International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

Hybridizing the Machine Learning Techniques for Prediction of Sediment Yield

Bezawada Supriya, Arvind Yadav, Guda Navya, Penke Satyannarayana

Abstract: ***Rivers are an integral part of the hydrologic cycle and are the major dynamic geologic agents that play major role for transformation of sediments from land to the oceans. Sedimentation is the biggest problem. Evaluation of suspended sediment yield is an essential parameter under the assessment on Dam filling, protecting of aquatic organism and wildlife habitats, understanding the flood capacity and hydroelectric equipment in hydro-electric power. The assurance of sediment yield through different traditional way isn't exactly correct because of the participation of different complex processes.There are many limitations of traditional methods but it can be overcome by artificial intelligence techniques. So, in this study, the MOGA-ANN (Multi-objective genetic algorithm based artificial neural network) hybrid artificial intelligence method is used to estimate the sediment yield in Krishna river basin, India. The research done for evaluation of the suspended sediment load by taking 20 years of data from Vijayawada, gauging station which is the downward station in Krishna river. The proposed MOGA-ANN model provided low root mean square error (0.03354) and high correlation coefficient (0.9214) during test phase. It exhibited satisfactory performance.***

Keywords: ***Water discharge; Artificial neural network; Multi-objective genetic algorithm; Suspended sediment yield; Krishna River.***

I. INTRODUCTION

Rivers are geological agents which transfers materials from continents to the ocean. Estimated sediment load provide the information of the mass balance between the land surface and ocean. In the river beds, sediment occur and that sediment disturb the organic food cycle and destroy the environment of the aquatic organisms which live under river causes huge failures in the fish populations. Sediment results in odour and taste problem in drinking water, nutrient cycle is also affected by these. While constructing of dams and reservoirs, it is important to estimate the sediment yield so that we can protect things like fish and wildlife organisms, hydroelectric equipment durability, watershed management and environmental impact. Measuring of suspend sediment yield is difficult because it needs more money and time. The sediment yield is affected by different number of variables and they are extremely nonlinear and complicated. Thus estimation of this is very difficult by traditional methods. Artificial intelligence techniques and mathematical models were used successfully by various researchers to solve complex non-linear worldwide problems in hydrology and other domains (Kebede et al. 2017;Yadav et al. 2018; Patel et al. 2018; Patel et al. 2019;Yadav 2019a; Yadav and Satyanarayana 2019b). Various researchers have been used artificial intelligence for sediment load estimation in river basin system (Yadav et al 2018). Recently, Artificial intelligence techniques are widely used by various researchers in various domain (Yadav et al. 2018;Patel et al. 2018, Patel et al. 2019, Yadav and Satyanarayana 2019b).Artificial Neural Network (ANN) is used for better result with accuracy in estimation of suspended sediment yield. The ANN is a computational based model in accordance with the structure and functions of biological neural network. The ANN understands in a sense based upon the input and output parameters. This is measured as nonlinear statistical data modelling devices where as the complex relationships among input and output patterns are found. The ANN is used by different researchers to estimate the sediment in rivers (Yadav et al. 2018). It takes data samples instead of the complete data sets to appear at solutions, it saves both time and money. The genetic algorithm (GA) is applied successfully by various researchers for model parameters selection of the ANN for overcome the limitation of parameters selection for ANN model. Mean square error (MSE) of ANN is affected by the bias and variance which contradict each other. It leads to a huge uncertainty in evaluation. Recently few studies have been carried out on estimation of sediment using multi objective optimization. ANN, GA and mathematical methodology are used for estimation and predicting of rainfall runoff, stream flow and flood in the Krishna river (Yadav et al. 2017). There are very few studies noticed on sediment yield which estimate in the Krishna river. In this paper, all ANN parameters are selected simultaneously by optimized bias and variance for estimating sediment yield efficiently using multi-objective genetic algorithm based artificial neural network (MOGA-ANN). Artificial neural networks are utilized in a broad range of various learning data applications and both the input–output correlations of non-linear processes. ANNs are used for hydrologic studies and those have been incorporated in the time series predictions of runoff or flow ,

Revised Manuscript Received on February 28, 2020.
* Correspondence Author
Arvind Yadav*, Department of Electronics and Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram-522502, (A.P.), India. Email: arvindyadav@kluniversity.in
Bezawada Supriya, Department of Electronics and Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram-522502, (A.P.), India. Email:160050027@kluniversity.in
Guda Navya, Department of Electronics and Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram-522502, (A.P.), India. Email:160050074@kluniversity.in
Penke Satyannarayana, Department of Electronics and Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram-522502, (A.P.), India. Email:satece@kluniversity.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

water table management and to estimate the runoff hydrograph parameters, water quality management, reservoir operation optimization, non-point source contamination, evaluating water quality parameters, sediment prediction and there are many other various applications in marine resources. No any study is found for estimating sediment load using multi objective based hybridising artificial intelligence techniques in Krishna river basin, India. So in this study, the multi objective genetic based artificial neural network has proved as it is useful artificial intelligence models for non-linear modelling applications or complex hydrologic systems for estimating sediment load in Krishna river basin, India.

II. STUDY AREA

In this paper Krishna River was selected for the study of suspended sediment yield estimation. India's fourth largest river basin having 70,614 sq.km total catchment area which accounts for 8% of the entire geographical region of India. The sewage area of 71,417 km^2 is completely surrounded by Vijayawada gauge station of the river. Krishna River originates at a height of 870m above the sea level near Hamsaladeevi, Krishna district, Andhra Pradesh. Krishna river remains with in the Western Ghats nearby the village Mahabaleshwar and the entire length of the total river from Mahabaleshwar to the Bay of Bengal is 1400 km, The Krishna basin is present in between the latitudes and longitudes(north latitude 13007' to 19020' and east longitudes 73022' to 81010'). The longitude and latitude of Krishna district is 80.7214° and 16.6100° respectively. Around t 90% of annual rainfall is collected in Krishna district during the Monsoon period of time. The maximum temperature noted in all parts varies from 30 C and 35 C. The Krishna Basin covers the states of Andhra Pradesh, Telangana (29.81%), Maharashtra (26.36%) and Karnataka (43.8%). The Krishna River elevates from the Western Ghats nearby the Jor village of Satara district in Maharashtra and outfall into the Bay of Bengal.

III. METHODOLOGY

ANNs are separate into recurrent networks and feed forward in according with the direction of information flow and processing. The multi layer perceptron (MLP) comes under feed forward network. In the current study MLP consist of different types of layers they are output layer , hidden layer and an input layer neuron. The data in each neuron from input way to output way will flow only in single direction and those neurons are which are present in the layers are been linked to all other neurons only in adjoining layer . The input layer neurons are represented as input variables. The number of transfer function, neurons, connection and bias weight are optimised by multi objective genetic algorithm. In this paper the networks are been considered and those networks are been trained by using Back-Propagation algorithm(BP). The BP has an information processing in both directions, one is feed forward of the input information and the other is back-propagation of the error

The training algorithm is done using Lavenberg Marquardt back propagation method (LM) for estimation of sediment yield. The poor selection of artificial neural network learning factors like combination coefficient of LM, and topology of network chosen which have size of hidden node, number of hidden layers and initial weight etc by trial and error approach which caused overfitting and underfitting problems. Optimization is important for these parameters of ANN. Genetic algorithm is based on population this is global search optimization algorithm which was invented by Darwin's theory of evolution, and this theory is used to estimate the best factors for Artificial Neural Network models. The ANN parameter and topology are been selected according to the optimizing of two conflicting responses bias and variance by MOGA-ANN model.

The main aim of using MOGA-ANN for four main factors of ANN models are especially neurons in hidden layer, connection and bias weights, combination coefficient, and transfer function. In this study, crossover rate 0.6 and mutation probability was utilized as 0.05. The greatest number of generation stopping based utilized for this study is 100. In this study maximum population size is 50. In this study, suspended sediment data ,water discharge and water level data during 1970-2015 are taken from Vijayawada gauge station to develop the MOGA-ANN model.

The data division and data normalization are required to solve the dissimilar units and dissimilar ranges between input data and output data, before neural networks training. To preserve the uniqueness in training phase and to remove dimension of variable this Data normalization is been used .The quick convergence and processing are done during the training those use to predict less errors. For normalization of data the between 0 and 1 has been done (Yadav et al. 2017, 2018). In this research Normalization is been done and checked for all output and inputs variables. The data is separated into different models which are testing, validation, and training. These are used for developing robust and generalised parameters for estimation of the sediment yield. For neural network training purposes the Training data is used, where as validation data is used for avoiding over fitting. Testing data is used as invisible data in this process model performance is tested. 64% of data includes training process and 36% of data is equally distributed among testing processes and validation. validation and Testing must have same properties it is used to prevent under and over estimation problems.

IV. RESULTS AND DISCUSSION

The MATLAB software is used for development of hybrid MOGA-ANN model. The MOGA-ANN model optimized the bias, variance with best chromosome and ANN parameter are selected accordingly. It is found that there are poptimum 26 hidden neurons. The S-shaped curve is known as tan sigmod transfer function, which is monotonically increasing and it was choosed as optimum transfer function for both the output layers and hidden layer. The Levenbarg LM algorithm model had selected 0.003 by the MOGA-ANN model by using the values of combinational coefficient(μ). The trade-off among the bias and variance of Pareto front, variation of crowding distance with individuals ranking of individuals and average spread against generation technique in the training process as it was shown in Figure 2.

Generalization capability and the performance evaluation of the MOGA-ANN model was performed with the test data set. To assess the exactness of the model different error statistical parameters for testing, validation and training data are used and those are been given in Table 1. It is observed that root mean square error (RMSE) is very low (0.029254-0.050559), and correlation coefficient (r) is very high (0.901365-0.992765) for validation, training, testing. Estimation of suspended sediment yield is done by MOGA-ANN. Amongst validation, training, testing data parameters has been recognized that developed model have more generalization capability. The Mean absolute error (MAE) and RMSE values are same during testing phase and validation phase. It can able to say that this model is not over-fitted

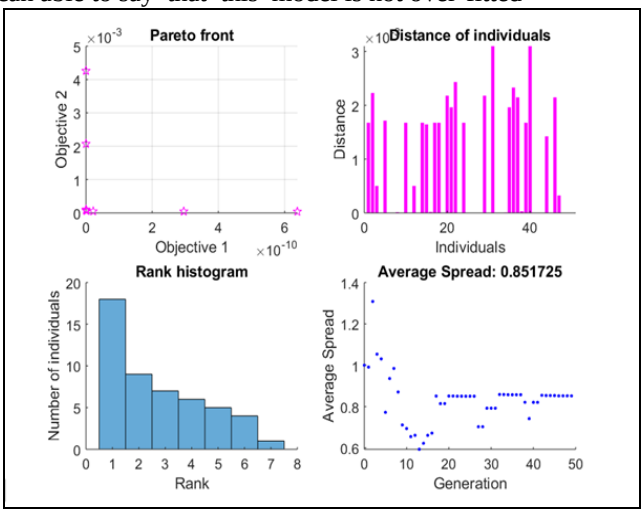

Figure 2.Variation of MOGA-ANN model parameters during training phase

Table 1. Training, validation and testing data error statistics of MOGA-ANN model

Data	RMSE	MSE	MAE	VAR	R
Validation	0.029254	0.000856	0.009224	0.000808	0.992765
Testing	0.03354	0.001125	0.012983	0.001148	0.921407
Training	0.050559	0.002556	0.016019	0.002571	0.901365

Moreover the quantitative evaluation using statistical measure, the efficiency of ANN technique in estimation of the suspended sediment yield is also assessed by graphical indicators. The comparisons between observed and modelled sediment yield during training, validation and testing phase are shown in form of hydrograph and scatter plots (Figure 3, 4, 5, 6, 7 and 8). The hydrographs indicate that the modelled sediment yield follow the variation in the observed data. During testing phase, it is shown that that the MOGA-ANN results are nearer to the observed sediment yield values(Figure 7 and Figure 8). Also, it is noticed that the model is unable to capture the low sediment yield as it is very clear in the scatter plot where negative value is predicted during low sediment data. Among all the that hydrographs, it is found that estimated sediment yield using the estimated MOGA-ANN model is closest to the observed data. This is also confirmed by scatter plot.

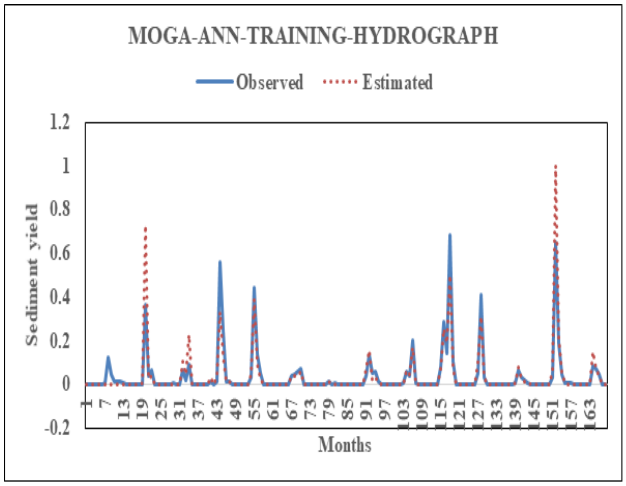

Figure 3. Hydrograph between estimated and observed sediment yield in testing the data by MOGA-ANN prediction model in training phase.

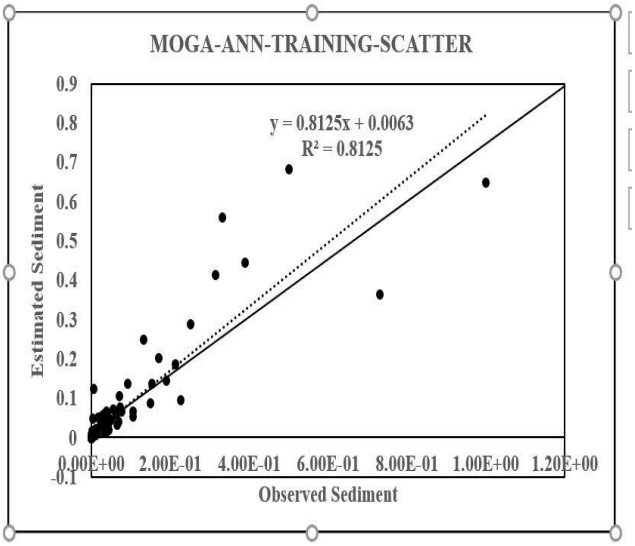

Figure 4. Scatter between estimated and observed sediment yield of testing the data by MOGA-ANN prediction model in training phase.

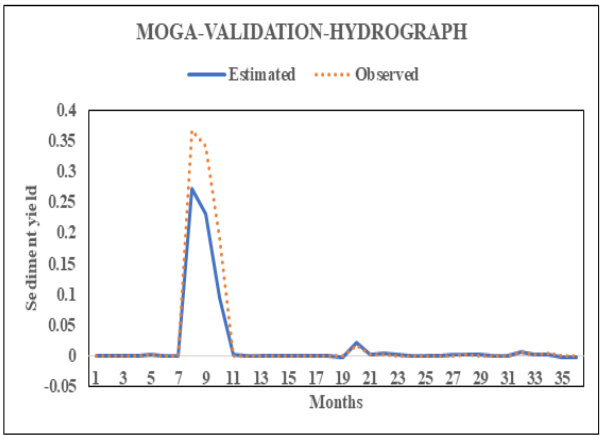

Figure 5. Hydrograph between estimated and observed sediment yield of testing the data by MOGA-ANN prediction model in validation phase

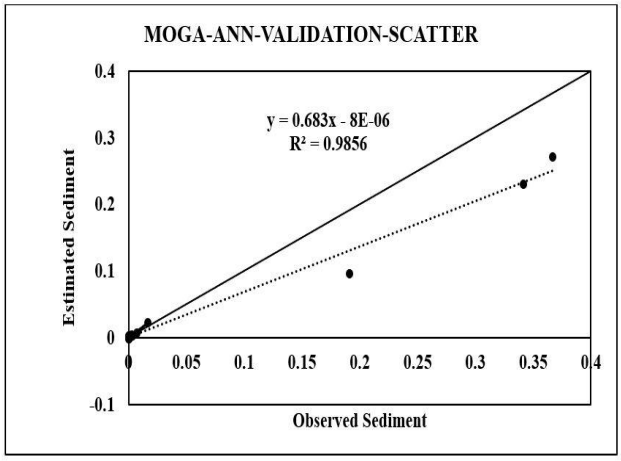

Figure 6. Scatter between estimated and observed sediment yield of testing the data by MOGA-ANN prediction model in validation phase.

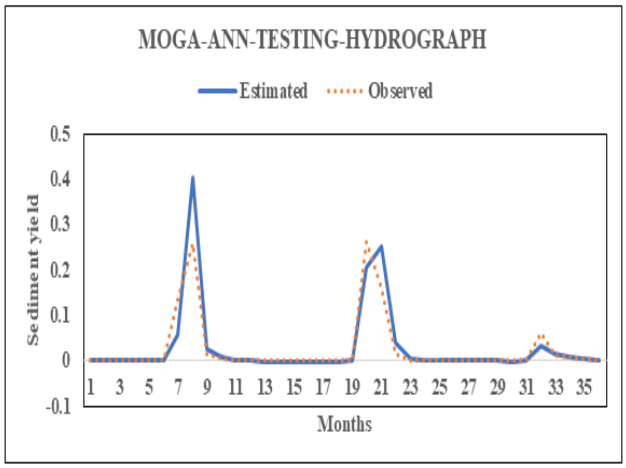

Figure 7. Hydrograph between the estimated and observed sediment yield of testing the data by MOGA-ANN prediction model in testing phase

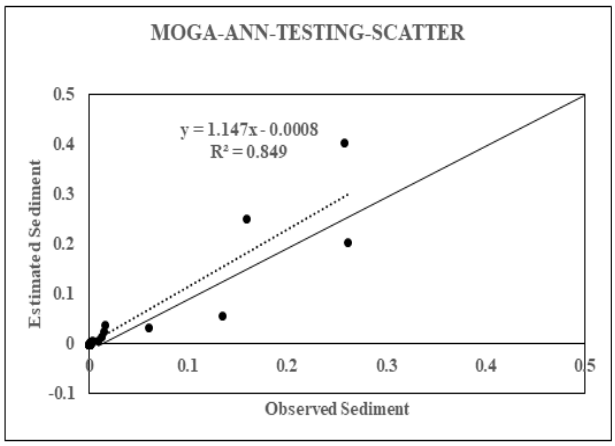

Figure 8. Scatter the estimated and observed sediment yield of testing the data by MOGA-ANN prediction model in testing phase

In this scatter plot, it can be observed that the ANN results are nearer to the 1:1 line and the data points are dispersed around the 1:1 line except the above mentioned gauging stations(Figure 4, 6 and 8) during, training, validation and testing phase. The 1:1 line is the bisector line (45 degree line), where both expected and observed values are similar. If scatter points lies closely along this line then expected and observed values are the approximately similar, that indicates that model is perfect. Also, it is observed that the model is incapable to capture the lower sediment yield and it is shown clearly in the figure where negative value is estimated during low suspended sediment data.

VI. CONCLUSION

By using different inputs of hydrological variables (water discharge and water level), present work reveals the evaluation of suspended sediment yield by hybrid MOGA-ANN approach at Vijayawada gauge station in Krishna river basin. For estimation of suspended sediment yield different ANN structure was developed by optimizing bias and variance in the MAGA-ANN model. The MOGA-ANN model is provided the satisfactory estimation for high, low and medium values of the suspended sediment yield. Implementation of ANN with multi objective genetic algorithm conclude the availability and flexibility model for handle the nonlinear relationships which are been present in between the output and input parameters for suspended sediment yield estimation. At last the observed magnitude values by MOGA-ANN model consist of upper ,medium and lower Suspend sediment yield are very nearer to the values which are observed. The MOGA-ANN estimated sediment yield and the observed suspended sediment yield are similar. This model is very efficient because it consist of low root mean square error and high value about correlation coefficient.

REFERENCES

1. Kebede, M., Chakravarti, A. and Adugna, T., 2017. Stream flow and land use lands cover changes in Finchaa Hydropower, Blue Nile river basin, Ethiopia" International Journal of Civil, Structural, Environmental, Infrastructures Engineering, Research & Development (IJCSEIERD), 7, 1-12, ISSN:2249-7978.
2. Patel, A.K., Chatterjee, S. and Gorai, A.K., 2018. Development of an expert system for iron ore classification. Arabian Journal of Geoscience, 11(15), p.401.
3. Patel, A.K., Chatterjee, S. and Gorai, A.K., 2019. Effect on the Performance of a Support Vector Machine Based Machine Vision System with Dry and Wet Ore Sample Images in Classification and Grade Prediction. Pattern Recognition and Image Analytics, 29(2), pp.309-324.
4. Yadav, A., Chatterjee, S., Equeenuddin, S.M., 2017. Prediction of Suspended Sediment Yield by Artificial Neural Network and Traditional Mathematical Model in Mahanadi River Basin, India, Journal of Sustainable Water Resource Management, 4(4), 745-759.
5. Yadav, A., Chatterjee, S. and Equeenuddin, S.M., 2018. Suspended sediment yield estimation using genetic algorithm-based artificial intelligence models: case study of Mahanadi River, India. Hydrological Sciences Journal 63:8, 1162-1182.
6. Yadav, A. and Satyanarayana, P.. 2019b. Multi-objective genetic algorithm optimization of artificial neural network for estimating suspended sediment yield in Mahanadi River basin, India. International Journal of River Basin Management, pp.1-21. DOI:10.1080/15715124.2019.1705317.
7. Yadav, A., 2019a. Estimation and Forecasting of Suspended Sediment Yield in Mahanadi River Basin: Application of Artificial Intelligence Algorithms, Ph.D. Thesis, National Institute of Technology, Rourkela, India.

AUTHORS PROFILE

Bezawada Supriya: I am a student of KL University studying final year belongs to Department of ECM , B-Tech. I have done my specification in Artificial Intelligence Technologies. I have worked on application of artificial intelligence in river basin system as my project.

Dr. Arvind Yadav: I am Dr. Arvind Yadav, Assistant Professor, Electronics and Computer Engineering Department, KL university Vijayawada. I have completed Ph.D. from National Institute of Technology (NIT) Rourkela. Artificial Intelligence Techniques are applied in interdisciplinary fields. Recently, I have published the "Prediction of suspended sediment yield by artificial neural network and traditional mathematical model in Mahanadi river, India" paper in Springer. Recently, I have also published SCI free journal paper "Suspended Sediment Yield Estimation using Genetic Algorithm-based Artificial Intelligence Models in Mahanadi River " of impact factor 2.3 in Taylor & Francis and others in river basin system. Many other papers which are based on application of artificial intelligence based techniques in interdisciplinary domains are under review in SCI journals. I have qualified GATE exam.

Guda Navya: I am a student of KL University studying final year belongs to Department of ECM , B-Tech. I have done my specification in Artificial Intelligence Technologies. I have worked on application of artificial intelligence in river basin system as my project.

Dr. Penke Satyannarayana: Professor, Electronics and Computer Science, Koneru Lakshmaiah University, Vaddeswaram, India. Ph.D JNTU. AP, India.