Prediction of Suspended Sediment Concentration Using ANFIS with the Bacterial Foraging Optimization Algorithm

Jinsheng Fan [1,2,3,4,*], Qiushi Luo [2,3,4], Yuchuan Bai [1]

[1] Institute for Sediment, River and Coast Engineering, Tianjin University, Tianjin 300072, China
[2] Yellow River Engineering Consulting Co. Ltd, Zhengzhou 450003, China
[3] Key Laboratory of Water Management and Water Security for Yellow River Basin, Ministry of Water Resources, Zhengzhou 450003, China
[4] Yellow River Engineering Consulting Co. Ltd, Zhengzhou 450003, China

The email of the corresponding author: fanjs.16b@igsnrr.ac.cn

Abstract. This study combined ANFIS with a bacterial foraging optimization algorithm (ANFIS-BFO) to predict the daily suspended sediment concentration based on the daily series data observed at the Rio Valenciano hydrological station near Puerto Rico, USA. Meanwhile, ANFIS with grid partition (ANFIS-GP), ANFIS with subtractive clustering (ANFIS-SC), ANFIS with fuzzy c-means clustering (ANFIS-FCM), artificial neural network (ANN), and the sediment rating curve (SRC) are utilized for the prediction of the same flow discharge-suspended sediment concentration (SSC) daily series data. The root mean square error (RMSE), mean root square error (MRSE), and coefficient of determination (R2) were adopted as the evaluation indicators of the prediction performance of each model. According to the different settings of the input and output variables, the predictions for four different scenarios were carried out. The comparative analysis results show that we can gain the best prediction results when the current day's flow discharge is used as the input and the current day's SSC is used as the output for the hydrological station in the study area. For the Rio Valenciano Station, the MRSE value of the ANFIS-BFO, ANFIS-FCM, ANFIS-GP, ANFIS-SC, ANN, and SRC is, respectively, 2.2172, 2.5389, 2.6627, 2.7549, 2.7994 and 3.7882. It can be inferred that ANFIS-BFO embodies better prediction results than all other models. ANFIS-SC and ANFIS-FCM have slightly superior prediction performance to ANFIS-GP. ANFIS-GP, ANFIS-SC, and ANFIS-FCM have slightly superior prediction performance to ANN.

Keywords: Bacterial foraging optimization algorithm, ANFIS, ANN, suspended sediment, modeling

D. Li et al. (eds.), *Proceedings of the 2023 9th International Conference on Architectural, Civil and Hydraulic Engineering (ICACHE 2023)*, Advances in Engineering Research 228,
https://doi.org/10.2991/978-94-6463-336-8_23

1 Introduction

Great emphasis has been attached to increasing the accuracy of the river suspended sediment concentration (SSC) prediction due to its remarkable impacts on the reservoir's functional life evaluation, river geomorphological evolution analysis, riverbed stability analysis, and river ecological environment evaluation[1]. The underestimation of sediments may cause excessive silting of deposits in the reservoir or channel, affecting the reservoir's useful life and riverbed stability, thus threatening the life and property safety of people living along the rivers. Contrarily, an overestimation of sediments could cause the unreasonableness of reservoir design and further undermine the economic benefits of the reservoir. Moreover, residues could also affect the ecological environment along the rivers, particularly when the sediments contain pollutants[2]. The earliest methods for predicting river SSC are empirical ones, whose prediction precision can hardly be guaranteed due to the complexity of input parameters[3-4]. Therefore, to enhance river SSC prediction accuracy, water resource scientists and hydrological engineers have devoted themselves to finding new prediction methods.

Among all empirical approaches, SRC has been widely used to predict the river SSC[5]. In recent years, as the development of computer science enables modeling of data with nonlinearity and allows convenient model operation, artificial neural network (ANN) and adaptive neural fuzzy inference network (ANFIS) have become two mainstream intelligent prediction models for hydrological forecasts, such as prediction of SSC[6-12].

To boost their performances, ANN and ANFIS are coupled with different data processing methods and intelligent optimization algorithms[13-19].This paper aims to construct a new model (ANFIS-BFO) combining ANFIS and the bacterial foraging optimization algorithm for SSC prediction with the flow- SSC daily series data observed at Rio Valenciano and Quebrada Blanca hydrological stations near Puerto Rico in the USA. In addition, the ANFIS model with three different clustering methods, including ANFIS-GP, ANFIS-SC, ANFIS-FCM, as well as ANN and SRC, are employed in this study for modeling prediction and compared with the results of ANFIS-BFO. This paper includes five sections. Section two illustrates the methods, including ANN, ANFIS, BFO, and ANFIS-BFO. Section three summarizes the situation of the study area in this paper and the statistical analysis of the data. Section four demonstrates the prediction results of various models and comparative analysis results. The last section is the conclusion.

2 Case study

The flow discharge and SSC daily series data observed at the Rio Valenciano hydrological station of the USGS at and near Jaguar are utilized in this study. The catchment area where the hydrological station is located covers 43.57 km^2, and the gauging datum is 98 m above sea level. The flow discharge and SSC daily series data observed at the hydrological station are retrieved from the website of the USGS. The data used to train the models are from January 1, 1994, to December 31, 1994. The data from January 1, 1995, to December 31, 1995, are adopted for testing. Table 1 lists the statistical parameters for discharge and SSC of the Rio Valenciano Station. S_x means standard deviation, x_{mean} represents the average, C_v is coefficient of variation, and C_{sx} is coefficient of deviation. According to Table 1, we notice that the flow and sediment data manifest a distribution of high skewness. All statistical parameters suggest a complex nonlinear relation between flow discharge and SSC.

Table 1. Statistical parameters of daily data observed at the Rio Valenciano Station

Data Set	Station	Basin Area (km^2)	Data Type	x_{mean}	S_x	C_v	C_{sx}	x_{max}	x_{min}
Training	Rio Valenciano	43.57	Flow (m^{3}s^{-1})	0.55	2.01	3.65	14.44	35.113	0.0396
			Sediment (mgl^{-1})	40.56	109.5	2.70	7.44	1200	2
Testing	Rio Valenciano	43.57	Flow (m^{3}s^{-1})	1.05	2.40	2.29	5.97	24.636	0.0510
			Sediment (mgl^{-1})	68.56	141	2.06	4.44	1090	2.5

3 Application and discussion

3.1 Methodology

This study uses different combinations of the current day flow discharge Q_t, previous day flow discharge Q_{t-1}, and previous day SSC S_{t-1} as inputs to train MLP, RBNN, GRNN, ANN-IGA, and previous ANN-PSO to evaluate the current day SSC. The input combinations of the SSC prediction model for these two hydrological stations include respectively i) Q_t; ii) Q_t, Q_{t-1}; iii) Q_t, S_{t-1}; and iv) Q_t, Q_{t-1}, S_{t-1}.

RMSE, MRSE, and R^2 were used as criteria for model prediction performance. Their respective calculation formulas are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left[S_{t_{measured,i}} - S_{t_{predicted,i}}\right]^2} \quad (1)$$

$$MRSE = \frac{1}{N}\sqrt{\sum_{i=1}^{n}\left[S_{t_{measured,i}} - S_{t_{predicted,i}}\right]^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left[S_{t_{measured,i}} - S_{t_{predicted,i}}\right]^2}{\sum_{i=1}^{N}\left[S_{t_{measured,i}} - \bar{S}_{t_{measured,i}}\right]^2}, \bar{S}_{t_{measured,i}} = \sum_{i=1}^{N}\frac{1}{N}S_{t_{measured,i}} \quad (3)$$

where N refers to the quantity of specimen. $S_{t_{measured,i}}$ and $S_{t_{predicted,i}}$ refer to the actual measured and predicted value for the i-th specimen, respectively.

3.2 Results

The prediction results of all models are shown in Table 2. We can note that, for all combinations, the performance of ANFIS-BFO is better than that of three unoptimized ANFIS models. In addition, ANFIS-SC and ANFIS-FCM exhibit higher performance than ANFIS-GP, and ANFIS-FCM is slightly better than ANFIS-SC. It can also be noted that all intelligent models display more robust prediction performance than the traditional SRC. Regarding the input combination i, the MRSE values of the ANFIS-BFO, ANFIS-FCM, ANFIS-SC, ANFIS-GP, ANN, and SRC are respectively 2.2172, 2.5389, 2.6627, 2.7549, 2.7994 and 3.7882; the RMSE values are respectively 42.3601, 48.5063, 50.2009, 52.6325, 53.4817 and 88.5519.

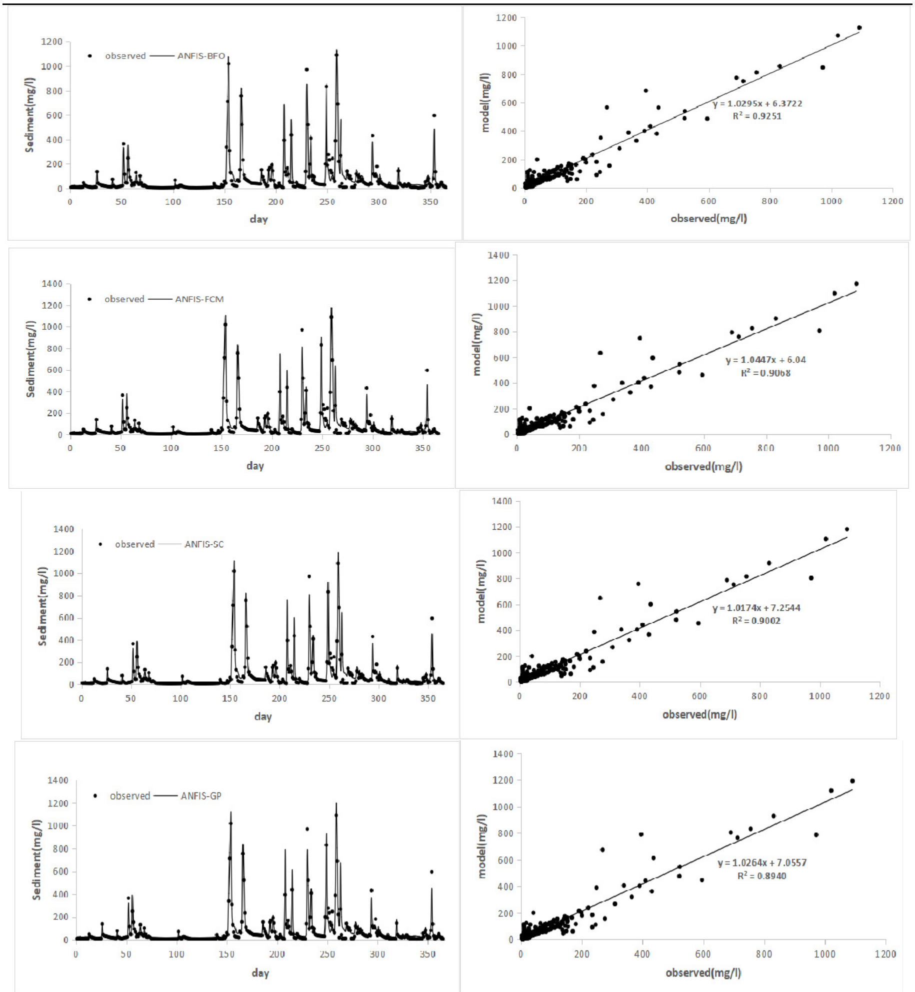
Figure 1 presents the comparison diagrams and scatter plots about the SSC predicted value and measured values of all models under input combination i. We could infer from the scatter plot and the R^2 value of the fitting curve of each model that ANFIS-BFO embodies superior prediction performance to all other models. The R^2 value is 0.9251. At the same time, SRC demonstrates the worst prediction results, with the R^2 value at 0.8304. The prediction performances of ANFIS-FCM, ANFIS-SC, ANFIS-GP, and ANN are in between, with the R^2 values at 0.9068, 0.9002, 0.8940, and 0.8822. As shown in the SSC prediction-measured value comparison diagrams of each model, the prediction values of ANFIS-BFO more closely approximate the measured values than all other models, particularly for high values (>500 mg/l).

Table 3 lists the errors of the peak prediction value (>500 mg/l) of all models. As shown in the table, regarding all input combinations, ANFIS-BFO manifests better performance for peak value prediction than all other models. Concerning the input combination i, the prediction value of ANFIS-BFO for the maximum peak of 1090 mg/l is 1127 mg/ll, overestimating by 3.4%. Comparatively, the prediction values of ANFIS-FCM, ANFIS-SC, ANFIS-GP, ANN, and SRC are respectively 1171 mg/l, 1181.2 mg/l, 1192.6 mg/l, 987.45 mg/l, and 874 mg/, suggesting that ANN and SRC register an underestimation by 9.4% and 19.8%. ANFIS-FCM, ANFIS-SC, and ANFIS-GP embody an overestimation of 7.43%, 8.37%, and 9.41%. With respect to the second maximum peak of 1020 mg/l, the prediction value of ANFIS-BFO is 1070.47, with an overestimation of 4.95%. The prediction values of ANFIS-FCM, ANFIS-SC, ANFIS-GP, ANN, and SRC are respectively 1098, 1104.2, 1119.2, 967.16, and 674.1. ANN and SRC underestimate the value by 5.1% and 33.9%, and ANFIS-FCM, ANFIS-SC, and ANFIS-GP overestimate the value by 7.65%, 8.25%, and 9.73%.

Table 2. Comparison of testing performances of all models for Rio Valenciano station

Model Inputs	ANFIS-BFO			ANFIS-FCM			ANFIS-SC		
	MRSE	RMSE	R^2	MRSE	RMSE	R^2	MRSE	RMSE	R^2
Q_t	2.2172	42.3601	0.9251	2.5389	48.5053	0.9068	2.6627	50.2009	0.9002

Q_t and Q_{t-1}	2.3321	43.2031	0.9107	2.5423	48.8210	0.9031	2.6951	50.4542	0.8987
Q_t and S_{t-1}	2.3731	43.8123	0.9034	2.5843	48.9521	0.9032	2.6861	50.7123	0.8979
Q_t , Q_{t-1} and S_{t-1}	2.4125	44.2132	0.9012	2.6334	49.0234	0.9021	2.7031	51.2642	0.8997
	ANFIS-GP			ANN			SRC		
Q_t	2.7549	52.6325	0.8940	2.7994	53.4817	0.8822	3.7882	88.5519	0.8304
Q_t and Q_{t-1}	2.7913	52.871	0.8987	2.8351	54.1542	0.8762			
Q_t and S_{t-1}	2.7842	52.912	0.8869	2.8561	55.0123	0.8654			
Q_t , Q_{t-1} and S_{t-1}	2.7715	52.220	0.8969	2.9431	56.5642	0.8465			



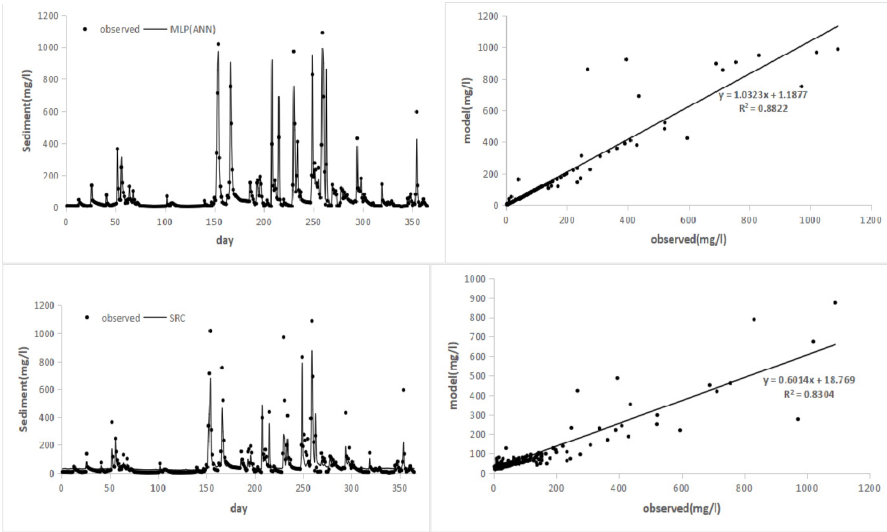


Fig. 1. Observation values and prediction values of each model in the testing stage for Rio Valenciano station

Table 3. Comparison of prediction peak values of all models for Rio Valenciano station

Observed Sediment Peaks	ANFIS-BFO	ANFIS-FCM	ANFIS-SC	ANFIS-GP	ANN	SRC	Relative Errors (%)					
							ANFIS-BFO	ANFIS-FCM	ANFIS-SC	ANFIS-GP	ANN	SRC
1090	1127	1171	1181.2	1192.6	987.45	874	3.4	7.43	8.37	9.41	-9.4	-19.8
1020	1070.47	1098	1104.2	1119.2	967.16	674.1	4.95	7.65	8.25	9.73	-5.1	-33.9
971	847.6	807	802	786	751	276	-12.7	-16.88	-17.4	-19.1	-22.6	-71.5
831	855	901	918	929	948	789.3	2.9	8.4	10.5	11.8	14.1	-5
755	811.7	824.3	815.1	832.4	904	461.3	7.5	9.2	7.96	10.25	19.7	-38.9
712	750.7	759.3	752.7	764.9	855.12	419.6	5.43	6.64	5.71	7.42	20.1	-41
690	774	792.7	787	804.7	895.8	451.4	12.2	12.9	14.1	16.6	29.8	-34.6
595	485.9	461.5	452.8	446	424	220	-18.3	-22.4	-23.8	-25	-28.7	-63

4 Conclusion

We put forward an ANFIS-BFO model utilizing the bacterial foraging optimization algorithm to seek the optimal structural parameters for ANFIS. ANFIS-BFO is built and trained with the flow discharge and SSC series data observed in the study area. A comparison of the three indicators of all models reveals that ANFIS-BFO has superior prediction performance to all other models. ANFIS-SC and ANFIC-FCM demonstrate slightly better prediction performance than ANFIS-GP. Three ANFIS models are slightly superior to ANN. The prediction peak values of all input combinations among the models are compared. The results suggest that ANFIS-BFO embodies a more robust prediction performance than all other models. The R2 of each model in terms of all combinations is compared, disclosing that the prediction values of ANFIS-BFO more closely approximate the actual values than all other models.

In this study, BFO is used to seek the optimal structural parameters of ANFIS. Other optimization algorithms, such as the differential evolution algorithm, flower pollination algorithm, and tabu search algorithm, may replace BFO for refining the optimal structural parameters of ANFIS. The comparison between the results of this study and the prediction performances of other algorithms for seeking the optimal structural parameters of ANFIS needs to be further investigated.

Acknowledgements

This research was financially supported by the National Key Research and Development Program of China (2016YFC0402503) and the National Natural Science Foundation of China (Grant Nos. 51979185,41576093).

References

1. Sivakumar, B., Jayawardena, A.W., 2002. An investigation of the presence of low-dimensional chaotic behaviour in the sediment transport phenomenon. Hydrol. Sci. J.-J. Sci. Hydrol., 47(3): 405-416. DOI:10.1080/02626660209492943
2. Kisi, O., 2013. Evolutionary neural networks for monthly pan evaporation modeling. J. Hydrol., 498: 36-45. DOI: 10.1016/j.jhydrol.2013.06.011
3. Ackers, P., White, W.R., 1973. Sediment transport: new approach and analysis. Journal of the Hydraulics Division, 99(11): 2041-2060. DOI: https://doi-org.xjpgl.80599.net/10.1061/JYCEAJ.0003791
4. Wu, B., van Maren, D.S., Li, L., 2008. Predictability of sediment transport in the Yellow River using selected transport formulas. International Journal of Sediment Research, 23(4): 283-298. DOI: https://doi.org/10.1016/S1001-6279(09)60001-9
5. Heng, S., Suetsugi, T., 2014. Comparison of regionalization approaches in parameterizing sediment rating curve in ungauged catchments for subsequent instantaneous sediment yield prediction. J. Hydrol., 512: 240-253. DOI: https://doi.org/10.1016/j.jhydrol.2014.03.003
6. Kisi, O., 2005. Suspended sediment estimation using neuro-fuzzy and neural network approaches/Estimation des matières en suspension par des approches neurofloues et à base de réseau de neurones. Hydrological Sciences Journal, 50(4). DOI: https://doi-org.xjpgl.80599.net/10.1623/hysj.2005.50.4.683
7. Alp, M., Cigizoglu, H.K., 2007. Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. Environ. Modell. Softw., 22(1): 2-13. DOI: https://doi.org/10.1016/j.envsoft.2005.09.009
8. Tayfur, G., Karimi, Y., Singh, V.P., 2013. Principle Component Analysis in Conjuction with Data Driven Methods for Sediment Load Prediction. Water Resour. Manag., 27(7): 2541-2554. DOI:10.1007/s11269-013-0302-7
9. Afan, H.A. et al., 2015. ANN Based Sediment Prediction Model Utilizing Different Input Scenarios. Water Resour. Manag., 29(4): 1231-1245. DOI:10.1007/s11269-014-0870-1
10. Kisi, O., Zounemat-Kermani, M., 2016. Suspended Sediment Modeling Using Neuro-Fuzzy Embedded Fuzzy c-Means Clustering Technique. Water Resour. Manag., 30(11): 3979-3994. DOI:10.1007/s11269-016-1405-8

11. Buyukyildiz, M., Kumcu, S.Y., 2017. An Estimation of the Suspended Sediment Load Using Adaptive Network Based Fuzzy Inference System, Support Vector Machine and Artificial Neural Network Models. Water Resour. Manag., 31(4): 1343-1359. DOI:10.1007/s11269-017-1581-1
12. Kisi, O., Yaseen, Z.M., 2019. The potential of hybrid evolutionary fuzzy intelligence model for suspended sediment concentration prediction. Catena, 174: 11-23. DOI: 10.1016/j.catena.2018.10.047
13. Kisi, O., Ozkan, C., Akay, B., 2012. Modeling discharge-sediment relationship using neural networks with artificial bee colony algorithm. J. Hydrol., 428: 94-103. DOI: 10.1016/j.jhydrol.2012.01.026
14. Asadi, S., Shahrabi, J., Abbaszadeh, P., Tabanmehr, S., 2013. A new hybrid artificial neural networks for rainfall–runoff process modeling. Neurocomputing, 121: 470-480. DOI: https://doi.org/10.1016/j.neucom.2013.05.023
15. Adnan, R.M. et al., 2021a. Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization. Knowledge-Based Syst., 230: 19. DOI: 10.1016/j.knosys.2021.107379
16. Adnan, R.M., Parmar, K.S., Heddam, S., Shahid, S., Kisi, O., 2021b. Suspended Sediment Modeling Using a Heuristic Regression Method Hybridized with Kmeans Clustering. Sustainability, 13(9): 21. DOI:10.3390/su13094648
17. Kaveh, K., Bui, M.D., Rutschmann, P., 2017. A comparative study of three different learning algorithms applied to ANFIS for predicting daily suspended sediment concentration. International Journal of Sediment Research, 32(3): 340-350. DOI: 10.1016/j.ijsrc.2017.03.007
18. Esmaeili-Gisavandani, H. et al., 2022. Evaluating ability of three types of discrete wavelet transforms for improving performance of different ML models in estimation of daily-suspended sediment load. Arab. J. Geosci., 15(1): 13. DOI:10.1007/s12517-021-09282-7
19. Tikhamarine, Y., Souag-Gamane, D., Ahmed, A.N., Kisi, O., El-Shafie, A., 2020. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. J. Hydrol., 582: 16. DOI: 10.1016/j.jhydrol.2019.124435