.S., Sadullah, A.F., Ismail, N., Ghadiri, S.M.R., "Evaluating the effects of road geometry, environment, and traffic volume on rollover crashes", ***Transport,*** Vol. 31, No. 2, (2016), 221-232. doi: 10.3846/16484142.2016.1193046

21. Sharifi, Y., Hosseinpour, M., "A predictive model based ANN for compressive strength assessment of the mortars containing metakaolin", ***Journal of Soft Computing in Civil Engineering,*** Vol. 4, No. 2, (2020), 1-11. doi: 10.22115/scce.2020.214444.1157

22. Naderpour, H., Rezazadeh, D., Fakharian, P., Rafiean, A.H., Kalantari, S.M., "A new proposed approach for moment capacity estimation of ferrocement members using group method of data handling", ***Engineering Science and Technology: An International Journal,*** Vol 23, No. 2, (2020), 382-391. doi: 10.1016/j.jestch.2019.05.013

23. Ghannadiasl, A., Rezaei Dolaghb, H. "Sensitivity analysis of vibration response of railway structures to velocity of moving load and various depth of elastic foundation", ***International Journal of Engineering, Transactions C: Aspects,*** Vol 33, No. 3, (2020), 401-409. doi: 10.5829/ije.2020.33.03c.04

24. Ghazvinian, H., Karami, H., Farzin, S., Mousavi, S. "Effect of MDF-cover for water reservoir evaporation reduction, experimental, and soft computing approaches", ***Journal of Soft Computing in Civil Engineering,*** Vol 4, No. 1, (2020), 98-110. doi: 10.22115/scce.2020.213617.1156

25. Naderpour, H., Rafiean, A.H., Fakharian, P. "Compressive strength prediction of environmentally friendly concrete using artificial neural networks", ***Journal of Building Engineering,*** Vol 16, (2018), 213-219. doi: 10.1016/j.jobe.2018.01.007

26. Ghasemi, S., Bahrami, H., Akbari, M. "Classification of seismic vulnerability based on machine learning techniques for RC frames", ***Journal of Soft Computing in Civil Engineering,*** Vol. 4, No. 2 (2020), 13-21. doi: 10.22115/scce.2020.223322.1186

27. Lashkenari, M.S., KhazaiePoul, A., Ghasemi, S., Ghorbani, M., "Adaptive neuro-fuzzy inference system prediction of Zn metal ions adsorption by γ-Fe2o3/Polyrhodanine nanocomposite in a fixed bed column", ***International Journal of Engineering, Transactions A: Basics,*** Vol. 31, No. 10, (2015), 1617-1623. doi: 10.5829/ije.2018.31.10a.02

Persian Abstract

چکیده

هدف از این مطالعه، ارزیابی تاثیر مشخصات فیزیکی رویه راه روی فراوانی تصادفات خروج از جاده در جاده‌های دوخطه مجزا می‌باشد. به همین منظور و برای ارائه یک مدل پیش‌بینی دقیق، نویسندگان سعی کردند مدل‌های خطی تعمیم‌یافته‌ای را ارائه دهند که شامل رگرسیون پواسون، رگرسیون دوجمله‌ای منفی و رگرسیون دوجمله‌ای منفی غیرخطی می‌شود. علاوه بر پارامترهای درمعرض‌بودن، برخی متغیرهای توصیفی مربوط به خصوصیات فیزیکی رویه راه مانند شاخص وضعیت رویه راه، شاخص ناهمواری بین‌المللی راه و عدد سواری نیز در مدلسازی لحاظ شدند. برای مدلسازی، از فرآیند پیش‌رونده بهره گرفته شده است که در آن، متغیرها به ترتیب به مدل اولیه (هسته اصلی) افزوده می‌شوند. در فرآیند مدلسازی غیرخطی و در هر مرحله، ۳۹ فرم ساختاری کنترل شدند تا مشخص شود که مدل جدید آیا برازش بهتری نسبت به مدل اولیه یا مدل قبلی داشته است یا خیر. ابزارهای متعددی برای تخمین نیکویی برازش مدل مورد آزمون قرار گرفتند. همچنین، از ابزارهای دیگری نیز برای تخمین اعتبار بیرونی و ساختار خطای مدل‌ها بهره گرفته شده است. نتایج نشان دادند که در مدل‌های رگرسیون پواسونی و رگرسیون دوجمله منفی، ضرایب متغیرها معنی‌دار نبودند. یافته‌های مدل غیرخطی پیشنهادی تایید کردند که متغیر شاخص وضعیت روسازی به عنوان یک متغیر عینی، از انتظارات متخصصان تبعیت می‌کند (به عبارتی، وضعیت رویه بهتر با تصادفات خروج از جاده کمتر همبستگی دارد). نهایتاً، شایان ذکر است که متغیر ناهمواری در سطح معنی‌داری مفروض، معنی‌دار نبود و بنابراین، سهمی در تصادفات خروج از جاده ندارد. نتایج تاکید دارند که بهبود وضعیت روسازی منجر به کاهش احتمالی بیشتری در فراوانی تصادفات خروج از جاده می‌شود.

International Journal of Engineering

Journal Homepage: www.ije.ir

Numerical Simulation of Hydrodynamic Properties of Alex Type Gliders

K. Divsalar, R. Shafaghat*, M. Farhadi, R. Alamian

Sea-Based Energy Research Group, Babol Noshirvani University of Technology, Babol, Iran

PAPER INFO

Paper history:
Received 04 May 2020
Received in revised form 19 may 2020
Accepted 12 June 2020

Keywords:
Underwater Gliders
Numerical Simulation
Finite Volume Method
Hydrodynamic Coefficients
Angle of Attack

ABSTRACT

This work presents a numerical Simulation of an underwater glider to investigate the effect of angle of attack on the hydrodynamic coefficients such as lift, drag, and torque. Due to the vital role of these coefficients in designing the controllers of a glider, and to obtain an accurate result, this simulation has been carried on at a range of operating velocities. The total length of the underwater glider with two wings is 900 mm with a 4-digits NACA0009 profile. The fluid flow regime is discretized and solved by computational fluid dynamics and finite volume method. Since the Reynolds number range for this study is in a turbulent flow state (up to 3.7e06), the κ-ω SST formulation was used to solve Navier-Stokes equations and continuity and the angles of attack ranging are from - 8 to 8 degrees. The main purpose of this research is to study the effect of each of the dynamics parameters of glider motion such as velocity and angle of attacks on the hydrodynamic coefficients. Based on the results, the drag and lift coefficients are enhanced with increasing the angle of attack. In addition, the drag coefficient enhanced with increasing the velocity however, when the glider velocity is increased, the lift coefficient does not change significantly except at the highest angle of attack that decreases. The highest drag coefficient is 0.0246, which corresponds to the angle of attack of -8 and the Reynolds number of 3738184. In addition to simple geometry, the glider studied in this paper shows relatively little resistance to flow.

***doi**: 10.5829/ije.2020.33.07a.26*

NOMENCLATURE

A	Reference area (m2)	p	Static pressure (Pa)
L	Reference Length (m)	AoA	Angle of attack (degree)
CD	Drag coefficient (-)	**Greek Symbols**	
CL	Lift coefficient (-)	κ	Turbulence kinetic energy (m2s-2)
CM	Moment coefficient (-)	ν	Kinematic viscosity (m2s-1)
Re	Reynolds number (-)	ρ	Density (kgm-3)
U	Mean stream velocity (m/s)	ω	Specific dissipation rate (s-1)
A	Reference area (m2)	p	Static pressure (Pa)
L	Reference Length (m)	AoA	Angle of attack (degree)
CD	Drag coefficient (-)	**Greek Symbols**	
CL	Lift coefficient (-)	κ	Turbulence kinetic energy (m2s-2)
CM	Moment coefficient (-)	ν	Kinematic viscosity (m2s-1)
Re	Reynolds number (-)	ρ	Density (kgm-3)

1. INTRODUCTION

The exploitation of the oceans and seas is of paramount importance in today's world in terms of transportation, trade, food and pharmaceutical resources, mineral resources and coastal security [1-6]. Underwater glider vehicles are widely used in the monitoring of oceans, exploring and studying various submarine related topics as well as the understanding of global oceanographic phenomena. Therefore, many researchers have investigated the parameters affecting the dynamics of glider motions [7]. They are categorized into three major groups: manned, remote-controlled and automatic underwater gliders [8, 9]. Manned underwater gliders,

*Corresponding Author Institutional Email: *rshafaghat@nit.ac.ir* (R. Shafaghat)

Please cite this article as: K. Divsalar, R. Shafaghat, M. Farhadi, R. Alamian, Numerical Simulation of Hydrodynamic Properties of Alex Type Gliders, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1387-1396

despite their favorable impact on marine research, are very expensive and time-consuming while gliders, remote control submarines, require less training to control and have fewer security problems. Besides, these types of underwater gliders are capable of onshore controlling and they show high maneuverability needed in surveys over extended periods. The third type of underwater gliders, i.e., automatic underwater gliders, carry their required energy and are capable of performing predefined operations [10]. Underwater gliders are a new generation of automatic gliders that are capable of moving vertically and horizontally underwater by changing their weight. The wings of these gliders help them to propel and control pitch behavior. The unique feature of these gliders is their ability to make exploring missions that take weeks and months. These gliders use buoyancy to propel them and require no further energy. Very low running costs, low power losses and low noise are some of the features that are useful in long time marine surveys. The main purpose of underwater gliders is to obtain information through sensors such as conductivity, temperature and depth gauges that are transmitted to the control center [11].

Over the past few decades, developments in the field of computational fluid dynamics have greatly helped to save costs and to adequately estimate issues such as the study of hydrodynamic behavior [12, 13]. Here are some of the studies that have been done for simulation of underwater gliders. Du et al. [14] presented an analysis of the hydrodynamic characteristics of gliders moving near the ocean floor. The importance of their method is in the glider maneuver, because of changes in vessels hydrodynamic characters near the ocean floor. They later investigated the impact of the underwater glider wings with the help of computational fluid dynamics to determine the effect of glider acceleration and stability [15]. Chen et al [11], hydro-dynamically analyzed a submersible glider for the role of the glider wing in a non-uniform flow. Jung et al. [16] studied the effects of changes in pitch of the propeller and its applications such as backward movement in tunnels. Gao et al. [17] studied a numerical model of a glider with two wings on each side that is angled to the body. The purpose of this research is to investigate and optimize the effect of underwater glider buoyancy motor size on its stability. Barros and Dantas [18] investigated the effect of the ductile propeller glider on the buoyancy forces at different angles of attack and its maneuverability, using κ-ω SST formulation. In a CFD study, they analyzed the combined effect of the control surface deviation and also the angle of attack on the hydrodynamic forces applied to the Pirajuba automated underwater gliders [19]. Ray et al. [20] evaluated the hydrodynamic characteristics of their gliders reaching speeds of up to 6 knots (3 m/s). In their research, the water flow velocity was different from the glider that reached up to 2 knots. Later in a study, Joung et al. [21] optimized the glider geometry to obtain a reduction in the drag coefficient and glider resistance. The hydrodynamic coefficients of a spherical underwater glider for three different types of motions have been investigated by Yu et al. [22]. Zheng et al. [23] studied their capsular underwater glider, using computational fluid dynamics, which has four thrust impellers. Singh et al. [24] also calculated their glider hydrodynamic coefficients using a numerical method and compared them with their experimental results.

Noman et al. [25] extracted their hydrodynamic coefficients, which were almost constant considering the variations in velocity and angle of attack. Lin et al. [26] optimized their 2.7 meters length vessel by using genetic algorithms. In this study, cloud points are obtained from the finite volume simulation method, and by modeling the vessel near the surface they were able to reduce the applied force by 28.9%. Nedelcu et al. [27], were inspired by the body of fish, proposed a glider and calculated the forces acting on it by CFD. Javaid et al. [28] studied the effects of glider wings numerically and experimentally and modeled glider movement in the rotational case. In their research, reducing the wing thickness along its length increases lift force and reduces dynamic stability. Liu et al. [29] applied the hydrodynamic coefficients obtained from the numerical method of rotational motion of the glider with good accuracy according to the experimental results using the dynamic model method. Javaid et al. [30] studied the hydrodynamic properties of their gliders in different conditions. The results of the numerical method were obtained with high accuracy compared to the experimental values.

Here, in this paper, the finite volume method is also used in computational fluid dynamics to study the hydrodynamic behavior of a model of underwater glider at different speeds and angles of attack to investigate the effects of these parameters on the hydrodynamic coefficients such as lift, drag, and torque.

It can be noted that the importance of this study is in the design of the controller for this system which plays an important role in the control and steering of the vehicle.

The remainder of this research is arranged as follows. In section 2, the geometry of underwater glider with a spherical nose is presented. The governing equations and the boundary conditions are derived. In section 3, a finite volume method for numerical solution is developed and the results are discussed. Finally, concluding remarks are provided in section 4.

2. METHOD and ASSUMPTIONS

2. 1. Geometry In this section, after defining the geometry of the glider under study, the assumptions for numerical analysis are given. Figure 1 shows the geometry of the proposed glider with a spherical nose. The overall length of the glider is 90 cm and its diameter to length ratio is 0.1. The aerodynamic center of the two wings with a 4-digit NACA0012 profile is located 43 cm from its nose tip. As shown in the figure, the width and length of the wings are 12.7 and 39.5 cm, respectively.

2. 2. The Governing Equations The two-equation perturbation model κ-ω SST has been used to analyze the proposed glider [31]. This formulation in the boundary layer portions makes this model directly applicable to the viscous substrate. Therefore, this model can be used as a model of low Reynolds perturbation without any additional damping function. In free flow, the SST formulation changes to κ-ε [32], thereby avoiding the common problem of the κ-ω model, which is sensitive to the perturbation properties of free-flow. In the present model, kinematic eddy viscosity is defined as follows [31]:

$$\vartheta_T = \frac{a_1 \kappa}{\max(a_1\omega, SF_2)} \tag{1}$$

where κ is the perturbation kinetic energy and ω the dissipation rate of this energy. a_1 is a constant value, $S = \frac{\partial u}{\partial y}$ and F_2 a function that adopts a value equal to 1 for boundary layer flow and zero for non-shear stress. The equations used are Navier-Stokes, continuity, and turbulence models, respectively, as follows [33]:

$$\rho\frac{\partial \boldsymbol{u}}{\partial t} + \rho(\nabla\cdot\boldsymbol{u})\boldsymbol{u} = \nabla\cdot[-p\boldsymbol{I} + (\mu+\mu_T)(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T)] \tag{2}$$

$$\rho\nabla\cdot\boldsymbol{u} = 0 \tag{3}$$

$$\frac{\partial\kappa}{\partial t} + U_j\frac{\partial\kappa}{\partial x_j} = P_\kappa - \beta^*\kappa\omega + \frac{\partial}{\partial x_j}\left[(\nu + \sigma_\kappa\nu_T)\frac{\partial\kappa}{\partial x_j}\right] \tag{4}$$

$$\frac{\partial\omega}{\partial t} + U_j\frac{\partial\omega}{\partial x_j} = \alpha S^2 - \beta\omega^2 + \frac{\partial}{\partial x_j}\left[(\nu + \sigma_\omega\nu_T)\frac{\partial\omega}{\partial x_j}\right] + 2(1-F_1)\sigma_{\omega^2}\frac{1}{\omega}\frac{\partial\kappa}{\partial x_j}\frac{\partial\omega}{\partial x_j} \tag{5}$$

The function F_1 is equal to one near the wall and zero at the other.

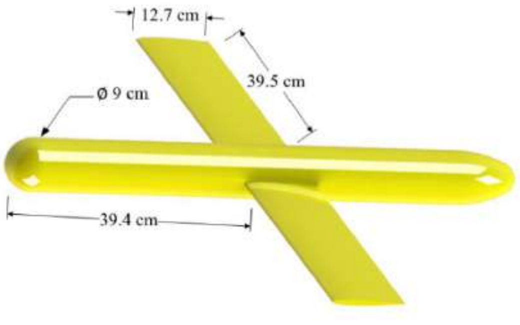


Figure 1. The shape and dimensions of the 90 cm long underwater glider with a spherical nose

Fluid effects on glider generally include drag force ($\mathrm{F_D}$), lift force ($\mathrm{F_L}$), and rotational torque (M), which are used to calculate the drag and lift coefficients, and torque in Equations (6), (7), and (8), respectively [24].

$$\mathrm{C_D} = \frac{\mathrm{F_D}}{\frac{1}{2}\rho\,\mathrm{U^2A}} \tag{6}$$

$$\mathrm{C_L} = \frac{\mathrm{F_L}}{\frac{1}{2}\rho\,\mathrm{U^2A}} \tag{7}$$

$$\mathrm{C_M} = \frac{M}{\frac{1}{2}\mathrm{\rho U^2Ac}} \tag{8}$$

where A is the reference area for calculations, which is considered as the total area of the glider. C_D is the drag coefficient, C_L the lift coefficient, U the mean fluid velocity, and c the length of the airfoil chord.

2. 3. Boundary Conditions The boundary conditions include velocity inlet, turbulent intensity, and turbulent characteristic length for the facing boundaries, zero relative pressure at the output, the symmetry condition for the underwater glider cutting plate, and the non-slip boundary condition for the underwater glider boundaries of the computational environment. It should be noted that the symmetry condition is intended to reduce the computation and its application creates the condition that the same flow pattern appears on the other side of the boundary with acceptable accuracy.

3. Results and Discussion

3. 1. Validation of Numerical Method In order to validate the numerical solution of the present study and to determine the hydrodynamic coefficients of underwater gliders, the studies of Isa et al. [34] were used. In their work, the hydrodynamic coefficients were evaluated by using a Strip theory and a computational fluid dynamics by modeling the k-ω SST turbulence theory with a Reynolds number greater than 10^6 and symmetry boundary condition. It can be noted that in the current research range of Reynolds number is wider than the that of Isa et al. [34]. For assuring the independence of meshing size, three patterns have been chosen. In all of these cases, the angle of attack is about 8 degrees and the velocity of flow is 4.26 m/s (the ultimate condition). These three models have 2.12 million, 3.15 million and 4.01 million elements, respectively. Table 1 shows hydrodynamic coefficients and error values in terms of the third level. As can be seen in Table 1, second level shows accurate results in the time-cost calculation.

The tetrahedron elements are used and the total number of elements in this mesh is 3150396. The pattern used is shown in Figure 2. Modeling the impact of half of the underwater glider on the surrounding flow is considered due to geometrical symmetry. Figure 3

TABLE 1. Results and comparison of different mesh sizes

Grid Level	Grid Number/10^6	Drag Coefficient		Lift Coefficient		Moment Coefficient	
		Value	Error	Value	Error	Value	Error
1	2.12	0.070038	6.7%	0.141256	3.6%	0.005584	7.8%
2	3.15	0.065694	1.6%	0.13633	1.0%	0.005189	2.4%
3	4.01	0.064634	-	0.134951	-	0.005069	-

compares the numerical results of the two methods, which shows the reasonable accuracy of the modeling performed in the present study by the ANSYS-FLUENT R.18.2 commercial software used to solve the flow in the domain.

A comparison of the results of the numerical method with the experimental tests is also provided to ensure the performance of the glider and confirm the accuracy of the simulation. The system needed to perform the test is called the towing tank. Figure 4 shows an overview of the designed glider sample. In addition to the traction system, the towing tank also has data measurement and reporting devices. In order to measure the drag coefficient in different Froude numbers in the towing tank with 38 m long, 3 m wide and 2.5 m deep, the tips of the wings are positioned sufficiently distant from the walls to minimize the effects of the walls on the wing flow. To measure hydrodynamic coefficients, a six-component dynamometer is used. For evaluation of the uncertainty of towing tank system, two types of uncertainty have been evaluated; in this regard, in order to assess the statistical uncertainty, the experiments were repeated 5 times. The overall uncertainty for the towing tank measurement system was evaluated to be 4%. After extracting the coefficients from the experiments, the results are presented along with the numerical simulation results in Figure 5. The convergence criteria is about 10^{-5} in this numerical simulation. As can be seen, the results of both numerical and experimental methods are in good agreement. The maximum error is evaluated to be 9%.

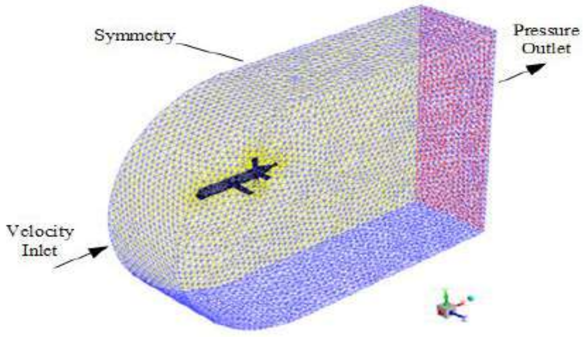


Figure 2. Meshing model investigated by Isa et al. [34] for validation

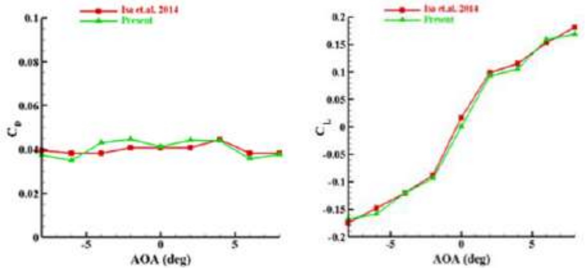


Figure 3. Comparison of the numerical results of the drag and lift coefficient in terms of angle of attack (AOA) at 2.5 m/s

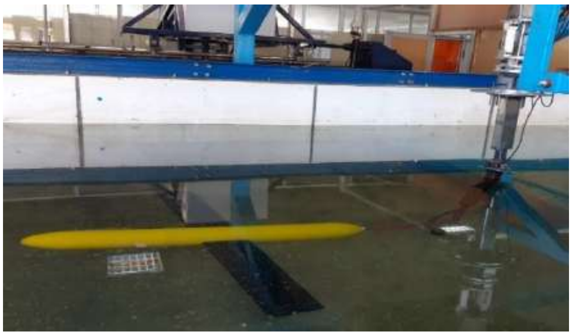


Figure 4. Underwater glider specimen made in a towing tank test

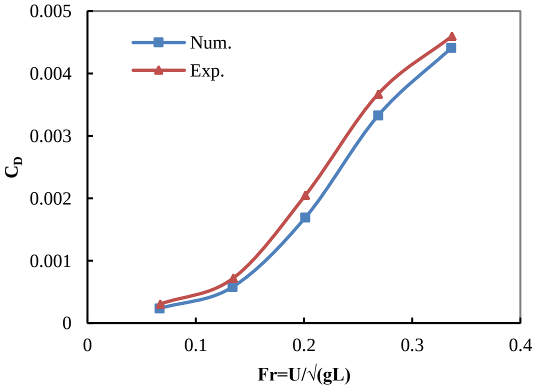


Figure 5. Comparison of numerical and experimental results for different Froude numbers at zero angle of attack

3. 2. Meshing

Figure 6 shows the meshing domain for the glider in the present study. The total number of elements in the domain is 3,894,591. Here, as in the model used in validation, symmetry is used and half of the underwater glider is modeled to reduce computation time and cost. As can be seen in this figure, the disorganized tetrahedron mesh is used for analysis. Due to the high gradient in the vicinity of the underwater glider and the turbulence of the stream, the mesh density is considered higher in that area. The diameter of the created domain is 5 times the length of the glider, which is taken to increase the resolution accuracy [22, 25, 35-37].

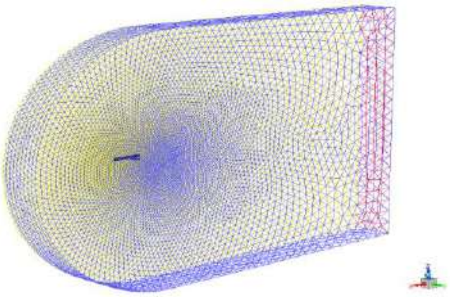


Figure 6. Computational environment and meshing intended for the numerical solution

3. 3. The Effect of Velocity and Angle of Attack

The effects of velocity and angle of attack on the drag and lift coefficient for five different Reynolds numbers with values of Re = 747,636.7, 1,495,273, 2,242,910, 2,990,547 and 3,738,184 in terms of 9 different angles of attack from -8 to 8 degrees (intervals 2 Degree) are presented. These values are selected to evaluate the vast range of operating conditions of the proposed glider. As in laboratory conditions, the sample is kept fixed and it moves with its defined velocity at different input angles of attack. Figures 7 and 8 show the glider lift coefficient in terms of angles of attack and velocity, respectively. The two diagrams show that the lift coefficient increases with increasing the angle of attack. However, when the glider velocity is increased, the lift coefficient does not change significantly in all other cases except at the highest angle of attack that decreases.

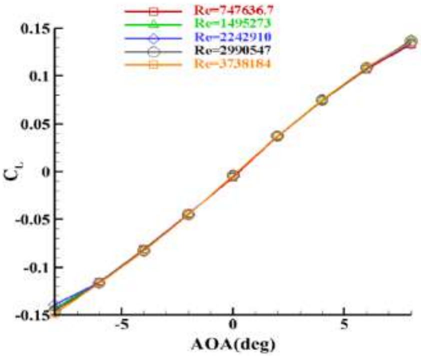


Figure 7. lift coefficient in term of the angle of attack for different flow velocities

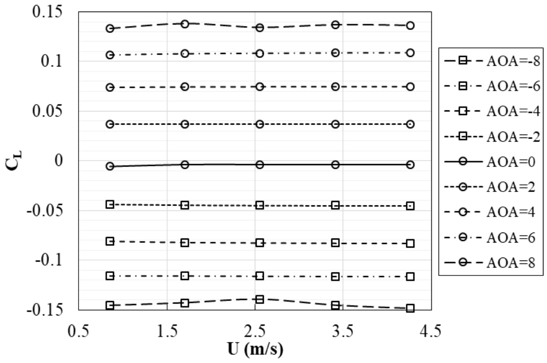


Figure 8. lift coefficient in terms of flow velocity for different angles of attack

The drag force coefficients in terms of angle of attack are shown in Figure 9. Based on this figure, at the same angle of attack, the glider drag coefficient enhances by increasing the Reynolds number. The drag force coefficients in terms of velocity are shown in Figure 10. Both figures (Figures 9 and 10) show that the drag coefficient increases with increasing the velocity and the angle of attack. Referring to the graphs, the highest drag coefficient is 0.0246 which corresponds to the angle of attack of -8 and the Reynolds number 3,738,184(equivalent to velocity of 4.26 m/s). On the other hand, the lowest drag coefficient is -0.0084 which corresponds to the angle of attack of 0 and Reynolds number of 747636.7 (equivalent to velocity of 0.852 m/s). The effect of velocity on the drag coefficient is more obvious, indicating that the drag coefficient is more dependent on the velocity compared to the lift coefficient. One of the major parameters that affects the drag force is geometry of the glider and its orientation that can be seen in the differences of drag coefficients at higher angle of attacks such as ±8 with others in Figure 10.

Due to specify the oriantation effects in various velocities, the difference of drag coefficient in terms of different attack angles is shown in Table 2.

The results obtained for the presented velocity and angle ranges showed that the drag and lift variations with the angle of attack are higher than the velocity, which indicates the importance of the angle of attack in the hydrodynamic studies. The wings have the most impact on the Lift force and the glider body has the least impact on it. Since the Reynolds number is inversely correlated with the fluid viscosity passing through the vicinity of the body, the viscous forces are less important than the compressive forces and are negligible. So here the drag

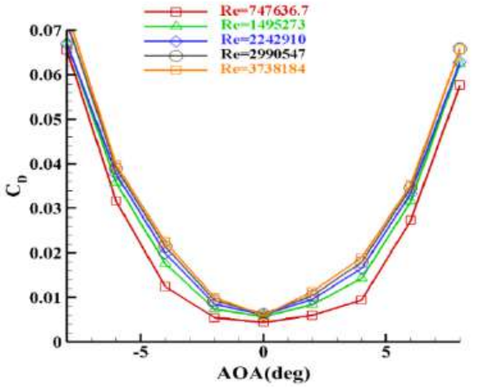


Figure 9. Drag coefficient in terms of angles of attack for different flow velocities

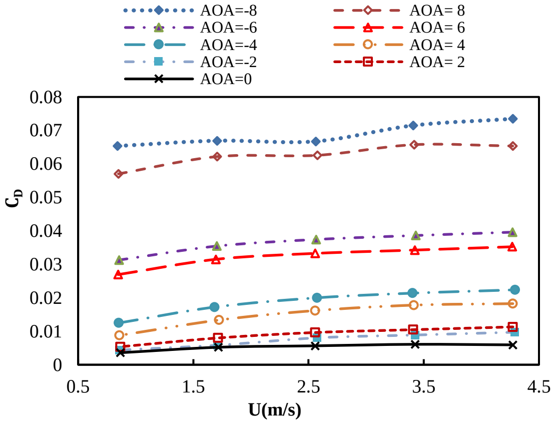


Figure 10. Drag coefficient in terms of flow velocity for different angles of attack

TABLE 2. The percentage difference of drag coefficient in terms of different attack angles

Velocity (m/s)	First angle	Second angle	Percentage difference of drag coefficient
1.7	0	4	156%
	4	8	365%
3.4	0	4	192%
	4	8	269%
4.3	0	4	207%
	4	8	258%

force is dominated by compressive forces. The drag force becomes more important as the angle of attack increases. As the surface area facing the direction of flow increases with increasing angle of attack, the compressive force increases. The torque coefficients in terms of angle of attack and velocity are shown in Figures 11 and 12. As it is shown in these figures, the torque coefficient is distributed approximately symmetrically with the angle of attack, indicating the relative uniformity of the drag and lift forces along with the glider.

3.4. Flow Analysis Around the Glider

According to the figures showing the effects of angles of attack on the drag and lift coefficients, as expected, the lift coefficient increases almost linearly with increasing angle of attack.

Due to the low-velocity range, the rate of change of the lift coefficient is low. As shown in Figure 8, the lift coefficient at the angle of attack of 8° is higher than the rest of the glider position due to the decrease in pressure adjacent to the upper surface of the glider wing (Figure 13). Figure 14 shows the flow lines around the glider wing at the angle of 8° and velocity of 4.26 m/s. As shown, the flow hits the airfoil at the angle of 8°, which reduces the pressure at the suction surface and increases the pressure at the pressure surface, thereby, the lifting force increases.

Figure 15 shows that at an angle of attack of 8°, the velocity on the front surface of the glider increases

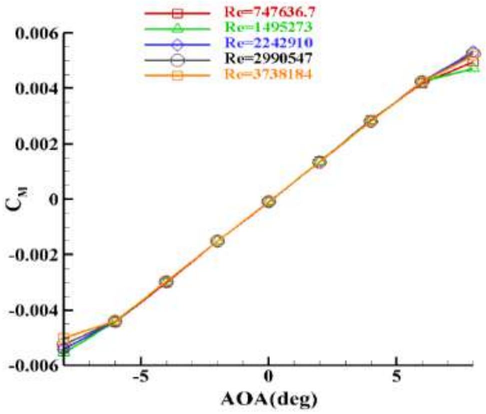


Figure 11. Torque coefficient in terms of angles of attack for different flow velocities

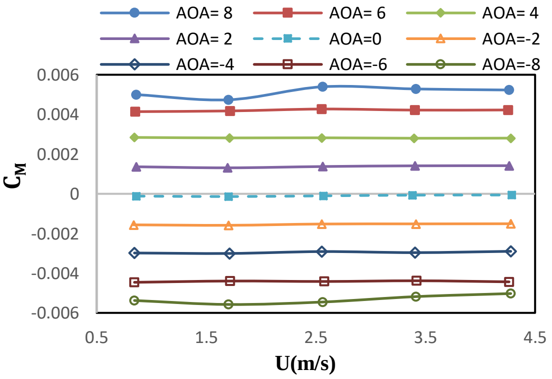


Figure 12. Torque coefficient in terms of velocity for different angles of attack

because of the pressure drop and decreases at the end of the glider. The cause of the deceleration at the end of the glider is illustrated in Figure 16.

Figure 16 illustrates the velocity vector for the angle of attack of 8° and the free-flow velocity of 4.26 m/s. The velocity distribution indicates that velocity is maximum around the underwater glider at the upper part of the nose, and behind the underwater glider, the velocity drops due to vortices.

The static pressure distribution, which represents the pressure applied to the body of the underwater glider, is shown at an angle of attack of 8° in Figure 17. In this figure, the free flow velocity is 4.26 m/s. The pressure is high at the tip of the underwater glider as well as its bottom, which is gradually reduced due to its hydrodynamic geometric shape. The pressure is also high on the front edge and lower part of the glider wings, which is an important factor that increases the lift coefficient at large attack angles. Because of the stagnation point, the maximum pressure is at the tip of the glider's nose. The static pressure is lower for the rest of the glider surface due to the uniform flow between the fluid and the cylindrical body.

Figure 18 illustrates the dynamic pressure distribution (fluid pressure applied to the underwater glider body) around the underwater glider. As shown, the front edge of the airfoil experiences the most dynamic pressure. Due to the different dynamic pressure caused by the angle of attack, the lift is created.

The effect of the negative angle of attack on the pressure distribution on the underwater glider is shown in Figure 19, which is obtained for the 8° angle of attack at a velocity of 4.26 m/s. As can be seen, due to the impact of the flow on the underwater glider, most of the pressure is applied to the upper part of the tip and the upper part of the lateral wings. This reduces the lift coefficient.

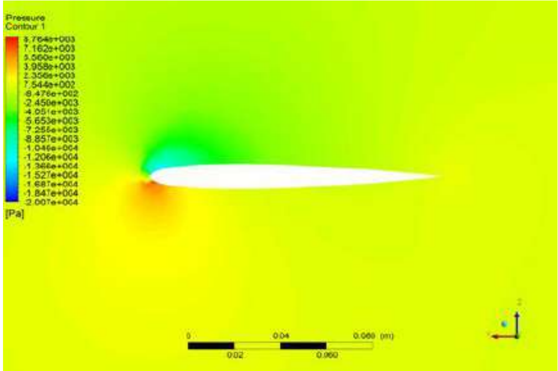


Figure 13. Pressure contour around the glider wing at an 8° angle of attack at the velocity of 4.26 m/s

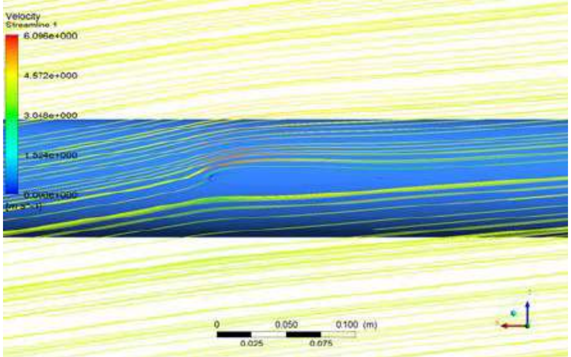


Figure 14. Flow lines around the glider wing at the 8° angle of attack at the velocity of 4.26 m/s

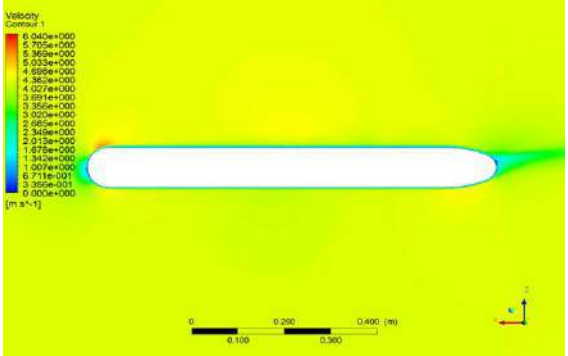


Figure 15. Velocity distribution around the glider body at an attack angle of 8° and a velocity of 4.26 m/s

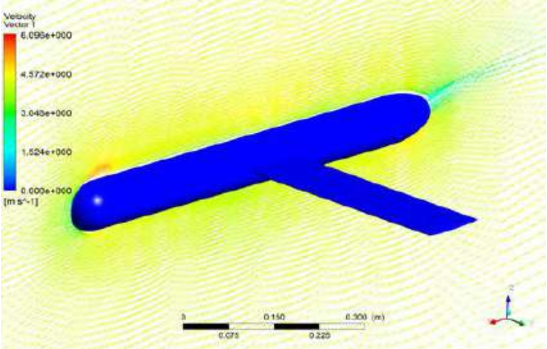


Figure 16. Flow velocity vector around the underwater glider

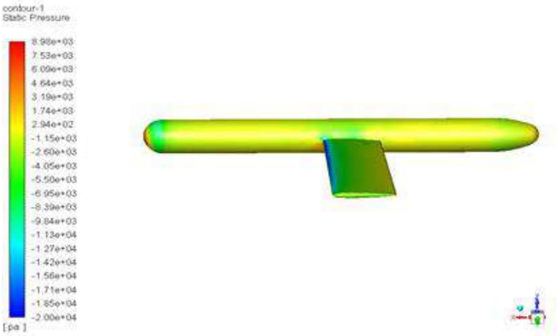


Figure 17. Static pressure contour on the glider body at an attack angle of 8° and a velocity of 4.26 m/s

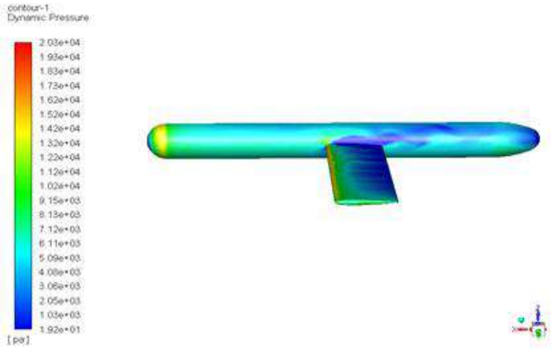


Figure 18. Dynamic pressure distribution on the glider body at an attack angle of 8° and a velocity of 4.26 m/s

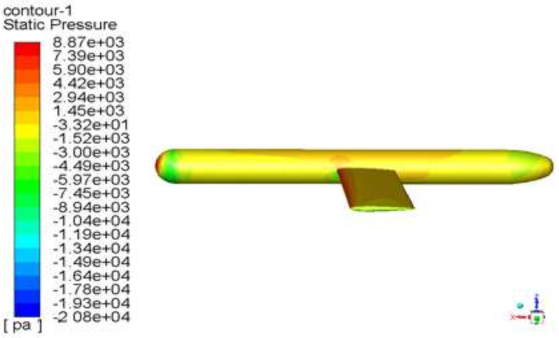


Figure 19. Static pressure contour on the glider body at an attack angle of -8° and a velocity of 4.26 m/s

4. CONCLUSION

In this paper, the numerical analysis of the performance of an automated underwater glider is presented in which numerical simulation has been used to evaluate the hydrodynamic coefficients of the underwater glider (drag, lift, and torque coefficients). In addition, the results also demonstrate the hydrodynamic response of the glider over the velocity and angle of attack. The selected solver method is finite volume. The SST κ-ε perturbation model is used to solve the Navier-Stokes equations and the continuity for fluid velocities up to 4.26 m/s. The accuracy of the method has been verified by comparing it with a similar study as well as with experimental data from testing the actual model. Velocity and pressure field distributions, as well as flow lines, are presented in the results section for underwater glider attack angles from -8 to 8 degrees and different operating velocities. The highest drag coefficient is 0.0246 which corresponds to the angle of attack of -8 and the Reynolds number 3,738,184.The underwater gliders investigated in this paper have suitable hydrodynamic coefficients compared to other gliders, which can be said to reduce the cost due to simpler geometry. This research can be a background for future studies on glider structural analysis and controller design for maneuverability.

5. REFERENCES

1. Alamian, R., Shafaghat, R., Amiri, H.A. and Shadloo, M.S., "Experimental assessment of a 100 w prototype horizontal axis tidal turbine by towing tank tests", ***Renewable Energy***, Vol. 155, (2020), 172-180. doi:10.1016/j.renene.2020.03.139

2. Alamian, R., Shafaghat, R., Bayani, R. and Amouei, A.H., "An experimental evaluation of the effects of sea depth, wave energy converter's draft and position of centre of gravity on the performance of a point absorber wave energy converter", ***Journal of Marine Engineering & Technology***, Vol. 16, No. 2, (2017), 70-83. doi:10.1080/20464177.20462017.21282718.

3. Alamian, R., Shafaghat, R., Farhadi, M. and Bayani, R., "Experimental evaluation of irwec1, a novel offshore wave energy converter", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 29, No. 9, (2016), 1292-1299. doi:10.5829/idosi.ije.2016.29.09c.15

4. Yazdi, H., Shafaghat, R. and Alamian, R., "Experimental assessment of a fixed on-shore oscillating water column device: Case study on oman sea", ***International Journal of Engineering, Transactions C: Aspects*** Vol. 33, No. 3, (2020), 494-504. doi:10.5829/IJE.2020.33.03C.14

5. Alamian, R., Shafaghat, R. and Safaei, M.R., "Multi-objective optimization of a pitch point absorber wave energy converter", ***Water***, Vol. 11, No. 5, (2019), 969. doi:10.3390/w11050969

6. Amiri, H.A., Shafaghat, R., Alamian, R., Taheri, S.M. and Shadloo, M.S., "Study of horizontal axis tidal turbine performance and investigation on the optimum fixed pitch angle using cfd", ***International Journal of Numerical Methods for Heat & Fluid Flow***, Vol. 30, No. 1, (2019), 206-227. doi:10.1108/hff-05-2019-0447

7. Wagawa, T., Kawaguchi, Y., Igeta, Y., Honda, N., Okunishi, T. and Yabe, I., "Observations of oceanic fronts and water-mass properties in the central japan sea: Repeated surveys from an underwater glider", ***Journal of Marine Systems***, Vol. 201, No., (2020), 103242. doi:10.1016/j.jmarsys.2019.103242

8. Leonard, N.E., Paley, D.A., Lekien, F., Sepulchre, R., Fratantoni, D.M. and Davis, R.E., "Collective motion, sensor networks, and ocean sampling", ***Proceedings of the IEEE***, Vol. 95, No. 1, (2007), 48-74. doi:10.1109/jproc.2006.887295

9. Woithe, H.C., Tilkidjieva, D. and Kremer, U., Towards a resource-aware programming architecture for smart autonomous underwater vehicles, in Technical Report DCS-TR-637. 2008, Rutgers University: Department of Computer Science. doi:10.1109/iros.2009.5354098

10. Stommel, H., "The slocum mission", ***Oceanography***, Vol. 2, No. 1, (1989), 22-25. doi:10.5670/oceanog.1989.26

11. Graver, J.G., "Underwater gliders: Dynamics, control and design", Princeton university Princeton, NJ, (2005).

12. Nosrati, K., Tahershamsi, A. and Taheri, S.H.S., "Numerical analysis of energy loss coefficient in pipe contraction using ansys cfx software", ***Civil Engineering Journal***, Vol. 3, No. 4, (2017), 288-300. doi:10.28991/cej-2017-00000091

13. Yamini, O.A., Mousavi, S.H., Kavianpour, M.R. and Movahedi, A., "Numerical modeling of sediment scouring phenomenon around the offshore wind turbine pile in marine environment", ***Environmental earth sciences***, Vol. 77, No. 23, (2018), 776. doi:10.1007/s12665-018-7967-4

14. Du, X.-x., Wang, H., Hao, C.-z. and Li, X.-l., "Analysis of hydrodynamic characteristics of unmanned underwater vehicle moving close to the sea bottom", ***Defence Technology***, Vol. 10, No. 1, (2014), 76-81. doi:10.1016/j.dt.2014.01.007

15. Du, X., Zhang, Z. and Cui, H., "Thrust performance of propeller during underwater recovery process of auv", in OCEANS 2017-

Aberdeen, IEEE., (2017), 1-5. doi:10.1109/oceanse.2017.8084603

16. Jung, H.-J., Kim, M.J., Lee, P.-Y. and Jung, H.-S., "A study on numerical analysis of controllable pitch propeller (CPP) using tunnel inspection auv", in 2012 Oceans-Yeosu, IEEE., (2012), 1-4. doi:10.1109/oceans-yeosu.2012.6263551

17. Gao, L., He, R., Li, Y. and Zhang, Z., "Analysis of autonomous underwater gliders motion for ocean research", in ASME 2014 33rd International Conference on Ocean, Offshore and Arctic Engineering, American Society of Mechanical Engineers Digital Collection., (2014). doi:10.1115/omae2014-24534

18. De Barros, E. and Dantas, J.L.D., "Effect of a propeller duct on auv maneuverability", ***Ocean Engineering***, Vol. 42, (2012), 61-70. doi:10.1016/j.oceaneng.2012.01.014

19. Dantas, J.L.D. and De Barros, E., "Numerical analysis of control surface effects on auv manoeuvrability", ***Applied Ocean Research***, Vol. 42, (2013), 168-181. doi:10.1016/j.apor.2013.06.002

20. Ray, S., Chatterjee, D. and Nandy, S., "Unsteady cfd simulation of 3d auv hull at different angles of attack", ***Journal of Naval Architecture and Marine Engineering***, Vol. 13, No. 2, (2016), 111-123. doi:10.3329/jname.v13i2.25849

21. Joung, T.-H., Sammut, K., He, F. and Lee, S.-K., "Shape optimization of an autonomous underwater vehicle with a ducted propeller using computational fluid dynamics analysis", ***International Journal of Naval Architecture and Ocean Engineering***, Vol. 4, No. 1, (2012), 45-57. doi:10.3744/jnaoe.2012.4.1.044

22. Yue, C., Guo, S. and Li, M., "Ansys fluent-based modeling and hydrodynamic analysis for a spherical underwater robot", in 2013 IEEE International Conference on Mechatronics and Automation, IEEE., (2013), 1577-1581. doi:10.1109/icma.2013.6618149

23. Zheng, H., Wang, X. and Xu, Z., "Study on hydrodynamic performance and cfd simulation of auv", in 2017 IEEE International Conference on Information and Automation (ICIA), IEEE., (2017), 24-29. doi:10.1109/icinfa.2017.8078877

24. Singh, Y., Bhattacharyya, S. and Idichandy, V., "Cfd approach to modelling, hydrodynamic analysis and motion characteristics of a laboratory underwater glider with experimental results", ***Journal of Ocean Engineering and Science***, Vol. 2, No. 2, (2017), 90-119. doi:10.1016/j.joes.2017.03.003

25. Noman, A.A., Tusar, M.H., Uddin, K.Z., Uddin, F., Paul, S. and Rahman, M., "Performance analysis of an unmanned under water vehicle using cfd technique", in AIP Conference Proceedings, AIP Publishing LLC. Vol. 2121, (2019), 040015. doi:10.1063/1.5115886

26. Lin, Y., Yang, Q. and Guan, G., "Automatic design optimization of swath applying cfd and rsm model", ***Ocean Engineering***, Vol. 172, No., (2019), 146-154. doi:10.1016/j.oceaneng.2018.11.044

27. Nedelcu, A.-T., Faităr, C., Stan, L.-C. and Buzbuchi, N., "Underwater vehicle cfd analyses and reusable energy inspired by biomimetic approach", (2018). doi:10.20944/preprints201808.0175.v1

28. Javaid, M.Y., Ovinis, M., Hashim, F.B., Maimun, A., Ahmed, Y.M. and Ullah, B., "Effect of wing form on the hydrodynamic characteristics and dynamic stability of an underwater glider", ***International Journal of Naval Architecture and Ocean Engineering***, Vol. 9, No. 4, (2017), 382-389. doi:10.1016/j.ijnaoe.2016.09.010

29. Liu, Y., Ma, J., Ma, N. and Huang, Z., "Experimental and numerical study on hydrodynamic performance of an underwater glider", ***Mathematical Problems in Engineering***, Vol. 2018, No., (2018). doi:10.1155/2018/8474389

30. Javaid, M.Y., Ovinis, M., Javaid, M. and Ullah, B., "Experimental study on hydrodynamic characteristics of underwater glider", ***Indian Journal of Geo-Marine Sciences (IJMS)***, Vol. 48, No. 7, (2019), 1091-1097.

31. Menter, F.R., "Two-equation eddy-viscosity turbulence models for engineering applications", ***AIAA Journal***, Vol. 32, No. 8, (1994), 1598-1605. doi:10.2514/3.12149

32. Sengupta, A.R., Gupta, R. and Biswas, A., "Computational fluid dynamics analysis of stove systems for cooking and drying of muga silk", ***Emerging Science Journal***, Vol. 3, No. 5, (2019), 285-292. doi:10.28991/esj-2019-01191

33. Boroomand, M.R. and Mohammadi, A., "Investigation of k-ε turbulent models and their effects on offset jet flow simulation", ***Civil Engineering Journal***, Vol. 5, No. 1, (2019), 127. doi:10.28991/cej-2019-03091231

34. Isa, K., Arshad, M. and Ishak, S., "A hybrid-driven underwater glider model, hydrodynamics estimation, and an analysis of the motion control", ***Ocean Engineering***, Vol. 81, No., (2014), 111-129. doi:10.1016/j.oceaneng.2014.02.002

35. Abbas, F.M., "Investigating role of vegetation in protection of houses during floods", ***Civil Engineering Journal***, Vol. 5, No. 12, (2019). doi:10.28991/cej-2019-03091436

36. Ahmad, M., Ghani, U., Anjum, N., Pasha, G.A., Ullah, M.K. and Ahmed, A., "Investigating the flow hydrodynamics in a compound channel with layered vegetated floodplains", ***Civil Engineering Journal***, Vol. 6, No. 5, (2020), 860-876. doi:10.28991/cej-2020-03091513

37. Zhang, S., Yu, J., Zhang, A. and Zhang, F., "Spiraling motion of underwater gliders: Modeling, analysis, and experimental results", ***Ocean Engineering***, Vol. 60, (2013), 1-13. doi:10.1016/j.oceaneng.2012.12.023

Persian Abstract

چکیده

در این مقاله از شبیه‌سازی عددی برای تحلیل ضرایب هیدرودینامیکی گلایدر زیرآبی (ضرایب پسا، لیفت و گشتاور) استفاده شده است. به دست آوردن این ضرایب نقش مهمی در طراحی کنترلر برای هدایت آن دارد. برای انجام این کار، گلایدر زیرآبی در معرض شرایط مختلف عملیاتی اعم از زاویه‌ی حمله و سرعت حرکت آن قرار داده شده است. طول گلایدر زیرآبی ۹۰ سانتی‌متر با دو باله با پروفیل NACA0012 است. سپس، رژیم جریان سیال به کمک دینامیک سیالات محاسباتی و روش حجم محدود گسسته‌سازی و حل شده است. از آن‌جا که محدوده‌ی عدد رینولدز کاری در حالت جریان آشفته است (تا ۳/۷×$۱۰^{٦}$)، از فرمولاسیون $\kappa - \omega$ SST برای حل معادلات ناویر استوکس و پیوستگی بهره گرفته شد. زوایای حمله بررسی شده بین ۸– و ۸ درجه است. هدف اصلی این تحقیق مطالعه‌ی تاثیر هر یک از پارامترهای دینامیکی حرکت گلایدر از جمله سرعت و زاویه‌ی حمله بر ضرایب هیدرودینامیکی است. براساس نتایج، ضریب پسا و بالابر با افزایش زاویه‌ی حمله افزایش می‌یابد. همچنین، با افزایش سرعت، ضریب پسا افزایش می‌یابد، اما با افزایش سرعت گلایدر، ضریب بالابر به جز در بالاترین زاویه‌ی حمله که کاهش می یابد، به طور چشم‌گیری تغییر نمی‌کند. بزرگترین ضریب پسا ۰.۰٢٤٦ است که مربوط به زاویه‌ی حمله‌ی ۸– درجه و مطابق با عدد رینولدز ۳۷۳۸۱۸٤ می‌باشد. گلایدر مطالعه شده در این مقاله علاوه بر هندسه‌ی ساده، در مقابل جریان مقاومت نسبتاً کمی از خود نشان داده است.

International Journal of Engineering

Journal Homepage: www.ije.ir

Metallurgical and Mechanical Properties of Laser Cladded AlFeCuCrCoNi-WC_{10} High Entropy Alloy Coating

A. Vyas*[a], J. Menghani[a], H. Natu[b]

[a] *Mechanical Engineering Department, SVNIT, Surat, India*
[b] *Magod Fusion Technologies Private Limited, Pune, India*

PAPER INFO

Paper history:
Received 17 May 2020
Received in revised form 06 June 2020
Accepted 12 June 2020

Keywords:
Laser-assisted Cladding
Microhardness
High Entropy Alloy
Microstructure

ABSTRACT

In spite of excellent corrosion resistance, good ductility and low cost of AISI 316 austenitic stainless steel, the low hardness and poor mechanical characteristic of material restricts its applicability in several industrial services. To improve the mechanical properties AlFeCuCrCoNi-WC_{10} high-entropy alloy coatings were deposited via laser cladding on austenitic stainless steel AISI 316 substrate. The influence of WC on phase constituents, microstructure, microhardness and elemental distribution were investigated using X-ray diffractometry, optical microscopy microhardness tester and FESEM-EDS (Energy Dispersive Spectroscopy), respectively. The XRD peaks revealed that as clad AlFeCuCrCoNi-WC_{10} multiple principal element alloy coating composed of BCC, FCC and W-rich phase. The cladding zone microstructure is mainly consisting of fine-grained non-directional and equiaxed crystals away from the base material and columnar grains near the base material. The energy dispersive spectroscopy indicated segregation of W and Cr in the interdendritic region. However, other elements of the multiple principal element alloy are observed to be uniformly distributed throughout the cladding. The microhardness of the AlFeCuCrCoNi-WC_{10} (670 $Hv_{0.5}$) high entropy alloy coating was 4.5 times greater than that of substrate AISI-316.

***doi**: 10.5829/ije.2020.33.07a.27*

1. INTRODUCTION

In order to overcome the restrictions such as elemental segregation and formation of various brittle phases in the conventional alloying system, Yin et al. [1] discovered the innovative idea of multi principal element alloy (high entropy alloy). The high entropy alloy (baseless alloy, multi principal element) can be defined as solid solution alloys which have a minimum of five principal elements, but not greater than thirteen elements, each of the primary elements having a contribution in between 5 to 35 at% [1-2]. The baseless alloys have some excellent characteristic such as superior wear resistance, distinctive magnetic as well as electrical properties, better resistance to erosion, excellent stability at various temperature range, high hardness and mechanical strength due to its four core effect namely cocktail effect, severe lattice distortion effect, sluggish diffusion effect as well as high entropy effect [3]. However, bulk products of high entropy alloys are usually made of casting as well as powder metallurgy route. To improve efficiency and working life of various components along with reduced cost of high entropy alloy (Ni, Cr and Co expensive elements), it can be used as a coating material instead of replacement of existing bulk material [4]. The HEA coating is deposited through different processes such as plasma transferred arc, magneto sputtering, electrochemical deposition and laser coating. Laser is a versatile tool which can be used not only for cladding but also for welding, cutting as well as forming [5-8]. The laser-assisted coating method has many advantages compared to other methods, including rapid solidification rate (in the range of 10^4 to 10^6 $^oC/s$), good metallurgical bonding, optimum dilution, narrow heat-affected zone, better control on process parameters, high repeatability and process stability. In addition, based on

*Corresponding Author Email: *akku.vyas2011@gmail.com* (A. Vyas)

Please cite this article as: A. Vyas, J. Menghani, H. Natu, Metallurgical and Mechanical Properties of Laser Cladded AlFeCuCrCoNi-WC_{10} High Entropy Alloy Coating, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1397-1402

kinetic theory, the reduction in nucleation growth of brittle intermetallics can be done through a quick solidification rate of the laser cladding process [9]. AISI 316 steel is a very important material for many tribological as well as marine applications due to its superior corrosive resistivity and ductility. However, its inferior tribological and wear characteristics impose a limitation on this material for a broader range of applications [10]. From the theoretical point of view and practical consideration, it is necessary to investigate the mechanical and metallurgical characteristics of HEA cladding on austenitic steel in brief. Recently, ceramic particles such as WC, TiC and SiC reinforced coatings prepared by laser cladding technology are mostly applied on Nickel, Cobalt or Ferrous alloys. However, the investigation regarding HEA coating strengthened by reinforcing phase particles are limited. WC has very high hardness (2600 HV) and melting point (2600 °C) making it ideal as a reinforcement material in HEA coating to enhance mechanical characteristics [11,16]. In present work, a novel high entropy alloy coating with an equiatomic composition of AlFeCuCrCoNi- and 10 wt.% WC (AlFeCuCrCoNi-WC_{10}) was deposited on AISI-316 substrate through laser cladding. Furthermore, the phase constituents, microstructure evaluation, elemental distribution and microhardness of high entropy alloy coating were investigated briefly.

2. EXPERIMENTAL PROCEDURE

Circular samples of AISI 316 stainless steel with a size of ϕ90mm x 10mm were chosen as the base material. The elemental composition of AISI 316 steel is as indicated in Table 1. As a cladding material, an equiatomic Fe, Cr, Co, Ni, Cu and Al metal powders with a reinforcement of 10 wt.% WC were used. The contribution of individual elements for making high entropy alloy is as listed in Table 2. The motive behind selecting a particular individual element for the preparation of high entropy alloy is mixing enthalpy of every principal element with each other, as listed in Table 3 [16].

All powders having 99.9% purity procured from Chrome special materials pvt ltd., Mumbai with an average mesh size of 80-100 were used in the present investigation. The substrate surface was polished through emery paper (320-1200 grit size) and rinsed with the acetone before laser cladding. The selected high entropy alloy powders (AlFeCuCrCoNi-WC_{10}) were mixed in an equiatomic composition using 3D Multi-Motion Mixer (Alphie 0.3 Hp) for 30 minutes (15 min forward and 15 min reverse cycle). The laser cladding experiment was performed through 4 kW CO_2 laser system situated at the Magod Fusion Technology, Pune, as shown in Figure 1. The operating parameters for laser cladding process were, spot diameter (d) = 2.4 mm, standoff distance (a) = 8 mm, laser power (p) = 1.1 kW, laser scanning velocity (v) = 500 mm/min, argon flow rate (r) = 6 lit/min and the powder feed rate (f) = 4 g/min. An overlapping ratio of 60% was chosen between successive tracks to generate a homogeneous cladding on the substrate. The laser cladded metallurgical specimens were cut cross-sectional, resin mounted, polished, and etched with an aqua regia solution. The difference in clad morphology and

TABLE 1. Chemical content of AISI 316 steel (wt.%).

Elements	**Ni**	**Mo**	**P**	**S**	**N_2**	**C**	**Mg**	**Si**	**Cr**	**Fe**
Composition	10.0 - 14.0	2.00 - 3.00	0.045	0.03	0.1	0.08	2.00	0.1	16.0 -18.0.	65-70

TABLE 2. Chemical composition of AlFeCuCrCoNi-WC_{10} high entropy alloy powder (wt.%).

Elements	**Al**	**Fe**	**Cu**	**Co**	**Cr**	**Ni**	**WC**
Composition	7.96± 0.40	15.90± 0.10	17.73± 0.05	16.52± 0.20	14.68± 0.05	17.14± 0.25	10.00 ± 0.05

TABLE 3. Binary mixing enthalpies, ΔH^{mix}ij (kJ/mol) of high entropy alloy [16]

Al-Fe	-11	Fe-Ni	2	Cu-Ni	4	Cr-Ni	-7	Co-Ni	0	Ni-W	-9
Al-Ni	-22	Fe-Cu	13	Cu-Co	6	Cr-Co	-4	Co-W	-5	Ni-C	-39
Al-Cu	-1	Fe-Co	-1	Cu-Cr	12	Cr-W	5	Co-C	-42		
Al-Co	-19	Fe-Cr	-1	Cu-W	23	Cr-C	-61				
Al-Cr	-10	Fe-W	4	Cu-C	-42						
Al-W	-2	Fe-C	-50								
Al-C	-30										

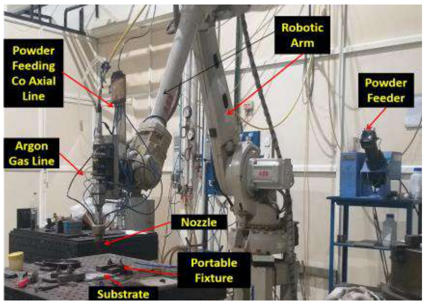


Figure 1. A 4 kW Co_2 laser cladding experimental setup equiped with robotic arm

microstructure along its length and height was revealed by optical microscopy (Leica S8APO). The phase constitution of the as cladded specimen was identified through X-ray diffractometer (XRD) using a Cu Kα radiation (PANalytical X'Pert Powder). The variation in microhardness from cladded zone to substrate metal was determined using a Micro Vickers Hardness Tester (Future Tech Corporation, Japan –FM 700) applying a load of 0.5 Kg with a dwell time of 15 Sec. The field emission scanning electron microscope (FESEM-JSM7100F) along with energy dispersive spectrometer (EDS) was used to analyze the elemental distribution in the coating as well as in the interface zone.

3. RESULTS AND DISCUSSION

3. 1. Phase Constituents

Figure 2 indicates the X-ray diffraction profile of the AlFeCuCrCoNi-WC_{10} high entropy alloy coating. The solid solution phases consisting of a combination of two phases FCC + BCC, an additional set of diffraction peaks corresponding to W rich carbide phases can be observed [12]. The results revealed that additional phases in the coating observed to be very limited and the laser rapid melting and solidification process effectively constrained the precipitation of undesired intermetallic compounds in the cladding. No oxide formation was revealed in XRD Spectra, which signifies that during laser cladding adequate protection against oxidation was maintained through continuous supply of Argon gas [13]. As stated by Gibbs phase rule, the number of phases for non-equilibrium solidification leads to be $p > n+1$ (p= number of phases, n= number of elements), while the phases formed in AlFeCuCrCoNi-WC_{10} high entropy alloy cladding is much less than 8 [14]. The reason behind this phenomenon is the influence of high mixing entropy generated by multi principal element. Considering intensity of XRD peaks, it is observed that BCC solid solution phase is much higher than that of FCC solid solution phase, and it can be concluded that the BCC solid solution phase is the primary phase [15].

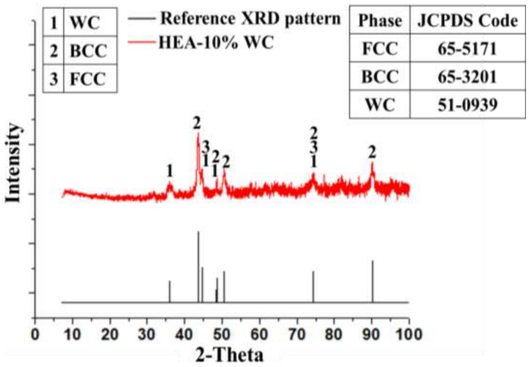


Figure 2. XRD spectra of AlFeCuCrCoNi-WC_{10} HEA

3. 2. Microstructure

Figure 3 shows the optical micrograph of AlFeCuCrCoNi-WC_{10} HEA coatings synthesized through the laser-assisted cladding. The micrograph indicates that cladding is dense enough and free from cracks as well as pores, with a coating thickness of 0.9 mm. As can be seen in Figure 3(a), there are three categorical regions in the cladding, that is clad zone (CZ), Bonding Zone (BZ) and Heat affected zone (HAZ) the span of which depends on variation in the microstructure. The cladding zone microstructure is mainly consisting of fine-grained non-directional and equiaxed crystals away from the base material and columnar grains near the base material. The columnar grains (Figure 3(c)) are transformed to equiaxed (Figure 3(b)) with a reduction in a temperature gradient in the center of the cladding zone. The layer of planar crystals with a thickness of approximately 20-25 µm is observed at the interface region of the high entropy alloy cladding, as indicated by "PC" in Figure 3 (a). The planar crystal is hard to corrode for metallurgical investigation. In addition, the curved line at the interface zone instead of a straight line also indicates an excellent metallurgical bonding between the cladding and the base material. When the high entropy alloy cools down during laser cladding process, agreeing to the constitutional undercooling criterion, the temperature difference at liquid solid interface and the cooling rate are the major factors to identify the microstructure behavior of HEA. In the laser cladding process, due to quick melting-solidification phenomena a thin layer of substrate melts with the cladding layer, and this diffused layer cools down through heat dissipation of the base material. Therefore, the temperature difference (Δt) is very high, while the solidification rate (r) is less, which tends to give large Δt/r. Therefore, the nucleation rate is much quicker than the growth rate of the crystal and it will lead to grow as the planar crystal at the interface [16-18]. The growth direction of the columnar grain is perpendicular to the interface zone due to the rapid directional solidification, typical of the laser cladding process [19].

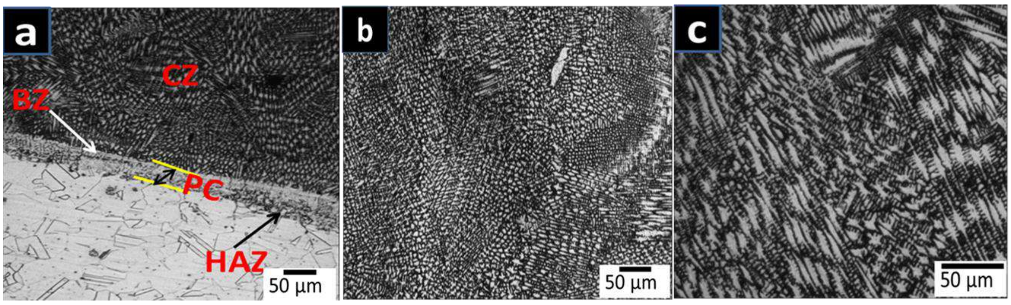


Figure 3. Optical micrograph of the HEA coatings (a) Interface zone (b) Cladding zone and (c) Grain orientation near the interface zone

3. 3. Compositional Analysis

The EDS (Energy Dispersive Spectroscopy) results (Figure 4(b)) for the elemental composition of the cladding layer revealed that all the individual principal element were uniformly distributed in approximately designed weight proportion across the coating. The quantitative composition of elements in the different areas as indicated in region 1 (Bright) and region 2 (Dark) in Figure 4(c) is shown in Table 5. The elemental distribution in the region 2 (Dendritic) is similar to the quantitative composition of the cladding, while region 1 (Interdendritic) is enriched with tungsten particles. [18]. Additionally, high proportion of Cr is observed in interdendritic region indicating partial dissolution of tungsten carbide particles and replacement of W with Cr. This could be majorly due to the capillary phenomena of melted material and concentrated laser energy, which may tend to increase in the temperature of the molten pool beyond the melting point of tungsten carbide particles and lead to the partial dissolution of tungsten carbide particles [16].

Figure 5 shows the EDS elemental mapping for the individual elements in the cladding layer. As in area EDS analysis mapping also revealed the segregation of tungsten rich particles in the inter-dendritic region, while the other elements of the HEA are observed to be uniformly distributed throughout the cladding. The rapid melting and solidification process of laser cladding along with the sluggish diffusion effect of HEA leads to the uniform distribution of alloying elements [20]. In order to study the degree of iron dilution and its effect on the elemental distribution of the cladding layer, line EDS analysis was conducted near the interface zone, and the results are indicated in Figure 6. The concentration of the iron is higher than design composition at the interface as compared to surface due to dilution from iron based substrate.

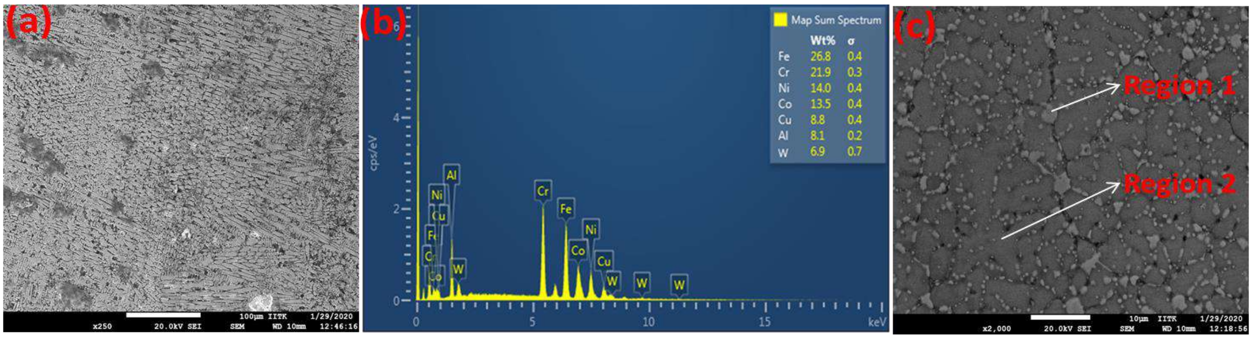


Figure 4. EDS elemental analysis of high entropy alloy coatings (a) SEM image of cladding zone (b) Elemental distribution of the cladding zone (c) High magnification image of cladding zone indicating various regions

TABLE 5.EDS results of AlFeCuCrCuNi-WC$_{10}$ coating

Region	Fe	Cu	Cr	Co	Ni	Al	W
Region 1	14.6	8.3	26.1	7.4	8.8	3.6	31.2
Region 2	29.1	9.5	19.2	15.2	15.3	5.1	6.6

3. 4. Microhardness

Figure 7 shows the measured values of microhardness for the AlFeCuCrCoNi-WC$_{10}$ high entropy alloy coatings along the cross-section from the cladding zone to the substrate. It was observed that the maximum

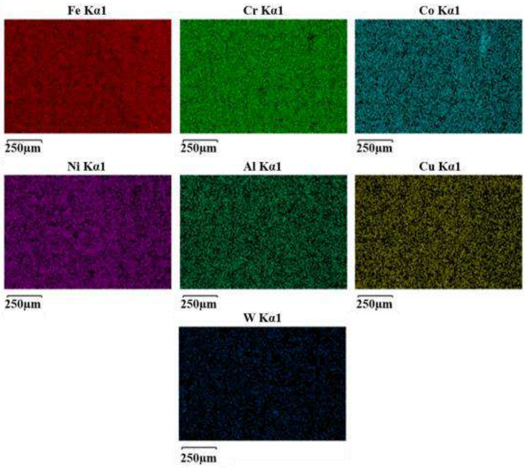


Figure 5. EDS maps of the HEA coating for elements Fe, Cr, Co, Ni, Al, Cu and W

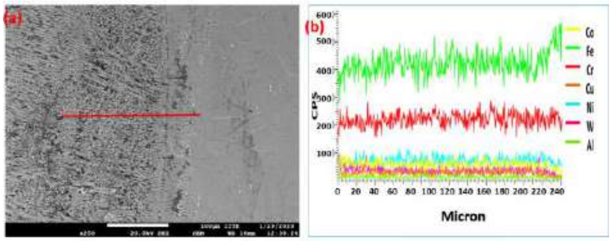


Figure 6. The line scan analysis of the AlFeCuCrCuNi-WC_{10} HEA coating (a) SEM micrograph of the interface zone of (b) Elemental distribution through line scan analysis

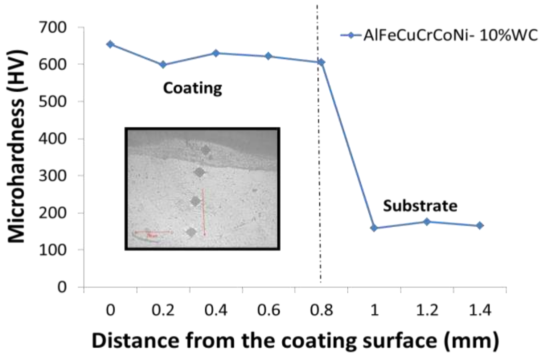


Figure 7. Microhardness distribution of the AlFeCuCrCoNi-WC_{10} coating

microhardness of AlFeCuCrCoNi-WC_{10} high entropy alloy coating reaches to 670 $HV_{0.5}$, which is approximately 4.5 times greater than the base material AISI-316 (155 $Hv_{0.5}$). This increasing hardness is attributed to the following reasons, (1) The ex-situ WC particles in the alloy coatings act as strengthening phase distributed in the solid solution phase due to the detachment of free carbon atoms from molten tungsten carbide particles into high entropy alloy matrix, increase lattice distortion and enhance solution strengthening. The substitutional and interstitial solution strengthening is the major phenomenon behind the solution strengthening of high entropy alloy. (2) The quick melting solidification behavior of the laser-assisted cladding process leads to grain size reduction, the development of nano-precipitates and the enhancement of solubility constrained in the coating. (3) The solution strengthening due to lattice distortion as a result of the difference in the atomic radius of the various metallic elements leads to large lattice strain distortion and solid solution effect in the AlCrFeNiCuCo-WC_{10} coatings. In addition, a high-density dislocation can be seen due to this variation in the atomic size. In AlCrFeNiCuCo-WC_{10}, Al contributes more to the lattice distortion because of its larger atomic radius compared to the other five elements. (4) Rapid solidification restricting grain growth leads to increasing the amount of grain boundary and tends to generate grain boundary strengthening [19].

4. CONCLUSION

The XRD peaks revealed that as clad AlFeCuCrCoNi-WC_{10} high entropy alloy coating composed of BCC, FCC and W-rich phase. The morphology of laser cladded AlFeCuCrCoNi-WC_{10} high entropy alloy includes cladding, bonding, planer crystal and heat-affected zones. The cladding zone microstructure is mainly consisting of fine-grained non-directional and equiaxed crystals away from the base material and columnar grains near the base material. The energy dispersive spectroscopy results revealed the segregation of W and Cr particles in the inter-dendritic region. At the same time, the other elements of the HEA are observed to be uniformly distributed throughout the cladding. The contribution of the iron is higher than that of the other elements near the substrate as compared to the cladding zone due to the rich iron content in the AISI-316substrate. Maximum microhardness of AlFeCuCrCoNi-WC_{10} high entropy alloy coating reaches 670 $HV_{0.5}$, which is approximately 4.5 times greater than that of the base material AISI-316 (155 $Hv_{0.5}$).

5. ACKNOWLEDGEMENT

The author appreciates the funding provided by the department of science and technology, India (DST-EMR/2016/000451) and ACMS department of IIT-Kanpur for metallurgical characterization.

6. REFERENCES

1. Yin, Shuo, Wenya Li, Bo Song, Xingchen Yan, Min Kuang, Yaxin Xu, Kui Wen, and Rocco Lupoi. "Deposition of FeCoNiCrMn high entropy alloy (HEA) coating via cold spraying." ***Journal of Materials Science & Technology***, Vol. 35, No. 6, (2019), 1003-1007. doi: 10.1016/j.jmst.2018.12.015
2. Wall, Michael T., Mangesh V. Pantawane, Sameehan Joshi, Faith Gantz, Nathan A. Ley, Rob Mayer, Andy Spires, Marcus L. Young, and Narendra Dahotre. "Laser-coated CoFeNiCrAlTi high entropy alloy onto a H13 steel die head." ***Surface and Coatings Technology***, Vol. 387 (2020), 125473. doi: 10.1016/j.surfcoat.2020.125473
3. Zhang, H. X., J. J. Dai, C. X. Sun, and S. Y. Li. "Microstructure and wear resistance of TiAlNiSiV high-entropy laser cladding coating on Ti-6Al-4V." ***Journal of Materials Processing Technology***, (2020), 116671. doi: 10.1016/j.jmatprotec.2020.116671
4. Guo, Yaxiong, Huilin Wang, and Qibin Liu. "Microstructure evolution and strengthening mechanism of laser-cladding MoFexCrTiWAlNby refractory high-entropy alloy coatings." ***Journal of Alloys and Compounds***, (2020), 155147. doi: 10.1016/j.jallcom.2020.155147
5. Safari, Mehdi, and Mahmoud Farzin. "Experimental investigation of laser forming of a saddle shape with spiral irradiating scheme." ***Optics & Laser Technology***, Vol. 66, (2015), 146-150. doi: 10.1016/j.optlastec.2014.09.003
6. Safari, Mehdi, Hosein Mostaan, and Mahmoud Farzin. "Laser bending of tailor machined blanks: Effect of start point of scan path and irradiation direction relation to step of the blank." ***Alexandria Engineering Journal***, Vol. 55, No. 2, (2016), 1587-1594. doi: 10.1016/j.aej.2016.01.010
7. Safari, Mehdi, and Hosein Mostaan. "Experimental and numerical investigation of laser forming of cylindrical surfaces with arbitrary radius of curvature." ***Alexandria Engineering Journal***, Vol. 55, No. 3, (2016), 1941-1949. doi: 10.1016/j.aej.2016.07.033
8. Safari, M., M. Farzin, and H. Mostaan. "A novel method for laser forming of two-step bending of a dome shaped part." ***Iranian Journal of Materials Forming*** Vol. 4, No. 2, (2017), 1-14. doi: 10.22099/IJMF.2017.4288
9. Zhang, Hui, Ye Pan, Yizhu He, and Huisheng Jiao. "Microstructure and properties of 6FeNiCoSiCrAlTi high-entropy alloy coating prepared by laser cladding." ***Applied Surface Science***, Vol. 257, No. 6, (2011), 2259-2263. doi: 10.1016/j.apsusc.2010.09.084
10. Khatak, H_S, and Baldev Raj, eds. Corrosion of austenitic stainless steels: mechanism, mitigation and monitoring. ***Woodhead Publishing***, 2002.
11. Jabbar Hassan, A., T. Boukharouba, D. Miroud, and S. Ramtani. "Metallurgical and Mechanical Behavior of AISI 316-AISI 304 during Friction Welding Process." ***International Journal of Engineering, Transactions B: Applications***, Vol. 32, No. 2 (2019), 306-312. doi: 10.5829/ije.2019.32.02b.16
12. Desale, Girish R., C. P. Paul, B. K. Gandhi, and S. C. Jain. "Erosion wear behavior of laser clad surfaces of low carbon austenitic steel." ***Wear***, Vol. 266, No. 9-10 (2009), 975-987. doi: 10.1016/j.wear.2008.12.043
13. Gu, Zhen, Shengqi Xi, and Chongfeng Sun. "Microstructure and properties of laser cladding and CoCr$_{2.5}$FeNi$_2$Ti$_x$ high-entropy alloy composite coatings." ***Journal of Alloys and Compounds***, Vol. 819, (2020), 152986. doi: 10.1016/j.jallcom.2019.152986
14. Fang, Shoushi, Xueshan Xiao, Lei Xia, Weihuo Li, and Yuanda Dong. "Relationship between the widths of supercooled liquid regions and bond parameters of Mg-based bulk metallic glasses." ***Journal of Non-Crystalline Solids***, Vol.321, No. 1-2 (2003), 120-125. doi:10.1016/S0022-3093(03)00155-8
15. Singh, Sheela, Nelia Wanderka, B. S. Murty, Uwe Glatzel, and John Banhart. "Decomposition in multi-component AlCoCrCuFeNi high-entropy alloy." ***Acta Materialia***, Vol.59, No. 1 (2011), 182-190. doi: 10.1016/j.actamat.2010.09.023
16. Peng, Y. B., W. Zhang, T. C. Li, M. Y. Zhang, L. Wang, Y. Song, S. H. Hu, and Y. Hu. "Microstructures and mechanical properties of FeCoCrNi high entropy alloy/WC reinforcing particles composite coatings prepared by laser cladding and plasma cladding." ***International Journal of Refractory Metals and Hard Materials***, Vol. 84 (2019), 105044. doi: 10.1016/j.ijrmhm.2019.105044
17. Kuldeep, B., K. P. Ravikumar, and S. Pradeep. "Effect of Hexagonal Boron Nitrate on Microstructure and Mechanical Behavior of Al7075 Metal Matrix Composite Producing by Stir Casting Technique." ***International Journal of Engineering, Transactions A: Basics***, Vol. 32, No. 7 (2019), 1017-1022. doi: 10.5829/IJE.2019.32.07A.15
18. Khajesarvi, A., and G. Akbari. "Effect of Mo addition on nanostructured Ni50Al50 intermetallic compound synthesized by mechanical alloying." ***International Journal of Engineering-Transection C: Aspects***, Vol. 28 (2015), 1328. doi: 10.5829/idosi.ije.2015.28.09c.10
19. Wen, Xin, Xiufang Cui, Guo Jin, Xuerun Zhang, Ye Zhang, Dan Zhang, and Yongchao Fang. "Design and characterization of FeCrCoAlMn0. 5Mo0. 1 high-entropy alloy coating by ultrasonic assisted laser cladding." ***Journal of Alloys and Compounds***, (2020), 155449. doi: 10.1016/j.jallcom.2020.155449
20. Qiu, Xing-wu. "Corrosion behavior of Al$_2$CrFeCo$_x$CuNiTi high-entropy alloy coating in alkaline solution and salt solution." ***Results in Physics***, Vol.12 (2019), 1737-1741. doi: 10.1016/j.rinp.2019.01.090

Persian Abstract

چکیده

علی‌رغم مقاومت در برابر خوردگی عالی، شکل‌پذیری خوب و هزینه‌ی پایین فولاد زنگ‌نزن آستنیتی AISI 316، به علت سختی کم و ویژگی‌های مکانیکی ضعیف، کاربرد آن در چند زمینه‌ی صنعتی محدود می‌شود. برای بهبود ویژگی‌های مکانیکی، پوشش‌های آلیاژی آنتروپی بالای AlFeCuCrCoNi-WC10 از طریق روکش‌کاری فلزی با استفاده از لیزر بر روی زیرلایه‌ی فولاد زنگ‌نزن آستنیتی AISI 316 قرار گرفتند. تاثیر ترکیب WC بر ترکیب دیگر فازها، ریزساختار، ریزسختی و توزیع عناصر به‌ترتیب با طیف‌نگار پرتو ایکس، طیف‌سنجی پراکندگی انرژی (FESEM-EDS) آزمایش‌گر میکرو سختی بررسی شد. قله‌های XRD نشان داد که به سطح پوشیده‌شده با AlFeCuCrCoNi-WC10 ، حاوی فاز غنی از W با ساختار BCC و FCC است ریزساختار ناحیه‌ی پوششی عمدتاً از بلورهای ریزدانه‌ی غیرجهت‌دار و مخلوط در نواحی دور از مواد پایه، و دانه‌های ستونی در نزدیکی مواد پایه تشکیل شده است. طیف‌سنجی پراکندگی انرژی، تفکیک W و Cr را در منطقه‌ی فصل مشترک نشان داد. البته، آن طورکه مشاهده شد، سایر عناصر آلیاژی به‌طور یک‌نواخت در سراسر روکش توزیع شده‌اند. ریزسختی پوشش آلیاژی آنتروپی بالا (AlFeCuCrCoNi-WC10) ۶۷۰ عدد ریزسختی ویکرز ۰/۵ (670 Hv0.5)، یعنی ٤/۵ برابر بیشتر از ماده‌ی زیرلایه (AISI-316) بود.

International Journal of Engineering

Journal Homepage: www.ije.ir

Numerical Analysis of Grease Film Characteristics in Tapered Roller Bearing Subject to Shaft Deflection

Z. H. Wu, Y. Q. Xu*, K. A. Liu, Z. Y. Chen

School of Mechatronical Engineering, Northwestern Polytechnical University, Xi'an 710072, China

PAPER INFO

Paper history:
Received 24 May 2020
Received in revised form 07 June 2020
Accepted 12 June 2020

Keywords:
Film Characteristics
Grease Lubrication
Shaft Deflection
Tapered Roller Bearing

ABSTRACT

The shaft deflection is one of critical factors that deteriorate the grease lubrication state inside the tapered roller bearings (TRBs). So, in this paper, on the premise of the TRB subjected to the combined loads and the shaft is deflected, the grease lubrication model at the TRB contacts was constructed, in which the interaction loads, linear and angular displacements of the bearing parts were involved. Then the grease film characteristics were numerically analyzed from the perspective of the whole bearing to clarify the negative effects of the shaft deflection on grease film characterizes. The results show that the effect of the shaft deflection on the load distribution in the absence of the radial load is greater than that in the presence of the radial load. The angular misalignment of the roller is mainly affected by the deflection. The deflection results in an irregular film shape and pressure profile at the TRB contacts, and induces a significant pressure spike and necking feature at the roller-end.

***doi**: 10.5829/ije.2020.33.07a.28*

NOMENCLATURE

E_r, E_b	Elastic modulus of roller and rings	$\dot{\gamma}$	Shear rate
F_a	Constant preload force	δ_a	Constant preload displacement
F_x, F_y, F_z	External forces	δ_f	Contact deformation at flange
M_y, M_z	External moments	δ_i	Contact deformations at cone
P_h	Hertzian contact pressure	δ_i	Contact deformations at cup
b_h	Half Hertzian contact width	δ^*	Geometric constant
c_v	Sum of the principal curvatures	ς_r, ς_{bi}	Radii of roller and cone
g, g_r, g_b	Geometric clearances	ρ	Grease density
h	Grease film thickness	τ	Grease shear stress
h_0	Rigid body displacement	τ_y	Grease yield viscosity
h_p	Thickness of plug flow layer	υ_b	Poisson's ratio of rings
l_e	Roller effective length	υ_r	Poisson's ratio of roller
m	Roller number	φ	Grease plastic viscosity
n	Grease flow index	ω_i	Cone rotation speed
p	Grease film pressure	ω_s, ω_p	Roller rotation and revolution speeds
s	Roller slice number	**Subscript**	
u_{ex}, u_{ex}	Entrainment speeds	i	Reference coordinate system o_i-$x_iy_iz_i$
u_x, u_y, u_z	Cone linear displacements	b	Cone fixed coordinate system o_b-$x_by_bz_b$
v_x, v_y, v_z	Roller linear displacements	r	Roller fixed coordinate system o_r-$x_ry_rz_r$
Greek Symbols		c	Contact coordinate systems o_c-$x_cy_cz_c$
Ω	Analysis area	**Superscript**	
α_h	Roller half cone angle	i	cone
β, θ, ϕ	Roller rotation, tilt and skew angles	o	cup
γ, λ	Deflection angle	f	flange

*Corresponding Author Institutional Email: *xuyngqng@nwpu.edu.cn* (Y. Q. Xu)

Please cite this article as: Z. H. Wu, Y. Q. Xu, K. A. Liu, Z. Y. Chen, Numerical Analysis of Grease Film Characteristics in Tapered Roller Bearing Subject to Shaft Deflection, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1403-1412

1. INTRODUCTION

Tapered roller bearings, as a separate bearing, can withstand combined radial and thrust loads, and widely exist in high-speed trains, automobiles, rolling mills and other machines. In most of these machines, the grease is commonly adopted as a lubricant to reduce bearing friction, for it can simplify designs of sealing devices and lubricating systems. [1] However, due to the installation error, the mass on the shaft, and the load from the gear system, etc., the bearing shaft would be more or less deflected. [2, 3] The deflected shaft in turn occasions bearing parts (rollers and rings) in a misaligned state, and results in disordered spatial position and uneven load distribution. Ultimately, it inevitably deteriorates the grease lubrication inside the bearing [4, 5]. In addition, coupled with the non-Newtonian properties of the grease, the lubrication problem of the TRB subjected to shaft deflection becomes more complicated. Therefore, the behavior of the grease film inside the TRB when the shaft is deflected should be clarified, which would expand lubrication theory and benefit lubrication design for the TRBs.

The literatures focused on negative effects of the shaft deflection on bearing performances mainly emphasized on the displacements, load-carrying, and fatigue life, etc., [2, 6-12] while on impacts of the misaligned shaft on oil or grease lubrication of the TRB are relatively deficient. [13-17] Harris [2] presented a representative load-deformation relationship to evaluate effects of the shaft misalignment on fatigue life reduction of cylindrical roller bearings. Similar to Harris's method, Zantopulos [6] studied the misaligned shaft's effects on bearing capacity and fatigue life of the TRBs. However, centrifugal force and gyroscopic moment of the rollers were neglected. Andréason [7] presented a vector method to investigate load distribution inside the misaligned TRB. Similarly, the inertial loads of the rollers were not addressed. Liu [8] further extended Andréason's works and considered the inertial loads. De Mul et al. [9] presented an analytical model for equilibrium and associated load distribution in TRBs. Although the model allowed five degrees of freedom for loads and displaces, the roller profile crowning is not considered. Based on De Mul's theory, Tong et al. [10-12] analyzed the stiffness, fatigue life and running torque with the effects of the combined loads and shaft deflection. The above-mentioned works can provide a reference for the grease lubrication analysis of the misaligned TRB. Kushwaha et al. [13] provided a solution for the finite line concentrated contact of a cylindrical roller-to-race under aligned and misaligned states. The oil-lubricated contact is subject to an elastohydrodynamic regime of lubrication under isothermal conditions. Panovko [14] presented a numerical method to study the effect of the roller misalignment on the film pressure and thickness of the oil lubricant in the contact region of the crowned roller pair. Liu et al. [15, 16] separately investigated effects of the roller tilt and skew on the oil-lubricated roller/race interface in the cylindrical roller bearing. The cylindrical roller is a constant cross-section structure, which is quite different from the tapered roller. Wu et al. [17] analyzed grease lubrication at misaligned tapered roller pair, and found that in the case of relatively large radius ratio of contact rollers, the tilt's effect on the film characterizes is more serious than the skew's.

All the above studies on lubrication use a simplified contact model of a pair of rollers, not from the perspective of the whole bearing. Therefore, in this paper, the responses of loads and displacements of the TRB parts with deflected shaft were analyzed first. Then, with regard of the TRB responses, the structure features of the bearing, and the grease-rich lubrication, the grease film characterizes inside the misaligned TRB were numerically studied in the whole bearing perspective.

2. THEORETICAL FORMULATION

2. 1. Load-carrying of TRB The TRB is a kind of roller bearings with strong structural nonlinearity. Assuming the cup is macroscopically stationary, as depicted in Figure 1, the reference system o_i-$x_iy_iz_i$ is arranged at the mass center of the cone, the cone fixed system o_b-$x_by_bz_b$ coincides with the o_i-$x_iy_iz_i$ system under unloaded condition. The external force and moment the TRB carried are $\mathbf{F}_e=[F_x,F_y,F_z]^T$ and $\mathbf{M}_e=[0,M_y,M_z]^T$. The shaft deflection is caused by the moment $\mathbf{M}_e$, and means that the geometric axis o_b-x_b of the bearing unparallel to the rotation axis o_i-x_i. In o_i-$x_iy_iz_i$, linear displacements of the cone and the roller are $\mathbf{w}_i=[u_x,u_y,u_z]^T$ and $\mathbf{v}_i=[v_x,v_y,v_z]^T$. Attitude angles of the cone and roller in their respective fixed systems are $\boldsymbol{\varepsilon}_b=[0,\gamma,\lambda]^T$ and $\boldsymbol{\kappa}_r=[\beta,\theta,\phi]^T$. The γ and λ are shaft deflection angles; the β, θ and ϕ are rotation, tilt and skew angles of the tapered roller, respectively.

For the roller/race assembly, as shown in Figure 1(b, c), the interaction is dealt with the slicing method [1]. For each slice, the Palmgren's load-deformation relationship is adopted [18], that is

$$q_{i,o} = 0.7117\delta_{i,o}{}^{\frac{10}{9}}\left(\frac{1-\upsilon_r^2}{E_r}+\frac{1-\upsilon_b^2}{E_b}\right)^{-1}\left(\frac{l_e}{s}\right)^{\frac{8}{9}} \tag{1}$$

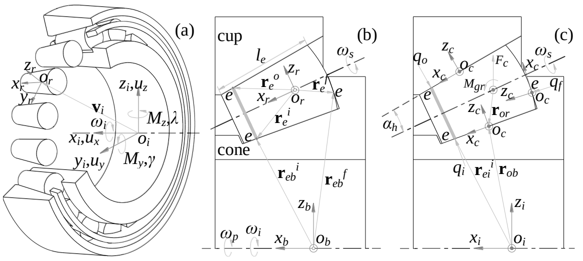


Figure 1. Description of geometric structure and coordinate system of the TRB

If profile crowning of the roller is considered, the crown drop should be subtracted from the roller radius, the crown drop σ along the roller length is [19]:

$$\sigma(l) = \begin{cases} \frac{k_0}{l_1+l_2}\left[l_1 + l_2\left(0.8864 + ln\frac{1}{\alpha} + ln\frac{l_2-l_1}{l_1}\right)\right], \ l = l_1 \\ k_0\left\{1 + \frac{l}{l_1+l_2}\left[1.7727 + 2\,ln\frac{1}{\alpha} + ln\frac{(l_2-l)(l-l_1)}{l^2}\right]\right\}, l_1 < l < l_2 \\ \frac{k_0}{l_1+l_2}\left[l_2 + l_1\left(0.8864 + ln\frac{1}{\alpha} + ln\frac{l_2-l_1}{l_2}\right)\right], l = l_2 \end{cases} \quad (2)$$

where l_1 and l_2 are generatrix lengths from the small end and the large end to the taper vertex of the roller, l_e=l_2-l_1; l stands for generatrix length at the contact point e; k_0 and α are coefficients and explicated in [19].

For the flange/roller-end assembly, it is a typical elliptical contact, the interaction force is [18]:

$$q_f = \frac{\pi c_v}{3}\left(\frac{1-v_r^2}{E_r} + \frac{1-v_b^2}{E_b}\right)^{-1}\left(\frac{2\delta_f}{\delta^* c_v}\right)^{\frac{3}{2}} \quad (3)$$

Since the roller's skew has little effect on grease film [18], in order to simplify the model, the skew and the friction at the bearing contacts were not considered. In the contact systems o_c-$x_cy_cz_c$, the interaction forces at the TRB contacts can be rewritten as vector forms.

$$\mathbf{Q}_i = \left[0,0,q_i\right]^T, \mathbf{Q}_o = \left[0,0,-q_o\right]^T, \mathbf{Q}_f = \left[0,0,q_f\right]^T \quad (4)$$

For equilibriums of tapered rollers, centrifugal force F_c and gyroscopic moment M_{gr} are included. Thus:

$$\sum_{k=1}^{s}\boldsymbol{T}_{ci}^{i}\boldsymbol{Q}_i + \sum_{k=1}^{s}\boldsymbol{T}_{ci}^{o}\boldsymbol{Q}_o + \boldsymbol{T}_{ci}^{f}\boldsymbol{Q}_f + \boldsymbol{T}_{ri}[0,0,F_c]^T = \boldsymbol{0} \quad (5)$$

$$\sum_{k=1}^{s}\boldsymbol{r}_e^i \times \boldsymbol{T}_{cr}^{i}\boldsymbol{Q}_i + \sum_{k=1}^{s}\boldsymbol{r}_e^o \times \boldsymbol{T}_{cr}^{o}\boldsymbol{Q}_o + \boldsymbol{r}_e^f \times \boldsymbol{T}_{cr}^{f}\boldsymbol{Q}_f + \left[0, M_{gr}, 0\right]^T = \boldsymbol{0} \quad (6)$$

where $\mathbf{T}_{ci}{}^{i,o,f}$ are conversion matrices from contact coordinate systems o_c-$x_cy_cz_c$ at the cone, cup, and flange to the o_i-$x_iy_iz_i$ system; $\mathbf{T}_{cr}{}^{i,o,f}$ are conversion matrices from the systems o_c-$x_cy_cz_c$ to the roller fixed system o_r-$x_ry_rz_r$; $\mathbf{T}_{ri}$ is conversion matrix from o_r-$x_ry_rz_r$ to o_i-$x_iy_iz_i$; $\mathbf{r}_e{}^{i,o,f}$ are locations of contact points e in o_r-$x_ry_rz_r$.

Simultaneously, the equilibrium for the cone is:

$$\sum_{j=1}^{m}\left[\sum_{k=1}^{s}\boldsymbol{T}_{ci}^{i}\boldsymbol{Q}_i + \boldsymbol{T}_{ci}^{f}\boldsymbol{Q}_f\right] + \boldsymbol{F}_e = \boldsymbol{0} \quad (7)$$

$$\sum_{j=1}^{m}\left[\sum_{k=1}^{s}\boldsymbol{r}_{eb}^{i} \times \boldsymbol{T}_{cb}^{i}\boldsymbol{Q}_i + \boldsymbol{r}_{eb}^{f} \times \boldsymbol{T}_{cb}^{f}\boldsymbol{Q}_f\right] + \boldsymbol{M}_e = \boldsymbol{0} \quad (8)$$

where $\mathbf{T}_{ci}{}^{i,f}$ are conversion matrices from the systems o_c-$x_cy_cz_c$ at the cone and flange to the o_i-$x_iy_iz_i$ system; $\mathbf{T}_{cb}{}^{i,f}$ are conversion matrices from the systems o_c-$x_cy_cz_c$ to o_b-$x_by_bz_b$; $\mathbf{r}_{eb}{}^{i,f}$ are locations of contact points e in o_b-$x_by_bz_b$.

The equilibrium Equations (5)-(8) contain 4m+5 displacement quantities and were solved with the Newton-Raphson method. The process can refer to [11]. The obtained TRB responses will be included in the lubrication model below.

2. 2. Governing Equations of Grease Lubrication

It is assumed that the grease adheres uniformly to the working surfaces of the TRB. The rheological behavior of the grease can be described with the Herschel-Bulkey (H-B) constitutive model, that is

$$\tau = \tau_y + \varphi\dot{\gamma}^n \quad (9)$$

As presented in Figure 2, the lubrication analysis at the roller/race interface was performed in a rectangular area Ω in the o-xy plane, and the o-xyz system is fixed in the reference system and coincides with the o_c-$x_cy_cz_c$ system under unloaded state. Based on the H-B model, Navier-Stokes equation and continuity condition of fluid, the steady state of the grease Reynolds equation is

$$\frac{\partial(\rho G_R)}{\partial x} + \frac{\partial(\rho V_R)}{\partial y} = \frac{\partial(\rho u_{ex}h)}{\partial x} + \frac{\partial(\rho u_{ey}h)}{\partial y} \quad (10)$$

where

$$G_R = \frac{n}{2n+1}\left(\frac{1}{\phi}\frac{\partial p}{\partial x}\right)^{\frac{1}{n}}\left(\frac{h-h_p}{2}\right)^{1+\frac{1}{n}}\left(h + \frac{nh_p}{2n+1}\right) \quad (11)$$

$$V_R = \frac{n}{2n+1}\left(\frac{1}{\phi}\frac{\partial p}{\partial y}\right)^{\frac{1}{n}}\left(\frac{h-h_p}{2}\right)^{1+\frac{1}{n}}\left(h + \frac{nh_p}{2n+1}\right) \quad (12)$$

where h_p indicates thickness of plug flow layer, which is generated due to yield behavior of the grease. Referring to the boundary condition between sheared and non-sheared flow layers, that is, τ =τ_y, h_p is deduced as:

$$h_p = \frac{2\tau_y}{\sqrt{\left(\partial p/\partial x\right)^2 + \left(\partial p/\partial y\right)^2}} \quad (13)$$

The film shape h at the roller/race interface is

$$h(x,y) = h_0 + g(x,y) + \frac{1}{\pi E^*}\iint_{\Omega}\frac{p(\xi,\zeta)d\xi d\zeta}{(x-\xi)^2+(y-\zeta)^2} \quad (14)$$

Since the cup is stationary, the clearance $g(x,y)$ at the roller/cup interface is easier to be determined. For the roller/cone interface, both the roller and the cone are disordered. To describe the clearance, the point in the o-xyz system is first marked on the roller/cone before the bearing is disordered. After the marked point moved with the roller/cone, its changed position in o-xyz can be known.

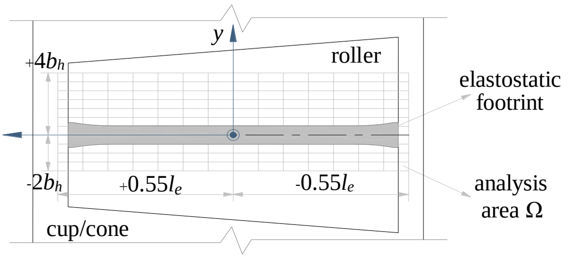


Figure 2. lubrication analysis area at the race contact

Similarly, the marked points $[x_r,y_r,z_r]^T$ and $[x_b,y_b,z_b]^T$ on the roller/cone can be deduced from the changed point $[x,y]^T$ in the regular analysis area Ω in the *o-xy* plane. That are:

$$[x_r, y_r, z_r]^T = \boldsymbol{T}'_{cr0}\{\boldsymbol{T}_{ir}(\boldsymbol{T}'_{ic}[x, y, 0]^T - \boldsymbol{v}_i) - \boldsymbol{r}_{or}\} \tag{15}$$

$$\left[x_b, y_b, z_b\right]^T = \mathbf{T}'_{cb0}\left\{\mathbf{T}_{ib}\left(\mathbf{T}'_{ic}\left[x, y, 0\right]^T - \mathbf{w}_i\right) - \mathbf{r}_{ob}\right\} \tag{16}$$

where $\mathbf{r}_{or,ob}$ are locations of the origin *o* in the roller and cone fixed systems, respectively; $\mathbf{T}_{cr0,cb0}$ denote conversion matrices for the ordered bearing.

The marked points $[x_r,y_r]$ and $[x_b,y_b]$ are the points in the *o-xy* plane before the bearing is disordered. Then, initial surface clearances of the roller and the cone to the *o-xy* plane are:

$$g_r(x_r, y_r)|_{(x,y)} = k_e\varsigma_e(x_r, y_r) - \sqrt{k_e^2\varsigma_r(x_r, y_r)^2 - x_r^2} \tag{17}$$

$$g_b(x_b, y_b)|_{(x,y)} = \sqrt{k_e^2\varsigma_{bi}(x_b, y_b)^2 - x_b^2} + k_e\varsigma_{bi}(x_b, y_b) \tag{18}$$

where k_e is equivalent coefficient of roller radius ς_r and inner-race radius ς_{bi} projected into the *o-yz* plane,

$$k_e = \frac{1 + 2\sin^2\alpha_h + \sin^4\alpha_h \sec 2\alpha_h}{\cos\alpha_h + \sin\alpha_h \tan 2\alpha_h} \tag{19}$$

After the bearing is loaded, the changed clearances $g_r(x,y)$ and $g_b(x,y)$ are:

$$[x, y, g_r(x, y)]^T = \boldsymbol{T}_{ic}\{\boldsymbol{T}'_{ir}(\boldsymbol{T}_{cr0}[x_r, y_r, g_r(x_r, y_r)]^T + \boldsymbol{r}_{cr}) + \boldsymbol{v}_i\} \tag{20}$$

$$[x, y, g_b(x, y)]^T = \boldsymbol{T}_{ic}\{\boldsymbol{T}'_{ib}(\boldsymbol{T}_{cb0}[x_b, y_b, g_b(x_b, y_b)]^T + \boldsymbol{r}_{cb}) + \boldsymbol{u}_i\} \tag{21}$$

Then the clearance $g(x,y)$ is $g_b(x,y)$-$g_r(x,y)$, and linear and angular displacements of bearing parts are contained.

Surface speeds on the roller and races are [20]:

$$\boldsymbol{u}_r^{i,o} = \boldsymbol{T}_{ir}\boldsymbol{T}_{ic}^{i,o'}\{[\omega_s, 0,0]^T \times \boldsymbol{r}_e^{i,o}\} \tag{22}$$

$$\boldsymbol{u}_b^i = \boldsymbol{T}_{ic}^i\left\{\left(\boldsymbol{T}'_{ib}[\omega_i, 0,0]^T - [\omega_p, 0,0]^T\right) \times \boldsymbol{r}_{ei}^i\right\} \tag{23}$$

$$\boldsymbol{u}_b^o = \boldsymbol{T}_{ic}^o\left\{[\omega_p, 0,0]^T \times \boldsymbol{r}_{eb}^o\right\} \tag{24}$$

Then entrainment speeds u_{ex} and u_{ey} with which the grease is swept into the race contact are

$$u_{ex} = \left.\frac{\mathbf{u}_r + \mathbf{u}_b}{2}\right|_x, u_{ey} = \left.\frac{\mathbf{u}_r + \mathbf{u}_b}{2}\right|_y \tag{25}$$

The viscosity- and density-pressure relationships are detailed in [17]. The grease film pressure over the area Ω should meet the normal contact force. That is:

$$\sum_{k=1}^{s} q_{i,o} = \iint_{\Omega} p(x, y)\,\mathrm{d}x\mathrm{d}y \tag{26}$$

Due to the spherical structure of the roller-end, the flange/roller-end assembly is still an elliptical contact under bearing state or shaft deflection. In fact, the bearing and deflection primarily affect the value of the interaction force q_f. Therefore, the lubrication at this assembly was not addressed in the article.

As shown in Figure 2, a rectangular computational domain for the analysis of film behavior at the race contact is defined, inlet and outlet limits are located at $4b_h$ and $-2b_h$, respectively, and side limits are located in $-0.55l_e \le x \le 0.55l_e$. Boundary conditions of the film pressure p are $p \ge 0$ and $p(x,4b_h)=p(\pm0.55l_e,y)=0$. The Reynolds equation was numerically simulated with the multi-grid method [21], in which the film pressure was iterated by the Gauss-Seidel method and the surface elastic deformation was solved with the DC-FFT method [22]. A total of 3 grid layers are arranged, where the number of nodes on the densest grid layer is 361(*o-y* axis)×1025(*o-x* axis). The iterative errors of the pressure and load are respectively required less than 1×10^{-4} and 1×10^{-5}.

3. COMPUTATIONAL RESULTS

The parameters of the TRB and the grease are detailed in Table 1. It is assumed that the rollers are evenly distributed in the bearing and symmetrically distributed about the o_i-z_i axis. In view of this, only radial load F_z in the o_i-z_i axis is considered, and F_y=0. The shaft deflection is an angular displacement caused by applied moment, so the load-carrying capacity and lubrication state of the TRB can be analyzed under different deflection angles. Generally, as a separable bearing, the TRB needs to be preloaded by a constant axial force F_a or displacement δ_a to ensure its normal service.

3. 1. Constant Force Preload Under the premise of constant force preloaded and no radial load applied, as depicted in Figure 3(a-c), the impact of the deflected shaft on the load-carrying of the TRB is manifest. The load on part of tapered rollers is increased several times compared to the bearing in the aligned state (γ=0), while the other part is in a 'relaxed' state. The distributions of the contact force at the races and flange are the same, and have same tendency to change with the shaft deflection. The contact force at the outer-race is slightly larger than that at the inner-race. This is due to the centrifugal force and gyroscopic moment of the roller. In Figure 3(d), the rollers are all in a negative tilted state ($\theta\approx$-0.005mrad) when the shaft is aligned. The farther away from the o_i-y_i axis, the more severe the roller is tilted when the shaft is

TABLE 1. Parameters of the TRB and the grease

Bearing parameters	**Value**
Bearing inner diameter [mm]	130.00
Bearing outer diameter [mm]	230.00
Bearing width [mm]	70.00
Number of rollers	17
Roller length [mm]	42.06
Roller average diameter [mm]	21.37
Roller half-cone angle [deg]	5.00
Radius of roller spherical end [mm]	153.95
Outer-race contact angle [deg]	15.00
Inner-race contact angle [deg]	5.00
Flange contact angle [deg]	83.00
Elastic modulus of roller and rings [GPa]	2.06
Poisson's ratio of roller and rings	0.30
Bearing speed [krpm]	3.50
Grease parameters	**Value**
Density [kg/m^3]	872
Yield stress [Pa]	153.14
Plastic viscosity [Pa·s^n]	1.7830
Flow index	0.6943

deflected. In Figure 3(e), the deflection shortens effective contact length l_c at the roller/cup interface. In [11, 12], the TRB's inner diameter is 30mm and deflection can reach 5mrad. In this analysis, that are 130mm and ±1mrad. Compared with the deflection's effect in [11, 12], it can be speculated that if the TRB with a large inner diameter, the allowable shaft deflection should be smaller.

The misalignment of the TRB induced by the shaft deflection or external loads would disturb the film forming at the bearing contacts. Figure 4 illustrates the film characteristics at the most-loaded tapered roller, i.e., No. 1 and 9 rollers (the rollers are numbered clockwise, and the No. 1 roller is at 0). The white arrow represents the grease flow direction, and the film thickness and pressure are dimensionless, i.e. $H= h\varsigma_a/b_h^2$, $P=p/P_h$, ς_a= 11.97mm, b_h=0.21mm, P_h=0.97GPa, the same below. It can be seen that the grease film deviated from the normal state (the case of γ=0). The misalignment of the shaft causes a serious tilt of the roller. As the contact area is reduced, it leads to a decrease in film thickness and an increase in film pressure. The thin-film or high-pressure area at tilted rollers exhibits an approximately triangular shape. Due to larger equivalent radius and higher swept speed (Equation (24)), the film at the roller/cup interface is thicker [20]. Note that typical necking feature appears at the outlet region, even if the tapered rollers is in a severely tilted state.

Figure 5 presents the grease film thickness when the shaft deflected at -1mrad. Under combined effect of the loads and displacements, the grease film around the race shows a regular change. Similarly, the film exhibits the necking feature at the outlet region. In particular, the load carried by No. 1~2 rollers are small, for degraded action area and roller's contact length, a severe tilt still induces a thinner film. Figure 6 reveals the film characterizes along the o_c-x_c axis. A notable necking feature and pressure spike occurs at the contact end, and it resembles the stress edge effect to some extent. [1] The pressure peak p_m fluctuates greatly and it is higher at the cone. Although the film at the outer-race is thicker, as shown in Figure 6, it is not much different in thickness at the inner-race. The pressure p_m may cause stress concentration in the subsurface of the roller or races, thereby reducing the fatigue life of the bearing.

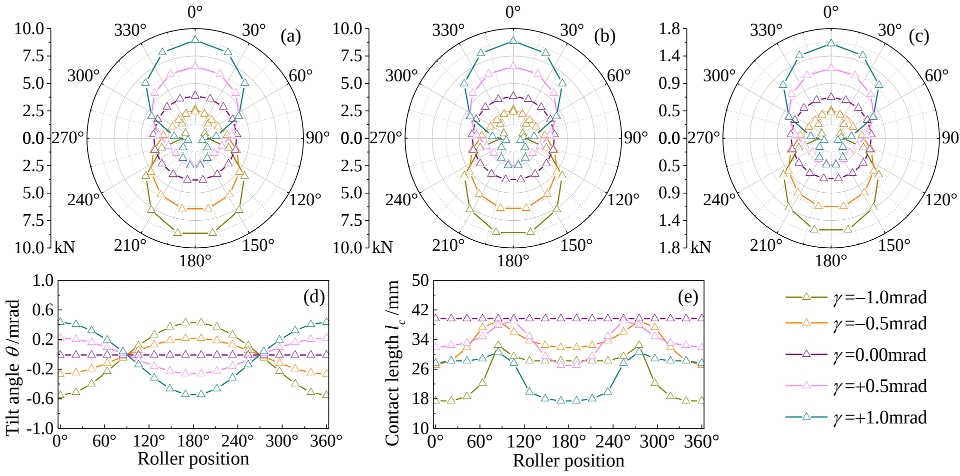


Figure 3. Bearing condition of TRB under constant force preload. F_a=17kN, F_z=0. Normal force (a) at the outer-race, (b) at the inner-race, (c) at the flange. (d) tilt angle, (e) contact length at the outer-race

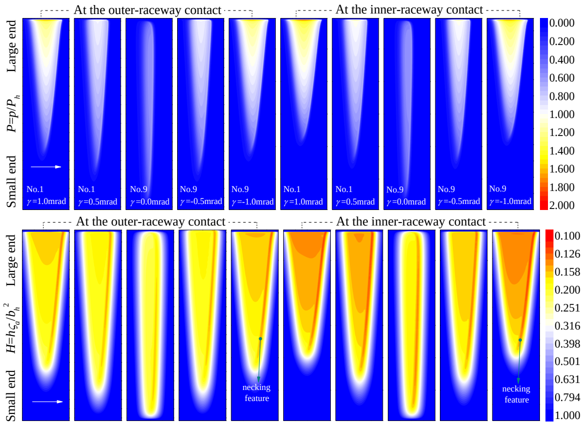


Figure 4. Grease film characterizes of the most loaded roller

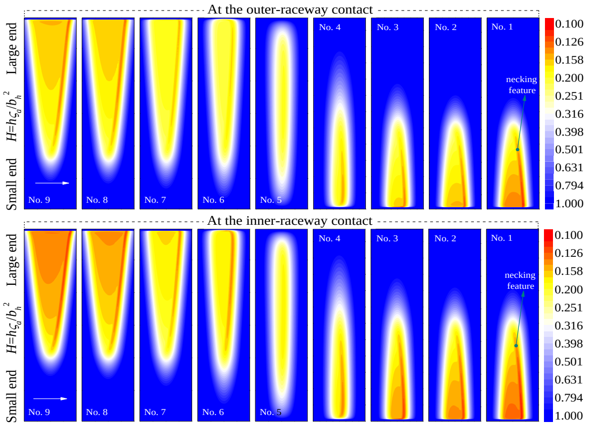


Figure 5. Film thickness at the race contacts when γ=-1mrad

In Figures 4 and 5, for one of the tilted tapered rollers, either the large or small end is in contact with the cone and the cup at the same time, instead of two ends being in contact with the cone and cup, respectively. Compared with Figure 3(e), the contact length of the roller at hydrodynamic contact is longer than that at dry contact; this confirms the bearing function of grease film.

When the TRB carries the radial load F_z, in Figure 7(a), it is substantially in a half-turn bearing state [10]. The force distribution is less affected by the misaligned shaft, which quite differs from the case with no radial load in Figure 3(a). Variations of the force distribution at the races and the flange with the shaft deflection are the same (the variations are not drawn here). In Figure 7(b), the tilted state of the tapered rollers coincides with the result in Figure 3(d), the rollers are still in a severe tilted state due to the shaft deflection. The amplitudes of the tilt angle are substantially similar, whether the radial load F_z is applied or not. However, in Figure 7(c), the effect on the effective contact length at the roller/race interface is remarkable. The contact length when the shaft is aligned is generally longer than that when the shaft is deflected. Although the radial load F_z reduces the contact length of the roller in the relaxed half-circle (γ=0), it has a slight effect on the roller tilt and the rollers are tilted in -0.015~-0.022mrad.

Under the loading state of F_a=28kN and F_z=-43kN, Figure 8(a) reveals the film conditions at the roller/cup interface of the highest loaded No. 9 roller. Due to the misalignment of the roller or shaft, the film thickness and pressure at the roller squeezed section are respectively reduced and increased. This is consistent with the results in [13]. With the shaft deflected from +1mrad to -1mrad, the tilt angle of the roller θ is -0.54, -0.27, -0.02, 0.22 and 0.44mrad. The film pressure peak p_m transfers from the small end to the large end and is 2.32, 1.80, 0.93, 1.53 and 1.88GPa. That is, the pressure p_m is positively correlated with the tilt angle θ. When γ=0, the tilt angle θ is -0.02mrad and is relatively small, the shape of the thin-film area is mainly related to the non-equal cross-section structure of the bearing. Figure 9(a~e) presents the film characteristics along the o_c-x_c axis. Although the tilt angle and the carried load of the roller are huge in the case of negative tilt (θ<0), the overall film pressure is smaller

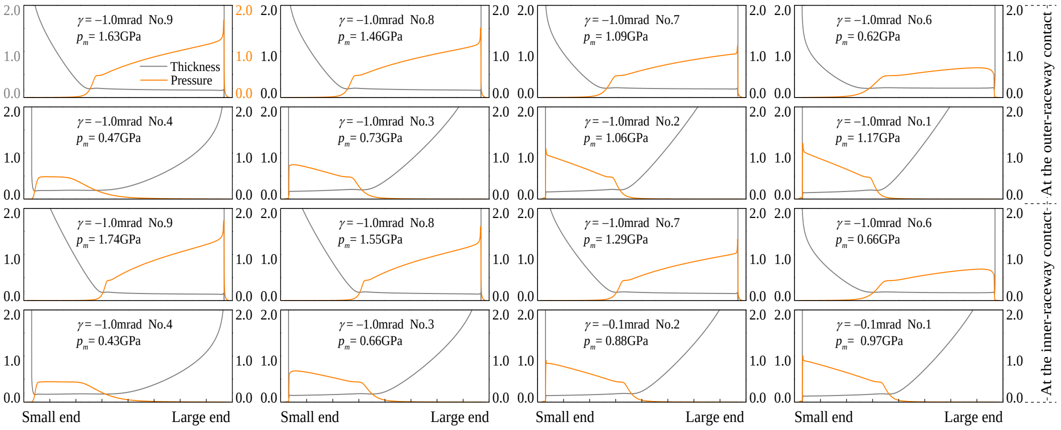


Figure 6. Grease film shape and pressure profile at y=0 when γ=-1mrad

than that in the positive tilted state (θ>0). This is due to the large radius section coming into contact when the tapered roller is in negative tilt. In fact, it indicates that the crown drop at the small radius section of the roller should be larger than the large radius section.

In addition, when the shaft deflection is -1mrad, film behaviors of the rollers at different positions are revealed in Figure 8(b). The film exhibits a regular change, and the necking feature is still remarkable at the outlet region. Even though some rollers subjected to a lower load, such as No. 1~3 rollers, the large tilt angle θ still induces a thinner grease film. The elastostatic footprint of light-loaded rollers is relatively narrow. In Figure 9(e~l), the film pressure spike and the necking feature appear at the roller-ends, especially for the rollers far away from the o_i-y_i axis. The pressure peak p_m is in the range of 0.63~ 1.88GPa. As with the results in Figure 6, the edge pressure is increased several folds. This is consistent with the results in [17]. It indicates that the efficacy of the roller profile crowning (Equation (2)) is weakened. Therefore, the misalignment of the shaft or roller should be properly considered when designing the profile of the roller. In the above analysis, it seems that the TRB's angular displacement has more pronounced effect on grease film than the linear displacement. In fact, any displacements will essentially affect the interaction at the contacts, thus affecting the film-forming performance. [13, 14]

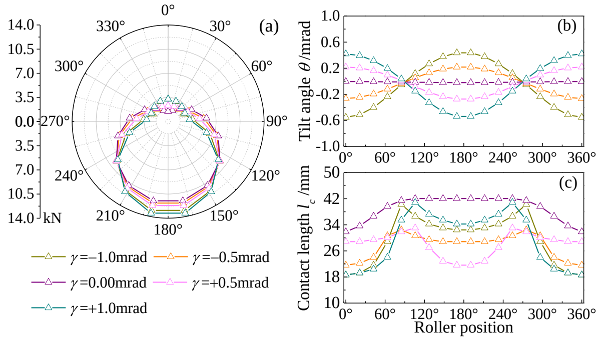


Figure 7. Bearing condition of the TRB under constant force preload. F_a=28kN, F_z=-43kN. (a) contact force at the outer-race. (b) tilt angle. (c) contact length at the outer-race

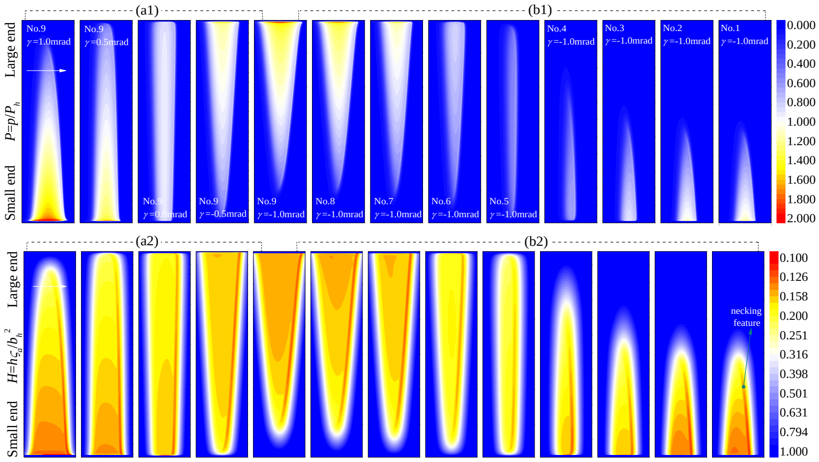


Figure 8. Film characteristics inside the TRB

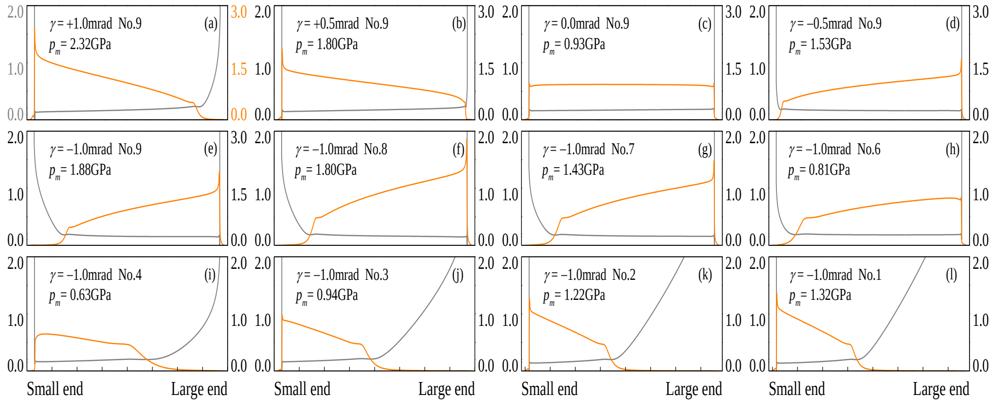


Figure 9. Grease film shape and pressure profile at y=0

3. 2. Constant Displacement Preload The constant displacement preload is a common preloading strategy for bearings. Figure 10(a, b) reveals the bearing condition of the TRB under the preload of δ_a=15μm and δ_a=30μm. In Figure 10(a), the load distribution varies with the shaft deflection is analogous to the constant force preload case, and the variation is more prominent. When the bearing with the radial load in Figure 10(b), it is substantially in a half-turn bearing state. If the shaft deflection continues to intensify, the half-turn bearing state would be changed.

In Figure 10(c, d), the variation and the amplitude of the roller tilt angle are basically consistent when the radial load F_z is 0 and -43kN. This is also reflected in Figures 3 and 7. Therefore, it can be inferred that the misaligned state of the roller is mainly produced by the external moment or shaft deflection, and is less affected by the radial load.

Using the results in Figure 10, the grease film at the TRB contacts was analyzed in three cases: 1) without the shaft deflection and with the radial load F_z, 2) with the shaft deflection and the load F_z, and 3) with the shaft deflection and without the load F_z. Figure 11(a) reveals the first case. The necking feature and pressure spike appear at the end of No. 9 roller, while the pressure profiles of other rollers are smooth. Note that the roller profile crowning can well ensure the film performance under this loading condition. The range of the pressure peak p_m is 0.01~0.85GPa. For No. 1 roller, the interaction of it with the cup is only 203.04N, but its deformation is obvious compared with the original profile.

Once the shaft is deflected, in Figure 11(b, c), the film characteristics are quite different from that without the shaft deflection, and their variations with the roller position are similar, whether or not the radial load F_z is applied. This is due to the fact that roller tilt is mainly affected by the deflected shaft. Meanwhile, significant film pressure spike and necking feature occur at roller-ends, especially for some light-loaded rollers, such as No. 1~3 rollers. The pressure peaks p_m at the contact edge

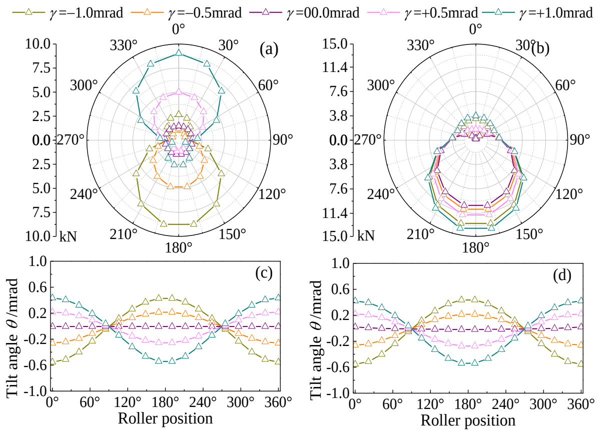


Figure 10. Bearing condition. (a, c) without radial load, F_z=0, δ_a=15μm. (b, d) with radial load, F_z=-43kN, δ_a=30μm

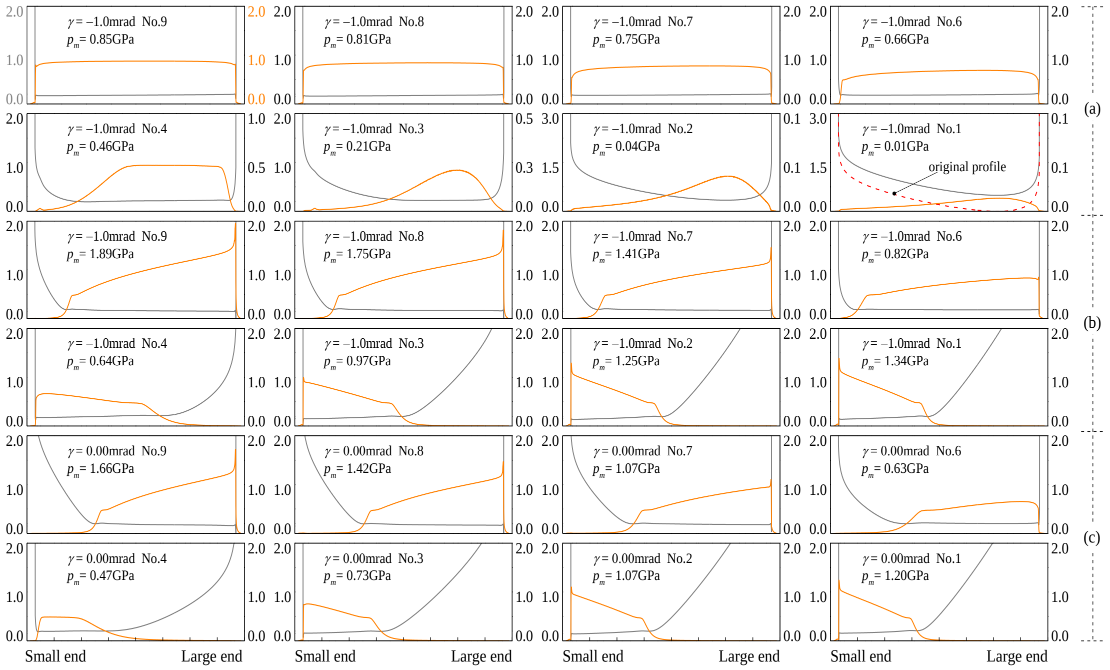


Figure 11. Grease film behavior under constant displacement preload. (a) F_z=-43kN, δ_a=30μm, γ=0mrad. (b) F_z=-43kN, δ_a=30μm, γ=-1mrad. (c) F_z=0, δ_a=15μm, γ=-1mrad

are in the ranges of 0.64~1.89GPa and 0.47~1.66GPa whether the force F_z is loaded or not. Since δ_a and F_z in Figure 11(b) are large, the roller contact length is generally longer than that in Figure 11(c), and the overall grease film is thinner. A common phenomenon is that the contact length of heavy-loaded rollers, as presented in Figure 6, 9 and 11(b, c), is greater than that of light-loaded rollers, even if they are in a severely tilted state. In summary, regardless of the TRB being preloaded by the force or displacement, the shaft misalignment has a negative impact on the bearing state and the film behavior at the bearing contacts.

4. CONCLUSION

The grease film characteristics inside the TRB subjected to the shaft deflection were addressed, negative effects of the misaligned shaft on bearing lubrication were clarified. The deflected shaft deteriorates the load-carrying capacity of the TRB. Roller tilt is mainly affected by the deflection and is less affected by the preload and radial load. When the shaft is in aligned state, rollers are all in a negative tilted state. The shaft deflection causes film characterizes at the TRB contacts to deviate from the normal state, and results in an irregular film shape and pressure profile. In addition, the deflection weakens the efficacy of the roller profile crowning, significant pressure spike and necking feature appear at the roller-end. Under loading conditions in this paper, the film thickness and pressure peak at the inner-race are thinner and greater than that at the outer-race, respectively. In the shaft-deflected state, the film thickness is reduced at some bearing contacts, but not much different in amplitude at all bearing contacts, and the film pressure peak has a large fluctuation.

The method in the paper is also applicable to the analysis of film behaviors inside oil-lubricated roller bearings in the shaft-deflected state. This study focused on grease-rich lubrication condition, the replenishment and starvation of the grease were neglected, which can be considered in the future work.

5. ACKNOWLEDGEMENT

The paper was supported by the National Natural Science Foundation of China (No. 51675427) and the Science and Technology Key Project of He'nan Province (No. 192102210026).

6. REFERENCES

1. Lugt, P. M., "A review on grease lubrication in rolling bearings", ***Tribology Transactions***, Vol. 52, No. 4, (2009), 470-480. doi: 10.1080/10402000802687940
2. Harris, T. A., "The effect of misalignment on the fatigue life of cylindrical roller bearings having crowned rolling members", ***Journal of Tribology***, Vol. 4, No. 2, (1969), 294-300. doi: 10.1115/1.3554918
3. Ebadi, P., Soleimani, M., and Beheshti, M., "Design methodology of base plates with column eccentricity in two directions under bidirectional moment", ***Civil Engineering Journal***, Vol. 4, No. 11, (2018), 2773-2786. doi: 10.28991/cej-03091197
4. Gurumoorthy, K. and Ghosh, A., "Failure investigation of a taper roller bearing: a case study", ***Case Studies in Engineering Failure Analysis***, Vol. 1, No. 2, (2013), 110-114. doi: 10.1016/j.csefa.2013.05.002
5. Dzyura, V. O., Maruschak, P. O., Zakiev, I. M., and Sorochaka, A. P., "Analysis of inner surface roughness parameters of load-carrying and support elements of mechanical systems", ***International Journal of Engineering, Transactions B: Applications***, doi. 30, No. 8, (2017), 1170-1175. Doi: 10.5829/ije.2017.30.08b.08
6. Zantopulos, H., "The effect of misalignment on the fatigue life of tapered roller bearings", ***Journal of Tribology***, Vol. 94, No. 2, (1972), 181-186. doi: 10.1115/1.3451678
7. Andréason, S., "Load distribution in a taper roller bearing arrangement considering misalignment", ***Tribology***, Vol. 6, No. 3, (1973), 84-92. doi: 10.1016/0041-2678(73)90241-8
8. Liu, J. Y., "Analysis of tapered roller bearings considering high speed and combined loading", ***Journal of Tribology***, Vol. 98, No. 4, (1976), 564-572. doi: 10.1115/1.3452933
9. De Mul, J. M., Vree, J. M., and Maas, D. A., "Equilibrium and associated load distribution in ball and roller bearings loaded in five degrees of freedom while neglecting friction-part II", ***Journal of Tribology***, Vol. 111, No. 1, (1989), 149-155. doi: 10.1115/1.3261865
10. Tong, V. C. and Hong, S. W., "Characteristics of tapered roller bearing subjected to combined radial and moment loads", ***International Journal of Precision Engineering and Manufacturing***, Vol. 1, No. 4, (2014), 323-328. doi: 10.1007/s40684-014-0040-1
11. Tong, V. C. and Hong, S. W., "The effect of angular misalignment on the running torques of tapered roller bearings", ***Tribology International***, Vol. 95, (2016), 76-85. doi: 10.1016/j.triboint.2015.11.005
12. Tong, V. C. and Hong, S. W., "The effect of angular misalignment on the stiffness characteristics of tapered roller bearings", ***Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science***, Vol. 231, No. 4, (2017), 712-727. doi: 10.1177/0954406215621098
13. Kushwaha, M., Rahnejat, H., and Gohar, R., "Aligned and misaligned contacts of rollers to races in elastohydrodynamic finite line conjunctions", ***ARCHIVE Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 1989-1996***, Vol. 216, No. 11, (2002), 1051-1070. doi: 10.1243/095440602761609434
14. Panovko, M. Y., "Numerical modeling an elastohydrodynamic contact of shaped rollers with allowance for misalignment of their axes", ***Journal of Machinery Manufacture & Reliability***, Vol. 38, No. 5, (2009), 460-466. doi: 10.3103/S1052618809050094
15. Liu, X. L., Yang, P., and Yang, P. R., "Analysis of the lubricating mechanism for tilting rollers in rolling bearings", ***Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology***, Vol. 225, No. 11, (2011), 1059-1070. doi: 10.1177/1350650111413839
16. Liu, X. L., Li, S. Y., Yang, P., and Yang, P. R., "On the lubricating mechanism of roller skew in cylindrical roller bearings", ***Tribology Transactions***, Vol. 56, No. 6, (2013), 929-942. doi: 10.1080/10402004.2013.812760

17. Wu, Z., Xu, Y., and Liu, K., "The analysis on grease lubrication at two tapered bodies contact considering surface roughness", ***Forschung im Ingenieurwesen***, Vol. 83, No. 3, (2019), 339-350. doi: 10.1007/s10010-019-00350-9

18. Palmgren, A., "Ball and Roller Bearing Engineering", 3rd Edition, SKF Industries Inc., (1959).

19. Lundberg, G., "Elastische Berührung Zweier Halbräume", ***Forschung Auf Dem Gebiet Des Ingenieurwesens***, Vol. 10, No. 5, (1939), 201-211. doi: 10.1007/BF02584950

20. Shi, X. and Wang, L. Q., "TEHL analysis of aero-engine mainshaft roller bearing based on quasi-dynamics", ***Journal of Mechanical Engineering***, Vol. 52, No. 3, (2016), 86-92. doi: 10.3901/jme.2016.03.086

21. Huang, C., Wen, S., and Huang, P., "Multilevel solution of the elastohydrodynamic lubrication of concentrated contacts in spiroid gears", ***Journal of Tribology***, Vol. 115, No. 3, (1993), 481-486. doi: 10.1115/1.2921663.

22. Liu, S., Wang, Q., and Liu, G., "A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses", ***Wear***, Vol. 243, No. 1-2, (2000), 101-111. doi: 10.1016/S0043-1648(00)00427-0

Persian Abstract

چکیده

خیز محور یکی از عوامل مهمی است که باعث تخریب حالت روان‌کاری گریس در یاتاقان‌های غلتکی مخروطی (TRB) می‌شود. بر این اساس، بنا بر فرضیه‌ی TRB، بارها و جابه‌جایی‌های خطی و پیچشی ایجاد شده در بخش‌هایی از یاتاقان تحت بارهای ترکیبی و پیچش محور در اثر بار حامل، تحلیل شد. سپس، ویژگی‌های لایه‌ی روان‌کار به منظور یافتن اثرات منفی خیز محور بر روی مشخصه‌های روان‌کار در کل یاتاقان بررسی شد. نتایج نشان داد که تاثیر خیز محور بر روی توزیع فشار در نبود بار شعاعی بیشتر از حالتی است که بار شعاعی بر غلتک وارد می‌شود. ناهم‌ترازی زاویه‌ای غلتک، معمولاً به دلیل خیز محور است. در نتیجه، خیز باعث نامنظم شدن شکل لایه‌ی روان‌کار و ایجاد پروفیل فشار در اتصالات TRB می‌گردد و فشار نقطه‌ای چشم‌گیر ایجاد می‌کند که منجر به بروز پدیده‌ی گلویی شدن در انتهای غلتک می‌شود.

International Journal of Engineering

Journal Homepage: www.ije.ir

A New Developed Model to Determine Waste Dump Site Selection in Open Pit Mines: An Approach to Minimize Haul Road Construction Cost

A. Hajarian, M. Osanloo*

Department of Mining and Metallurgical Engineering, Amirkabir University of Technology, Tehran, Iran

PAPER INFO

Paper history:
Received 15 Febrauray 2020
Received in revised form 09 June 2020
Accepted 11 June 2020

Keywords:
Mine Haul Road
Waste Dump
Earthwork Optimization
Linear Programming

ABSTRACT

Today, during the life of an open pit mine, million tons of materials, including waste and ore, are displaced by truck fleets. In the case of a shallow ore deposit, which is located up to 300 meters to the ground surface, depending on preliminary equipment size and capacity, it will take three to five years to remove overburden and waste rocks to expose the ore body. In that period, the main waste dump site will be used as a disposal of waste dump. Apart from considering the characteristics of the waste dump location such as geological and geotechnical properties, the major factors influencing the hauling process are topography, hauling length and construction cost of the haul road. Truck transportation cost depending on the circumstances comprises 45 to 60% of the cost of mining of one tonne ore. Thus, well site selection of waste dump in coordination with the main haul road path confidently leads to a significant saving of economic resources. In this research, while identifying the effective factors in selecting the waste dump sites, a linear mathematical model is developed to find a suitable site for waste dump disposal considering minimizing haul road construction cost.

***doi**: 10.5829/ije.2020.33.07a.29*

1. INTRODUCTION

Waste dump site selection is of significant importance due to economic, technical, and environmental concerns. Environmental restrictions/law or regulations and also the location of the mine exit point will restrict the site path for road construction ending to waste dump. For example, if mine pit has two exit points for carrying materials, the dump location should be in a balanced position to both of them. If it has only one road exit point, then according to mine expansion direction, it is possible to consider other exit points. In this case, this can be viewed as a problem of the allocation of facilities [1]. The location site in this regard needs to be substructural resistant while respecting technical and economic issues such as proximity to the pit. The road properties factor, such as distance, is of particular importance between others, like geotechnical characteristics, final pit limit, and landform, due to their long-term and indirect impact on the productivity of mine fleet. Traveling cycle time of mine fleet is undeviatingly linked to traversing distance. According to the typical classification of mine haul road, main hauling distances are from pit extraction face toward (average the first five years), crusher, processing plant, and mine facilities [2]. Figure 1 displays a schematic path length of the truck's trip through its main directions inside and outside of the pit.

Customary, one of the main places for carrying materials after extraction from the pit, is the waste dump. Therefore, the shorter the length of the route, leads to a reduction in transportation time and relevant factors such as fuel consumption, maintenance cost, as well as the productivity of machinery increases. To have an idea about the main travel route overpass by trucks, they categorized in Table 1, according to the beginning and ending locations. Defined periods are merely corresponding to hauling distance; we ignored other periods regarding the truck cycle.

Looking for more efficiency in mining operations has many aspects. One aspect is the hauling of the rock/overburden fleet in the shortest period toward destinations. Moreover, transportation costs are

*Corresponding Author Institutional Email: *osanloo@aut.ac.ir* (M. Osanloo)

Please cite this article as: A. Hajarian, M. Osanloo, A New Developed Model to Determine Waste Dump Site Selection in Open Pit Mines: An Approach to Minimize Haul Road Construction Cost, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1413-1422

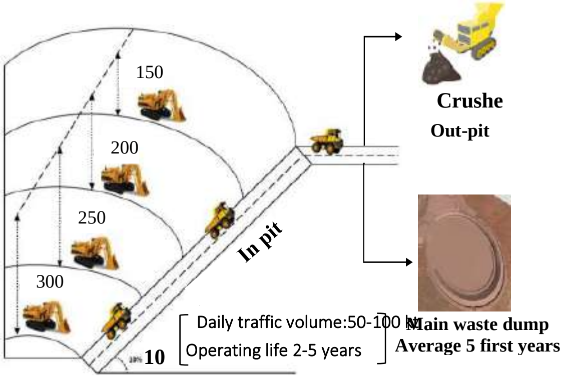


Figure 1. Schematic main places of material handling in open pit mines

TABLE 1. Main hauling travel route of mine fleet

Location	Definition	Corresponding time definition
Truck entrance location	Distance from truck entrance location to the extraction site	Time elapsed from the truck entrance point to extraction site
Truck enter elevation change	Distance from ramp to next level	Time elapsed to reach the next level
Truck loading location	Distance from the extraction point to the loading point	Time elapsed from extraction point to loading point
Exit point	Distance from the loading point to the pit exit point	Time from loading location to exit point location
Processing plant location	Distance from exit point to processing plant	Time from the exit point to reach the processing plant
Mine facility location	Distance to the mine facility	Time to reach to mine facility
Waste dump location	Distance from the extraction to waste dump	Time from extraction point to reach waste dump

approximately 45 up to 60% of mining cost based on Equation (1). It is apparent that the less hauling cost, the less mining cost. One of the most visited places is the main waste dump; therefore, connecting the haul road should consider location and subsequent active factors. The main steps of waste dump site allocation should follow the diagram in Figure 2 and Table 2.

$$C_M = \frac{C_D+C_B+C_L+C_H}{Pr} \quad (1)$$

where:

C_M: Cost of mining ($/ton)
C_D: Cost of drilling ($/h)
C_B: Cost of blasting($/h)
C_H: Cost of hauling ($/h)
Pr: Production rate(ton/h)
CL: Cost of loading ($/h)

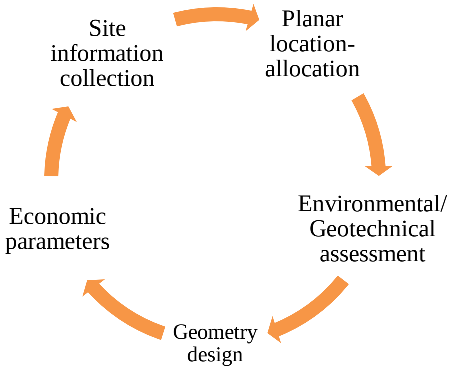

Site information collection
Planar location-allocation
Environmental/ Geotechnical assessment
Geometry design
Economic parameters

Figure 2. Main steps of mine waste dump design

TABLE 2. Main stages of mine waste dump design

Stage	Description
Site information	Vehicle type, Traffic volume, Haulage unit cost, Road life, Construction material available, Waste volume
Planar-location-allocation	Considering the location of the pit exit point, Minimum distance, Minimum cut and fill cost of connecting road, proximity to a waterbody, Limits of pit
Environmental/Geotechnical assessment	Layer works material strengths, Mechanical quality parameters, Environmental restrictions (tree/vegetation), Seepage, Flood safety
Geometry design	Waste dump Capacity, Repose Angle, Shape
Economic parameters	Capital cost, Operational cost

Nowadays, the selection of a preferred waste disposal site is based on multiattribute decision making (MADM) methods. However, very little decentralized research has been done on the selection of waste dumps location using mathematical methods. Summarizing the above points, well location selection of waste dumps in alignment with the main road construction cost confidently contributes to significant economic resource savings during mine planning stage. Therefore, posing a mathematical method to determine the right place, regardless of qualitative methods, is at the highest priority in this stage.

2. BACKGROUND HISTORY

Optimization of target route from extraction point inside the pit to any facility location, waste dump, and processing plant should consider the following factors:

1. Location of other facilities relative to each other 2. Minimum earthwork moving 3. Environmental, geometry, stability control, constraints 4. Fixed cost such as a) building bridges b) tunnels (in case of need) and c) path/road repair and maintenance. Depending on types of mines, the cost of haul road construction varies from mine to mine. The majority costs associated with road building are including 1. Pre-road construction preparation (sub-grade, sub-base, base placement and preparation, berm placement and ditching), 2. Preparation of raw materials [3] which is the excavation of soil from the cut or borrow part and haul to fills or waste dump and compact to shape the ground. As a result of these operations, imposed costs arise. The first model of earthwork allocation was developed based on previous model by considering accommodation setup cost of the external source of material and landfill. In the proposed model, the costs were considered constant. Further research was carried out by Easa [4] for linear programming and quadratic programming. Son et al. [5] presented their achievements for the period of 1990 to 2005. Horizontal alignment [6-11] and environmental consideration [12-17] are other aspects of this subject. During the recent decade, some researchers have developed models in rock waste dump management, aiming to reduce the cost associated with waste rock haulage from the pit toward proposed destinations [18-20]. Based on previous studies, various quantitative and qualitative factors are involved in the selection of mine waste dump locations (see Table 3). Recommended underlined parameters need an adjustment to match modern mining activity and minimize total cost; thus, a new column added to carry out this task. Also, multi-objective papers in other fields based upon mathematical models or MADM studied this problem. MADM studies main goal is to select a qualified place among several pre-defined locations (see Table 4).

All the above studies disregard the earthwork costs are only base on qualitative parameters. In this regard, some researchers focused on scheduling waste dumping sites (see Table 5). As can be seen in Table 5, more than 88% of the studies have been formulated using the MIP method. This is due to the nature of the type of problem (removing or placing a block). Besides, some researchers combined mine planning somehow into waste dumping site selection [28-34]. Their concept of waste dump management is base on stockpiling such that to allocate low-grade material to appropriate stockpiles. In the

TABLE3. Effective factors in waste dump site selection

Main criterion	Sub criteria	To match modern mining and minimize cost
Topographic conditions	The shape of the ground, Capacity, Hauling distance	Combine with earthwork management and pit expansion direction
Hydrology and weather conditions	Precipitation amount, Wind speed and direction ،Acid Mine Drainage, Regional water regime, Quality of surface water, Downstream conditions	
Geology and geomechanical conditions	Subgrade condition, Porosity and seepage, Fault, Waste material strength, Earthquake /Blast vibration	Interrelation with layer works material strengths of the road path
Technical aspects	Mining method, Ultimate pit limit, haulage system, Volume of waste material	
Environmental aspects	Physical and chemical properties of waste materials, Reclamation	Sustainable development issues, Future land use
Economic parameters	Capital cost/Operational cost	Earthwork cost, Tunnel and bridge cost

TABLE 4. History of mine waste dump sites selection [21-27]

Author	Article	Year	MADM context	Method
Osanloo	Factors Affecting the Selection of Site for Arrangement of Pit Rock-Dumps [21]	2003	Pit rock dump site selection	SAW
Mensah	Integrating Global Positioning Systems and Geographic Information Systems in Mine Waste disposal:[22]	2007	Waste dump site selection	GIS
Hekmat	New approach for selection of waste dump sites in open pit mines [23]	2008	Waste dump site selection	SAW, TOPSIS, AHP
Yazadni-Chamzini	Waste dump site selection by using fuzzy vikor [24]	2012	Waste dump site selection	Vikor
Suleman	Selecting Suitable Sites for Mine Waste Dumps Using GIS Techniques [25]	2017	Waste dump site selection	GIS
Oggeri	Overburden management in open pits [26]	2019		Multi-disciplinary
Fazeli, Osanloo	Mine Facility Location Selection in Open-Pit Mines Based on a New Multistep-Procedure [27]	2014	Disposal site selection	Envirinmental impact assessment

current study, unlike existing methods that rely on expert opinion, which-firmly fixed on the specified location, a base MIP model is formulated to minimize hauling cost during the waste dumping period. It can be a tool for experts to have an evaluation of road construction costs before or after choosing any places for waste dumping.

3. METHODOLOGY

One of the main components of choosing a dump site location is the connecting road beginning from the pit and ending to the entrance of dump site. If the selection of location considering the waste dump is according to the qualitative factors (see Table 3), then the optimal location must contain the shortest distance, but always the shortest route is not the least cost path. It is due to many factors, such as building construction. The haulage route is from the mining point to landfill location through the pit exit point. This road must obey vertical alignment in such a way that road profile fit to the ground profile concerning grade constraints. The major problem is to detect an area outside the pit with appropriate size to encompass waste from mining blocks for the specific period, such that to minimize associated haulage distance, cost of building, and preparation cost. In this situation, it is an excellent strategy to use waste material in connecting road path construction as a source of filling material in case of possibility. Excavation of soil from the cut or borrow part and haul to fills or waste dump and compact to shape the ground imposes a cost which is called earthwork cost. The proposed model should consider the earthwork cost model while minimizing distance. The main steps of the methodology are as follows: a) Input: Highlighting the candidate route using existing techniques such as satellite images or photogrammetry (Figure 3a), b) Process: In the first step; 3D blocking the path with a safety margin and defining forbidden area (Figure 3b) (Natural protected areas, Location of buildings, Plant and crusher location and final pit limit), next step; applying model, c) Output: Find a suitable location for waste dump according to the capacity required and optimize haul road construction cost and length.

TABLE 5. Literature survey of mathematical model for scheduling of mine waste dumping site [18-20, 26-32]

Author	Article	Year	Research feature	Method
M. Kumral	Selection of waste dump sites using a tabu search [28]	2008	minimization of dump transportation costs	MIP tabu search
Yu Li	Waste rock dumping optimisation using (MIP) [29]	2013	Optimizing dumping plans	MIP
Yu Li	Optimisation of waste rock placement using (MIP) [18]	2014	Optimizing dumping plans	MIP
Zhao Fu	A New Tool for Optimisation of Mine Waste Management in Potential Acid Forming Conditions [30]	2015	planning of mine waste rock movement	MIP
Jorge Puell	Methodology for a dump design optimization in large-scale open pit mines [31]	2017	Optimizing dumping plans	MIP
Yu Li	Optimising the long-term mine waste management and truck schedule in a large-scale open pit mine [20]	2017	Optimizing dumping plans	MIP
Yuksel Asli	A landfill based approach to surface mine design [19]	2018	Combining mine scheduling with waste dump filling	MIP
M. Adrien	A stochastic optimization method with in-pit waste and tailings disposal for open pit life-of-mine production planning [32]	2018	Combining production scheduling with waste dump managing	Two-stage stochastic MIP
Claudio Oggeri	Overburden management in open pits [26]	2019	Waste dump site selection	Multi-disciplinary

3. 1. Proposed Model

To complete the mathematical model of waste dump site selection, incorporating the earthwork cost model into the hauling cost model must be considered. The main steps of road design can be broken into three principal components: a) Horizontal alignment, which is a trajectory from a satellite's eye view, and using surveying that can introduce candidate routes as input for optimization, b) Vertical alignment, which is a profile of curve from beginning to the ending point of the road. It fits road profile to the ground profile by respecting to terrain grade constraint. c) Earthwork activity which moves blocks into/out of the terrain to determine a smooth surface.

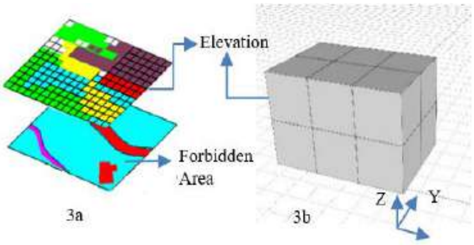


Figure 3. a) Digital elevation layout information of ground b) Blocks layout including information

Mathematical modeling framework begins by applying blocks into connection road from exit point of the pit to the entrance of waste dump. Depend on block position relative to terrain, they classify into the cut and fill blocks. The decision variable $T_{b,b'}$ is the tonnage to be cut from block b and move to block b'. For each b ϵ cut∪ $fill$ the amount of required change in the volume is computed: If this change is negative, then it would be considered as a cut and should be removed, and in case of positive, it is considered as fill. For each pair of b, b' ϵ cut∪ $fill$ (b $\neq$ b'), $D_{b,b'}$ parameter is defined as the distance between the middle point of two blocks. Other parts called waste dump and borrow pit (waste blocks inside the pit) to dump or supply material, are required to be introduced in this problem with the sign of Ш and Ɓ to ensure that there is at least one pit and one waste dump location with substantial capacity. Partial transfer of material from the pit to a block of the road can be shown with the variable $T_{b,b'}$ where bϵƁ and $b'\epsilon B^{+}$.

Similarly, transfer a part of the material from the road part to the waste dump or fill section can also be shown with the variable $T_{b,b'}$, bϵB^{-} and $b'(B^{+} \cup \text{Ш})$. Movement of materials other than these two places is prohibited. Usually, the cost of moving materials from the pit is higher than the cost of moving materials from the road section. Since payment depends on the amount hauled tonnage per distance, this cost factor is neglected in the model. The primary model is to minimize total distance and movement of all material from mining block to nominated 3D block domain considering for waste dump location and keeping away the proximity to water bodies (Equation (2)). It also must be noted to create a logical path which, means those included blocks must be adjacent.

$$Min\colon \sum_{t\in T}\sum_{b\in(B^{-}\cup \text{Ɓ})}\sum_{b'\in(B^{+}\cup \text{Ш})}\{(D_{b,b'}\times T_{b,b'}^{t})\times a_{b,b'}\}/(1+r)^{t} \qquad (2)$$

where:

$D_{b,b'}$	Flat distance between the middle point of two blocks;
$T_{b,b'}^{t}$	Volume to be cut from block b and move to block b′ during period t
$a_{b,b'}$(binary variable)	1 if block b is adjacent to block b′ and have directed path, 0 otherwise
r	Discount rate
B^{-}	Set of cut blocks
Ɓ	Set of pit blocks
B^{+}	Set of fill blocks
Ш	Set of waste dump blocks
T	Set of time period
b	Block model index

Allocation of material in the earthwork problem must be logical. If the unit of the material belongs to road section, then the place of transfer must be either fill sections or the place of the waste dump:

if $b\epsilon B^{-}$ (cut section) then $b'\epsilon B^{+}\cup \text{Ш}$

Similarly, if the unit of the material belongs to mine pit, then the place of transport can be either road fill sections or waste dump:

if $b\epsilon$Ɓ then $b'\epsilon B^{+}\cup \text{Ш}$

If the unit of the material belongs to the waste dump, then the target location is empty, or there is no transferring location:

if $b\epsilon$Ш then $b'\epsilon\phi$

They eliminate the pair of indices b and that are not logical moves. For example, transfer from the road block to the pit is unacceptable. The above definition will be provided mathematically during the text.

a) Location Constraint

To use mine pit and waste dump option, they should have been previously created with sufficient slack. When a cube block extracted, it becomes a square frustum when dumping on the ground (Figure 4). For the convenience of computation, it considered a pyramid. Thus, let C_w denotes the capacity of blocks in the waste domain.

$$C_w = \frac{1}{3}\times h_w \times S_b \qquad (3)$$

$$LW_e = \frac{2h_w}{tan\,(\alpha_s)} \qquad (4)$$

$$S_b \geq 0 \; and \; integer \qquad (5)$$

where:

C_w	Maximum capacity of waste dump section
h_w	Height of waste dump
S_b	Number of blocks existing in the length of the waste dump location

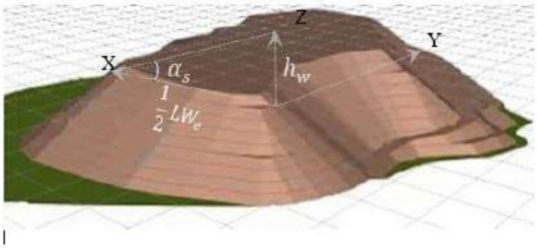


Figure 4. Terminology of location constraint in a mine dump

α_s	Angle of tailing dump
LW_e	Equivalent length of waste dump
Δx_b	Dimension of a block

Remark 1: Environmental regulations determine the height of the pyramid.
Remark 2: Only one landfill entry point is considered.
Remark 3: An aggregation of waste dump is considered if more than one exists.

$$C_w = \sum_{i\in \text{Ш}} C_i,\ C_b = \sum_{i\in \text{Ɓ}} C_i,\ T_{cut} = \sum_{i\in B^-} T_i,\ T_{fill} = \sum_{i\in B^+} T_i \tag{6}$$

$$T_{cut} \le C_w \tag{7}$$

$$T_{fill} \le C_b \tag{8}$$

where:

T_{cut}	Total amount of cut tonnage
T_{fill}	Total amount of fill tonnage
S_b	As described before
α_s	Angle of tailing dump
R	Equivalent length of waste dump
$C_w, \text{Ш}, \text{Ɓ}$	As described before
Δx_b	Dimension of a block in x-direction
C_b	Maximum capacity of waste blocks in pit

b) Capacity Constraint

The following equations enforce the maximum capacity for the pit, and the waste dump is not over-utilized.

$$\sum_{t\in T} \sum_{b'\epsilon B^+\cup \text{Ш}} T_{b,b'}^t \le C_b \qquad \forall b \in \text{Ɓ} | g_b \le g_o \tag{9}$$

$$\gamma_s^E \times \gamma_s^F \sum_{t\in T} \sum_{b\epsilon B^-\cup \text{Ɓ}|g_b\le g_o} T_{b,b'}^t \le C_w \qquad \forall b' \in \text{Ш} \tag{10}$$

where:

γ_s^{E}	Expansion factor of material in excavation
γ_s^F	Compaction factor of material in filling
g_b	Grade of mining block
g_o	Cut-off grade
$T_{b,b'}^t, b, B^-, \text{Ɓ}, C_w, \text{Ш}, C_b$	As described before

c) Material Balance

Material hauled from or into each part must be equal to a defined amount of cut or fill.

$$s_b^t = \sum_{t\in T} \sum_{b'\in B^+\cup \text{Ш}} T_{b,b'}^t \qquad \forall b \in B^- \cup \text{Ɓ} | g_b \le g_o \tag{11}$$

$$\gamma_s^E \times \gamma_s^F \times \sum_{t\in T} \sum_{b\in B^-\cup \text{Ɓ}|g_b\le g_o} T_{b,b'}^t = d_b^t \ \forall b' \in B^+ \cup \text{Ш} \tag{12}$$

where:

s_b^t	Amount of cut in each block (supply) in period t;
d_b^t	Amount of fill in each block (demand) in period t;
$T, B^+, \text{Ш}, B^-$	As described before
$\text{Ɓ}, g_b, g_o, \gamma_s^E, \gamma_s^F$	As described before

d) Block Constraints

If a waste block is extracted from mine pit or road section, then it must be hauled to a single adjacent fill block. Similarly, each fill block can receive material from one single cut block.

$$\sum_{b\in B^-\cup \text{Ɓ}|g_b\le g_o} (a_{b,b'}) = 1 \qquad \forall b' \in B^+ \tag{13}$$

$$\sum_{b'\in (B^+\cup \text{Ш})} (a_{b,b'}) = 1 \qquad \forall b \in B^- \tag{14}$$

where:

$a_{b,b'}$	As described before
$T, B^+, \text{Ш}, B^-, \text{Ɓ}, g_b, g_o$	As described before

Movement of waste material into a mine pit or out of a waste dump site is not permitted.

$$a_{b,b'} = 0 \qquad \forall b \in \text{Ɓ}, b' \in \text{Ɓ} \tag{15}$$

$$a_{b,b'} = 0 \qquad \forall b \in \text{Ш}, b' \in \text{Ш} \tag{16}$$

e) Access Constraints

Overlying blocks must be extracted to access a block in the pit during the time period or earlier time. In the case of the filling block, underlying blocks must be filled during the time period or earlier time.

$$\sum_{w=1}^{t} x_b^w \le \sum_{w=1}^{t} x_{\hat{b}}^w \qquad \forall t, b \in \text{Ɓ}, \hat{b} \tag{17}$$

where:

w	Time period
$\hat{b}$	Overlying block index (1,…,9)

$b, \text{Ɓ}$	As described before
$x_b^{t(w)}$(binary variable)	1 if block b is extracted at time t, 0 otherwise

f) Vertical Cut-Fill Precedence

Vertical precedence assigning of cut blocks to fill blocks must be considered to make the resource allocation feasible. If the model assigns material according to Figure 5, then to cover the space of the block b_1 using block b material, we must wait until $\mathrm{b'}_1$ is filled using material from $\mathrm{b'}$. Otherwise, the assignment is violated. block b must land out and set aside, extract block $\mathrm{b'}$ and haul to $\mathrm{b'}_1$ location Later, pick up material of b and move to $\mathrm{b'}_1$.

Top-down cutting and the bottom-up filling equations are as follow:

$$a_{a,d} + a_{b,c} \leq 1 \qquad \forall (a,b,c,d) \in \psi \qquad (18)$$

where:

$$\psi = \{(a,b,c,d) | (a,b) \in B_{p-q}^{-}, (c,d) \in B_{p-q}^{+}, [(z_a > z_b) \wedge (z_c < z_d)] \vee [(z_a < z_b) \wedge (z_c > z_d)]\} \qquad \forall (a,b,c,d) \in \psi \qquad (19)$$

where:

$$B_{p-q}^{-} = \{(b,b') \in B^{-} \cup \text{Ш}\},$$
$$B_{p-q}^{+} = \{(b,b') \in B^{+} \cup \text{Ɓ}\} \qquad (20)$$

where:

$a_{a,d}$	As described before
z_a	Elevation of block a
z_b	Elevation of block b
ψ	Set of blocks with specific precedence
$T, B^{+}, \text{Ш}, B^{-}$	As described before
$\text{Ɓ}, g_b, g_o$	As described before

g) Proximity to Waterbody

A boundary is proposed in a set of ϕ to consider not passing through a forbidden area like a potential waterbody zone.

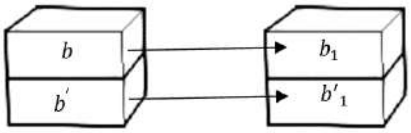


Figure 5. A possible model of allocating block to the locations

$$\sum_{b\in B^{-}\cup \text{Ɓ}} a_{b,b'} + \sum_{b'\in \phi} a_{b,b'} = 0 \qquad \forall b' \in \phi, \quad b \in B^{-} \cup \text{Ɓ} \qquad (21)$$

where:

ϕ	Set of blocks in the forbidden area (like water body)
$a_{b,b'}, b, B^{-}, \text{Ɓ}$	As described before

Equation (22) ensures that if excavated block is belonging to cut sections and destination is belong to the forbidden area, no volume of material is hauled.

4. NUMERICAL ANALYSIS

A hypothetical block model representing terrain complexities and the same 3D blocks of dimension for pit and 3D blocks for cut and fill section were defined to demonstrate the efficiency of the model. This combination layout depicted in Figure 6. Details are summarized in Table 6. Other parameters like compaction and expansion factor, cut-off grade, and the rest were considered in the normal range within the block model. Also, different block sizes applied to the road path and waste dump location section. The blocks in the pit must be removed and haul to waste dump during their scheduled time, according to Table 7.

Referring to given equations, those blocks with the grade less than cut-off grade sent to dump or filling position. Besides, cut blocks located in the proposed connecting road must add up to this set, with the above

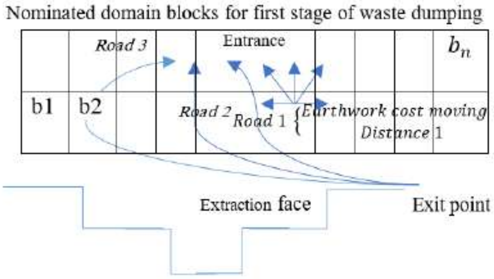


Figure 6. Conceptual layout of nominated domain blocks for waste dumping showing the different connecting path

TABLE 6. Parameter values for study

Parameter	Value
Number of periods (years)	5
Discount rate(prercent)	10
Maximum road grade (percent)	13
Minimum road grade (percent)	10

assumptions, the MIP model designed for lingo software. The model was solved using a PC with the specification of 3.2 GHz CPU with 16 GB of RAM.

5. RESULTS AND DISCUSSION

The solution results of decision variables shape the schedule of hauling blocks, including waste in the pit and those in the proposed connection road toward fill or waste location. By this method, it is guaranteed to use the waste material of pit blocks in construction road as a filling material. The objective function also promises minimum cost (hauled material per distance) during the life-of-mine. All block contains elevation data. The terrain profile in line with the proposed connecting road is shown in Figure 7.

The resulting accuracy intensely depends on block size. A Different block size (20, 20, 1) also was planned to examine the model. Different planning in block size leads to distinct cut and fill volumes and hence the other results. Table 8 compares the results of (5, 20, 1) and (20, 10, 1) block size. Figure 8 shows a profile section for cut and fill blocks.

Density for all blocks was considered hemogenous, but it can be defined in the model as a variable. Truck capacity has a remarkable effect on the result. More capacity leads to more hauling tonnage, but further maintenance and fuel costs must be into consideration to adjust for fleet selection. Here, we consider fleet capacity the same, which means tonnage per distance has no irregular rising and falling.

TABLE 7. Number of waste blocks in mining schedule

Time	Number of blocks
1	320
2	289
3	405
4	310
5	280
Total	1604

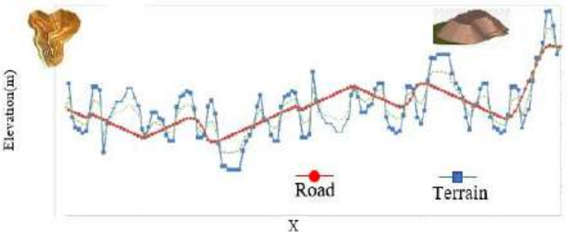


Figure 7. Profiles of the ground and proposed road surface

The result in column Distance×Tonnage shows a more compacted block size, improves the quality of the solution, but these scores do not have a linear relation to block size. It can be concluded that the smaller the dimensions of the blocks, the higher the accuracy of the path determination, which is due to the increase of grid resolution. The reduction of costs is also due to the increase in the resolution of the grid. In both terms of length and cost, grid size-reduction gives us a more accurate evaluation. Otherwise, the whole route and the location of the route will not change. However, natural physical features of an area and terrain have anonymous effects on the percentage of change. To deal with smaller block size, enough memory, and better configuration is also needed. To achieve a more accurate solution, the assignment of blocks must obey the realistic configuration. Removal of significant obstacles before the movement of a block to the destination is necessary. An obstacle is those blocks in a large area like a topographical feature or lake. Consider cut block 4 in Figure 9; to access it, fill block 3 needs to be removed first. Only in rare cases, this occurs in mines because of the proximity of the site of waste dump to the pit. However, this should not be overlooked. This issue can be handled before optimization by modifying such considerable barriers or considering it in the model. To extend the linear program constraints, we can incorporate time-steps into the removal stages.

The proposed model needs additional variables with temporal properties to represent the logical movement of

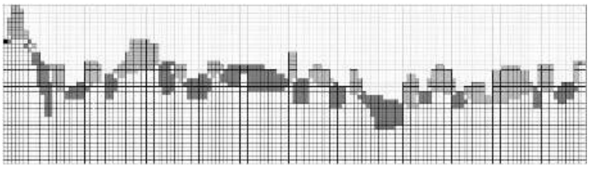


Figure 8. A profile section for cut and fill blocks ($\Delta x = 50, \Delta y = 20, \Delta z = 1$. Cuts are light grey, and fills are dark grey)

TABLE 8. Different block size analysis

Block Size	Number of Blocks (Pit+Road)	Distance (km)	Distance×Tonnage ($)
(50, 20, 1)	2554	15.870	128000
(20, 20, 1)	3720	12.940	72000

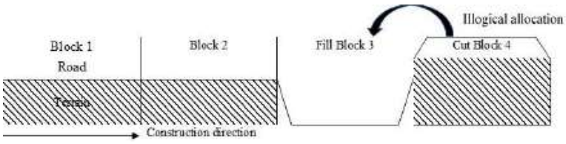


Figure 9. Unrealistic removal solution of obstacle

blocks via the access road. Such blocks without access road cannot be operated on in this situation before at least the obstacle is eliminated. The time-step idea can schedule the delay and precedence the removal of blocks.

6. CONCLUSION

The purpose of this paper is to find the optimal location of the main waste dump in such a way that balances the trade-off among hauling cost, connecting road construction cost and environmental impact. In this research, we investigate the factors that influence waste dump location and review the past activities of other researchers. It focuses on incorporating earthwork moving plans to locate a waste dump location. Unlike previous multi-objective models that were only concentrating on rock dump placement optimization and management, the current model finding a more realistic waste dump position. According to Table 4, most researches on waste dump site selection are based on the screening or ranking methods. First, the potential sites, alternative, and their attributes such as dump capacity and haulage distance are defined. Next, by using a conventional way, the qualitative attribute converts to quantitative. At the last step by ranking alternatives, the best one that fits on the applied method is chosen. The current model finds sufficient capacity using Equation 3 for waste dumping and determine minimum haulage distance road. It tries to use mine waste dump blocks as a filling material, and by scheduling an assignment of cut and fill parts, reduce the costs. Since transportation costs are approximately 45 to 60% of mining cost, it also addresses the reader that haul road construction cost is a good point of beginning for waste dump site selection, and other factors could follow it. To clear up the subject, we consider the given approach in most articles on the selection of the waste dump location in open pit mines. First, they are using MADM methods that obtain the overall preference value for each alternative, and then the best alternative is selected. The preparation expenditure that mostly includes the construction of access roads is per dollar. It only considers the length of the direct route per kilometer. The restrictions such as forbidden area or topographical conditions not considered. There is no view on the road construction operation section. Therefore, the applied scores are not accurate enough. As discussed, the cost of transportation plays an essential role in the mining economy, so choosing the location of the waste dump by mistake leads to loss of capital expenditures. It is necessary to use trucks to carry out material in real work. To have a real schedule, integration of earthwork planning and truck selection also seems to be very necessary. For future works, combining of time scheduling and capacity constraints of trucks is necessary. In part 4, by solving a numerical example, we also showed the effects of block size on the results but discussed to have a better sight; more different analysis is needed. It noted to overcome the restriction movement of blocks to remove untrue allocation; time-steps approaches need to incorporate into the model. A constraint is added to the model not to pass through blocks to consider environmental restrictions, but more investigation must consider to handle the real-world problem. If, for example, we only consider not passing through a woodland area, but near it, most likely, continuity of animal life is put on danger. That is why to consider this restriction carefully. To improve this topic for future, sustainable development and future land use issues in the mining area in addition to the processing plant location and their impacts on ex-pit road location enriches this research.

7. REFERENCES

1. Love, R.F., "Facilities location: models & methods", ***Publications in Operations Research Series***, Vol. 7, (1988), New York: North-Holland. https://nla.gov.au/nla.cat-vn436294

2. Song, S., E. Marks, N. Pradhananga, "Impact Variables of Dump Truck Cycle Time for Heavy Excavation Construction Projects", ***Journal of Construction Engineering and Project Management***, Vol. 7, (2017). 11-18**.** doi:10.6106/JCEPM.2017.7.2.011

3. Regensburg, B., D. Tannant," Guidelines for Mine Haul Road Design", University of British Columbia Library, (2001). doi:10.14288/1.0102562

4. Easa, S.M., "Earthwork Allocations with Linear Unit Costs", ***Journal of Construction Engineering and Manageme***, Vol. 114, (1988), 641-655. doi:10.1061/(ASCE)0733-9364(1988)114:4(641)

5. Son, J., K.G. Mattila, and D.S. Myers, "Determination of Haul Distance and Direction in Mass Excavation", ***Journal of Construction Engineering and Management***, Vol. 131, (2005). 302-309. doi:10.1061/(ASCE)0733-9364(2005)131:3(302)

6. Jong, J.-C., M. K. Jha, and P. Schonfeld, "Preliminary highway design with genetic algorithms and geographic information systems", ***Computer-Aided Civil and Infrastructure Engineering***, Vol. 15, (2000), 261-271. 10.1111/0885-9507.00190

7. Easa, S., A. Mehmood, "Optimizing Design of Highway Horizontal Alignments", ***Computer-Aided Civil and Infrastructure Engineering***, Vol. 23, (2008), 560-573. doi:10.1111/j.1467-8667.2008.00560.x

8. Lee, Y., Y.-R. Tsou, and H.-L. Liu, "Optimization Method for Highway Horizontal Alignment Design", ***Journal of Transportation Engineering***, (2009), 217-224. doi:10.1061/(ASCE)0733-947X(2009)135:4(217)

9. Mondal, S., Y. Lucet, and W. Hare, "Optimizing horizontal alignment of roads in a specified corridor", ***Computers & Operations Research***, Vol. 64, (2015), 130-138. https://doi.org/10.1016/j.cor.2015.05.018

10. Pushak, Y., W. Hare, and Y. Lucet, "Multiple-path selection for new highway alignments using discrete algorithms", ***European Journal of Operational Research***, Vol. 248, (2016), 415-427. https://doi.org/10.1016/j.ejor.2015.07.039

11. Casal, G., D. Santamarina, and M.E. Vázquez-Méndez, "Optimization of horizontal alignment geometry in road design and reconstruction", ***Transportation Research Part C: Emerging***

Technologies, Vol. 74, (2017), 261-274. https://doi.org/10.1016/j.trc.2016.11.019

12. Kim, B., H. Lee, H. Park, H. Kim, "Framework for Estimating Greenhouse Gas Emissions Due to Asphalt Pavement Construction", ***Journal of Construction Engineering and Management***, Vol. 138, (2012), 1312-1321. doi:10.1061/(ASCE)CO.1943-7862.0000549
13. Hajji, A.M. and M.P. Lewis, "How to estimate green house gas (GHG) emissions from an excavator by using CAT's performance chart", AIP Conference Proceedings, Vol. 1887, (2017), 20047. doi:10.1063/1.5003530
14. Carmichael, D.G., B.J. Bartlett, and A.S. Kaboli, "Surface mining operations, coincident unit cost and emissions", ***International Journal of Mining, Reclamation and Environment***, Vol. 28, (2013), 47-65. doi:10.1080/17480930.2013.772699
15. Norgate, T., N. Haque, "Energy and greenhouse gas impacts of mining and mineral processing operations", ***Journal of Cleaner Production***, (2010). Vol. 18, 266-274**.** https://doi.org/10.1016/j.jclepro.2009.09.020
16. Avetisyan, H.G., E. Miller-Hooks, and S. Melanta, "Decision Models to Support Greenhouse Gas Emissions Reduction from Transportation Construction Projects", ***Journal of Construction Engineering and Management***, Vol. 138, (2012), 631-641. doi:10.1061/(ASCE)CO.1943-7862.0000477
17. Lewis, P. and A. Hajji, "Estimating the Economic, Energy, and Environmental Impact of Earthwork Activities", Construction Research Congress 2012. (2012). doi:10.1061/9780784412329.178
18. Li, Y., E. Topal, and D.J. Williams, "Optimisation of waste rock placement using mixed integer programming", ***Mining Technology***, Vol. 123, (2014), 220-229. doi:10.1179/1743286314Y.0000000070
19. Sari, Y.A. and M. Kumral, "A landfill based approach to surface mine design", ***Journal of Central South University***, Vol. 25, (2018), 159-168. doi:10.1007/s11771-018-3726-7
20. Li, Y., E. Topal, and S. Ramazan, "Optimising the long-term mine waste management and truck schedule in a large-scale open pit mine", ***Mining Technology***, Vol. 125, (2016), 35-46. doi:10.1080/14749009.2015.1107343
21. Osanloo, M., M. Ataei, "Factors Affecting the Selection of Site for Arrangement of Pit Rock-Dumps", ***Journal of Mining Science***, Vol. 39, (2003). 148-153. doi:10.1023/B:JOMI.0000008460.62695.44
22. Mensah, F., "Integrating Global Positioning Systems and Geographic Information Systems in Mine Waste disposal: The Case of Goldfields Ghana Limited", University of Mines and Technology, Tarkwa, (2007).
23. Hekmat, A., M. Osanloo, A.M. Shirazi, "New approach for selection of waste dump sites in open pit mines", ***Mining Technology***, Vol. 117, (2008), 24-31. doi:10.1179/174328608X343768
24. Yazadni-Chamzini, A., "Waste dump site selection by using fuzzy vikor", SME Annual Meeting and Exhibit 2012, SME 2012, Meeting Preprints, (2012), 145-152.
25. Suleman, H. and P. Baffoe, "Selecting Suitable Sites for Mine Waste Dumps Using GIS Techniques at Goldfields, Damang Mine", ***Ghana Mining Journal***, Vol. 17, (2017), 9-17. doi:10.4314/gm.v17i1.2
26. Oggeri, C., T. Fenoglio, A. Godio, and R. Vinai, "Overburden management in open pits: options and limits in large limestone quarries", ***International Journal of Mining Science and Technology***, Vol. 29, (2019), 217-228. https://doi.org/10.1016/j.ijmst.2018.06.011
27. Fazeli, M. and M. Osanloo, "Mine Facility Location Selection in Open-Pit Mines Based on a New Multistep-Procedure", ***Mine Planning and Equipment Selection***, (2014), 1347-1359. https://link.springer.com/chapter/10.1007/978-3-319-02678-7_129
28. Kumral, M. and R. Dimitrakoponlos, "Selection of waste dump sites using a tabu search algorithm", ***Journal of the Southern African Institute of Mining and Metallurgy***, Vol. 108, (2008), 9-13. https://www.saimm.co.za/Journal/v108n01p009.pdf
29. Li, Y., E. Topal, and D. Williams, "Waste rock dumping optimisation using mixed integer programming (MIP)", ***International Journal of Mining, Reclamation and Environment***, Vol. 27, (2013), 425-436. https://doi.org/10.1080/17480930.2013.794513
30. Fu, Z., Y. Li, E. Topal, D.Williams, "A New Tool for Optimisation of Mine Waste Management in Potential Acid Forming Conditions", Tailings and Mine Waste Management for the 21st Century, (2015). https://espace.library.uq.edu.au/view/UQ:403073
31. Puell Ortiz, J., "Methodology for a dump design optimization in large-scale open pit mines", ***Cogent Engineering***, Vol. 4, (2017). doi:10.1080/23311916.2017.1387955
32. Adrien Rimélé, M., R. Dimitrakopoulos, and M. Gamache, "A stochastic optimization method with in-pit waste and tailings disposal for open pit life-of-mine production planning", ***Resources Policy***, Vol. 57, (2018), 112-121. https://doi.org/10.1016/j.resourpol.2018.02.006
33. Rezakhah, M. and A. Newman, "Open pit mine planning with degradation due to stockpiling", ***Computers & Operations Research***, Vol. 115, (2020). https://doi.org/10.1016/j.cor.2018.11.009
34. Koushavand, B., H. Askari-Nasab, and C.V. Deutsch, "A linear programming model for long-term mine planning in the presence of grade uncertainty and a stockpile", ***International Journal of Mining Science and Technology***, Vol. 24, (2014), 451-459. https://doi.org/10.1016/j.ijmst.2014.05.006

Persian Abstract

چکیده

در طول عمر یک معدن روباز میلیون‌ها تن از مواد اعم از باطله و ماده معدنی توسط ناوگان کامیونی جابجا می‌شود. در مواردی که ماده معدنی در عمق تا سیصد متر از سطح زمین قرار گرفته است، بسته به اندازه و ظرفیت اولیه تجهیزات، بین سه تا پنج سال جهت روباره و باطله برداری نیاز است تا به ماده معدنی دسترسی پیدا شود. جدا از در نظر گرفتن ویژگی‌های محل دفع باطله مانند خصوصیات زمین‌شناسی و ژئوتکنیکی، مهمترین عوامل موثر در قسمت حمل شامل توپوگرافی، طول مسیر و هزینه ساخت مسیر می‌باشد. هزینه حمل با کامیون بسته به شرایط مختلف بین ٤٥ تا ٦٠ درصد از هزینه استخراج یک تن سنگ را به خود اختصاص می‌دهد. لذا مکانیابی مناسب مکان دامپ باطله که در هماهنگی با مسیر جاده معدن باشد در اقتصاد معدن تاثیر زیادی دارد. در این تحقیق ضمن شناسایی عوامل موثر در مکان احداث دامپ باطله، یک مدل ریاضی خطی با هدف پیدا کردن محل مناسب دفع باطله و کمینه‌سازی هزینه ساخت جاده توسعه داده شده است.

IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1423-1430

International Journal of Engineering

Journal Homepage: www.ije.ir

Development of Mathematical Model for Controlling Drilling Parameters with Screw Downhole Motor

M. Dvoynikov, A. Kunshin*, P. Blinov, V. Morozov

Department of Wells Drilling, Saint-Petersburg Mining University, Saint-Petersburg, Russian Federation

PAPER INFO

Paper history:
Received 05 March 2020
Received in revised form 09 June 2020
Accepted 12 June 2020

Keywords:
Drilling Control
Drilling Dynamics
Directional Drilling
Horizontal Well
Bit Shock
Bit Vibration

ABSTRACT

Present article results of study on possibility of increasing the efficiency of drilling directional straight sections of wells using screw downhole motors (SDM) with a combined method of drilling with rotation of drilling string (DS). Goal is to ensure steady-state operation of SDM with simultaneous rotation of DS by reducing the amplitude of oscillations with adjusting the parameters of drilling mode on the basis of mathematical modeling for SDM – DS system. Results of experimental study on determination of extreme distribution of lateral and axial oscillations of SDM frame depending on geometrical parameters of gerotor mechanism and modes ensuring stable operation are presented. Approaches to develop a mathematical model and methodology are conceptually outlined that allow determining the range of self-oscillations for SDM – DS system and boundaries of rotational and translational wave perturbations for a heterogeneous rod with an installed SDM at drilling directional straight sections of well. This mathematical model of SDM – DS system's dynamics makes it possible to predict optimal parameters of directional drilling mode that ensure stable operation of borehole assembly.

doi: *10.5829/ije.2020.33.07a.30*

NOMENCLATURE

z_1, z_2	Number of rotor / stator teeth	H	Well depth (m)
ω_r	Angular rotation velocity of rotor around its own axis (s^{-1})	λ_{Ln}	Speed of rotation oscillation transmission (m/s)
M_{ind}	Indicator moment (kN·m)	$\varphi_n(s_n,t)$	Angular deflection current cross-section column on according part (rpm)
e	Eccentricity (m)	s_n	Current position cross-section (m)
m	Rotor mass (kg)	$u_n(s_n, t)$	Translational movements of current cross-sections of string in corresponding sections (rad)
ω	Angular velocity (s^{-1})	h	Value of translational movement with transfer
D	Stator diameter at tooth cavities (m)	$f_{\tau_{n1}}\left(\frac{\partial\varphi_n}{\partial t}\right)$	Dissipative term described to resist of drilling string rotation on according part
P_p	Pressure difference (Pa)	$f_{\tau_n}\left(\frac{\partial u_n}{\partial t}\right)$	Dissipative members characterizing the resistance of drilling string translational movement
t	Rotor pitch (m)	n_0	Rotation speed for upper end of string (rad/s, rpm)
A	Amplitude (mm)	G_1, G_2, G_3	Modulus of rigidity material according part (N·m^2)
L_a	Level vibration acceleration	E_1, E_2, E_3	Elastic modules of materials in corresponding sections under tension or compression (kg/m^2)
L_v	Level vibration speed	J_1, J_2, J_3	Polar moment of inertia cross-section column on according part (m^4)
$v_0 = 5\cdot10^{-8}$	Backup value vibration speed (m/s)	$M_H(P, n_H)$	Moment of resistance bottom composite rod rotation from side of rock (N·m)
$a_0 = 1\cdot10^{-6}$	Backup value vibration acceleration (m/s^2)	P	Axial load on the end of composite rod lower section (N)
a	Medium–square value vibration acceleration (m/s^2)	F_1, F_2, F_3	cross-sectional area of string in corresponding sections (m^2)
v	Medium–square value vibration speed (m/s)	n_H	Rotation speed bottom part composite rod (rpm /s)

*Corresponding Author's Email: *kunshin.a.a@gmail.com* (Andrey Kunshin)

Please cite this article as: M. Dvoynikov, A. Kunshin, P. Blinov, V. Morozov, Development of Mathematical Model for Controlling Drilling Parameters with Screw Downhole Motor, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1423-1430

n	Frequency rotation (s^{-1}, rpm, Hz)	$\theta=\frac{G_1J_1}{G_2J_2}$, $\theta=\frac{E_1F_1}{E_2F_2}$	Coefficient of moment-force ratio of the first and the second sections during rotation / translational movement
f	Medium geometrical frequency octave filter (dB)	$\varepsilon=\frac{G_2J_2}{G_3J_3}$, $\varepsilon=\frac{E_1F_1}{E_2F_2}$	Coefficient of moment-force ratio of the second and the third sections during rotation / translational movement
M	Resistance moment (kN·m)	μ_1, μ_2, μ_3	Coefficient of dissipation on according part composite rod
n_0	Rotor rotation speed (rpm/s)	k, k_1, k_2	Coefficient spect wave circling indignation on boundary heterogeneous parts composite rod
t	Time (s)	$\Delta M_H=M_H(P,0)-M_H(p,n_0)$	Margin between evict moment end bottom part and nominee moment that end (N·m)
L_1, L_2, L_3	DS length, drill collar length, length of SDM body and navigation and measuring tool (m)	P_b, P_H	Axial loads according top and bottom boundary auto-oscillation (N)
D_{L1}, D_{L2}, D_{L3}	DS, drill collar, SDM body and navigation and measuring tool external diameters (mm)	n_0^*	Rotor speed rotation, when $P_b = P_H$.
d_{L1}, d_{L2}, d_{L3}	DS, drill collar, SDM body and navigation and measuring tool internal diameters (mm)		

1. INTRODUCTION

Increasing hydrocarbon production due to developing offshore fields, as well as additional development of previously drilled areas is only achieved by construction of complex well profiles which trajectories may contain curved or straight sections of long distances [1]. Implementation of such profiles involves the use of a rotary steerable system (RSS), or SDM as a drill bit (DB) drive. The use of expensive RSS [2] is economically unsuitable. Therefore, in most cases (70-80 %) such wells are drilled by using SDM in Russia.

SDM – hydraulic downhole motor of volumetric type, multiple thread working bodies of which are made according to the scheme of the gerotor planetary mechanism, driven by the energy of the drill fluid.

During drilling of extended directional and horizontal sections of wells using volumetric principle engines, part of axial load on the bit is not transmitted due to frictional force arising between walls of well and drilling tool [3].

To ensure required load on the bit a combined drilling method is used in production process. Distinctive feature of this method is in the joint operation of drilling string (DS) and screw down- hole motor (SDM) [4]. In the process of their joint work, torsional, lateral and axial oscillations can occur depending on type of SDM, its energy characteristics and DS, which acts as an elastic unbalanced rod [5].

It should be noted that SDM that is located in lower part of DS has its own beating of the frame, nature of occurrence of which is associated with work of its power section, represented by a planetary reductor. Moreover, frequency, amplitude and direction of the frame beats depend on design of gerotor mechanism, hydraulic component of drilling mud flow, as well as load on the bit [6].

To determine parameters of well drilling mode by a combined method, it is necessary to develop a technique that allows providing forecast and control of stable operation of bottom hole assembly (BHA), based on mathematical modeling of elastic properties of DS stress-strain state and characteristics of SDM [7, 8].

Specifics of the installed SDM in the BHA, power section lead to independence of axial and lateral oscillations characteristics from DS and drill bit.

Rotation of long elastic rod with different rigidity is limited by the wellbore wall with alternating stress-deformation state (SDS) causing the occurrence of oscillations. DB amplitude changes are difficult to determine by math methods. In practice to measure DB vibration a three-position accelerometer is placed into telemetric tool, which allows to control the acceleration of BHA. According passport, for geophysical well logging to avoid damage of column elements, the vibration acceleration should be in the range of 30-45 G. Resulting from the large values of vibration acceleration, it is difficult to accurately determine the load on the bit. Maintaining the required vibration acceleration is possible due to control dynamic of system «SDM-DS».

Oscillations in the DB, SDM and DS in critical range of vibration [9], acceleration leads to tool instability, adversely affect the formation wellbore walls, decreased quality of well trajectory control and increased risk of accidents as a result of decoupling in screw connections damages to BHA elements.

The dynamics of the DS is due to mechanical energy transmitted from the top power drive while the SDM is based on the conversion of the energy of the process fluid stream pumped by the drilling pump units.

Scientists from the beginning of 17th century studied the dynamics of DS while drilling. The method for research the elastic-deformed state of a drill string represented by a one-dimensional core was formulated by Euler and continues to improve [10]. The increment of the potential energy of the drilling tool in the form of a mechanical system and deviations thereof from the balance position are recorded and proved in the form of the theorem by Laplace-Dirichlet. For the considered section of the DS, the method proposed by Leibenzon, which determined the nature of rotation, is known in literature [11]. The definition of the area of stable operation of the DS by the mechanical analogue method was developed by Yunin and Khegai [10].

The great contribution to the design, creation and improvement of gerotor machines, as well as to the research of the working processes of the SDM for drilling and workover, was made by domestic and international scientists [11, 12].

Presently, there is a large amount of information on the research of increasing the operating time of the SDM and on ways to increase the efficiency of the motor [11-14]. Much attention is also paid to search the dynamics of the DS and SDM during drilling of deviated and horizontal wells with geomechanical studies [12, 14-17]. The results of these researches shown that the negative effect of vibration and oscillations on the transmission of axial load on the bit can be reduced due to the operational control of the dynamics of the DS, SDM and DB. Therefore, the main objectives of this work are to increase drilling efficiency and reduce the likelihood of accidents inside the well by controlling the dynamics of the DB-SDM-DS system by optimizing drilling parameters (rotation per minute, mathematical models of axial stress) during drilling deviated and horizontal wells.

To achieve these objectives, following analysis should be conducted:

1. researches of SDM wobbling with different work modes;
2. investigate the oscillation of the DS considering its SDS and drilling parameters;
3. develop technology to regulate torque – power and frequency of DS and SDM.

As a result of the conducted research, a model for regulating and controlling the dynamics of the «SDM-DS» system was developed while drilling directional straight sections of the well, which makes it possible to increase the efficiency of drilling directional straight sections of the well.

2. MATERIALS AND METHODS

Stability of SDM operation is characterized by working mode of power section, in which there is no intensive decrease in rotor rotation frequency with increasing torque on motor shaft.

It is known that axis of rotor rotates around its own axis, and also makes a transferring movement around axis of stator, directed counterclockwise. Moreover, frequency of transferring (planetary) rotation of rotor's axis relative to stator's axis is higher than rotor rotation frequency around its own axis.

Angular rotation velocity of rotor's axis relative to stator's axis, which determines beat frequency of the frame,

$$\omega_n = z_1 \cdot \omega_r . \tag{1}$$

Motor's frame beats depend on inertial F_{in} and hydraulic F_h forces acting on rotor,

$$F_{in} = m z_2 \omega^2 e; \tag{2}$$

$$F_h = M_{ind} e z_1 . \tag{3}$$

During engine start, a skew moment arises, causing instability of rotor rolling along stator teeth and leading to additional beating of SDM frame.
Skew moment is:

$$M_n = \frac{P \cdot D \cdot t^2}{4\pi} . \tag{4}$$

Experimental study of motor's frame beats is performed at the test bench (Figure 1) [6]. Bench is equipped with an automatic control system that provides real-time output of SDM main energy characteristics to panel of a personal computer. To study beats of SDM, oscillation sensors are installed on frame.

30 SDM with diameters from 156 to 195 mm were tested to determine their optimal stable operation. As an example, the results of investigations on the energy characteristics of the motor with shortened spindle and adjustable unit curvature – 178M.7/8.37 with synchronous measurement of the wobbling on the body (Figures 2-4). Vibration measurement sensors (you can see on Figure 1: 10, 11 and 12 positions) has been installed on three points of the motor. Two sensors

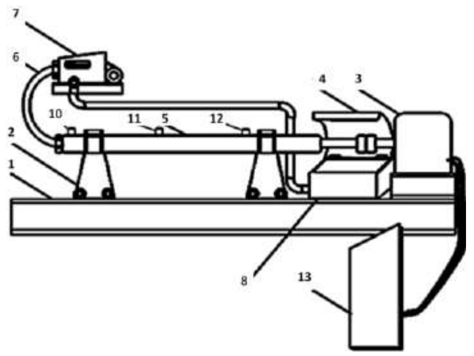


Figure 1. Test bench: 1 – installation platform; 2 – compressor; 3 – electromagnetic powder breaker; 4 – hydro-chisel; 5 – SDM; 6 – pipe line; 7 – pump; 8 – receiver tank; 10, 11, 12 – vibration measurement sensors; 13 – hardware and software complex (information processing module)

are installed in top and middle parts power section, and the third one is installed in the top part spindle (in the place of its coupling with hinge coupling). Measurements of energy characteristics and body wobbling were carried out from maximum frequency of shaft rotation 5 s^{-1} (300 rpm) to 0.5 s^{-1} (30 rpm). In the test process, the fluid flow was held at constant value of 0.03 m^3/s. Upon reaching the shaft rotation frequency of 5 s^{-1}, a moment of resistance was created by the brake 3, leading to a complete stop of the engine.

The SDM oscillations were measured in frequency bands with a constant relative width with the possibility of representing on a single graph a wide frequency range with a fairly narrow resolution at low frequencies.

Vibration acceleration at different frequencies from 1 to 43 Hz was recorded in three mutually perpendicular directions *x*, *y*, *z* with simultaneous measurement of the energy characteristics of the SDM. The levels of vibration velocity, vibration acceleration and amplitude are related by the following equations:

$$L_v = 20 \cdot \lg\left({}^{v}\!/_{v_0} \right), L_a = 20 \cdot \lg\left({}^{a}\!/_{a_0} \right); \tag{5}$$

$$A = {}^{1}\!/_{2\pi f} \cdot v(a). \tag{6}$$

Based on experimental study, shaft rotation frequency is determined, which ensures minimal lateral oscillations and optimal axial beats of motor.

Modeling of tool operation is carried out on an advanced mathematical model of Yunin and Khegai [10]. Mathematical modeling was performed in engineering mathematical software MathCAD.

At well drilling, it is required to determine the combination of load on the bit along depth *P* and rotor rotation frequency *n₀* so that drilling time *t* of specified interval is minimal under condition of optimal energy costs [8].

DS can be represented as a composite rod, interval of drill collar and interval, represented by SDM frame and navigation. Current well depth $H = L_1 + L_2 + L_3$ in process of drilling a certain interval increases due to deepening of the BHA. At this stage, let us assume L_2, L_3 = const and due to increasing of $L_1 + \Delta L$, H also rises.

Let us consider that sections are made of various materials. Therefore, first, second and third section corresponds to propagation velocity of rotational oscillations, propagation velocity of translational oscillations. Computational scheme for analyzing DS behavior during rotational and translational motion is shown in Figure 2.

Differential equations of rotational and translational motion of composite heterogeneous rod with initial and boundary conditions are given as follows:

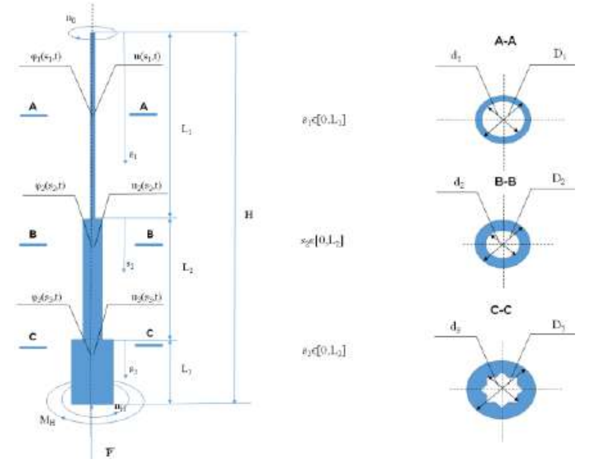


Figure 2. Computational scheme for study of rotational and translational oscillations of SDM – DS system operation

$$\begin{cases} \dfrac{\partial^2 \varphi_1}{\partial t} + f_{\tau_1}\left(\dfrac{\partial \varphi_1}{\partial t}\right) = \lambda_{L_1}{}^2 \dfrac{\partial^2 \varphi_1}{\partial s_1{}^2}, s_1 \in [0, L_1]; \\ \dfrac{\partial^2 \varphi_2}{\partial t} + f_{\tau_2}\left(\dfrac{\partial \varphi_2}{\partial t}\right) = \lambda_{L_2}{}^2 \dfrac{\partial^2 \varphi_2}{\partial s_2{}^2}, s_2 \in [0, L_2]; \\ \dfrac{\partial^2 \varphi_3}{\partial t} + f_{\tau_3}\left(\dfrac{\partial \varphi_3}{\partial t}\right) = \lambda_{L_3}{}^2 \dfrac{\partial^2 \varphi_3}{\partial s_3{}^2}, s_3 \in [0, L_3]. \end{cases}$$

$$\begin{cases} \dfrac{\partial^2 u_1}{\partial t} + f_{\tau_1}\left(\dfrac{\partial u_1}{\partial t}\right) = \chi_{L_1}{}^2 \dfrac{\partial^2 u_1}{\partial s_1{}^2}, s_1 \in [0, L_1]; \\ \dfrac{\partial^2 u_2}{\partial t} + f_{\tau_2}\left(\dfrac{\partial u_2}{\partial t}\right) = \chi_{L_2}{}^2 \dfrac{\partial^2 u_2}{\partial s_2{}^2}, s_2 \in [0, L_2]; \\ \dfrac{\partial^2 u_3}{\partial t} + f_{\tau_3}\left(\dfrac{\partial u_3}{\partial t}\right) = \chi_{L_3}{}^2 \dfrac{\partial^2 u_3}{\partial s_3{}^2}, s_3 \in [0, L_3]. \end{cases} \tag{7}$$

Boundary conditions for rotational motion are:

$1. s_1 = 0; \varphi = n_0 t, M = G_1 J_1 \dfrac{\partial \varphi_1}{\partial s_1},$

$2. s_1 = L_1; s_2 = 0; G_1 J_1 \dfrac{\partial \varphi_1}{\partial s_1} = G_2 J_2 \dfrac{\partial \varphi_2}{\partial s_2},$

$3. s_1 = L_1; s_2 = 0; \varphi_1 = \varphi_2,$

$4. s_2 = L_2; s_3 = 0; G_2 J_2 \dfrac{\partial \varphi_2}{\partial s_2} = G_3 J_3 \dfrac{\partial \varphi_3}{\partial s_3},$

$5. s_2 = L_3; s_3 = 0; \varphi_2 = \varphi_3,$

$6. s_3 = L_3; G_3 J_3 \dfrac{\partial \varphi_3}{\partial s_3} = -M_H(P, n_H).$

Boundary conditions for translational motion are:

$1. s_1 = 0; u_1 = h, N = E_1 F_1 \dfrac{\partial u_1}{\partial s_1}$

$2. s_1 = L_1; s_2 = 0; E_1 F_1 \dfrac{\partial u_1}{\partial s_1} = E_2 F_2 \dfrac{\partial u_2}{\partial s_2},$

$3. s_1 = L_1; s_2 = 0; u_1 = u_2,$

$4. s_2 = L_2; s_3 = 0; E_2 F_2 \dfrac{\partial u_2}{\partial s_2} = E_3 F_3 \dfrac{\partial u_3}{\partial s_3},$

$5. s_2 = L_3; s_3 = 0; u_2 = u_3,$

$6. s_3 = L_3; E_3 F_3 \dfrac{\partial u_3}{\partial s_3} = P(n_H).$

Initial conditions for rotational motion at *t* = 0:

$$7.\varphi_1(s_1,t=0)=\frac{f_{\tau_1}(n_0)}{2\lambda_{L_1}{}^2}\cdot s_1{}^2-\left\{\frac{f_{\tau_1}(n_0)L_1}{\lambda_{L_1}{}^2}+\theta\left[\frac{f_{\tau_2}(n_0)L_2}{\lambda_{L_2}{}^2}+\varepsilon\left(\frac{f_{\tau_3}(n_0)L_3}{\lambda_{L_3}{}^2}+\frac{M_H(P,n_H)}{G_3J_3}\right)\right]\right\}s_1,$$

$$8.\varphi_2(s_2,t=0)=f_1(L_1)+\frac{f_{\tau_2}}{2\lambda_{L_2}{}^2}\cdot s_2{}^2-\left[\frac{f_{\tau_2}(n_0)L_2}{\lambda_{L_2}{}^2}+\varepsilon\left(\frac{f_{\tau_3}(n_0)L_3}{\lambda_{L_3}{}^2}+\frac{M_H(P,n_H)}{G_3J_3}\right)\right]s_2,$$

$$9.\varphi_3(s_3,t=0)=f_1(L_1)+f_2(L_2)+\frac{f_{\tau_3}(n_0)}{2\lambda_{L_3}{}^2}\cdot s_3{}^2-\left(\frac{f_{\tau_3}(n_0)L_3}{\lambda_{L_3}{}^2}+\frac{M_H(P,n_H)}{G_3J_3}\right)s_3.$$

$$s_1\in[0,L_1],s_2\in[0,L_2],s_3\in[0,L_3].$$

$$\frac{\partial\varphi_1}{\partial t}=n_0,\frac{\partial\varphi_2}{\partial t}=n_0,\frac{\partial\varphi_3}{\partial t}=n_0.$$

Initial conditions for translational motion at t = 0:

$$7.u_1(s_1,t=0)=\frac{f_{\tau_1}(n_0)}{2\chi_{L_1}{}^2}\cdot s_1{}^2-\left\{\frac{f_{\tau_1}(n_0)L_1}{\chi_{L_2}{}^2}+\theta\left[\frac{f_{\tau_2}(n_0)L_2}{\chi_{L_2}{}^2}+\varepsilon\left(\frac{f_{\tau_3}(n_0)L_3}{\chi_{L_3}{}^2}+\frac{P(n_H)}{E_3F_3}\right)\right]\right\}s_1,$$

$$8.u_2(s_2,t=0)=f_1(L_1)+\frac{f_{\tau_2}}{2\chi_{L_2}{}^2}\cdot s_2{}^2-\left[\frac{f_{\tau_2}(n_0)L_2}{\chi_{L_2}{}^2}+\varepsilon\left(\frac{f_{\tau_3}(n_0)L_3}{\chi_{L_3}{}^2}+\frac{P(n_H)}{E_3F_3}\right)\right]s_2,$$

$$9.u_3(s_3,t=0)=f_1(L_1)+f_2(L_2)+\frac{f_{\tau_3}(n_0)}{2\chi_{L_3}{}^2}\cdot s_3{}^2-\left(\frac{f_{\tau_3}(n_0)L_3}{\chi_{L_3}{}^2}+\frac{P(n_H)}{E_3F_3}\right)s_3.$$

$$s_1\in[0,L_1],s_2\in[0,L_2],s_3\in[0,L_3].$$

$$\frac{\partial u_1}{\partial t}=\chi_{L_1}u_1,\frac{\partial u_2}{\partial t}=\chi_{L_2}u_2,\frac{\partial u_3}{\partial t}=\chi_{L_3}u_3.$$

where $n_H=\left.\frac{\partial\varphi_3}{\partial t}\right|_{s_3=L_3}$ – rotation frequency for end of composite rod lower section.

This problem is most clearly solved for case in which the values of dissipative terms of system are equal to zero. Following equations are used:

$$\begin{cases} H=\dfrac{\lambda_1}{\mu_1}\ln\dfrac{\lambda_1\left(\dfrac{\lambda_2}{\mu_2}\ln\dfrac{\lambda_2\left[\dfrac{\lambda_3}{\mu_3}\ln\dfrac{\lambda_3\Delta M_H+G_3J_3\cdot n_H}{\lambda_3\Delta M_H-G_3J_3\cdot n_H}\right]+G_2J_2\cdot n_0}{\lambda_2\left[\dfrac{\lambda_3}{\mu_3}\ln\dfrac{\lambda_3\Delta M_H+G_3J_3\cdot n_H}{\lambda_3\Delta M_H-G_3J_3\cdot n_H}\right]-G_2J_2\cdot n_0}\right)+G_1J_1\cdot n_0}{\lambda_1\left(\dfrac{\lambda_2}{\mu_2}\ln\dfrac{\lambda_2\left[\dfrac{\lambda_3}{\mu_3}\ln\dfrac{\lambda_3\Delta M_H+G_3J_3\cdot n_H}{\lambda_3\Delta M_H-G_3J_3\cdot n_H}\right]+G_2J_2\cdot n_0}{\lambda_2\left[\dfrac{\lambda_3}{\mu_3}\ln\dfrac{\lambda_3\Delta M_H+G_3J_3\cdot n_H}{\lambda_3\Delta M_H-G_3J_3\cdot n_H}\right]-G_2J_2\cdot n_0}\right)-G_1J_1\cdot n_0}, \\ H\le\dfrac{\lambda_1}{\mu_1}\ln\dfrac{\left|\lambda_1\left|\dfrac{\lambda_2}{\mu_2}\ln\dfrac{\lambda_2\left[\dfrac{\lambda_3}{\mu_3}\ln\dfrac{\lambda_3\Delta M_H+G_3J_3\cdot n_H}{\lambda_3\Delta M_H-G_3J_3\cdot n_H}\right]+G_2J_2}{\lambda_2\left[\dfrac{\lambda_3}{\mu_3}\ln\dfrac{\lambda_3\Delta M_H+G_3J_3\cdot n_H}{\lambda_3\Delta M_H-G_3J_3\cdot n_H}\right]-G_2J_2}\right|+G_1J_1\cdot n_0\right|}{\left|\lambda_1\left|\dfrac{\lambda_2}{\mu_2}\ln\dfrac{\lambda_2\left[\dfrac{\lambda_3}{\mu_3}\ln\dfrac{\lambda_3\Delta M_H+G_3J_3\cdot n_H}{\lambda_3\Delta M_H-G_3J_3\cdot n_H}\right]+G_2J_2}{\lambda_2\left[\dfrac{\lambda_3}{\mu_3}\ln\dfrac{\lambda_3\Delta M_H+G_3J_3\cdot n_H}{\lambda_3\Delta M_H-G_3J_3\cdot n_H}\right]-G_2J_2}\right|-G_1J_1\right|}. \end{cases} \tag{8}$$

That equation defines the occurrence possibility of torsional auto-oscillation in the composite DS during its rotational movement for well deepening purpose.

Equations (8) are found for the special case, when dissipative terms in Equation (7) are equal to zero. Results of Equations (8) are converged with the investigation results. Therefore, it is now possible to determine the value of parameters in the steady work mode of active dynamic system «SDM-DS».
Then we have:

$$\begin{cases} P=\dfrac{G_3J_3}{\lambda_3}\left(n_0{}^2\right)\dfrac{1+ke^{\frac{\mu_3L_3}{2\lambda_3}}}{1-ke^{\frac{\mu_3L_3}{2\lambda_3}}}, \\ P_b=\dfrac{1+ke^{\frac{\mu_3L_3}{\lambda_3}}P\cdot\left(ch\left(\dfrac{\mu_1L_1}{2\lambda_1}+\dfrac{\mu_2L_2}{2\lambda_2}+\dfrac{\mu_3L_3}{2\lambda_3}\right)+kch\left(\dfrac{\mu_1L_1}{2\lambda_1}-\dfrac{\mu_2L_2}{2\lambda_2}-\dfrac{\mu_3L_3}{2\lambda_3}\right)\right)}{1-ke^{\frac{\mu_3L_3}{\lambda_3}}n_0\left(sh\left(\dfrac{\mu_1L_1}{2\lambda_1}+\dfrac{\mu_2L_2}{2\lambda_2}+\dfrac{\mu_3L_3}{2\lambda_3}\right)+ksh\left(\dfrac{\mu_1L_1}{2\lambda_1}-\dfrac{\mu_2L_2}{2\lambda_2}-\dfrac{\mu_3L_3}{2\lambda_3}\right)\right)}, \\ P_H=\dfrac{1+ke^{\frac{\mu_3L_3}{\lambda_3}}P\cdot\left(sh\left(\dfrac{\mu_1L_1}{2\lambda_1}+\dfrac{\mu_2L_2}{2\lambda_2}+\dfrac{\mu_3L_3}{2\lambda_3}\right)+ksh\left(\dfrac{\mu_1L_1}{2\lambda_1}-\dfrac{\mu_2L_2}{2\lambda_2}-\dfrac{\mu_3L_3}{2\lambda_3}\right)\right)}{1-ke^{\frac{\mu_3L_3}{\lambda_3}}n_0\left(ch\left(\dfrac{\mu_1L_1}{2\lambda_1}+\dfrac{\mu_2L_2}{2\lambda_2}+\dfrac{\mu_3L_3}{2\lambda_3}\right)+kch\left(\dfrac{\mu_1L_1}{2\lambda_1}-\dfrac{\mu_2L_2}{2\lambda_2}-\dfrac{\mu_3L_3}{2\lambda_3}\right)\right)}, \\ n_0{}^*=\dfrac{1-k^2}{sh^2\left(\dfrac{\mu_1L_1}{2\lambda_1}+\dfrac{\mu_2L_2}{2\lambda_2}+\dfrac{\mu_3L_3}{2\lambda_3}\right)+kch^2\left(\dfrac{\mu_1L_1}{2\lambda_1}-\dfrac{\mu_2L_2}{2\lambda_2}-\dfrac{\mu_3L_3}{2\lambda_3}\right)}; \end{cases} \tag{9}$$

Task for case, in which the values of dissipative members of system are equal to zero, and propagation depth of translational oscillations of drilling tool, represented as a composite rod of three heterogeneous sections, is solved by system (8). At the same time G_1, G_2, G_3 are replaced by E_1, E_2, E_3 and J_1, J_2, J_3 by F_1, F_2, F_3, and also $\lambda_{L_1},\lambda_{L_2},\lambda_{L_3}$ propagation velocity of rotational oscillations by propagation velocity of translational oscillations by χ_{L1}, χ_{L2}, χ_{L3} in corresponding sections. Obtained equations determine conditions for occurrence possibility of translational self-oscillations of DS, represented as a composite rod in process of translating to deepen bottomhole of the well [19]. Axial loads on lower end of SDM frame, corresponding to upper and lower boundaries of self-oscillations during translational movement of P_B and P_H, are determined by Equation (9). At the same time propagation velocity of rotational oscillations $\lambda_{L_1},\lambda_{L_2},\lambda_{L_3}$ is replaced by χ_{L1}, χ_{L2}, χ_{L3}, elastic modulus G_1, G_2, G_3 and polar moments of inertia in cross-section J_1, J_2, J_3 are replaced by E_1, E_2, E_3 and F_1, F_2, F_3, respectively.

Developed methodology of determining required parameters for drilling mode of inclined sections in a well, ensuring stable operation of BHA, is as follows.

SDM is started and pressure drop is determined during its operation in idle mode. Then, required load on the bit (according to work plan and geological and technical schedule) is created and pressure drop is fixed taking into account loading of gerotor mechanism. On the basis of SDM test bench diagram, optimal range of shaft rotation frequency with corresponding pressure drop is graphically determined. At the same time, maximum allowable decrease in rotation frequency of SDM shaft is noted, which corresponds to optimum amplitudes of frame lateral oscillations.

According to mathematical model developed, boundaries of DS self-oscillation onset are calculated.

After constructing the graphical dependencies, required frequency and load on the bit are determined, at which DS is in permissible range of stable operation. Noting modes of DS stable operation, correlation is made with load on the bit, at which SDM will also be in mode of optimal en- ergy characteristics. If rotation frequency of SDM shaft (according to test bench diagram), determined by pressure drop, has decreased by more than 70 %, load on the bit is reduced. Based on graphical dependences for boundaries range of self-oscillations' onset at given rotation frequencies of DS and load on the bit, rotation frequency of top drive is adjusted to ensure stable operation of the system while maintaining mechanical drilling speed [18].

3. RESULTS AND DISCUSSION

Results of ordeal work process hydro machine with consideration of its vibration acceleration and amplitudes body wobbling on different work mode are illustrated in Figure 3.

The results of investigation on the motor vibration with shortened spindle and adjustable unit curvature – 178M.7 / 8.37 engine body showed that the values of vibration accelerations along the body are different from each other. So, for example, in the upper part of the power section for a shaft rotation frequency of 5 s^{-1} (300 rpm), the values of vibration accelerations of the transverse vibrations of the housing vary from 24 to 219 mm/s^2, which corresponds to a change in the beat amplitudes from 0, 19 to 1.75 mm, and the values of vibration accelerations of the longitudinal vibrations of the body vary from 49 to 251 mm/s^2 (Figure 3*b*), which corresponds to a change in the amplitudes of the wobbling from 0.39 to 2 mm (Figure 3*a*).

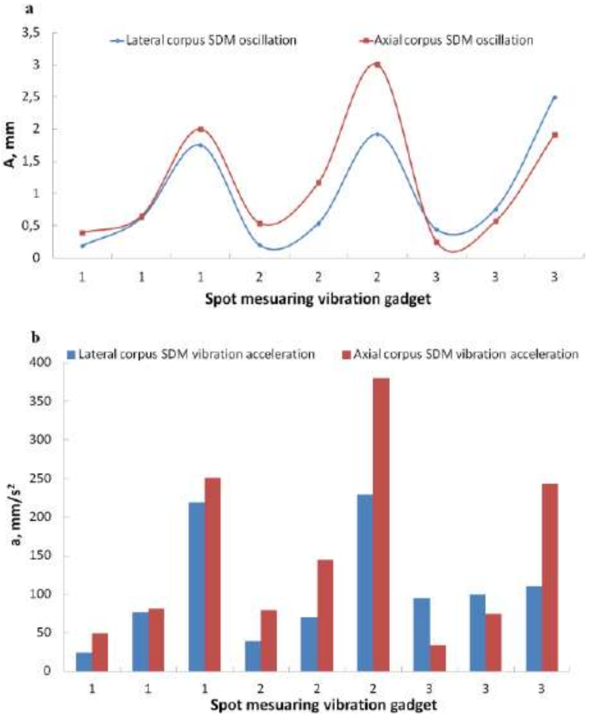


Figure 3. Amplitude (*a*) and vibration acceleration (*b*) of frame depending on sensor installation location at the motor′s frame: 1 – upper SDM sub; 2 – middle of working bodies' active part of SDM; 3 – upper spindle sub

The maximum values of the amplitudes and vibration accelerations of the lateral and axial body vibrations are determined at a frequency of 5 s^{-1} (300 rpm) in the middle part of the power section are 1.92 and 3 mm, 229 and 380 mm/s^2, respectively.

Results of research changes amplitudes lateral and axial oscillation depending on moment shaft of SDM as shown in Figure 4.

Analysis of investigation results showed optimum frequency interval of SDM body beats to be from 35 to 24.5 Hz. Axial and lateral vibrations depend on the moment on shaft SDM. In the regime of motor work, amplitude of lateral body wobbling is 5 mm, while the amplitude of axial oscillations is not more than 3 mm. This is due to the act of skew moments on the working elements. Resistance moment on shaft SDM made descend amplitude lateral vibration corpus to 3.5-4 mm, amplitude axial vibration increases to 7.3-8 mm. Increasing resistance moment from 1 to 4.5 kN·m motor body lateral vibration to 5-6 mm and descend corpus motor axial vibration to 7.3-8 mm. The wobbling frequency is reduced to 24.5 Hz (210 rpm), it is 30% power of idle SDM – optimum exploitation SDM (see Figure 4). With an increase in torque from 4.5 to 9 kN·m, the engine enters the braking (extreme) mode of operation. The wobbling frequency is reduced to 3.5 Hz (30 rpm). As a result, there is an intensive increase in the amplitude of the lateral vibrations of the SDM body from 6 to 10 mm which a corresponding decrease in the amplitude of axial vibrations from 8 to 2 mm.

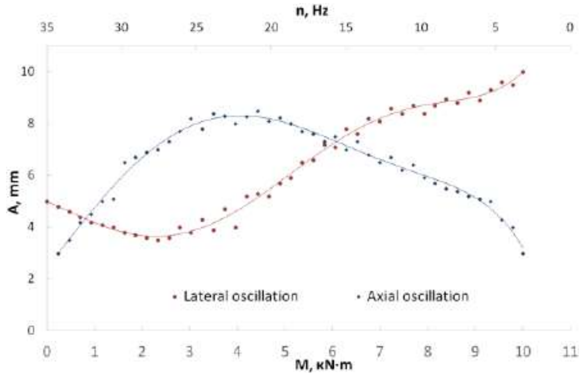


Figure 4. Average amplitude (with respect to the body length) in lateral and axial oscillations in dependence of SDM shaft moment

As a result of calculations based on developed mathematical model (7), range of self-oscillation onset during rotation and translational movement of SDM – DS system was revealed.

Input parameters for calculating rotational and translational movements:

L_1 = 1800 m; L_2 = 190 m; L_3 = 10 m; J_1 = 5.841·10^{-6} m^4; J_2 = 1.941·10^{-6} m^4; J_3 = 4.928·10^{-6} m^4; k = 0.106; G_1 = G_2 = G_3 = 8·10^{10} Pa; λ_{L1}, λ_{L2}, λ_{L3} = 3200 m/s;
n_0 = [0; 7] rad/s; $\mu_1 = 0.1, \mu_2 = 0.2, \mu_3 = 0.3$.

L_1 = 1800 m; L_2 = 190 m; L_3 = 10 m; F_1 = 1.018·10^{-3} m^2; F_2 = 1.81·10^{-3} m^2; F_3 = 8.042·10^{-4} m^2; k = 0.106; E_1 = E_2 = E_3 = 2·10^{10} Pa; χ_{L1}, χ_{L2}, χ_{L3} = 5320 m/s;
n_0 = [0; 7] rad/s; $\mu_1 = 0.1, \mu_2 = 0.2, \mu_3 = 0.3$.

Results of mathematical modeling are illustrated in Figure 5. Comparison of obtained study results for SDM frame oscillations in bench conditions with calculated values of boundaries of DS self-oscillations allows determining the range of stable operation for SDM – DS system. Values located under the line indicated by lower boundary of P_b self-oscillations mean absence of vibration – uniform translational and rotational movement of tool, between upper P_B and lower P_B boundaries – a temporary stop (jamming), above the upper P_H – braking (no rotation).

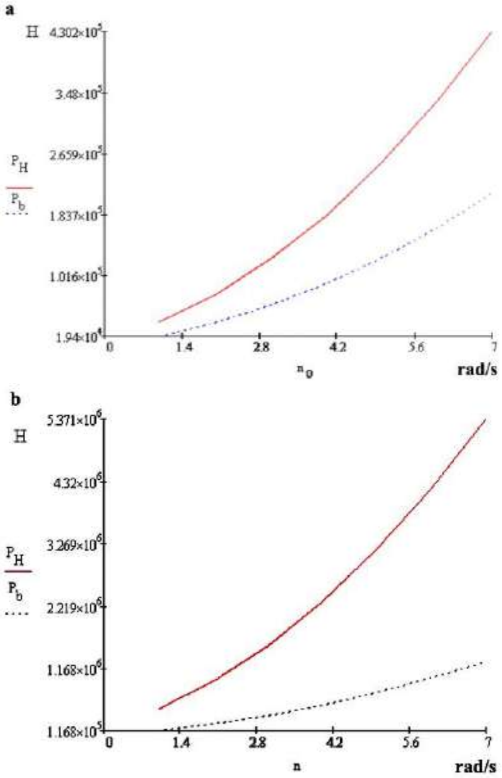


Figure 5. The boundaries of the rotational (a) and translational self-oscillations (b) of the «SDM-DS» system

The results of mathematical modeling are presented in Figure 5 (a and b).

Figure 5(a) shows the boundaries of the self-oscillations of the «SDM-DS» system due to the dynamic axial load resulting from the rotational and translational movement of the system.

Comparison of the obtained research results of the SDM oscillations (Figure 1) with the calculated boundaries values of the DS self-oscillations allow to determine the range of «SDM-DS» system stable work. The values located under the line indicated by the lower boundary of self-oscillations (P_b) means the absence of vibration – uniform translational and rotational movement of the tool, between the upper (P_H) and lower (P_b) boundaries – shut-down (jamming), above the upper (P_H) – braking (deficiency of rotation).

To perform stable operation, the «SDM-DS» system determines the rotor speed and axial load on the end face of the lower part of the composite rod of the represented of SDM, equal to the upper and lower self-oscillation boundaries that are recorded in the system (9).

By knowing the value P_b and P_H and considering the optimum frequency of drill string rotation n_0, the values of SDM working energy characteristics in the load condition (as shown in Figure 4) can be determined.

4. CONCLUSION

(1) The data of experimental researches aimed at determining the vibration values of the SDM body showed that a decrease in the rotational speed of not more than 30% of the maximum rotational speed of the shaft in idle mode allows for the stable operation of the «SDM-DS» system. At the same time, the moment in the indicated range varies from 1 to 4.5 kN m, which is sufficient for the implementation of volumetric destruction of rock.
(2) In the case of operation of the SDM in extreme mode (maximum power mode), a sharp increase in the amplitude of oscillations occurs in the lower part of the SDM, which leads to the appearance of half-waves in the BHA and loss of tool stability.
(3) The mathematical model has been improved, which allows to determine the diameters of the drilling operational parameters that ensure the BHA stability by controlling the dynamics of the «SDM-DS» system during its joint operation.

The developed methodology and technical recommendations aimed at ensuring stable operation of SDM with simultaneous rotation of drilling string at drilling directional wells are used in the branch of «LUKOIL-Engineering» LLC – «KogalymNIPIneft».

Further research will look at technical solutions that reduce vibrations during drilling. Namely, new models of screw downhole engines and justification of their technical and technological efficiency.

5. REFERENCES

1. Nutskova, M.V., Kupavykh, K.S., Sidorov, D.A., Lundaev, V.A., «Research of oil-based drilling fluids to improve the quality of wells completion». ***IOP Conf. Series: Materials Science and Engineering***, Vol. 666, No. 1, (2019). DOI: 10.1088/1757-899X/666/1/012065
2. Wang, H., Guan, Z.-C., Shi, Y.-C., Liu, Y.-W., Liang, D.-Y., «Drilling Trajectory Prediction Model for Push-the-bit Rotary Steerable Bottom Hole Assembly», ***International Journal of Engineering (IJE)***, TRANSACTIONS B: Applications Vol. 30, No. 11, (2017), 1800-1806. DOI: 10.5829/ije.2017.30.11b.23
3. Neskoromnykh, V.V., Popova, M.S., «Development of a drilling process control technique based on a comprehensive analysis of the criteria», ***Journal of Mining Institute***, Vol. 240, (2019), 701-710. DOI: 10.31897/PMI.2019.6.701
4. Liu, X.H., Liu, Y.H., Feng, D., «Downhole Propulsion/Steering Mechanism for Wellbore Trajectory Control in Directional Drilling». ***Applied Mechanics and Materials***, Vol. 318, (2013), 185-190. DOI: 10.4028/www.scientific.net/AMM.318.185
5. Izadia, M., Tabatabaee Ghomi, M., Pircheraghib, G., «Mechanical Strength Improvement of Mud Motor's Elastomer by Nano Clay and Prediction the Working Life via Strain Energy», ***International Journal of Engineering (IJE)***, Transactions B: Applications Vol. 32, No. 2, (2019), 338-245. DOI: 10.5829/ije.2019.32.02b.20
6. Simonyants, S.L, Al Taee, M., «Stimulation of the Drilling Process with the Top Driven Screw Downhole Motor», ***Journal of Mining Institute***, Vol. 238, (2019), 438-442. DOI: 10.31897/PMI.2019.4.438
7. Leine, R.I., van Campen, D.H., Keultjes, W.J.G., «Stick-slip whirl interaction in drillstring dynamics». ***ASME Journal of Vibration and Acoustics***, Vol. 124, No. 2, (2002), 209-220. DOI: 10.1007/1-4020-3268-4_27
8. Li, Z., Guo, B., «Analysis of longitudinal vibration of drillstring in air and gas drilling». ***Rocky Mountain Oil and Gas Technology Symposium***, (2007). DOI: 10.2128/107697-MS
9. Lui, H., Irving, P. G., Jaspreet, S. D., «Identification and control of stick-slip vibrations using Kalman estimator in oil-well drill string», ***Journal of Petroleum Science and Engineering***, Vol. 140, (2016), 119-127. DOI: 10.1016/j.petrol.2016.01.017
10. Khegai, V., Savich, V., Mihitarov, A., ***Determination of torsional resonant frequencies of a tree***, IOP Conference Series: Earth and Environmental Science, 316, (2019). DOI: 10.1088/1755-1315/316/1/012020
11. Lyagov, I.A., Baldenko, F.D., Lyagov, A.V., Yamaliev, V.U., Lyagova, A.A. Methodology for calculating technical efficiency of power sections in small-sized screw downhole motors for the «Perfobur» system. ***Journal of Mining Institute***, Vol. 240, (2019), 694-700. DOI: 10.31897/PMI.2019.6.694
12. Dvoynikov, M.V., «Technology of Drilling Oil and Gas Wells by Modernized Screw Downhole Motors», ***Doct. Techn. Sci. Diss.***, Tyumen', (2010), 54 p. Source link: http://www.geokniga.org/books/14667
13. Dvoynikov, M.V., Syzrantsev, V., Syzrantseva, K., «Designing a High Resistant, High-torque Downhole Drilling Motor», ***International Journal of Engineering (IJE)***, Transactions A: Applications Vol. 30, No. 10, (2017), 1615-1621. DOI: 10.5829/ije.2017.30.10a.24
14. Tikhonov, V.S., Baldenko, F.D., Bukashkina, O.S., Liapidevskii V.Y., «Effect of hydrodynamics on axial and torsional oscillations of a drillstring with using a positive displacement motor», ***Journal of Petroleum Science and Engineering***, Vol. 183, (2019). DOI: 10.1016/j.petrol.2019.106423
15. Huang, M., Wang, Y., Liu, B., Gao, M., Wang L., «Development of Downhole Motor Drilling Test Platform», ***Geological Engineering Drilling Technology Conference (IGEDTC)***, Procedia Engineering 73, (2014), 71-77. DOI: 10.1016/j.proeng.2014.06.172
16. Moradi, S.S.T., Nikolaev, N.I., Chudinova, I.V., Martel, A.S., «Geomechanical study of well stability in high-pressure, high-temperature conditions», ***Geomechanics and Engineering***, Vol. 16, No. 3, 331-339. DOI: 10.12989/gae.2018.16.3.331

17 Gospodarikov, A.P., Zatsepin, M.A., «Mathematical modeling of boundary problems in geomechanics», ***Gornyi Zhurnal***, No. 19, (2019). DOI: 10.17580/gzh.2019.12.03

18. Zhu, X., Liping, T., Qiming, Y., «A literature review of approaches for stick-slip vibration suppression in oil well drill string». ***Advances in Mechanical Engineering***, (2014), No. 6. 1-17. DOI: 10.1155/2014/967952
19. Samuel, R., Robertson, J.E., «Vibration Analysis and Control with Hole-Enlarging Tools», ***Annual Technical Conference and Exhibition***, (2010). DOI: 10.2118/134512-MS

Persian Abstract

چکیده

مقاله ارائه شده در مورد احتمال افزایش بهره وری از بخش های مستقیم چاه های حفاری با استفاده از موتورهای سرپیچ پیچ (***SDM***) با روش ترکیبی از حفاری با چرخش رشته حفاری (***DS***). هدف حصول اطمینان از عملکرد پایدار ***SDM*** با چرخش همزمان ***DS*** با کاهش دامنه نوسانات با تنظیم پارامترهای حالت حفاری بر اساس مدل سازی ریاضی برای سیستم ***SDM – DS*** است. نتایج مطالعه تجربی در مورد تعیین توزیع اکسترا از نوسانات جانبی و محوری قاب ***SDM*** بسته به پارامترهای هندسی مکانیسم حرکت و حالت های حصول اطمینان از عملکرد پایدار ارائه شده است. رویکردهای توسعه یک مدل و روش ریاضی به صورت مفهومی بیان شده است که اجازه می دهد طیف وسیعی از نوسانات خود را برای سیستم ***SDM-DS*** و مرزهای اختلال موج چرخشی و ترجمه ای برای یک میله ناهمگن با یک ***SDM*** نصب شده در حفاری بخش های مستقیم چاه ، مشخص کند. این مدل ریاضی پویایی سیستم ***SDM-DS*** امکان پیش بینی پارامترهای بهینه از حالت حفاری جهت را فراهم می کند که عملکرد پایدار مونتاژ گمانه را تضمین می کند.

International Journal of Engineering

Journal Homepage: www.ije.ir

Investigations on Material Composition of Iron-containing Tails of Enrichment of Combined Mining and Processing in Kursk Magnetic Anomaly of Russia

K. R. Argimbaev*, D. N. Ligotsky, K. V. Mironova, E. V. Loginov

Faculty of Mining, Saint Petersburg Mining University, St. Petersburg, Russia Federation

PAPER INFO

Paper history:
Received 21 January 2020
Received in revised form 22 March 2020
Accepted 11 June 2020

Keywords:
Material Composition
Polished Section
Samples
Tailings
Thin Section

ABSTRACT

The inevitable depletion of mineral resources, the constant deterioration of the geological and mining conditions for the development of mineral deposits and the restoration of raw materials from mining waste by recycling are all urgent problems we faced today. The solution to this problem may ensure: a considerable extension of raw material source; decrease of investments in opening new deposits; cost savings for dumping and handling of tailing dumps, disturbed land remediation; obtaining social and economic effect due to a considerable reduction in pollution of the environment. This article deals with the study of iron-containing tailings dumped at the tailing dumps of ore-refinery and processing facilities located in Kursk Magnetic Anomaly (KMA GOKs), where samples were taken for this study. The article contains the results of the materials composition study, namely: chemical composition, the mineral-petrographic study of thin and polished sections, grain size distribution and physical-mechanical properties of tailing samples. Regularities were revealed for the change of the useful component content due to gravity differentiation. It was also noted that the sulphur content increased near the pulp discharge outlet due to pyrite accumulation. The ratio of ore minerals in tailings and the fineness ratio of the sand fraction were measured. The examination with a focused beam microscope with x90 to x600 magnification showed a variety of grain sizes and shapes that facilitate using tailing materials after additional processing in the construction industry as sand.

***doi**: 10.5829/ije.2020.33.07a.31*

1. INTRODUCTION

Processing of wastes from ore-refining facilities that accumulate at tailing dumps is among the major issues in complex treatment of mineral resources, environmental protection and remediation [1-3].

The idea of processing tailing was first conceived at the beginning of the 20th century in the 80s [4]. This idea is still being developed till date. Its main tasks comprise of the disposal of mining and metallurgical waste, development and implementation of measures toward considerable loss decrease and increase of the mineral recovery quality during ore mining and processing [5-10]. Despite the available vast technical potential and scientific-methodological background, efficient flowcharts for valuable component recovery from mineral processing waste have not been substantiated yet [11-15]. Big interest in elaboration technogenic formations started to manifest at the same time with new recycling technologies emergesing and partial depletion of large field.

Many investigators [9-15] have searched on process to systemize subsoil explorations and rationally resourse using potential techniques. The best scholars in this area were Trubetskoy et al. [11] and Trubetskoy [13] academician of russian science academy. They were the first who systemized and categorized the technological fields and process formations. The results of systems approaching to the questions about systemised subsoil explorations are:

- The intelligent method of subsoil explorations at the non-ferrous and ferrous metallurgy enterprises was generalized.

*Corresponding Author Institutional Email: *diamond-arg@mail.ru* (K. R. Argimbaev)

Please cite this article as: K. R. Argimbaev, D. N. Ligotsky, K. V. Mironova, E. V. Loginov, Investigations on Material Composition of Iron-containing Tails of Enrichment of Combined Mining and Processing in Kursk Magnetic Anomaly of Russia, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1431-1439

- The features of the internally structure and spatial variabillity of useful components content were identifieded.
- The recommended intelligent methodogy for evaluation of technological fields were designed. Also it includes the forecasting method spatial variabillity of useful components content in tailings, which is based on information about aprobaration.
- The economical feasibility of using perspective metalic and nonmetalic minerals was studied.
- The economic and mathematical model and softwere were developed. It helped us to optimize the volume of using different metalic and nonmetalic technological raw materials, to determine minimal industrial content which are useful components in balance stocks. Besides, to make a choise more effective directions to use technological raw materials and technological circuits of the process development.

Complex aggregate circuit of complex development technological mineral recourses which covers all stages: from intelligent till finished product. That is used to devide to connected cybsystems [8-15]:

- Geo-technological scrutiny.
- Techno-economical assessment and grounding the conditions.
- Designing the technological fields.
- Recycling of the raw minerals.
- Targeted formation of technological field with adjusted parametrs and specifications.
- Ground recultivating which were destroyed by technological fields.

Two independent directions were made by virtue of analysis of this subsystems:

- Familiarization of two technological formations.
- Making technological fields with adjusted parametrs and in view of this issue we have the following exploration.

Research and calculations which are made by different laboratories in the world, shows us the principal possibility to work off and complex rework ferrous tailings. But we may ascertay that fundamental switching over principles of transiting to the low-waste technologies, which can promote developing minerals fields through technological fields formation, and also development of current technological formations [14-29].

In spite of extensive theoretical capacity, effective technological circuit of extraction useful components weren't unfounded.

Since the issue of processing man-induced deposits formed by tailing dumps of ore-dressing facilities is the nearest future problem. The study of the material composition of iron-containing tailings formed at ore-refining and processing facilities of Russian Kursk Magnetic Anomaly is required a very long duration of time research. It will provide the ground for further scientific study dedicated to highly efficient processing ensuring minimal loss of useful components with waste.

2. NOVELITY RESEARCH AND PRACTICAL VALUE

The novelity of this research work is to establish the scientific approach to regulate and handling the enrichment of magnetite-containing tails. It depends on the content rating -0.044 mm in them, which makes it thus to tedermine optimal number 15 25% in which enrichment indicators are maximal.

The practical value of the work reflects that the main results have been used to design the processing enrichment tails.

3. METHODS

The objective of this work was to enrich substances of iron content and to justify the opportunity of using technological fields which are based on wasting results from enrichment Kursk Magnetic Anomaly (KMA GOKs). Such issues were adressesed during the process:

- Due to statistical analysis for the information in processing about quality and quantity of taillings.
- Recearch conducted on substances enrichment tails.
- Identification of the kind relationship of composition and enrichment.

To solve the tasks, research was carried out in three stages: field, laboratory and analytical parts. In the field period, gross samples were taken weighing 40-50 kg of waste from the taillings of GOKs: at Lebedinskiy Ore-Refinery and Processing Facility (LGOK) - 6 samples, at Stoilensky Ore-Refining and Processing Facility (SGOK) - 4 samples, at Mikhailovsky Ore-Refinery and Processing Facility (MGOK) - 6 samples. Samples were taken at different distances from the pulp outlet. That is due to the formation of spatially isolated sections of large fractions, as well as fractions with a high iron content in large taillings especially with unilateral discharge of pulp and to a lesser extent with contour. Taillings of Kursk Magnetic Anomaly (KMA GOKs) have all the prerequisites for the formation of such sites. As a result of the gravitational differentiation of the solid part of the pulp, the stored material is redistributed into the taillings dump and areas near the pulp outlets are formed with an increased iron content (compared to other taillings storages dump sections).

Significant influence at the iron lossing and enrichment tails render: imperfection of the existing technology for the enrichment of quartzite, which leads to incomplete extraction of iron in concentrate; such as emergency equipment shutdowns, especially during commissioning, accompanied by as a rule, by emergency discharges of enrichment products with an abnormally high iron content; imperfection or lack of schemes for the disposal and capture of spills and industrial wash products; insufficient organization production and low qualification staff.

During the analytical phase, quantitative and qualitative statistics were collected about enrichment tails, their use and recommendations were made for further use in tails, in various industries. The researching were carried out on certified equipment manufactured in the USA, Australia, Japan and Russia according to international standard methods. Chemical analyzes are performed in accordance with current GOSTs. Technological tests were carried out according to standard enrichment schemes, taking into account characteristics of the material composition of the enrichment tails.

4. RESULTS

4. 1. General Information about Deposit and Tailings of KMA

The Kursk Magnetic Anomaly (KMA) is the most powerful iron ore basin on earth, where KMA mining and processing plants (GOKs) are located. In mineralogical terms, the ores in the deposit are two-component or three-component formations consisting the hematite (and its morphological variety - martite), magnetite (and its morphological variety - musketovite), goethite, less often hydrohematite and carbonates. Minor minerals are: berthierin, chamosite, apatite, quartz, mica. Hematite is often hydrated in red cystral formation, sometimes brown, iron hydroxides, often staining ores in red and brown; the content in such interlayers of various hydroxides is very different.

The prevalence of mushy ores (hematite varieties) increases in those places where gentle paleoscopes along quartzites are noted. By genetic characteristic, all the minerals of weathering zone (oxidation) quartzites are divided into three groups: 1) relict, metamorphogenic - hematite, magnetite, quartz; 2) weathering minerals - martite, goethite, hydrohematite, hematite, berthierin (chamosite), marshall; 3) infiltration - siderite, calcite, glandular chlorite, pyrite, marcasite and iron hydroxides.

Among them, ore-forming ones are hematite, martite, goethite, hydrohematite and magnetite; minor - carbonates, chamosite (ferrous chlorite and berthierin) and quartz. A complex structures and equipment for hydraulic transport and hydraulic tailing of enrichment tails plants exists for the storage of enrichment waste at GOKs.

To store the wasting production of the ore processing plant, GOKs use large capacities of natural formations - ravine beams with the construction of dams and enclosing dams - tailings. Tailings ponds are filled in the initial period of GOKs operation by gravity hydraulic transport with subsequent application as production capacities increase and volumes of taillings pressuring hydraulic conveyors are increased by gradually increasing the length of slurry pipelines and dumping tails around the taillings perimeter.

4. 2. Tailings Chemical Composition Study

The process of the material composition tail's scrutiny for the gerruginous quartzites enrichment whick were taken from taillings KMA GOKs which was consisted of two sequentially carried out operations: taking an average sample from a certain amount of the mass of the product being tested; laboratory analysis of the sample substance. To obtain an average laboratory sample, the initial ore was crushed, mixed and reduced to the minimum allowable mass. The prepared chemical samples were subjected to spectral, chemical analyzes (according to the content of the main rock-forming oxides (SiO_2, Al_2O_3, Fe_2O_3, MgO, CaO, Na_2O, K_2O and Fe_{total}, $Fe_{mag.}$).

The results of the chemical composition of the initial samples are summarized in Tables 1-3.

Tables 1-3 show that the tailings chemical composition is caused by the initial rock properties. All tailings contain silicon dioxide SiO_2, iron oxide Fe_2O_3, and ferrous iron FeO as basic components.

Silicon dioxide content in Stoilensky and Lebedinsky tailings varied from 47.50 to 75.08%; mill tailings of Mikhailovsky GOK featured lower SiO_2 content. Mostly, silicon dioxide was bonded with quartz, and only a small part of constituted silicates. The highest content silicon dioxide was in SGOK tailings (up to 75.08%), the lowest – in MGOK tailings (up to 36.13%).

Iron oxides formed ore minerals – magnetite and hematite. Silicates contained small amounts of iron oxides. Their ratio in mill tailings was different. The highest Fe_2O_3 content was typical for MGOK tailings (39.91 to 38.12%) where hematite prevailed. LGOK tailings featured high FeO content (6.26 to 10.71%), it

TABLE 1. Chemical composition of mill tailing initial samples taken at LGOK, %

Components	ChVL -1	ChVL -2	ChVL -3	ChVL -4	ChVL -5	ChVL -6
SiO_2	67.58	68.02	66.37	54.07	65.77	65.47
Al_2O_3	1.98	2.04	1.77	2.36	3.37	3.60
Fe_2O_3	10.18	9.91	10.75	18.24	8.56	8.26
FeO	7.18	6.26	7.35	10.71	7.14	7.30
MgO	3.50	3.96	3.70	2.93	4.61	4.24
CaO	2.58	3.08	3.37	2.50	2.65	2.86
Na_2O	0.67	0.65	0.67	1.21	1.11	1.08
K_2O	0.44	0.45	0.49	0.70	0.72	0.75
Other	5.30	4.70	4.89	6.05	5.23	5.50
Total	99.41	99.07	99.07	98.77	99.16	99.06
Fe_{total}	12.70	11.76	13.09	21.07	11.53	11.45
$Fe_{mag.}$	3.02	2.18	3.17	8.06	2.10	2.0

TABLE 2. Chemical composition of tailing initial samples taken at SGOK, %

Components	ChVS -1	ChVS -2	ChVS -3	ChVS -4
SiO_2	63.89	71.63	47.50	75.08
Al_2O_3	3.13	1.84	1.92	2.69
Fe_2O_3	16.62	9.85	32.44	6.93
FeO	5.90	4.79	7.03	5.04
MgO	2.11	2.95	2.10	2.18
CaO	2.07	2.55	2.05	2.18
Na_2O	0.58	0.39	0.41	0.54
K_2O	0.77	0.53	0.49	0.73
Other	5.05	4.94	5.59	4.46
Total	100.36	99.92	100.05	100.19
Fe_{total}	16.20	10.61	28.15	8.76
$Fe_{mag.}$	1.65	1.14	4.02	1.13

TABLE 3. Chemical composition of mill tailing initial samples taken at MGOK, %

Components	ChVM -1	ChVM -2	ChVM -3	ChVM -4	ChVM -5	ChVM -6
SiO_2	47.48	51.75	36.13	42.28	33.84	47.46
Al_2O_3	0.04	0.07	0.07	0.04	0.14	0.04
Fe_2O_3	44.69	39.91	56.76	50.21	58.12	44.36
FeO	2.90	3.38	2.91	3.10	3.57	3.20
MgO	0.82	0.82	0.91	0.97	0.97	1.12
CaO	0.98	0.98	0.68	0.68	0.68	0.68
Na_2O	0.50	0.55	0.49	0.55	0.56	0.60
K_2O	0.44	0.40	0.33	0.37	0.32	0.38
Other	2.01	2.40	1.85	1.96	2.0	2.25
Total	99.86	100.6	100.3	100.6	100.2	100.9
Fe_{total}	33.5	30.5	41.9	37.5	43.4	33.5
$Fe_{mag.}$	1.78	1.58	1.94	2.19	2.5	1.74

was somewhat less in SGOK tailings (5.04 to 7.03%). Aluminium oxide Al_2O_3 as a component of mica, feldspar, amphiboles was in LGOK and SGOK mill tailings in approximately equal amounts (1.84 to 3.6%). Its content is minor in MGOK tailings – hundredths of a percent. The rest of the components – CaO, MgO – prevailed in LGOK tailings (1.77 to 3.60%). Na_2O and K_2O content was approximately the same for all tailing sites.

The regularities in the component content change were caused by gravity separation; they were similar for Lebedinsky and Stoilensky tailing dumps – iron-containing minerals (magnetite, hematite) accumulate near the pulp discharge outlet while the content of SiO_2 (pure quartz without joints), CaO, MgO, K_2O, Na_2O (silicates, amphiboles, carbonates) increases in remote areas.

Increased sulphur content due to pyrite accumulation was also noted near the discharge outlet.

4. 3. Tailings Mineralogical-Petrographic Study

Thin and polished sections of tailings were studied in transmitted and reflected light using EM3900M microscope.

Visually, LGOK and SGOK mill tailings were dark grey mineral of varied grain sizes, MGOK tailings were dark-brown. Grain size varied 5 mm to -0.05mm.

Tailing material composition was caused by mineral composition of the initial ore, specific features of the refining process and the nature of the material separation during the tailing dump filling.

Coarse fractions +5 mm, +2.5 mm were represented by ferrous quartzite fragments with amphibole-quartz composition with magnetite impregnations. Fine fractions were represented by separate minerals. Basic ore minerals comprised of magnetite, hematite; barren minerals – quartz, amphiboles, mica, carbonates, feldspar. Auxiliary minerals were represented by ilmenite, rutile and apatite.

Quartz was the main mineral that formed tailings; its content in samples varied from 35.0% (ChVM-3) to 58.4% (ChVL-2). Sample ChVS-3 was the exception, its quartz content was low – 27.3%. Quartz presented as sharply-angular irregular fragments with magnetite impregnations (Figure 1a). In finer fractions, quartz had round shape and virtually did not contain ore impregnations. Magnetite was the main ore mineral, which was presented as impregnations in quartz grains, in silicates, less often – as separate thin interlayers in ferrous quartzite fragments. The free magnetite amount increased with the fraction size decrease. Crystals had regular isometric shapes (Figure 1b).

Hematite was presented in two varieties: as fine impregnations in quartz and as small tabular flakes.

The second hematite variety was typical for Mykhailovsky tailings. Here, hematite contained in all fractions: as aggregates with irregular shape with translucent ruby-red edges (Figure 1c) in coarse fractions and as separate flakes of regular shape in fine ones. Most often, hematite was found as shots in green mica. Hematite content in MGOK tailings was 39 to 53.5%.

Mica existed of two minerals. Biotite-phlogopite type mica was typical for Lebedinsky and Stoilensky deposits. It formed black aggregates in coarse fractions and separate round or long tabular brown flakes in fine ones. Magnetite impregnations were rarely observed in biotite. Green mica was typical for the tailings of Mikhailovsky deposit. Fine fractions contained mica as dissipated lath-like emerald-green flakes; it often contained hematite impregnations.

Amphiboles existed as cummingtonite, alkaline amphibole, riebeckite, actinolite. Crystals had elongated prismatic or tabular shape with uneven edges at end faces (Figure 1d). Polysynthetic twins were typical for cummingtonite.

Carbonates presented as two minerals: yellow-brown ferrous siderite with ore impregnations and white dolomite with pearly lustre. Crystal shape was irregular, most often round. Pyrite had golden-yellow colour with tarnish, with irregular grains; its content was 0.5 %. Apatite, ilmenite, rutile were met as single auxillary minerals.

There was the following specific ratio of ore minerals: LGOK tailings featured the highest magnetite and low hematite content; SGOK tailings contained somewhat less magnetite (2.9 to 5.2%). However, discharged materials contained significant magnetite amount (2.1 to 3.3%) and maximum hematite amount (35 to 53.5 %).

Since samples were taken in the tailing dumps at different distance from discharge outlets, grain size distribution within one site considerably differed.

Grain size distribution in the initial material had the following specific features: most fine tailings material formed fraction -0.14 mm; the rest small portion was distributed between fractions -0.315+0.14 mm and -0.630+0.315 mm. The majority was distributed between fractions -0.315+0.14 mm and -0.63+0.315 mm with the material prevailing in the first one. As a rule, the main part of coarse tailings was formed by -0.315+0.14 mm and -0.63+0.315 mm fractions. LGOK and SGOK tailings were less uniform in terms of grain size, while MGOK tailings feature more uniform material distribution by grain structure.

Fineness ratio of the tailing sand fraction (i.e., fraction -5+0.14 mm) varied: 1.06 to 2.45 (average value 1.75) for LGOK, 1.12 to 2.57 (average value 1.85) for SGOK, and 1.69 to 2.79 (average value 2.24) for MGOK.

Sand fraction yield varied within 4.6 to 77.5 % for LGOK tailing dump, 10.2 to 80.8 % at SGOK dump and 48.5 to 77.1% at MGOK dump.

Thus, sand fraction of tailings was attributed to fine (fineness modulus < 2), middle (fineness modulus =2.0 to 2.5) and coarse sand (fineness modulus > 2.5) in terms of fineness ratio, which are suitable for construction works according to GOST 8736-2014 Sand for Construction Works. Specifications (Russia).

Tailing examination under the focused beam microscope has shown the presence of mineral particles ranging from less than a micron to several hundred microns (Figure 1e) in size. Quartz grains shape varied, but isometric sharply-angular particles prevailed. Elongated prismatic particle shape was less common (Figure 1f).

Elongated, prismatic, flaky particles were found more often in MGOK tailings. A detailed study of the mineral grain surface topography for the mill tailings at KMA GOKs showed that it had changed significantly due to shock loads during ferrous quartzite crashes. Grain surface is rough, irregular, with numerous defects (Figure 1g). Natural sand has even and smooth surfaces, with rare pits and cavities (Figure 1h).

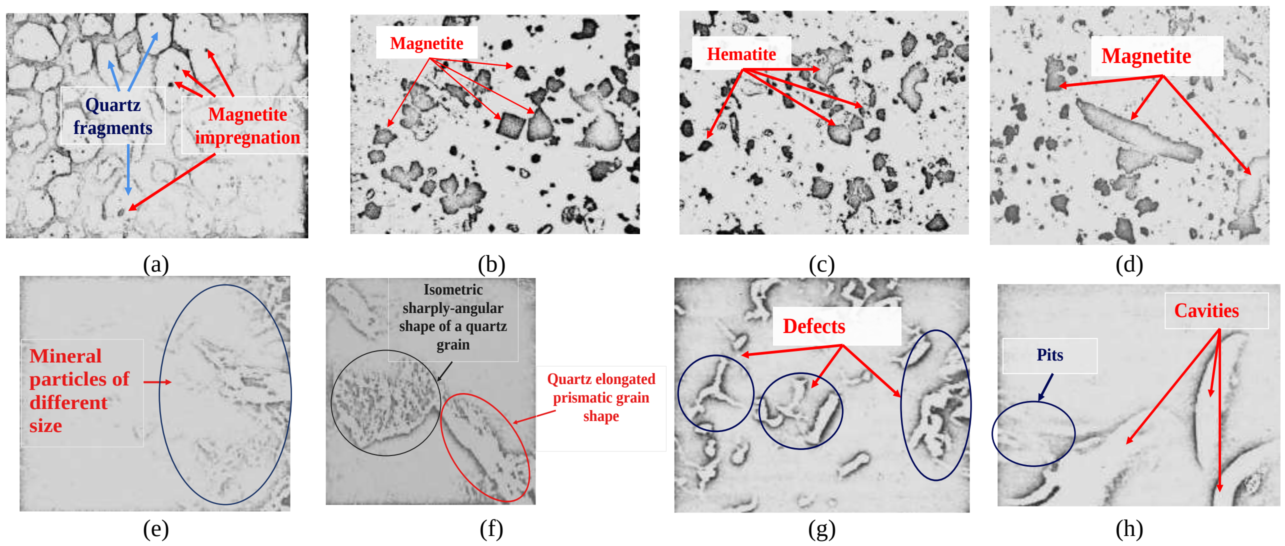

Quartz fragments
Magnetite impregnation
(a)
Magnetite
(b)
Hematite
(c)
Magnetite
(d)
Mineral particles of different size
(e)
Isometric sharply-angular shape of a quartz grain
Quartz elongated prismatic grain shape
(f)
Defects
(g)
Pits
Cavities
(h)

Figure 1. Scanning electron microscopy results: a - quartz grains with magnetite impregnations (Sample ChVM-2, Quartz is grey, magnetite is black. Transmitted light, x125 magnification); b - magnetite automorphic shape (Sample ChVL-2. Transmitted light, x125 magnification); c - fine hematite impregnations in green mica (Sample ChVM-6. Transmitted light, x 200 magnification); d - alkaline amphibole with magnetite impregnations (Sample ChVS-3. Transmitted light, x150 magnification); e - Lebedinsky GOK tailings (Sample ChVL-5. Transmitted light, x190 magnification); f - topography of quartz grain surface at Mikhailovsky GOK (Sample ChVM-5. Transmitted light, x190 magnification); g - the surface of the tailing quartz grain (Sample ChVS-1. Transmitted light, x600 magnification); h - Volsk sand deposit (Sample 1. Transmitted light, x450 magnification)

Specific features of grain micro-pattern for tailings caused their increased adhesive ability as compared with traditional sand. In this connection, using tailings as fine filler for concrete should ensure an additional increase in concrete strength due to better adhesive bonds with a binding agent.

The examination of polished samples of tailings with Epiquant structural analyser has shown that magnetite particle linear sizes were 20.0 μm to 42.9 μm (average value 30.1 μm) for LGOK tailings, 21.3 μm to 44.3 μm (average value 39.4 μm) for SGOK and 14.4 μm to 34.1 μm (average value 22.2 μm) for MGOK. It followed that the most coarse magnetite was contained in SGOK tailings while the fine one – in MGOK tailings.

Analysis of the samples taken from well No. 2 at LGOK has revealed that magnetite particles of 44.5 μm to 45.0 μm in size was contained in tailing coarse fractions (+5 mm, +2.5 mm). Fine fractions (+0.63 mm - 0.14 mm) featured magnetite grains with prevailing size 22.3 μm to 22.8 μm.

4. 4. The Influence of Material Composition of Tailings on Enrichment

The researchings of tailability tails samples for enrichment were carried out on samples in which the content ranges from 3.07 to 10.88%, from 13.25 to 20.06%, which makes it possible to compare the results of the enrichment of samples with different iron contents in the initial product.

It was established that tails enrichment are significantly affected by their material composition, in particular, the iron content in tails (especially magnetite) and the content of fine particle size fractions. The maximum indicators of tails enrichment are achieved with a grade of -0.044 mm in the initial product at the range of 15–25% (Figures 2 and 3).

This is due to the fact that this class contains the largest number of discovered magnetite grains, while the larger fractions contain quartz and aggregates of non-metallic minerals with magnetite, while the finely dispersed fractions contain barren sludge and over-crushed magnetite particles, which are weakly captured during enrichment. Based on the results of studying the enrichment of enrichment in laboratory conditions, the possibility of obtaining iron ore concentrate with iron content for the tailings of KMA GOK was established (initial tailings content of 5.95%) - 61.47%.

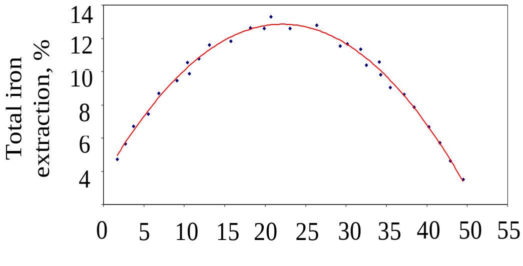


Figure 2. Dependence of the extraction of total iron on the grade content of -0.044 mm in the initial samples

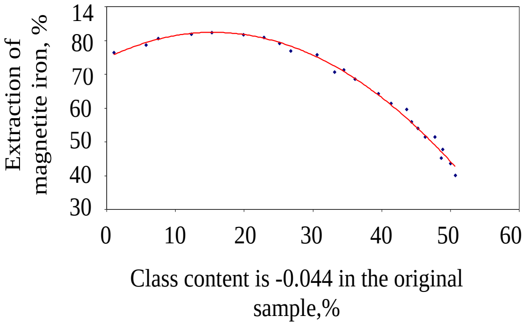


Figure 3. Dependence of the extraction of magnetite iron on the grade content of -0.044 mm in the initial samples

The proposed scheme to have a production of concentrate from tails according to the summary indicators of laboratory tests includes the following operations: screening, desliming, preliminary magnetic separation, the intermediate product of which has a yield of 34.8%. The total iron content is 30.6% and the recovery is 68.7%. After regrinding up to 98% of the class -0.044 mm and magnetic separation, the final product has qualitative characteristics: yield - 15%, total iron content 65.2%, recovery - 62.7%, with the initial parameters of the tailings: total iron content - 15.5 %, magnetite iron - 5.2%.

The output of building sands by fractionation the preliminary magnetic separation is 30.0%, the total iron content is 8.4%, the recovery is 16.2%. The general tailings according to this scheme have a yield of 55.1%, the total iron content in them is 5.9%, and the recovery is 21.1%.

4. 5. Physical and Mechanical Properties of Tailings

The physicomechanical properties of taillings are based on the results of testing tailings. The basic laboratory results are shown in Table 4 and in Figures 4 to 5, which illustrated that the weighted average diameter d_{wa} of the studied tailing materials changed from 0.02 mm to 0.21 mm. Within the specified rangedd_{wa}, the internal friction angle ϕ varied within the wide range 26° to 38° while adhesion value C changed from 0.002 to 0.011 MPa. In particular, shearing angle ψ (at $P = 1$ MPa) was 32 in the sample with the minimal angle ϕ and C.

Natural gravimetric humidity W of the studied tailing samples varied from 15 to 35% with average value 25 %. Porosity factor E was sufficiently low and changed from 0.67 to 0.81 depending the distance from the pulp discharge outlet L.

TABLE 4. Shearing test results for Lebedinsky GOK tailings

$d_{св}$, mm	ϕ, rad.	C, MPa	ψ, grad.	W, %	E	L, m
0.02	26	0.0021	32	35	0.777	300
0.03	27.21	0.003	33	28	0.726	250
0.051	28.42	0.0042	33	25	0.691	200
0.091	30.48	0.005	34	21	0.666	150
0.13	32.51	0.007	35	17	0.814	100
0.15	35.05	0.009	36	15	0.717	40
0.21	37.94	0.011	40	–	–	–

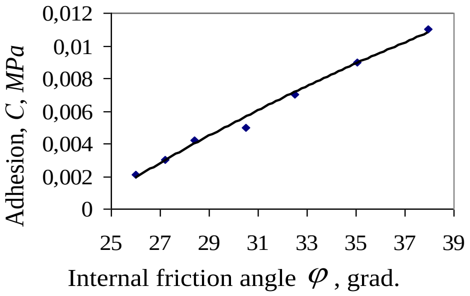


Figure 4. Adhesion vs. the internal friction angle plot

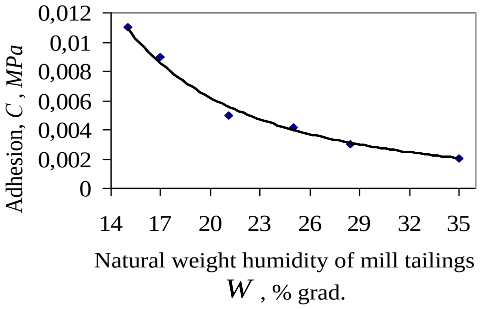


Figure 5. Adhesion vs. natural gravimetric humidity plot

The research results showed that the increase of the distance from the tailing discharge outlet resulted in the growth of humidity percentage for iron-containing mill tailings and, hence, to the decrease of physical and mechanical properties and the bearing capacity of the surface tails when mining equipment is located on it.

5. DISCUSSION

The recearching of the material composition to assess the possibility of using enrichment tailings for KMA GOKs to obtain iron ore concentrate showed that the selected samples differ in chemical, mineralogical and granulometric composition both within the same taillingd dump and between taillings of different GOKs.

When the tail pulp is stratified, larger and heavier particles are placed in the pulp duct in the lower part, and the pulp concentration is higher here than in the upper part of the stream. As a result, through the first alluvial outlets from the distribution slurry conduit, larger and heavier particles enter the beach tail zone dump than through the subsequent ones (for example, to tailings in the MGOK).

Further segregation of particles occurs on the alluvial beach and in the taillings pond under the influence of changes in the hydrodynamic characteristics of the carrier flow. On the surface taillings dump during the alluvial formation, taillings zones with a similar composition are formed, and the configuration of the zones depends on the pattern of tails storages dump. The general regularity of the formation of tachy zones is the accumulation of heavy iron minerals and large tiles articles near the outlets, while the light quartz, micaceous, silicate minerals and the finely dispersed part of tiles cover a considerable distance, forming another technogenic mineral association.

Based on these features of tail accumulation, the process of fractionation by size and density might be considered as the process of enrichment of certain areas with any mineral or fraction of particles. The differences in taillings of ore dressing plants are determined by the type of quartzite mined and by the specific features of the dressing technology. The characteristic differences between the GOK taillings can be attributed to the following: in taillings of the MGOK enrichment, the main iron-containing mineral is hematite (39-53.5%), while for the storages of the GOKK and SGOK it is magnetite (2.9-10.7%); in terms of particle size distribution, the wastes of MGOK enrichment are more finely dispersed (fraction content is -0.044 mm 53%) than LGOK, SGOK (23.5%).

The moisture content of taillings in the explored sections of taillings varies widely. Humidity of taillings increases along the height of the washout with the depth of sampling. In the near-surface stratum at a depth of 1.0 m, humidity varies from 15 to 20% with an average value of 17.5%. At great depths, humidity ranges from 21 – 35%. The porosity coefficient of the stocked tails in the considered area has an average value of 0.74.

During the process of studying the material composition, the following was established: taillings might be used in the construction industry as sand or gravel, to obtain additional products as a main or associated useful component in mining enterprises of KMA, and also used in agriculture and forestry (crop production).

6. CONCLUSIONS

An assessment of the utilization of enrichment waste in agriculture and forestry requires a preliminary study of the mineralogical, chemical, and particle size distribution, agrophysical, physical, and agrochemical properties. The main condition for using waste is the presence of necessary components for the soil. Recommended doses of waste are established on the basis of data from agrochemical analyzes, field experiments, economic and environmental indicators. The use of waste as fertilizers and substances that improve the physical and chemical properties of soils should ensure an increase in the productivity of plants, the quality of their yield, as well as an economic or environmental-social effect.

Waste which was introduced into peat bog soils contributes to the improvement of their hydrothermal regime. Taillings of processing plants can also have an increased absorption capacity due to the presence of residual amounts of clay minerals in them. The factors limiting waste utilization in crop production are reduced to three categories: 1) the presence of concomitant impurities and elements that cause soil pollution and damage; 2) the possibility of re-processing enrichment waste in order to extract the main or related substances; 3) the possibility of disposal in other industries with great economic effect.

Thus, the researching showed that the iron-containing taillings of the KMA GOKs are promising the involvement in development with the allocation of the main and associated useful components, as well as their using as sand, especially since during the reprocessing of iron-containin taillings, energy-intensive operations such as crushing are not necessary and partial grinding [14]. These research results will become a good reserve for conducting experimental studies of the possibility of processing iron-containing taillings of enrichment of KMA GOKs.

The research results will become a good basis for experimental study of the possibility to process iron-containing tailings accumulated from ore-refining and processing facilities in Kursk Magnetic Anomaly (Russia).

7. REFERENCES

1. Kempton, H., Bloomfield, T.A., Hanson, J.L. and Limerick, P., "Policy guidance for identifying and effectively managing perpetual environmental impacts from new hardrock mines", ***Environmental Science & Policy***, Vol. 13, No. 6, (2010), 558-566. doi: 10.1016/j.envsci.2010.06.001
2. Franks, D.M., Davis, R., Bebbington, A.J., Ali, S.H., Kemp, D., and Scurrah, M., "Conflict translates environmental and social risk into business costs". In Proceedings of the National Academy of Sciences of the United States of America, Vol. 111, (2014), 7576-7581. doi: 10.1073/pnas.1405135111
3. Adiansyah, J.S., Rosano, M., Vink, S., Keir, G., "A framework for a sustainable approach to mine tailings management: Disposal strategies", ***Journal of Cleaner Production***, Vol. 108, (2015), 1050-1062. doi:10.1016/j.jclepro.2015.07.139
4. Hekmat, A., Osanloo, M. and Shirazi, A.M., "New approach for selection of waste dump sites in open pit mines", ***Mining Technology***. (2008). 117. 24-31. doi: 10.1179/174328608X343768
5. Bakhtavar, E., Shahriar, K. and Osanloo, M., "Old tailing rehabilitation with regard to environmental impacts at the Mooteh Gold Mine, Iran", In 6th International Conference SGEM 2006, Bulgaria, (2006).
6. Zhang, J.R., Wang, W.Z., Li, F.P. and Wang, A.D., "Comprehensive Utilization and Resources of Metal Mine Tailings", Metallurgical Industry Press. Beijing, (2002).
7. Argimbayev, K.R., Bovdui, M.O. and Mironova, K.V., "Prospects for exploitation of tailing dumps", ***International Journal of Ecology and Development***, Vol. 31, (2016), 117-124.
8. Karu V., "European Union Baltic Sea region project; Min-Novation". ***Oil Shale***, Vol. 28, № 3, (2015), 464-465.
9. Karu V., Valgma I., and Rahe T., "Mining Waste Reduction Methods", In 13th International Symposium PÄRNU 2013 "Topical Problems in the Field of Electrical and Power Engineering" and "Doctoral School of Energy and Geotechnology II", Estonia, (2013).
10. Blight G., "Geotechnical Engineering for Mine Waste Storage Facilities". CRC Press/Balkema, The Netherlands, (2010), 652.
11. Trubetskoy K.N., Chanturia V.A., Kaplunov D.R., and Rylnikova M.V., "Integrated development of deposits and deep processing of mineral raw materials (monograph)", Russian Academy of Sciences, Moscow, (2010). 446.
12. Trubetskoy K.N., "The main directions and ways of solving the problems of resource conservation in the integrated development of mineral resources", ***Mine Surveying and Subsoil Use***, No. 3, (2010), 22-29.
13. Trubetskoy K.N., "Development of science, engineering and technology in the field of integrated development of deposits in the open pit", ***Mining Journal***, No. 3, (2009), 4-7.
14. Trufanov D.V., Leizerovich S.G., and Uskov A.K., "Prospects for the application of non-waste technology in the development of the Lebedinsky deposit of ferruginous quartzites", ***Mining Informational and Analytical Bulletin (Scientific and Technical Journal)***, No. 8, (2000), 163-166.
15. Arkhipov A.N. "Complexity and environmental friendliness of mineral processing technologies", ***Mining Informational and Analytical Bulletin (Scientific and Technical Journal)***, No. 388, (2005), 100-106.
16. Blight, G., Mine Waste: A Brief Overview of Origins, Quantities, and Methods of Storage, In Waste, Trevor M., Letcher, Daniel A. Vallero, (2011), Academic Press: USA, 77-88. doi: 10.1016/B978-0-12-381475-3.10005-1
17. Fomin, S.I, "Foundations for technical solutions in organizing excavation of open ore pits", ***Journal of Mining Institute***, Vol. 221, (2016), 644-650.
18. Fitzpatrick, P., Fonseca, A. and McAllister, M.L., "From the Whitehorse Mining Initiative Towards Sustainable Mining: lessons learned", ***Journal of Cleaner Production***, Vol. 19, No. 4, (2011), 376-384. doi: 10.1016/j.jclepro.2010.10.013
19. Rafkatovich, A.K. and Mironova, K.V., "Methods for the Reduction of Loss and Optimization Processes Open Pit Mining Operations When Mining Man-Made Deposits Formed by Sections", ***Journal of Engineering and Applied Sciences***, Vol. 13, (2018), 1624-1631.
20. Tayebi-Khorami, M., Edraki, M., Corder, G. and Golev, A., "Re-Thinking Mining Waste Through an Integrative Approach Led by Circular Economy Aspirations", ***Minerals***, Vol. 9, (2019), 2-12. doi: 10.3390/min9050286
21. Argimbayev, K. R., Mironova, K. V., Bovdui, M. O., and Podlesnyj, P. V., "Method for forming and developing a technogenic deposit and device for its implementation", RU Patent 2661510, No.d July 17, (2018).

22. Kuskov V.B. and Vasilyev A.M., "Specific features of the concentration process for fine–grained materials in a short–cone hydrocyclone", ***Obogashchenie Rud***, Vol. 2, (2018), 30-34. doi: 10.17580/or.2018.02.06

23. Guezennec, A.G., Bru, K., Jacob, J. and d'Hugues, P., "Co-processing of sulfidic mining wastes and metal-rich post-consumer wastes by biohydrometallurgy", ***Minerals Engineering***, Vol.75, (2015), 45-53. doi: 10.1016/j.mineng.2014.12.033

24. Franks, D.M., Boger, D.V., Côte, C.M. and Mulligan, D.R., "Sustainable development principles for the disposal of mining and mineral processing wastes", ***Resources Policy***, Vol. 36, No. 2, (2017), 114-122. doi: 10.1016/j.resourpol.2010.12.001

25. Morenov, V. and Leusheva, E., "Influence of the Solid Phase\'s Fractional Composition on the Filtration Characteristics of the Drilling Mud", ***International Journal of Engineering-Transactions B: Applications***, Vol. 32, No. 5, (2019), 794-798. doi: 10.5829/ije.2019.32.05b.22

26. Gorelikov, V.G., Lykov, Y.V., Gorshkov, L.K. and Uspechov, A.M.,"Investigation of thermal operational regimes for diamond bit drilling operations (technical note)", ***International Journal of Engineering-Transactions B: Applications***, Vol. 32, No. 5, (2019), 790-793. doi: 10.5829/ije.2019.32.05b.21

27. Argimbaev, K.R. and Kholodjakov, H.A., "Tailings development and their utilization in the National Economy", ***International Journal of Ecology and Development***, Vol.31, (2016), 94-100.

28. Sizyakov, V.M., Kawalla, R. and Brichkin, V.N., "Geochemical aspects of the mining and processing of the large-tonne mineral resources of the hibinian alkaline massif", ***Geochemistry***, (2019). doi: 10.1016/j.chemer.2019.04.002

29. Louwrens, E., Napier-Munn, T., and Keeney, L. "Geometallurgical characterisation of a tailings storage facility - A novel approach", In Tailings and Mine Waste Management for the 21st Century, Sydney, NSW, Australia, Australasian Institute of Mining and Metallurgy, (2015), 125-132.

Persian Abstract

چکیده

هدر رفتن اجتناب‌ناپذیر از منابع معدنی، وخامت مداوم شرایط زمین‌شناسی و معدن برای توسعه ذخایر معدنی و احیای مواد اولیه از زباله‌های معدن با بازیافت، همه مشکلات متعددی هستند که امروزه با آن روبرو هستیم. راه‌حل این مشکل می‌تواند گسترش قابل ملاحظه مواد اولیه خام، کاهش سرمایه‌گذاری در بازیابی از منابع‌های جدید، صرفه‌جویی هزینه برای دورریزهای معدن و رسیدگی به زباله‌های باطله، به دست آوردن اثرات اجتماعی و اقتصادی به دلیل کاهش قابل توجه آلودگی محیط زیست را تضمین کند.. در این مقاله به بررسی مواد دوریز که حاوی آهن در حوضچه‌های نهایی پالایش سنگ معدن و تاسیسات فرآوری‌شده واقع در ناهنجاری مغناطیسی کورسک (KMA GOKs) است، پرداخته شده و نمونه هایی از این مطالعه گرفته شده است. مقاله حاوی نتایج حاصل از مطالعه ترکیب شیمیایی مواد، اجزاء مواد، مطالعه مواد معدنی-پتروگرافی مقاطع انتهایی و صیقل، توزیع اندازه دانه و خصوصیات فیزیکی و مکانیکی نمونه‌ها است. به دلیل تمایز گرانش، مقررات مربوط به تغییر محتوای مؤلفه مفید است. همچنین خاطر نشان شد که به دلیل تجمع پیریت میزان گوگرد در نزدیکی محل تخلیه تفاله‌ها افزایش یافته است. نسبت مواد معدنی سنگ در بافندگی و نسبت ظرافت کسری ماسه اندازه‌گیری شد. بررسی با میکروسکوپ پرتو متمرکز با بزرگنمایی x90 تا x600، و تعیین انواع اندازه و اشکال دانه نشان داد که مواد انتهایی معادن بعد از پردازش اضافی در صنعت ساخت و ساز به عنوان ماسه قابل استفاده می‌باشند.

AIMS AND SCOPE

The objective of the International Journal of Engineering is to provide a forum for communication of information among the world's scientific and technological community and Iranian scientists and engineers. This journal intends to be of interest and utility to researchers and practitioners in the academic, industrial and governmental sectors. All original research contributions of significant value focused on basics, applications and aspects areas of engineering discipline are welcome.

This journal is published in three quarterly transactions: Transactions A (Basics) deal with the engineering fundamentals, Transactions B (Applications) are concerned with the application of the engineering knowledge in the daily life of the human being and Transactions C (Aspects) - starting from January 2012 - emphasize on the main engineering aspects whose elaboration can yield knowledge and expertise that can equally serve all branches of engineering discipline.

This journal will publish authoritative papers on theoretical and experimental researches and advanced applications embodying the results of extensive field, plant, laboratory or theoretical investigation or new interpretations of existing problems. It may also feature - when appropriate - research notes, technical notes, state-of-the-art survey type papers, short communications, letters to the editor, meeting schedules and conference announcements. The language of publication is English. Each paper should contain an abstract both in English and in Persian. However, for the authors who are not familiar with Persian, the publisher will prepare the latter. The abstracts should not exceed 250 words.

All manuscripts will be peer-reviewed by qualified reviewers. The material should be presented clearly and concisely:

- *Full papers* must be based on completed original works of significant novelty. The papers are not strictly limited in length. However, lengthy contributions may be delayed due to limited space. It is advised to keep papers limited to 7500 words.
- *Research* notes are considered as short items that include theoretical or experimental results of immediate current interest.
- *Technical notes* are also considered as short items of enough technical acceptability with more rapid publication appeal. The length of a research or technical note is recommended not to exceed 2500 words or 4 journal pages (including figures and tables).

Review papers are only considered from highly qualified well-known authors generally assigned by the editorial board or editor in chief. Short communications and letters to the editor should contain a text of about 1000 words and whatever figures and tables that may be required to support the text. They include discussion of full papers and short items and should contribute to the original article by providing confirmation or additional interpretation. Discussion of papers will be referred to author(s) for reply and will concurrently be published with reply of author(s).

INSTRUCTIONS FOR AUTHORS

Submission of a manuscript represents that it has neither been published nor submitted for publication elsewhere and is result of research carried out by author(s). Presentation in a conference and appearance in a symposium proceeding is not considered prior publication.

Authors are required to include a list describing all the symbols and abbreviations in the paper. Use of the international system of measurement units is mandatory.

- On-line submission of manuscripts results in faster publication process and is recommended. Instructions are given in the IJE web sites: www.ije.ir-www.ijeir.info
- Hardcopy submissions must include MS Word and jpg files.
- Manuscripts should be typewritten on one side of A4 paper, double-spaced, with adequate margins.
- References should be numbered in brackets and appear in sequence through the text. List of references should be given at the end of the paper.
- Figure captions are to be indicated under the illustrations. They should sufficiently explain the figures.
- Illustrations should appear in their appropriate places in the text.
- Tables and diagrams should be submitted in a form suitable for reproduction.
- Photographs should be of high quality saved as jpg files.
- Tables, Illustrations, Figures and Diagrams will be normally printed in single column width (8cm). Exceptionally large ones may be printed across two columns (17cm).

PAGE CHARGES AND REPRINTS

The papers are strictly limited in length, maximum 6 journal pages (including figures and tables). For the additional to 6 journal pages, there will be page charges. It is advised to keep papers limited to 3500 words.

Page Charges for Papers More Than 6 Pages (Including Abstract)	
For International Author ***	**$55 / per page**
For Local Author	**100,000 Toman / per page**

AUTHOR CHECKLIST

- Author(s), bio-data including affiliation(s) and mail and e-mail addresses).
- Manuscript including abstracts, key words, illustrations, tables, figures with figure captions and list of references.
- MS Word files of the paper.

International **J**ournal of **E**ngineering, P.O. Box 13145-1659, Tehran, I.R. Iran
Materials and Energy Research Center, 5 Ahuramazda St., Alvand Ave., Argentina Sq., Tehran, Islamic Republic of Iran
www.merc.ac.ir - www.emerc.ac.ir

INDEXED IN EMERGING SOURCES CITATION INDEX CLARIVATE ANALYTICS

THOMSON REUTERS
WEB OF SCIENCE

INDEXED IN EMERGING SOURCES CITATION INDEX THOMSON REUTERS

Scopus®

CAS®
A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

ISC
Islamic World Science Citation Center

Crossref

SIS
Scientific Indexing Services

Google
Scholar